

**UNIVERSIDAD NACIONAL DEL SANTA
ESCUELA DE POSTGRADO**

**DOCTORADO EN INGENIERÍA DE SISTEMAS E
INFORMÁTICA**



TESIS DOCTORAL

**SISTEMA INTEGRAL BAJO EL ENFOQUE DE MINERÍA DE
DATOS Y REDES NEURONALES PARA LA PREDICCIÓN Y
CONTROL DE LA CONTAMINACIÓN ATMOSFÉRICA POR PM10
EN LA CIUDAD DE CHIMBOTE**

AUTOR : Ms. Sixto Díaz Tello

ASESOR : Dr. Carlos Raymundo Ibañez

**CHIMBOTE PERÚ
OCTUBRE, 2014**



**UNIVERSIDAD NACIONAL DEL SANTA
ESCUELA DE POSTGRADO**

**TESIS PARA OPTAR EL GRADO DE
DOCTOR EN INGENIERÍA DE SISTEMAS E INFORMÁTICA**

**“SISTEMA INTEGRAL BAJO EL ENFOQUE DE MINERÍA DE
DATOS Y REDES NEURONALES PARA LA PREDICCIÓN Y
CONTROL DE LA CONTAMINACIÓN ATMOSFÉRICA POR PM10
EN LA CIUDAD DE CHIMBOTE”**

TESISTA: SIXTO DÍAZ TELLO

REVISADO Y APROBADO POR

**DR. CARLOS RAYMUNDO IBAÑEZ
ASESOR**

NUEVO CHIMBOTE, OCTUBRE 2014

“SISTEMA INTEGRAL BAJO EL ENFOQUE DE MINERÍA DE DATOS Y REDES NEURONALES PARA LA PREDICCIÓN Y CONTROL DE LA CONTAMINACIÓN ATMOSFÉRICA POR PM10 EN LA CIUDAD DE CHIMBOTE”

AUTOR : Mg. SIXTO DÍAZ TELLO

ASESOR : DR. CARLOS RAYMUNDO IBAÑEZ

TESIS DE DOCTORADO APROBADO POR LOS SIGUIENTES MIEMBROS.

JURADO EVALUADOR


Dr. JAVIER GAMBOA CRUZADO
PRESIDENTE


Dr. IGOR AGUILAR ALONSO
SECRETARIO


Dra. DIANA CECILIA MUÑOZ CASANOVA
MIEMBRO

Dedicatorias

*A la memoria de mi padre: José María y a mi madre:
Telesvina por su amor y ejemplo.*

*A mi familia: Agustina, Boris, Karen y
Sixto José, por su comprensión y apoyo.*

Agradecimientos

A mis profesores del Doctorado por su valioso aporte en mi formación académica.

A Ingeniero Wilder Leonardo Valverde Silva por su apoyo en la realización de este proyecto de investigación.

A la Universidad Nacional Del Santa por su constante preocupación para mejorar la educación

RESUMEN

La contaminación atmosférica es una amenaza aguda, constituye un fenómeno que tiene particular incidencia sobre la salud del hombre. Los cambios que se producen en la composición química de la atmosfera pueden cambiar el clima, producir lluvia ácida o destruir el ozono, fenómenos todos ellos de una gran importancia global. La Organización Mundial de la Salud (OMS) considerara la contaminación atmosférica como una de las más importantes prioridades mundiales.

Chimbote; ha sido catalogada por la Dirección General de Salud Ambiental (DIGESA) como la tercera ciudad más contaminada en este país. Las principales fuentes de contaminación son fuentes fijas como la industria pesquera y siderúrgica. Los contaminantes que se han registrado como preocupantes son el monóxido de Azufre (SO_2) y principalmente las Partículas Menores a 10 micrómetros (PM_{10}).

En la presente tesis doctoral se propone la implementación de un sistema integral, que permita realizar la predicción de la contaminación atmosférica por concentración del contaminante crítico PM_{10} para Chimbote, haciendo uso de redes neuronales y minería de datos.

El modelo implementado plantea el uso de las variables meteorológicas y variables estacionales; como factores que influyen en la concentración de los contaminantes. La Red Neuronal Artificial utilizada es el Perceptrón Multicapa, con una capa de entrada de 09 neuronas (correspondiente a cada variable meteorológica), 02 capas ocultas y la capa de salida de 05 neuronas (correspondiente al índice de calidad del aire). Las variables meteorológicas disponibles para realizar la predicción fueron: Dirección predominante del viento, Velocidad del viento predominante, Velocidad promedio Velocidad máxima, Temperatura ambiente promedio, Presión promedio, Humedad promedio y las variables estacionales: día y mes del año.

Para la implementación del sistema integral, se tuvo que hacer uso de diversas tecnologías, siendo los sistemas desarrollados los siguientes: **Sistema de Registro y Validación de Data (Desarrollado en Power Builder)**, registra la data meteorológica diaria, valida dicha información, y la procesa para que pueda ser exportada, usada por el sistema de Matlab y por el software Knime. Dicho sistema, adicionalmente sirve para validar las variables estacionales y meteorológicas, en comparación con el PM10, de tal manera que se verifique si dichas variables influyen en el factor contaminante (PM10). **El Sistema de Procesamiento de Red Neuronal (Desarrollado en Matlab)**, realiza el procesamiento de la red, a través de Interfaces gráficas (GUI), en Matlab; el cual importa la data del sistema de Power Builder y realiza la configuración, creación, entrenamiento, validación y simulación (predicción) de la red neuronal modelada para este proyecto. **El Sistema Web de Visualización de Predicciones (Desarrollado en PHP)**, sistema que importa los resultados de Matlab, los procesa e interpreta, de acuerdo a los rangos cualitativos del Índice de Calidad del Aire (ICA). De tal manera que el público en general pueda hacer consultas. **Para el Procesamiento de Minería de Datos (Se hace uso del Software Knime)**, el sistema de Power Builder, exporta la data para realizar el procesamiento de minería de datos a través del Knime. Los resultados obtenidos muestran la importancia de las variables meteorológicas en la predicción de la concentración del contaminante PM₁₀ en la ciudad de Chimbote informando a la población con anticipación del posible riesgo que corre y proteger a los sectores más vulnerables de la exposición a altas concentraciones del contaminante, además de notificar a las autoridades correspondientes para el control de fuentes de emisión.

ABSTRACT

Air pollution is an acute threat is a phenomenon that has particular impact on human health. The changes that occur in the chemical composition of the atmosphere can change the climate, produce acid rain or destroy ozone, all phenomena of great global importance. The World Health Organization (WHO) consider air pollution as one of the world's most important priorities.

Chimbote; It has been classified by the General Directorate of Environmental Health (DIGESA) the third most polluted city in the country. The main pollution sources are stationary sources such as fishing and steel industry. Contaminants that have been recorded as concern are sulfur monoxide (SO₂) and mainly the Lesser Particles 10 microns (PM₁₀).

In this thesis the implementation of a comprehensive system that allows for the prediction of air pollution by PM₁₀ pollutant concentration critical for Chimbote, using neural networks and data mining is proposed.

The implemented model makes use of meteorological variables and seasonal variables; as factors affecting the concentration of pollutants. Artificial Neural Network Multilayer Perceptron is used with an input layer of 09 neurons (Corresponding to each weather variable), 02 hidden layer and output layer of 05 neurons (for the air quality index). The meteorological variables available for the prediction were predominant wind direction, speed of the prevailing wind, high speed average speed, average ambient temperature, average pressure, average humidity and seasonal variables: day and month of the year.

To implement the integrated system, it had to make use of various technologies, systems being developed include: System Registration and Validation Data (developed in Power Builder), records the daily weather data, validate the information, and processes so it can be exported used by the system and the KNIME Matlab software. Said system

further serves to validate seasonal and weather variables, compared to PM10, so as to check whether these variables influence the pollutant factor (PM10). The Processing System Neural Network (Developed in Matlab), performs the processing of the network, through graphical interfaces (GUI) in Matlab; which imports the data system Power Builder and performs configuration, creation, training, validation and simulation (prediction) of the neural network modeled for this project. The Web Predictions Display System (developed in PHP) system that results matter Matlab, processes and interprets according to the qualitative range of Air Quality Index (ICA). So that the general public can make inquiries. Processing for Data Mining (KNIME use Software it done), Power Builder system, export the data processing for data mining through KNIME. The results show the importance of meteorological variables in predicting the contaminant concentration PM10 in the city of Chimbote informing the public in advance of the potential risk to and protect the most vulnerable sectors of exposure to high concentrations of pollutants also notify the relevant authorities for the control of emission sources.

ÍNDICE

Carátula	i
Dedicatoria	ii
Agradecimiento	iii
Resumen	iv
Abstract	vi
Índice	viii
Lista de gráficos	xiii
Lista de Tablas	xviii
CAPÍTULO I: INTRODUCCIÓN	1
1.1. REALIDAD PROBLEMÁTICA	2
1.2. ESTADO DEL ARTE	6
1.3. CARACTERIZACIÓN Y NATURALEZA DEL OBJETO DE INVESTIGACIÓN	9
1.4. FORMULACIÓN DEL PROBLEMA	10
1.5. FORMULACIÓN DE LA HIPOTESIS	10
1.6. OBJETIVOS DE LA INVESTIGACIÓN	10
1.6.1. Objetivo General	10
1.6.2. Objetivos Específicos	10
1.7. IMPORTANCIA Y JUSTIFICACIÓN DE LA INVESTIGACIÓN	11
1.7.1. Importancia	11
1.7.2. Justificación	12
1.8. MOTIVACIÓN	14
CAPÍTULO II: MARCO TEÓRICO	15
2.1. FUNDAMENTOS FILOSÓFICOS	15
2.2. MARCO CONCEPTUAL	18
2.2.1. CONTAMINACIÓN ATMOSFÉRICA	18
2.2.1.1. Definición	18
2.2.1.2. Fuentes de Contaminación Atmosférica	19
2.2.1.3. Clasificación de los Contaminantes Atmosféricos	21
2.2.1.4. Variables que Intervienen en la Contaminación Atmosférica	26
2.2.1.5. Efectos de la Contaminación Atmosférica	28

2.2.1.6.	Monitorización de la Calidad del Aire	28
2.2.1.7.	Estándares y Normas de la Contaminación Atmosférica	31
2.2.2.	PREDICCIÓN DE LA CONTAMINACIÓN ATMOSFÉRICA	34
2.2.2.1.	Introducción	34
2.2.2.2.	Métodos de Predicción	36
2.2.2.3.	Predicción de la Contaminación Atmosférica	40
2.2.2.4.	Modelos de la Contaminación Atmosférica	42
2.2.2.5.	Modelos de Redes Neuronales aplicadas a la contaminación atmosférica	43
2.2.3.	REDES NEURONALES ARTIFICIALES	47
2.2.3.1.	Introducción	47
2.2.3.2.	Neurona Biológica	49
2.2.3.3.	Neurona Artificial	50
2.2.3.4.	Funcionamiento de una RNA	53
2.2.3.5.	Arquitectura de las RNA	53
2.2.3.6.	Proceso de Aprendizaje	56
2.2.3.7.	El Perceptrón	57
2.2.3.8.	El Perceptrón Multicapa	59
2.2.4.	MINERÍA DE DATOS	62
2.2.4.1.	Definición	62
2.2.4.2.	Tipos de minería de datos	63
2.2.4.3.	Proceso de explotación de información	64
2.2.4.4.	Técnicas de minería de datos	66
	 CAPÍTULO III: METODOLOGÍA	 68
3.1.	MÉTODOS EMPLEADOS EN LA INVESTIGACIÓN	68
3.1.1.	Tipo de investigación	68
3.1.2.	Método de investigación	70
3.2.	METODOLOGÍA PARA LA PRUEBA DE HIPOTESIS	71
3.2.1.	Hipótesis	71
3.2.2.	Variables	71
3.2.3.	Indicadores	71
3.2.4.	Diseño	71
3.3.	TÉCNICAS E INSTRUMENTOS EMPLEADOS	72

3.3.1. Técnicas	72
3.3.2. Instrumentos	72
3.4.PROCEDIMIENTO DE LA RECOLECCIÓN DE DATOS	72
3.5.TÉCNICAS DE PROCESAMIENTO Y ANÁLISIS DE DATOS.	73
3.6.PROCEDIMIENTO METODOLÓGICO PARA LA RED NEURONAL	74
CAPITULO IV: ANÁLISIS E INTERPRETACION Y DISCUSIÓN DE RESULTADOS	77
4.1. RESULTADOS DEL PROCESAMIENTO DE RED NEURONAL	78
4.1.1. Análisis de la contaminación atmosférica por partículas en suspensión menores de 10 micras (PM10)	78
4.1.1.1 Identificación de estaciones de monitoreo.	78
4.1.1.2. Identificación de rangos de ICA vs PM10:	83
4.1.1.3. Variables influyentes en la Calidad del Aire	84
4.1.2. Análisis y diseño de la red neuronal perceptrón	85
4.1.2.1. Selección de tecnología	85
4.1.2.2. Identificación de Procesos para la red neuronal	86
4.1.2.3. Identificación de parámetros de red neuronal	87
4.1.3. Implementación del sistema para procesamiento de la red neuronal	90
4.1.3.1. Implementación de la red neuronal a través de la herramienta nntool.	90
4.1.3.2. Implementación de la red neuronal por medio de un sistema GUI en Matlab	95
4.2. RESULTADOS SISTEMA INTEGRAL (RUP)	119
4.2.1. FASE DE INICIO	119
4.2.1.1. Documento visión del negocio	119
4.2.1.2. Plan de desarrollo de software	121
4.2.1.3. Modelado del negocio	124
4.2.1.4. Requerimientos funcionales y no funcionales	129
4.2.1.5. Modelado de casos de uso	130
4.2.2. FASE DE ELABORACIÓN	135
4.2.2.1. Modelado de casos de uso	135

4.2.2.2.	Modelo de Análisis	163
4.2.2.3.	Modelo de Diseño	169
4.2.2.4.	Modelo de Implementación	189
4.2.3.	FASE DE CONSTRUCCIÓN	193
4.2.3.1.	Interfaces del sistema	193
4.2.3.2.	Codificación del sistema	218
4.2.3.3.	Pruebas del sistema	225
4.2.4.	FASE DE TRANSICIÓN	226
4.2.4.1.	Preparación de la versión beta	226
4.2.4.2.	Instalación de la versión beta	227
4.2.4.3.	Reacción a los resultados de prueba	227
4.2.4.4.	Finalización del análisis de negocio	227
4.3.	RESULTADOS PROCESAMIENTO DE MINERÍA DE DATOS	227
4.3.1.	Descripción de la herramienta.	227
4.3.2.	Preparación de data para realización de minería de datos.	229
4.3.3.	Modelos implementados en knime	230
4.3.4.	Procesamiento y resultados de los modelos	233
CAPÍTULO V: EVALUACIÓN		246
5.1.	Demostración de la hipótesis	246
5.2.	Determinación del nivel porcentual	253
5.3.	Discusión de resultados	256
CAPÍTULO VI: CONCLUSIONES Y SUGERENCIAS		261
5.1.	CONCLUSIONES	261
5.2.	SUGERENCIAS	262
5.3.	LÍNEAS DE INVESTIGACIÓN	263
5.4.	TRABAJOS PUBLICADOS	264
REFERENCIA BIBLIOGRÁFICAS		265

ANEXOS	275
ANEXO A: DISEÑO DEL CUESTIONARIO DE PRUEBAS DE SOFTWARE	275
ANEXO B: RANGO DEL ÍNDICE DE CALIDAD DE AIRE	276
ANEXO C: DISEÑO DEL USUARIO PARA MEDIR LA SATISFACCIÓN DEL USUARIO	277
ANEXO D: CARTAS DE VALIDACIÓN POR EXPERTOS	279

ÍNDICE DE GRÁFICOS

2.1. Capas que forman la atmósfera	19
2.2. Fuentes de Contaminación de la Atmósfera	21
2.3. Proceso de mejora continua de la calidad del aire	29
2.4. Métodos de predicción.	38
2.7. Estructura de una neurona biológica típica	50
2.8. Modelo neuronal	50
2.9. Funciones de transferencia	52
2.10. Arquitectura unidireccional con tres capas de neuronas	54
2.11. Taxonomía de las redes neuronales artificiales	55
2.12. Arquitectura de un perceptrón multicapa con una capa oculta	60
2.13. Fases del proceso KDD	63
3.1. Etapas de la metodología utilizada	74
4.1. Mapa del Santa	80
4.2. Mapa de Coishco	81
4.3. Mapa de Chimbote	82
4.4. Mapa de Nuevo Chimbote	83
4.5. Procesos de la red neuronal	87
4.6. Creación de red neuronal	90
4.7. Herramienta nntool	91
4.8. Nueva red neuronal	91
4.9. Entrenamiento de red neuronal	92
4.10. Resultados de entrenamiento	93
4.11. Simulación de red neuronal	94
4.12. Selección de estación	96
4.13. Pantalla principal	96
4.14. Importar data entrada	97
4.15. Resultados de importación	98
4.16. Importación de data salida	98
4.17. Resultados data salida	99
4.18. Gráficas entrada y salida	100
4.19. Configuración de red neuronal	101
4.20. Crear red neuronal	101

4.21. Mensaje satisfactorio	101
4.22. Esquema de red neuronal	102
4.23. Entrenar red neuronal	103
4.24. Ntraintool	104
4.25. Simulación y validación de red neuronal	105
4.26. Resultados de simulación	106
4.27. Gráfica salidas deseadas vs estimadas	107
4.28. Gráfica de funcionamiento	108
4.29. Gráfica ajustes entre salidas	108
4.30. Valores de predicción	109
4.31. Importar data para predicción	110
4.32. Simular y predecir ICA	111
4.33. Resultados simulación y aproximación	113
4.34. Exportación de salida	114
4.35. Mensaje satisfactorio	114
4.36. Gráfica salida obtenida	115
4.37. Procesamiento automático	118
4.38. Mensaje de confirmación	118
4.39. Actores del negocio	125
4.40. Diagrama de paquetes del Negocio	125
4.41. CU: Reglamentación de estándares	127
4.42. CU: Monitoreo de la calidad del Aire	127
4.43. CU: Planes de Acción	128
4.44. CU: Estructura del Modelo de Dominio	128
4.45. Modelo de casos de Uso	133
4.46. Diagrama de casos de Uso	139
4.47. Identificación de paquetes de análisis – Casos de Uso	164
4.48. Identificación de paquetes de análisis – Clases de Dominio	165
4.49. Realización del CU: Iniciar Sesión	166
4.50. Realización del CU: Registrar parámetros ICA	166
4.51. Realización del CU: Registrar información meteorológica diaria	167
4.52. Realización del CU: Validar variables	167
4.53. Realización del CU: Generar archivo input entrada	168
4.54. Realización del CU: Generar archivo input salida	168

4.55. Diseño de la Arquitectura	169
4.56. Diagrama de Secuencia CU: Iniciar Sesión	170
4.57. Diagrama de Secuencia CU: Registrar parámetros de ICA	170
4.58. Diagrama de Secuencia CU: Registrar información meteorológica	171
4.59. Diagrama de Secuencia CU: Variables estacionales y meteorológicas para red	171
4.60. Diagrama de Secuencia CU: Generar archivo Input: Entrada para Matlab	172
4.61. Diagrama de Secuencia CU: Generar archivo Input: Salida para Matlab	172
4.62. Diagrama de Secuencia CU: Exportar archivo Input para Matlab	173
4.63. Diagrama de Secuencia CU: Seleccionar estación de monitoreo	173
4.64. Diagrama de Secuencia CU: Importar y visualizar data de entrada	174
4.65. Diagrama de Secuencia CU: Importar y visualizar data de salida	174
4.66. Diagrama de Secuencia CU: Crear red neuronal	175
4.67. Diagrama de Secuencia CU: Generar Entrenamiento	175
4.68. Diagrama de Secuencia CU: Simular y validar red con data entrenada	176
4.69. Diagrama de Secuencia CU: Importar data de entrada para predicción	176
4.70. Diagrama de Secuencia CU: Simular y predecir ICA	177
4.71. Diagrama de Secuencia CU: Realizar aproximación	177
4.72. Diagrama de Secuencia CU: Procesar resultados Matlab	178
4.73. Diagrama de Secuencia CU: Visualizar monitoreo de predicción de ICA	178
4.74. Diagrama de Colaboración CU: Iniciar Sesión.	179
4.75. Diagrama de Colaboración CU: Registrar parámetros de ICA	179
4.76. Diagrama de Colaboración CU: Registrar información meteorológica	180
4.77. Diagrama de Colaboración CU: variables estacionales y meteorológicas para red.	180
4.78. Diagrama de Colaboración CU: Generar archivo Input: Entrada para Matlab	181
4.79. Diagrama de Colaboración CU: Generar archivo Input: Salida para Matlab	181
4.80. Diagrama de Colaboración CU: Exportar archivo Input para Matlab.	182
4.81. Diagrama de Colaboración CU Seleccionar estación de monitoreo.	182
4.82. Diagrama de Colaboración CU Importar y visualizar data de entrada	183
4.83. Diagrama de Colaboración CU: Importar y visualizar data de salida	183
4.84. Diagrama de Colaboración CU: Crear red neuronal.	184
4.85. Diagrama de Colaboración CU: Generar Entrenamiento	184

4.86. Diagrama de Colaboración CU: Simular y validar red con data entrenada	185
4.87. Diagrama de Colaboración CU: Importar data de entrada para predicción	185
4.88. Diagrama de Colaboración CU: Simular y predecir ICA	186
4.89. Diagrama de Colaboración CU: Realizar aproximación	186
4.90. Diagrama de Colaboración CU: Procesar resultados Matlab	187
4.91. Diagrama de Colaboración CU: Visualizar monitoreo de predicción-ICA	187
4.92. Diagrama de Clases de Diseño	188
4.93. Diagrama de Componentes	189
4.94. Diagrama de despliegue	190
4.95. Resultados de cuestionario	225
4.96. Preparación de data para knime	229
4.97. Árbol de clasificación	230
4.98. Algoritmo de asociación	231
4.99. Clauster k-means	231
4.100. Clauster weka	231
4.101. Gráficos minería de datos	232
4.102. Selección de atributos	233
4.103. Configuración de algoritmos	0 233
4.104. Resultado Algoritmo	234
4.105. Resultado árbol	234
4.106. Algoritmo J48 graft	235
4.107. Configuración algoritmo a priori	235
4.108. Filtro de columnas	235
4.109. Visualización de nodos weka	236
4.110. Configuración de cluster	236
4.111. Etiqueta de entradas	237
4.112. Modelo de cluster	237
4.113. Tabla con colores	238
4.114. Visualización gráficos de dispersión	238
4.115. Configuración cluster weka	239
4.116. Resultados cluster	239
4.117. Procesamiento Box Plot	240
4.118. Gráfico Histograma	240
4.119. Gráfico Line Plot	241

4.120. Gráfico parallel coordinates	242
4.121. Gráfico Pie Chart	242
4.122. Gráfico scatter matrix	243
4.123. Gráfico scatter plot	243
4.124. Procesamiento spark line appender	244
4.125. Selección de atributos	244
4.126. Resultados de procesamiento	245
5.1. Gráfica de Tiempos de procesos antes de aplicar el sistema	249
5.2. Tiempos de procesos después de aplicar el sistema	250
5.3. Gráfica de Tiempos de procesos después de aplicar el sistema	257

ÍNDICE DE TABLAS

1.1. Concentraciones promedio de contaminantes en Chimbote.	4
2.1. Fuentes biogénicas de contaminación atmosférica.	20
2.2. Efectos de la contaminación atmosférica.	28
2.3. Estándares Nacionales de Calidad ambiental del Aire (ECA)	33
2.4. Valores de tránsito	33
2.5. Valores referenciales	34
4.1. Características principales de los distritos del área de estudio	83
4.2. Rango del ICA	84
4.3. Variables estacionales	84
4.4. Variables meteorológicas	85
4.5. Data entrada de red	88
4.6. Estado ICA vs salida de red	89
4.7. Data de salida de la red	90
4.8. Plan de Fases	122
4.9. Tiempo de duración	123
4.10. Cronograma del proyecto	123
4.11. Cronograma de fases del proyecto	123
4.12. Clasificación de Casos de Uso-Inicio	132
4.13. Clasificación de Casos de Uso-Elaboración	137
4.14. Iniciar sesión	140
4.15. Registrar parámetros de ICA	141
4.16. Recolectar data automáticamente	142
4.17. Registrar información meteorológica diaria	143
4.18. Validar variables estacionales y meteorológicas para red	144
4.19. Generar archivo Input: Entrada para Matlab	145
4.20. Generar archivo Input: Salida para Matlab	146
4.21. Exportar archivo Input para Matlab	147
4.22. Generar y exportar archivo Input para predicción en Matlab	148
4.23. Consultar data estacional/ meteorológica	149
4.24. Generar y exportar archivo Input para Knime	150
4.25. Seleccionar estación de monitoreo	151
4.26. Importar y visualizar data de entrada	152

4.27. Importar y visualizar data de salida	153
4.28. Crear red neuronal	154
4.29. Generar entrenamiento	155
4.30. Simular y validar red con data entrenada	156
4.31. Importar data de entrada para predicción	157
4.32. Simular y predecir ICA	158
4.33. Realizar aproximación	159
4.34. Exportar resultado de predicción obtenida	160
4.35. Procesar resultados Matlab	161
4.36. Visualizar monitoreo de predicción de ICA	162
5.1. Resultados de cuestionario	246
5.2. Tiempos de procesos antes de aplicar el sistema	248
5.3. Tiempos de procesos después de aplicar el sistema	249
5.3. Comparación de tiempos	250
5.4. Prueba de comparación de varianzas	259
5.5. Coeficiente de variación y rango	251
5.6. Prueba de comparación de varianzas	252
5.7. Etapa previa – Satisfacción del usuario	253
5.8. Etapa previa – Tiempo	254
5.9. Etapa Posterior – Satisfacción del usuario	254
5.10. Etapa Posterior – Tiempo	255
5.11. Comparación de Ambas etapas	256

CAPÍTULO I: INTRODUCCIÓN

La contaminación atmosférica es actualmente uno de los problemas ambientales más severos a nivel mundial. Está presente en todas las sociedades, independientemente del nivel de desarrollo socioeconómico y constituye un fenómeno que tiene particular incidencia sobre la salud del hombre, además, los cambios que se producen en la composición química de la atmósfera pueden cambiar el clima, producir lluvia ácida o destruir el ozono, fenómenos todos ellos de una gran importancia global.

El crecimiento económico y la urbanización, asociados al desarrollo de diversas actividades como la industria petrolera, la industria pesquera, los servicios, la agroindustria y el incremento de las unidades automotoras, traen como resultado un consumo intenso de combustibles fósiles; al mismo tiempo, la práctica de actividades agropecuarias no apropiadas inciden en la generación de elevados volúmenes de contaminantes, que al relacionarse con las condiciones ambientales pueden dañar la salud humana, los ecosistemas y los recursos materiales.

El transporte y dispersión de contaminantes atmosféricos están influenciados por complejos factores. Las variaciones globales y regionales del clima, las condiciones topográficas locales afectan el transporte y dispersión de los contaminantes. Todos estos factores hacen que el problema de la calidad del aire sea un tema complejo.

Como resultado de las afectaciones de la contaminación atmosférica, ha crecido el interés del público, administraciones y gobiernos por la mejora de la calidad del aire, dedicando recursos a la investigación para conocer mejor el fenómeno de contaminación atmosférica y para entender la atmósfera como un sistema dinámico y complejo. Las redes de monitorización atmosférica juegan un papel fundamental en la mitigación de este problema aportando información valiosa sobre las emisiones de contaminantes. La información obtenida de las redes de monitorización atmosférica ha

permitido la modelización y estudio de los contaminantes atmosféricos en diferentes partes del mundo.

Desde los años 70 y 80 el incremento en la investigación de plataformas computacionales permite de una forma robusta y eficiente el control y previsión de la contaminación atmosférica. Paralelo a este desarrollo computacional se ha producido el desarrollo de modelos y sistemas meteorológicos de dispersión y transformación de contaminantes atmosféricos, sobre todo en la última década.

El tema de la calidad del aire es un área de investigación totalmente inter disciplinario en el que intervienen varias materias como son las Matemáticas, Física, Química y la Informática. De hecho, la Informática actúa como puente de unión entre las disciplinas anteriores, situándonos en el área del conocimiento de Ciencias de la Computación e Inteligencia Artificial.

La presente Tesis Doctoral aborda la implementación de un sistema integral, bajo el enfoque de Minería de Datos y Redes Neuronales, para la predicción y control de la contaminación atmosférica por PM10 en la ciudad de Chimbote, Perú. El modelo propuesto utiliza información de las concentraciones de los contaminantes y las variables meteorológicas obtenidas de la ciudad de Chimbote. El sistema implementado utiliza técnicas de inteligencia artificial y reconocimiento de patrones que permitan utilizar la información obtenida por la concentración de contaminantes en la ciudad de Chimbote y las variables meteorológicas para predecir las concentraciones de los contaminantes que se han reportado como preocupantes en la región.

1.1. REALIDAD PROBLEMÁTICA

La ciudad de Chimbote, que ha crecido desde los años 60 a un ritmo muy acelerado (1956 = 30,000 hab.; 1993 = 265,100 hab.; 2007 = 500,000 hab.), es la tercera ciudad más contaminada del país. Según DIGESA (2012), señala que en la cuenca atmosférica

de Chimbote se emiten en total 23,411 Toneladas/año de contaminantes atmosféricos de los cuales el 53% es generado por fuentes móviles, el 44% por fuentes fijas puntuales del sector industrial y el 3% por emisiones de fuentes fijas de área debido a las actividades económicas de menor escala.

Los principales contaminantes identificados fueron el monóxido de carbono (31%), dióxido de azufre (29%), óxidos de nitrógeno (16%), partículas totales en suspensión (16%) y compuestos orgánicos volátiles 7% de emisiones.

Según, la Dirección General de Salud Ambiental (DIGESA) [DIGESA. 2008], [DIGESA.2012], órgano técnico normativo del Ministerio de Salud del Perú, entre las fuentes puntuales de contaminación atmosférica identificadas en la ciudad de Chimbote, se tienen dos rubros principales: las empresas pesqueras y la siderúrgica. Afectando a una población de 434,646 habitantes ubicados en los distritos del Santa, Coishco, Chimbote y Nuevo Chimbote, según, la Oficina de Gestión de la Información y Estadística de la Dirección General Parlamentaria (DGP) [DGP.2014]

La contaminación atmosférica en Chimbote, es generada por la industria pesquera, se da principalmente por la combustión de Diesel 2, Residual 500 y Bunker 6, y está conformada principalmente por la emisión de gases tales como el hidrógeno sulfurado, el anhídrido sulfuroso, el material particulado y el hollín proveniente de las calderas de vapor. Emitiéndose 1763 Tn./año de Partículas Menores de 10 micras (PM₁₀).

La planta siderúrgica SIDERPERU emite principalmente contaminantes atmosféricos como el dióxido de azufre, el material particulado, así como óxidos de hierro a la atmósfera. Cuyo volumen de emisión de Partículas Menores de 10 micras es de 1117 Tn/año de PM₁₀.

Las concentraciones promedio registradas correspondiente a la ciudad de Chimbote según DIGESA [DIGESA. 2012], se muestran en la tabla 1.1.

Tabla N° 1.1. Concentraciones promedio registradas en Chimbote

Contaminante	PTS (ug/m ³)	PM ₁₀ (ug/m ³)	SO ₂ (ug/m ³)	NO ₂ (ug/m ³)	Pb (ug/m ³)
Concentración	129.22	155.36	127.66	44.48	0.132
Estándar	120	150	124	150	0.5
Fuente	OMS (24h)	EPA (24h)	OMS (24h)	OMS (24h)	Nac. Anual

Fuente DIGESA

Según la tabla 1.1, podemos determinar que el agente contaminante PM10 en Chimbote sobrepasa los límites establecidos por las normas de calidad del aire a nivel nacional e internacional lo cual representa un problema para la salud humana.

El procesamiento de harina de pescado, así como la operación del horno eléctrico de arco en la siderúrgica constituye la principal fuente de emisión de partículas totales y menores a 10 micras. Las emisiones de material particulado de diámetro aerodinámico inferior a 10 micras, de importancia para la salud dada su facilidad para ingresar al sistema respiratorio humano, representan aproximadamente el 30% de las emisiones totales del material particulado en la cuenca atmosférica de Chimbote.

Somos conscientes del grado de contaminación que posee Chimbote, a las enfermedades que nos exponemos diariamente, estos son algunos motivos que nos impulsaron a realizar la presente investigación con la finalidad de encontrar soluciones idóneas a la contaminación ambiental, como prevenir y cómo actuar frente a este problema.

Según el estudio realizado por Consejo nacional del Medioambiente (CONAM) y el Instituto Meteorológico e Hidrológico de Suecia (SMHI) en el 2007, determina que Huancayo y Chimbote son las zonas más contaminadas en el Perú y que el uso masivo de combustible con alto contenido de azufre, como el residual 500, ha causado en Chimbote altos niveles de dióxido de azufre (SO₂), así como las emisiones de las

pesqueras y la siderúrgica local que contaminan el aire, poniendo en grave riesgo a la salud de la población. Si se reducen los niveles de contaminación ambiental que afectan las ciudades de Huancayo (Junín) y Chimbote (Áncash) se evitaría, cada año, la muerte prematura de 22 personas.

Fuentes consultadas de la Dirección Regional de Salud Ambiental-Caleta de Chimbote señalan que en, los últimos años se ha observado incremento de los problemas respiratorios en la ciudad, siendo entre las primeras causas de morbilidad las enfermedades respiratorias, que en el 2005, registró 36 984 episodios en menores de 5 años (HIS).

El estudio realizado en Chimbote en el período del 2002-2003 halló prevalencia 2: Asma acumulada: 14%; Asma diagnosticado por médico: 10,8%; Asma actual: 6,7%; Rinitis : 13,3%,; Faringitis : 18,7%.

A la fecha, no se han encontrado ningún programa de inspección y mantenimiento efectivo para la disminución de la contaminación ambiental por fuentes puntuales en la ciudad de Chimbote.

Para poder gestionar correctamente la calidad del aire en Chimbote, tal como se establece en la legislación dado por el Ministerio del Medio Ambiente (MINAM) [MINAM.2014], resulta necesario disponer de herramientas sistémicas que permitan a las autoridades realizar actuaciones efectivas tendentes a reducir los niveles de contaminación es decir, un sistema de ayuda a la toma de decisiones, mediante cuyo empleo sea posible interpolar datos de las concentraciones de contaminantes obtenidos por las redes de inmisión. Estas herramientas, denominadas modelos de calidad del aire consisten en una serie de algoritmos matemáticos (redes neuronales) que a partir de unos datos de partida (minería de datos) interpretan un proceso físico y químico complejo, que se aproxima a lo que acontece en la realidad. Gracias a estos modelos, se

consigue un entendimiento del fenómeno observado en la naturaleza, la atmósfera, ya que permiten la reproducción matemáticamente del comportamiento de dicho fenómeno.

1.2. ESTADO DEL ARTE

La contaminación atmosférica es un problema fundamental e importante en muchas partes del mundo, en particular en la ciudad de Chimbote, donde existe una alta concentración de población, Así como de unidades industriales y de transporte, factores que originan una alta concentración de emisiones y que además son el resultado de un crecimiento económico e industrial [DIGESA.2012].

La predicción de contaminantes atmosféricos no es una área nueva, sin embargo, es una área con pocos desarrollos. En la literatura científica indexada se encuentran obras desde los años sesenta [Moses.1968], donde se aplican métodos estadísticos y matemáticos. Posteriormente se marca un crecimiento en los años setenta [Sheih-Moroz.1973], [Tokumar.1978], [Carlson.1979], donde dichos trabajos en su mayoría utilizan modelos estadísticos. Continuando de manera cronológica en los años ochenta es todavía mayor el interés por la predicción de contaminantes atmosféricos [Pascual.et.al.1989], [Robeson and Stey.1989], [Williams and Yamada.1990], mismo que ha seguido creciendo hasta la fecha, donde se han hecho presentes diversas técnicas de inteligencia artificial.

En la actualidad, la preocupación por la salud humana y la correcta orientación de las políticas públicas que conducen el desarrollo de estándares para minimizar los efectos de la contaminación del aire en la salud humana, ha llevado a distintos investigadores a desarrollar técnicas de cómputo inteligente, que les permitan generar predicciones confiables sobre el índice de calidad del aire o bien sobre los niveles de concentración

en ciertos contaminantes atmosféricos como el ozono(O₃), el dióxido de nitrógeno(NO), el monóxido de carbono(CO), partículas menores de 10 micras (PM10) entre otros.

Varias técnicas de inteligencia artificial se han aplicado a la predicción de contaminantes atmosféricos, [Kumar and Goyal.2011], [Ozbay,et.al.2011], [Paschlidou,et.al.2011], entre las cuales se encuentran: Los Métodos de Regresión, la Lógica Difusa, la Programación Genética, los Sistemas Expertos, las Redes Neuronales Artificiales, las Máquinas de Soporte Vectorial y el Clasificador Gamma

Los modelos basados en el uso de redes neuronales artificiales (RNAs) son modelos matemático-computacionales de tratamiento de la información, cuyo origen parte de la simulación del cerebro humano. Estos modelos se han utilizado con fines de predicción, de aproximación de funciones y de clasificación de patrones. La ventaja de los modelos basados en la utilización de redes neuronales artificiales reside en su elevada capacidad de generalización y de tratamiento de problemas no lineales. Además, no requieren del conocimiento riguroso de la distribución de la variable de estudio. En el área de la calidad del aire, durante los últimos años se han elaborado diversos modelos de predicción basados en el uso de las redes neuronales artificiales [Agirre.et al.2006].

Estudios comparativos entre las diferentes técnicas de predicción de los niveles de contaminantes atmosféricos han probado la eficiencia de los modelos basados en el uso de las redes neuronales, así lo explican los siguientes trabajos de investigación: “Relevancia de las variables Meteorológicas en el Diseño de un Modelo de Predicción de los niveles de Ozono, en Tiempo Real, basado en el uso de Redes Neuronales” [Agirre.2006]; “Time Series Analysis of Atmosphere Pollution Data Using Artificial Neural Networks Techniques” [Salini-Perez.2006]. Modelización de problemas ambientales en entornos urbanos utilizando sistemas de información geográfica y métodos multivariantes. [Londoño-Linares.2007], “Implementación de un modelo

híbrido para la predicción de niveles de PM10, en Temuco”, Chile. [Díaz y Robles.2008], “Pronóstico de la concentración promedio en Suecia” [Mesin.et.al.2010] “Predicción de la concentración de contaminantes atmosféricos basado en un clasificador asociativo de patrones”, México, [Itzama.2011], y “Aplicación de técnicas de inteligencia artificial a la predicción de contaminantes atmosféricos”, España. [Cortina.2012].

Rivera [Rivera.2012], en su tesis “Modelo de identificación de factores contaminantes atmosféricos”, realizado en la Universidad Nacional Mayor de San Marcos, Lima-Perú; se determinan que los diferentes factores de contaminación que influyen en la calidad del aire, mediante un análisis estadístico de los principales contaminantes para un periodo de 10 años. Luego se establece los criterios de medición de la calidad del aire los relaciona con los diferentes factores de la población y termina con una proyección del 2011 al 2021 en condiciones reales, mediante un modelo dinámico.

Cabrera [Cabrera.2012], en su tesis: “Aplicación de cómputo no convencional en la predicción de contaminantes atmosférico”, realizado en el Instituto politécnico Nacional, México, desarrolla y propone el Modelo de Predicción por Aproximación de Cadenas (MPAC), el cual se aplica en la predicción de contaminantes atmosféricos presentes en la ciudad de México.

De la Paz Cortez [De la Paz.2013], en su tesis: “Evaluación de la contaminación atmosférica en el aire ambiente de Manizales por COPs y PM10”, realizado en la Universidad Nacional de Colombia, realiza un estudio de la calidad del aire en cuatro puntos de la ciudad de Manizales, en la cual se analizó la dinámica de dispersión de las dioxinas (PCDD, PCDFs, PCBs) y el material particulado inferior a 10 micras de diámetro, PM10. Se encontró que el material particulado PM10, se encuentra fuertemente relacionado con la influencia de las emisiones provenientes de la zona

industrial y vehicular. Del análisis estadístico se determinó que la precipitación, la humedad relativa, la dirección y velocidad del viento como las variables meteorológicas influyentes en la dinámica dispersiva de este contaminante.

Sin embargo es importante hacer notar que a pesar de que existen diversas técnicas de inteligencia artificial para la predicción de contaminantes atmosféricos, a la fecha no se ha encontrado trabajo alguno de investigación científica que utilice minería de datos con redes neuronales para la predicción de contaminantes atmosféricos.

1.3. CARACTERIZACION Y NATURALEZA DEL OBJETO DE INVESTIGACION

El objeto del sistema predictivo es facilitar información a los usuarios y autoridades ambientales de la previsión de la calidad del aire, de forma que en una segunda fase se puedan arbitrar medidas – de acuerdo con modelos especialmente diseñados para ello – de reducción de emisiones cuyo objetivo sea el cumplimiento de los límites marcados en las directivas nacionales.

Este sistema de predicción y control, suministrará información sobre la evolución de los niveles de contaminación atmosférica (PM_{10}) y los focos de contaminación en la ciudad de Chimbote. Es desarrollado para utilizarlo en modo de pronóstico, diariamente se simulará a futuro, obteniendo predicciones sobre el estado de la contaminación atmosférica para las próximas 24 horas. Permitiendo actuaciones preventivas.

En esta línea se informará diariamente al ciudadano vía web del índice de predicción de contaminantes (PM_{10}). La calidad del aire se medirá directamente a través de cuatro estaciones localizadas en Santa, Coishco, Chimbote y Nuevo Chimbote.

1.4. FORMULACIÓN DEL PROBLEMA

¿En qué medida la implementación de un Sistema Integral, bajo el enfoque de Minería de Datos y Redes Neuronales, permitirá realizar la Predicción y Control de la Contaminación Atmosférica por PM_{10} , en la ciudad de Chimbote?

1.5. FORMULACIÓN DE LA HIPÓTESIS

La implementación de un Sistema Integral, bajo el enfoque de Minería de Datos y Redes Neuronales, permitirá realizar la predicción y control de la contaminación atmosférica por PM_{10} , en la ciudad de Chimbote, en un 91%, a través de información clara y transparente.

1.6. OBJETIVOS DE LA INVESTIGACIÓN

1.6.1. Objetivo General

Implementar un sistema integral, basado en el uso de minería de datos con redes neuronales (concretamente, en el uso del perceptrón multicapa), para la predicción de los niveles de contaminación atmosférica producido por PM_{10} (procesos industriales: industria pesquera y siderúrgica) en la ciudad de Chimbote, que permita anticiparse a situaciones de riesgo.

1.6.2. Objetivos Específicos

- ✓ Recopilar y Organizar la información y requerimientos necesarios para el desarrollo del trabajo en el área de dominio (Santa, Coishco, Chimbote y Nuevo Chimbote), comprendiendo el funcionamiento de redes neuronales y del Sistema integral.
- ✓ Analizar y procesar la información de la contaminación atmosférica en la ciudad de Chimbote, estableciendo y seleccionando las variables (meteorológicas y estacionales), que serán utilizada para la implementación del modelo de predicción.

- ✓ Implementación de un modelo de red neuronal artificial para la predicción del contaminante atmosférico (PM10).
- ✓ Evaluar y validar el modelo de predicción implementado.
- ✓ Aplicar la Metodología del Proceso Unificado de Rational (RUP), estableciendo el Modelado del Negocio, Modelado de Requisitos, Análisis y Diseño e Implementación del Sistema Integral.
- ✓ Implementar el sistema integral: Sistema de registro y validación de data (Desarrollado en Power Builder), Sistema de procesamiento de red neuronal (Desarrollado en Matlab), Sistema web de visualización de predicciones (Desarrollado en Php) y de Procesamiento de minería de datos (Se hace uso del software Knime).
- ✓ Diseñar las Pruebas de Software, para probar el funcionamiento de los componentes y establecer las condiciones de ejecución, las entradas de la prueba, y los resultados esperados.
- ✓ Obtener y revisar los resultados obtenidos mediante la prueba del Sistema integral.

1.7. IMPORTANCIA Y JUSTIFICACION DE LA INVESTIGACION

1.7.1. Importancia

Este proyecto se fundamenta en la necesidad de comprender la dinámica de contaminación atmosférica y, especialmente, de material particulado PM10, en la ciudad de Chimbote con el fin de poder obtener una precisa estimación de la calidad del aire.

Para ello, se desarrollará, implementará y validará un sistema integral para la predicción control , utilizando el enfoque de minería de datos y redes neuronales, que contribuya a incrementar el conocimiento científico de la dinámica de contaminantes atmosféricos en

Chimbote con un alto grado de detalle o profundización y que contribuya a ser una de herramienta de ayuda en: (1) La selección de los emplazamientos de estaciones de redes de calidad del aire (material particulado menores de 10 micras); (2) la interpretación de los datos de dichas redes; y (3) la toma de decisiones para poder cumplir con las directivas de actuales sobre calidad del aire.

Asimismo, el proyecto contribuirá con un importante aporte científico técnico en el rendimiento de los modelos de calidad del aire en Chimbote. Gracias a este nuevo enfoque se podrán determinar los patrones dinámicos de contaminación por material particulado menores de 10 micras (en un ciclo anual, para observar un comportamiento estacional y distinto en diferentes zonas de la ciudad de Chimbote). Así mismo se podrá:

- **Determinar los niveles** de los principales contaminantes y focos de contaminación en cada zona geográfica de Chimbote y en tiempo real.
- **Denunciar** episodios de contaminación y vigilar la evolución acotando situaciones de riesgo.
- **Proporcionar información** fiable y a tiempo a la población y a los municipios de los diferentes distritos de Chimbote.
- Establecer medidas necesarias para **corregir** situaciones crónicas.
- Establecer y **analizar** el resultado de medidas correctivas coyunturales capaces de evitar situaciones de crisis, así como planes de actuación ante dichas situaciones.
- **Elaborar modelos** que permitan prever la concentración de contaminantes analizando la influencia de factores topográficos, climáticos, meteorológicos (modelos de dispersión de la contaminación atmosférica en tiempo real).
- **Proporcionar datos** para el desarrollo de planes de diseño y planificación urbana.

1.7.2. Justificación

El impacto de la contaminación atmosférica es un tema de vital importancia en el clima y el medio ambiente. La atmósfera es esencial para la vida por lo que sus alteraciones

tienen una gran repercusión en el hombre, otros seres vivos y en general en todo el planeta. Pero además los cambios que se producen en la composición química de la atmósfera pueden cambiar el clima, producir lluvia ácida o destruir la capa de ozono, fenómenos todos ellos de importancia mundial.

Chimbote es una ciudad con grandes industrias pesqueras que dedican en mayor parte sus actividades a la elaboración de harina de pescado, aceites y conservas de pescado. Sus procesos dan por resultado la emisión de contaminantes a la atmósfera que pueden provocar situaciones de riesgo para la salud de la población y el ecosistema en general. Uno de los contaminantes de mayor preocupación en la zona es el PM10, el cual sumado con las condiciones meteorológicas y climatológicas que dificultan la dispersión de contaminantes, hacen indispensable la aplicación de medidas de impacto en la región.

Las PM10 por su radio aerodinámico puede ser absorbidas vía nariz/ boca y pasar por el tracto respiratorio, por lo que afectan a más personas que cualquier otro contaminante. Además es conocido que el potencial de daños a la salud a causa de las partículas suspendidas está directamente relacionado con el tamaño de éstas, ya que penetran con mayor facilidad por el tracto respiratorio llegando a los pulmones, donde permanecen depositadas en los alvéolos pulmonares. Si bien éste es absorbido principalmente por el sistema nasal, la exposición de altas concentraciones por cortos períodos de tiempo puede irritar el tracto respiratorio, causar bronquitis y congestionar los conductos bronquiales de quienes sufren de problemas respiratorios (asma).

Se implementa un sistema integral para predecir la concentración del contaminante PM10, con la finalidad de alertar a la población de posibles riesgos en la salud y advertir a las empresas que contribuyen a las emisiones para que tomen las medidas pertinentes.

Un modelo de predicción de concentración de contaminantes ambientales permitiría tomar decisiones apropiadas en cuanto a la activación de contingencias ambientales y minimizar los riesgos a la salud de la población.

La predicción de los contaminantes en el aire, evitaría posibles contingencias y alertaría a las empresas contaminantes a reducir sus emisiones y esto ayudaría a prevenir a la población con antelación.

1.8. MOTIVACIÓN.

El problema bajo estudio en la presente investigación sobre la contaminación atmosférica en la ciudad de Chimbote es de una actualidad e importancia de alto impacto. Los principales sectores industriales como la industria pesquera y la industria siderúrgica han sido las principales empresas para el progreso de la ciudad, sin embargo son los principales agentes contaminantes de la ciudad de Chimbote.

Hoy en día según los estudios realizados por DIGESA [DIGESA.2012], muestran a la ciudad de Chimbote como la tercera ciudad más contaminada del Perú, debido a altas concentraciones, de PM10, de Monóxido de Azufre (SO₂), Monóxido de Carbono (CO₂) y de Plomo (Pb) y de los efectos que éstos producen en la salud de las personas.

La deficiente calidad del aire que existe en la ciudad de Chimbote revela una planificación y gestión urbana faltas de coherencia y compromiso con los principios de un desarrollo urbano saludable y sostenible.

Esta Tesis Doctoral, se ve motivada por la necesidad de advertir a la población de Chimbote con anticipación del posible riesgo que corre y proteger a los sectores más vulnerables de la exposición de altas concentraciones de contaminantes, además de notificar a las autoridades correspondientes para que controlen las fuentes de emisión.

CAPÍTULO II: MARCO TEÓRICO

2.1. FUNDAMENTOS FILOSÓFICOS

La presente investigación se enmarca dentro de los paradigmas y fundamentos filosóficos de la siguiente manera:

- Postura ante la búsqueda de conocimiento: Pragmatismo
- Corriente Filosófica: Positivismo
- Paradigma de investigación: Funcionalismo
- Método para formular objetivos: Taxonomía de B. Bloom y proceso administrativo
- Posibilidad de conocimiento: Empirismo – Racionalismo

En principio se debe señalar que, según Barreras [Barreras.1999], el pragmatismo alude a la postura epistémica soportada sobre la noción de que el valor del conocimiento radica en la práctica, siendo la actividad lo que permite conocer lo desconocido. El pragmatismo parte del uso de criterios con posibilidad de aplicación práctica. Pensar y conocer son instrumentos de la actividad práctica y es la forma como ha evolucionado el pensar del hombre y su preocupación sobre la verdad. La investigación se orienta por la lógica y la estructura de la realidad. La verdad de los aspectos planteados la decide la praxis.

En cuanto a la corriente filosófica dentro de la cual se enmarca la presente investigación, se debe señalar que se trata del positivismo; el cual se fundamenta en hechos objetivos y verificables; dejando de un lado cualquier tesis metafísica, para contribuir al progreso material y moral de los humanos. El conocimiento sistémico debe usar un lenguaje lógico, formal, riguroso y preciso, retratando la realidad con el

uso de formulaciones simples y elementales. Barreras [Barreras.1999] sostiene que el positivismo como movimiento intelectual se centra en el predominio de la experiencia sobre las ideas y la razón. La comprobación emerge como condición necesaria para determinar la validez de lo conocido o por conocer. Por otra parte, Burk [Burk.2001] plantea que el positivismo es una filosofía clara y definida en cuanto a principios y metas; bajo parámetros realistas y prácticos.

En este sentido, la idea de la investigación no es enunciar la verdad absoluta, sino enunciar el conocimiento de manera racional y científica con el uso de un lenguaje lógico. De esta manera, el progreso científico sistémico dependerá de la actitud crítica ante los eventos acontecidos, estudiados pragmáticamente bajo la razón científica, que permita un acercamiento a la exacta formulación del conocimiento.

En relación con el paradigma de la investigación, se tiene que el funcionalismo, de acuerdo con Barreras [Barreras.1999], concibe la realidad a partir de relaciones, procesos y funciones, contexto en el cual opera el conocimiento. La realidad está determinada por funciones y cada proceso contiene actividades formales propias. Es así como en la presente investigación los tópicos planteados en los objetivos cumplen funciones dentro del contexto social y ambiental.

Para Bloom, el investigador recorre un camino ascendente en la búsqueda del conocimiento, pasando por varios niveles de una menor complejidad a otro cada vez mayor; partiendo de un conocimiento de la realidad, pasando por los respectivos niveles de comprensión, aplicación, análisis, síntesis y evaluación del problema. En este sentido, esta investigación pasa necesariamente por algunas de esas fases para llegar a los resultados planteados.

Según Barreras [Barreras.1999] el empirismo acepta que la experiencia es la única y real fuente del saber, en tanto que el racionalismo reconoce como fundamento del conocimiento a la razón. El racionalismo descansa en el pensar, la inteligencia y la razón. La aplicación de los sistemas inteligentes en la solución de problemas ambientales es el factor que impulsa el proceso investigativo de este estudio y el razonamiento representa la fuente primaria para generar el conocimiento necesario. Por otra parte, a través de la percepción y experiencias vividas en el estudio de campo se podrán formular opiniones sobre los hechos que acontecen en el medio.

A medida que la protección ambiental y la preocupación por el cambio global son considerado como elementos de calidad de vida, se puede esperar una valoración progresiva del medio ambiente. Para las Organización de las Naciones Unidas [ONU.2012] el principal desafío que enfrentan los países es el de aplicar políticas de gestión ambiental para alcanzar un desarrollo sustentable, por ello se debe lograr un crecimiento económico, equidad y sustentabilidad ambiental de manera articulada e incluyentes dentro de los programas de gestión.

Son éstas pues, las bases epistemológicas sobre las cuales se fundamenta la presente investigación, que le dan sustento al carácter científico de la misma de acuerdo con las distintas corrientes filosóficas, las cuales constituyen las bases fundamentales para el desarrollo y el crecimiento de las ciencias.

2.2. MARCO CONCEPTUAL

2.2.1.LA CONTAMINACIÓN ATMOSFÉRICA

La Organización Mundial de la Salud (OMS) considera la contaminación atmosférica como una de las más importantes prioridades mundiales. Los efectos en la salud de la población son variados, dependiendo del tipo de contaminante y tiempos de exposición; siendo los más afectados el sistemas respiratorio y el cardiocirculatorio [ONU. 2012]. Por otro lado, pueden percibirse los efectos de la contaminación atmosférica en áreas alejadas de los focos emisores, tal es el caso de la lluvia ácida, el incremento de los gases de efecto invernadero y el agujero de la capa de ozono.

2.2.1.1. Definición

Se llama contaminante cualquier sustancia que no forme parte de los elementos gaseosos normales del aire. Se entiende por contaminación atmosférica cualquier alteración de la atmósfera terrestre susceptible de causar impacto ambiental por la adición de gases o partículas sólidas o líquidas suspendidas en proporciones distintas a las naturales, que pueda poner en riesgo a personas, animales y plantas, así como atacar a distintos materiales, reducir la visibilidad o producir olores desagradables [Ballester.2005].

La atmósfera terrestre es una capa envolvente gaseosa, compuesta de 78% de nitrógeno, 20% de oxígeno y el resto formado de otros gases en pequeñas cantidades, como vapor de agua, dióxido de carbono, hidrógeno, helio y algunos gases nobles. Tiene unos 2000 km., de espesor, cuya densidad disminuye con la altura. En función de la distribución de temperatura, la atmósfera se divide en: troposfera, estratosfera, mesosfera y termosfera. La figura 2.1., muestra dichas capas. Las propiedades de cada capa guardan relación con la actividad química de sus contaminantes. La troposfera contiene el aire que respiran los seres vivos y en la que se desarrollan todos los procesos meteorológicos.

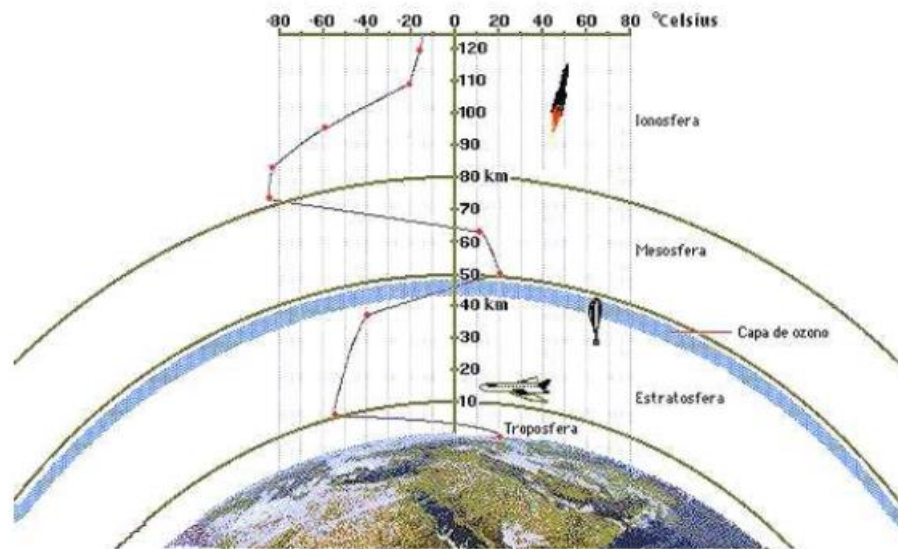


Figura N° 2.1: Capas que forman la atmósfera

2.2.1.2. Fuentes de Contaminación Atmosférica

Los cambios en la calidad del aire pueden suceder con intervención o no, del ser humano, por lo que las fuentes de generación de contaminantes pueden clasificarse como fuentes antropogénicas y fuentes biogénicas.

Las fuentes antropogénicas son originadas por el ser humano, podemos encontrar ejemplos como la combustión causada por el hombre, procesos industriales o cambios de uso de suelos; las fuentes biogénicas son causadas por fenómenos naturales como son: las erupciones volcánicas, los incendios provocados por los rayos, las tormentas de arena, la descomposición de plantas, o las concentraciones de desechos animales, por citar algunos casos. Por su parte, las principales fuentes antropogénicas se pueden clasificar como fuentes móviles y fuentes fijas. La figura 2.2., muestra la clasificación de las fuentes de contaminación atmosférica y lo que las origina [Jiménez, 2006].

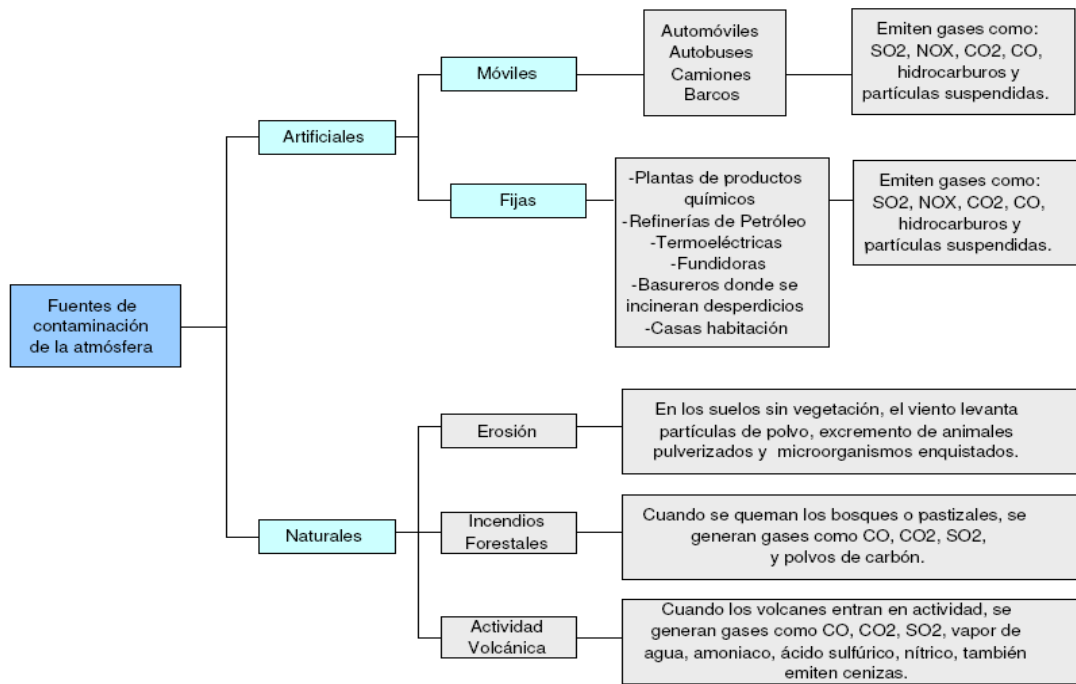


Figura N° 2.2: Fuentes de contaminación de la atmósfera.

Fuentes	Contaminantes
Volcanes	Óxidos de azufre, Partículas
Fuegos Forestales	Monóxido de carbono, dióxido de carbono Óxidos de nitrógeno, partículas
Vendavales	Polvo
Plantas(vivas)	Hidrocarburos, polen
Plantas(en descomposición)	Metano, sulfuro de hidrógeno
Suelo	Polvo
Mar	Partículas de sal

Tabla N° 2.1: Fuentes Biogénicas de contaminación atmosférica.

2.2.1.3. Clasificación de los Contaminantes Atmosféricos

Los principales contaminantes de la atmósfera se clasifican en:

- ◆ **Primarios.** Son los que permanecen en la atmósfera tal y como fueron emitidos por la fuente. Para fines de evaluación de la calidad del aire se consideran: óxidos de azufre, monóxido de carbono, óxido de nitrógeno, hidrocarburos y partículas.
- ◆ **Secundarios.** Son los que han estado sujetos a cambios químicos, o bien, son el producto de la reacción de dos o más contaminantes primarios en la atmósfera. Entre ellos destacan los oxidantes fotoquímicos y algunos radicales de corta existencia como el ozono.

Tomando en cuenta el impacto que tienen los contaminantes atmosféricos sobre la salud de los seres humanos, la Agencia de Protección Ambiental de los Estados Unidos (EPA–Environmental Protection Agency de sus siglas en inglés), clasificó los contaminantes atmosféricos como: contaminantes criterio y contaminantes no criterio.

Los contaminantes criterio se han identificado como perjudiciales para la salud y el bienestar de los seres humanos. Se les llamó contaminantes criterio porque fueron objeto de evaluaciones publicadas en documentos de calidad del aire en los Estados Unidos [Cortina.2012], donde se establecieron los niveles permisibles que protegieran la salud, el medio ambiente y el bienestar de la población.

2.2.1.3.1. Contaminantes Criterio

Los contaminantes criterios son:

- ✓ Ozono (O₃).
- ✓ Material Particulado (PM).
- ✓ Bióxido de Nitrógeno (NO₂).
- ✓ Monóxido de Carbono (CO).
- ✓ Bióxido de Azufre (SO₂).

✓ Plomo (Pb).

Para cada contaminante criterio se han desarrollado guías y normas. Las guías son recomendaciones que establecen los niveles de exposición a contaminantes atmosféricos, a fin de reducir los riesgos o proteger de los efectos nocivos. Las normas establecen las concentraciones máximas de los contaminantes atmosféricos que se permiten durante un período definido. Estos valores límite son diseñados con un margen de protección ante los riesgos y tienen la finalidad de proteger la salud humana y el medio ambiente.

Ozono (O₃)

El ozono es un compuesto gaseoso natural de la atmósfera que se produce en presencia de luz solar. El ozono situado en la estratosfera, protege la vida en la tierra al absorber los rayos ultravioleta del sol. El ozono es un gas incoloro, cuyo olor se detecta a niveles muy bajos. Formado principalmente por reacciones atmosféricas de hidrocarburos y óxidos de nitrógeno bajo la influencia de la luz solar. De acuerdo a estudios del cambio climático, el ozono constituye el tercer gas efecto invernadero más importante después del dióxido de carbono y el metano.

Por su naturaleza reactiva principalmente a la radiación solar, es un contaminante generado esencialmente durante el día y con la presencia de corrientes de aire. El ozono es uno de los principales contaminantes atmosféricos presentes en zonas altamente industrializadas y en las ciudades con un número alto de automóviles.

Material Particulado (PM)

Son una mezcla de partículas sólidas y líquidas (vapor) que se encuentran en el aire emitidas directamente desde la fuente (formación directa) o bien son formadas por la condensación de contaminantes gaseosos (formación indirecta).

Forman una compleja mezcla de partículas sólidas y vapores con sustancias orgánicas e inorgánicas en suspensión, incluyendo frecuentemente sustancias ácidas como sulfatos y nitratos.

Todos estos compuestos en el aire tienen la capacidad de ser absorbidos vía nariz o boca y pasar por el tracto respiratorio, por lo que afectan a más personas que cualquier otro contaminante. Además es conocido que el potencial de daños a la salud a causa de las partículas suspendidas está directamente relacionado con el tamaño de éstas. Entre más pequeño sea su diámetro aerodinámico, más dañinas están consideradas, ya que penetran con mayor facilidad por el tracto respiratorio llegando a los pulmones, donde permanecen depositadas en los alvéolos pulmonares.

Clasificación del Material Particulado

En primera instancia, las partículas suspendidas son reconocidas en función de su diámetro aerodinámico.

- ◆ PM10 (diámetro $< 10 \mu\text{g}/\text{m}^3$): Partículas gruesas. Permanecen menos tiempo en la atmósfera y pueden depositarse cerca de la fuente de emisión debido a su peso, por lo que no viajan lejos de su fuente de emisión.
- ◆ PM2.5 (diámetro $< 2,5 \mu\text{g}/\text{m}^3$): Partículas finas. Ayudadas por corrientes de aire, pueden desplazarse grandes distancias y alejarse de la fuente de emisión; por su diámetro tan pequeño, tienen mayor facilidad para permanecer suspendidas por tiempos mucho más largos que las partículas gruesas y mayor probabilidad de ser absorbidas por el ser humano durante la respiración.

Bióxido de Nitrógeno (NO₂)

Los óxidos de nitrógeno (NO_x) generados por el hombre se producen por la oxidación de nitrógeno atmosférico presente en los procesos de combustión a altas temperaturas.

El contaminante generado en forma primaria es el monóxido de nitrógeno (NO), parte

del cual rápidamente se oxida y se convierte a dióxido de nitrógeno (NO_2). Ambos óxidos, liberados a la atmósfera, participan activamente en un conjunto de reacciones fotoquímicas que en presencia de hidrocarburos reactivos, generan ozono (O_3). Además, en su proceso de transformación, este contaminante forma nitratos, es decir, sales que pueden ser transportadas en el material particulado respirable y que en presencia de humedad, forman ácidos. Estos ácidos son una parte importante del material particulado secundario ($\text{PM}_{2.5}$), que tiene efectos nocivos en la salud tal y como se ha especificado en párrafos anteriores. Es imprescindible el control de esta sustancia gaseosa, ya que contribuye tanto a la formación de ozono como a la de material particulado fino. La principal fuente antropogénica de óxidos de nitrógeno, conocidos como NO_x que incluye óxido nítrico (NO) y dióxido de nitrógeno (NO_2), es el transporte automotor y cualquier otro tipo de combustión de hidrocarburos.

Monóxido de Carbono (CO)

El monóxido de carbono es un gas inodoro, incoloro, inflamable y altamente tóxico que en concentraciones altas puede ser letal, pues impide el transporte del oxígeno a la sangre, lo que puede ocasionar una reducción significativa en la dotación de oxígeno al corazón, siendo particularmente peligroso para las personas con problemas de salud del corazón y vías respiratorias.

El monóxido de carbono se forma en la naturaleza mediante la oxidación del metano (CH_4), que es un gas común producido por la descomposición de la materia orgánica. La principal fuente antropogénica de monóxido de carbono es la quema incompleta de combustibles como la gasolina por falta de oxígeno.

Una manera de reducir el CO en la atmósfera, es que los automóviles sean afinados debidamente para asegurar la mezcla del combustible con el oxígeno.

Bióxido de Azufre (SO₂)

Los óxidos de azufre son una fuente importante de contaminantes atmosféricos y son producidos por los gases de los escapes de los automóviles, las refinerías de petróleo, los procesos de fabricación de papel y las industrias químicas.

El SO₂ es un gas incoloro y con sabor ácido picante, es percibido por el olfato en concentraciones hasta de 3 ppm (partículas por millón). Cuando se encuentra a Niveles de 1 a 10 ppm induce al aumento de frecuencia respiratoria y pulso.

El bióxido de azufre es un gas irritante y tóxico. Afecta sobre todo las mucosidades y los pulmones provocando ataques de tos. Si bien éste es absorbido principalmente por el sistema nasal, la exposición de altas concentraciones por cortos períodos de tiempo puede irritar el tracto respiratorio, causar bronquitis y congestionar los conductos bronquiales de quienes sufren de problemas respiratorios (asma).

El SO₂ tiene relación directa como componente de la lluvia ácida, cuyos efectos son nocivos tanto en las grandes ciudades, acelerando la corrosión de edificios, monumentos y reduciendo significativamente la visibilidad; como en el campo, produciendo acidez de lagos, ríos y suelos. En la naturaleza, el bióxido de azufre se encuentra en las proximidades de los volcanes y las erupciones pueden liberar cantidades importantes de este compuesto químico.

Plomo (Pb)

El Plomo existe de forma natural en el ambiente generalmente en cantidades muy pequeñas, pero las concentraciones significativas son el resultado de las actividades humanas. Por ejemplo, la utilización del plomo en gasolinas. Las partículas grandes se precipitan al suelo o en la superficie de las aguas, mientras que las partículas pequeñas viajarán largas distancias a través del aire y permanecerán más tiempo en la atmósfera. De esta manera, este contaminante también puede terminar en el agua y suelos. Otras

actividades humanas, como la combustión del petróleo, procesos industriales, corrosión de pinturas que contienen plomo, combustión de residuos sólidos, que también contribuyen. El plomo es un elemento químico particularmente peligroso, y se puede acumular en organismos individuales, pero también entrar en las cadenas alimenticias causando graves problemas de salud.

2.2.1.4. Variables que Intervienen en la Contaminación Atmosférica

El transporte y dispersión de contaminantes atmosféricos están influenciados por complejos factores. Las variaciones globales y regionales del clima, las condiciones topográficas locales afectan el transporte y dispersión de los contaminantes [EPA.2011].

La temperatura ambiental, la humedad, la pluviosidad, la velocidad y dirección del viento, la estabilidad atmosférica, las altas y bajas presiones, sumado a la topografía (valles, montañas) condiciona la concentración final de los contaminantes atmosféricos.

La dispersión de contaminantes de una fuente depende de la cantidad de turbulencia en la atmósfera cercana. La turbulencia puede ser creada por el movimiento horizontal y vertical de la atmósfera. El movimiento horizontal es lo que comúnmente se llama viento. La velocidad del viento puede afectar en gran medida la concentración de contaminantes en un área. Mientras mayor sea la velocidad del viento, menor será la concentración de contaminantes. El viento diluye y dispersa rápidamente los contaminantes en el área circundante.

Desde 1930 una serie de episodios de contaminación aguda en diferentes partes del mundo demostraron los riesgos para la salud, con aumento de la morbimortalidad, cuando se dan determinadas condiciones atmosféricas en lugares con altas concentraciones de contaminantes atmosféricos (EPA, 2011). La combinación de niebla intensa, situación anticiclónica y altas concentraciones de contaminantes en los meses fríos del año, fueron los factores comunes en los episodios del Valle del Mosa (Bélgica,

diciembre 1930), Donora (Pensilvania, EE.UU., octubre 1948), Londres (diciembre, 1952), Nueva York (noviembre, 1953). Todos ellos son casos típicos de lo que se ha denominado Wintersmog (EPA, 2011).

Las concentraciones de los contaminantes atmosféricos dependen de su producción y también, de manera determinante, de su dispersión. En invierno, la suma de calefacciones, tráfico rodado, industrias, etc. provoca un aumento de la producción de SO₂ y partículas. En situación de inestabilidad atmosférica el viento y la lluvia contribuyen a limpiar la atmósfera y a reducir su concentración. Sin embargo, en situación anticiclónica, es decir, con estancamiento de la atmósfera, al aumento de emisión de estos contaminantes primarios hay que añadir la persistencia de los mismos en el ambiente y por consiguiente el aumento de su concentración, y con ello el incremento de efectos perjudiciales para la salud.

En verano, con muchas horas de sol y temperaturas altas, el tipo de contaminación que predomina se denomina Summersmog y se refiere a la contaminación procedente de las reacciones de los hidrocarburos y los óxidos de nitrógeno, estimuladas por la luz solar intensa. El ozono es considerado como el elemento esencial de este tipo de contaminación.

Otros factores meteorológicos básicos que afectan la concentración de contaminantes en el aire ambiental son:

- ✓ Radiación solar. La radiación solar contribuye a la formación de ozono y contaminantes secundarios en el aire.
- ✓ Precipitación. La precipitación puede tener un efecto beneficioso porque lava las partículas contaminantes del aire y ayuda a minimizar las partículas provenientes de actividades como la construcción y algunos procesos industriales.

- ✓ Humedad. La humedad y la precipitación también pueden favorecer la aparición de contaminantes secundarios peligrosos, tales como las sustancias responsables de la lluvia ácida.

2.2.1.5. Efectos de la Contaminación Atmosférica

Los efectos producidos por la contaminación atmosférica dependen principalmente de la concentración de contaminantes, del tipo de contaminantes presentes, del tiempo de exposición y de las fluctuaciones temporales en las concentraciones de contaminantes [Rodríguez y Domínguez.2011]. En la tabla 2.2., se resumen algunos de los diferentes efectos de la contaminación atmosférica.

Efectos	
Salud Humana	Mal Olor, enfermedades crónicas, incapacidad laboral, bronquitis, asma, malestar general
Plantas	Alteración de diversos mecanismos vitales, parálisis en el crecimiento.
Materiales	Corrosión de aceros al carbono, compuestos del cobre, níquel y aluminio Daños en la estructura mecánica de mármol, la caliza y la argamasa Envejecimiento prematuro y agrietamiento
Visibilidad	Reducción de la visibilidad, pérdida de luz solar, accidentes
Globales	Ecosistemas (lluvias ácidas) Clima (efecto invernadero) Estratosfera (disminución de la concentración de ozono en la estratosfera.)

Tabla N° 2.2: Efectos de la contaminación atmosférica.

2.2.1.6. Monitorización de la Calidad del Aire

En las últimas décadas se ha desarrollado una preocupación creciente por los efectos de la contaminación del aire en la salud humana y el ambiente. La fuerte presión social junto con una legislación más rigurosa, requieren de información imparcial y confiable sobre la calidad del aire. La monitorización es una herramienta para identificar y evaluar

los problemas de la calidad del aire [SENAMHI.2013]. La monitorización ambiental junto con los modelos de predicción y los inventarios de emisiones, son parte integral de la gestión de la calidad del aire. Esta se encuentra directamente vinculada con el cumplimiento de objetivos económicos, de evaluación y de regulación.

El propósito más importante de la monitorización atmosférica es el de proveer la información necesaria a los científicos, legisladores y planificadores para que éstos tomen las decisiones adecuadas para propósitos de gestión y mejora del ambiente. La monitorización juega un papel regulador en este proceso, proporcionando la base científica para el desarrollo de las políticas y estrategias, en el establecimiento de objetivos durante la evaluación del cumplimiento de las metas y en la ejecución de las acciones. En la figura 2.3., se presenta un esquema del papel de la monitorización atmosférica dentro del ciclo de la gestión de la calidad del aire.

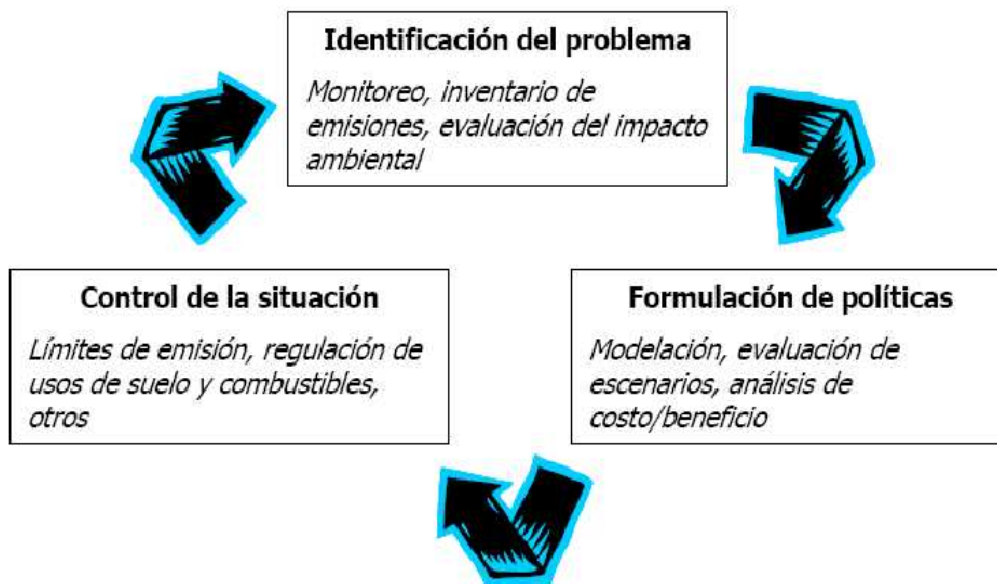


Figura N° 2.3: Proceso de mejora continua de la calidad del aire.

2.2.1.6.1. Estaciones de Monitorización

Para lograr la gestión y mejora del ambiente se han establecido estaciones de monitorización de la calidad del aire, en diferentes partes del mundo. Una estación de monitorización consiste en una caseta que contiene diversos instrumentos destinados a monitorizar las concentraciones de uno o más contaminantes del aire y, por lo general, algunos parámetros meteorológicos. La información (series de tiempo) obtenida permite evaluar respecto a las normas establecidas y en casos pertinentes tomar acción para la disminución de emisiones de contaminantes a la atmósfera.

Los criterios de localización de las estaciones son diferentes según los objetivos de monitorización que hayan sido establecidos. Es común que las estaciones de monitorización se encuentren agrupadas en redes que pueden cubrir grandes extensiones geográficas, formando así redes de monitorización que sirven como fuente fundamental de información sobre el estado actual del entorno. Algunos países que cuentan con redes de monitorización de la calidad del aire son: EE.UU., Canadá, México, Chile, Reino Unido, Francia, España, entre otros.

Series de Tiempo de Contaminantes Ambientales

Una Serie de Tiempo (ST) es un conjunto de observaciones ordenadas en el tiempo, las cuales pueden pertenecer a la evolución de un fenómeno o variable a lo largo de tiempo. Así, una serie de tiempo S se puede definir como un conjunto de valores reales consecutivos, $S = \{x_i \in \mathfrak{R} | x_1, x_2, \dots, x_n\}$. El objetivo del análisis de una serie temporal, de la que se dispone de datos de períodos regulares de tiempo, es el conocimiento de su patrón de comportamiento para prever la evolución futura. Si al conocer la evolución de la serie en el pasado se pudiese predecir su comportamiento [Manahan.2007].

Las series de tiempo de contaminantes ambientales consisten en complejos patrones lineales y no lineales que son difíciles de pronosticar. Estas ST, generalmente son

utilizadas para el análisis de tendencias, evaluación de emisiones de contaminantes a la atmósfera, evaluación de los daños en la población y en el hábitat, así como la predicción.

2.2.1.6.2. Objetivos de la Monitorización de la Calidad del Aire

En términos generales la monitorización de la calidad del aire se realiza con cualquiera de los siguientes objetivos:

- ✓ Determinar el grado de cumplimiento con las normas de calidad del aire.
- ✓ Activar los procedimientos de emergencia para prevenir o mitigar los episodios de contaminación.
- ✓ Diseñar estrategias de control a mediano y largo plazo.
- ✓ Observar las tendencias de la contaminación en una región.
- ✓ Proveer una base de datos para hacer la gestión de la calidad del aire y con fines de investigación, con relación al transporte, el tránsito vehicular y la planeación del uso de suelo.
- ✓ Disminuir e identificar fuentes de emisión.
- ✓ Desarrollo y evaluación de estrategias de control y reducción de la contaminación.
- ✓ Desarrollar y validar herramientas de gestión (modelos de simulación, sistemas de información geográfica, etc.).
- ✓ Evaluar tendencias para identificar los problemas futuros o la evolución de las actividades de acuerdo con los objetivos de gestión y control.

2.2.1.7. Estándares y Normas Vigentes con la Contaminación Atmosférica.

Según el Ministerio del Medio Ambiente, de Perú [MINAM.2014], los criterios para la evaluación de la calidad del aire están dados por dos tipos de instrumentos legales. Los Estándares de Calidad Ambiental (ECA) proporcionan los criterios de calidad que se aplican al aire ambiental en su condición de cuerpo receptor de emisiones de

contaminantes atmosféricos (i.e. emisiones gaseosas o de material particulado). Los Límites Máximos Permisibles (LMP), proporcionan los criterios de calidad exigidos para las fuentes puntuales de emisión de contaminantes atmosféricos. Los ECA y los LMP están definidos por las siguientes normas:

- Reglamento de Estándares Nacionales de Calidad Ambiental del Aire – DS 074-2001-PCM.
- Valor Anual de Concentración de Plomo – DS 069-2003-PCM.
- Niveles Máximos Permisibles de elementos y compuestos presentes en emisiones gaseosas provenientes de las unidades minero-metalúrgicas - RM 315-96-EM/VMM (19/07/96).

Los Estándares de Calidad Ambiental del Aire (ECA), se definen como aquellos "niveles de concentración máxima de contaminantes del aire que en su condición de cuerpo receptor es recomendable no exceder para evitar riesgo a la salud humana. Como estos estándares protegen la salud, son considerados estándares primarios. [MINAM.2014]

El valor correspondiente para la concentración del plomo fue establecido por DS 069-2003-PCM (15/06/03). En la Tabla 2.3 se muestran los parámetros regulados.

Los ECA son "referencia obligatoria en el diseño y aplicación de las políticas ambientales y de las políticas, planes y programas públicos en general. Ninguna autoridad judicial o administrativa podrá hacer uso de los estándares nacionales de calidad ambiental del aire, con el objeto de sancionar bajo forma alguna a personas jurídicas o naturales".

Tabla 2.3. Estándares Nacionales de Calidad ambiental del Aire (ECA)

(Todos los valores son concentraciones en microgramos por metro cúbico. NE significa no exceder)

Contaminantes	Periodo	Valor	Forma del estándar Formato	Método de análisis
Dióxido de Azufre	Anual	80	Media aritmética anual	Fluorescencia UV (método automático)
	24 horas	365	NE más de 1 vez al año	
PM-10	Anual	50	Media aritmética anual	Separación inercial/filtración (Gravimetría)
	24 horas	150	NE más de 3 veces/año	
Monóxido de Carbono	8 horas	10000	Promedio móvil	Infrarrojo no dispersivo (NDIR) (Método automático)
	1 hora	30000	NE más de 1 vez/año	
Dióxido de Nitrógeno	Anual	100	Promedio aritmético anual	Quimiluminiscencia (Método automático)
	1 hora	200	NE más de 24 veces/año	
Ozono	8 horas	120	NE más de 24 veces/año	Fotometría UV (Método automático)
Plomo	Anual ²			Método para PM10 (Espectrofotometría de absorción atómica)
	Mensual	1.5	NE más de 4 veces/año	
Sulfuro de Hidrógeno	24 horas ²			Fluorescencia UV (Método automático)

¹ O método equivalente aprobado² A determinarse según lo establecido en el Artículo 5 del presente reglamento

Fuente: MINAM

Tabla: 2.4. Valores de Tránsito

Contaminantes	Periodo	Valor	Forma del estándar Formato	Método de análisis
Dióxido de Azufre	Anual	100	Media aritmética anual	Fluorescencia UV (método automático)
PM-10	Anual	80	Media aritmética anual	Separación inercial/filtración (Gravimetría)
	24 horas	200	NE más de 3 veces/año	
Dióxido de Nitrógeno	8 horas	250	NE más de 24 veces/año	Quimiluminiscencia (Método automático)
Ozono	Anual	160	NE más de 24 veces/año	Fotometría UV (Método automático)

Fuente: MINAM

Tabla 2.5. Valores Referenciales

Contaminantes	Periodo	Forma del estándar	Método de análisis
		Valor	
PM-10	Anual	80	Separación inercial/filtración (Gravimetría)
	24 horas	65	

Fuente: MINAM

Zonas de Atención Prioritaria

- | | |
|-------------|--------------------|
| 1. Arequipa | 8. La Oroya |
| 2. Chiclayo | 9. Lima-Callao |
| 3. Chimbote | 10. Pisco |
| 4. Cusco | 11. Piura |
| 5. Huancayo | 12. Trujillo |
| 6. Ilo | 13. Cerro de Pasco |
| 7. Iquitos | |

2.2.2. PREDICCIÓN DE LA CONTAMINACIÓN ATMOSFÉRICA

2.2.2.1. Introducción

El aire que respiramos cambia constantemente. La contaminación atmosférica es un problema regional, debido a que la atmósfera se encuentra en continuo movimiento los contaminantes son transportados hasta lugares situados a cientos o miles de kilómetros del punto de origen.

El desarrollo urbano, la industrialización, la modificación de la superficie de la tierra y el posible cambio climático, fenómenos derivados de una explosión demográfica mundial, están alterando la composición de nuestro aire. Esos cambios pueden afectar drásticamente al clima y como consecuencia deteriorar nuestra salud y la de nuestros ecosistemas.

La atmósfera deposita contaminantes en los cursos de agua y de la tierra, lo que perjudica no solamente a las personas sino también a los animales y las plantas del ecosistema. Algunos de esos problemas ambientales son la erosión, la acidificación de

los océanos, lagos, ríos, bosques y la acumulación de componentes tóxicos en las plantas y en la fauna salvaje [Bernstein.2004].

Los daños causados a la salud y al ecosistema debido a altas concentraciones de contaminantes originaron el interés de los gobiernos, los cuales han establecido normas y directrices para cada contaminante atmosférico.

Las redes de monitorización atmosférica juegan un papel fundamental en la mitigación de este problema aportando información valiosa sobre las emisiones de contaminantes.

La información obtenida de las redes de monitorización atmosférica ha permitido la modelización y estudio de los contaminantes atmosféricos en diferentes partes del mundo. Los modelos matemáticos son herramientas que tratan de reproducir matemáticamente los procesos observados en la naturaleza con la mayor fidelidad posible. En el caso de los modelos de calidad del aire, los investigadores usan técnicas matemáticas para simular los procesos físicos y químicos que afecta a los contaminantes en su dispersión y transformación en la atmósfera. Estos modelos tienen aplicaciones no solamente científicas, sino que son útiles en la gestión de la calidad del aire. Una de esas aplicaciones son el control y la predicción en tiempo real de la contaminación atmosférica en una ciudad o región, en una zona industrial, etc. La aplicación de un modelo, es una tarea compleja que demanda la integración sistemática de una gran cantidad de información generada por diferentes fuentes, tales como: inventario de emisiones, meteorología, topografía, usos de suelo, etc.

Predecir de manera efectiva un fenómeno de contaminantes atmosféricos es una potente herramienta de análisis, que tiene un amplio espectro de aplicación que va desde la evaluación de estrategias de control de emisiones y su impacto en la calidad del aire, soporte en la toma de decisiones de carácter ambiental, hasta la generación de

información científica valiosa para comprender mejor la dinámica de la atmósfera y la contaminación en un lugar o región determinada.

En esta sección se mencionan brevemente los métodos de predicción en general y su clasificación. En particular se citan los métodos usados para la predicción de la contaminación atmosférica en diferentes partes del mundo.

2.2.2.2. Métodos de Predicción

La predicción es una actividad esencial cuya importancia ha crecido en las últimas décadas, convirtiéndose en imprescindible en la gran mayoría de los procesos de toma de decisiones. La predicción en términos generales es conocer por anticipado algún suceso, basándose en el análisis y consideraciones necesarias dependiendo del campo de aplicación.

El avance de la tecnología ha permitido recopilar información de diferentes fenómenos a lo largo del tiempo, a esta información se le llama serie de tiempo. Una serie tiempo o cronológica es una secuencia de datos, observaciones o valores, medidos en determinados momentos del tiempo, espaciados entre sí de manera uniforme y por tanto ordenados cronológicamente. Uno de los usos más habituales de las series de datos temporales es su análisis para predicción y pronóstico.

El análisis de series temporales comprende métodos que ayudan a interpretar este tipo de datos, extrayendo información representativa, tanto referente a los orígenes o relaciones subyacentes como a la posibilidad de extrapolar y predecir su comportamiento futuro. Planificadores y decisores disponen de una gran variedad de técnicas para realizar sus predicciones, que van desde las más subjetivas e intuitivas hasta métodos cuantitativos más complejos.

La elección de un método apropiado de predicción depende del fenómeno o situación que se quiera predecir. Los científicos han diseñado diferentes modelos para explicar estructuras y prever la evolución de las variables observadas a lo largo del tiempo.

Son innumerables las aplicaciones que se pueden citar en distintas áreas del conocimiento, dependiendo del tipo de variables que se utilicen, por ejemplo: variables económicas (índice de precios al consumo, demanda de electricidad, series de exportaciones o importaciones), variables físicas, (estado del tiempo [Ellwood,et.al,2005], velocidad del viento en una central eólica, temperatura en un proceso, caudal de un río, concentración en la atmósfera de un agente contaminante), variables químicas (determinación de la cantidad de azúcares en soluciones) o variables sociales (número de nacimientos, matrimonios, defunciones, o votos a un partido político, operaciones militares [Hamilton, 2009]), etc.

2.2.2.2.1. Clasificación de los Métodos de Predicción

Los métodos de predicción se clasifican ampliamente en cualitativos y cuantitativos. Los primeros son subjetivos e intuitivos, la posición central en estos métodos no la tiene los datos pasados, sino la experiencia, el conocimiento y el buen juicio de los expertos, son útiles cuando los datos son escasos. Generalmente estos pronósticos no pueden ser reproducidos por alguien más, puesto que el pronosticador no indica claramente como fue incorporada la información disponible en el pronóstico. Según la metodología seguida, los métodos cualitativos pueden ser:

- ✓ **Exploratorios o Prospectivos:** Cuando se parte de las experiencias pasadas o presentes para proyectar al futuro, sopesando las diferentes posibilidades.
- ✓ **Normativos o Deductivos:** cuando se procede al revés, parte de las metas u objetivos a lograr en el futuro y se analiza qué se necesita para lograrlos y eso da una pauta para pensar cuando sucederán los eventos previstos para el futuro.

Los pronósticos basados en modelos matemáticos o estadísticos son los cuantitativos; una vez que se señala el modelo y la técnica se puede determinar automáticamente los pronósticos correspondientes para ser reproducidos cuando se necesiten. La información utilizada consiste en datos históricos, que se encuentran en forma numérica y corresponden a los registros de las propias empresas o información oficial de diversos orígenes. Los métodos cuantitativos son de dos tipos según la información en que se basen:

- ✓ **Métodos de series de tiempo:** cuando se usa información de la misma variable que se va a pronosticar, buscando encontrar el patrón del pasado y proyectarlo al futuro.
- ✓ **Métodos Causales:** cuando se utiliza información de las variables que se van a pronosticar y de las otras variables que influyen en ella o que se relacionan y cuyo pronóstico es más simple. Se puede apreciar en la figura 2.4.

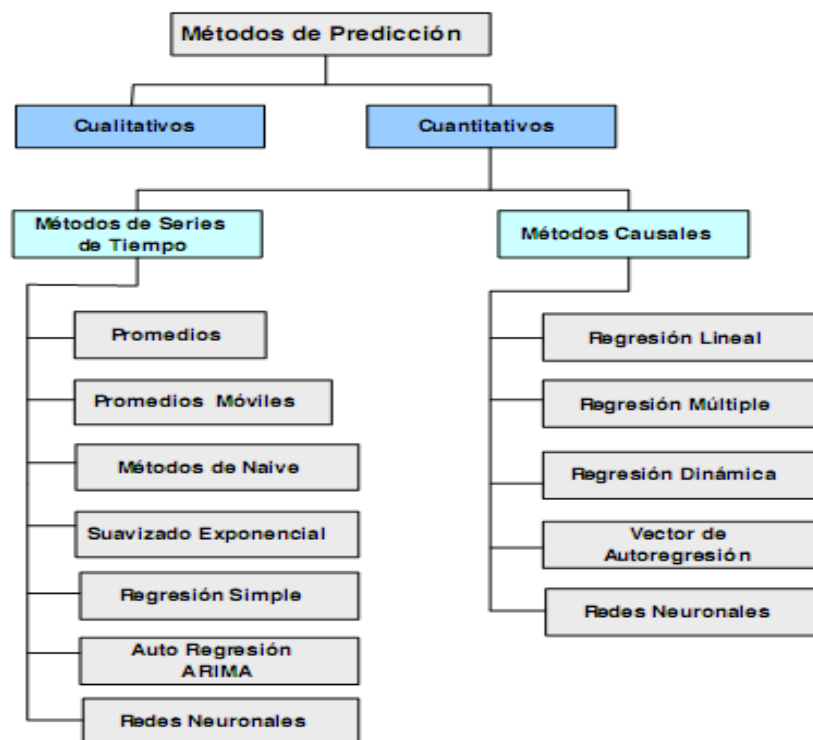


Figura N° 2.4: Métodos de predicción.

2.2.2.2.2. Evaluación de la Predicción

Un aspecto importante cuando se mide el desempeño de un método de predicción es tener una métrica adecuada que permita comparar el desempeño de diferentes métodos. A continuación se mencionan algunas de las métricas más comunes.

- ✓ **Error Cuadrático Medio** (MSE-Mean Square Error, por sus siglas en inglés), mide las diferencias entre el valor real y el estimado, si el MSE es un valor pequeño, las previsiones serán más exactas que si el resultado es grande.

$$MSE = \frac{1}{N} \sum_{i=1}^N (Y_p - Y_o)^2$$

- ✓ **Raíz del Error Cuadrático Medio** (RMSE-Root Mean Square Error, por sus siglas en inglés), que proporciona una idea global de la diferencia entre los valores predichos y reales.

$$RMSE = \left[\frac{1}{N} \sum_{i=1}^N (Y_p - Y_o)^2 \right]^{1/2}$$

- ✓ **Error Cuadrático Medio Normalizado** (NMSE - Normalized Mean Square Error, por sus siglas en inglés), es una versión del MSE, pero normalizada para establecer la relación entre diferentes modelos.

$$NMSE = \frac{\mu(Y_o - Y_p)^2}{\mu(Y_o)\mu(Y_p)}$$

- ✓ **Media del Error Absoluto** (MAE - Mean Absolute Error, por sus siglas en inglés), cuantifica los errores residuales.

$$MAE = \frac{1}{N} \sum_{i=1}^N |Y_p - Y_o|$$

- ✓ **Coefficiente de Correlación**, provee una descripción global del modelo.

$$R = \frac{\mu((Y_o - \mu(Y_o))(Y_p - \mu(Y_p)))}{\sigma_{Y_o}\sigma_{Y_p}}$$

- ✓ Varianza Geométrica, cuantifica la varianza geométrica.

$$VG = e^{(\mu(\ln(Y_o)) - \ln(Y_p))}$$

- ✓ Factor 2 (FAC2), cuantifica el porcentaje de casos predichos en los cuales los valores del radio Y_o/Y_p está en el rango $[0,5, 2]$.

$$0,5 \leq Y_o/Y_p \leq 2$$

- ✓ Varianza Fraccional (FV - Fractional Variance, por sus siglas en inglés), medida normalizada para comparar la diferencia entre la varianza predicha y la varianza observada. Un modelo con $FV = 0$ es un modelo cuya varianza es igual a la varianza de los valores observados.

$$FV = 2 \frac{\sigma_{Y_o} - \sigma_{Y_p}}{\sigma_{Y_o} + \sigma_{Y_p}}$$

- ✓ índice de Acuerdo (Index of Agreement), indica la congruencia entre los datos predichos y los datos observados, es una medida de evaluación descriptiva (Willmott.1982).

$$d = 1 - \frac{\sum_{i=1}^n (Y_p - Y_o)^2}{\sum_{i=1}^n (|Y_p - \mu(Y_p)| + |Y_o - \mu(Y_o)|)^2}$$

Donde Y representa los valores observados, Y_p representa los valores pronosticados, μ es la media aritmética.

2.2.2.3. PREDICCIÓN DE LA CONTAMINACIÓN ATMOSFÉRICA

2.2.2.3.1. Introducción.

El estudio de las formas contaminantes del medio ambiente que afectan el hábitat terrestre, así como la aplicación de las técnicas y procedimientos para su monitorización, prevención y remediación han sido objeto de estudio en los últimos

años. Países como España, Italia, Francia, Grecia, Japón, China, Chile, Argentina, Brasil, India, Turquía entre otros, han instalado redes de monitorización atmosférica para medir la cantidad de contaminantes que se emiten a la atmósfera.

Numerosos estudios científicos vinculan la contaminación del aire con enfermedades respiratorias y cardiovasculares, el cáncer, trastornos del sistema nervioso, así como enfermedades transmitidas por el aire e inducidas por el calor [Salazar y Álvarez. 2013]. A medida que los científicos recopilan más datos sobre el aire, se han percatado de lo estrecho que es el vínculo entre la calidad del aire y el sistema tiempo-clima. El viento, la lluvia, la luz solar y las variaciones de la temperatura controlan, conjuntamente, el transporte y la duración de los agentes contaminantes en todo el mundo. Por esta razón los científicos han hecho esfuerzos por predecir la distribución de partículas y gases atmosféricos que son dañinos para la salud y el medio ambiente.

2.2.2.3.2. Objetivos de la Predicción de Contaminantes Atmosféricos

El modelado matemático es una herramienta imprescindible en el estudio de la contaminación atmosférica. Los modelos de predicción de contaminantes pueden variar de acuerdo a la región de estudio y tipo de contaminante atmosférico e información disponible, sin embargo todos persiguen alguno de los siguientes objetivos:

- ✓ Realizar notificaciones a favor de la salud pública.
- ✓ Planeación y gestión de políticas ambientales.
- ✓ Planificar programas para tomar el control en episodios críticos.
- ✓ Predecir los cambios en las concentraciones ambientales.
- ✓ Detectar cambios en los patrones de emisión.

Partiendo de estos objetivos, se establecen las características deseadas en un modelo de predicción de contaminantes atmosféricos y lo que se considera como características deseadas de predicción de cualquier otro contaminante.

- **Límites mínimos de error.** Los modelos actualmente existentes, manejan índices de error suficientemente grandes, que dan resultados poco fiables.
- **Predicciones de corto, mediano y largo plazo.** Dependiendo de las necesidades específicas, el usuario podría necesitar los resultados de la predicción como parte de sus planes, en intervalos de tiempo de horas, días o semanas.
- **Predicciones multizonales.** Con ello se busca que el área geográfica que abarque la predicción sea amplia y suficientemente eficiente, para poder ser aplicada a toda el área metropolitana de una ciudad, o entre ciudades cercanas.
- **Predicciones de valores máximos del contaminante.** Este tipo de predicción sirve de ayuda para mantener el control de los niveles más altos de contaminantes a vigilar y poder tomar acciones en torno a ello. Uno de los principales fallos de los modelos de predicción actuales es que los valores más altos de las concentraciones no son predichos correctamente.

2.2.2.4. MODELOS DE LA CONTAMINACIÓN ATMOSFÉRICA

Los científicos han hecho esfuerzos por modelar y pronosticar la contaminación atmosférica. Diversos modelos de predicción han sido desarrollados con el propósito de obtener las concentraciones de los contaminantes atmosféricos. Una primera diferencia debe hacerse en cuanto a los dos tipos básicos de modelos existentes: modelos matemáticos y modelos físicos [Barrón. 2012].

Los modelos físicos son representaciones a pequeña escala de la dispersión atmosférica; por ejemplo, réplica de una determinada área mediante túneles de viento. Los modelos matemáticos son aquellos que se basan en determinados sistemas conceptuales basados en principios físico-químicos, e implementados en programas de ordenador.

Entre los modelos matemáticos se pueden distinguir dos grandes grupos:

- ✓ **Modelos Determinísticos**, basados en las ecuaciones matemáticas (principalmente, la ecuación de advección difusión) que representan los procesos atmosféricos. Estos modelos no requieren una gran cantidad de datos, pero exigen un conocimiento de las fuentes de contaminación, las emisiones, la composición química de los gases y los procesos físicos de la atmósfera [Barron.2012]. Entre ellos tenemos a los modelos de Caja, modelo Gaussiano, modelo Euleriano, modelo Lagrangiano,
- ✓ **Modelos empíricos**, basados en relaciones estadísticas y empíricas entre los datos relativos a la contaminación y otras variables que puedan tener influencia sobre ella. Estos modelos por lo general requieren más datos que los modelos determinísticos, limitándose generalmente a una zona y condiciones específicas. Así, en el caso de la predicción de los contaminantes atmosféricos se han desarrollado modelos de predicción basados en el análisis de series temporales o la regresión lineal múltiple, modelos basados en Inteligencia Artificial, etc.

2.2.2.5. MODELOS DE REDES NEURONALES APLICADAS A LA CONTAMINACIÓN ATMOSFÉRICA

A lo largo de los años estos modelos diferentes de RNA se han utilizado con fines de predicción, de aproximación de funciones y de clasificación de patrones.

Las RNA han sido utilizadas por distintos autores para predicción de contaminantes atmosféricos (gases y Material Particulado). Los modelos propuestos predicen concentraciones promedio de contaminantes entre una y 48 horas en avanzada. Algunos de estos modelos usan los datos de contaminación pasada, más información meteorológica, información de tráfico, día de la semana, etc. Entre los últimos aportes podemos citar a: Díaz-Robles [Díaz-Robles, et.al.2008], implementa un modelo híbrido

para la predicción de niveles de PM_{10} en Temuco, Chile. Los modelos Box-Jenkins y una MLP son combinados para mejorar la precisión de los pronósticos, este modelo híbrido es comparado con el modelo de regresión lineal múltiple, Box-Jenkins y una MLP. Utiliza la concentración diaria y horaria de PM_{10} y las variables meteorológicas de los años 2000 al 2006, calculando adicionalmente los valores máximos horarios de PM_{10} del día anterior y los promedios móviles de 6, 12 y 24 horas de las concentraciones de PM_{10} . En el modelo híbrido la predicción se realiza con el modelo Box-Jenkins y la RNA calcula el residuo del modelo. Los resultados obtenidos en este trabajo muestran un coeficiente de correlación de 0,77 para la MLP y 0,76 para el modelo Box-Jenkins.

Salazar-Ruíz, [Salazar-Ruíz, et.al.2008], realiza la previsión del día siguiente de O_3 de las ciudades de Mexicali, Baja California (México) y Calexico, California (US), utilizando datos meteorológicos y variables atmosféricas de los años 1999-2004. Para la implementación de los modelos construyen dos matrices de datos, ambas con información de un día anterior. La primera matriz contiene el valor medio de O_3 , T, NO_2 , NO, NO_x , VV, DV, CO, PB, RS, y la concentración máxima de O_3 registrada un día antes. La segunda matriz de datos es formada por los valores de las primeras 6 horas del día de las variables O_3 , T, NO_2 , NO, CO, V, RH, y la concentración máxima de O_3 registrada un día antes. Los modelos implementados fueron persistencia, regresión lineal, regresión Ridge, una MLP, máquinas de vector soporte, RNA recurrente (ELMAN), probando 290 estructuras de RNAs. Los mejores resultados se obtienen con la primera matriz, para todos los casos la MLP muestra mejor desempeño, seguida de la RNA recurrente.

Mesin, [Mesin.et.al.2010], hace el pronóstico de la concentración promedio de PM_{10} para Suecia. Utiliza la información del criterio de Información Mutua Promedio (PMI-

the Partial Mutual Information por sus siglas en inglés) para eliminar los datos no relevantes y redundantes. El pronóstico se hace mediante la implementación de una MLP con una capa oculta de 8 neuronas, utilizando como entrada de la RNA variables del promedio diario, valores máximos y mínimos de tres días de meteorológicas (T, HR, PB, RS, lluvia, VV, DV) y la concentración de contaminantes (SO₂, NO_x, N O₂, O₃, y CO). Los resultados obtenidos por los autores muestran un coeficiente de correlación de 0,91.

Sfetsos, [Sfetsos, et.al.2010], predice la concentración promedio diario de P M₁₀ en Atenas, los autores plantean una metodología para utilizar períodos de tiempo más cortos de datos, capaces de captar mejor el comportamiento subyacente de las series diarias. Para dicha metodología utilizan datos de cinco estaciones de monitorización recogidos en un período de 5 años. Implementa tres métodos: regresión lineal, Vecino más cercano (NN - Nearest Neighbors por su nombre en inglés) y una RNA de propagación hacia adelante, la metodología es implementada para datos univariantes y multivariantes. Las variables utilizadas fueron, las componentes de DV, HR, VV y los promedios diarios de PM₁₀. Los autores concluyen que los modelos multivariantes que consideran las variables meteorológicas dan mejores resultados que los modelos univariantes.

Barrón [Barrón.2010], en sus tesis doctoral: "Modelado de un sistema de supervisión de la calidad del aire usando técnicas de fusión de sensores y redes neuronales", realizado en la Universidad Politécnica de Madrid. España, propone un modelo de supervisión inteligente que combina las técnicas de Fusión de Datos de Sensores y Redes Neuronales Artificiales y contribuye a determinar en forma automática el Índice de Calidad Ambiental en la ciudad de Salamanca.

Itzama [Itzama.2011], en el artículo: “Predicción de la concentración de contaminantes atmosféricos basado en un clasificador asociativo de patrones”, publicado en la Revista de la facultad de Ingeniería de la Universidad de Antioquía de México, determina la utilidad del clasificador Gamma en la predicción de valores futuros de la concentración de contaminantes utilizados para el cálculo del Índice Metropolitano de la Calidad del Aire de la Ciudad de México (IMECA), específicamente en la predicción de valores horarios del contaminante criterio Monóxido de Carbono.

Cortina [Cortina.2012], en su Tesis Doctoral: “Aplicación de técnicas de inteligencia artificial a la predicción de contaminantes atmosféricos”, realizado en la Universidad Politécnica de Madrid. España, aplica técnicas de inteligencia artificial a la aplicación de contaminantes atmosféricos, propone un modelo de predicción de concentración de los contaminantes más críticos SO₂ y PM₁₀ para cada caseta de monitorización de la Red de Monitorización Atmosférica de Salamanca, México (REDMAS). El modelo propuesto es el Perceptrón Multicapa, hacen uso de las variables meteorológicas como factores que influyen en la concentración de los contaminantes. Estos modelos pueden ser implementados en un sistema que permita realizar la predicción en tiempo real y analizar el impacto en la salud de la población.

Rivera [Rivera.2012], en su tesis “Modelo de identificación de factores contaminantes atmosféricos”, realizado en la Universidad Nacional Mayor de San Marcos, Lima-Perú; se determinan que los diferentes factores de contaminación que influyen en la calidad del aire, mediante un análisis estadístico de los principales contaminantes para un periodo de 10 años. Luego se establece los criterios de medición de la calidad del aire los relaciona con los diferentes factores de la población y termina con una proyección del 2011 al 2021 en condiciones reales, mediante un modelo dinámico.

Cabrera [Cabrera.2012], en su tesis: “Aplicación de cómputo no convencional en la predicción de contaminantes atmosférico”, realizado en el Instituto politécnico Nacional, México, desarrolla y propone el Modelo de Predicción por Aproximación de Cadenas (MPAC), el cual se aplica en la predicción de contaminantes atmosféricos presentes en la ciudad de México.

De la Paz Cortez [De la Paz.2013], en su tesis: “Evaluación de la contaminación atmosférica en el aire ambiente d Manizales por COPs y PM10”, realizado en la Universidad Nacional de Colombia, realiza un estudio de la calidad del aire en cuatro puntos de la ciudad de Manizales, en la cual se analizó la dinámica de dispersión de las dioxinas (PCDD, PCDFs, PCBs) y el material particulado inferior a 10 micras de diámetro, PM10. Se encontró que el material particulado PM10, se encuentra fuertemente relacionado con la influencia de las emisiones provenientes de la zona industrial y vehicular. Del análisis estadístico se determinó que la precipitación, la humedad relativa, la dirección y velocidad del viento como las variables meteorológicas influyentes en la dinámica dispersiva de este contaminante.

Molina [Molina.2014], en su tesis doctoral: “Simulación numérica de un problema de contaminación atmosférica”, realizado en la Universidad Politécnica de Madrid. España, presenta la implementación de un método conservativo para la resolución de la parte advectiva de la ecuación en derivadas parciales que modela la dispersión de contaminantes dentro del modelo meso escalar de transporte químico

2.2.2. REDES NEURONALES ARTIFICIALES

2.2.3.1. Introducción

La Inteligencia Artificial (IA) es un área del conocimiento compuesta por un conjunto de técnicas que se basan en imitar computacionalmente las distintas habilidades relacionadas con la inteligencia del ser humano, como por ejemplo: reconocimiento de

patrones, diagnóstico, clasificación, entre otros. Una de estas técnicas imita, específicamente el comportamiento de las neuronas en el cerebro humano, por lo cual se le han denominado Redes Neuronales Artificiales (RNA) [Barrón. 2010]

Las RNA son modelos matemático-computacionales que poseen una elevada capacidad de generalización y de tratamiento de problemas tanto lineales como no lineales, no requiere un conocimiento exhaustivo de la distribución de las variables en estudio.

Las RNA tratan de emular el sistema nervioso, de forma que son capaces de reproducir algunas de las principales tareas que desarrolla el cerebro humano, al reflejar las características fundamentales de comportamiento del mismo. Lo que realmente intenta modelizar las redes neuronales, es una de las estructuras fisiológicas de soporte del cerebro, la neurona y los grupos estructurados e interconectados de varias de ellas, conocidos como redes de neuronas. De este modo, construyen sistemas que presentan un cierto grado de inteligencia

Una de las principales propiedades de este sistema es la capacidad de aprender y generalizar a partir de ejemplos reales. Es decir, la red aprende a reconocer la relación que existe entre el conjunto de entradas proporcionadas como ejemplos y sus correspondientes salidas, de modo que, finalizado el aprendizaje, cuando a la red se le presenta una nueva entrada, en base a la relación funcional establecida en el mismo, es capaz de generalizarla ofreciendo una salida.

Actualmente las RNA se emplean en diferentes campos y problemas muy diversos, como:

- ✓ Modelación financiera y económica
- ✓ Aplicaciones médicas
- ✓ Minería de datos
- ✓ Optimización de procesos industriales y control de calidad

A lo largo de este capítulo se expondrán los fundamentos básicos de las RNA, describiendo la estructura de la neurona artificial, se mencionarán las diferentes arquitecturas y la clasificación de las RNA. En particular se introducirá el Perceptrón y el Perceptrón Multicapa, este último será utilizado en la metodología. Como punto final en este capítulo se menciona como han sido utilizadas por diferentes autores las RNA en los problemas ambientales, en particular la predicción de contaminantes atmosféricos.

2.2.3.2. Neurona Biológica

La historia de las redes neuronales artificiales comenzó con el científico aragonés Santiago Ramón y Cajal [Salini.2009], descubridor de la estructura neuronal del sistema nervioso. Ramón y Cajal, demostró que el sistema nervioso está compuesto por una red de células individuales, las neuronas, ampliamente conectadas entre sí. Estableció que la información fluye en la neurona desde las dendritas hacia el axón, atravesando el soma. Este descubrimiento fue básico en el desarrollo de la neurociencia en el siglo XX.

Se estima que el sistema nervioso contiene alrededor de cien mil millones de neuronas. Las neuronas constituyen procesadores de información sencillos y están formadas por tres componentes principales: dendritas, el cuerpo celular y el axón. Ver figura 2.7.

Las dendritas son como un árbol de redes de fibras nerviosas que llevan señales eléctricas al cuerpo de la célula (canal de entrada de información). El cuerpo celular, contiene el núcleo que es donde se ejecutan todas las transformaciones necesarias para la vida de la neurona (órgano de cómputo). El axón transmite la señal de salida a otras neuronas (canal de salida), el intercambio químico de información entre una neurona y otra se hace a través de la sinapsis, la cual es el punto de interconexión entre neuronas.

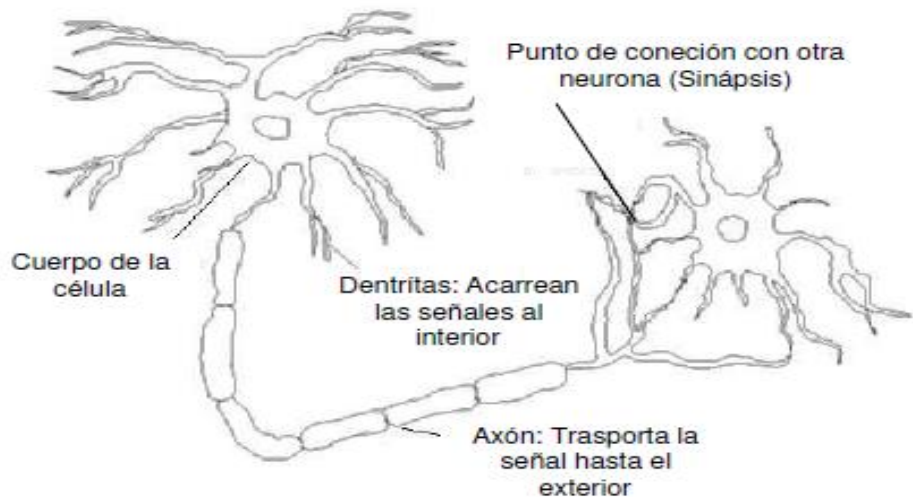


Figura N° 2.7: Estructura de una neurona biológica típica (Fuente: Google Image)

2.2.3.3. Neurona Artificial

Las RNA actuales se basan en el modelo matemático de neurona, propuesto por McCulloch y Pitts en 1943. Se denomina neurona a un modelo simple de cálculo que a partir de un vector de entrada $\{x_1, x_2, \dots, x_n\}$ que procede del exterior o de otras neuronas, que proporciona una única respuesta o salida y dentro de una RNA existen numerosas conexiones entre las distintas neuronas que la forman. Estas conexiones simulan las conexiones neuronales del cerebro y al igual que éstas pueden establecerse con mayor o menor intensidad. En el caso de las RNA estas intensidades son llamadas pesos sinápticos, de modo que cada entrada x_i de una neurona se encuentra afectada por un peso w_i . Ver figura 2.8.

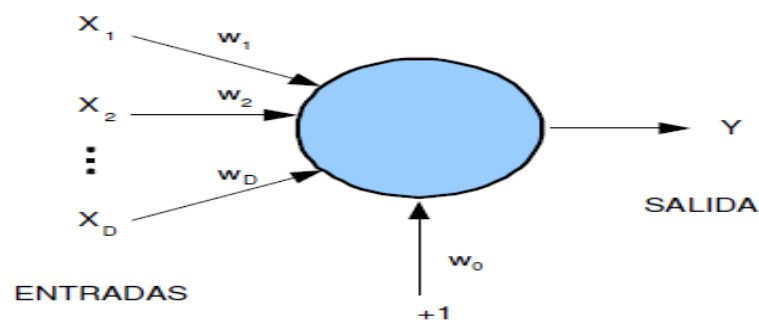


Figura N° 2.8: Modelo neuronal

El primer paso para obtener la salida y de una neurona es calcular la suma ponderada a de las entradas, llamada activación de la neurona:

$$a = \sum_{i=1}^D w_i x_i + w_0$$

dónde w_0 es un umbral o sesgo que se utiliza para compensar la diferencia entre el valor medio de las entradas, sobre el conjunto de entrenamiento, y el correspondiente valor medio de las salidas deseadas [Haykin,1994]. La función de activación o transferencia $g(a)$ permite obtener la salida mediante la ecuación:

$$y = g(a) = g\left(\sum_{i=1}^D w_i x_i + w_0\right) = g\left(\sum_{i=0}^D w_i x_i\right)$$

dónde w_0 es un umbral como un peso más si se supone una entrada añadida x_0 con un valor fijo de 1

La función de transferencia usada en el modelo básico de McCulloch-Pitts es la función escalón, la cual es una función discontinua cuyo valor es 0 para un argumento negativo y 1 para cualquier argumento positivo, está definida por la ecuación:

$$g(a) = \begin{cases} 0 & \text{cuando } a < 0 \\ 1 & \text{cuando } a > 0 \end{cases}$$

En los modelos actuales se escogen otro tipo de funciones, normalmente monótonas y derivables. Las funciones más comunes son:

- ✓ La lineal o función identidad:

$$g(a) = a$$

- ✓ La sigmoidea:

$$g(a) = \frac{1}{1 + e^{-a}}$$

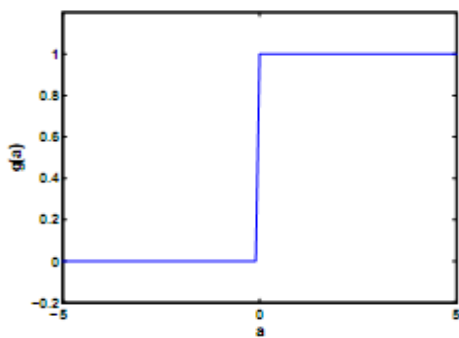
✓ La tangente hiperbólica:

$$g(a) = \tanh = \frac{e^a - e^{-a}}{e^a + e^{-a}}$$

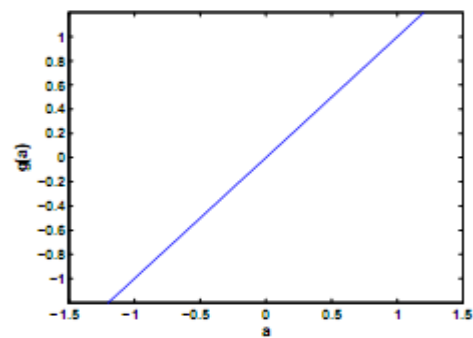
✓ La gaussiana:

$$g(a) = \exp\left(-\frac{(a - \mu)^2}{2\sigma^2}\right)$$

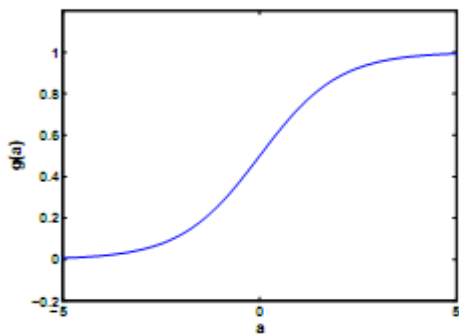
Estas ecuaciones se ven representadas en la figura 2.10.



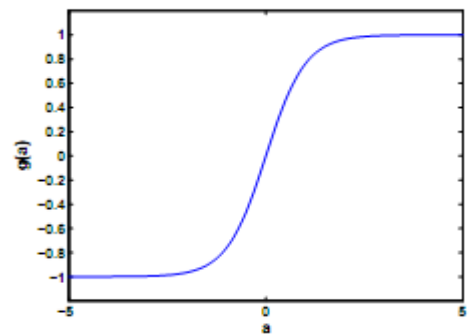
(a) Escalón



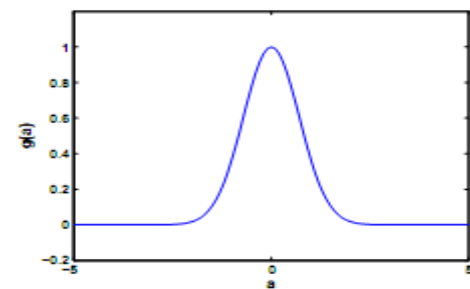
(b) Lineal



(c) Sigmoidea



(d) Tangencial hiperbólica



(e) Gaussiana

Figura N° 2.10: Funciones de transferencia

2.2.3.4. Funcionamiento de una RNA

Para la creación y aplicación de una red neuronal a un problema concreto, se distinguen los siguientes pasos:

- ✓ *Conceptualización del modelo para el estudio del problema concreto.* Se deben señalar las entradas, salidas y la información de que se dispone.
- ✓ *Adecuación de la información de que se dispone a la estructura de la red a crear.* Es decir, construir los patrones de aprendizaje, donde parte de la información va a ser utilizada como validación de la red.
- ✓ *Fase de aprendizaje.* Se le van presentando a la red los patrones adecuados y la red proporciona salidas, este proceso se repite un cierto número de etapas, estas salidas se comparan con las salidas esperadas y los diversos algoritmos de aprendizaje de la red y los patrones de validación.
- ✓ *Fase de validación.* Se presenta a la red entrenada el conjunto de patrones de validación, y se ve el error cometido por la red en este conjunto, este error es una medida de la bondad de la red.
- ✓ *Fase de generalización.* Cuando se ha conseguido la red adecuada se utiliza la red como modelo predictor, aportándole una nueva entrada, la procesará y dará una salida.

2.2.3.5. Arquitectura de las RNA

El número de neuronas que componen una RNA y la forma en que están conectas entre sí se conoce como la topología o arquitectura de la red. En general las neuronas suelen agruparse en unidades estructurales que se denominan capas. Las conexiones entre las neuronas llevan asociadas un peso, w , que determina cuantitativamente el efecto que produce cada neurona sobre las otras neuronas, de este modo en los pesos se almacenan la información de la red.

Se distinguen tres tipos de capas: entrada, salida y ocultas. La capa de entrada, está compuesta por neuronas que reciben datos o señales del entorno. Una capa de salida es aquella cuyas neuronas proporcionan la respuesta de la red neuronal. Una capa oculta es aquella que no tiene conexión directa con el entorno. Este tipo de capa proporciona a la red neuronal grados de libertad adicionales, gracias a los cuales es capaz de representar mejor determinadas características del entorno que trata de modelar. Ver figura 2.10.

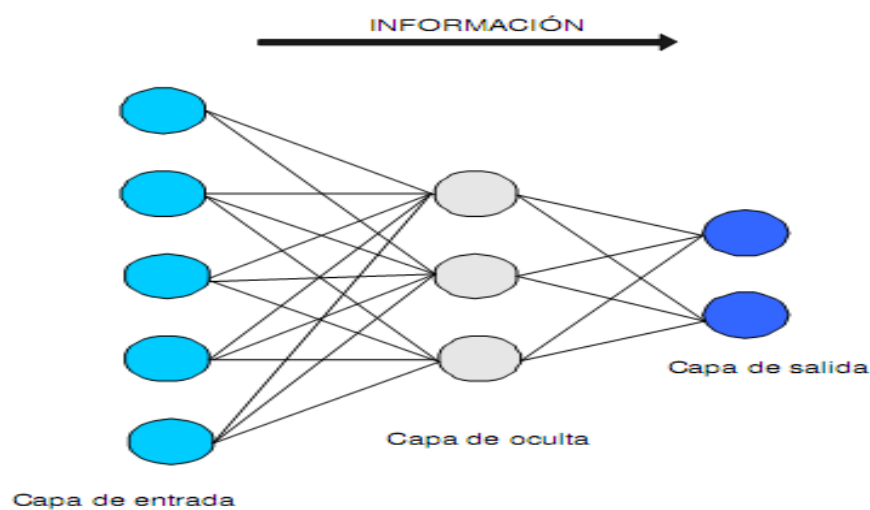


Figura N° 2.10: Arquitectura unidireccional con tres capas de neuronas: una capa de entrada, capa oculta y capa de salida.

Se establecen diferentes tipos de arquitecturas neuronales. En relación a su estructura de capas pueden ser redes monocapa y redes multicapa. Las redes monocapa son aquellas compuestas por una única capa de neuronas, por ejemplo: el perceptrón simple, la red de Hopfield, las máquinas estocásticas de Botzmann y Cauchy. Las redes multicapa son aquellas cuyas neuronas se organizan en varias capas, por ejemplo: el perceptrón multicapa, Adaline y Madaline.

Tomando en cuenta el flujo de datos de la red neuronal, se clasifican en: redes unidireccionales (feedforward) y redes recurrentes (feedback). En las redes unidireccionales, la información se propaga en un único sentido, desde la neurona de

entrada hacia las de salida. En las redes recurrentes o realimentadas la información puede circular entre las capas en cualquier sentido, incluido salida-entrada.

En función de cómo está constituido el mecanismo de aprendizaje, se clasifican como: redes supervisadas y redes no supervisadas. La figura 2.11., muestra la taxonomía de las RNA.

Las redes supervisadas necesitan un conjunto de datos de entrada previamente clasificados o cuya respuesta de salida sea conocida. Entre las redes supervisadas se encuentran las que tienen reglas de aprendizaje basadas en la corrección del error, como la regla de retropropagación del error en el caso del Perceptrón y el algoritmo de mínimos cuadrados muy utilizados en problemas de clasificación y predicción.

Las redes no supervisadas, no tienen un conjunto de datos previamente clasificados y deben encontrar las relaciones que se pueden establecer entre los datos que se presentan en la entrada. La red por sí misma buscará su comportamiento más adecuado atendiendo a cierto criterio y encontrará estructuras o prototipos en el conjunto de patrones de entrenamiento. Como ejemplo tenemos la regla de aprendizaje competitivo no supervisado (utilizada en problemas de agrupación de patrones y obtención de prototipos), la regla de Kohonen (utilizada en reconocimiento e identificación de patrones) y la regla de Hebb.

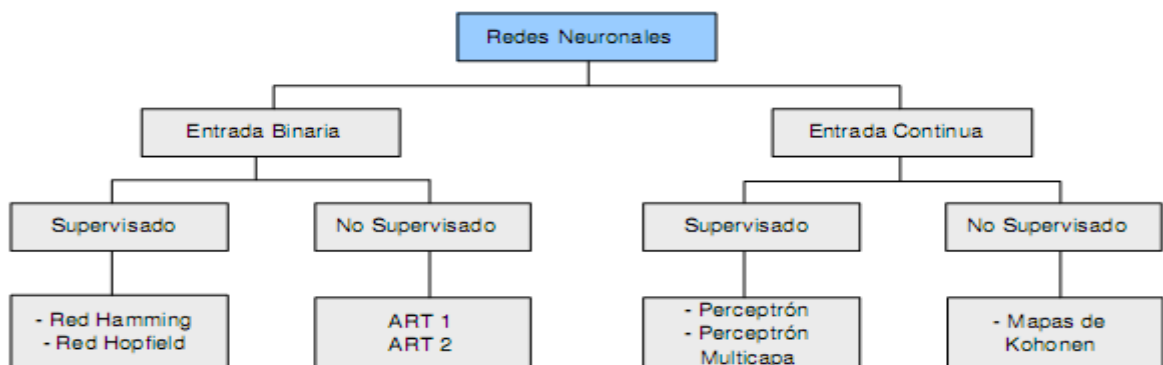


Figura N° 2.11: Taxonomía de las redes neuronales artificiales

2.2.3.6. Proceso de Aprendizaje

Una de las principales propiedades de las RNA es su capacidad de aprender a partir de datos, a través de un proceso de aprendizaje o entrenamiento. Como el proceso de aprendizaje humano, el de las RNA se basa en el uso de ejemplos que representan el problema. A este conjunto se le conoce como conjunto de entrenamiento. Como se mencionó anteriormente, en función de cómo esté constituido el conjunto de entrenamiento se distinguen dos tipos de aprendizaje: supervisado y no supervisado.

En el aprendizaje supervisado se parte de una serie de observaciones o entradas y unas salidas deseadas, que la red debería obtener y el objetivo es aprender la correspondencia entre ambas. Las redes multicapa y las redes recurrentes son apropiadas para problemas de aprendizaje supervisado, donde se dispone de un conjunto de patrones de entrenamiento de la forma:

$$X = \{(x^n, t^n)\}_{n=1}^N$$

dónde x es el vector de entrada, t el de la salida deseada y N es el tamaño del conjunto.

El entrenamiento está basado en que la red sea capaz de reproducir estos patrones con el menor error posible. Este esquema responde a problemas de clasificación y ajuste de datos, desde un punto de vista estadístico se puede considerar como una generalización de la regresión no lineal. La aproximación que se sigue es partir del modelo genérico $f(\bullet)$ que representa la red, totalmente definido excepto por una serie de parámetros o pesos w , y con él producir una buena aproximación de t :

$$y = f(x|w)$$

dónde y representa la salida proporcionada por la red. El algoritmo de aprendizaje optimizará los parámetros para que la salida sea lo más parecida al conjunto de entrenamiento. Es decir, trata de encontrar el conjunto de parámetros w^* que minimiza

el error de aproximación, por esta razón se define una función de error E, que en cada paso del aprendizaje indica que tan cerca se está de la solución.

La función de error más utilizada es el Error Cuadrático Medio (MSE-Mean Square Error, por sus siglas en inglés)

$$MSE = \frac{1}{N} \sum_{n=1}^N (y^n - t^n)^2$$

donde n es el número de patrones, y^n es la salida obtenida por la red, y la salida deseada t^n que debería haber obtenido la red.

Por otra parte, en el aprendizaje no supervisado el conjunto de entrenamiento los datos de entrada x no vienen acompañados de una salida deseada y el objetivo no es encontrar el mapeo entrada-salida, sino encontrar patrones que definan regularidades en x. Las redes competitivas responden a problemas de aprendizaje no supervisado, la red tiene que auto-organizarse para hallar las salidas óptimas para cada entrada en base a algún criterio de aprendizaje que corresponda a maximizar la entropía de la salida. Normalmente este tipo de redes se utilizan en problemas de agrupamiento, donde cada una de las neuronas de salida representa un grupo distinto, y la red tiene que distribuir los patrones de entrenamiento entre los grupos de forma automática.

La red neuronal que se utiliza en este trabajo es un perceptrón multicapa. Por esta razón en las siguientes secciones se hace una introducción al perceptrón simple y al perceptrón multicapa.

2.2.3.7. El Perceptrón

El perceptrón es la RNA más sencilla, fue introducida por Frank Rosen-Blatt a finales de los años cincuenta y emplea como regla de aprendizaje una modificación de la propuesta por Hebb.

- ✓ La regla de Hebb. Una de las primeras reglas de aprendizaje fue propuesta por D. Hebb en 1949 basándose en un hecho biológico: cuando dos neuronas se activan simultáneamente su conexión se refuerza. Esto es, cuando una neurona tiene un nivel positivo de actividad y se siente estimulada por otra neurona, tiende a producirse un refuerzo en la conexión que las enlaza. Por el contrario, si una neurona tiende a inhibir a otra, la fuerza de conexión entre ambas tiende a disminuir. Esta idea se expresa matemáticamente como:

$$w_i(\tau + 1) = w_i(\tau) + \eta x_i y$$

dónde la entrada x_i es la salida de otra neurona, w_i es el peso en el instante $\tau + 1$ y el peso en el instante anterior, η es la tasa de aprendizaje.

El perceptrón simple es un modelo unidireccional, compuesto por dos capas, una capa entrada de j neuronas con funciones ϕ_j que se encarga de transformar los datos de entrada. La salida del perceptrón se expresa como:

$$y = g(a) = g\left(\sum_{j=0}^M w_j \phi_j(\mathbf{x})\right) = g(\mathbf{w}^T \phi)$$

dónde g es la función escalón y ϕ es el vector formado por las funciones de activación. La función de activación de la neurona de salida es tipo escalón.

El perceptrón puede utilizarse tanto como clasificador, como para la representación de las funciones booleanas, pues su neurona es esencialmente del tipo McCulloch-Pitts, de salida binaria. En un clasificador, una neurona tipo perceptrón solo permite discriminar entre dos clases linealmente separables.

El algoritmo de aprendizaje del perceptrón simple pertenece al grupo de los algoritmos que se fundamentan en la corrección de errores. Los algoritmos de este tipo ajustan los pesos de manera proporcional a la diferencia existente entre la salida actual de la red

neuronal y la salida deseada, con el objetivo de minimizar el error actual de la red, es por tanto una RNA de aprendizaje supervisado. Dado un vector de entrada x^n , asociado a un valor objetivo t^n , teniendo en cuenta que cada entrada genera su correspondiente vector de activación ϕ^n , y considerando la ecuación 4.12, el criterio que sigue el perceptrón es:

$$E^{perc}(\mathbf{w}) = - \sum_{\phi^n \in M} (\mathbf{w}^T \phi^n) t^n$$

dónde M es el conjunto de vectores ϕ^n que han sido clasificados incorrectamente de modo que E^{perc} será igual a 0, cuando los vectores han sido correctamente clasificados. Con el objetivo de minimizar el error los pesos w que son inicializados aleatoriamente, se modifican iterativamente cuando la salida del perceptrón no coincide con la salida deseada. La modificación de los pesos se realiza con la regla de aprendizaje del perceptrón simple dada por:

$$w_j(\tau + 1) = w_j(\tau) + \eta \phi_j^n t^n$$

η es la tasa de aprendizaje.

2.2.3.8. El Perceptrón Multicapa

El perceptrón de una sola capa resulta muy útil como clasificador de patrones, pero resulta de utilidad limitada debido a que requiere que los patrones que clasifica sean linealmente separables (i.e. separables por un hiperplano). Las limitaciones de las redes de una sola capa hicieron que se planteara aumentar el número de capas introduciendo capas intermedias entre la entrada y la salida, de modo que se pudiese implementar cualquier función con el grado y precisión deseada. El perceptrón multicapa (MLP- MultiLayer Perceptron, por su siglas en inglés) incluye en su arquitectura una o varias capas intermedias, llamadas capas ocultas. Esta arquitectura suele entrenarse mediante

el algoritmo de retropropagación de errores o BackPropagation (BP-por sus siglas en inglés).

La MLP está formada por D entradas, M neuronas en la capa oculta y C neuronas en la capa de salida. El nivel de activación a_j de la neurona j de la capa oculta, se calcula como una combinación lineal de las D entradas x_i , tras aplicar una función de transferencia g se obtiene la salida z_j de dicha neurona:

$$z_j = g(a_j) = g \left(\sum_{i=1}^D w_{ji} x_i \right); \quad k = 1, 2, \dots, M$$

dónde w_{ji} es el peso asociado a la neurona j y la entrada x_i . La salida de la red es obtenida como la suma ponderada de las salidas de las neuronas de la capa oculta, sobre la cual se aplica la función de transferencia, de tal modo que la salida de la neurona k viene dada por:

$$y_k = \tilde{g} \left(\sum_{j=0}^M w_{kj} z_j \right) = \tilde{g} \left(\sum_{j=0}^M w_{kj} g \left(\sum_{i=1}^D w_{ji} x_i \right) \right); \quad k = 1, 2, \dots, C$$

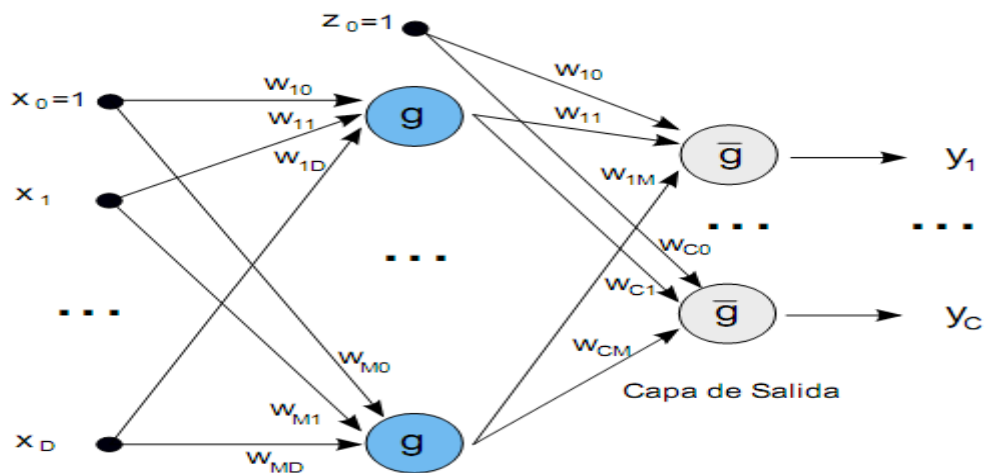


Figura N° 2.12: Arquitectura de un perceptrón multicapa con una capa oculta.

En la ecuación anterior es importante destacar que la función g y \tilde{g} no tienen que ser iguales.

2.2.3.8.1. Algoritmo de Retropropagación

El MLP basa el aprendizaje de sus pesos en una regla de ajuste del error, para determinar los pesos de las conexiones que sean lo más próximas posibles a las salidas esperadas. El algoritmo de aprendizaje de una MLP utiliza el método de descenso del gradiente para ajustar los pesos de la red de modo que, el aprendizaje se realiza en sentido opuesto a la computación, es decir, hacia atrás. El ajuste comienza de la capa de salida, según el error cometido y propagado en las capas anteriores, hasta llegar a la capa de entrada, de ahí que se denomine algoritmo de retropropagación. Básicamente consiste en dos fases que se repiten hasta minimizar el error:

- ✓ La primera fase, o paso hacia adelante, se aplica un patrón a las entradas de la red y su efecto se propaga a través de la misma, capa a capa. Finalmente la red representa un conjunto de salidas como respuesta a dicho patrón de entrada. Normalmente, en la primera iteración del algoritmo se parte de un conjunto de pesos con un valor inicial pequeño aleatorio.
- ✓ La segunda fase, o paso hacia atrás, los pesos de la red se recalculan de acuerdo a la regla de ajuste de error, donde se compara la respuesta actual de la red y la respuesta deseada, propagando el error hacia atrás.

La función de error más ampliamente utilizada es el Error Cuadrático Medio, así para cada patrón x^n se tiene el siguiente error:

$$E^n = \frac{1}{2} \sum_{k=0}^C (e_k^n)^2$$

el error e_k^n se define como:

$$e_k^n = y_k^n - t_k^n$$

El algoritmo de retropropagación corrige el peso sináptico w_{kj} , añadiéndole un incremento Δw_{kj} , en este caso el incremento es proporcional al gradiente $\delta E / \delta w_{kj}$. El

gradiente determina la dirección de búsqueda en el espacio de pesos para obtener el valor del peso w_{kj} que conduce al mínimo de E.

La actualización de los pesos en el algoritmo de retropropagación, está dado por:

$$\Delta w_{ji} = -\eta \delta_j x_i$$

dónde δ_j es el gradiente local para una neurona en la capa oculta, definida como:

$$\delta_j = g'(a_j) \sum_k \delta_k w_{kj}$$

dónde $g'(a_j) = \frac{\delta z_j}{\delta a_j}$.

2.2.4. MINERIA DE DATOS

2.2.4.1. Definición

Se define a la minería de datos [Castellanos.2013] como el proceso de extracción de conocimiento no trivial que se encuentra de manera implícita en los sets de datos disponibles, considerando que estos datos pueden provenir de diferentes y variadas fuentes; este conocimiento es previamente desconocido y debe ser de utilidad. Es decir que la minería de datos plantea dos desafíos, por un lado trabajar con grandes bases de datos y por el otro aplicar técnicas que conviertan en forma automática estos datos en conocimiento.

La minería de datos es un elemento dentro de un proceso más amplio que tiene como objetivo el descubrimiento de conocimiento en grandes bases de datos, en inglés “Knowledge Discovery in Databases” (KDD), este proceso, como lo muestra la figura 2.13, tiene una primera etapa de preparación de datos, luego el proceso de minería de datos, la obtención de patrones de comportamiento, y la evaluación e interpretación de los patrones descubiertos.

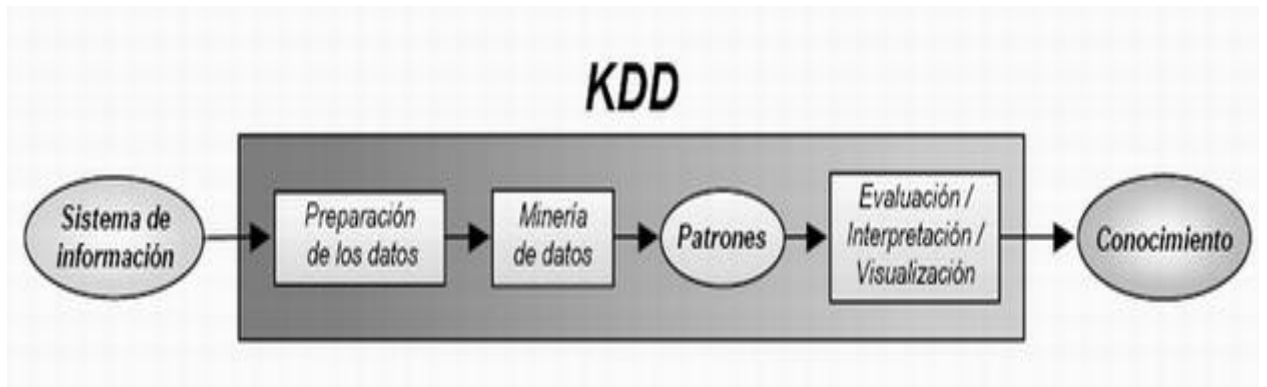


Figura 2.13: Fases del proceso KDD

2.2.4.2. Tipos de modelos de minería de datos.

Los modelos de minería de datos pueden ser de dos tipos, predictivos o descriptivos. Dentro del modelo predictivo se encuentra la clasificación y la regresión. En el modelo descriptivo se encuentra el agrupamiento o clustering, las reglas de asociación, las reglas de asociación secuenciales.

- Modelos predictivos: Este tipo de modelo tiene como objetivo la estimación de valores desconocidos de variables de interés.

Fundamentalmente en los modelos predictivos se utiliza la clasificación, donde dada una base de datos se indica mediante un valor en cada tupla a que clase pertenece, el objetivo es predecir a que clase pertenece una nueva instancia, considerando que los atributos pueden asumir valores discretos.

También se utiliza la regresión, en este caso el valor a predecir es numérico.

- Modelos Descriptivos: Estos modelos exploran las propiedades de los datos examinados con el objetivo de generar etiquetas o agrupaciones.

Se utiliza el clustering, para analizar datos y generar etiquetas. La correlación se utiliza para determinar el grado de similitud de los valores de dos variables numéricas.

También se utilizan las reglas de asociación, que son similares a la correlación y tienen como objetivo encontrar relaciones no explícitas entre atributos.

Las reglas de asociación secuencial se utilizan para determinar los patrones secuenciales en los datos basados en el tiempo.

2.2.4.3 Proceso de explotación de información.

La minería de datos es considerada como una etapa dentro de un proceso de explotación de información, este proceso consta de las siguientes etapas:

A. Selección de datos (data selection).

B. Extracción de datos relevantes para obtener el objetivo planteado, estos datos pueden encontrarse en diferentes formatos y provenir de distintas fuentes.

- Integración de datos (data integration). El objetivo de esta etapa es crear una única fuente, para lograr esta meta se deben integrar todas las fuentes de datos.
- Limpieza de datos (data cleaning). Se deben identificar, procesar y de ser necesario eliminar de datos anómalos, erróneos, faltantes, ruidosos o sin relevancia.
- Transformación de datos (data transformation). Los datos son transformados para optimizar el posterior proceso de minería de datos, algunas de las técnicas que generalmente se aplican son:
 - Agregación, con esta técnica se aplican funciones de resumen de los datos.
 - Suavización, el objetivo es eliminar el ruido de los datos.
 - Construcción de atributos, con el objetivo de mejorar el proceso de obtención de conocimiento se pueden crear nuevos atributos.
 - Normalización, con esta técnica se pretende crear categorías en determinados atributos y modificar el contenido de esas columnas en función de las categorías definidas.
- Minería de datos (Data Mining). Obtención de patrones de comportamiento de los datos en forma automática, que son previamente desconocidos y que son de utilidad.

- Evaluación de patrones (Pattern Evaluation). Identificación y análisis de patrones interesantes, esta tarea debe ser realizada por el experto en el dominio junto al especialista en minería de datos.
- Presentación del conocimiento (Knowledge presentation). Visualización y representación de los conocimientos obtenidos.

Existen varias metodologías para implementar procesos de explotación de información, las más difundidas son CRISP-DM, SEMMA y P3TQ:

- CRISP-DM [Hanumanth y Prasada.2013] se encuentra definida en función a un modelo jerárquico de procesos, donde se define un ciclo de vida de los proyectos de explotación de información cuyas fases son: entendimiento de negocios, entendimiento de los datos, preparación de los datos, modelado, evaluación y despliegue. Se trata de una metodología de uso público y es la más difundida y utilizada.
- SEMMA [Britos.et al.2008] toma su nombre de las etapas que esta metodología define para procesos de explotación de información, estas etapas son: muestreo (sample), exploración (explore), modificación (modify), modelado (model) y valoración (assess). La metodología fue desarrollada por SAS Institute Inc.12, uno de las mayores organizaciones relacionadas con el desarrollo de software de inteligencia de negocios. SEMMA está desarrollada para aplicarla sobre la herramienta de minería de datos “SAS Enterprise Miner”.

Esta metodología es aplicada fundamentalmente por quienes usan esta específica herramienta software y su aplicación es limitada.

- P3TQ [(Britos.et al.2008)], su nombre proviene de los principales elementos que considera la metodología: producto (product), lugar (place), precio (price), tiempo (time) y cantidad (quantity). La metodología propone desarrollar el modelo de

negocios y el modelo de explotación de información. Su aplicación aún es incipiente.

2.2.4.4. Técnicas de la minería de datos.

Existen diferentes paradigmas detrás de las técnicas que se aplican en el proceso de minería de datos. La minería de datos utiliza técnicas basadas en el análisis estadístico y técnicas basadas en los sistemas inteligentes. En relación al análisis estadístico [Britos.et al.2008] algunas de las tecnologías más utilizadas son:

- **Análisis de agrupamiento:** en base a la característica común de una determinada cantidad de grupos, este tipo de análisis se utiliza para la clasificación de individuos.
- **Análisis de varianza:** es utilizado para la búsqueda de diferencias entre las medias de una cantidad determinada de variables de tipo continuas.
- **Análisis discriminante:** con este tipo de análisis es posible clasificar individuos en grupos establecidos previamente, para después encontrar la regla de clasificación para los elementos de cada grupo y por último identificar aquellas variables que definen la inclusión en un grupo.
- **Prueba de Chi-cuadrado:** Prueba que permite medir la probabilidad de que la frecuencia observada de una muestra sea debida sólo a la variación de dicha muestra.
- **Regresión:** es utilizada para definir relaciones entre un conjunto de variables y un conjunto asociado de otras variables que son utilizadas para la predicción de las primeras.
- **Series de tiempo:** utilizadas para el estudio de la evolución de una variable a lo largo del tiempo y para poder realizar predicciones.

Se pueden mencionar entre otras a las siguientes tecnologías basadas en sistemas inteligentes (García Martínez et al., 2007):

- Algoritmos de Inducción. TDIDT (Top Induction Decision Trees), donde a partir de ejemplos pre-clasificados es posible clasificar nuevos ejemplos.
- Perceptron Multicapa. Es un algoritmo basado en redes de neuronas artificiales de backpropagation.
- Maquinas de vector soporte. Se trata de un conjunto de algoritmos de aprendizaje supervisado.
- Algoritmos a priori. Que se basa en el conocimiento previo de los datos.
- Redes neuronales SOM. Conocidas como mapa auto organizado de Kohonen.
- Algoritmos genéticos. Que se basan en la evolución biológica.
- Redes bayesianas. Mediante la estimación de las probabilidades que utilizan el Teorema de Bayes.
- Algoritmo del vecino más próximo. Utilizado en la solución del problema del viajante.
- Existen otras técnicas como las estocásticas, relacionales, declarativas, difusas, etc. y una variedad de técnicas híbridas.

CAPÍTULO III: METODOLOGÍA

3.1. MÉTODOS EMPLEADOS EN LA INVESTIGACIÓN

3.1.1. Tipo de Investigación

De acuerdo con Grajales [Grajales.2002], la investigación se puede clasificar desde tres perspectivas diferentes: según la fuente de datos, según la finalidad y según la profundidad del estudio.

Según la fuente de datos una investigación se divide en documental, de campo o experimental.

- ✓ La investigación documental se realiza apoyándose en fuentes de carácter documental, esto es, en documentos de cualquier especie.
- ✓ La investigación de campo se apoya en informaciones que provienen entre otras, de entrevistas, cuestionarios, encuestas y observaciones.
- ✓ La investigación experimental obtiene su información de la actividad intencional realizada por el investigador y que se encuentra dirigida a modificar la realidad con el propósito de crear el fenómeno mismo que se indaga, y así observarlo.

Según la finalidad las investigaciones se dividen en básica y aplicada.

- ✓ La investigación básica también recibe el nombre de investigación pura o teórica. Se caracteriza porque parte de un marco teórico y permanece en él; la finalidad radica en formular nuevas teorías o modificar las existentes, es decir, en incrementar los conocimientos científicos.
- ✓ La investigación aplicada, también llamada práctica o empírica; se caracteriza porque busca la aplicación o utilización de los conocimientos que se adquieren. Se vincula con la investigación básica, pues depende de los resultados y avances de esta última.

Según la profundidad del estudio, la investigación se divide en exploratoria, descriptiva o explicativa.

- ✓ La investigación exploratoria se realiza con el propósito de destacar los aspectos fundamentales de una problemática determinada y encontrar los procedimientos adecuados para elaborar una investigación posterior.
- ✓ La investigación descriptiva utiliza el método de análisis; mediante esta se logra caracterizar un objeto de estudio o una situación concreta, señalar sus características y propiedades. Además, puede servir de base para investigaciones que requieran un mayor nivel de profundidad.
- ✓ La investigación explicativa requiere la combinación de los métodos analítico y sintético, en conjugación con el deductivo y el inductivo. Mediante este tipo de investigación se trata de responder o dar cuenta del porqué del objeto que se investiga.

Con base en lo anterior, la presente investigación es de tipo documental, básica, descriptiva y explicativa. Es documental porque para iniciar es necesario hacer una revisión detallada de libros, tesis y artículos con el fin de establecer los elementos teóricos y metodológicos relacionados con redes neuronales y minería de datos. Es aplicada ya que después de la investigación teórica se desarrolla una aplicación práctica que consiste en un sistema integral para predecir la contaminación atmosférica por PM_{10} . Es descriptiva ya que para diseñar los modelos de redes neuronales es necesario describir detalladamente el comportamiento del fenómeno de estudio; y finalmente es explicativa, ya que el modelo busca explicar cómo las variables se relacionan para generar una salida determinada, es decir, tiene un componente de inferencia estadística.

Por otro lado, de acuerdo con [Pita y Pértegas.2002], la investigación cuantitativa es aquella en la que se recogen y analizan datos cuantitativos sobre variables; trata de determinar la fuerza de asociación o correlación entre variables, la generalización y objetivación de los resultados a través de una muestra para hacer inferencia a una población de la cual toda muestra procede. Por otra parte, según los mismos autores, la investigación cualitativa hace la descripción de las cualidades de un fenómeno que son estudiados mediante técnicas como la observación participante y las entrevistas no estructuradas. Se trata de obtener un entendimiento lo más profundo posible.

Este proyecto corresponde a una investigación cuantitativa, ya que, el objetivo fue el desarrollo de un sistema de pronósticos mediante redes neuronales, lo cual implica la relación de diferentes variables para obtener un resultado numérico de la proyección; además se optimizaron las redes y se compararon con otros modelos tradicionales analizando los errores de pronósticos, medido todo en términos cuantitativos.

3.1.2. Método de investigación.

Para la formulación del modelo de predicción bajo el enfoque de redes neuronales y en la simulación de éste en el proceso de contaminación atmosférica por PM10, es un proceso complejo y dinámico se utilizó el método analógico.

Para el desarrollo del algoritmo, amparado en la Teoría de la Computabilidad y en la teoría de la lógica-matemática se ha empleado el método deductivo.

En el desarrollo del sistema integral se ha utilizado el método inductivo.

3.2.METODOLOGÍA PARA LA PRUEBA DE HIPOTESIS

3.2.1. Hipótesis

La implementación de un sistema integral, bajo el enfoque de minería de datos y redes neuronales, permitirá realizar la predicción de la contaminación atmosférica por PM10, en la ciudad de Chimbote, en un 91%, a través de información clara y transparente.

3.2.2. Variables:

VARIABLE INDEPENDIENTE: Sistema Integral, bajo el enfoque de minería de datos y redes neuronales

VARIABLE DEPENDIENTE: Predicción de la contaminación atmosférica por PM10, en la ciudad de Chimbote.

3.2.3. Indicadores

Variable Independiente:

- Confiabilidad
- Seguridad
- Accesibilidad
- Flexibilidad

Variable Dependiente:

- Tiempo
- Satisfacción del usuario

3.2.4. Diseño Descriptivo:

El diseño de estudio de la investigación es el diseño descriptivo de grupo único con pre prueba y post prueba

G1 O₁ X O₂

Dónde:

O_1 : Medición previa antes de aplicar el estímulo. (Pre pruebas)

O_2 : Medición posterior después de aplicar la variable estímulo.

(Post pruebas)

X: Variable Independiente o variable estímulo. (Sistema inteligente).

3.3.TÉCNICAS E INSTRUMENTOS EMPLEADOS

Para la recolección de datos se hará uso de las siguientes técnicas e instrumentos:

3.3.1. Técnicas

- Observación Directa
- Revisión de documentos
- Entrevista
- Encuesta
- Investigación Bibliográfica

3.3.2. Instrumentos

- Ficha de Observación
- El cuaderno de Notas.
- Cuestionario.
- Ficha de Encuesta
- Ficha Bibliográfica

3.4. PROCEDIMIENTO DE LA RECOLECCIÓN DE DATOS

La recolección de datos abarca desde el análisis para la selección y validación de las variables que se usarán para el procesamiento de la red neuronal, que se dividen en dos grupos variables meteorológicos (07) y variables estacionales (02), a estos se añaden los datos contaminantes. Una vez definida dichas variables, la recolección

de datos se procesará a través del propio sistema integral, para lo cual se pueden dar a través de 3 opciones: (1) A través de una interfaz que se conecta con una estación de monitoreo meteorológica, de tal manera que la data se irá recolectando automáticamente y en tiempo real, (2) La información se va ingresando diariamente, a través de una interfaz, donde se ira digitando los valores por cada variable. (3) A través de un archivo CSV, permitirá importar los datos en bloque, que contenga la información meteorológica de otros días. Todo esto, es necesario con la finalidad de que toda la información se validada y sirva para poder realizar el entrenamiento de la red neuronal, de tal manera que luego se puedan realizar las simulaciones (predicción).

La edición de dichos datos, se hizo con el fin de comprender mejor la información en cuanto a la legibilidad, consistencia y totalidad de la información, clasificando esta, para poder hacer un análisis comparativo

3.5. TÉCNICAS DE PROCESAMIENTO Y ANÁLISIS DE DATOS.

Con la finalidad de contrastar la hipótesis de trabajo, se empleará la técnica de análisis estadístico t-student, para la demostración de la hipótesis en función de la satisfacción del usuario; y para la demostración de la hipótesis en función del tiempo, se utilizó la prueba de comparación de varianzas. Se harán las Pruebas del Sistema Integral, donde incluye las pruebas de caja negra y pruebas de caja blanca.

Las técnicas de análisis de datos representa la forma de cómo será procesada la información recolectada según las técnicas de recolección de datos utilizados, ésta será procesada de dos formas *Cuantitativa y Cualitativamente*:

- El análisis cuantitativo se define como una operación que se efectúa con toda la información numérica resultante de la investigación, ésta luego del proceso que se le haya aplicado, se presentará como un conjunto de cuadros y medidas con

porcentajes ya calculados, nos permitirá representar gráficamente los resultados de los datos obtenidos para tener la información ordenada con representaciones visuales que nos permitan su posterior estudio.

- El análisis cualitativo es una técnica que indaga para conseguir información de sujetos, comunidades, contextos, variables o ambientes en profundidad, asumiendo una actitud abierta y previniendo a toda costa no involucrar sus afirmaciones o práctica.

3.6. PROCEDIMIENTO METODOLOGICO PARA LA RED NEURONAL

La metodología utilizada para el uso del modelo de la red neuronal se basa en el proceso KDD (Knowledge Discovery in Databases). Este proceso consiste en: “Un proceso no trivial de identificar patrones previamente desconocidos, válidos, nuevos, potencialmente útiles y comprensibles dentro de los datos” [Sánchez.2012]

Es posible desglosar este proceso en una serie de etapas, las que se describen en la figura 3.1.

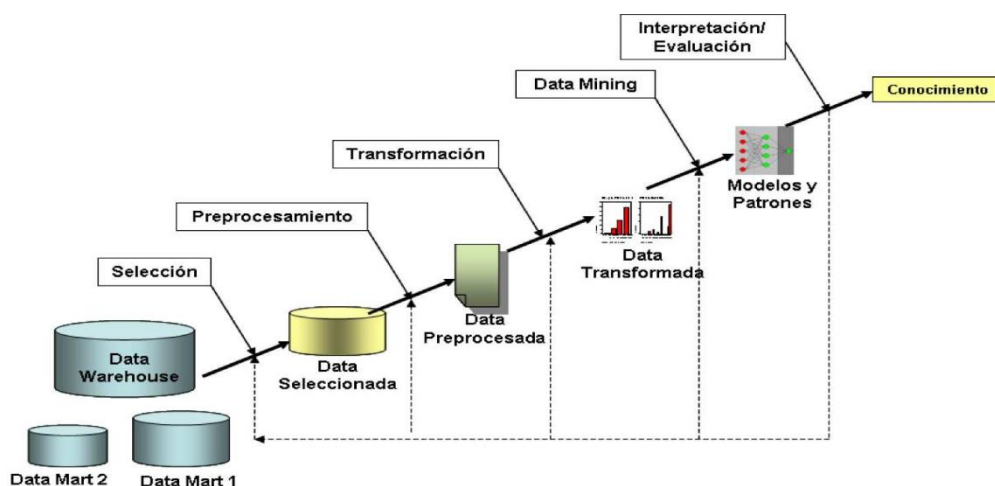


Figura N° 3.1: Etapas de la metodología utilizada.

3.6.1. Selección y Descripción de las Variables del Estudio

En esta etapa se seleccionan las variables que se utilizan como entrada para el modelo predictivo. Las fuentes de información utilizadas provienen de diversas bases de datos y Datawarehouses que tiene la ciudad de Chimbote sobre contaminación ambiental al momento del estudio.

La selección del conjunto inicial de variables de entrada se realizará en conjunto con los expertos del negocio, quienes transmitirán su experiencia y conocimientos sobre temas relacionados con la contaminación atmosférica.

3.6.2. Preprocesamiento de la Información

En esta etapa se definen los criterios para identificar y dar solución a las inconsistencias que tenga la base de datos. Existen dos tipos de inconsistencias dentro de los registros de la base de datos: valores faltantes y valores fuera de rango. Los valores faltantes corresponden a la inexistencia del valor de un registro en cierta variable y los valores fuera de rango corresponden a valores que se escapan de los rangos normales de las variables.

Las inconsistencias producen distorsiones sobre las distribuciones empíricas de las variables. El resultado final de esta etapa, es una base de datos sin inconsistencias y totalmente poblada para su posterior análisis.

3.6.3. Transformación de la Información

El objetivo central de esta etapa es obtener una base de datos totalmente numérica tanto para la aplicación de diversos análisis estadísticos, como para la aplicación de las técnicas de minería de datos necesarias para el desarrollo del modelo predictivo.

Las transformaciones utilizadas fueron:

- ✓ Conversión de variables texto o string a variables categóricas numéricas.
- ✓ Escalamiento y estandarización del valor del registro sobre un rango específico.
- ✓ Creación de nuevas variables a partir de las variables originales.

3.6.4. Selección de la Técnica de Minería de Datos

En esta etapa se desarrolla el modelo predictivo, el cual es posible construir por diversas técnicas de minería de datos que modelan un problema de predicción. Algunos ejemplos posibles de mencionar son: redes neuronales artificiales, arboles de decisión o los support vector machines.

Para este caso en particular, motivados por la efectividad y robustez, se seleccionó como técnica de modelación las RNA (redes neuronales artificiales) ofrecen un medio eficiente para modelar problemas complejos en los que existen cientos de variables predictoras en las que se dan multitud de interacciones [Crows,1999]. Pueden ser utilizadas en problemas de clasificación (la variable de salida es cualitativa) o en predicción (la variable de salida es cuantitativa).

3.6.5. Evaluación e Interpretación de los Resultados

En esta etapa se evaluarán los resultados del modelo para varios meses y se estimará el nivel de acierto en la predicción y control de la contaminación ambiental en Chimbote.

El resultado de esta etapa es la caracterización del patrón de agentes contaminantes siendo posible identificar las causas o razones que generan esta acción. Al identificar dichos motivos será posible implementar mejoras en la calidad del aire.

CAPÍTULO IV: RESOLUCIÓN DEL PROBLEMA DESARROLLO DEL SISTEMA INTEGRAL

Como parte de este capítulo se presenta los resultados del presente proyecto: “Sistema integral para un modelo de predicción y control de la contaminación atmosférica por partículas en suspensión menores de 10 micras (PM10) bajo el enfoque de minería de datos y redes neuronales”; cabe mencionar que para la implementación del sistema y de acuerdo a la complejidad de este, se tuvo que hacer uso de diversas tecnologías, siendo los sistemas desarrollados los siguientes:

- **Sistema de registro y validación de data (Desarrollado en Power Builder).-**

Se ha desarrollado un sistema por el cual se recoleta la data meteorológica diaria, y también se da las opciones para que se conecte directamente a estaciones de monitorio en tiempo real, de tal manera que este sistema valida dicha información, y la procesa para que pueda ser exportada, usada por el sistema de Matlab y por el software Knime. Dicho sistema, adicionalmente sirve para validar las variables estacionales y meteorológicas, en comparación con el PM10, de tal manera que se verifique si dichas variables influyen en el factor contaminante (PM10). Así mismo sirve para realizar todo tipo de consultas, reportes y filtros de data meteorológica.

- **Sistema de procesamiento de red neuronal (Desarrollado en Matlab).-**

Se ha realizado el sistema para el procesamiento de la red, a través de Interfaces gráficas(GUI), en Matlab; el cual importa la data del sistema de Power Builder y realiza la configuración, creación, entrenamiento, validación y simulación (predicción) de las red neuronal modelada para este proyecto. Como resultado de este sistema, exporta el archivo que contiene el resultado de la predicción del ICA a 24 horas, a través de datos binarios.

- **Sistema web de visualización de predicciones (Desarrollado en Php).-**

Sistema que importa los resultados de Matlab, los procesa e interpreta, de acuerdo a los rangos cualitativos del ICA. De tal manera que cuando el público en general desea hacer la consulta del ICA a 24 horas siguientes, el sistema muestra la predicción a nivel de estado, color, rango PM10, rango ICA, efectos en la salud, acciones preventivas.

- **Procesamiento de minería de datos (Se hace uso del software Knime).-**

El sistema de Power Builder, exporta la data para realizar el procesamiento de minería de datos a través del Knime.

4.1. RESULTADOS DEL PROCESAMIENTO DE RED NEURONAL

4.1.1. Análisis de la contaminación atmosférica por partículas en suspensión

menores de 10 micras (PM10)

Los contaminantes atmosféricos considerados en las normas de calidad del aire a nivel internacional son: partículas suspendidas (**PS**), plomo (**Pb**), monóxido de carbono (**CO**), óxidos de azufre (**SO₂**), óxidos de nitrógeno (**NO_x**) e hidrocarburos y oxidantes fotoquímicos; los seis primeros son emitidos en forma directa por la combustión, los procesos industriales, la erosión y los incendios.

Las partículas suspendidas (**PS**) abarcan un amplio intervalo de sólidos y líquidos, sutilmente divididos que pueden estar dispersos en el aire, las partículas menores de 10 μm constituyen la fracción respirable (**PM₁₀**), que por su tamaño pueden llegar a los alvéolos pulmonares y afectar a la vida humana.

El procesamiento de harina de pescado, así como la operación del horno eléctrico de arco en la siderúrgica constituyen las principales fuentes de emisión de partículas totales y menores a 10 micras. Las emisiones de material particulado de diámetro aerodinámico inferior a 10 micras, de importancia para la salud dada su facilidad para ingresar al

sistema respiratorio humano, representan aproximadamente el 30% de las emisiones totales del material particulado en la cuenca atmosférica de Chimbote.

4.1.1.1. Identificación de estaciones de monitoreo.-

La topografía y los factores climáticos de la Zona de Estudio, han dado como resultado la formación de pendientes planas con ligeras inclinaciones y otras con accidentadas pendientes, las cuales han permitido definir pequeñas micro cuencas con diferentes características meteorológicas y topográficas, prestándole atención a sus desarrollos sociales, urbanos, industriales, económicos, entre otros; pero también observar la variación del impacto ambiental producto de las emisiones de los diferentes rubros industriales que laboran en cada una de estas microcuencas o también denominadas en este estudio estaciones de monitoreo.

- **Estación de Santa**

La Sub cuenca de Santa cuenta con una población de 19,010 habitantes en el 2013 [INE.2013], está localizado al norte de Chimbote, posee extensas áreas de cultivo, localizadas en la margen izquierda del río Santa, dedicadas a la mayor producción de arroz, marigol, maíz, algodón, entre otras. El desarrollo del agro se ha convertido en la principal actividad del desarrollo económico de la zona, llegando a instalarse varias empresas para el procesamiento del pilado de arroz, así como desmotadoras de algodón y una empresa procesadora de la flor de marigol; todas ellas causantes de emisión de material particulado por la falta de aplicación de filtros y gases producidos en sus calderas o por el trabajo a cielo abierto (proceso del marigol), afectando a la población circundante. Ver figura 4.1.

La industria pesquera es otra actividad industrial instalada en la zona, pero de bajo desarrollo, siendo causante directo de efluentes hacia el océano y de emisión de gases

tóxicos por la falta de aplicación de tecnologías limpias en el proceso de combustión para el procesamiento del pescado.



Figura N° 4.1: Mapa de Santa

- **Estación de Coishco**

Coishco fue constituido como distrito en el año 1988, y limita por el sur con el distrito de Chimbote, por el norte con el distrito de Santa y por el Oeste con el Océano Pacífico, el cual le ha permitido desarrollarse como Puerto Industrial Pesquero, cobijando en sus periferias a empresas productoras de harina y conserva de pescado, que lo han convertido en la principal actividad laboral con un escaso margen poblacional dedicado a la agricultura. La empresa más importante es la Empresa Pesquera Hayduk S.A, de Harina y Conserva de Pescado que es una de las mayores productoras de derivados de pescado a nivel nacional e internacional.

El distrito de Coishco está alojado dentro la Cuenca Atmosférica, tiene una superficie aproximada de 9.21 km², con una población total de 15,832 hab. [INEI.2013], está rodeado de cerros de alturas promedio 200 a 300 m.s.n.m., lo cual origina que los vientos que van en dirección noreste y los provenientes del oeste, se mantengan por periodos cortos y continuos concentrando los diversos contaminantes emitidos por el

sector industrial, el transporte y el material particulado que es suspendido de las diferentes avenidas sin pavimento y de los cerros que flanquean la ciudad.



Figura N° 4.2: Mapa de Coishco

- **Estación de Chimbote**

Con el desarrollo industrial pesquero y siderúrgico desde los años 50 se dio realce y reconocimiento a la ciudad de Chimbote, pero el crecimiento desordenado y acelerado de su población hicieron de la ciudad, el principal problema de focos infecciosos y contaminación de la provincia por acumulación de desperdicios. El desarrollo de la actividad pesquera (Industria de Harina y Conserva) se ha convertido en el principal sector industrial que genera divisas para el desarrollo económico de la ciudad, pero también es el principal emisor de contaminantes a las zonas urbanas y rurales de la ciudad, acompañado del transporte urbano, compuesto por unidades vehiculares deterioradas sin revisiones técnicas por parte de una empresa, que determine su libre circulación o cese. Posee vientos dominantes en dirección noreste por la mañanas y muy variado al atardecer, que permite que las emisiones del material particulado producidas por las pesqueras se dirijan hacia el distrito de Nuevo Chimbote, este mismo problema lo afronta el distrito de Coishco por las emisiones producidas por la empresa SIDERPERU, que está ubicada al norte de la ciudad de Chimbote. Los problemas de

contaminación de aire, que afronta el distrito son por la emisión del PM10, del Dióxido de Azufre (SO₂) y Dióxido de Nitrógeno (NO₂), emitidos en su mayoría por el consumo de grandes volúmenes de combustibles fósiles como el Petróleo Residual 500 por parte del Sector Industrial Pesquero y Siderúrgico y el Petróleo Diesel2 por el transporte motorizado.



Figura N° 4.3: Mapa de Chimbote

- **Estación de Nuevo Chimbote**

En los años 50, cuando el desarrollo industrial pesquero se manifestó a nivel nacional y se hablaba de una sobrepoblación en el distrito de Chimbote, sus residentes decidieron migrar al sur, donde la presencia de arenales era la única barrera a impedir su población; con el terremoto de 1970, buscando seguridad y estabilidad muchas familias se mudaron a estos arenales, que años más tarde (1994) sería reconocido como distrito con el nombre de Nuevo Chimbote.

La presencia industrial en este distrito solo hace referencia a las empresas conserveras y de tratamiento de aceite crudo de pescado, instalados en las fronteras con el distrito de Chimbote; el problema de contaminación que afronta el distrito se debe a las emisiones de estas empresas que son llevadas a este zona, por el cambio de dirección de los vientos, el material particulado que es trasladado desde las zonas arenosas que rodea al

distrito y el transporte que es el mismo que transita en el distrito vecino de Chimbote. No obstante el diseño de la ciudad, con sus avenidas amplias y asfaltadas, rodeadas de áreas verdes, no permiten la concentración de los contaminantes emitidos o acarreados hacia el interior de la ciudad.



Figura N° 4.4: Mapa de Nuevo Chimbote

Tabla N° 4.1: Características principales de los distritos del área de estudio

Distritos	Capital	Altitud (m.s.n.m)	Superficie (Km ²)	Densidad Poblac. (Hab/Km ²)
Santa	Santa	6	42.23	426
Coishco	Coishco	15	9.21	1610
Chimbote	Chimbote	4	1,467.00	147
Nuevo Chimbote	Buenos Aires	25	389.73	290

4.1.1.2. Identificación de rangos de ICA vs PM10:

El índice de calidad del aire (ICA) es un valor adimensional, calculado a partir de información procedente de la normatividad vigente relacionada con los distintos contaminantes atmosféricos, el cual para nuestro estudio es el PM10, cuyo principal objetivo es facilitar la comprensión de la información relacionada con la contaminación del aire.

Cuanto mayor sea el índice, peor será la calidad del aire. A nivel cualitativo, el rango del Índice de Calidad del Aire, para nuestro país está determinado en la tabla 4.2.

Id	ICA		Estado	PM10		Efectos	Acciones Preventivas
	Minimo	Maximo		Minimo	Maximo		
1	0	50	Bueno	0	50.00	No hay riesgos en la salud	
2	50	100	Moderado	50.00	150.00	Niños, ancianos y personas con enfermedades respiratorias y cardiacas, pueden presentar sintomas tales como tos y cansancio. La población en general no se ve afectada.	
3	100	200	Malo	150.00	250.00	Toda la población puede padecer de sintomas tales como tos seca, ojos cansados, ardor en la nariz y garganta, niños, ancianos y personas con enfermedades respiratorias y cardiacas, los efectos podrían ser mas graves para la salud.	Reducir el esfuerzo físico y pesado al aire principalmente de personas con dolencias cardiacas y pulmonares niños y adultos.
4	200	300	Muy Malo	250.00	420.00	Toda la población puede presentar agravamiento de los sintomas tales como tos seca, ojos cansados, ardor en nariz y garganta. Efectos aun mas graves en la salud de los niños ancianos y personas con problemas cardiovasculares.	Evitar cualquier esfuerzo físico al aire libre, principalmente personas con dolencias cardiacas o pulmonares niños y adultos.
5	300	1000	Alerta Máxima	420.00	1000.00	Toda la población puede padecer riesgos graves y manifestaciones de enfermedades respiratorias y cardiovasculares. Aumento de las muertes prematuras en personas de los grupos mas sensibles.	Todas las personas deben evitar actividades al aire libre.

Tabla N° 4.2. Rango del ICA

4.1.1.3. Variables influyentes en la Calidad del Aire

Para el desarrollo del proyecto se escogieron 9 variables que están directamente vinculadas a la emisión del PM10 y la calidad del aire de las diversas estaciones de monitoreo, que servirán para el modelo de predicción de la red neuronal. Las cuales se encuentran divididas en variables estacionales y variables meteorológicas, que son las siguientes:

Variables estacionales:

Tabla N° 4.3: Variables estacionales

N°	Nombre	Simbología	Unidad
1	Día de la semana	DS	Número
2	Mes del año	MA	Número

Variables Meteorológicas

Tabla N° 4.4: Variables Meteorológicas

N°	Nombre	Simbología	Unidad
1	Dirección predominante del viento:	DV	N°
2	Velocidad del viento predominante	VV	m/s
3	Velocidad promedio viento	V	m/s
4	Velocidad máxima viento	V _{máx}	m/s
5	Temperatura ambiente promedio	T	°C
6	Presión promedio	P	Kpa
7	Humedad promedio	H	%

4.1.2. Análisis y diseño de la Red Neuronal Perceptrón

4.1.2.1. Selección de tecnología

Para el procesamiento de la red neuronal, se hará uso del software Matlab, este brinda la opción a través de un APP (Neural Network Toolbox) para realizar dicho procedimiento o también a través de la línea de comandos (Command Window).

Por tanto, las capacidades de la red van a depender en gran medida de esta fase de entrenamiento. En la fase de entrenamiento es necesario controlar muchos parámetros y distintos algoritmos de optimización, por lo que el usuario de una red neuronal debe tener conocimiento suficiente de cuáles son estos parámetros y cómo funcionan. Por lo que en este proyecto se desarrollará una aplicación dirigida, principalmente, a usuarios que no han trabajado nunca con redes neuronales o no muy expertos, con el fin de que puedan realizar el procesamiento de la red neuronal de manera sencilla y cómoda, y que de esta manera se integre a los otros sistemas que estamos desarrollando (Power Builder y PHP). Dicha aplicación se desarrollará utilizando herramientas de Matlab, tanto las de

la toolbox de redes neuronales, como las herramientas para la programación de ventanas gráficas que dispone (GUI), es decir desarrollaremos este sistema en Matlab.

Por ello, en el desarrollo de la aplicación se ha perseguido que sea fácil de utilizar, que no sea necesario tener vastos conocimientos de redes neuronales y que la visualización de resultados sea clara y concisa.

Pero con la finalidad de realizar una mejor integración con los otros sistemas desarrollados (Power Builder y PHP), y dar una mejor facilidad respecto al uso del sistema, se desarrollará en Matlab un sistema GUI, con interfaces gráficas amigables, de tal manera que este sea el que procese la red neuronal.

4.1.2.2. Identificación de Procesos para la Red Neuronal

- ✓ Seleccionar la estación de monitoreo
- ✓ Importar y visualizar data de entrada
- ✓ Importar y visualizar data de salida
- ✓ Configurar Red
- ✓ Crear red Neuronal
- ✓ Visualizar esquema de red
- ✓ Generar entrenamiento
- ✓ Visualizar esquema de entrenamiento
- ✓ Simular y validar red con data entrada
- ✓ Graficar resultados de procesamiento de red
- ✓ Importar data de entrada para predicción
- ✓ Simular y predecir ICA
- ✓ Realizar aproximación
- ✓ Exportar resultado de predicción obtenida
- ✓ Realizar proceso automático

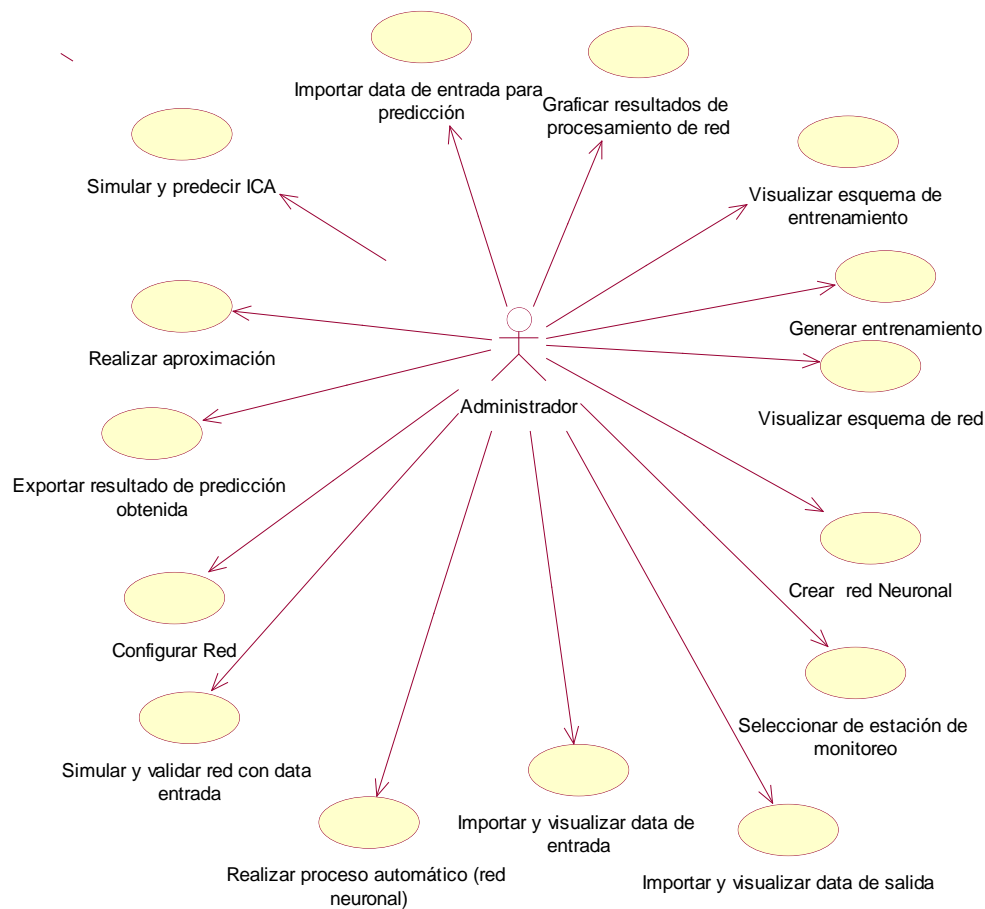


Figura N° 4.5: Procesos red neuronal

4.1.2.3. Identificación de Parámetros de Red Neuronal

Antes de la creación de un Perceptrón Multicapa debe determinarse, de acuerdo a la aplicación, el número de capas, el número de neuronas en cada capa y las funciones de activación. Se puede obtener buenos resultados utilizando dos o tres capas de neuronas.

En la capa de salida se selecciona un número de neuronas igual al número de clases a diferenciar; el número de neuronas en las capas intermedias se selecciona de acuerdo al criterio del diseñador, generalmente se selecciona el doble de las neuronas de la capa de salida más uno; otros criterio son, seleccionar la mitad de la suma del número de entradas con el número de

neuronas en la capa de salida o también seleccionar el número de neuronas en la capa de salida multiplicado por diez.

Generalmente la función de activación usada en las capas de salida es la función sigmoideal logarítmica; en las capas intermedias se usa la función sigmoideal logarítmica o la función tangente hiperbólica. Cuando en la capa de salida se usa la función sigmoideal logarítmica, los valores de salida deben estar entre 0 y 1. Por esta misma razón, cuando se usa en la salida la función sigmoideal tangente hiperbólica deben usarse con valores de salida entre -1 y 1.

Los datos meteorológicos de los diferentes días, que se han venido almacenando y procesando en el sistema de Power Builder, son importados por el sistema de Matlab colocados en un arreglo bidimensional de 9 x n, donde n representa el número de días y 9 el número de variables que han sido identificadas anteriormente (**días, mes, Dirección predominante del viento, Velocidad del viento predominante, Velocidad promedio, Velocidad máxima, Temperatura ambiente promedio, Presión promedio, Humedad promedio**), cada arreglo contiene valores numéricos diferentes, de acuerdo al tipo de datos que contiene, los datos se muestran en la figura 4.5.

Tabla N° 4.5: Data entrada de red

	1	2	3	4	5	6	7	8						
1	4	3	4	2	2	7	1	3						
2	1	7	2	2	3	4	3	3						
3	50	10	30	90	20	50	960	500						
4	60	10	60	60	520	250	950	500						
5	50	10	50	50	510	250	850	500						
6	20	10	10	20	520	30	420	500						
7	50	10	20	50	40	60	20	500						
8	10	10	1020	5520	50	20	6550	500						
9	20	10	10	50	60	50	690	500						
10														
11														

Se creó un Perceptrón Multicapa con un vector de entrada de tamaño 9, dos capas ocultas y la capa de salida de 5 neuronas, la primera capa oculta tenía 90 neuronas y la segunda 50, para calcular el tamaño de la primera capa oculta se tomó el criterio del número de clases multiplicado por diez; el tamaño de la segunda capa oculta se asumió por criterio propio. A todas las neuronas les fue asignada una función de activación sigmoideal logarítmica, ya que las entradas van de cero a uno y las salidas se tomarán en ese rango.

El Perceptrón Multicapa es, entrenada por medio del método gradiente descendiente con velocidad de aprendizaje variable y momento. Ya que se disponían de cinco tipos diferentes de estado para los ICA obtenidos (bueno, moderado, malo, muy malo, alerta máxima), se utilizaron cinco tipos para entrenar la red.

Los arreglos vectoriales que contenían los datos fueron colocados en un arreglo matricial, donde cada columna contenía los datos de entrada, de cada variable seleccionada. El arreglo matricial de salida usado en el entrenamiento contenía dos tipos de valores (0 y 1); los valores de este arreglo se colocaron de acuerdo a los valores deseados de salida, éstos se describen de la siguiente manera:

Tabla N° 4.6: Estado ICA vs salida de red

Estado ICA	Valor de Salida de Red
Bueno	0 0 0 0 1
Moderado	0 0 0 1 0
Malo	0 0 1 0 0
Muy Malo	0 1 0 0 0
Alerta Máxima	1 0 0 0 0

Por lo tanto en base a la clasificación señalada anteriormente, como ejemplo de los datos de la salida para el funcionamiento de la red neuronal queda determinada en la tabla 4.7.

Tabla N° 4.7: Data salida de red

	1	2	3	4	5	6	7	8
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	1	0
4	0	0	0	0	1	0	0	0
5	1	1	1	1	0	0	1	1
6								
7								
8								
9								
10								
11								

La red será entrenada después de 2300 iteraciones, con un error máximo permisible de 9.978×10^{-6}

4.1.3. Implementación del Sistema para Procesamiento de la Red Neuronal

La red neuronal, se implementará primero usando la herramienta nativa de Matlab nntool, para poder validar y generar la red neuronal, que genere resultados óptimos para luego migrarlos a través de un sistema en Matlab con interfaces de usuario amigables para que a través de este se realice el procesamiento.

4.1.3.1. Implementación de la Red Neuronal a través de la herramienta nntool.

4.1.3.1.1. Creación de la Red Neuronal.

En la figura 4.6., se muestra Matlab 2013, en su pantalla principal.

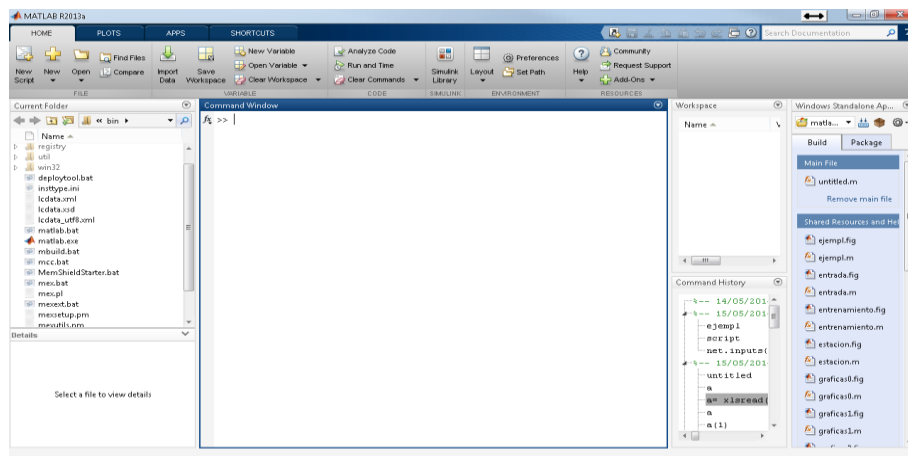


Figura N° 4.6: Creación de red neuronal

Enseguida se procede a llamar a la herramienta nntool, para crear y procesar la red neuronal.

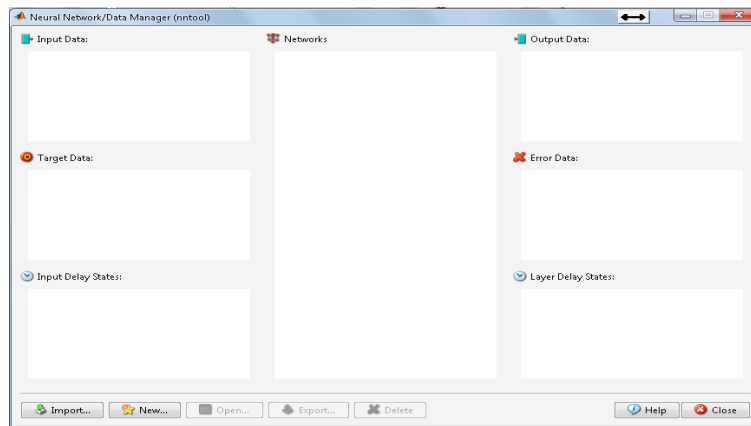


Figura N° 4.7: Herramienta nntool

Se muestra la pantalla principal de la herramienta, donde se aprecia el campo para data de entrada, data de salida redes neuronales, data objetivo, data error, estado de espera de entrada y estado de espera de capa. Ver figura 4.7.

Para crear una red Perceptrón Multicapa debe presionarse el botón New Network, luego aparecerá la siguiente ventana.

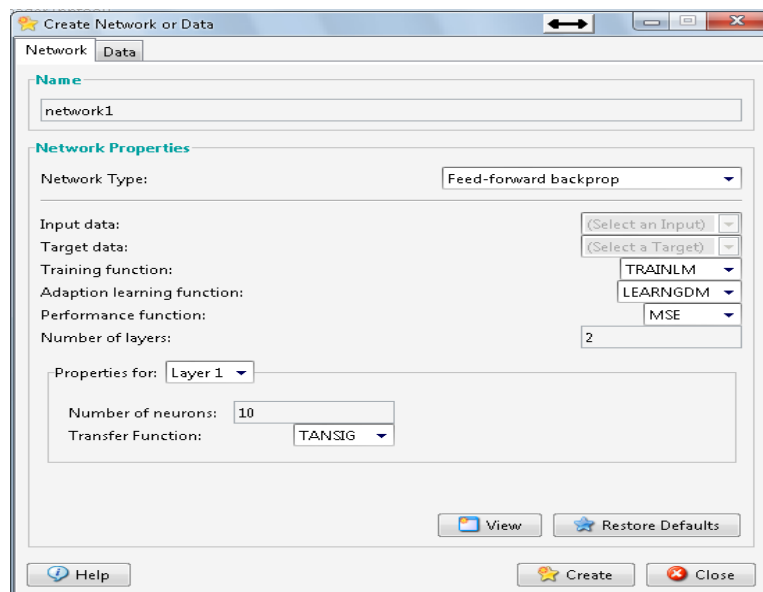


Figura N° 4.8: Nueva red neuronal

Se le asigna un nombre a la red por medio del cuadro de texto Network Name, que se encuentra en la parte superior de la ventana. Figura 4.8.

En el menú descendente Network Type, se selecciona Feed-forwrd backpropagation (Alimentación hacia adelante, propagación hacia atrás), que es el nombre que Matlab da al Perceptrón Multicapa.

En el cuadro de texto Input ranges (Rangos de entrada), se ingresa un arreglo $n = 2$, que indica los rangos de los valores del vector de entrada.

También puede seleccionarse en el menú descendente Get From Input, uno de los arreglos de entrada disponibles en la interfaz, para obtener los valores máximos y mínimos a partir de este. Además se seleccionó el algoritmo de entrenamiento (Training funtion). En el cuadro de texto number of layers (número de capas), se escribe la cantidad de capas que tendrá la red. Mediante el menú descendente properties for layer (propiedades de la capa) se indica el número de neuronas en cada capa, así como su función de activación.

4.1.3.1.2. Entrenamiento de la Red

Después de crear la red se procede a su entrenamiento; para entrenar un Perceptrón Multicapa debe seleccionarse la red en el cuadro Networks y luego presionar el botón train; al hacer esto aparecerá la figura 4.9.

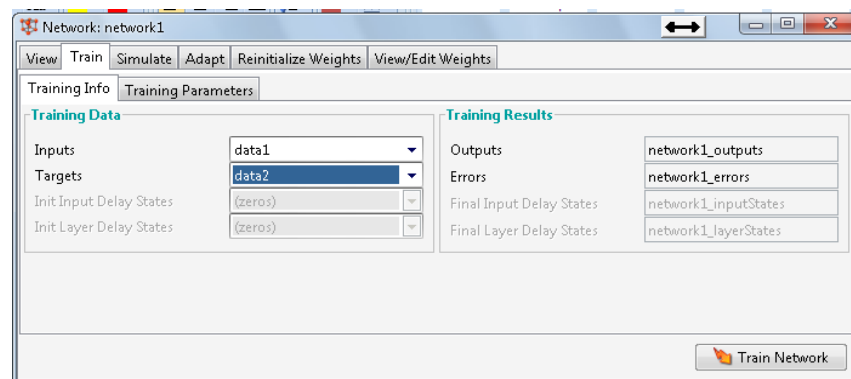


Figura N° 4.9: Entrenamiento de red neuronal

Para el entrenamiento es necesario un grupo de entradas y salidas de ejemplo; para nuestro proyecto esta data, deberá ser procesada a través del sistema inicial de Power Builder, el cual exportará ambas datas en archivos de Excel, para su importación a arreglos. En el menú descendente Inputs (entradas), se selecciona el arreglo que contiene todas las entradas, cada columna en el arreglo representa una entrada, en el menú descendente Targets (objetivos), se selecciona el arreglo que contiene todas las salidas correspondiente a las entradas, si en el caso los arreglos de entrada y salida no poseen el mismo número de columnas se generará un mensaje de error al tratar de entrenar la red. Después de seleccionar los arreglos que contienen las entradas y salidas, se presiona el botón Train Network (Entrenar red) para iniciar el entrenamiento del Perceptrón Multicapa.

Durante el entrenamiento se muestra una gráfica que muestra el error total en las salidas en función de las iteraciones realizadas. Por medio de la pestaña, Training Parameters (Parámetros de entrenamiento), se puede cambiar los parámetros del algoritmo de entrenamiento, entre estos están el epochs: que indica el número máximo de iteraciones, por defecto es cien iteraciones. También está el parámetro goal (meta), que indica el error máximo aceptable, por defecto es cero, pero generalmente es muy difícil llegar a este valor de error, por lo que se recomienda cambiarlo. Al finalizar el entrenamiento se muestra en la figura 4.10.

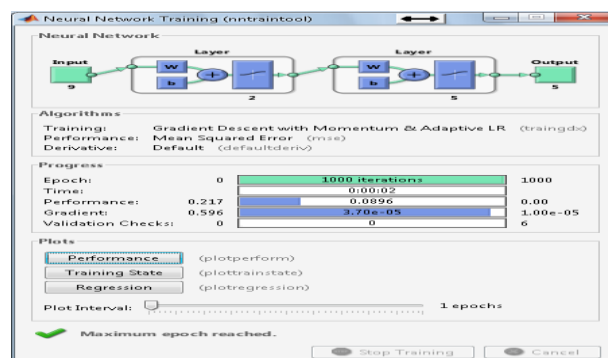


Figura N° 4.10: Resultados de entrenamiento

4.1.3.1.3. Simulación de la Red.

La simulación sirve para poder realizar la predicción, y consiste en aplicar una entrada a la red ya entrenada, para obtener una salida, poniendo en funcionamiento la red neuronal. Antes de simular la red deberá observarse que esta ya ha sido entrenada exitosamente, de forma que el error promedio en las salidas sea tolerable, este valor es mostrador en la ventana de comandos al finalizar el entrenamiento de la red.

Para simular el Perceptrón Multicapa, debe seleccionarse la red ya entrenada y luego presionar el tab simulate (simular); al hacer esto aparecerá la figura 4.11.

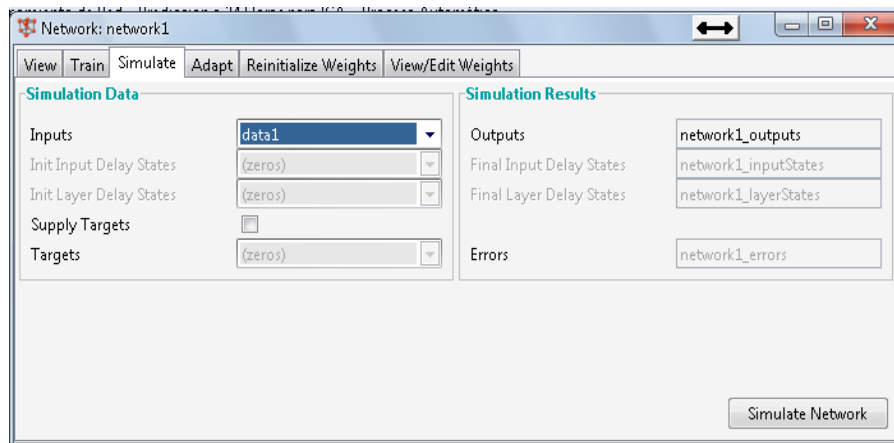


Figura N° 4.11: Simulación de red neuronal

En el menú descendente Inputs (entrada), se selecciona el arreglo que contiene todas las entradas, cada columna es una entrada diferente.

El cuadro de texto Outputs (Salidas), se escribe el nombre del arreglo que contendrá las salidas de la red, correspondientes a la entrada de la simulación. El arreglo salida tendrá el mismo número de columnas que el arreglo entrada, ya que cada columna en el arreglo de entrada genera una columna en el arreglo salida. Para proceder con la simulación se presiona el botón Simulate Network (Simular red). La interfaz gráfica de redes neuronales también permite importar o exportar redes neuronales o arreglos al espacio de trabajo de Matlab.

La herramienta nntool, es una buena opción para procesar redes neuronales, pero el tiempo necesario para crear, entrenar o simular una red neuronal, es mucho mayor que si contáramos con un sistema con interfaces implementado en Matlab, para procesar la red neuronal para el desarrollo del proyecto, por lo que habiendo validado la red neuronal, ahora se procede a implementarlo pero a través del sistema GUI.

4.1.3.2. Implementación de la Red Neuronal por medio de un Sistema GUI en Matlab.

Con la finalidad de que el sistema integral pueda ser usado por personal, con poco de conocimiento en redes neuronales se implementa la red neuronal a través de un sistema con interfaces.

4.1.3.2.1. Funcionamiento del sistema

- a. **Selección de estación de monitoreo.-** Al cargar el Sistema de Matlab, se tendrá que seleccionar la estación de trabajo, con la cual se procesará toda la red neuronal.

- **Código:**

```
% --- De acuerdo al botón que se presione almacenará el nombre de la
estacion en una variable.
```

```
function pushbutton1_Callback(hObject, eventdata, handles)
global estaciones
estaciones='nuevo_chimbote';
close(estacion);
```

```
function pushbutton2_Callback(hObject, eventdata, handles)
global estaciones
estaciones='chimbote';
close(estacion);
```

```
function pushbutton3_Callback(hObject, eventdata, handles)
global estaciones
estaciones='coishco';
close(estacion);
```

```
function pushbutton4_Callback(hObject, eventdata, handles)
global estaciones
estaciones='santa';
```

close(estacion);

- **Interface:**

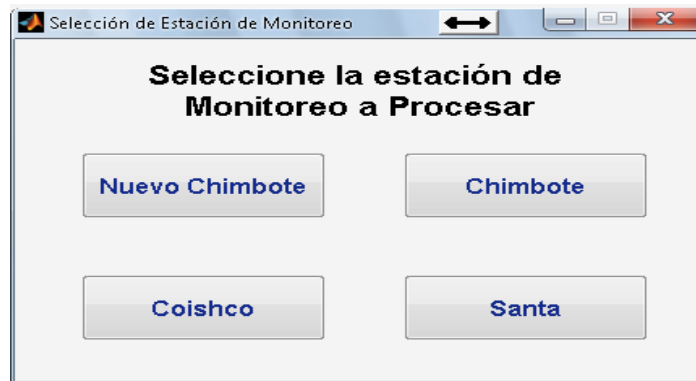


Figura N° 4.12: Selección de estación

- b. Pantalla principal del Sistema en Matlab:** Una vez seleccionado la estación el sistema cargará todo el panel principal, figura 4.13.

- Código:

```
pantalla_principal % ---abre panel de control
```

- Interface:

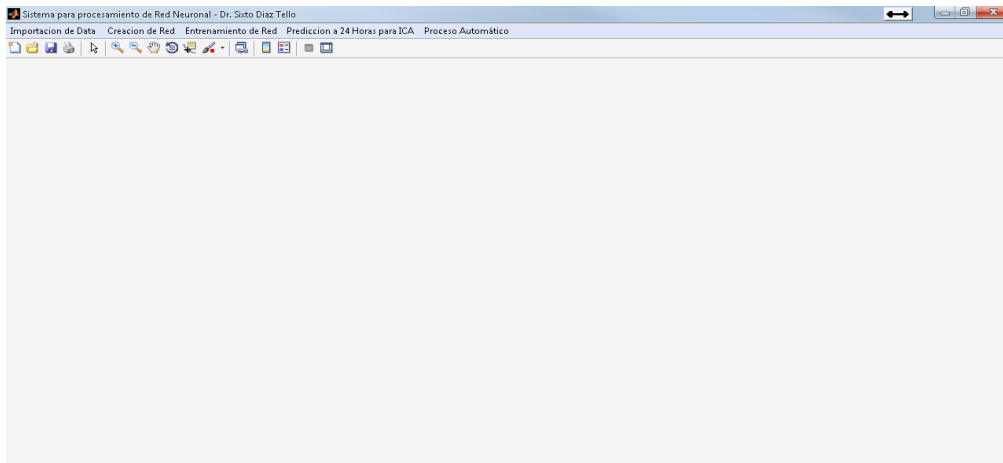


Figura N° 4.13: Pantalla principal

- c. Importar y visualizar Data entrada.-** importa la data input de entrada para que sea usada en el entrenamiento de la red neuronal, la cual ha sido exportado por el sistema de Power Builder, en un archivo excel.

- **Código:**

```
function entrada_Callback(hObject, eventdata, handles)
global A;
global estaciones;
% --- Lee el archivo que ha sido exportado por el sistema de PB
ruta=strcat('C:\ambiental\exportaciones\matlab\',estaciones,'\entrada.xls');
% --- Almacena los datos del archivo en un arreglo
A= xlsread(ruta);
% --- Abre la ventana de entrada para mostrar resultados
entrada;
```

- **Interface:**

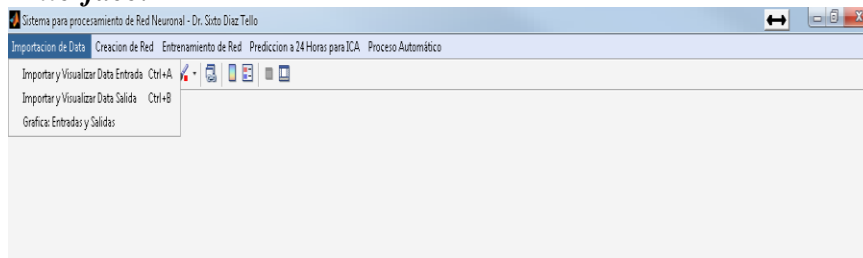


Figura N° 4.14: Importar data entrada

d. Resultado de la importación data entrada: como resultado de la importación, el sistema almacenará dicho arreglo en una variable global y mostrará en pantalla el arreglo “9 x n” y un mensaje satisfactorio.

- **Código:**

```
function entrada_OpeningFcn(hObject, eventdata, handles, varargin)

global A;
global estaciones;
% --- invierte filas por columnas.
A=A';
% --- Imprime en pantalla el arreglo.
set(handles.uitable1,'Data',A);
% --- mensaje satisfactorio
warndlg(['Archivo de entrada:
C:\ambiental\exportaciones\matlab\',estaciones,'\entrada.xls', ' importado con
exito'],'Sistema');
```

- **Interface:**

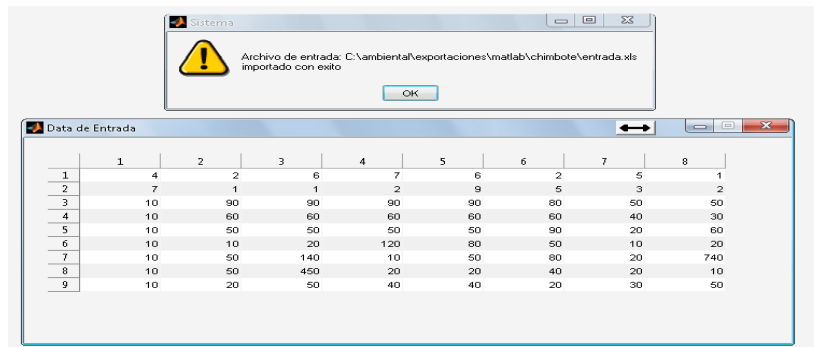


Figura N° 4.15: Resultados de importación

e. **Importación y visualización de data salida.**- importa la data input de salida para que sea usada en el entrenamiento de la red neuronal, la cual ha sido exportado por el sistema de Power Builder en un archivo de excel.

- **Código:**

```
function salida_Callback(hObject, eventdata, handles)
global B
global estaciones;
% --- Lee archive que contiene data de salida, que ha sido exportado por
PB
ruta=strcat('C:\ambiental\exportaciones\matlab\',estaciones, 'salida.xls');
% --- Almacena los datos en variable global
B= xlsread(ruta);
salida;
```

- **Interface**

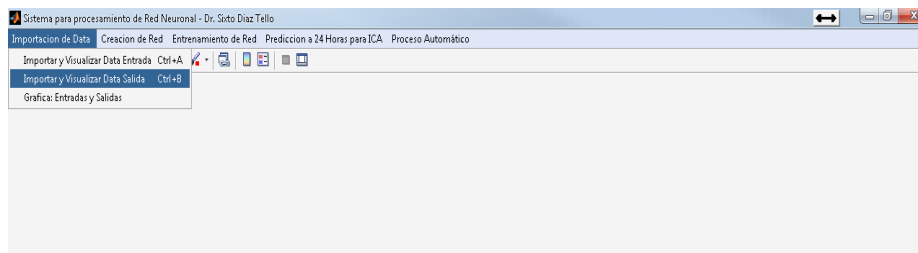


Figura N° 4.16: Importación de data salida

f. **Resultado de la importación de data salida:** como resultado de la importación, el sistema almacenará dicho arreglo en una variable global y

mostrará en pantalla el arreglo “5 x n” y un mensaje satisfactorio. Cabe mencionar que a cada salida importada le corresponde cada entrada.

- **Código:**

```
function salida_OpeningFcn(hObject, eventdata, handles, varargin)
% --- Recupera variable que contiene datos de salida
global B;
global estaciones;
% --- Invierte filas por columnas
B=B'
% --- Muestra en pantalla el arreglo
set(handles.uitable1,'Data',B);
% --- Mensaje satisfactorio.
warndlg(['Archivo de salida: C:\ambiental\exportaciones\matlab\' de salida:
C:\ambiental\exportaciones\matlab\',estaciones,\"salida.xls\",' importado
con éxito'],'Sistema');
```

- **Interface:**

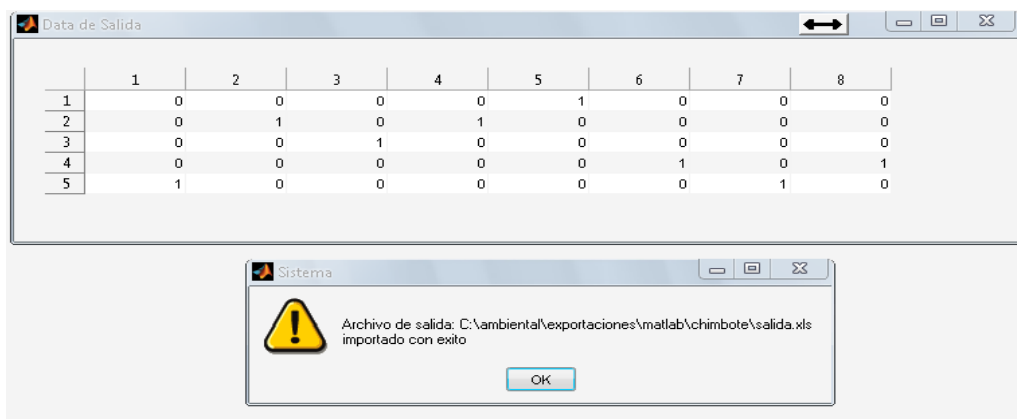


Figura N° 4.17: Resultados data salida

g. Gráfica entradas y salidas: Se realiza una primera gráfica de las entradas vs. las salidas importadas para la estación seleccionada

- **Código:**

```
% --- recupera arreglo de entrada
global A;
% --- recupera arreglo de salida
global B;
% --- grafica salidas
plot(B, 'o')
hold on
% --- adiciona a gráfica entradas
plot(A, '+r')
```


- **Interface:**

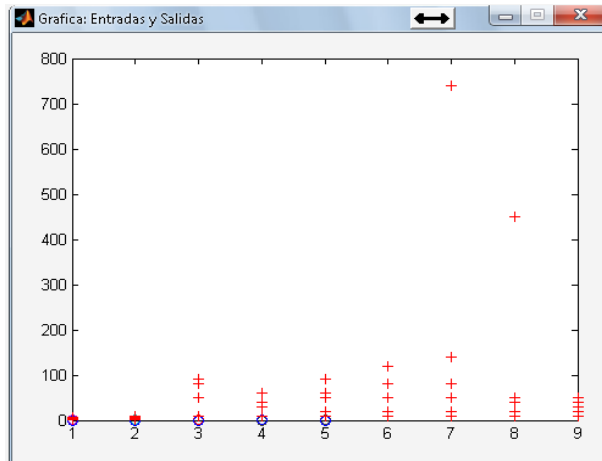


Figura N° 4.18: Gráficas entrada y salida

h. Configuración de red: sirve para que en el caso, el administrador desee, configure de la manera que desee la red neuronal, con los parámetros óptimos que se crea conveniente. Cabe mencionar que la red, ya está inicializada con los parámetros óptimos, para su correcto funcionamiento. Se puede modificar el número de interacciones, la tasa de aprendizaje, el factor momento y el error máximo permitido, los cuales son utilizados al realizar el entrenamiento.

- **Código:**

```
function pushbutton3_Callback(hObject, eventdata, handles)
% --- Crea variables de configuración
global iteraciones;
global aprendizaje;
global factor;
global error;
% --- Almacena valores ingresados en variables
iteraciones= get(handles.edit1,'String');
iteraciones= str2double(iteraciones);
aprendizaje= get(handles.edit2,'string');
aprendizaje= str2double(aprendizaje);
factor= get(handles.edit3,'string');
factor= str2double(factor);
error= get(handles.edit4,'string');
error= str2double(error);
```

- **Interface:**



Figura N° 4.19: Configuración de red neuronal

- i. **Creación de red:** esta opción sirve para poder crear la red neuronal, que servirá luego para realizar el entrenamiento y simulación (predicción), para esto almacena las variables y parámetros en variables globales.

- **Código:**

```
function c_crear_Callback(hObject, eventdata, handles)
% --- Crea variable para red neuronal que será trabajada en todo
el sistema
global red
% --- Invoca a la funcion newff de creación de red
red = newff([0 1; 0 1], [2, 5], {'logsig','logsig'}, 'traingdx');
red.inputs{1}.size=9
% --- Mensaje satisfactorio
msgbox('Red Neuronal "RED" creada con exito','Sistema');
```

- **Interface:**

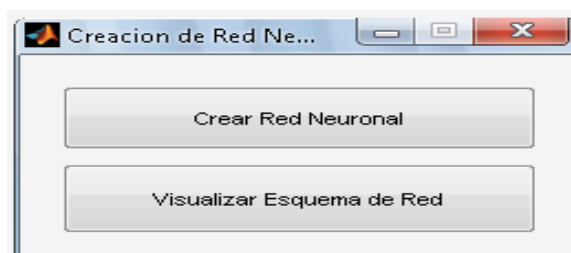


Figura N° 4.20: Crear red neuronal

- j. **Mensaje satisfactorio de creación de red**

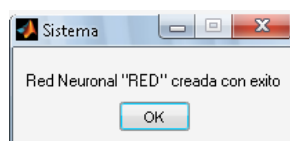


Figura N° 4.21: Mensaje satisfactorio

k. **Visualización del esquema de red:** para esta opción se puede acceder a través de la ventana anterior o también por medio del menú superior. Mediante esta se llama a la función `view()`, la cual permite visualizar el esquema de la red (número de neuronas en la capa de entrada, salida oculta, entre otros parámetros).

- **Código**

```
% --- Recupera variable de red
global red;
% --- Invoca a función view, para ver esquema
view(red);
```

- **Interface**

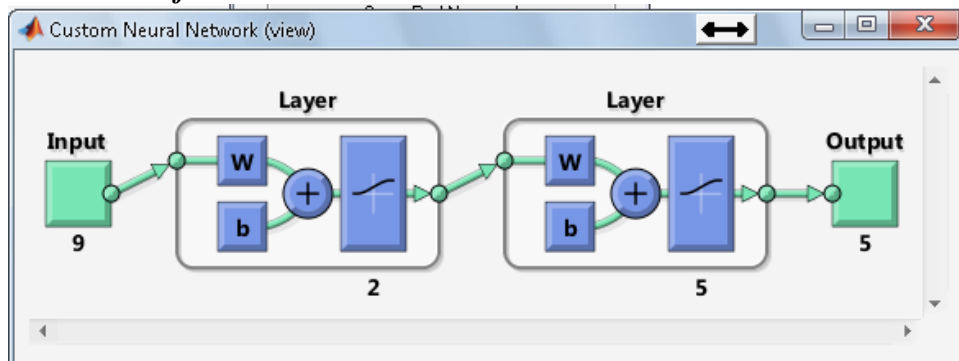


Figura N° 4.22: Esquema de red neuronal

l. **Entrenar red.-** tiene dos opciones la primera realiza el entrenamiento con los parámetros por defecto que tiene asignado ya el sistema, y la segunda genera el entrenamiento con la configuración que asignó el administrador, en una ventana que se presentó anteriormente. Al momento de realizar el entrenamiento hace el llamado a las variables que contienen los arreglos de entrada y salida.

- **Código:**

```
% --- Entrenamiento con parámetros estándares
function c_entrenar_Callback(hObject, eventdata, handles)
% --- Recupera variables y data de entrada y salida
global A;
global B;
global red;
global tr;
```

```

% --- Llama a function train(), que sirve para realizar el
entrenamiento
[red,tr] = train(red, A, B);
% --- Entrenamiento con parámetros de configuración del usuario
function pushbutton2_Callback(hObject, eventdata, handles)
% hObject handle to pushbutton2 (see GCBO)
% eventdata reserved - to be defined in a future version of
MATLAB
% handles structure with handles and user data (see GUIDATA)
global iteraciones;
global aprendizaje;
global factor;
global error;
global A;
global B;
global red;
global tr;
% --- Configura parámetros en variables de red
red.trainParam.epochs=iteraciones;
red.trainParam.lr=aprendizaje;
red.trainParam.mc=factor;
red.trainParam.goal=error;
% --- Entrena red con function train()
[red,tr] = train(red, A, B);

```

- **Interface:**

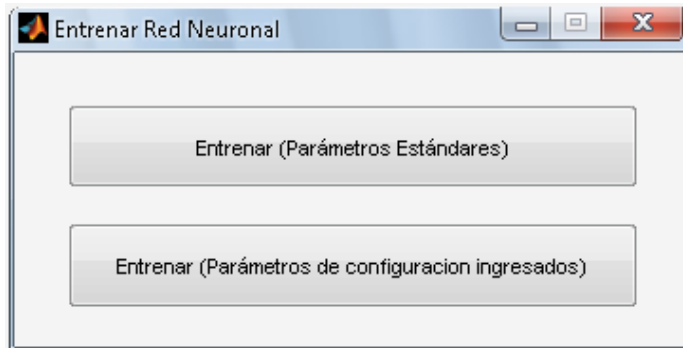


Figura N° 4.23: Entrenar red neuronal

m. Visualización de esquema de entrenamiento de red (nntraintool):

Automáticamente cuando se está realizando el entrenamiento, se invoca a la función `nntraintool()`, mediante el cual se visualiza el esquema de entrenamiento, donde se observa el progreso de las iteraciones, entre otros. A esta opción también puede ser luego accedida a través del menú superior.

- **Código:**

```
% --- Invoca a la function nntaintool(), que sirve para visualizar el progreso de entrenamiento.
nntaintool
```
- **Interface:**

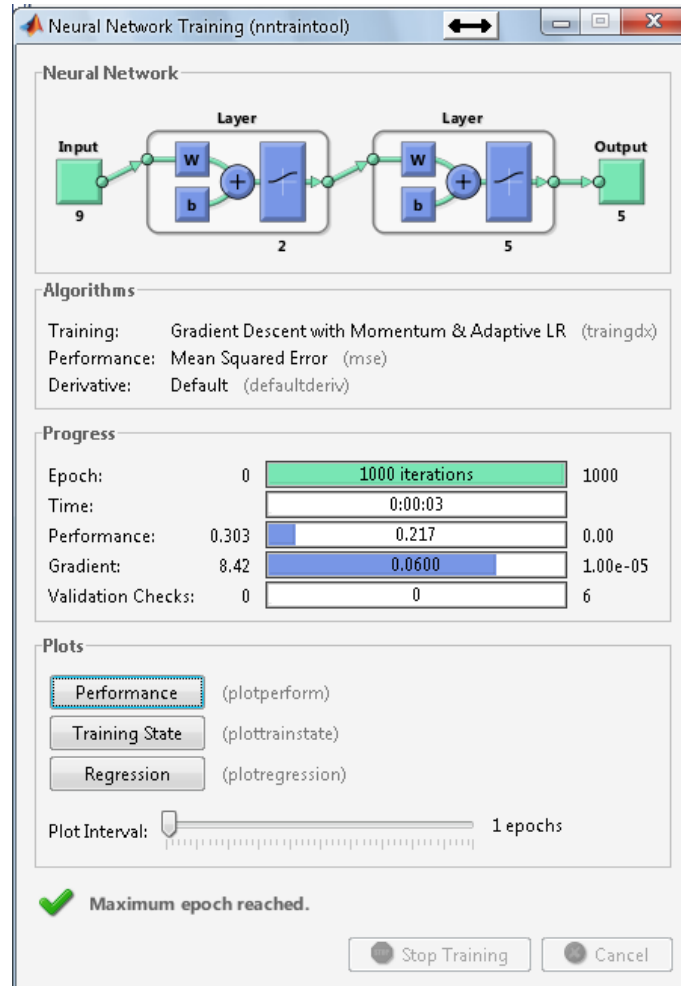


Figura N° 4.24: nntaintool

- Simulación y validación de red con datos de entrada:** sirve para realizar una primera simulación de la red que ya ha sido entrenada, pero como dato de entrada se tendrá a los mismo valores input entrada, a fin de poder verificar que las salidas estimadas, sean semejantes a las salidas deseadas. De ser el caso que ambas salidas se asemejan, quiere decir que la red se encuentra en óptimas condiciones para realizar la predicción.

- **Código:**

```
% --- Abre ventana de simulación y validación
Simulation
```
- **Interface:**

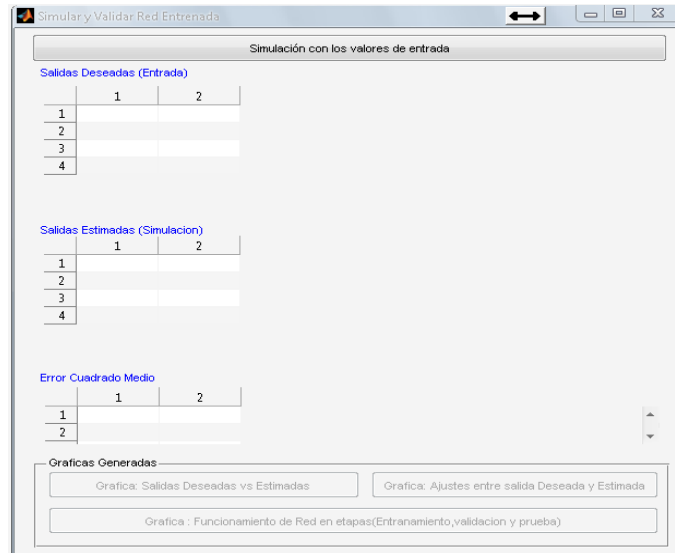


Figura N° 4.25: Simulación y validación de red neuronal

- o. **Resultado de la simulación y validación:** Como resultado de dicha simulación y validación, se tendrá las salidas deseadas, salidas estimadas, y el error cuadrado medio para cada columna del arreglo; adicionalmente se habilitarán los botones de las gráficas de los resultados.

- **Código:**

```
function simulacion_Callback(hObject, eventdata, handles)
% --- Recupera variables globales
global A;
global B;
global Y;
global red;
global d;
global mse;
% --- Invoca a la function sim(), para realizar la simulacion con la
data de entrada
Y = sim(red,A);
% --- Muestra en pantalla salida deseada
set(handles.uitable1,'Data',B);
% --- Muestra en pantalla salida estimada
set(handles.uitable2,'Data',Y);
% --- Calcula error cuadrado medio
```

```

d=(Y-B).^2;
mse=mean(d);
% --- Imprime resultados
set(handles.uitable3,'Data',mse);
% --- Habilita graficas
set(handles.pushbutton3,'Enable','on')
set(handles.pushbutton4,'Enable','on')
set(handles.pushbutton5,'Enable','on')

```

- **Interface:**

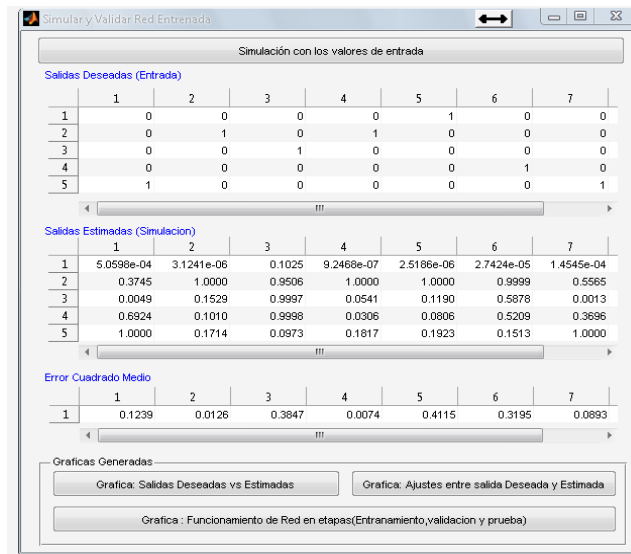


Figura N° 4.26: Resultados de simulación

p. **Resultado: Gráfica Salidas deseadas vs estimadas:** Estos gráficos muestran gráficamente la ubicación de los valores estimados y las salidas deseadas que sean producidas por la red como aparece en la figura (*nota:* sus resultados pueden diferir ligeramente de acuerdo con el valor de los pesos calculados)

- **Código:**

```

function graficas1_OpeningFcn(hObject, eventdata, handles,
varargin)
% --- recupera datos de variables de salidas deseadas y estimadas
global Y;
global B;
% --- Grafica salidas deseadas
plot(Y,'o');
hold;
% --- Adiciona a gráfica salidas estimadas
plot(B,'*');

```

```
legend('Valores  
Deseadas','Location','NorthWest');
```

```
Estimados','Salidas
```

- **Interface:**

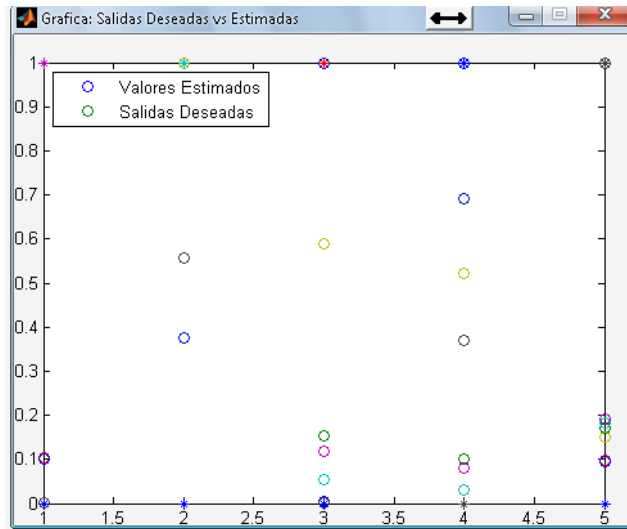


Figura N° 4.27: Gráfica salidas deseadas vs estimadas

q. **Resultado: Gráfica funcionamiento de red en etapas.:** Este comando muestra gráficamente el funcionamiento de la red en las etapas de entrenamiento, validación y prueba, así como el mejor resultado durante estas fases como se muestra en la figura

- **Código:**

```
function graficas2_OpeningFcn(hObject, eventdata, handles,  
varargin)  
% --- Obtiene variable de entrenamiento  
global tr;  
% --- Grafica resultados de las etapas de entrenamiento  
plotperform(tr),
```

- **Interface:**

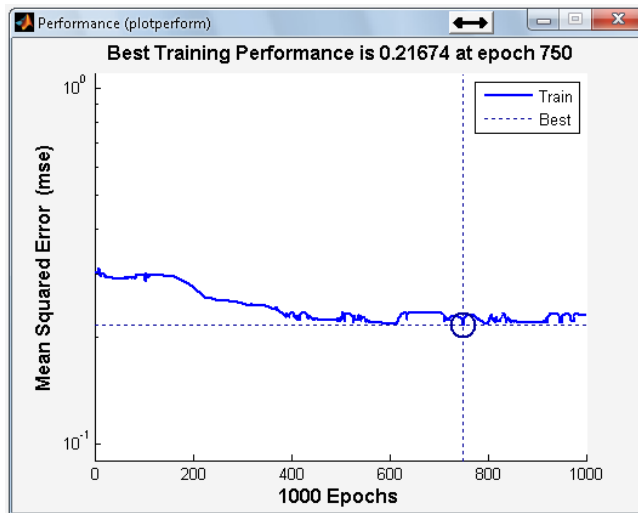


Figura N° 4.28: Gráfica de funcionamiento

r. **Resultado** : Gráfica Ajustes entre salida deseada y estimada: para observar el ajuste entre los valores estimados y predichos

- **Código:**

```
function graficas3_OpeningFcn(hObject, eventdata, handles,
varargin)
% --- Recupera arreglos de salidas deseadas y estimadas
global Y;
global B;
% --- Grafica ajuste entre salidas estimadas y deseadas
plotregression(Y,B);
```

- **Interface:**

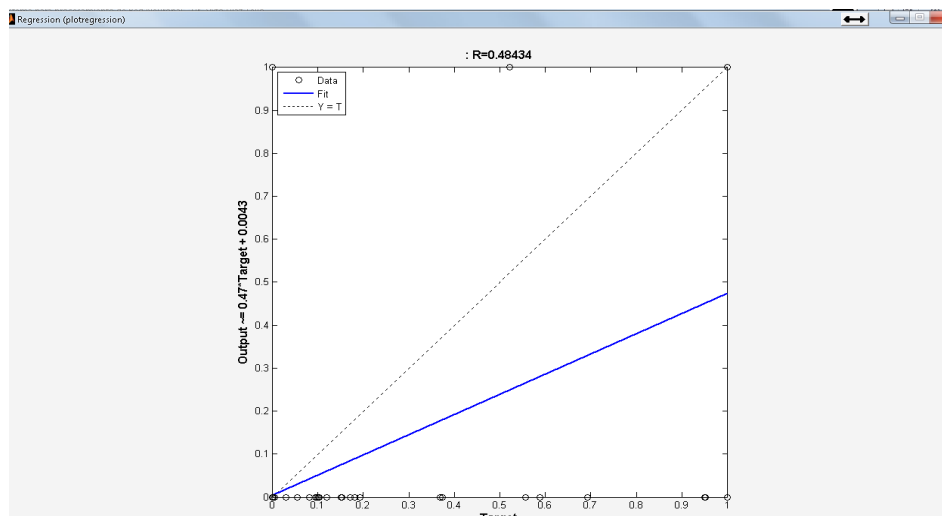


Figura N° 4.29: Gráfica ajustes entre salidas

s. **Valores y parámetros obtenidos de simulación:** se visualizan los valores obtenidos de la simulación, que resulta el valor que toman los pesos de entrada y de capa así como los biases una vez realizado el ajuste de la red que se encuentran almacenados en las variables (pesos de Entrada, Pesos de Capa, red.b{1} y red.b{2}).

- **Código:**

```
function parametroso_OpeningFcn(hObject, eventdata, handles,
varargin)
% --- Crear variables para almacenamiento de resultados
global pesosEntrada;
global pesosCapa;
global red;
% --- Valor que toman los pesos de entrada y de capa así como
los biases una vez realizado el ajuste de la red que se encuentran
almacenados en las variables
pesosEntrada=red.IW{1,1};
pesosCapa=red.LW{2,1};
set(handles.uitable1,'Data',pesosEntrada);
set(handles.uitable2,'Data',pesosCapa);
set(handles.uitable3,'Data',red.b{1});
set(handles.uitable4,'Data',red.b{2});
```

Cabe mencionar que si se modifica la arquitectura de la red, esto es el número de neuronas en la capa oculta (net.layers{1}.size) así como el número de iteraciones (net.trainParam.epochs), los valores de la tasa de aprendizaje (net.trainParam.lr) y el coeficiente momento (net.trainParam.mc), se observará que la variación que se refleja en el error cuadrado medio y en los pesos de entrada y de capa.

- **Interface:**

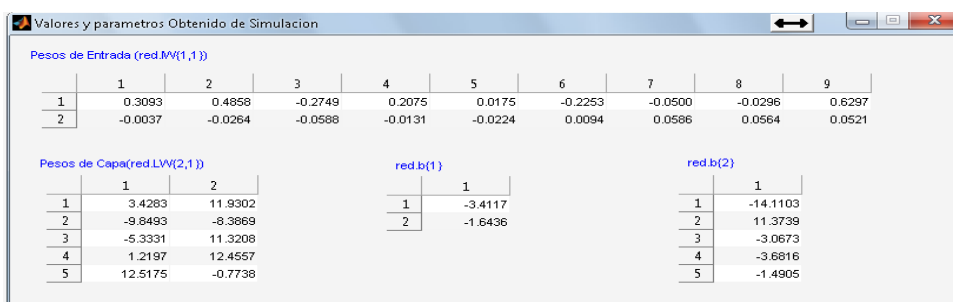


Figura N° 4.30: Valores de predicción

t. **Importación de data para predicción:** Sirve para importar la data que contiene las entradas para realizar la simulación y predicción a 24 horas, que a la vez es almacenada en un arreglo de 9 x n. Cabe mencionar que dicha data ha sido generada por el sistema de Power Builder, y que realizó la exportación de dicha data a través de un archivo Excel.

- **Código**

```
% --- Importación de data de predicción
function Untitled_10_Callback(hObject, eventdata, handles)
% --- Crear variable para almacenar datos de entrada
global C;
global estaciones;
% --- Recupera ruta de archivo que exporto el sistema de PB
ruta=strcat('C:\ambiental\exportaciones\matlab\',estaciones,'prediccion.xls');
% --- Lee y guarda los datos en la variable global
C= xlsread(ruta);
% --- Abre ventana para visualización de datos
prediccion;

% --- Visualización de data de entrada para predicción
function prediccion_OpeningFcn(hObject, eventdata, handles, varargin)
% --- recupera variable que contiene data para predecir
global C;
% --- Intercambia filas por columnas
C=C';
% --- Muestra en pantalla el arreglo
set(handles.uitable1,'Data',C);
```

- **Interface**

	1
1	5
2	6
3	20
4	20
5	20
6	10
7	10
8	120
9	10

Figura N° 4.31: Importar data para predicción

u. Simulación y predicción de ICA: Sirve para realizar la simulación (predicción a 24hrs), con respecto a la data de entrada que ha sido importada anteriormente.

- **Código:**

```
function pushbutton1_Callback(hObject, eventdata, handles)
% --- Recupera variables
global C;
global D;
global red;
% --- invoca a función sim() y realiza simulación como data de
entrada la que se importó para la predicción
D = sim(red,C);
% --- Muestra en pantalla resultados de simulacion
set(handles.uitable1,'Data',D);
% --- Habilita botones de aproximación y visualización de gráfica
set(handles.pushbutton2,'Enable','on')
set(handles.pushbutton3,'Enable','on')
```

- **Interface:**

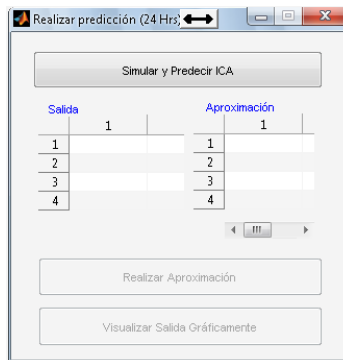


Figura N° 4.32: Simular y predecir ICA

v. Resultados de simulación y realización de aproximación: Una vez que se ha realizado la simulación, se cuenta con resultados, pero que no están en función de 0 y 1; por lo que se realiza el procedimiento ‘Aproximación’, a fin de ya tener un resultado más entendible por el sistema web, quien es el que procesará los resultados para informarlo al público en general, a través de un sistema geográfico.

- **Código:**

```
function pushbutton3_Callback(hObject, eventdata, handles)
global D;
global E;
global indica;
indica='falso';
% --- Parámetros mínimos y máximos que servirán como
indicadores, para determinar si el digito resultado es aceptado
como resultado o no.
min=0.30;
max=0.75;
% --- Analiza el primer digito
if D(1,1) < min
    E(1,1)=0;
elseif D(1,1) > max
    E(1,1)=1;
else
    E(1,1)=9.9999;
    indica='verdad';
end
% --- Analiza el segundo digito
if D(2,1) < min
    E(2,1)=0;
elseif D(2,1) > max
    E(2,1)=1;
else
    E(2,1)=9.9999;
    indica='verdad';
end
% --- Analiza el tercer digito
if D(3,1) < min
    E(3,1)=0;
elseif D(3,1) > max
    E(3,1)=1;
else
    E(3,1)=9.9999;
    indica='verdad';
end
% --- Analiza el cuarto digito
if D(4,1) < min
    E(4,1)=0;
elseif D(4,1) >max
    E(4,1)=1;
else
    E(4,1)=9.9999;
    indica='verdad';
end
```

```

% --- Analiza el quinto dígito
if D(5,1) < min
    E(5,1)=0;
elseif D(5,1) > max
    E(5,1)=1;
else
    E(5,1)=9.9999;
    indica='verdad';
end
% --- Imprime todo el arreglo de aproximación en pantalla
set(handles.uitable2,'Data',E);
% --- Validación para muestra de mensaje
if strcmp(indica,'verdad')
    msgbox('Datos ingresados no se pudieron validar con la Red
Neuronal(Fuera de rango)','Sistema');
else
    warndlg('Aproximaciones realizadas con éxito','Sistema');
end

```

- **Interface**

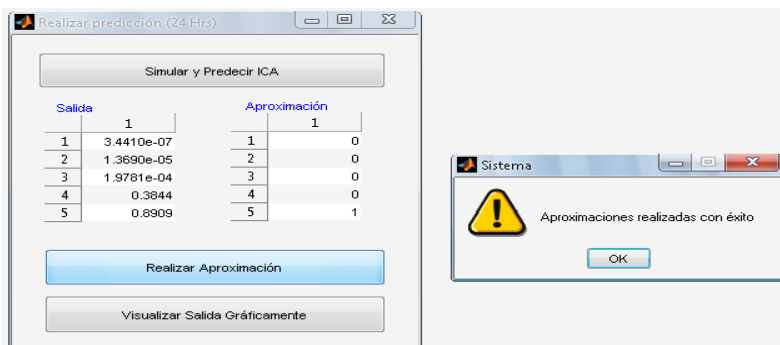


Figura N° 4.33: Resultados simulación y aproximación

w. Exportación de salida obtenida (resultado y aproximación): esta opción sirve para exportar los resultados de la simulación (resultado inicial y aproximado), en archivos separados y que serán importados por el sistema web.

- **Código:**

```

function Untitled_13_Callback(hObject, eventdata, handles)
% --- recupera variables de arreglo de resultado inicial y resultado
aproximación
global D;
global E;
global estaciones;
global indica;

```

```

% --- Verifica si la aproximación obtenida cumple con los rangos
del ICA, realizará este procedimiento solo si cumple dicha
coincidencia
if strcmpi(indica,'verdad')
    msgbox('No se puede exportar, porque datos ingresados no se
    pudieron validar con la Red Neuronal(Fuera de rango)','Sistema');
else
    % --- Captura rutas para guardar ambos archivos
    ruta=strcat('C:\ambiental\exportaciones\matlab\',estaciones,'\result
    ado.csv');
    ruta2=strcat('C:\ambiental\exportaciones\matlab\',estaciones,'\res_
    inicial.csv');
    % --- Escribe en ambos archivos y los guarda en el disco duro
    xlswrite(ruta, E, 'Hoja1','A1');
    xlswrite(ruta2, D, 'Hoja1','A1');
    % --- Muestra mensaje satisfactorio de exportación.
    warndlg(['Archivo de RESULTADO(Inicial y aproximado),
    generado con exito.',ruta2, ' y ',ruta'],'Sistema');
end

```

- **Interface:**

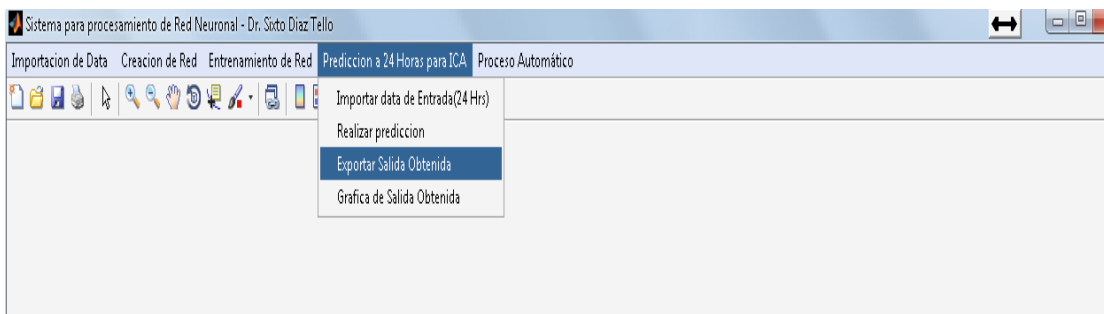


Figura N° 4.34: Exportación de salida

x. Mensaje satisfactorio

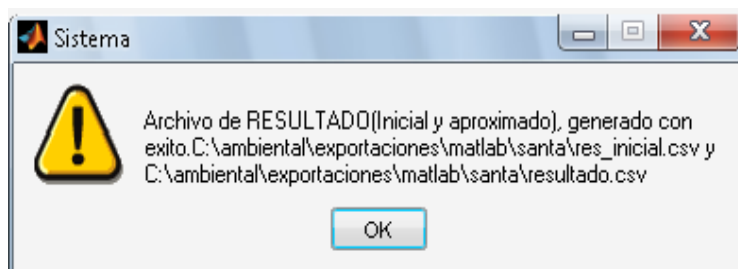


Figura N° 4.35: Mensaje satisfactorio

y. Resultado: gráfica salida obtenida

- **Código:**

```
function graficas4_OpeningFcn(hObject, eventdata, handles,  
varargin)  
% --- Recupera variable de arreglo de data resultado  
global D;  
% --- Grafica resultado  
plot(D,'o')
```

- **Interface:**

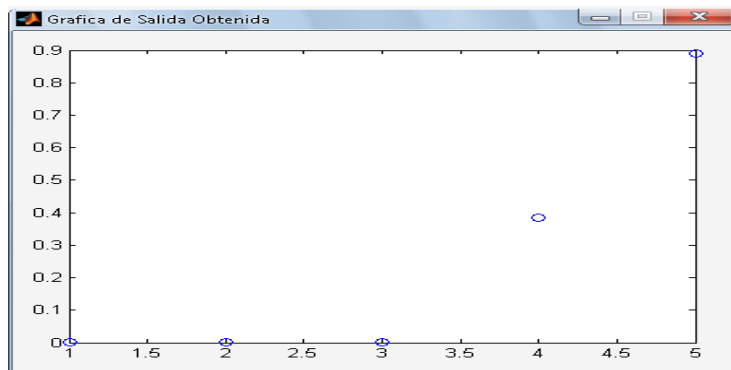


Figura N° 4.36: Gráfica salida obtenida

z. Procesamiento automático: Esta opción sirve para que realice la importación, procesamiento, creación, entrenamiento, simulación, predicción, hasta los resultados de la red neuronal en un solo click. Con la finalidad de abreviar tiempo, pero ya no se verán los resultados que van emitiendo en cada paso, sino solo el resultado final. Al momento de realizar dicho proceso el sistema, emitirá un mensaje de confirmación, a fin de aceptar el funcionamiento automático

- **Código:**

```
% --- Todo este código fuente es la concatenación de los códigos  
expuestos anteriormente, para que puedan ser realizados en un  
solo paso.  
function Untitled_20_Callback(hObject, eventdata, handles)  
opcion=questdlg('¿Seguro que desea ejecutar el proceso  
automático, para la red neuronal?',",...  
'Yes','No','cancel');  
switch opcion,
```



```
case 'Yes'
```

```
%%aca hay que poner el codigo para inicializar
```

```
global A;  
global estaciones;
```

```
ruta=strcat('C:\ambiental\exportaciones\matlab\',estaciones,'\entra  
da.xls');
```

```
A= xlsread(ruta);  
A=A';  
global B;
```

```
ruta=strcat('C:\ambiental\exportaciones\matlab\',estaciones,'\salid  
a.xls');
```

```
B= xlsread(ruta);
```

```
B=B';
```

```
global red;
```

```
red = newff([0 1; 0 1], [2, 5], {'logsig','logsig'}, 'traingdx');
```

```
red.inputs{1}.size=9
```

```
global tr;
```

```
[red,tr] = train(red, A, B);
```

```
global C;
```

```
ruta=strcat('C:\ambiental\exportaciones\matlab\',estaciones,'\predi  
ccion.xls');
```

```
C= xlsread(ruta);
```

```
C=C';
```

```
global D;
```

```
D = sim(red,C);
```

```
global E;
```

```
global indica;
```

```
indica='falso';
```

```
min=0.30;
```

```
max=0.75;
```

```
if D(1,1) < min
```

```
    E(1,1)=0;
```

```
elseif D(1,1) > max
```

```
    E(1,1)=1;
```

```
else
```

```
    E(1,1)=9.9999;
```

```
    indica='verdad';
```

```
end
```

```
if D(2,1) < min
```

```
    E(2,1)=0;
```

```
elseif D(2,1) > max
```

```
    E(2,1)=1;
```

```
else
```

```
    E(2,1)=9.9999;
```

```

        indica='verdad';
    end
    if D(3,1) < min
        E(3,1)=0;
    elseif D(3,1) > max
        E(3,1)=1;
    else
        E(3,1)=9.9999;
        indica='verdad';
    end
    if D(4,1) < min
        E(4,1)=0;
    elseif D(4,1) >max
        E(4,1)=1;
    else
        E(4,1)=9.9999;
        indica='verdad';
    end
    if D(5,1) < min
        E(5,1)=0;
    elseif D(5,1) > max
        E(5,1)=1;
    else
        E(5,1)=9.9999;
        indica='verdad';
    end
    if strcmpi(indica,'verdad')
        msgbox('Datos ingresados no se pudieron validar con la
Red Neuronal(Fuera de rango)','Sistema');
    else

ruta=strcat('C:\ambiental\exportaciones\matlab\',estaciones,'result
ado.csv');

ruta2=strcat('C:\ambiental\exportaciones\matlab\',estaciones,'res_
inicial.csv');
        xlswrite(ruta, E, 'Hoja1','A1');
        xlswrite(ruta2, D, 'Hoja1','A1');
        warndlg(['Archivo de RESULTADO(Inicial y
aproximado), generado con exito.',ruta2,' y ',ruta],'Sistema');
    end
    case 'No'
end
end

```

- **Interface:**

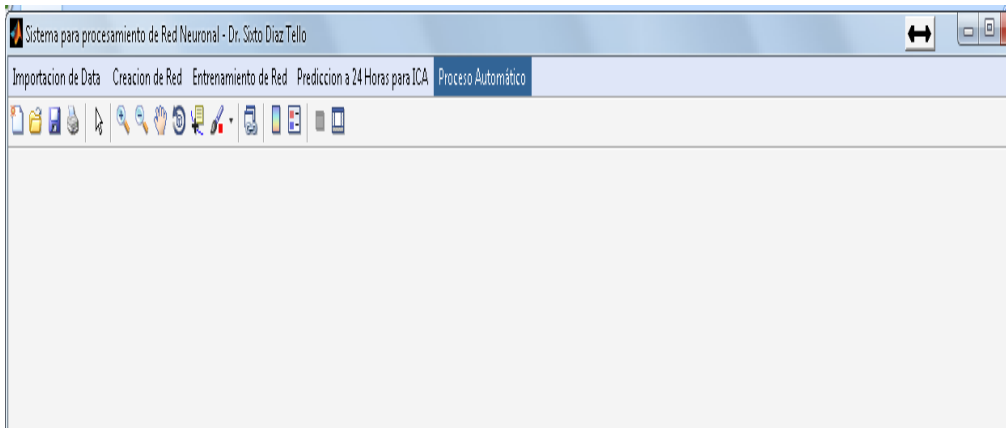


Figura N° 4.37: Procesamiento automático

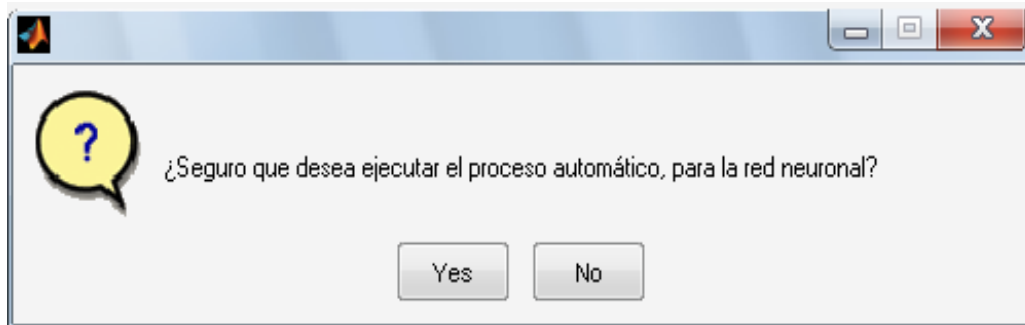


Figura N° 4.38: Mensaje de confirmación

4.2. RESULTADOS: SISTEMA INTEGRAL (RUP)

4.2.1. FASE DE INICIO

4.2.1.1. DOCUMENTO VISIÓN DEL NEGOCIO

a) Propósito

El propósito de este documento es recoger, analizar y definir las necesidades de alto nivel y las características de la aplicación para la integración y automatización de un Modelo de predicción y control de la contaminación atmosférica por partículas en suspensión menores de 10 micras (PM10) bajo el enfoque de minería de datos y redes neuronales, en la ciudad de Chimbote. El documento se centra en la funcionalidad requerida por los participantes en el proyecto y los usuarios finales. Los detalles de cómo el sistema propuesto cubre los requerimientos se pueden observar en la especificación de los casos de uso y otros documentos adicionales.

b) Alcance

El documento visión se ocupa del sistema integral para la predicción del Índice de Calidad del Aire, bajo el enfoque de minería de datos y redes neuronales. El sistema permitirá implementar y validar un modelo de predicción y control, utilizando el enfoque de minería de datos y redes neuronales, que contribuya a incrementar el conocimiento científico de la dinámica de contaminantes atmosféricos en la Provincia del Santa con un alto grado de detalle o profundización y que contribuya a ser una de herramienta de ayuda en: La selección de los emplazamientos de estaciones de redes de calidad del aire (PM10); y la toma de decisiones para poder cumplir con las directivas de actuales sobre calidad del aire.

El presente sistema incluye los siguientes módulos: Modulo de Preparación de data, Modulo para procesamiento de red neuronal, Modulo para Minería de Datos y Módulo

para publicación de resultados. Para los cual el sistema interactúa, por medio de 04 estaciones de trabajo (Santa, Coishco, Chimbote y Nuevo Chimbote), por medio de un Administrador general y un administrador por estación con sus respectivos usuarios y claves.

Se le brindará información oportuna, útil, fiable y a tiempo a la Administración, la opinión pública y los centros de investigación, a fin de tomar las medidas preventivas del caso.

c) Oportunidad Del Negocio

Las oportunidades del negocio que serán encontradas con el desarrollo del proyecto son:

- ✓ Oportunidad para informar episodios de contaminación y vigilar la evolución acotando situaciones de riesgo.
- ✓ Oportunidad para determinar los niveles del principal contaminante PM10 y focos de contaminación en cada zona geográfica que se ha establecido.
- ✓ Oportunidad para establecer medidas necesarias para corregir situaciones crónicas.
- ✓ Oportunidad para apoyar al proceso de toma de decisiones, a través de la explotación de los reportes emitidos por la minería de datos.

d) Panorama del Producto

• Perspectiva del producto

El producto a implementar es un sistema integral para un modelo de predicción y control de la contaminación atmosférica por partículas en suspensión menores de 10 micras (PM10) bajo el enfoque de minería de datos y redes neuronales que nos permitirá brindar información clara y transparente sobre la calidad del aire en la ciudad de Chimbote.

e) Características del Producto

El sistema a desarrollar contará con las siguientes características:

- ✓ Interfaces gráficas amigables y fáciles de manejar, que facilitarán los procesos a implementar.
- ✓ El sistema interactuará con el SGBD (SQL Anywhere)
- ✓ La Red Neuronal, se implementará a través de Matlab 2013.
- ✓ La minería de datos, se realizará a través de KNIME 2.9.2
- ✓ Las consultas y reportes serán de manera precisa, exacta y clara.
- ✓ Acceso al sistema a Administrador y a las 04 estaciones

f) Requerimientos del Producto

Los requerimientos mínimos para el uso del sistema son:

- ✓ Un servidor Web Apache Web Server
- ✓ Un servidor de Base de datos SQL Anywhere
- ✓ 1 Computadora Intel Core I7. 8 GB RAM y SO Windows Windows Seven.
- ✓ Contar con las herramientas; Power Builder, Matlab, Knime, WAMP Server, Macromedia

4.2.1.2. PLAN DE DESARROLLO DEL SOFTWARE

4.2.1.2.1. Introducción al Plan de Desarrollo

La finalidad de éste plan es precisar la definición de actividades en términos de fases e iteraciones para implementar el Sistema integral para un modelo de predicción y control de la contaminación atmosférica por partículas en suspensión menores de 10 micras (PM10) bajo el enfoque de minería de datos y redes neuronales, en la ciudad de

Chimbote. Los detalles individuales de las iteraciones serán descritos en los planes de iteración

4.2.1.2.2. Plan del proyecto

a) Plan de Fases

El desarrollo del Sistema Integral se llevará a través de la utilización de un número de iteraciones en cada una de las fases y del tiempo de duración aproximado por cada fase. Los hitos que marcan el final de cada fase se describen en la siguiente tabla:

Tabla N° 4.8. Plan de Fases

FASES	DESCRIPCIÓN	HITO
Iniciación	En la Fase de Iniciación se desarrollarán los requerimientos del producto y se establecerán los casos de uso del negocio. Además se desarrollará el Plan de Desarrollo del Software. Se definirán los nuevos procesos, necesarios a implementar para el nuevo producto	El Hito Revisión de Casos del Negocio marcará la decisión de seguir o cancelar el proyecto.
Elaboración	En la Fase de Elaboración se analizarán los requerimientos y se desarrollará el prototipo arquitectónico. Al término de la Fase de Elaboración todos los casos de uso seleccionados serán completados en el análisis y diseño.	El Hito Prototipo arquitectónico marcará la verificación de los principales componentes arquitectónicos
Construcción	Durante la Fase de Construcción, se analizarán y se diseñarán los casos del uso restantes. Además se desarrollará y se distribuirá la versión Beta del producto para su evaluación respectiva. Por último se completarán las actividades de prueba e implementación de v1.0.	El Hito Capacidad Operacional de la versión beta y de la v1.0 marcará la disponibilidad del Software.
Transición	En la Fase de Transición se preparará la versión 1.0 del producto para su distribución. Además se proporcionará el apoyo necesario y se distribuirán los manuales para la instalación y configuración del sistema y la capacitación del usuario.	El Hito Puesta en marcha de la v1.0 marcará la disponibilidad de las capacidades del producto

b) Tiempo de duración:

El tiempo de duración aproximado por cada fase es el siguiente:

Tabla N° 4.9. Tiempo de Duración

Fases	N° Iteraciones	Comienza	Finaliza
Inicio	3	Semana 1	Semana 2
Elaboración	3	Semana 3	Semana 5
Construcción	25	Semana 6	Semana 15
Transición	10	Semana 16	Semana 18

c) Cronograma del Proyecto:

El Cronograma del Desarrollo de las Fases e Hitos del Proyecto es el siguiente:

Tabla N° 4.10. Cronograma del Proyecto

Hitos del Proyecto	Fecha de Inicio	Fecha de Terminación
Revisión del negocio	11/01/2014	24/01/2014
Prototipo arquitectural	25/01/2014	14/02/2014
Capacidad operacional de la versión final del sistema	15/02/2014	25/04/2014
Puesta en marcha del sistema	26/04/2014	16/05/2014

d) Cronograma de Fases del Proyecto

El cronograma para cada fase del proyecto se detalla a continuación:

Tabla N° 4.11. Cronograma de Fases del proyecto

FASES DEL PROYECTO	FECHA DE INICIO	FECHA DE TERMINO
FASE DE INICIO	11/01/2014	24/01/2014
Requerimientos	11/01/2014	24/01/2014
Modelo de Negocio	11/01/2014	15/01/2014
Modelo de Dominio	15/01/2014	20/01/2014
Modelo de Casos de Uso	20/01/2014	24/01/2014
FASE DE ELABORACION	25/01/2014	14/02/2014

Análisis	25/01/2014	14/02/2014
Diagrama de Paquetes	25/01/2014	05/02/2014
Análisis de Casos de Uso	04/02/2014	14/02/2014
FASE DE CONSTRUCCION	15/02/2014	25/04/2014
Diseño	15/02/2014	06/03/2014
Diagrama de Colaboración	15/02/2014	22/02/2014
Diagrama de Colaboración	23/02/2014	01/03/2014
Diagrama de Clases	02/03/2014	06/03/2014
Implementación	27/02/2014	25/04/2014
Diagrama de Componentes	27/02/2014	01/03/2014
Diagrama de Despliegue	02/03/2014	05/03/2014
Diagrama de Base de Datos	05/03/2014	12/03/2014
Interfaces y codificación	27/02/2014	25/04/2014
FASE DE TRANSICION	26/04/2014	16/05/2014
Pruebas	26/04/2014	16/05/2014

e) Recursos del Proyecto :

El proyecto será desarrollado totalmente por el autor del presente informe, pero algunas actividades contarán con el apoyo de profesionales especialista en el área de Sistemas.

4.2.1.3. MODELO DEL NEGOCIO

a) Actores Del Negocio

- DIGESA
- MINSA
- SENAMHI
- GESTA DEL AIRE
- Consejo de Ministros
- Entidades públicas y privadas

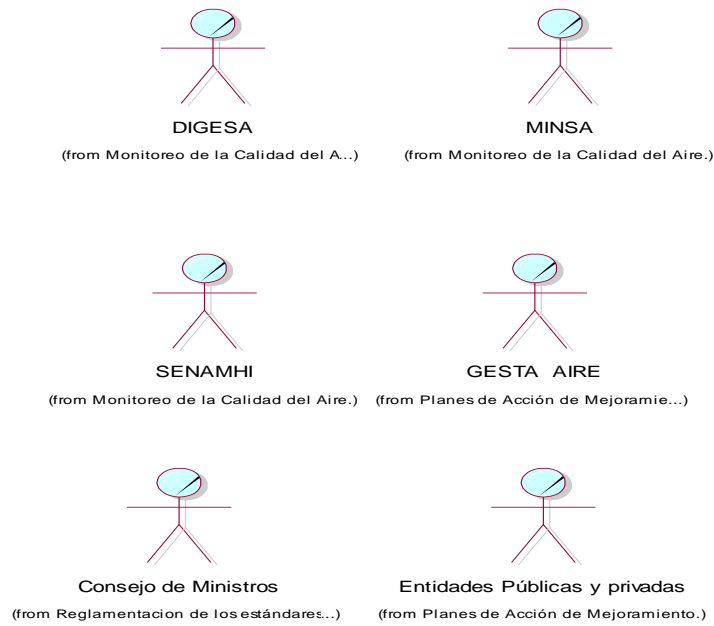


Figura N° 4.39: Actores del negocio

b) Paquetes del Negocio

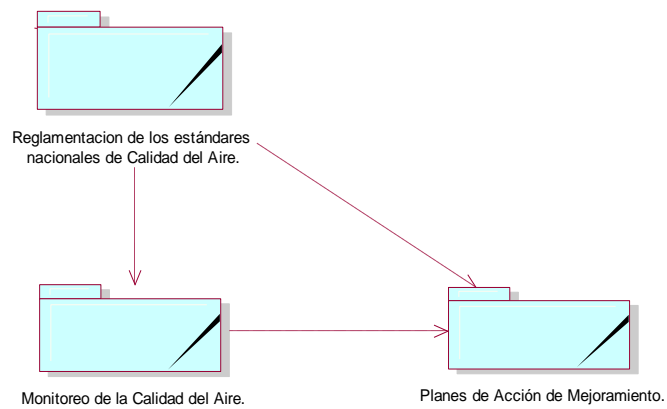


Figura 4.40. Diagrama de paquetes del Negocio

Proceso: Reglamentación de los estándares nacionales de Calidad del Aire

- ✓ Realización de estudios técnico ambiental
- ✓ Determinación de Estándares
- ✓ Aprobación de reglamentación

- ✓ Publicación de estándares

Proceso: Monitoreo de la Calidad del Aire

- ✓ Identificación de áreas potenciales
- ✓ Aplicación de Instrumentos y medidas
- ✓ Recopilación de informes meteorológicos
- ✓ Inventario de Emisiones
- ✓ Realización de Estudios epidemiológicos
- ✓ Entrega del informe final de Monitoreo de Calidad del Aire

Proceso: Planes de Acción de Mejoramiento

- ✓ Diagnóstico de línea base
- ✓ Análisis costo- beneficio
- ✓ Elaboración de propuesta de planes de acción
- ✓ Elaboración de programas de vigilancia
- ✓ Aprobación de planes de acción
- ✓ Ejecución del plan de acción

1. Estructura del Modelo de Negocio

Diagrama de Casos de Uso: Proceso - Reglamentación de los estándares nacionales de Calidad del Aire

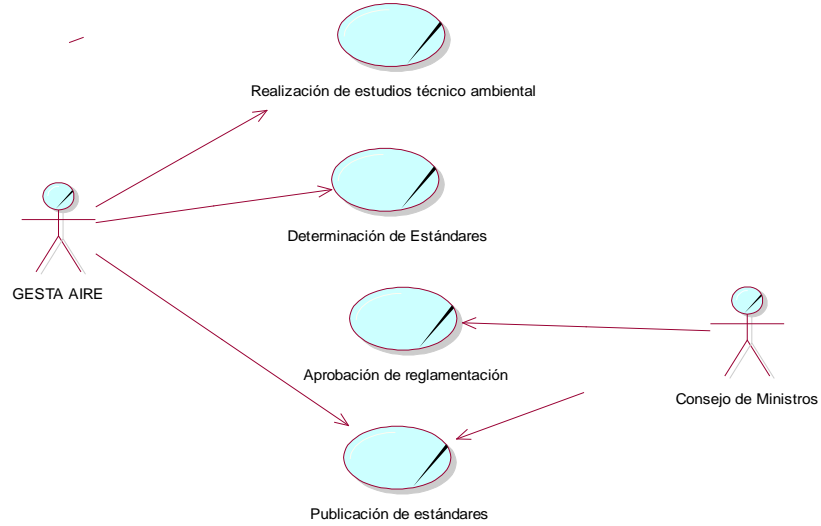


Figura 4.41. CU: Reglamentación de estándares

Diagrama de Casos de Uso: Proceso- Monitoreo de la Calidad del Aire

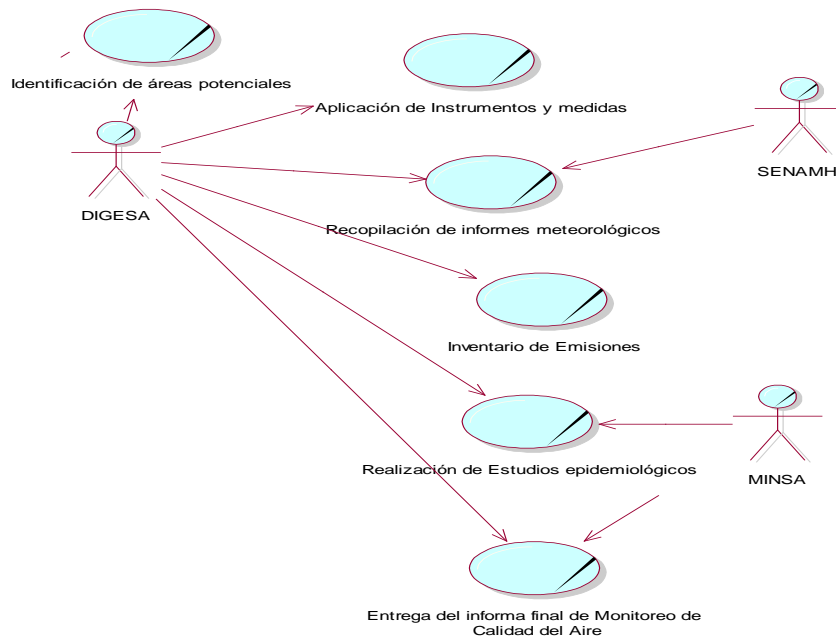


Figura 4.42. CU: Monitoreo de la calidad del Aire

Diagrama de Casos de Uso: Proceso - Planes de Acción de Mejoramiento

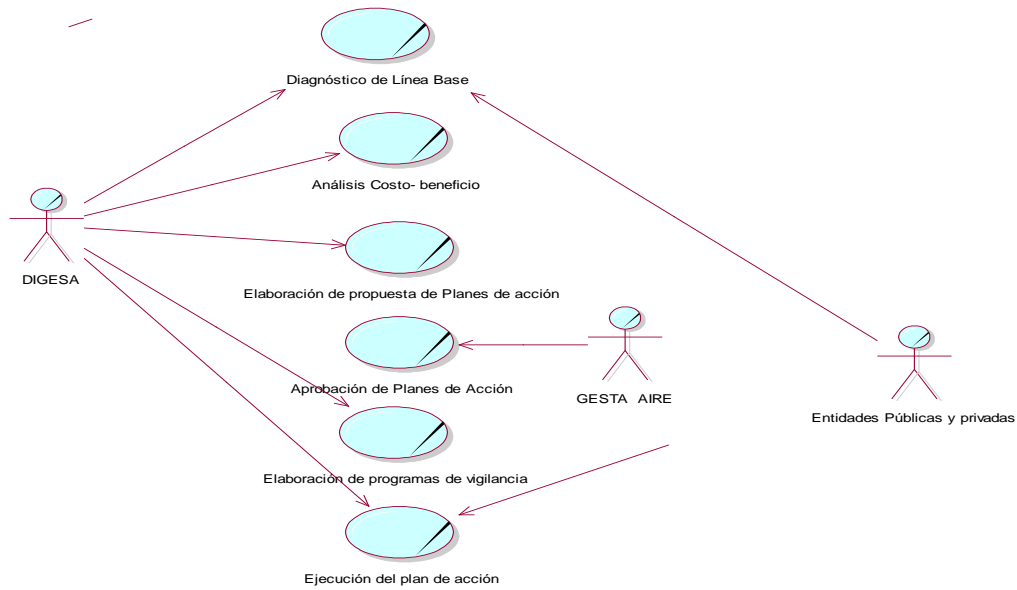


Figura 4.43. CU: Planes de Acción

f) Estructura del Modelo de Dominio

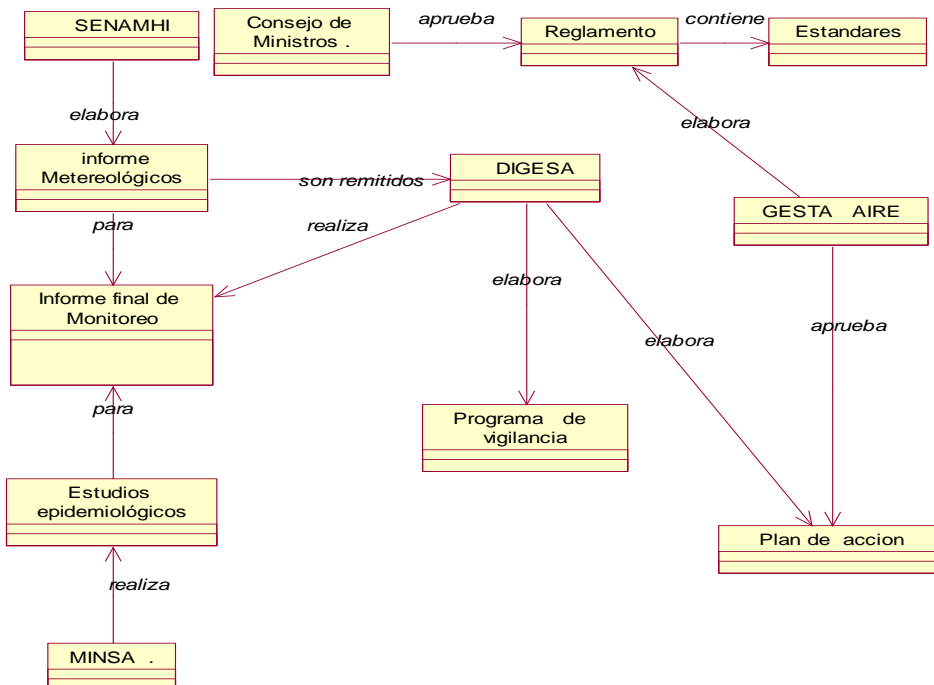


Figura 4.44 CU: Estructura del Modelo de Dominio

4.2.1.4. REQUERIMIENTOS FUNCIONALES Y NO FUNCIONALES

REQUERIMIENTOS FUNCIONALES	
ID	DESCRIPCIÓN
1.	Registro de los parámetros de ICA.
	1.1 Permitir la inserción de parámetros. 1.2 Permitir la consulta y modificación después del registro.
2.	Registro de información meteorológica.
	2.1 Permitir el registro de la data meteorológica, donde se incluirán la información de las variables estacionales y meteorológicas. 2.2 Permitir la modificación y eliminación de la información que se ve registrando diariamente. 2.3 Con la finalidad de poder tener una mejor interpretación de la información que se va registrando, permitir el ordenamiento de la data de acuerdo a los criterio que serán establecidos por el usuario
3.	Validar las variables para la red neuronal
	3.1 Permitir obtener y procesar estadísticamente la información registrada. 3.2 Permitir seleccionar cada variable y validarla, de acuerdo a la estación y según el rango de fechas seleccionadas. 3.3. Permitir visualizar gráficamente y estadísticamente la información procesada, con la finalidad de verificar si la variable analizada influye en nuestra variable de contaminación PM10.
4.	Crear y procesar Red Neuronal.
	4.1 Permitir exportar la data meteorológica del sistema de Power Builder 4.2 Permitir importar la data meteorológica para el programa de Matlab. 4.3 Permitir generar la red neuronal, con los parámetros óptimos establecidos 4.4 Permitir entrenar la red neuronal con la data importada. 4.5 Permitir simular la red, con la data a 24 horas.
5.	Visualizar a través de la web, las predicciones de la estaciones
	5.1 Permitir generar un mapa geográfico. 5.2 Permitir procesar y visualizar las diversas predicciones.
6.	Permitir consultar la información meteorológica a través de filtros.
7.	Permitir procesar gráficas de interpretación de la red neuronal
9.	Emitir Reporte de la información meteorológica y de la red neuronal

REQUERIMIENTOS NO FUNCIONALES	
ID	DESCRIPCION
1.	Flexibilidad para el registro, proceso, modificación y mantenimiento, de los datos meteorológicos.
2.	Facilidad de uso debido a su interfaz amigable y sencilla y la utilización de teclas abreviadas de fácil manejo para todos los usuarios.
3.	El tiempo de recuperación de datos debe ser menor o igual a los 5 segundos, para interfaces de mantenimiento y registro y menor o igual a 1 minuto para procesos de operaciones masivas y de reportaría.
4.	Seguridad de la información registrada.
5.	Interfaz intuitiva y segura de usar.
6.	Emisión de buenos reportes.
7.	Integración de todos los sistemas desarrollados.

4.2.1.5. MODELO DE CASOS DE USO

a) Actores

Despues de realizar el analisis se encontraron los siguientes actores:

- ✓ **Administrador del sistema:** Representa la persona que es el responsable de mantener actualizada la base de datos del sistema, administrando la informacion de las variables metereológicas, para la realizacion de la predicción y control de la contaminación atmosférica por partículas en suspensión menores de 10 micras (PM10) bajo el enfoque de minería de datos y redes neuronales,

b) Casos de Uso

Despues del analisis establecido en el analisis de negocio se encontraron los siguientes casos de uso para el sistema:

- ✓ Iniciar sesión
- ✓ Registrar parámetros de ICA

- ✓ Registrar información meteorológica diaria
- ✓ Validar variables estacionales para red.
- ✓ Validar variables meteorológicas para red
- ✓ Generar archivo Input: Entrada para Matlab
- ✓ Generar archivo Input: Salida para Matlab
- ✓ Exportar archivos input para Matlab
- ✓ Generar y exportar archivo Input para predicción en Matlab
- ✓ Generar y exportar archivo input para Knime
- ✓ Seleccionar de estación de monitoreo
- ✓ Importar y visualizar data de entrada
- ✓ Importar y visualizar data de salida
- ✓ Crear red Neuronal
- ✓ Visualizar esquema de red
- ✓ Generar entrenamiento
- ✓ Visualizar esquema de entrenamiento
- ✓ Simular y validar red con data entrada
- ✓ Importar data de entrada para predicción
- ✓ Simular y predecir ICA
- ✓ Realizar aproximación
- ✓ Exportar resultado de predicción obtenida

c) Clasificación de Casos de Uso

A continuación se presenta una clasificación de los casos de uso en base a su priorización en los procesos para la predicción y control de la contaminación atmosférica por partículas en suspensión menores de 10 micras (PM10), esta clasificación se basa en los patrones establecidos por la metodología RUP.

Tabla N° 4.12. Clasificación de Casos de Uso-Inicio

Clasificación	Casos De Uso	Justificación
Alto	<ul style="list-style-type: none"> ✓ Iniciar sesión ✓ Registrar parámetros de ICA ✓ Registrar información meteorológica diaria ✓ Generar archivo Input: Entrada para Matlab ✓ Generar archivo Input: Salida para Matlab ✓ Exportar archivos input para Matlab ✓ Generar y exportar archivo Input para predicción en Matlab ✓ Importar y visualizar data de entrada ✓ Importar y visualizar data de salida ✓ Crear Red Neuronal ✓ Generar entrenamiento ✓ Importar data de entrada para predicción ✓ Simular y predecir ICA ✓ Realizar aproximación ✓ Exportar resultado de predicción obtenida 	<p>Son los casos de uso de gran importancia en el modelo del negocio, es por ende la base para obtener una Arquitectura Estable.</p>
Medio	<ul style="list-style-type: none"> ✓ Validar variables estacionales para red. ✓ Validar variables meteorológicas para red ✓ Seleccionar de estación de monitoreo ✓ Simular y validar red con data entrada 	<p>Son los casos de uso que tienen mediano impacto en la arquitectura, pero sirve de apoyo a los casos de uso crítico.</p>
Bajo	<ul style="list-style-type: none"> ✓ Generar y exportar archivo input para Knime ✓ Visualizar esquema de red ✓ Visualizar esquema de entrenamiento 	<p>Son los casos de uso que tienen impacto bajo en la arquitectura del sistema y algunas veces van a depender de otros casos de uso de nivel medio y alto</p>

d) Modelo de Casos de Uso

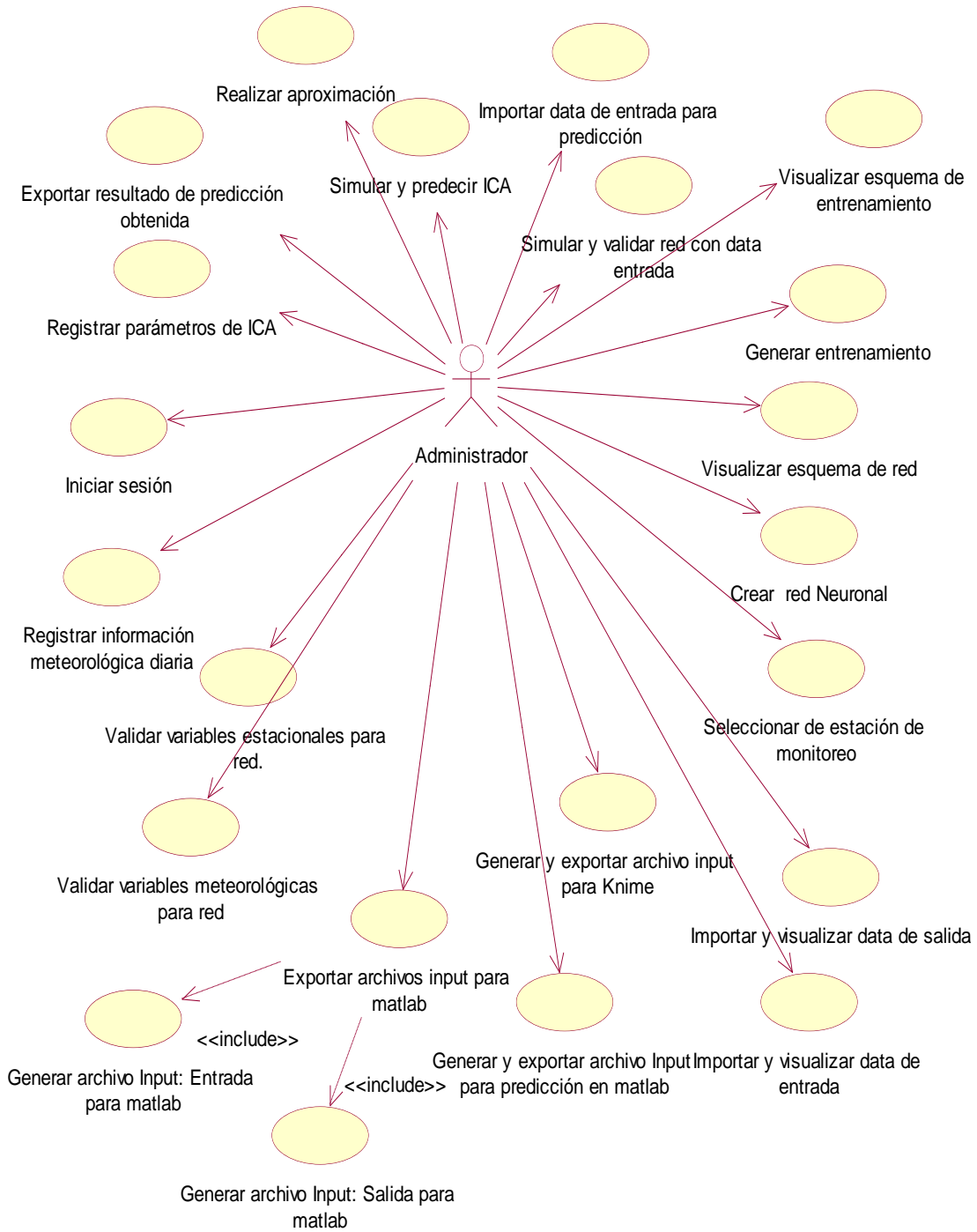


Figura 4.45. Modelo de casos de Uso

e) Glosario

Se presentarán los términos manejados a lo largo del proyecto de desarrollo del sistema integral, consta de un diccionario de datos y definiciones de la nomenclatura que se utiliza, con el objetivo de que se convierta en un lenguaje común dentro de nuestro proyecto para mayor entendimiento.

El glosario está organizado por definición de términos ordenados de forma ascendente según el orden alfabético.

- ✓ **GESTA AIRE:** Grupo de Estudio Técnico Ambiental de la Calidad del Aire encargado de formular y evaluar los planes de acción para el mejoramiento de la calidad del aire en una Zona de Atención Prioritaria.
- ✓ **Índice de Calidad del Aire (ICA):** El índice de calidad del aire (ICA) se considera un indicador global de la calidad del aire en un momento determinado o día y en una estación de monitoreo concreta.
- ✓ **PM10:** Pequeñas partículas sólidas o líquidas de polvo, cenizas, hollín, partículas metálicas, cemento o polen, dispersas en la atmósfera, y cuyo diámetro es menor que 10 μm (1 micrómetro corresponde la milésima parte de 1 milímetro).
- ✓ **SENAMHI:** El Servicio Nacional de Meteorología e Hidrología del Perú brinda información sobre el pronóstico del tiempo en el Perú, así asesoría y estudios científicos en las áreas de hidrología, meteorología, agrometeorología y asuntos ambientales.
- ✓ **Servidor Proxy:** Es un ordenador que intercepta las conexiones de red que un cliente hace a un servidor de destino.

- ✓ **Variable Estacional:** Para el desarrollo del proyecto, se esta considerando, como variables al dia de la semana y mes del año.
- ✓ **Variable Meteorológica:** Para el desarrollo del proyecto, se esta considerando, como variables a la dirección del viento, velocidad predominante, velocidad promedio, velocidad máxima, temperatura, presion humedad.

f) Analisis de Riesgo.

Los riesgos a los que puede estar expuesto el Sistema son los siguientes:

- ✓ Ataques a la aplicación, mediante ejecución de código remoto, inyecciones SQL, inyecciones PHP (usado para la visualización de los resultados de predicción), etc.
- ✓ Pérdida de información importante por algún tipo de desastre (incendios, terremotos, etc.)
- ✓ Pérdida de información por virus informáticos.
- ✓ La resistencia al cambio por parte de alguna entidad, relacionada al control de la calidad del aire, para hacer uso del nuevo sistema.
- ✓ Configuración insegura del servidor de base de datos.

4.2.2. FASE DE ELABORACION

4.2.2.1. MODELO DE CASOS DE USO

a) Encontrando Casos de Uso

- ✓ Recolectar data automáticamente
- ✓ Importar data meteorológica por CSV
- ✓ Consultar parámetros de ICA
- ✓ Consultar data estacional/ meteorológica
- ✓ Configurar Red
- ✓ Graficar resultados de procesamiento de red

- ✓ Realizar proceso automático
- ✓ Visualizar monitoreo de predicción de ICA
- ✓ Procesar minería de datos en knime
- ✓ Procesar resultado Matlab

b) Restableciendo Casos de Uso

Módulo registro y validación de data

- ✓ Iniciar sesión
- ✓ Registrar parámetros de ICA
- ✓ Recolectar data automáticamente
- ✓ Registrar información meteorológica diaria
- ✓ Importar data meteorológica por CSV
- ✓ Consultar parámetros de ICA
- ✓ Validar variables estacionales y meteorológicas para red
- ✓ Generar archivo Input: Entrada para Matlab
- ✓ Generar archivo Input: Salida para Matlab
- ✓ Exportar archivos input para Matlab
- ✓ Generar y exportar archivo Input para predicción en Matlab
- ✓ Consultar data estacional/ meteorológica

Módulo minería de datos (Knime)

- ✓ Generar y exportar archivo input para Knime
- ✓ Procesar minería de datos en knime

Módulo red neuronal (Matlab)

- ✓ Seleccionar de estación de monitoreo
- ✓ Importar y visualizar data de entrada
- ✓ Importar y visualizar data de salida
- ✓ Configurar Red

- ✓ Crear red Neuronal
- ✓ Visualizar esquema de red
- ✓ Generar entrenamiento
- ✓ Visualizar esquema de entrenamiento
- ✓ Simular y validar red con data entrada
- ✓ Graficar resultados de procesamiento de red
- ✓ Importar data de entrada para predicción
- ✓ Simular y predecir ICA
- ✓ Realizar aproximación
- ✓ Exportar resultado de predicción obtenida
- ✓ Realizar proceso automático

Módulo web para visualización de predicciones

- ✓ Procesar resultados Matlab.
- ✓ Visualizar monitoreo de predicción de ICA

c) Clasificación de los Casos de Uso

Una vez descubiertos los casos de uso, clasificaremos de acuerdo a su priorización, es decir desde los que tienen fuerte impacto hasta un mínimo impacto en la arquitectura, esta clasificación se basa en los patrones establecidos por la metodología RUP y se representa en la tabla 4.13.

Tabla N° 4.13. Clasificación de Casos de Uso-Elaboración

Clasificación	Casos De Uso	Justificación
Alto	<ul style="list-style-type: none"> ✓ Iniciar sesión ✓ Registrar parámetros de ICA ✓ Registrar información meteorológica diaria ✓ Generar archivo Input: Entrada para Matlab ✓ Generar archivo Input: Salida para Matlab 	<p>Son los casos de uso de gran importancia en el modelo del negocio, es por ende la base para obtener una Arquitectura Estable.</p>

	<ul style="list-style-type: none"> ✓ Exportar archivos input para Matlab ✓ Generar y exportar archivo Input para predicción en Matlab ✓ Importar y visualizar data de entrada ✓ Importar y visualizar data de salida ✓ Crear red Neuronal ✓ Generar entrenamiento ✓ Importar data de entrada para predicción ✓ Simular y predecir ICA ✓ Realizar aproximación ✓ Exportar resultado de predicción obtenida ✓ Recolectar data automáticamente ✓ Realizar proceso automático ✓ Procesar resultados Matlab. ✓ Visualizar monitoreo de predicción de ICA 	
Medio	<ul style="list-style-type: none"> ✓ Validar variables estacionales para red. ✓ Validar variables meteorológicas para red ✓ Seleccionar de estación de monitoreo ✓ Simular y validar red con data entrada ✓ Importar data meteorológica por CSV ✓ Procesar minería de datos en knime ✓ Configurar Red ✓ Graficar resultados de procesamiento de red ✓ Procesar resultados Matlab 	<p>Son los casos de uso que tienen mediano impacto en la arquitectura, pero sirve de apoyo a los casos de uso crítico.</p>
Bajo	<ul style="list-style-type: none"> ✓ Generar y exportar archivo input para Knime ✓ Visualizar esquema de red ✓ Visualizar esquema de entrenamiento ✓ Consultar parámetros de ICA ✓ Consultar data estacional/ meteorológica 	<p>Son los casos de uso que tienen impacto bajo en la arquitectura del sistema y algunas veces van a depender de otros casos de uso de nivel medio y alto</p>

d) Diagrama de Caso de Uso

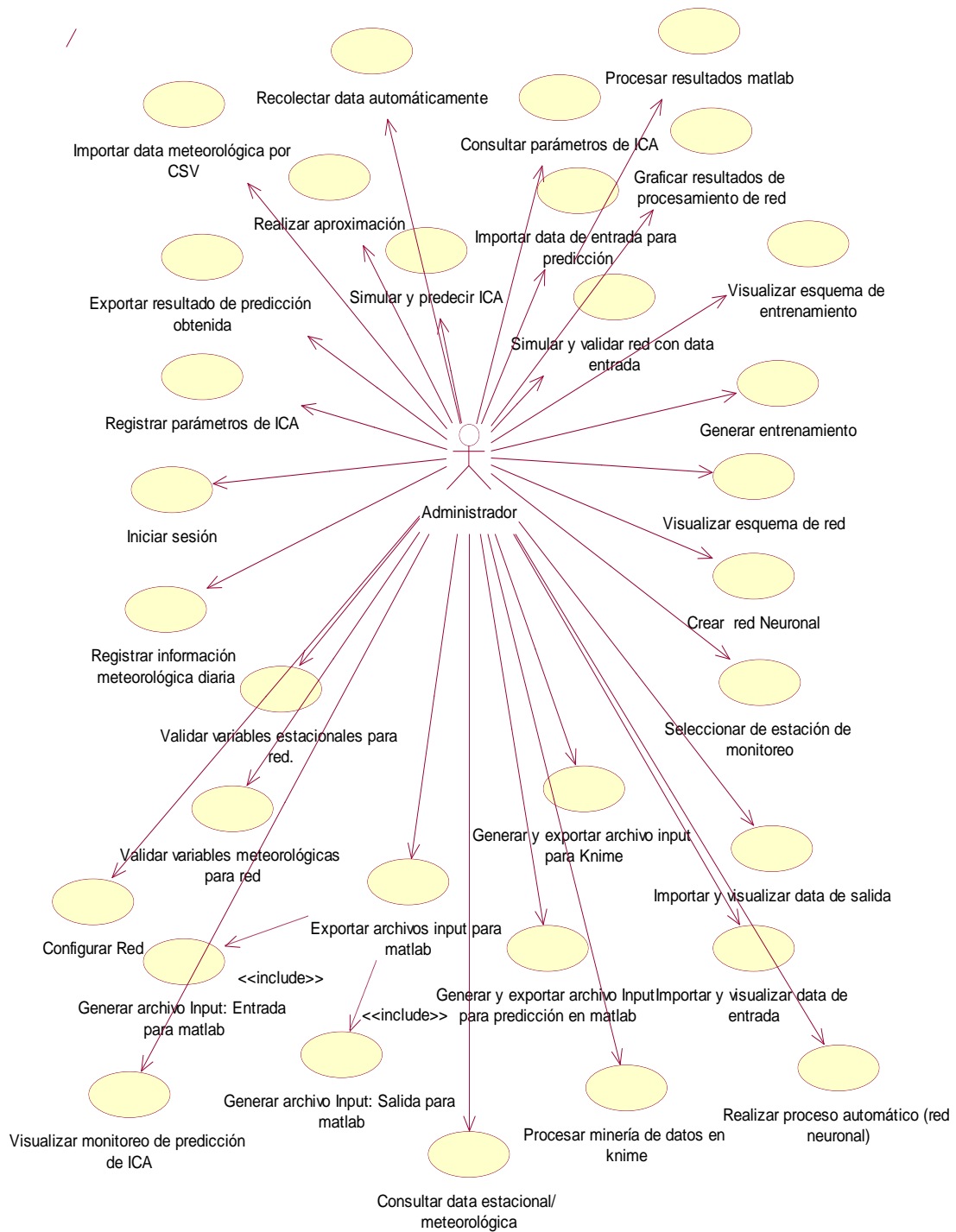


Figura 4.46. Diagrama de casos de Uso

e) **Especificación de Casos de Uso**

CASO DE USO: INICIAR SESION

Tabla N° 4.14. Iniciar Sesión

Caso de Uso:	Iniciar Sesión	
Actores:	Administrador	
Propósito:	Ingresar al sistema	
Resumen:	El usuario a lo sistemas, e ingresa el usuario y clave asignado, si sus datos ingresados, son incorrectos, se negará el acceso al sistema; caso contrario, ingresará al sistema y podrá realizar las opciones que se le presenten.	
Acciones de los eventos:		
	Acción de los Actores	Respuesta del Sistema
	1. El usuario ingresa al sistema integral	2. El sistema muestra la ventana de acceso.
	3. El usuario ingresa su usuario y clave	4. Verifica la identificación
		5. Establece una sesión de autenticación.
		6. El sistema muestra el panel de control, con las diversas opciones.
Casos Alternos		
Línea 4: El sistema rechaza el inicio de sesión debido a que los datos no coinciden.		

CASO DE USO: REGISTRAR PARÁMETROS DE ICA

Tabla N° 4.15. Registrar parámetros de ICA

Caso de Uso:	Registrar parámetros de ICA	
Actores:	Administrador	
Propósito:	Registrar parámetros de ICA	
Resumen:	El administrador, ingresará los datos referente a los parámetros del índice de calidad del aire, de acuerdo a la normatividad vigente, relacionando principalmente, el ICA con los valores del PM10.	
Acciones de los eventos:		
	Acción de los Actores	Respuesta del Sistema
	1. El administrador ingresa a la ventana “Parámetro del ICA”	2. El sistema muestra la ventana seleccionada
	3. Presionará el botón “Insertar / registrar Modificación”	4. El sistema muestra la venta de registro
	5. Ingresar los datos para el registro	6. El sistema valida los datos ingresados
		7. El sistema verifica los colores de acuerdo al estado seleccionado
		8. El sistema registra en la base de datos el parámetro.
		9. El sistema muestra un mensaje satisfactorio.
Casos Alternos		
Línea 6: El sistema rechaza el registro, debido a que no se han ingresado los datos correctamente.		

CASO DE USO: RECOLECTAR DATA AUTOMÁTICAMENTE

Tabla N° 4.16. Recolectar data automáticamente

Caso de Uso:	Recolectar data automáticamente	
Actores:	Administrador	
Propósito:	Registrar la información meteorológicas, a través de estaciones de monitoreo en tiempo real.	
Resumen:	El administrador, activará la recolección de data, que se irá actualizando en el sistema, para su posterior procesamiento.	
Acciones de los eventos:		
	Acción de los Actores	Respuesta del Sistema
	1. El administrador selecciona la ventana de recolección.	2. El sistema muestra la ventana seleccionada.
		3. El sistema sincronizará con la estación.
	4. Activa la recolección de data.	5. El sistema captura la información y la procesa.
		6. El sistema verifica los datos.
		7. Graba los datos temporales y diarios.
		8. Muestra mensaje de captura de paquetes de data.
Casos Alternos		
	Línea 3: El sistema emite un error debido a que no existe la conexión con la estación.	

CASO DE USO: REGISTRAR INFORMACION METEREOLÓGICA DIARIA.

Tabla N° 4.17. Registrar información meteorológica diaria

Caso de Uso:	Registrar información meteorológica diaria	
Actores:	Administrador	
Propósito:	Registrar la información meteorológica cada día.	
Resumen:	El administrador, por medio de esta interfaz, registrará cada día la información, que servirá para realizar el entrenamiento de la red.	
Acciones de los eventos:		
	Acción de los Actores	Respuesta del Sistema
	1. Ingresar a la ventana “Ingreso data individual / por formulario”	2. Muestra la ventana seleccionada
	3. Presiona el botón nuevo	4. El sistema muestra la ventana de registro de nuevo día
	5. Selecciona la estación.	
	6. Ingresar data de variables meteorológica	
	7. Ingresar data de variables estacionales.	
	8. Ingresar variables de contaminación y graba el registro	9. El sistema valida la información registrada.
		10. El sistema graba los datos en la BD
		11. Mensaje satisfactorio
Casos Alternos		
Línea 9: El sistema rechaza la grabación, debido a que los datos están mal ingresados o falta algún campo.		

CASO DE USO: VALIDAR VARIABLES ESTACIONALES Y METEOROLÓGICAS PARA RED

Tabla N° 4.18. Validar variables estacionales y meteorológicas para red

Caso de Uso:	Validar variables estacionales y meteorológicas para red	
Actores:	Administrador	
Propósito:	Validar que las variables que se usaran influyen respecto al PM10.	
Resumen:	El administrador, validará las 9 variables (2 estacionales y 7 meteorológicas), con respecto al PM10. A fin de que si puedan ser usadas para la red neuronal.	
Acciones de los eventos:		
	Acción de los Actores	Respuesta del Sistema
	1. Selecciona la ventana que contenga a la variable que desea validar.	2. El sistema muestra la ventana seleccionada
	3. Selecciona la estación	
	4. Selecciona el rango de fechas	
	5. Presiona el botón visualizar	6. El sistema filtra la data de acuerdo a los parámetros seleccionados y la variable.
		7. El sistema procesa la data filtrada.
		8. Muestra en pantalla los datos estadísticos, grafica y el análisis.
Casos Alternos		
Línea 6: El sistema no filtra información, debido a que no existen coincidencias con dichos parámetros.		

CASO DE USO: GENERAR ARCHIVO INPUT: ENTRADA PARA MATLAB

Tabla N° 4.19. Generar archivo Input: Entrada para Matlab

Caso de Uso:	Generar archivo Input: Entrada para Matlab	
Actores:	Administrador	
Propósito:	Preparar la información que será proceda para la red neuronal en Matlab	
Resumen:	El administrador, prepara la data, de entrada para el entrenamiento de la red neuronal.	
Acciones de los eventos:		
	Acción de los Actores	Respuesta del Sistema
	1. Ingresa a la pantalla generar archivo input para Matlab.	2. Muestra la ventana seleccionada.
	3. Selecciona la estación a analizar.	
	4. Presiona el botón “Procesar datos registrados.”	5. El sistema filtra la información de acuerdo a la estación seleccionada.
		6. El sistema procesa la información de las 9 variables y la muestra en pantalla
Casos Alternos		
Línea 5: El sistema no encuentra data registrada en dicha estación.		
Línea 6: El sistema no puede procesar la data, porque encuentra información incompleta de alguna variable.		

CASO DE USO: GENERAR ARCHIVO INPUT: SALIDA PARA MATLAB**Tabla N° 4.20.** Generar archivo Input: Salida para Matlab

Caso de Uso:	Generar archivo Input: Salida para Matlab	
Actores:	Administrador	
Propósito:	Preparar la información que será proceda para la red neuronal en Matlab	
Resumen:	El administrador, prepara la data, de salida para el entrenamiento de la red neuronal.	
Acciones de los eventos:		
	Acción de los Actores	Respuesta del Sistema
	1. Ingresa a la pantalla generar archivo input para Matlab.	2. Muestra la ventana seleccionada.
	3. Selecciona la estación a analizar.	
	4. Presiona el botón “Procesar datos registrados.”	5. El sistema filtra la información de acuerdo a la estación seleccionada.
		6. El sistema procesa la información de la salida objetivo para cada día registrado.
Casos Alternos		
Línea 5: El sistema no encuentra data registrada en dicha estación.		
Línea 6: El sistema no puede procesar la data, porque encuentra información incompleta de alguna variable.		

CASO DE USO: EXPORTAR ARCHIVO INPUT PARA MATLAB

Tabla N° 4.21. Exportar archivo Input para Matlab

Caso de Uso:	Exportar Archivo Input para Matlab	
Actores:	Administrador	
Propósito:	Exportar la información en archivo Excel.	
Resumen:	El administrador, exporta la data, de entrada y salida para el entrenamiento de la red neuronal, a través de un archivo de Excel, que será importado por Matlab para su procesamiento.	
Acciones de los eventos:		
Acción de los Actores	Respuesta del Sistema	
	1. El sistema tiene que haber procesado, los archivos input de entrada y salida.	
2. El administrador presiona el botón 'Exportar data para archivos input'	3. El sistema escribe los datos en un archivo Excel	
	4. El sistema guarda el archivo generado en el disco duro de la PC.	
	5. El sistema muestra un mensaje satisfactorio	
Casos Alternos		
Línea 1: El sistema no realizará ninguna acción, porque aun no se ha realizado el procesamiento de la data..		
Línea 3: El sistema no genera el archivo Excel porque no existe data registrada.		

CASO DE USO: GENERAR Y EXPORTAR ARCHIVO INPUT PARA PREDICCIÓN EN MATLAB

Tabla N° 4.22. Generar y exportar archivo Input para predicción en Matlab

Caso de Uso:	Generar y exportar archivo Input para predicción en Matlab	
Actores:	Administrador	
Propósito:	Exportar la información que servirá como entrada para predecir, el ICA.	
Resumen:	El administrador, prepara y guarda la información sobre la información de las variables en un archivo Excel, que servirá como entrada para predecir, el ICA, a través de la red neuronal en Matlab.	
Acciones de los eventos:		
Acción de los Actores	Respuesta del Sistema	
1. Ingresar a la pantalla generar archivo input para Matlab.	2. Muestra la ventana seleccionada.	
3. Selecciona la estación a analizar.		
4. El administrador ingresa la información única sobre las variables.		
5. Selecciona la fecha para la predicción (24 horas)	6. El sistema procesa la fecha	
7. Presiona el botón "Exportar data"	8. El sistema valida la información registrada.	
	9. El sistema prepara la información de acuerdo a la estación seleccionada	
	10. El sistema guarda la información en un archivo Excel	
	11. El archivo generado es guardado en el disco duro de la PC	
Casos Alternos		
Línea 8: El sistema no registra la grabación, porque falta ingresar algún dato.		

CASO DE USO: CONSULTAR DATA ESTACIONAL/ METEOROLÓGICA

Tabla N° 4.23. Consultar data estacional/ meteorológica

Caso de Uso:	Consultar data estacional/ meteorológica	
Actores:	Administrador	
Propósito:	Consultar y filtrar la información seleccionada.	
Resumen:	El administrador, consulta la información de la data que se ha registrado o capturado a través del sistema, para esto hace uso de filtros y parámetros, de acuerdo a los que se dese visualizar.	
Acciones de los eventos:		
	Acción de los Actores	Respuesta del Sistema
	1. El administrador, ingresa a la ventana : “Filtro de datos”	2. El sistema muestra la ventana seleccionada
	3. El administrador selecciona, los filtros que desee utilizar.	
	4. Presiona el botón “Consultar con filtros establecidos”	5. El sistema Valida los filtros seleccionados.
		6. El sistema procesa la información
		7. El sistema muestra en pantalla el resultado de la consulta
		8. Mensaje satisfactorio
Casos Alternos		
	Línea 7: El sistema no muestra nada, porque no encuentra data, con los filtros seleccionados.	

CASO DE USO: GENERAR Y EXPORTAR ARCHIVO INPUT PARA KNIME

Tabla N° 4.24. Generar y exportar archivo Input para Knime

Caso de Uso:	Generar y exportar archivo Input para knime (Minería de datos)	
Actores:	Administrador	
Propósito:	Exportar la información que servirá para realizar la minería de datos	
Resumen:	El administrador, prepara y guarda la información sobre la información de las variables en un archivo Excel, que servirá como entrada para Knime, para realizar la minería de datos.	
Acciones de los eventos:		
	Acción de los Actores	Respuesta del Sistema
	1. Ingresa a la pantalla generar archivo input para knime.	2. Muestra la ventana seleccionada.
	3. Selecciona la estación a analizar.	
	4. El administrador presiona el botón “Procesar datos registrado”	5. El sistema procesa la información de acuerdo a la estación seleccionada.
		6. El sistema muestra en pantalla la información registrada.
	7. Selecciona el tipo de archivo a exportar	
	8. El administrador presiona el botón “Exportar data”	9. Crea un archivo de acuerdo al tipo de archivo seleccionado.
		10. El sistema graba y guarda la información en el archivo
		11. El archivo generado es guardado en el disco duro de la PC
		12. Muestra mensaje satisfactorio.
Casos Alternos		
Línea 6: El sistema no muestra nada, porque no existe información registrada.		

CASO DE USO: SELECCIONAR ESTACIÓN DE MONITOREO

Tabla N° 4.25. Seleccionar estación de monitoreo

Caso de Uso:	Seleccionar estación de monitoreo	
Actores:	Administrador	
Propósito:	Selecciona la estación con la cual va a trabajar la red.	
Resumen:	El administrador, se encuentra en Matlab, y seleccionará la estación (Chimbote, Nuevo Chimbote, Coishco y Santa) con la cual trabajará para la configuración, creación, entrenamiento y simulación de la red neuronal.	
Acciones de los eventos:		
	Acción de los Actores	Respuesta del Sistema
		1. El sistema muestra la pantalla, donde se encuentran las 4 estaciones
	2. El administrador selecciona una estación de trabajo, con la cual se va a trabajar.	3. El sistema guarda el nombre de la estación en una variable global de Matlab, para su tratamiento en todo el sistema.
		4. El sistema cierra la pantalla inicial.
Casos Alternos		
Línea 1: El sistema no muestra la pantalla inicial, debido a que no se encuentran instalados los componentes de Matlab respectivos.		

CASO DE USO: IMPORTAR Y VISUALIZAR DATA DE ENTRADA

Tabla N° 4.26. Importar y visualizar data de entrada

Caso de Uso:	Importar y visualizar data de entrada	
Actores:	Administrador	
Propósito:	Importar la data de entrada, para el funcionamiento de la red neuronal.	
Resumen:	El administrador, seleccionará el archivo que contiene la data de entrada para la red modelada en Matlab, que servirá para el entrenamiento respectivo.	
Acciones de los eventos:		
Acción de los Actores	Respuesta del Sistema	
1. Selecciona la opción “Importar y visualizar data de entrada”	2. El sistema captura la dirección de la ruta del archivo que contenga dicha data	
	3. El sistema lee el archivo	
	4. El sistema procesa la data para la red y almacena en una variable global.	
	5. Imprime en pantalla la data preparada	
	6. Mensaje satisfactorio	
Casos Alternos		
Línea 3: El sistema rechaza la lectura, debido a que no se ha generado la información respectiva desde el sistema inicial.		
Línea 4: El sistema no puede procesar el archivo, porque no está de acuerdo al formato establecido.		

CASO DE USO: IMPORTAR Y VISUALIZAR DATA DE SALIDA

Tabla N° 4.27. Importar y visualizar data de salida

Caso de Uso:	Importar y visualizar data de salida	
Actores:	Administrador	
Propósito:	Importar la data de salida, para el funcionamiento de la red neuronal.	
Resumen:	El administrador, seleccionará el archivo que contiene la data de salida para la red modelada en Matlab, que servirá para el entrenamiento respectivo.	
Acciones de los eventos:		
Acción de los Actores	Respuesta del Sistema	
1. Selecciona la opción “Importar y visualizar data de salida”	2. El sistema captura la dirección de la ruta del archivo que contenga dicha data	
	3. El sistema lee el archivo	
	4. El sistema procesa la data para la red y almacena en una variable global.	
	5. Imprime en pantalla la data preparada	
	6. Mensaje satisfactorio	
Casos Alternos		
Línea 3: El sistema rechaza la lectura, debido a que no se ha generado la información respectiva desde el sistema inicial.		
Línea 4: El sistema no puede procesar el archivo, porque no está de acuerdo al formato establecido.		

CASO DE USO: CREAR RED NEURONAL

Tabla N° 4.28. Crear red neuronal

Caso de Uso:	Crear red neuronal
Actores:	Administrador
Propósito:	Crear la red neuronal, que realizará la predicción del ICA
Resumen:	El administrador, procederá a crear la red neuronal, de acuerdo a los parámetros óptimos de configuración, que permitirán realizar la correcta predicción del ICA.
Acciones de los eventos:	
Acción de los Actores	Respuesta del Sistema
1. Selecciona la opción “Creación de red”	2. El sistema muestra la ventana seleccionada
3. Presiona el botón “Crear red neuronal”	4. El sistema captura las variables globales
	5. El sistema asigna los parámetros para la función newff
	6. El sistema crea la red neuronal
	7. Mensaje satisfactorio
Casos Alternos	
Línea 6: El sistema no puede crear la red neuronal, debido a que no se establecieron los parámetros correctos.	

CASO DE USO: GENERAR ENTRENAMIENTO

Tabla N° 4.29. Generar Entrenamiento

Caso de Uso:	Generar Entrenamiento	
Actores:	Administrador	
Propósito:	Entrenar la red neuronal, que ha sido creada	
Resumen:	El administrador, procederá a realizar el entrenamiento de la red neuronal, de acuerdo a la data input: Entrada y salida que ha sido exportada del sistema inicial.	
Acciones de los eventos:		
	Acción de los Actores	Respuesta del Sistema
	1. Selecciona la opción “Generar entrenamiento”	2. El sistema muestra la ventana seleccionada
	3. Presiona el botón “Entrenar(Parámetros estándares)”	4. El sistema captura las variables globales
		5. El sistema asigna lee la data input Entrada y Salida
		6. El sistema entrena la red a través de la función train.
		7. El sistema muestra una nueva pantalla “Neural Network Training (nntraintool)”
		8. Mensaje satisfactorio
Casos Alternos		
Línea 6: El sistema no puede entrenar la red neuronal, debido a que no ha sido creada.		

CASO DE USO: SIMULAR Y VALIDAR RED CON DATA ENTRENADA

Tabla N° 4.30. Simular y validar red con data entrenada

Caso de Uso:	Simular y validar red con data entrenada	
Actores:	Administrador	
Propósito:	Simular la red con la data que ya ha servido para el entrenamiento.	
Resumen:	El administrador, con la finalidad de validar la red que ha sido entrenada, realizará una simulación con la data de entrada, de tal manera que la salida estimada obtenida deberá ser similar a la salida deseada.	
Acciones de los eventos:		
	Acción de los Actores	Respuesta del Sistema
	1. Selecciona la opción “Simular y validar red entrenada”	2. El sistema muestra la ventana seleccionada
	3. Presiona el botón “Simulación con valores de entrada”	4. El sistema captura las variables globales
		5. El sistema invoca a la función sim().
		6. El sistema realiza la simulación
		7. El sistema calcula el error cuadrado medio de la simulación realizada
		8. El sistema muestra los resultados obtenidos en pantalla
		9. El sistema habilita las gráfica de los resultados
Casos Alternos		
Línea 6: El sistema no puede realizar la simulación, debido a que no ha sido creada o no ha sido entrenada.		

CASO DE USO: IMPORTAR DATA DE ENTRADA PARA PREDICCIÓN

Tabla N° 4.31. Importar data de entrada para predicción

Caso de Uso:	Importar data de entrada para predicción	
Actores:	Administrador	
Propósito:	Importar la data que servirá para realizar la predicción del ICA.	
Resumen:	El administrador, importará la data que ha sido preparada, en el sistema inicial de Power Builder, para poder realizar la predicción (Simulación) del ICA a 24 horas, a través de la red neuronal.	
Acciones de los eventos:		
	Acción de los Actores	Respuesta del Sistema
	1. Selecciona la opción “Importar data de entrada(24 Hrs.)”	2. El sistema mostrará la ventana seleccionada.
		3. El sistema lee el archivo
		4. El sistema procesa la data y almacena en una variable global.
		5. Imprime en pantalla la data preparada
		6. Mensaje satisfactorio
Casos Alternos		
	Línea 3: El sistema rechaza la lectura, debido a que no se ha generado la información respectiva desde el sistema inicial de Power Builder.	
	Línea 4: El sistema no puede procesar el archivo, porque no está de acuerdo al formato establecido (9 variables).	

CASO DE USO: SIMULAR Y PREDECIR ICA

Tabla N° 4.32. Simular y predecir ICA

Caso de Uso:	Simular y predecir ICA	
Actores:	Administrador	
Propósito:	Simular la red con la data importada, para predecir el ICA	
Resumen:	El administrador, con la data que ha sido importada anteriormente, procederá a realizar la simulación con dicha data, de tal manera que el resultado obtenido servirá como dato de la predicción del ICA.	
Acciones de los eventos:		
	Acción de los Actores	Respuesta del Sistema
	1. Selecciona la opción “Realizar predicción”	2. El sistema muestra la ventana seleccionada
	3. Presiona el botón “Simular y predecir ICA”	4. El sistema captura las variables globales, y la de la data importada.
		5. El sistema invoca a la función sim().
		6. El sistema realiza la simulación
		7. El sistema calcula el error cuadrado medio de la simulación realizada
		8. El sistema muestra los resultados de la simulación en pantalla
		9. El sistema habilita los botones de aproximación y gráficas
Casos Alternos		
Línea 6: El sistema no puede realizar la simulación, debido a que no ha sido creada o no ha sido entrenada.		

CASO DE USO: REALIZAR APROXIMACIÓN

Tabla N° 4.33. Realizar aproximación

Caso de Uso:	Realizar aproximación	
Actores:	Administrador	
Propósito:	Convertir el resultado de la predicción en datos entendibles por el sistema web	
Resumen:	El administrador, con los resultados obtenidos, los convierte en función de “0” y “1”, de tal manera que pueda haber una correspondencia con los datos de salida que se están trabajando.	
Acciones de los eventos:		
	Acción de los Actores	Respuesta del Sistema
		1. El sistema realizó la simulación para la predicción.
	2. Presiona el botón “Realizar aproximación”	3. El sistema captura la variable global, que contiene la data resultado.
		4. El sistema verifica que el resultado obtenido, sea un resultado aceptado por el modelo
		5. El sistema aproxima cada uno de los 5 valores a 0 o 1,
		6. El sistema muestra en pantalla el resultado y lo almacena en variable global.
		7. Mensaje satisfactorio
Casos Alternos		
Línea 3: El sistema no realizará la aproximación debido a que aun no se invoca a la función sim().		

CASO DE USO: EXPORTAR RESULTADO DE PREDICCIÓN OBTENIDA

Tabla N° 4.34. Exportar resultado de predicción obtenida

Caso de Uso:	Exportar resultado de predicción obtenida	
Actores:	Administrador	
Propósito:	Exportar el resultado de la predicción para su tratamiento por el sistema web	
Resumen:	El administrador, exportará el resultado y la aproximación en un archivo Excel, para que pueda ser interpretado por el sistema web y muestre las predicciones a los usuarios finales (Público en general).	
Acciones de los eventos:		
	Acción de los Actores	Respuesta del Sistema
	1. El administrador invoca a la opción "Exportar salida obtenida"	2. El sistema captura la variable global, que contiene la data resultado inicial y resultado aproximación.
		3. El sistema escribe las matrices resultado en archivo Excel.
		4. El sistema graba los archivos en el disco duro
		5. El sistema muestra mensajes satisfactorios de exportación
Casos Alternos		
Línea 3: El sistema no realizará la exportación debido a que aún no se realiza la aproximación.		

CASO DE USO: PROCESAR RESULTADO MATLAB

Tabla N° 4.35. Procesar resultados Matlab

Caso de Uso:	Procesar resultados Matlab	
Actores:	Administrador	
Propósito:	Procesar los resultados de Matlab, para que pueda ser mostrado en el portal.	
Resumen:	El administrador, procesará los resultados a través del sistema web para que muestre las predicción a 24 horas, a los usuarios finales (Público en general).	
Acciones de los eventos:		
	Acción de los Actores	Respuesta del Sistema
	1. El administrador invoca a la función "Procesar"	2. El sistema web verifica la existencia de los archivo que exporto Matlab.
		3. El sistema verifica la data que contiene dichos archivos.
		4. El sistema procesa los archivos.
Casos Alternos		
	Línea 2: El sistema rechaza el proceso, porque no se generaron dichos archivos.	

CASO DE USO: VISUALIZAR MONITOREO DE PREDICCIÓN DE ICA

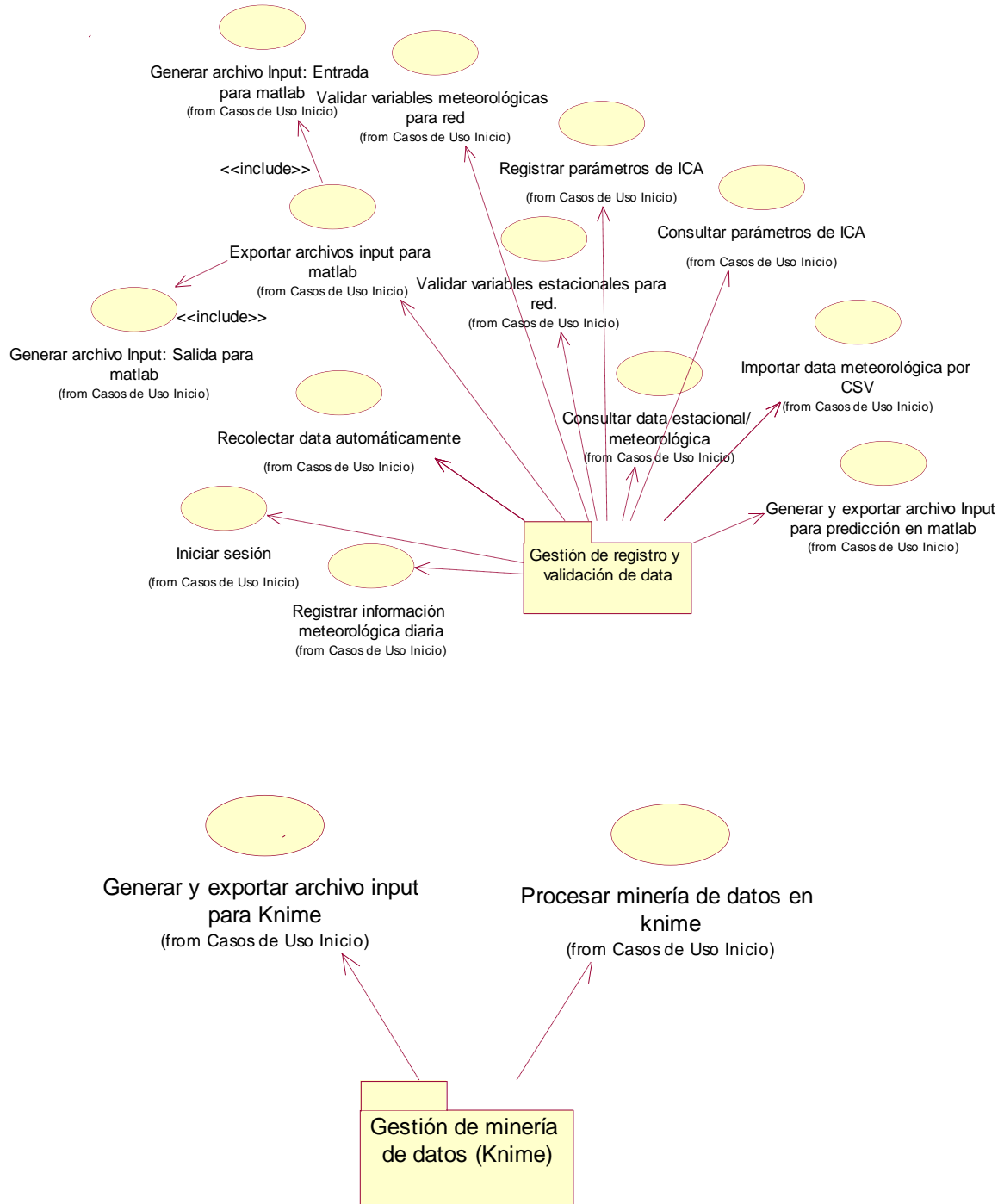
Tabla N° 4.36. Visualizar monitoreo de predicción de ICA

Caso de Uso:	Visualizar monitoreo de predicción de ICA	
Actores:	Administrador, público en general (usuario)	
Propósito:	Muestra los resultados de predicción a través del portal.	
Resumen:	El administrador, mostrará los resultados del ICA, con una predicción de 24 horas, a través del portal web, para que puedan ser visualizados por el público en general, así como la interpretación de este, sus colaboraciones y las acciones que se pueden tomar.	
Acciones de los eventos:		
	Acción de los Actores	Respuesta del Sistema
		1. El sistema genera un mapa geográfico, donde se ubican las 4 estaciones.
	2. El usuario solicita visualizar el ICA, para 24 horas siguientes, para una determinada estación	3. El sistema web captura el archivo resultado, que ha sido exportado de Matlab.
		4. Verifica que corresponda con algún parámetro registrado en la base de datos.
		5. De acuerdo a la correspondencia, procesa los resultados, con todos los valores que se van a mostrar: PM10, ICA, resultado, Colaboraciones, acciones, color.
		6. Muestra resultado en el mapa
Casos Alternos		
Línea 3: El sistema rechaza el proceso, porque no se generaron los archivos de predicción a través de Matlab, para las 24 horas.		

4.2.2.2. Modelo de Análisis del Sistema

4.2.2.2.1. Análisis de la Arquitectura

a) Identificación de paquetes de Análisis a partir de los Casos de Uso



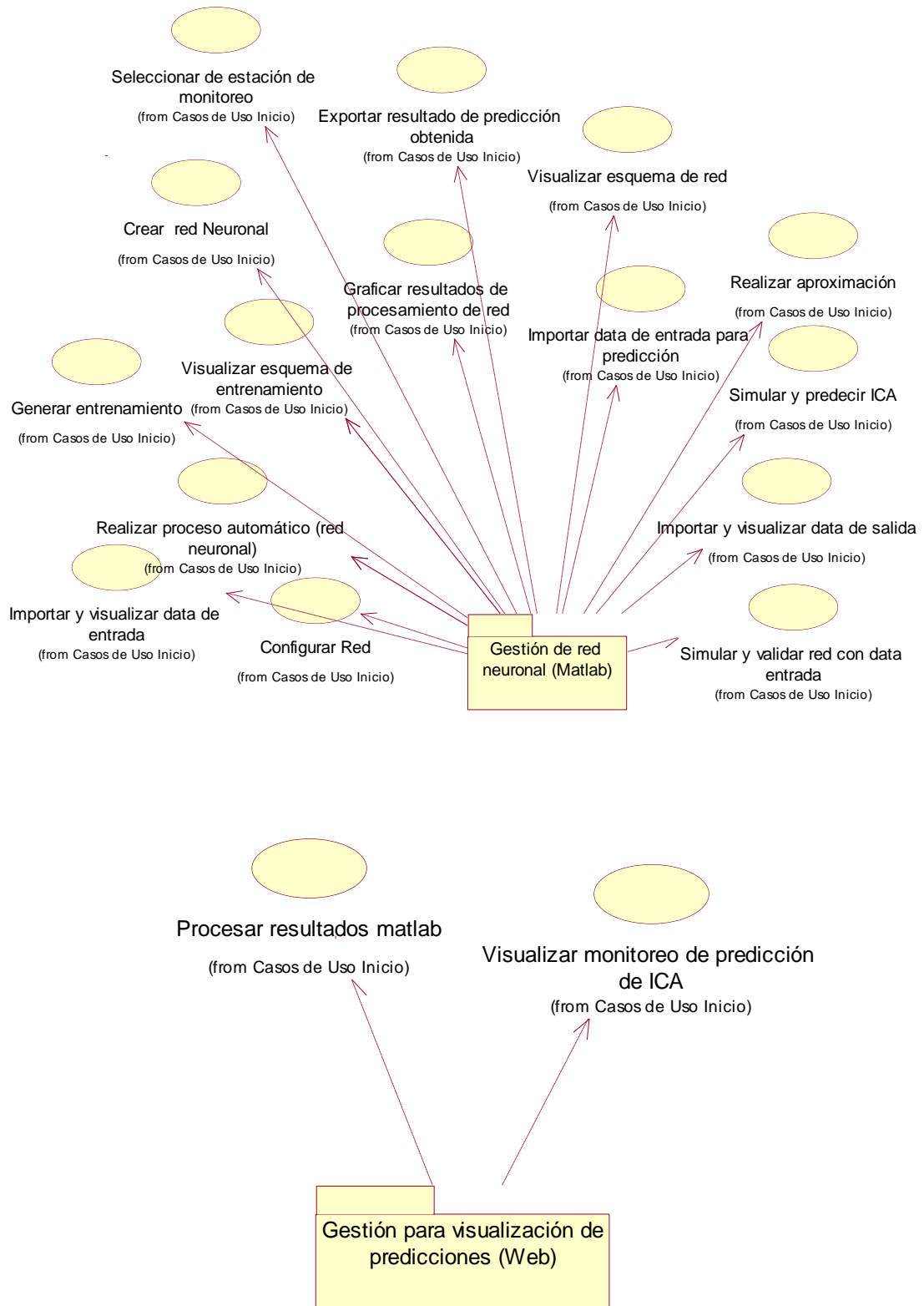


Figura 4.47. Identificación de paquetes de análisis – Casos de Uso

b) Identificación de paquetes de Análisis generales a partir de Clases de Dominio

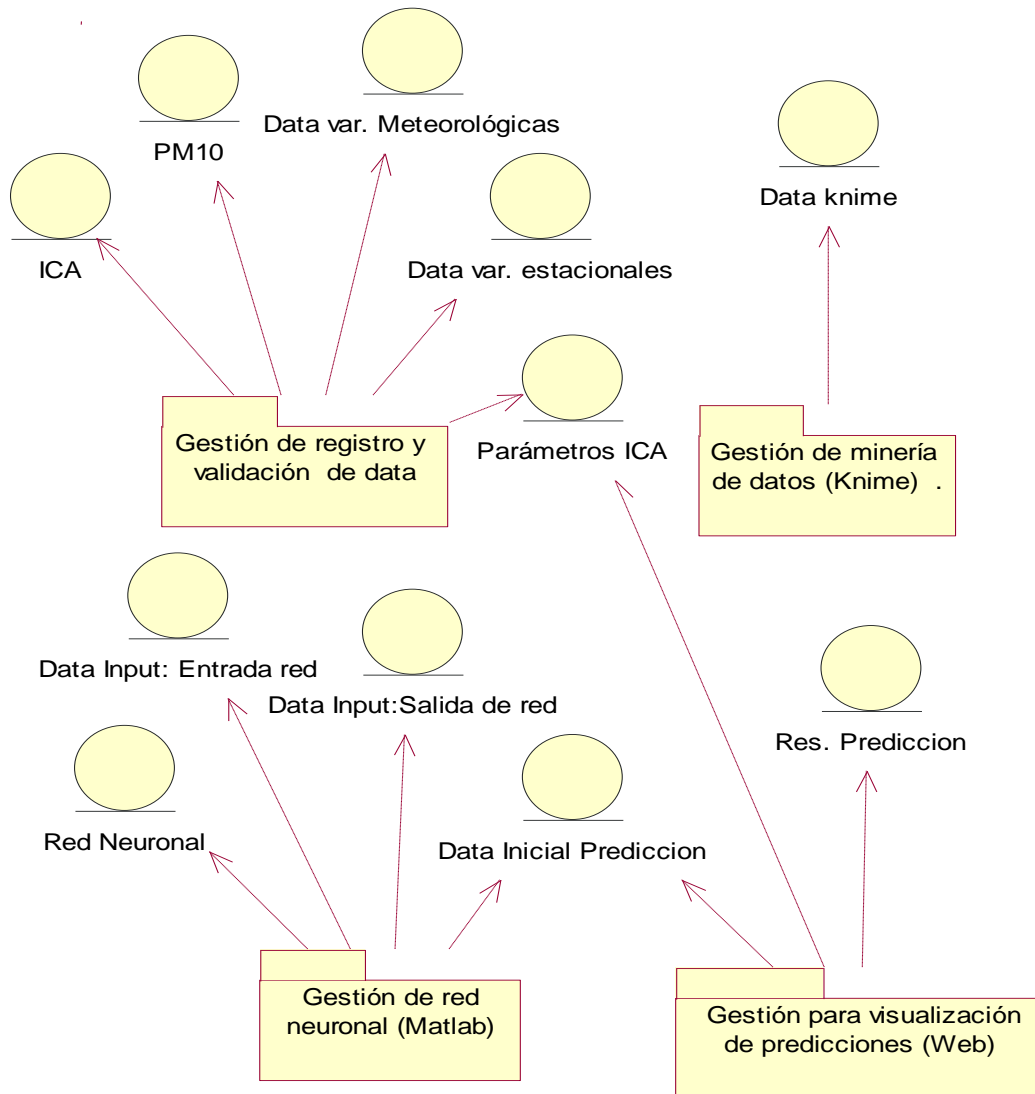


Figura 4.48. Identificación de paquetes de análisis – Clases de Dominio

4.2.2.2. Analizar Casos de Uso

Realización del Caso de Uso Iniciar sesión.

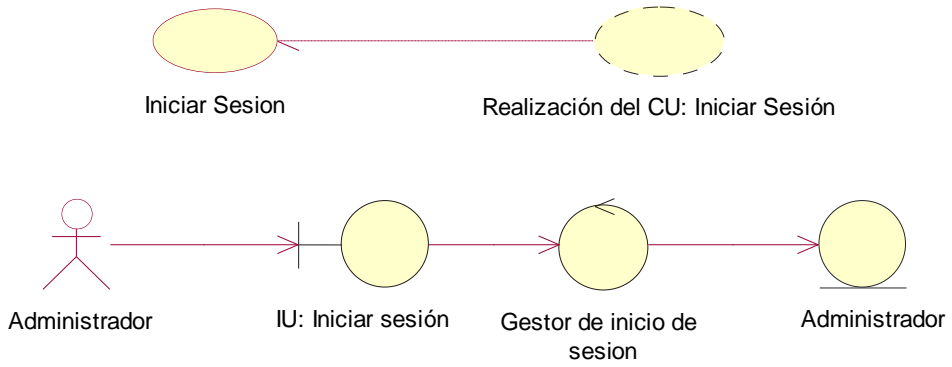


Figura 4.49. Realización del CU Iniciar Sesion

Realización del Caso de Uso Registrar parámetros ICA

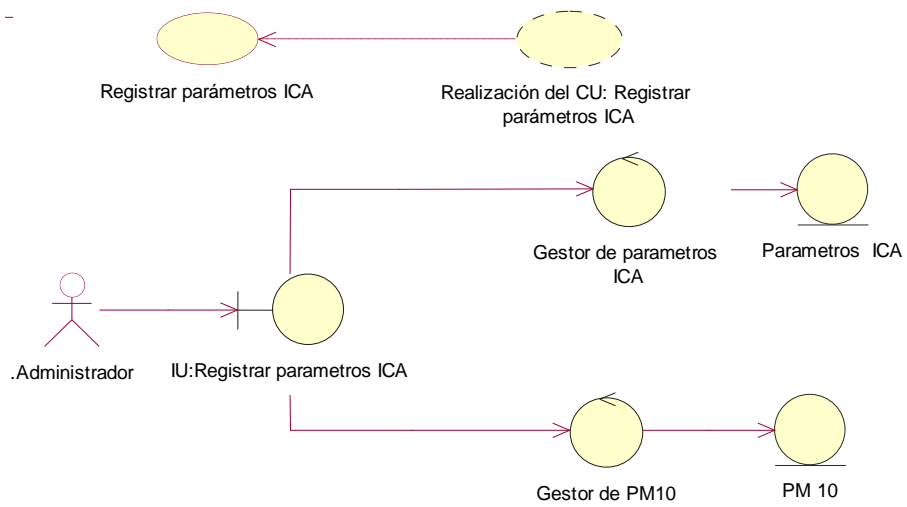


Figura 4.50. Realización del CU Registrar parámetros ICA

Realización del Caso de Uso registrar Información meteorológica diaria

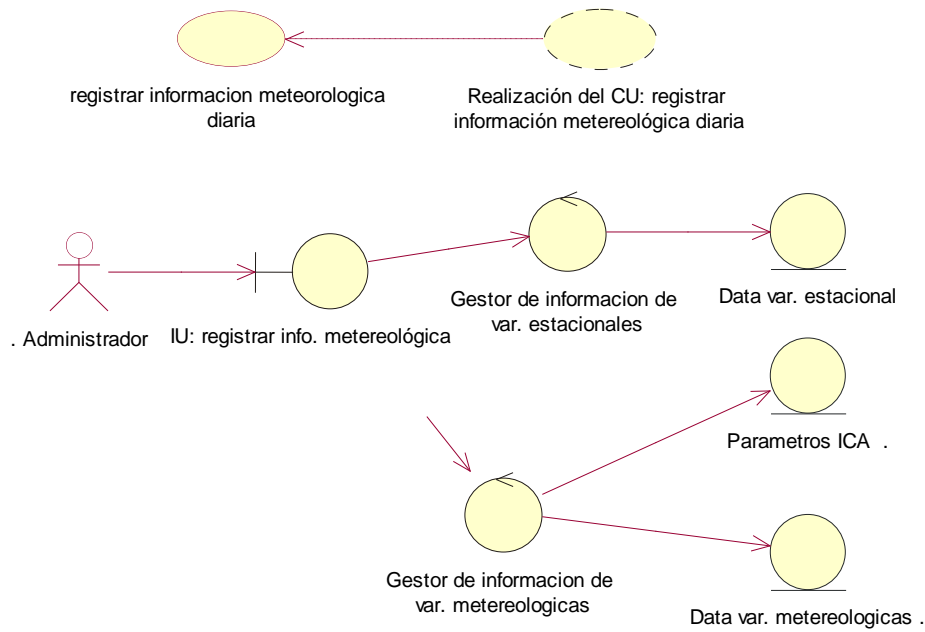


Figura 4.51. Realización del CU registrar información meteorológica diaria

Realización del Caso de Uso validar variables

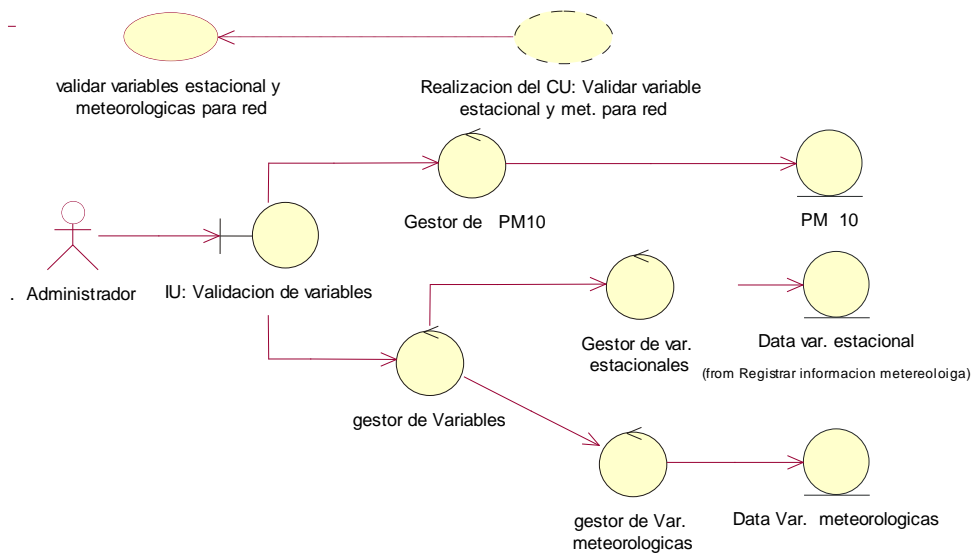


Figura 4.52. Realización del CU validar variables

Realización del Caso de Uso generar Archivo input entrada

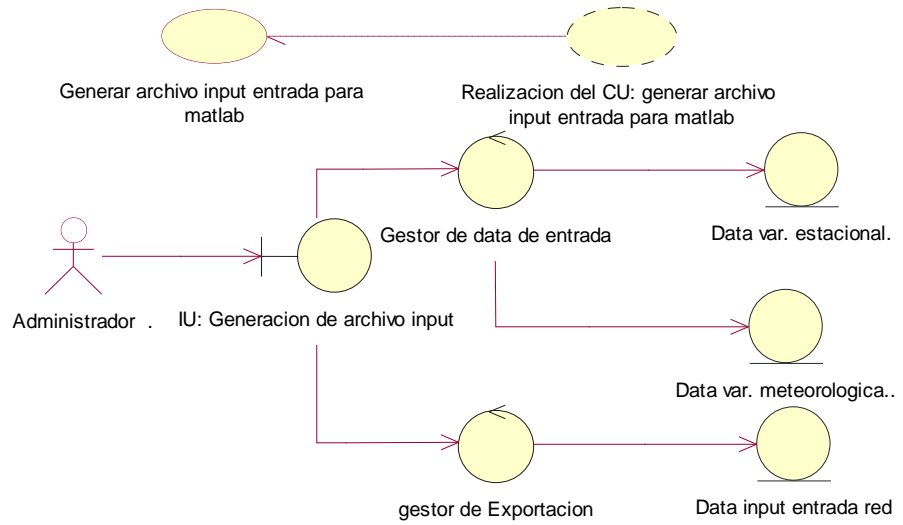


Figura 4.53. Realización del CU generar Archivo input entrada

Realización del Caso de Uso generar Archivo input salida

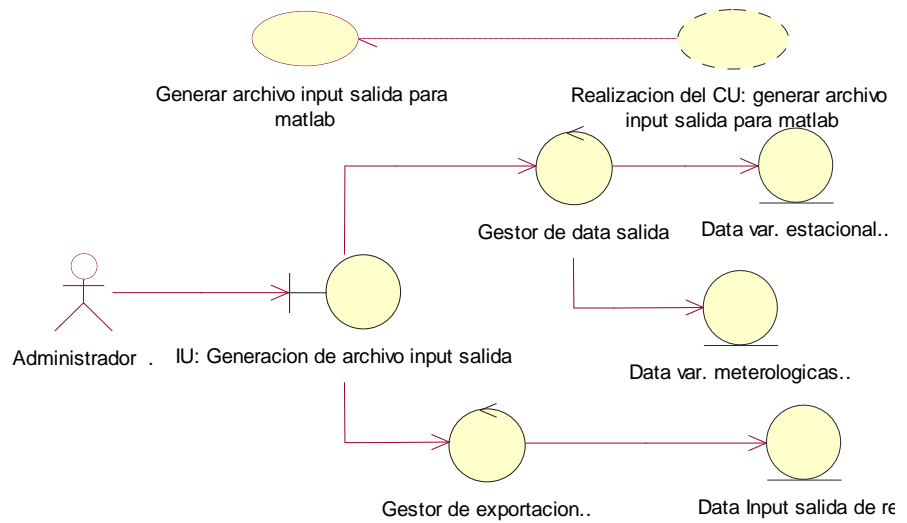


Figura 4.54. Realización del CU generar Archivo input salida

4.2.2.3. MODELO DE DISEÑO

a) Diseño de la Arquitectura

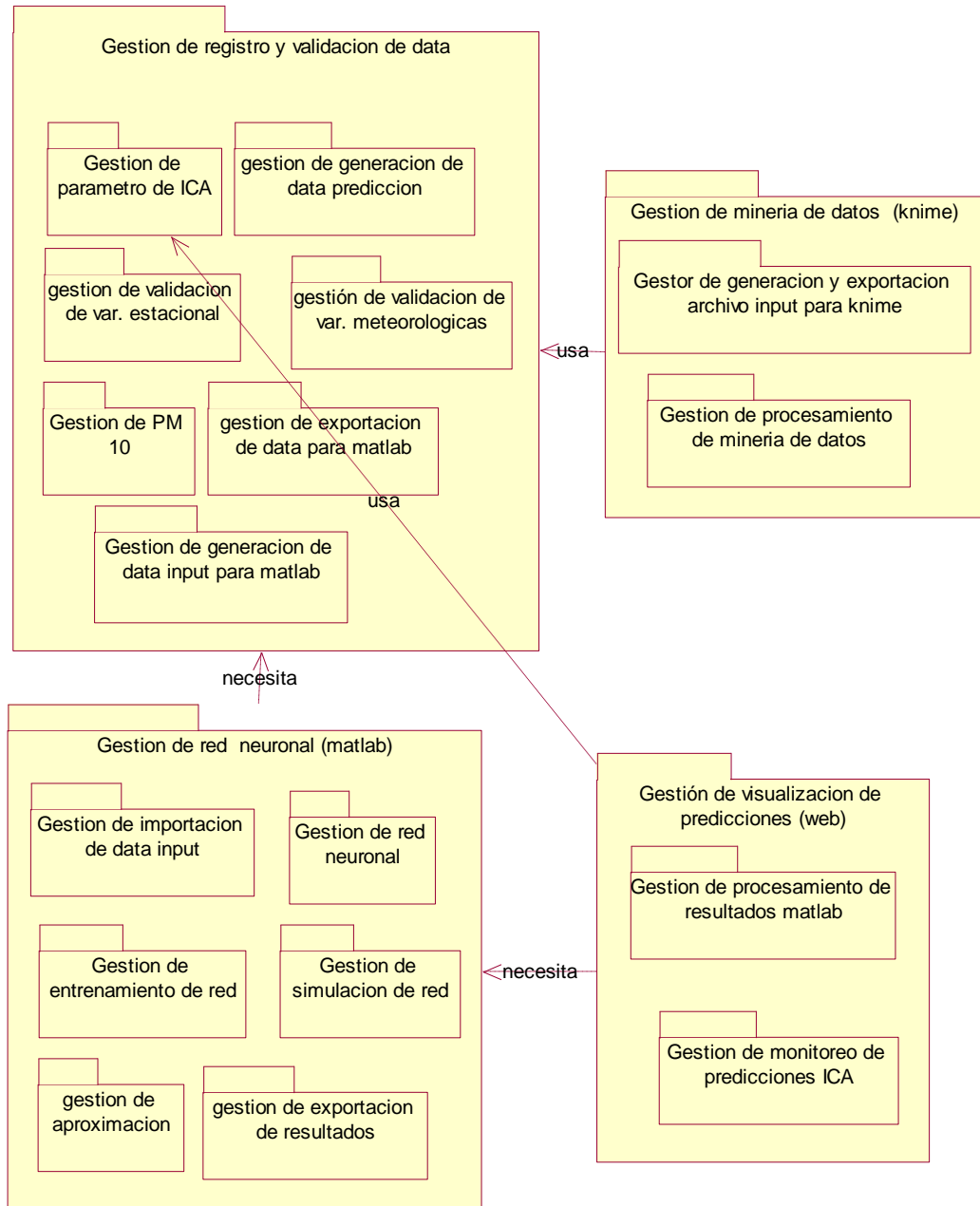


Figura 4.55. Diseño de la Arquitectura

b) Diseño de casos de uso

Diagrama de Secuencia del CU: Iniciar sesión.

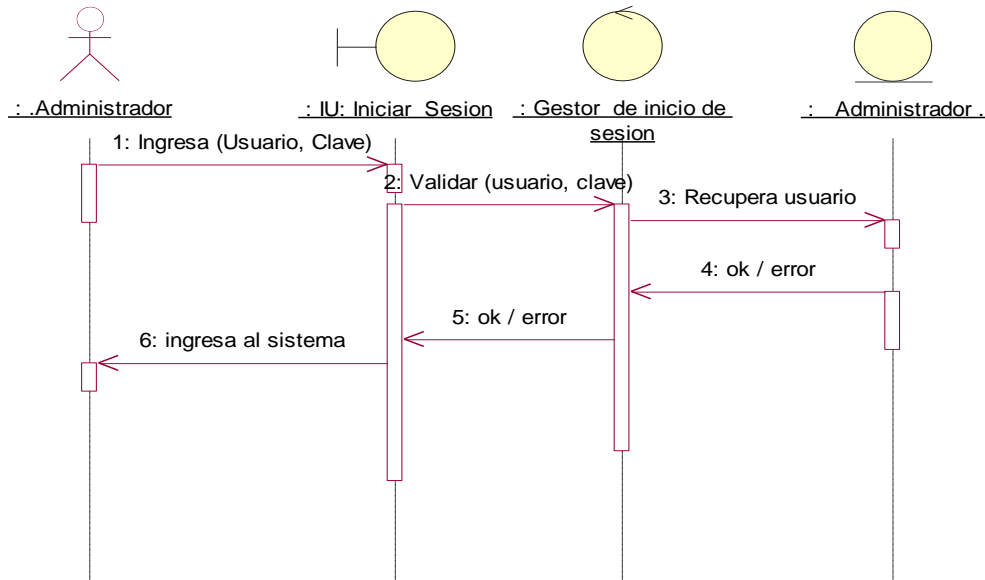


Figura 4.56. Diagrama de Secuencia CU Iniciar Sesión

Diagrama de Secuencia del CU: Registrar parámetros de ICA

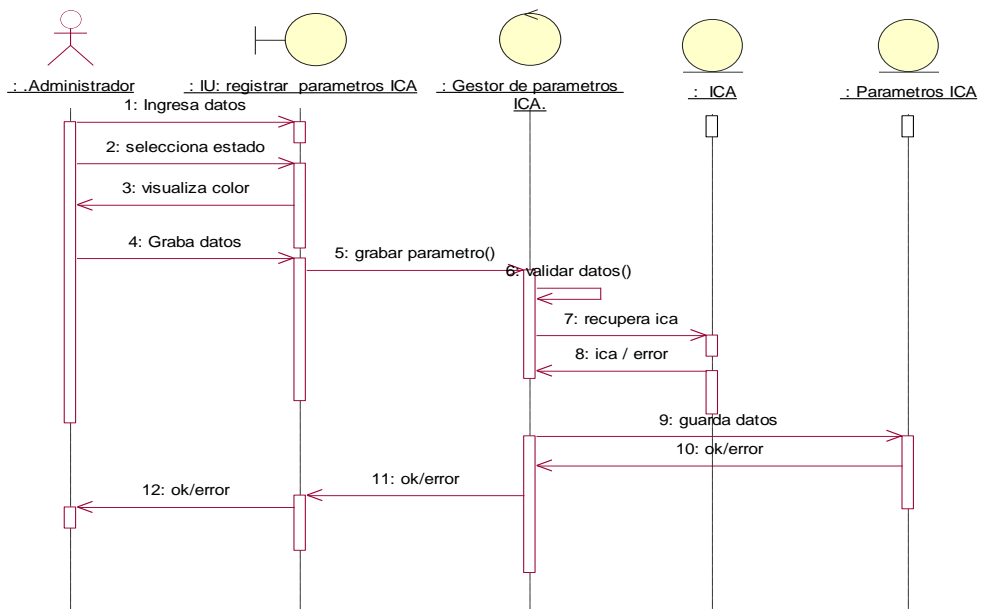


Figura 4.57. Diagrama de Secuencia CU: Registrar parámetros de ICA

Diagrama de Secuencia del CU: Registrar información meteorológica diaria

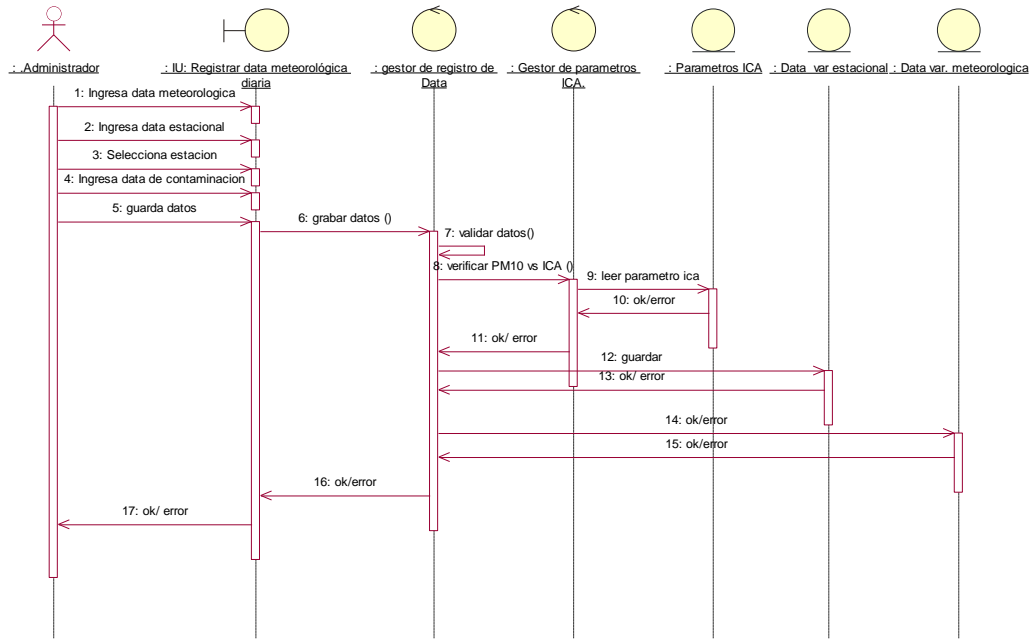


Figura 4.58. Diagrama de Secuencia CU: Registrar información meteorológica

Diagrama de Secuencia del CU: variables estacionales y meteorológicas para red

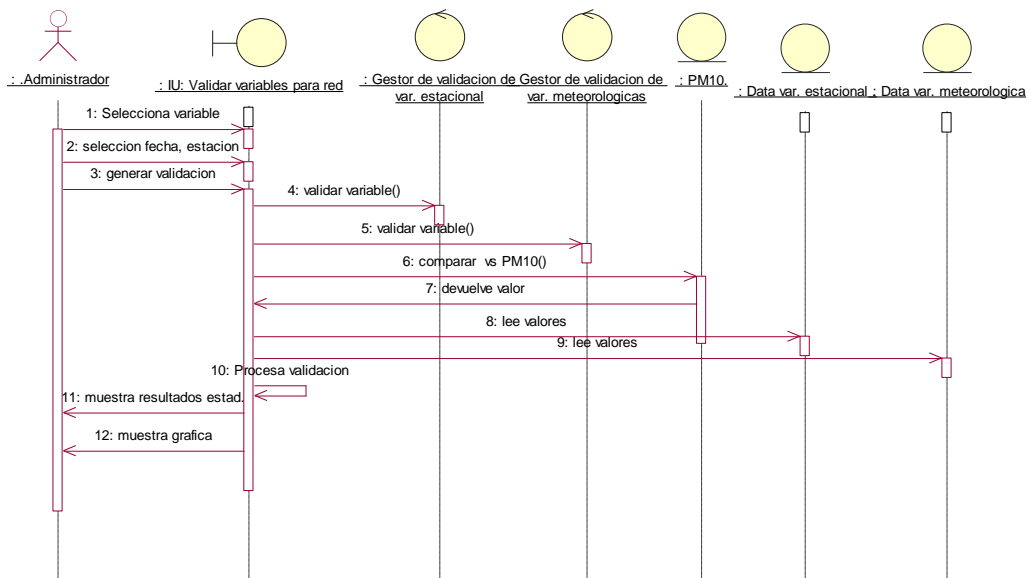


Figura 4.59. Diagrama de Secuencia CU: variables estacionales y meteorológicas para red

Diagrama de Secuencia del CU: Generar archivo Input: Entrada para Matlab

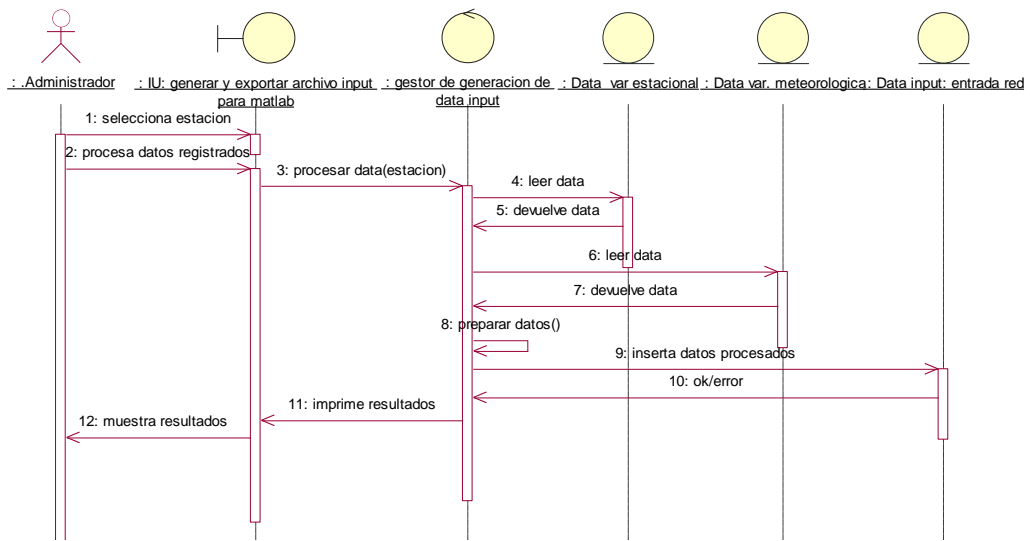


Figura 4.60. Diagrama de Secuencia CU: Generar archivo Input: Entrada para Matlab

Diagrama de Secuencia del CU: Generar archivo Input: Salida para Matlab

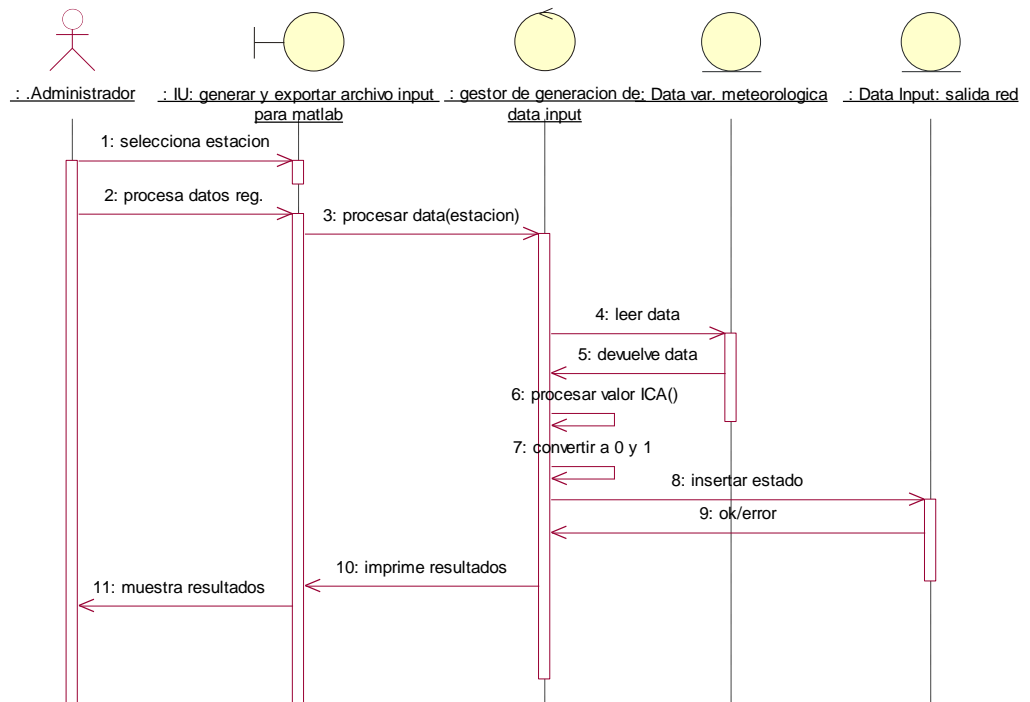


Figura 4.61. Diagrama de Secuencia CU: Generar archivo Input: Salida para Matlab

Diagrama de Secuencia del CU: Exportar archivo Input para matlab

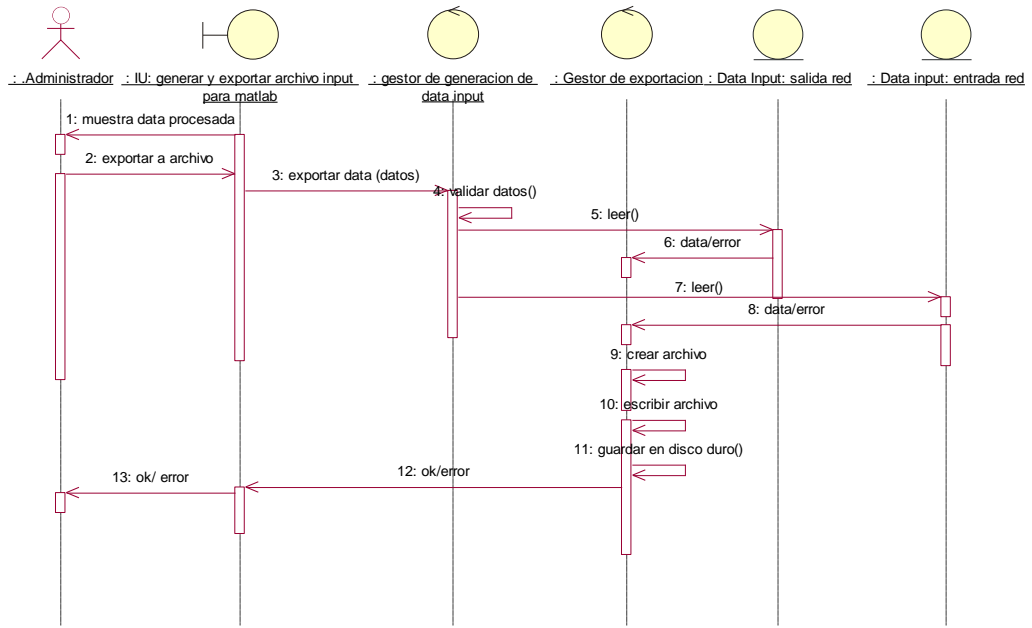


Figura 4.62. Diagrama de Secuencia CU: Exportar archivo Input para Matlab

Diagrama de Secuencia del CU: Seleccionar estación de monitoreo

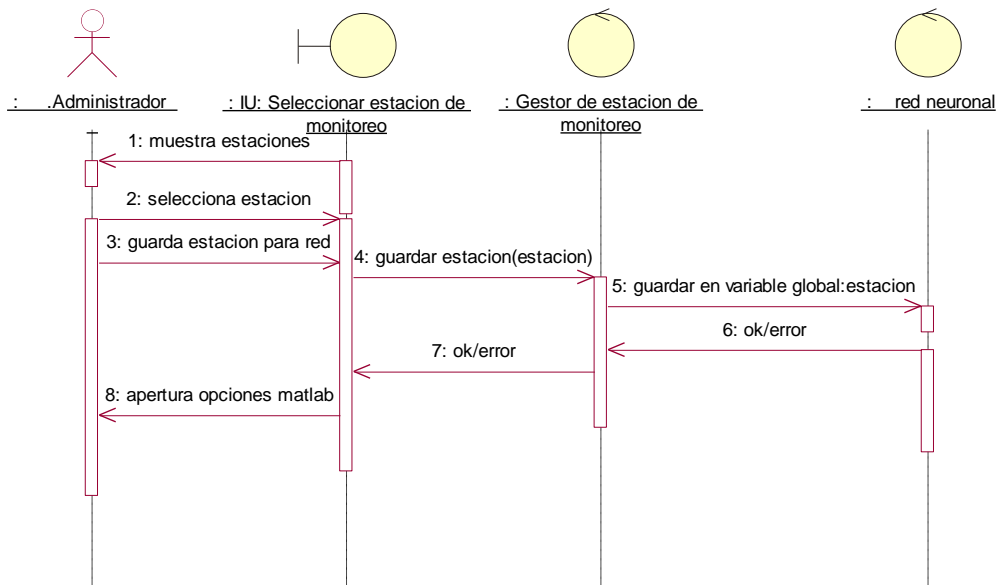


Figura 4.63. Diagrama de Secuencia CU: Seleccionar estación de monitoreo

Diagrama de Secuencia del CU: Importar y visualizar data de entrada

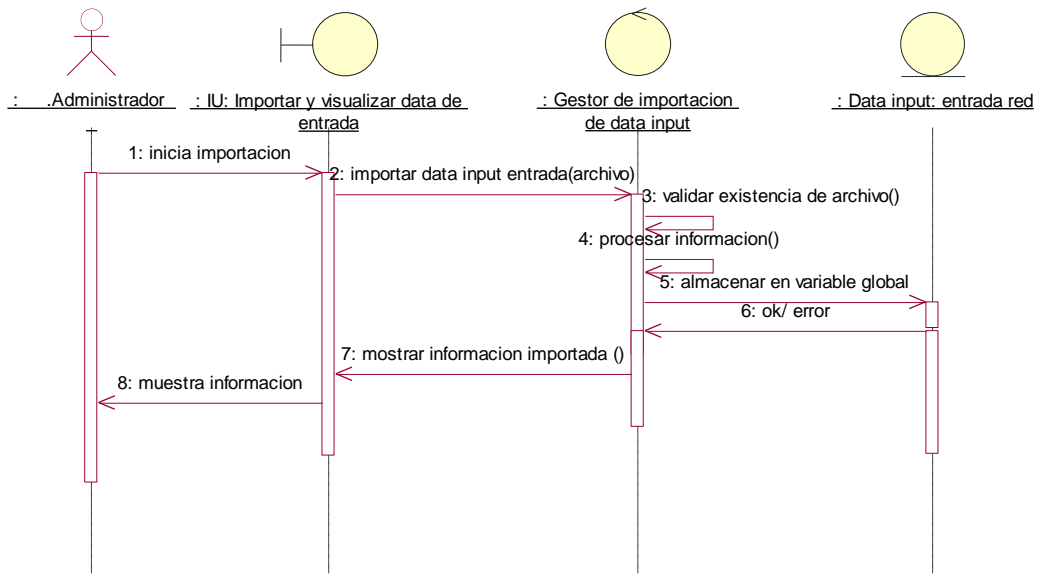


Figura 4.64. Diagrama de Secuencia CU: Importar y visualizar data de entrada

Diagrama de Secuencia del CU: Importar y visualizar data de salida

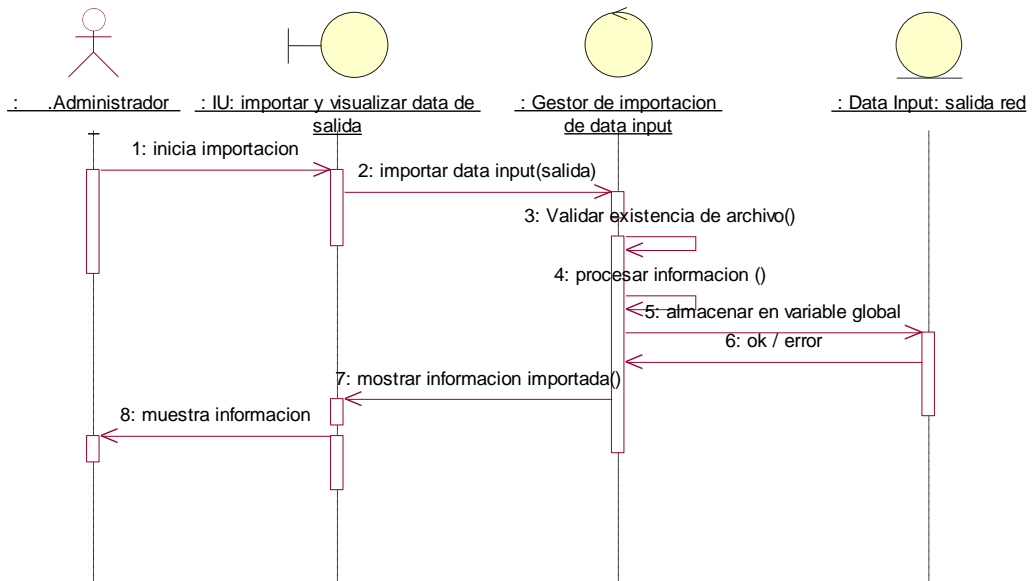


Figura 4.65. Diagrama de Secuencia CU: Importar y visualizar data de salida

Diagrama de Secuencia del CU: Crear red neuronal.

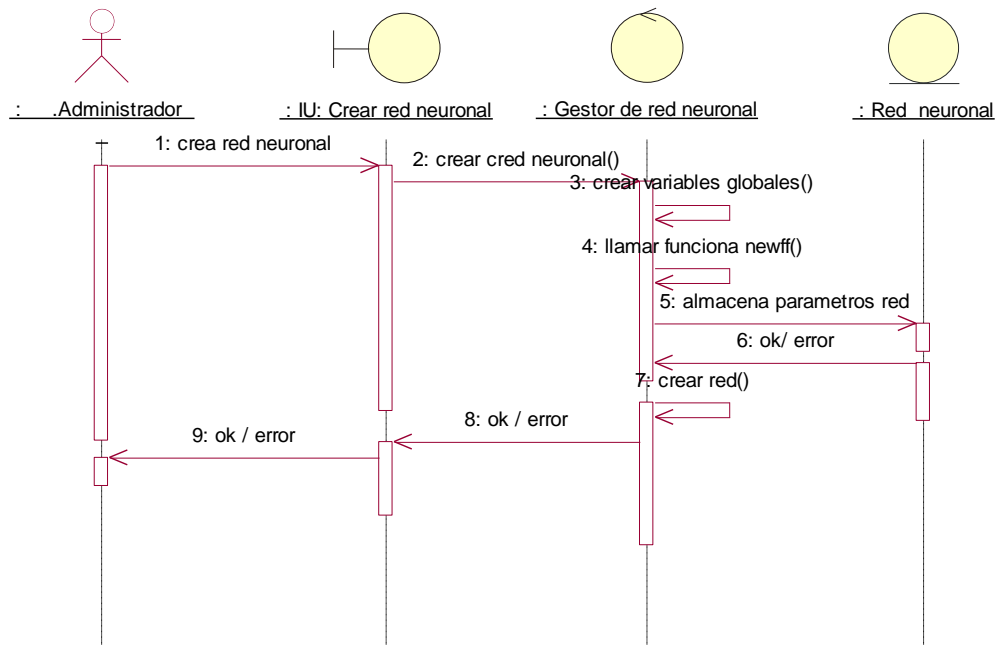


Figura 4.66. Diagrama de Secuencia CU: Crear red neuronal

Diagrama de Secuencia del CU: Generar Entrenamiento

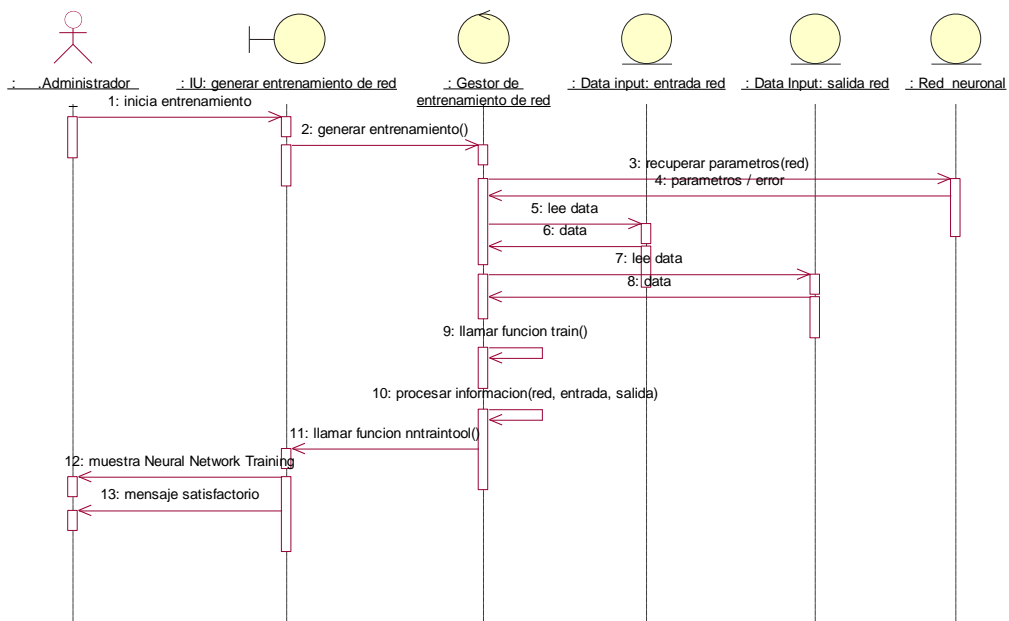


Figura 4.67. Diagrama de Secuencia CU: Generar Entrenamiento

Diagrama de Secuencia del CU: Simular y validar red con data entrenada

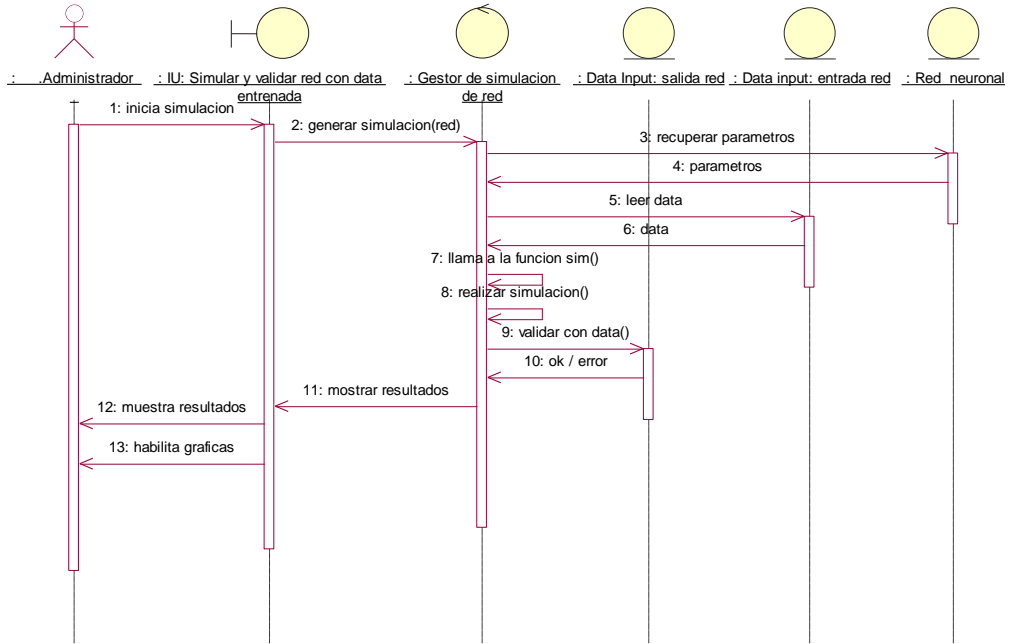


Figura 4.68. Diagrama de Secuencia CU: Simular y validar red con data entrenada

Diagrama de Secuencia del CU: Importar data de entrada para predicción.

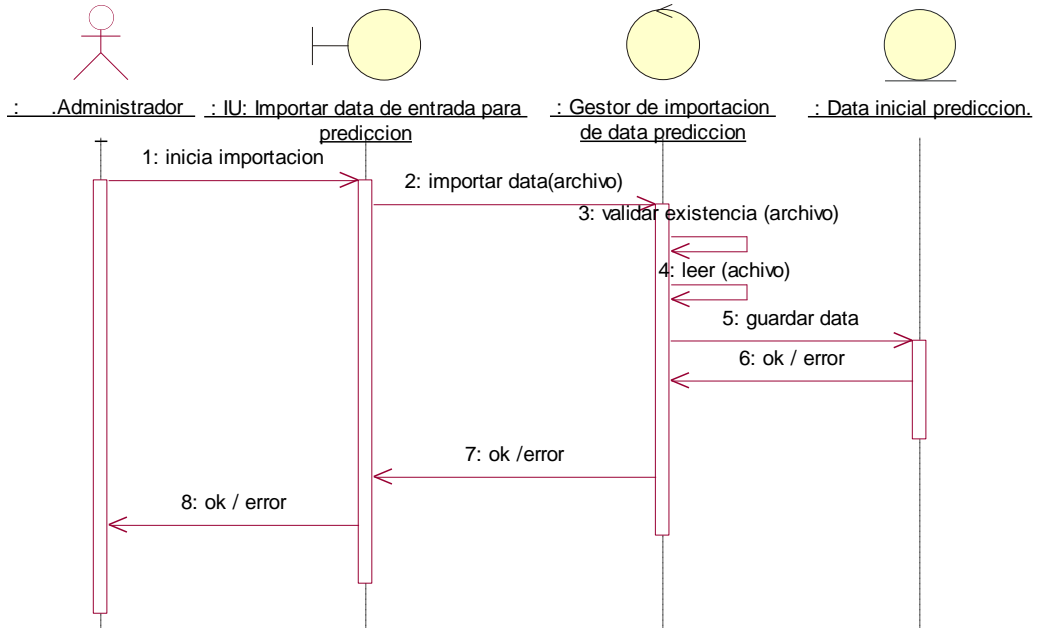


Figura 4.69. Diagrama de Secuencia CU: Importar data de entrada para predicción

Diagrama de Secuencia del CU: Simular y predecir ICA

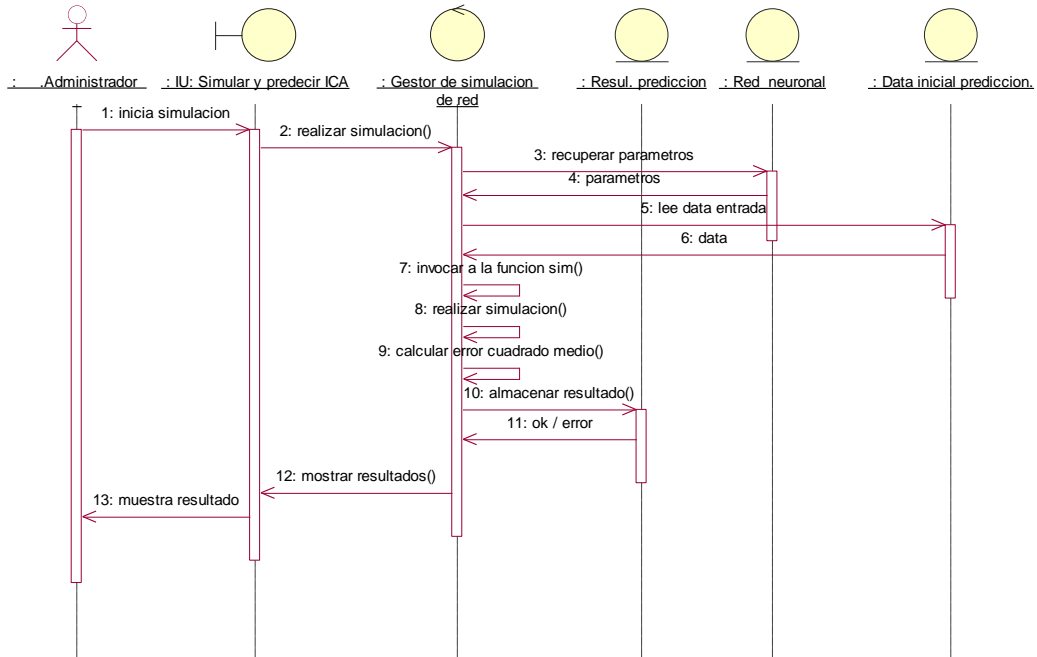


Figura 4.70. Diagrama de Secuencia CU: Simular y predecir ICA

Diagrama de Secuencia del CU: Realizar aproximación

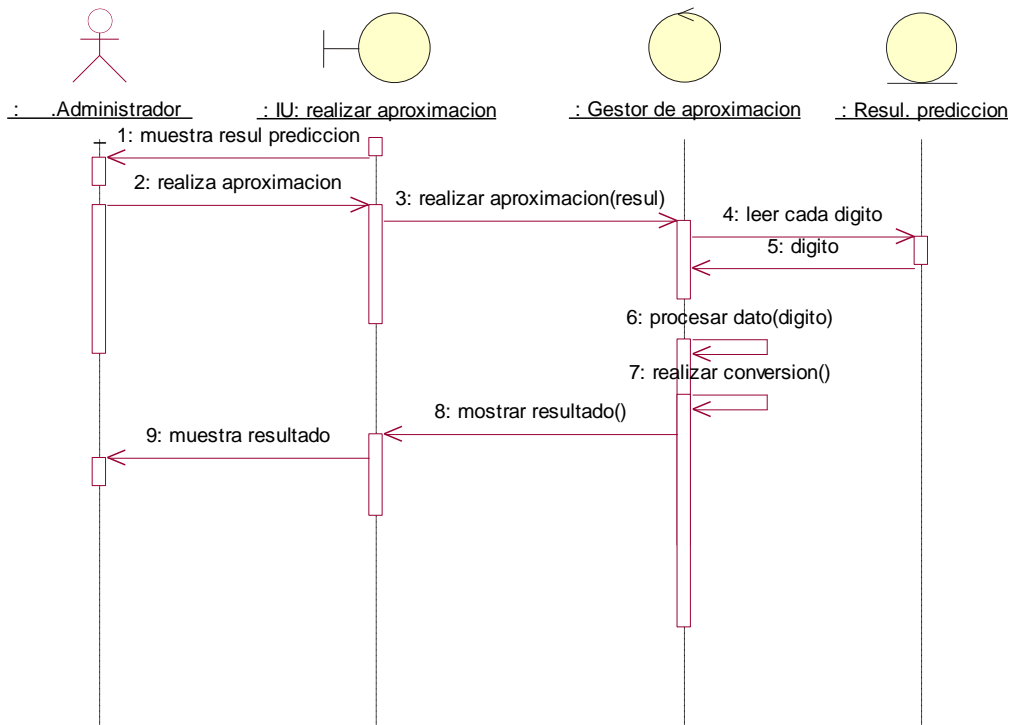


Figura 4.71. Diagrama de Secuencia CU: Realizar aproximación

Diagrama de Secuencia del CU: Procesar resultados Matlab

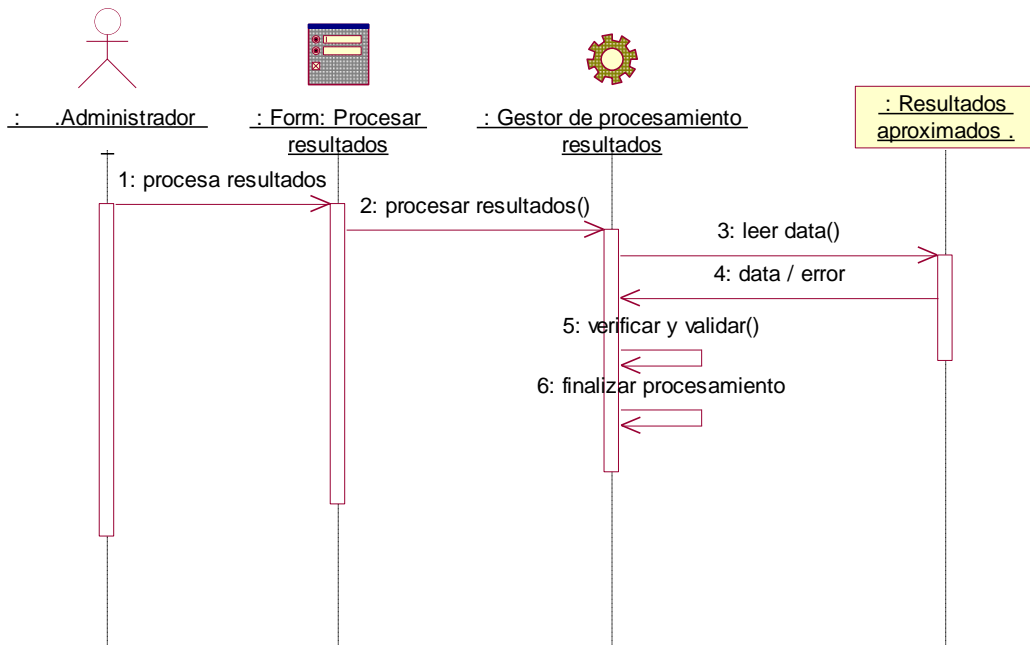


Figura 4.72. Diagrama de Secuencia CU: Procesar resultados Matlab

Diagrama de Secuencia del CU: Visualizar monitoreo de predicción de ICA

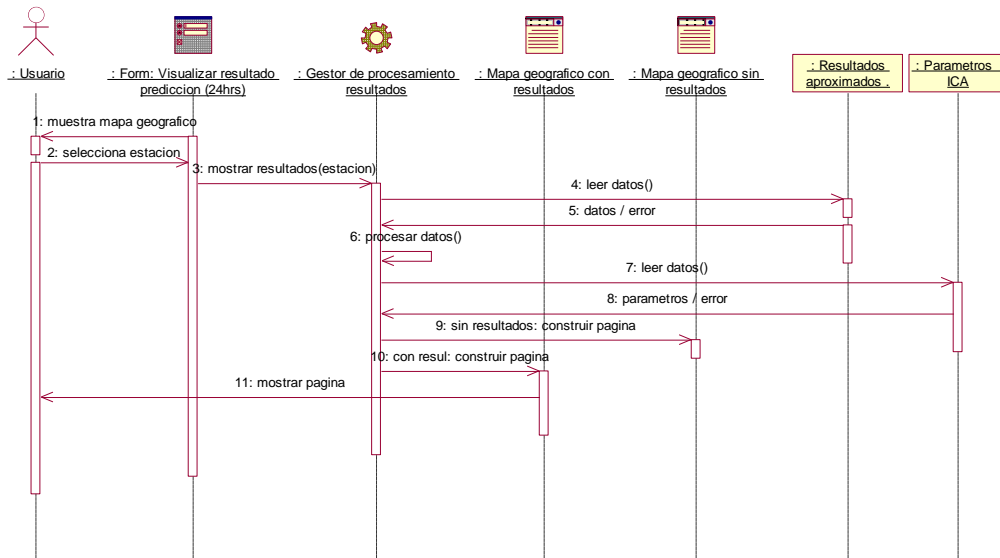


Figura 4.73. Diagrama de Secuencia CU: Visualizar monitoreo de predicción de ICA

c) Diagramas de Colaboración

Diagrama de Colaboración del CU: Iniciar sesión.

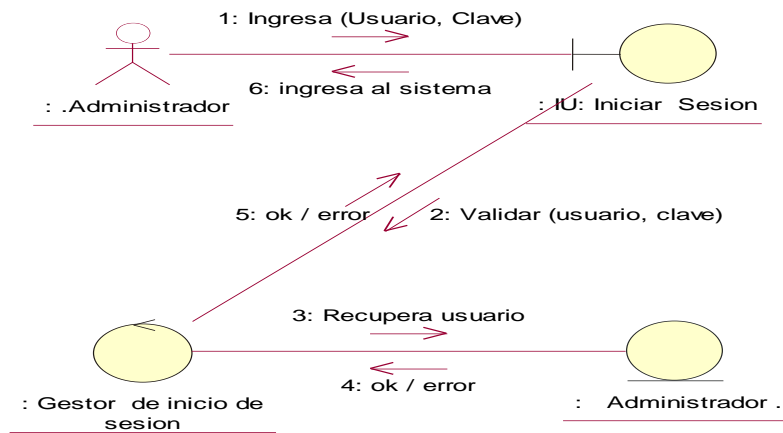


Figura 4.74. Diagrama de Colaboración CU: Iniciar Sesión

Diagrama de Colaboración del CU: Registrar parámetros de ICA

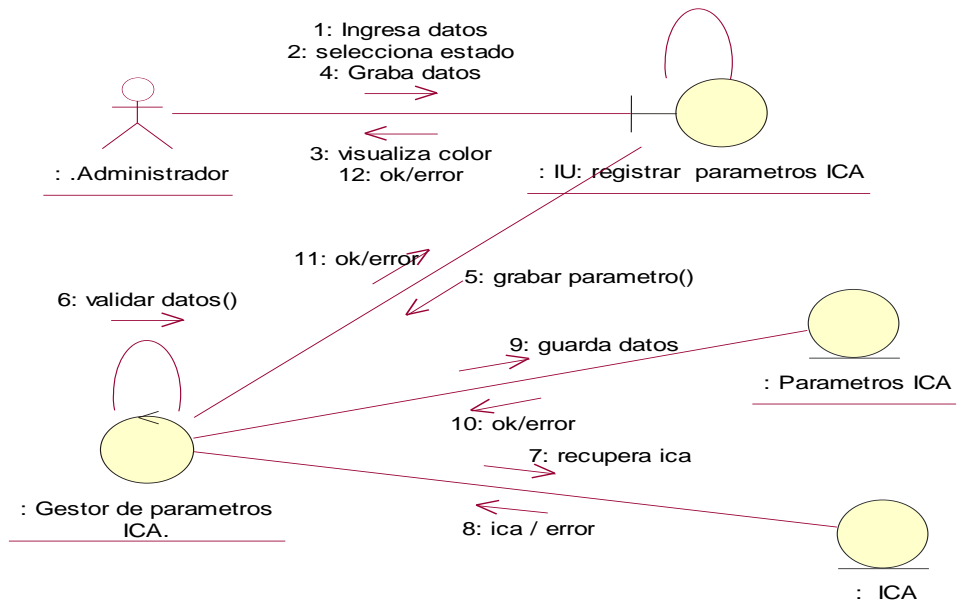


Figura 4.75. Diagrama de Colaboración CU: Registrar parámetros de ICA

Diagrama de Colaboración del CU: Registrar información meteorológica diaria

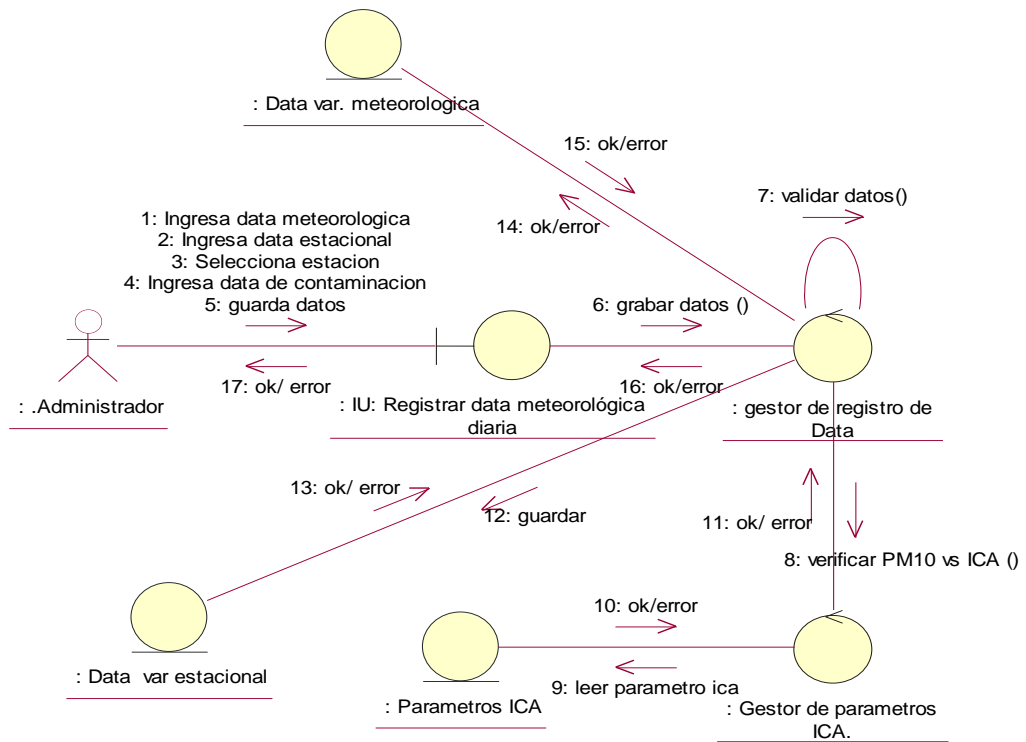


Figura 4.76. Diagrama de Colaboración CU: Registrar información meteorológica

Diagrama de Colaboración del CU: variables estacionales y meteorológicas para red

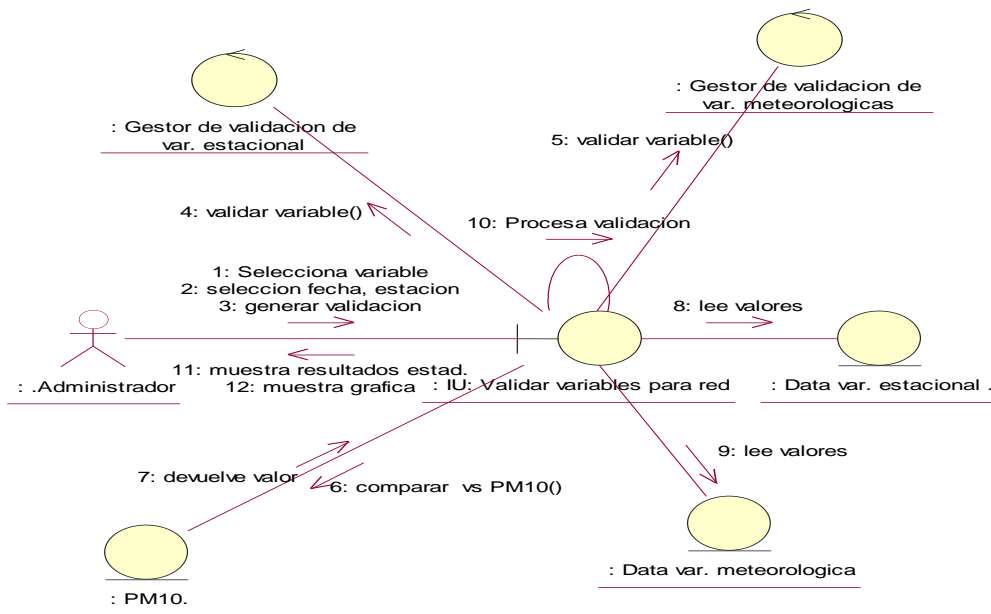


Figura 4.77. Diagrama de Colaboración CU: variables estacionales y meteorológicas para red.

Diagrama de Colaboración del CU: Generar archivo Input: Entrada para matlab

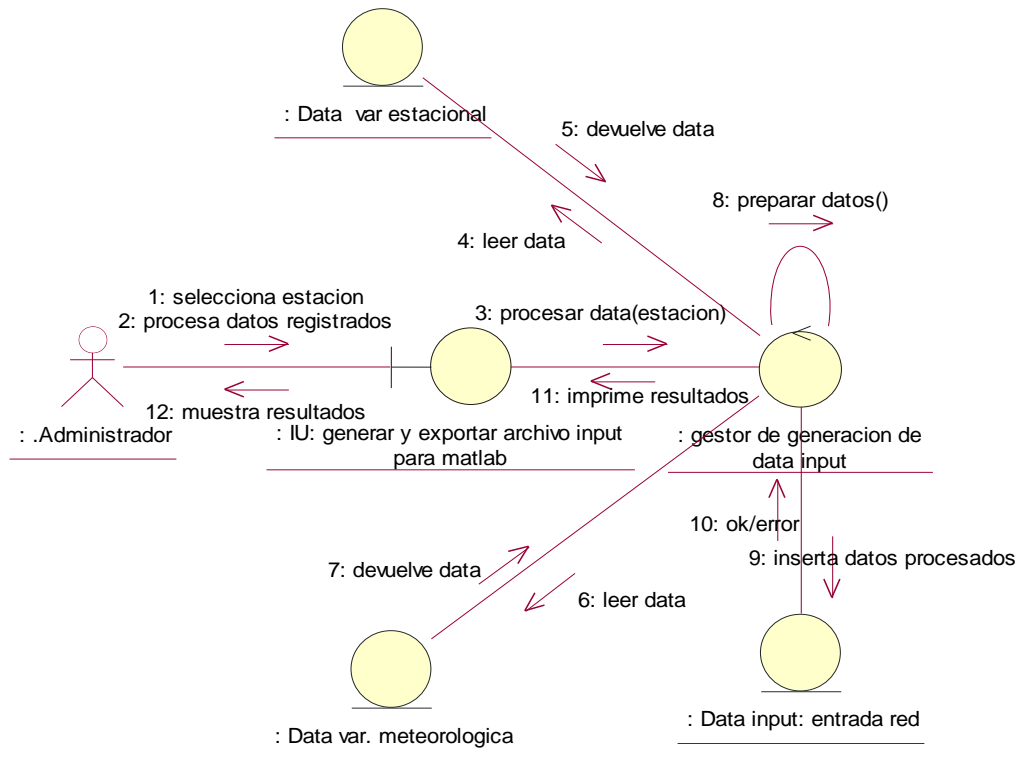


Figura 4.78. Diagrama de Colaboración CU: Generar archivo Input: Entrada para matlab

Diagrama de Colaboración del CU: Generar archivo Input: Salida para matlab

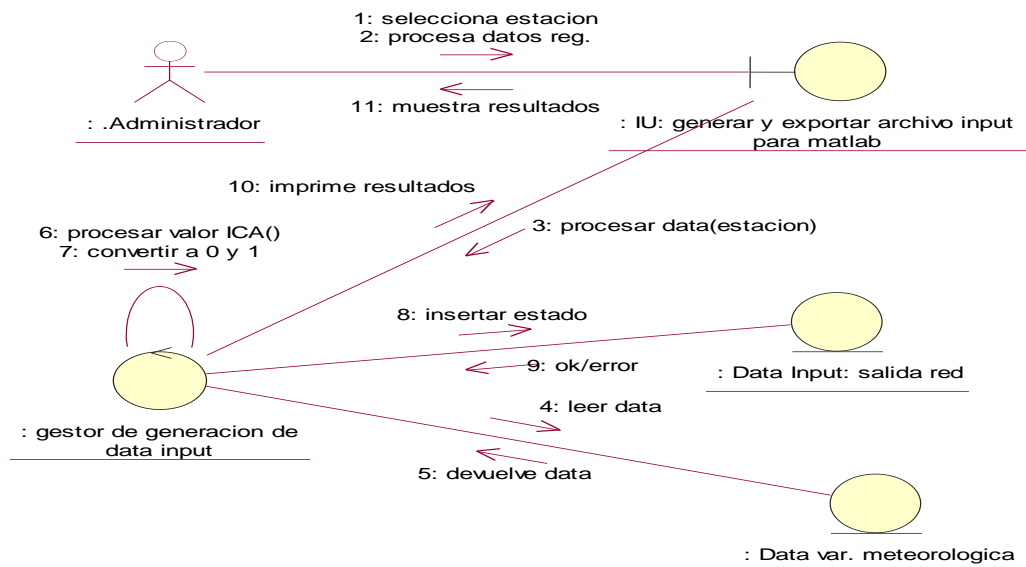


Figura 4.79. Diagrama de Colaboración CU: Generar archivo Input: Salida para matlab

Diagrama de Colaboración del CU: Exportar archivo Input para matlab

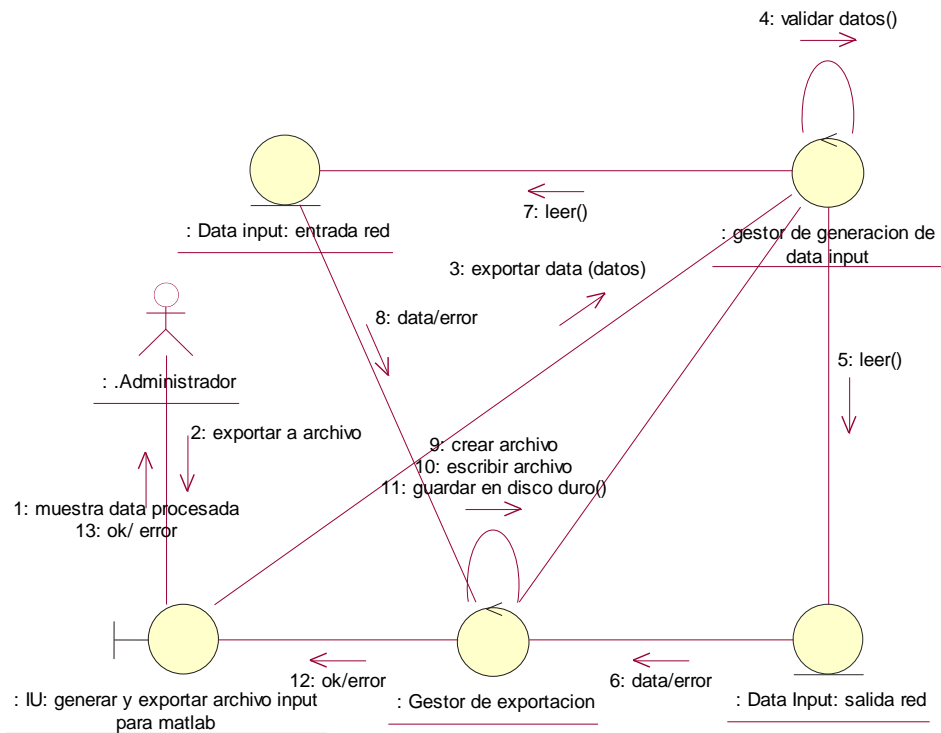


Figura 4.80. Diagrama de Colaboración CU: Exportar archivo Input para Matlab

Diagrama de Colaboración del CU: Seleccionar estación de monitoreo

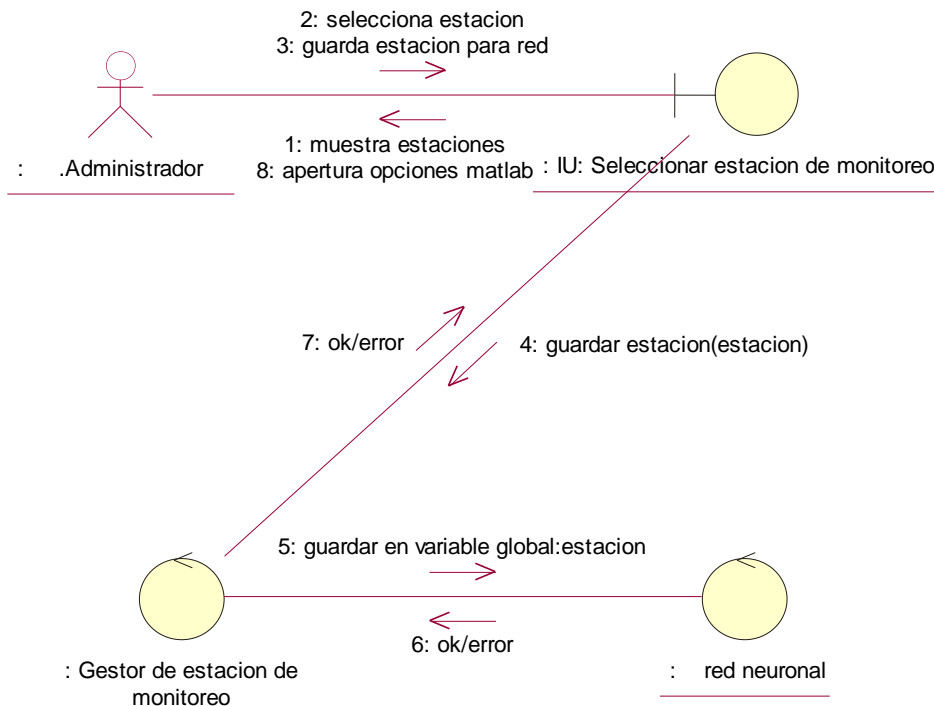


Figura 4.81. Diagrama de Colaboración CU: Seleccionar estación de monitoreo

Diagrama de Colaboración del CU: Importar y visualizar data de entrada

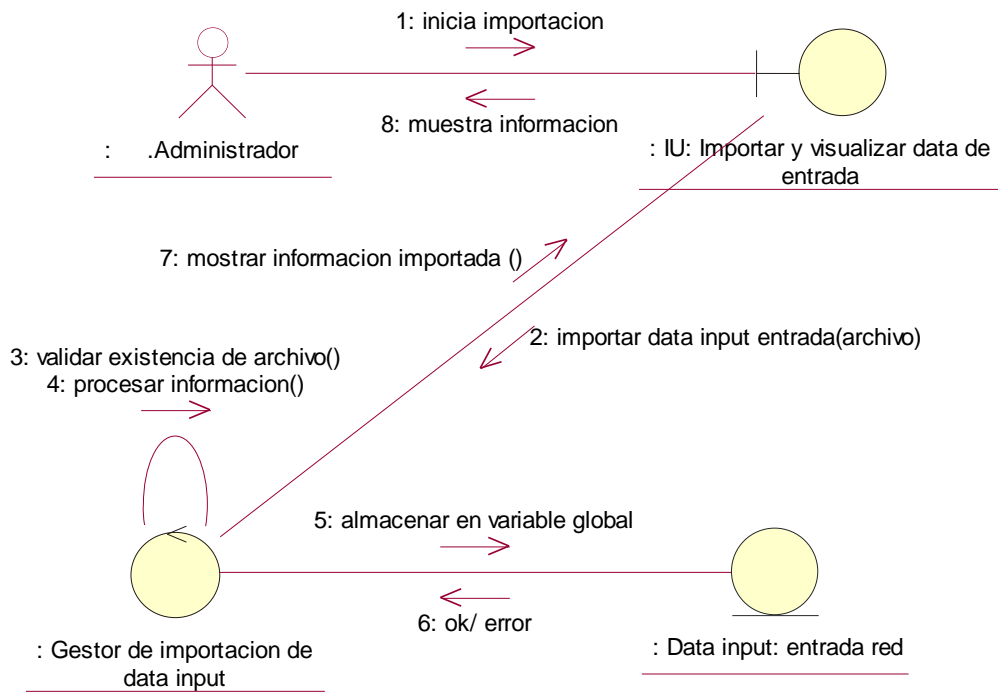


Figura 4.82. Diagrama de Colaboración CU: Importar y visualizar data de entrada

Diagrama de Colaboración del CU: Importar y visualizar data de salida

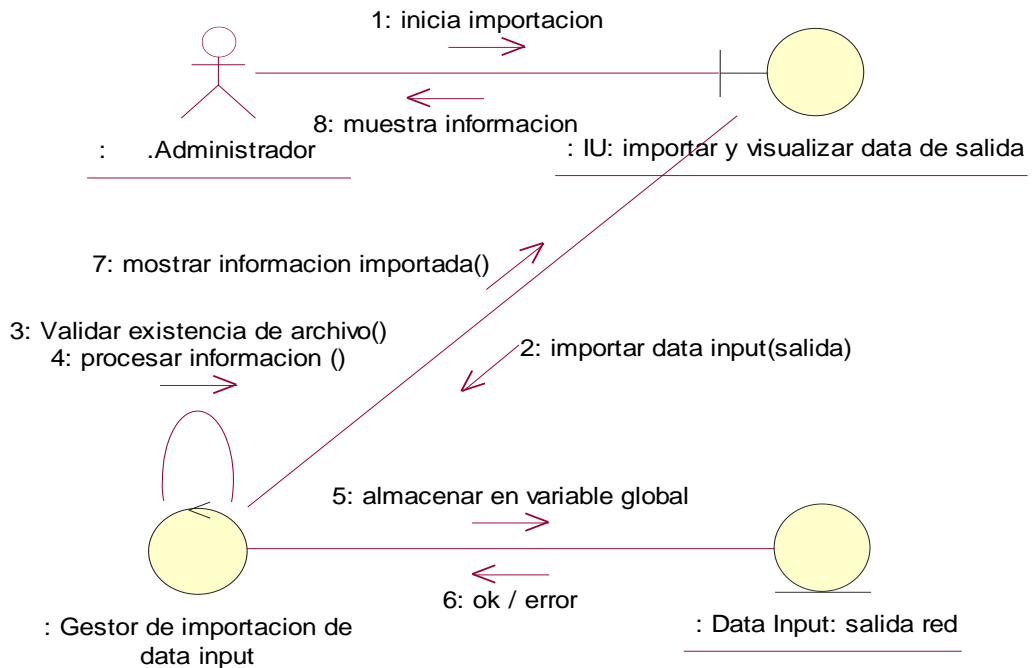


Figura 4.83. Diagrama de Colaboración CU: Importar y visualizar data de salida

Diagrama de Colaboración del CU: Crear red neuronal.

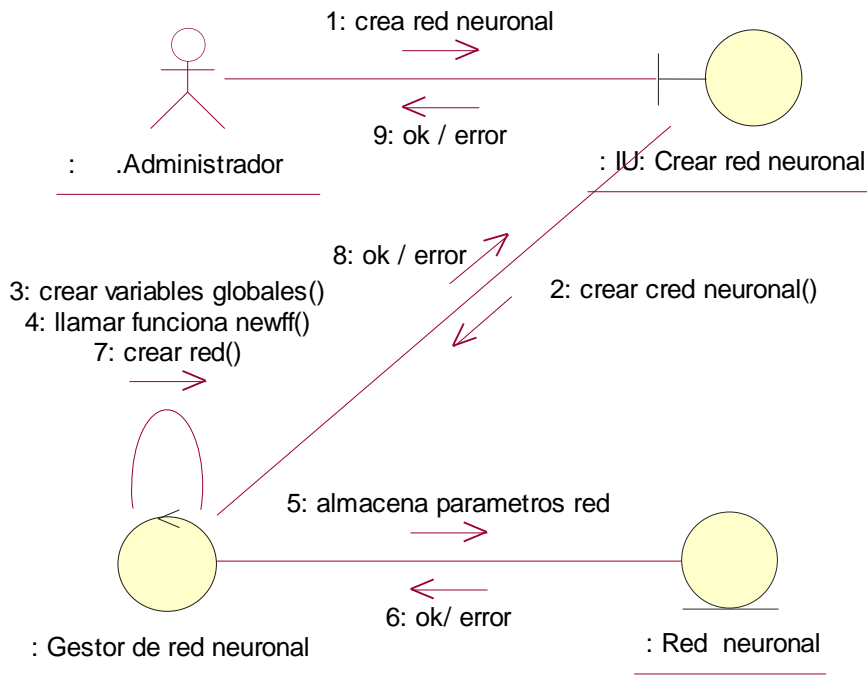


Figura 4.84. Diagrama de Colaboración CU: Crear red neuronal

Diagrama de Colaboración del CU: Generar Entrenamiento

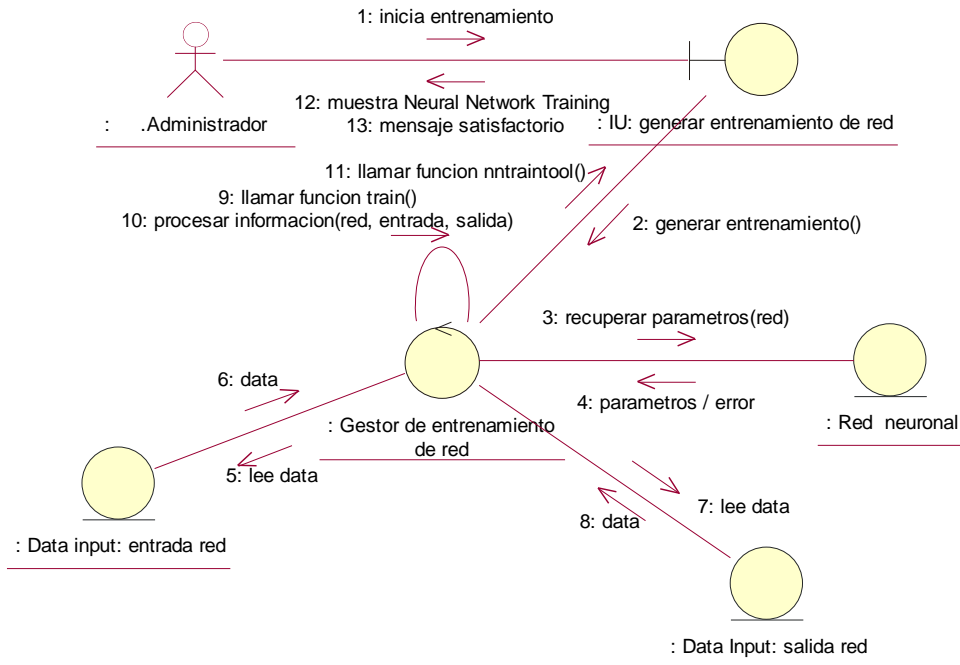


Figura 4.85. Diagrama de Colaboración CU: Generar Entrenamiento

Diagrama de Colaboración del CU: Simular y validar red con data entrenada

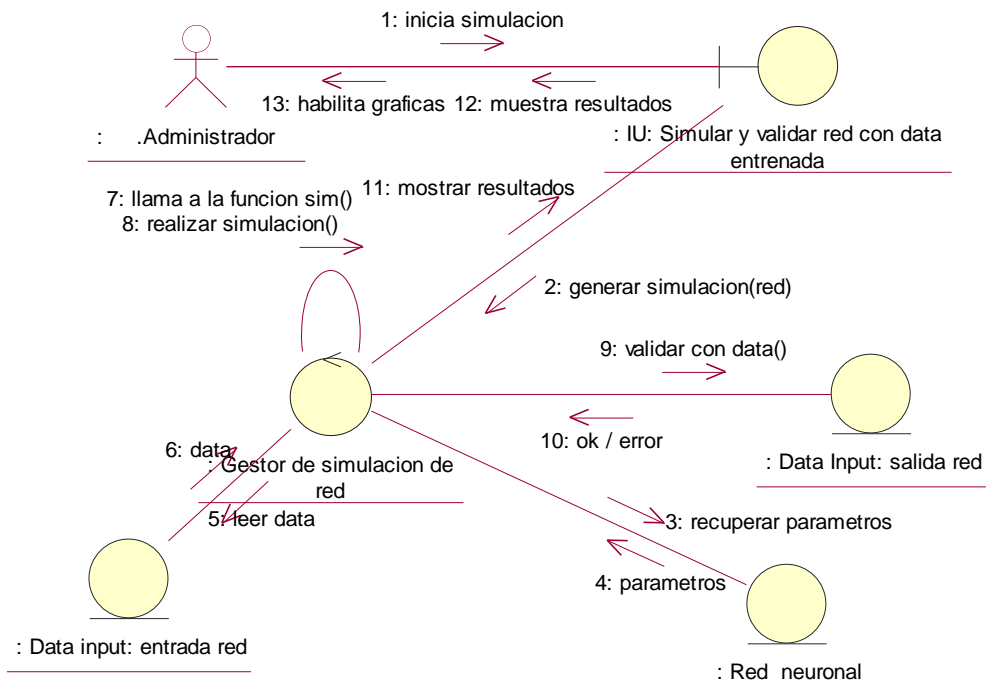


Figura 4.86. Diagrama de Colaboración CU: Simular y validar red con data entrenada

Diagrama de Colaboración del CU: Importar data de entrada para predicción.

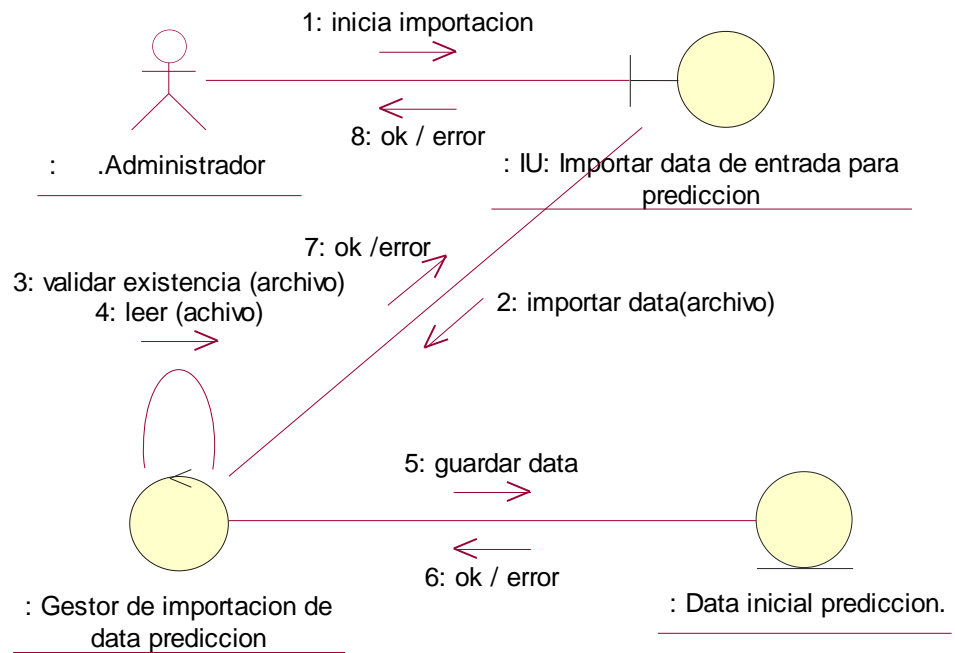


Figura 4.87. Diagrama de Colaboración CU: Importar data de entrada para predicción

Diagrama de Colaboración del CU: Simular y predecir ICA

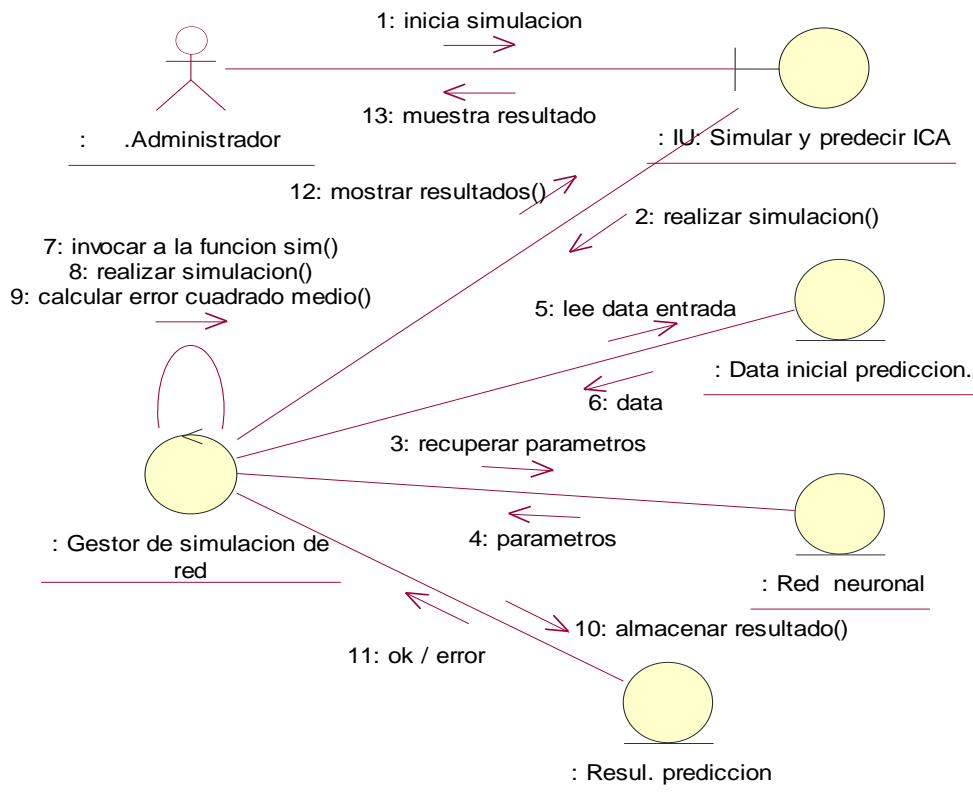


Figura 4.88. Diagrama de Colaboración CU: Simular y predecir ICA

Diagrama de Colaboración del CU: Realizar aproximación

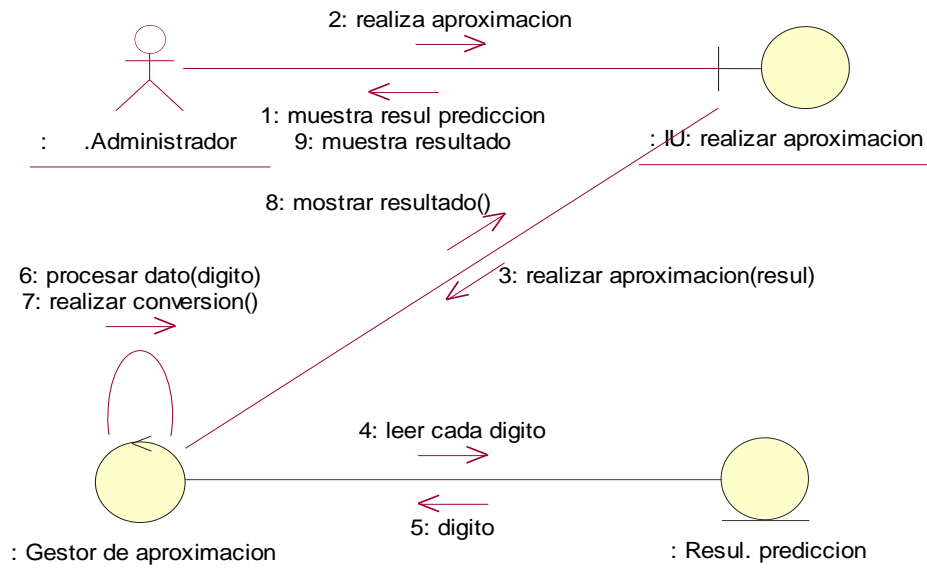


Figura 4.89. Diagrama de Colaboración CU: Realizar aproximación

Diagrama de Colaboración del CU: Procesar resultados matlab

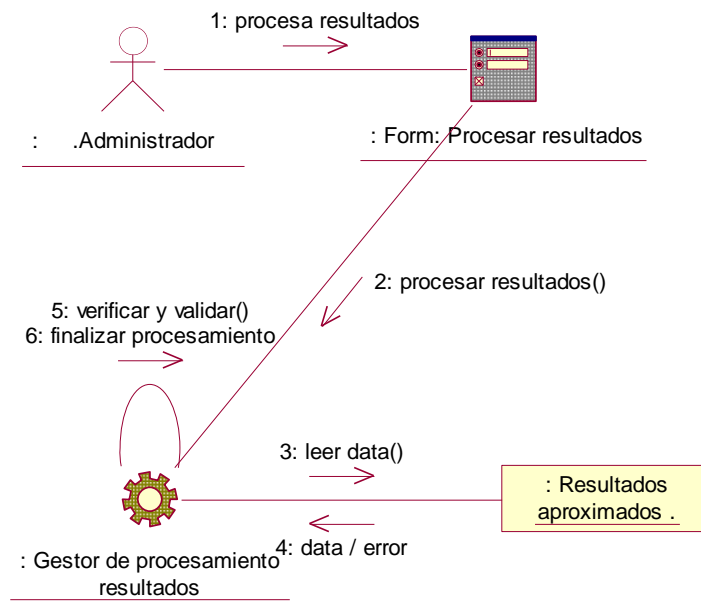


Figura 4.90. Diagrama de Colaboración CU: Procesar resultados matlab

Diagrama de Colaboración del CU: Visualizar monitoreo de predicción de ICA

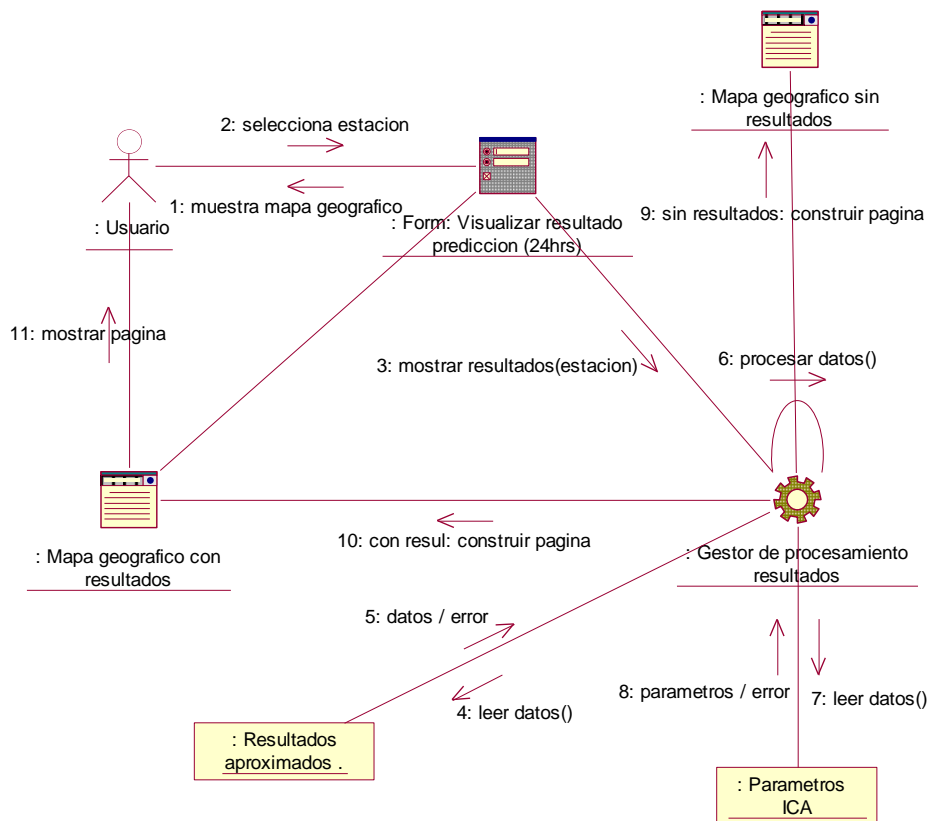


Figura 4.91. Diagrama de Colaboración CU: Visualizar monitoreo de predicción de ICA

d. Diagrama de Clases de Diseño.-

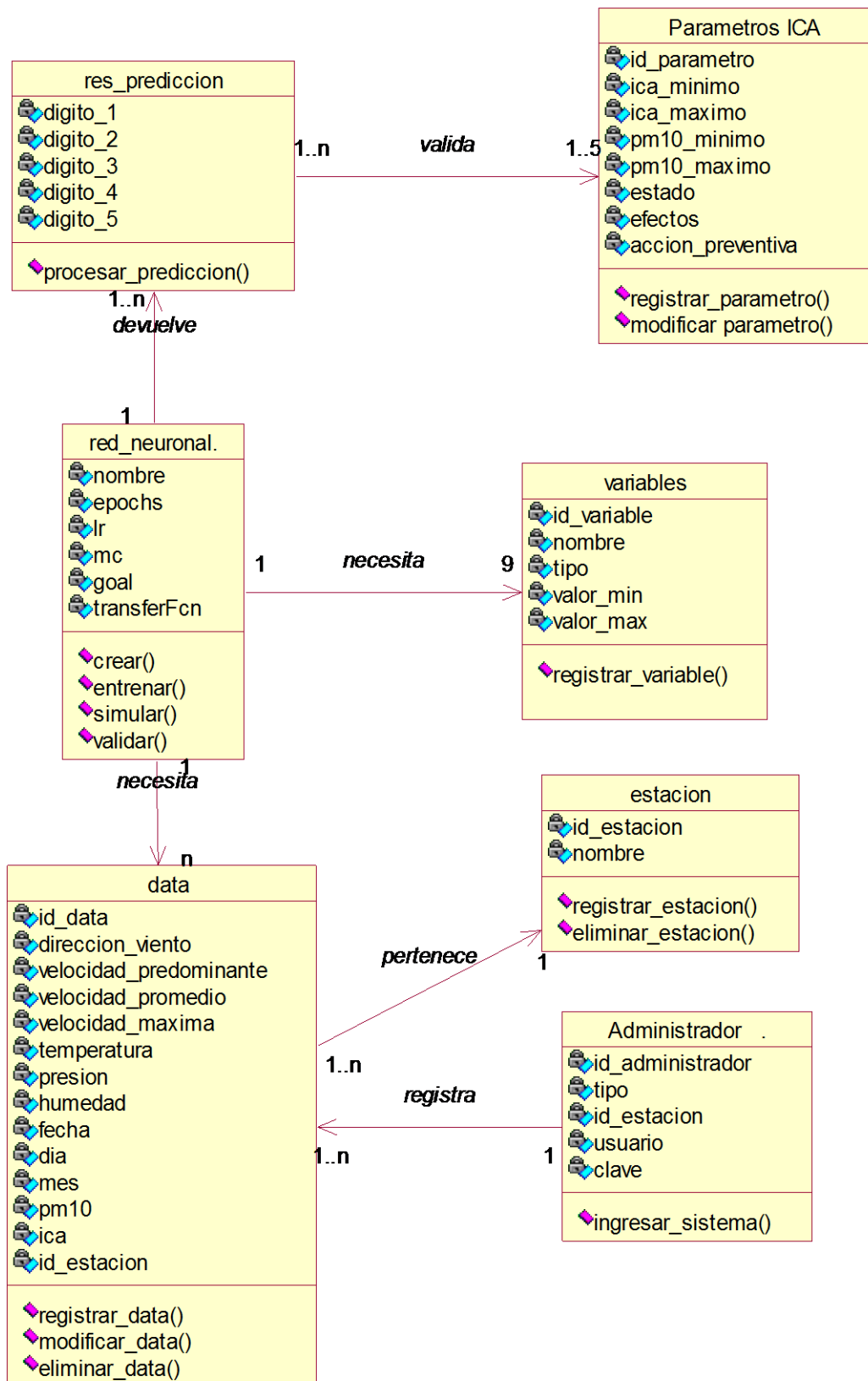


Figura 4.92. Diagrama de Clases de Diseño

4.2.2.4. MODELO DE IMPLEMENTACION

4.2.2.4.1. Diagrama de Componentes

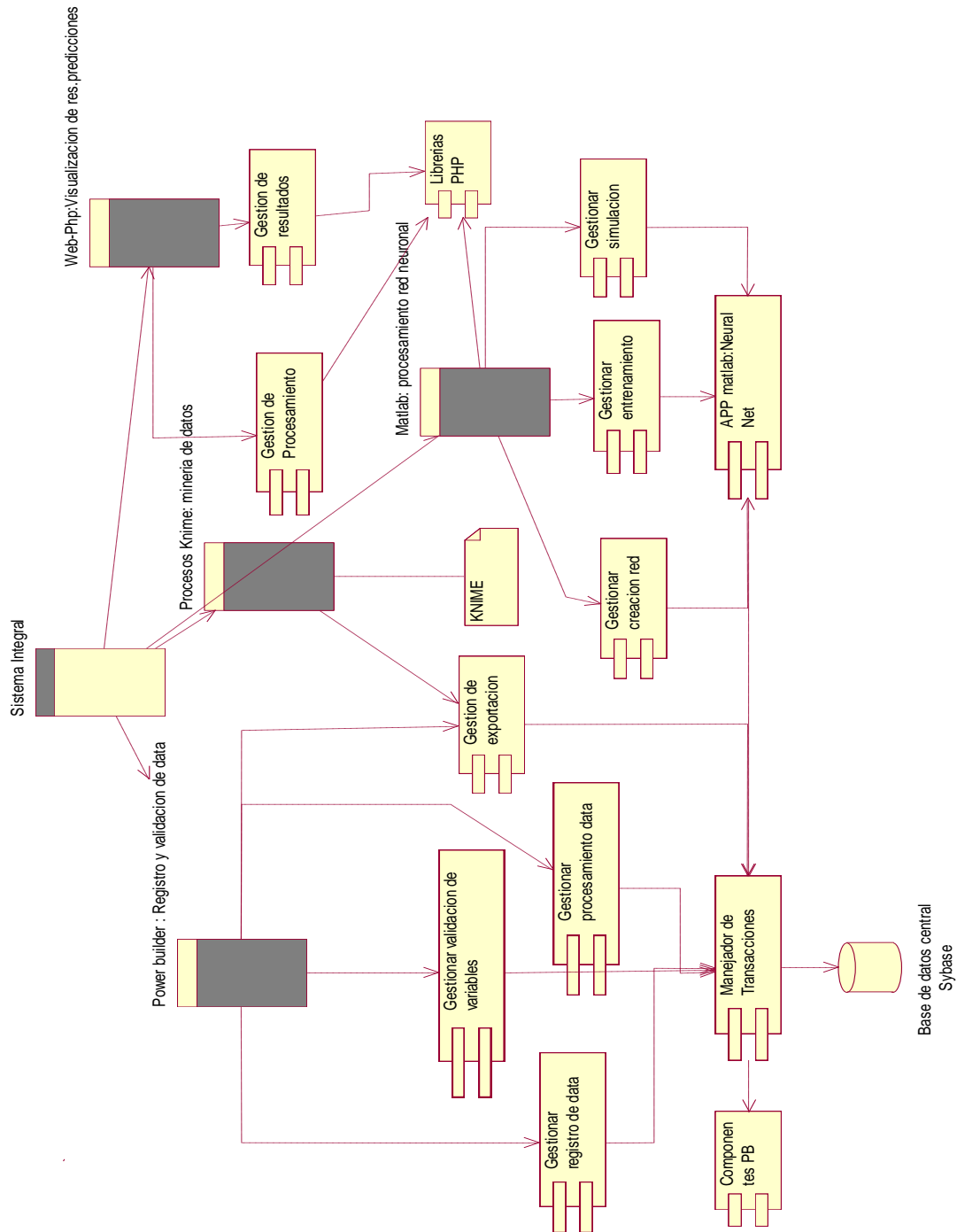


Figura 4.93. Diagrama de Componentes

4.2.2.4.2. Diagrama de Despliegue

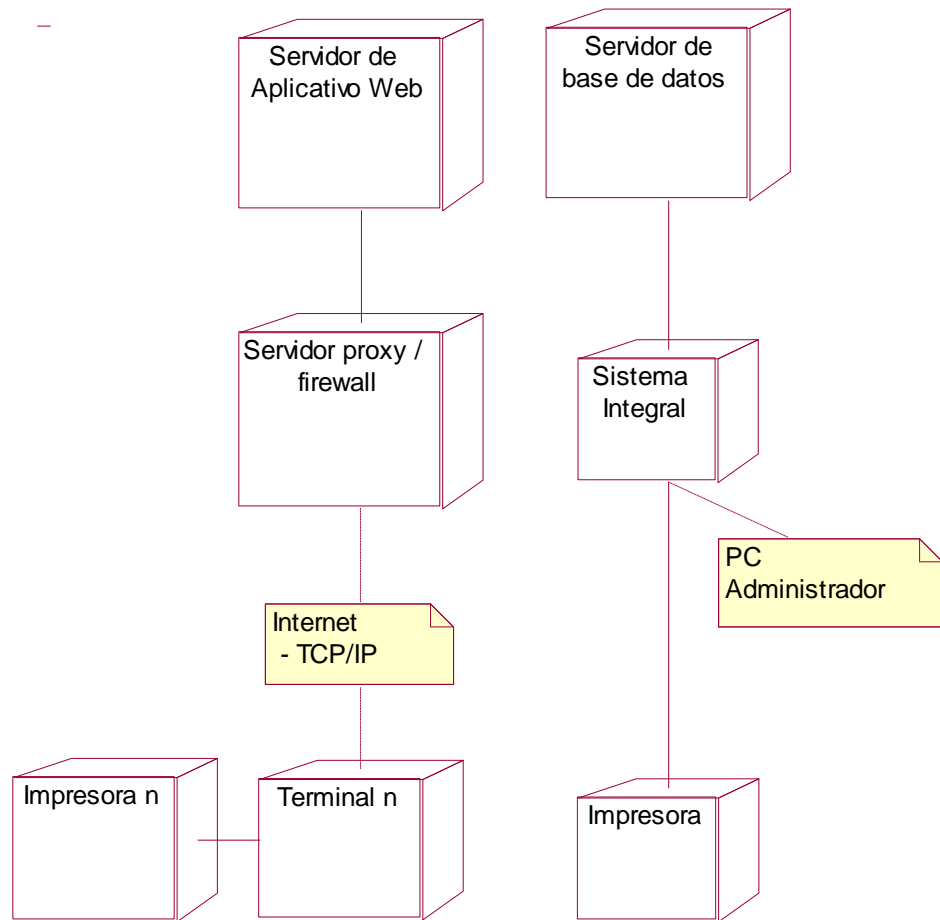


Figura 4.94. Diagrama de despliegue

4.2.2.5. LISTA DE RIESGOS:

Nuevos Riesgos

- No Diseñar una Arquitectura Estable

Descripción: El desarrollador no logre capturar los casos de uso más importantes y por ende no se pueda satisfacer las necesidades de los usuarios.

Prioridad: Critico

Impacto: El impacto se da dentro de la Fase de Elaboración y Construcción.

Responsabilidad: Desarrollador del Sistema

Contingencia: Realizar la mayor cantidad de abstracciones y desde distintos ángulos de los casos de uso y diseñar pensando en que no soy yo (el desarrollador) para quien se implementará el sistema

- No Obtener Interfaces Amigables o de Fácil Uso

Descripción: Diseñar e implementar el sistema usando interfaces que el usuario no esté contento con estas o sienta que no cumple con las funciones requeridas.

Prioridad: Significativo

Impacto: El impacto se da dentro de la Fase de Elaboración y Construcción.

Responsabilidad: Desarrollador del Sistema

Contingencia: Se realizará Presentaciones prototipos a los usuarios finales del sistema integral para que estos

se vayan familiarizado y sugiriendo algunos cambios o mejoras.

- No Obtener un tiempo de respuesta esperado del Sistema

Descripción: El Desarrollador no logre implementar los procesos de consulta en líneas de código de manera eficiente.

Prioridad: Critico

Impacto: El impacto se da dentro de la Fase de Construcción y Transición.

Responsabilidad: Desarrollador del Sistema

Contingencia: Verificar los requerimientos mínimos en hardware para el eficiente funcionamiento del sistema y testear constantemente los procesos más largos.

4.2.3. FASE DE CONSTRUCCION

4.2.3.1. INTERFACES DEL SISTEMA

A. SISTEMA DE REGISTRO Y VALIDACION DE DATA (POWER BUILDER)

1. Acceso al sistema



SISTEMA INTEGRAL PARA UN MODELO DE PREDICCIÓN Y CONTROL DE LA CALIDAD DEL AIRE BAJO EL ENFOQUE DE MINERIA DE DATOS Y REDES NEURONALES

Estación : Administrador

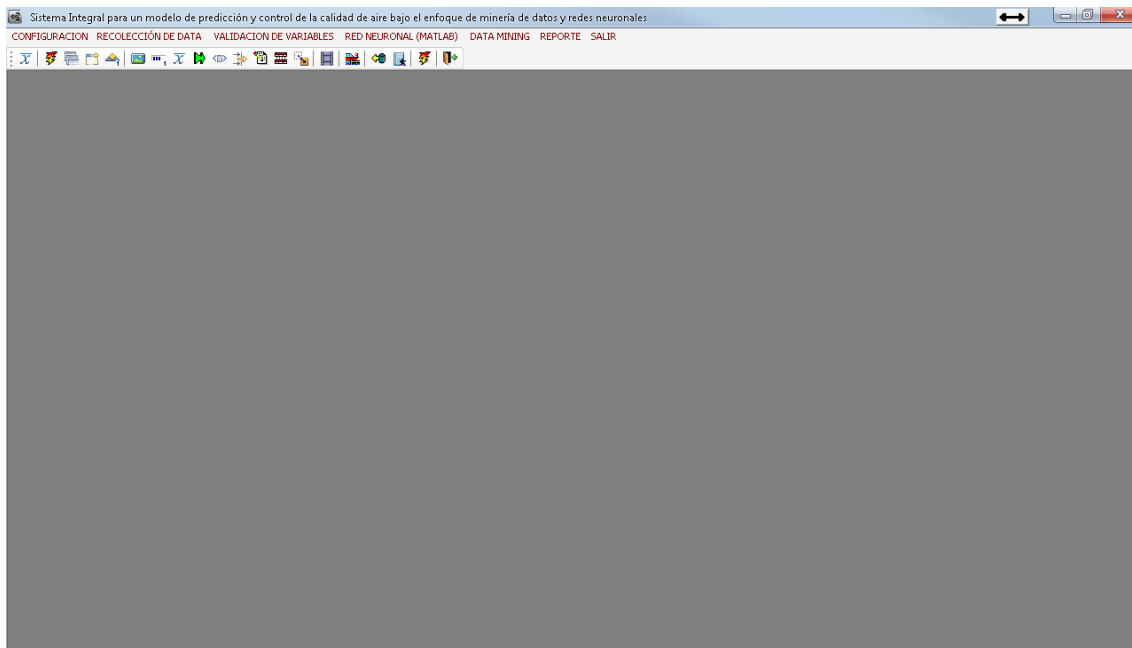
Usuario : SIXTO

Password : ●●●●●●

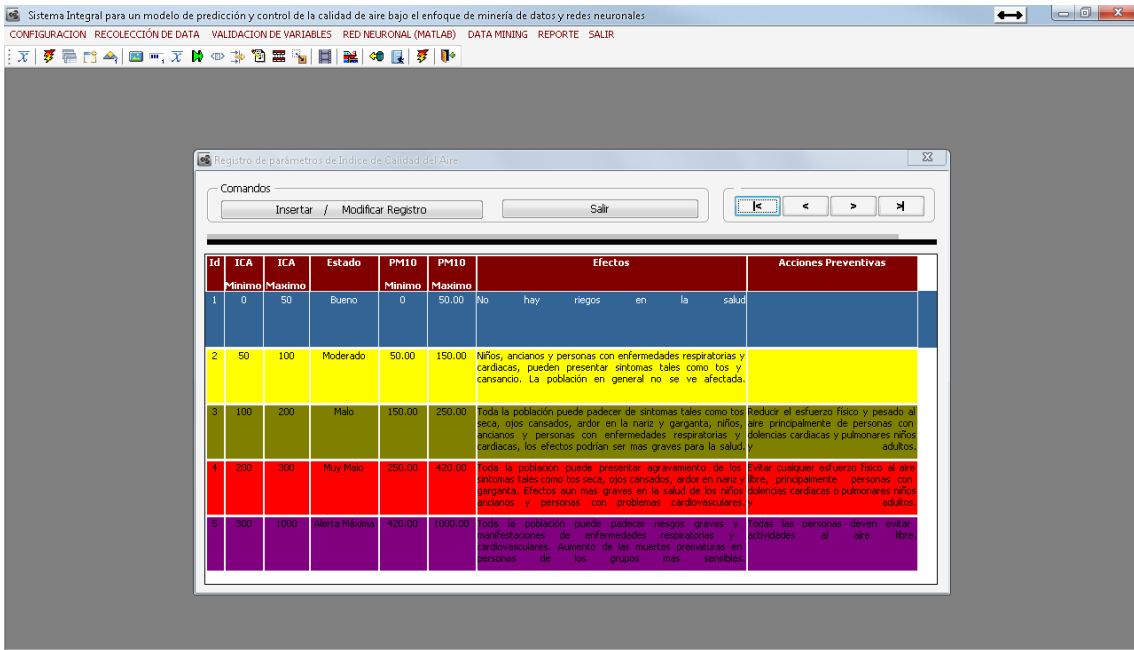
Ingresar Salir

Autor Tesis: Dr. Sixto Diaz Tello

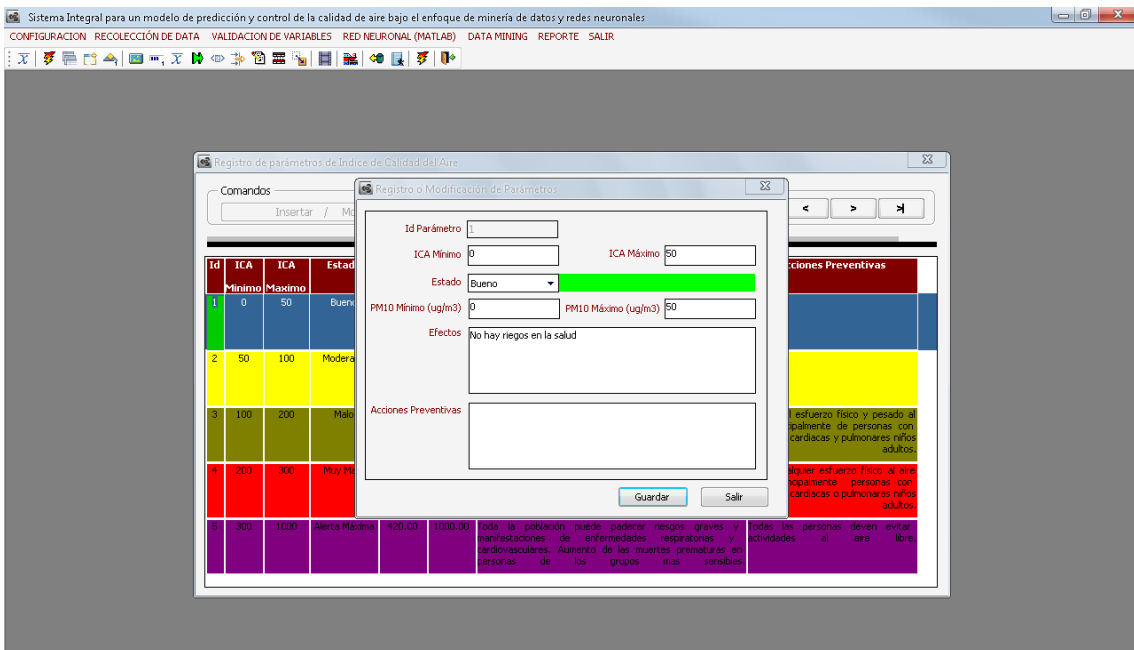
2. Pantalla principal del sistema



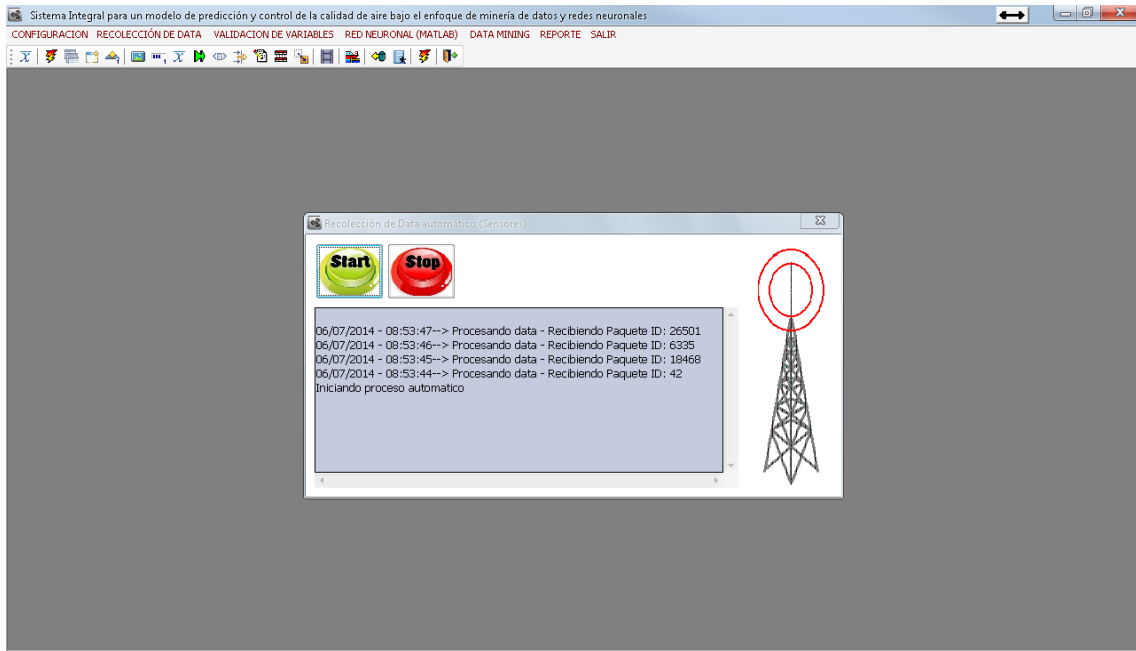
3. Registro de Parámetros de ICA.- Sirve para registrar



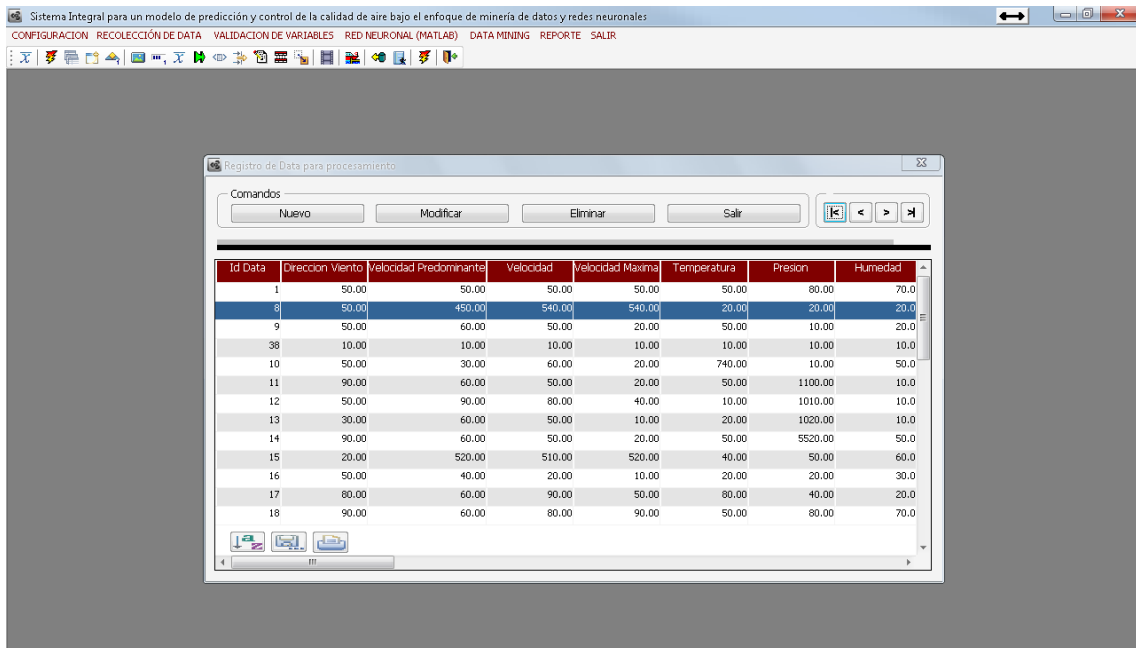
4. Registro o modificación de los datos del nuevo parámetro



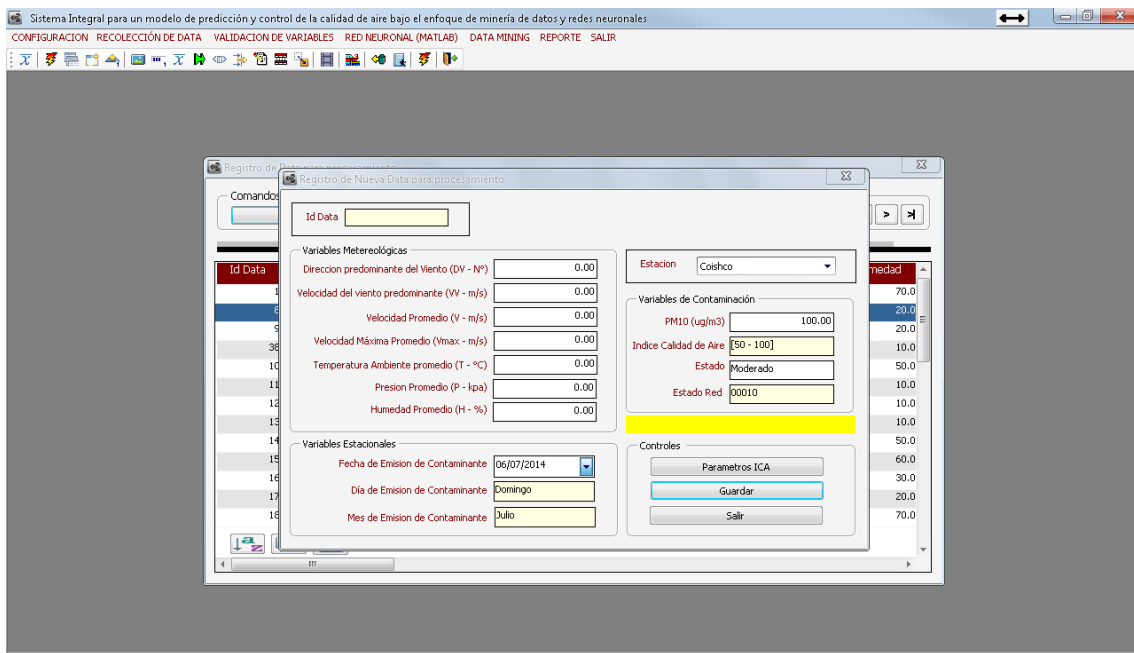
5. Recolección de data automática.- Sirve para realizar la conexión a través de una estación de monitoreo en tiempo real.



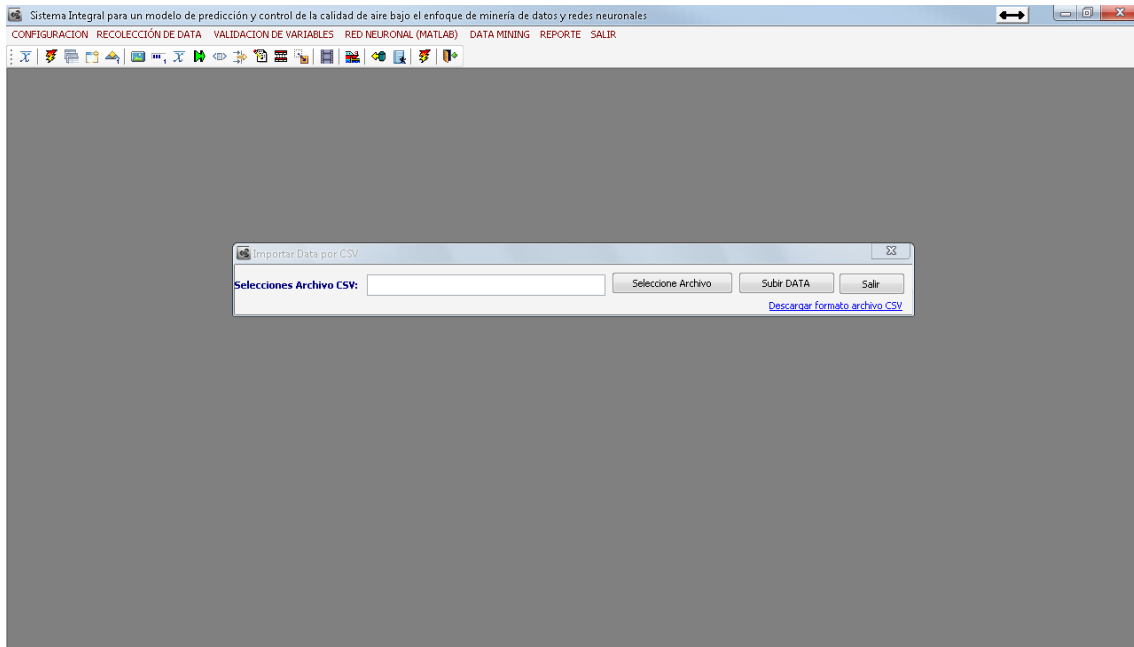
6. Registro de data meteorológica diaria



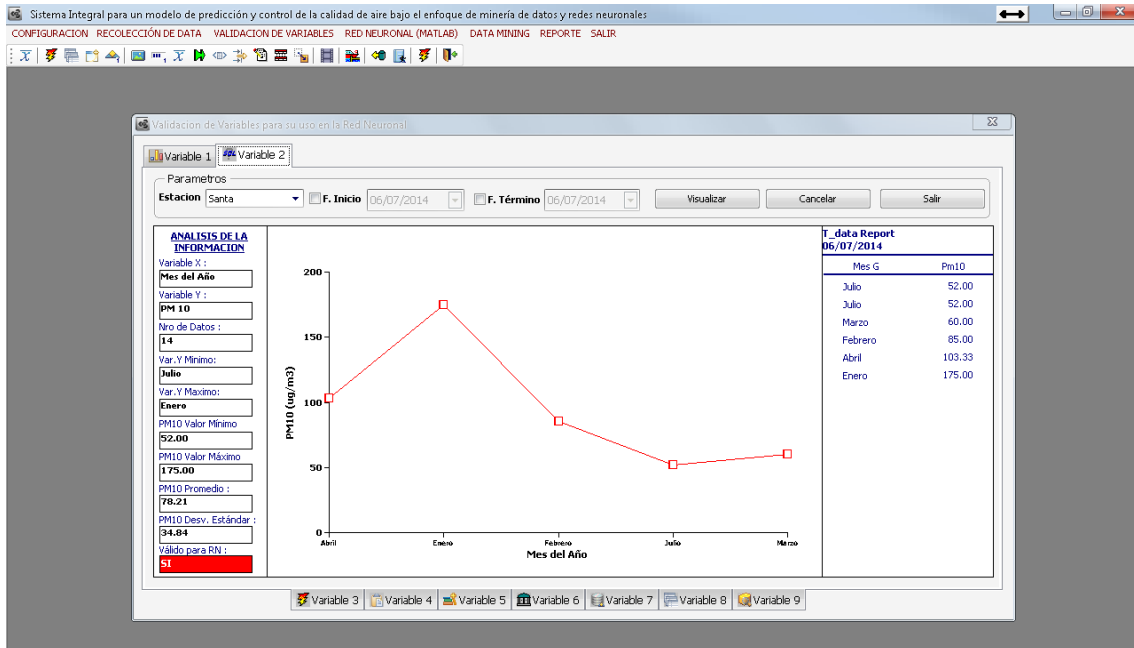
7. Registro de los datos de la información meteorológica diaria



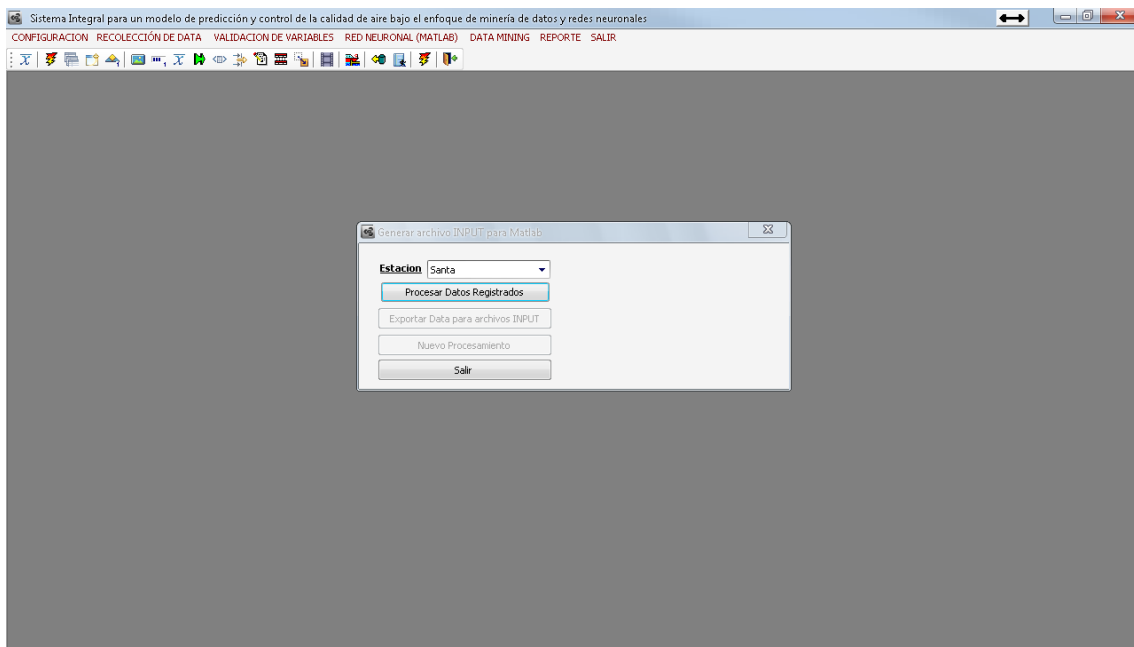
8. Importación de data meteorológica por CSV: Sirve para importar data meteorológica por bloques a través de un archivo.



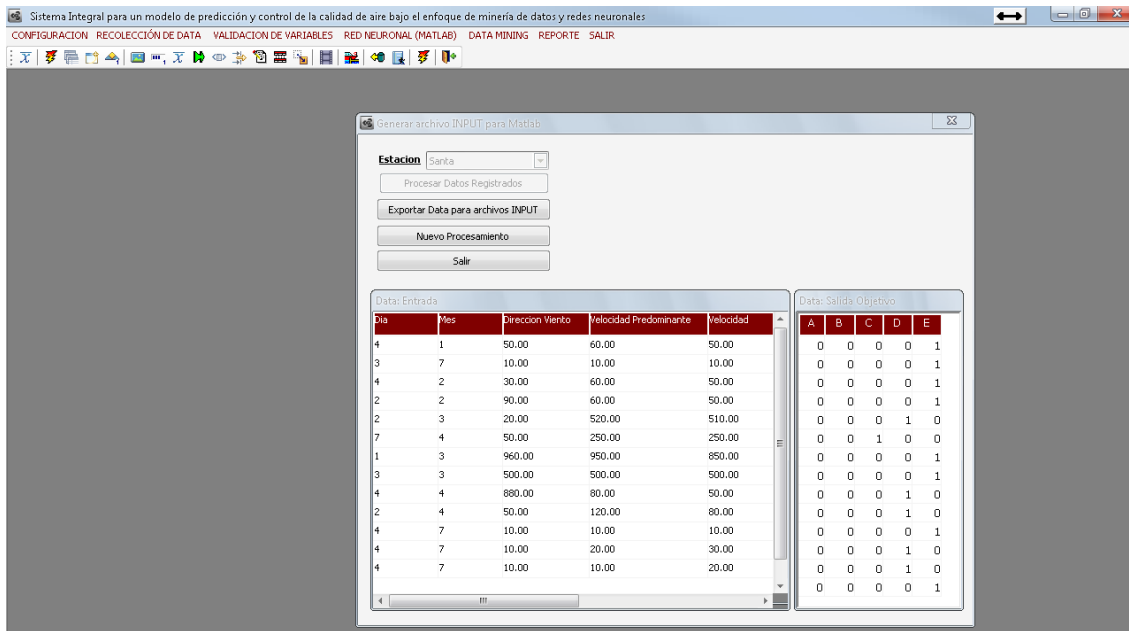
9. Validación de variables meteorológicas y estacionales: Sirve para validar las 9 variables (2 estacionales y 7 meteorológicas), cada una de estas validadas respecto al PM10.



10. Generar archivos input para Matlab: sirve para generar los archivo input que servirán para el entrenamiento de la red neuronal.



11. Resultados del procesamiento de data



Generar archivo INPUT para Matlab

Estacion: Santa

Procesar Datos Registrados

Exportar Data para archivos INPUT

Nuevo Procesamiento

Salir

Data: Entrada

Dia	Mes	Direccion Viento	Velocidad Predominante	Velocidad
4	1	50.00	60.00	50.00
3	7	10.00	10.00	10.00
4	2	30.00	60.00	50.00
2	2	90.00	60.00	50.00
2	3	20.00	520.00	510.00
7	4	50.00	250.00	250.00
1	3	960.00	950.00	850.00
3	3	500.00	500.00	500.00
4	4	880.00	80.00	50.00
2	4	50.00	120.00	80.00
4	7	10.00	10.00	10.00
4	7	10.00	20.00	30.00
4	7	10.00	10.00	20.00

Data: Salida Objetivo

A	B	C	D	E
0	0	0	0	1
0	0	0	0	1
0	0	0	0	1
0	0	0	0	1
0	0	0	1	0
0	0	1	0	0
0	0	0	0	1
0	0	0	0	1
0	0	0	1	0
0	0	0	1	0
0	0	0	1	0
0	0	0	0	1

12. Mensaje satisfactorio de exportación de archivos input.



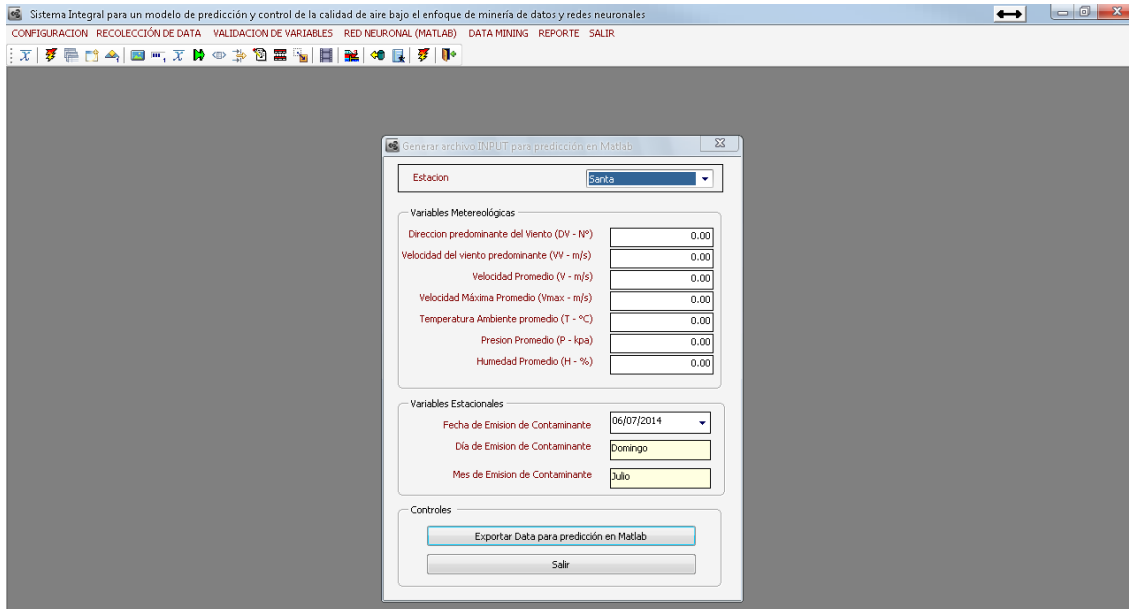
Sistema

Archivo INPUTS para MATLAB exportado con exito.

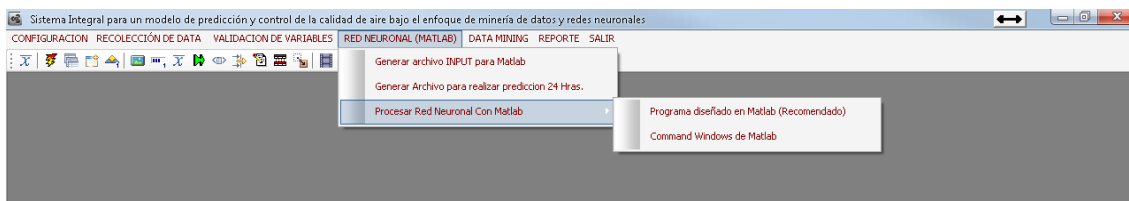
Archivo INPUT:Entrada - C:/ambiental/exportaciones/matlab/santa/entrada.xls
Archivo INPUT:Salida - C:/ambiental/exportaciones/matlab/santa/salida.xls

Aceptar

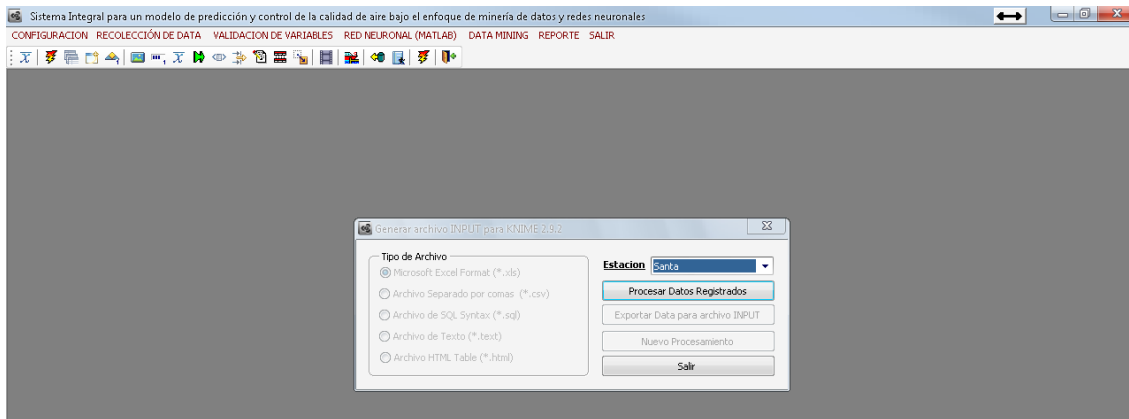
13. Generar archivo input para predicción en Matlab: generar el archivo con la data que servirá para predecir el ICA a 24 horas, el cual será importado luego por el sistema de Matlab



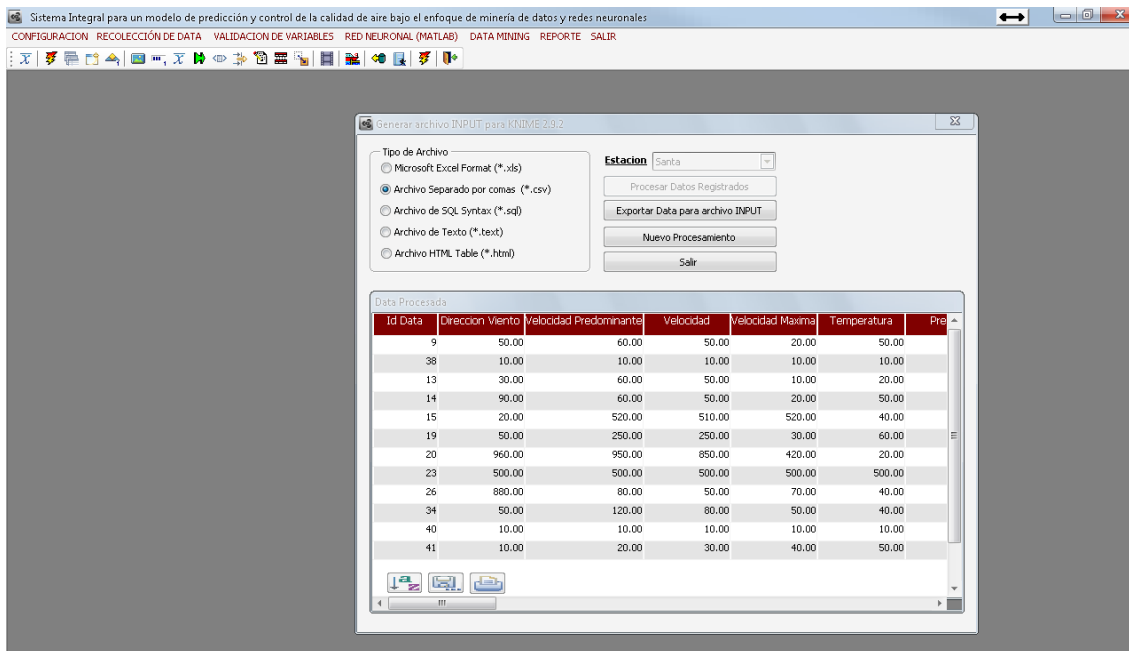
14. Abrir el sistema de procesamiento de red neuronal en Matlab: sirve para ejecutar el sistema desarrollado en Matlab, también existe la opción “Command Window de Matlab”, el cual sirve cuando se desee trabajar la red neuronal desde la línea de comandos de Matlab.



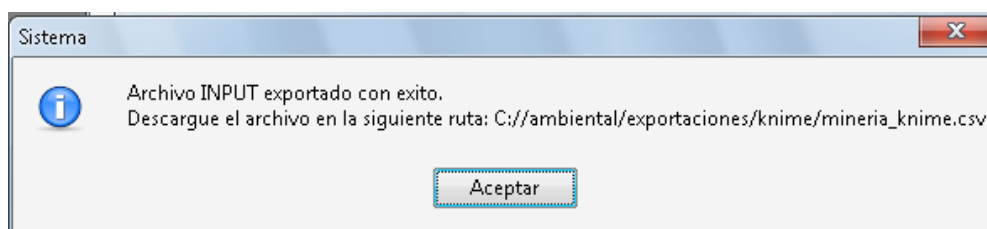
15. Generar archivo input para Knime: Sirve para procesar y exportar la data, de acuerdo a la estación seleccionada, para realizar la minería de datos a través del Software Knime.



16. Resultado del Procesamiento de la data: Una vez procesada se seleccionada el tipo de archivo con el cual se desea trabar en knime (Se recomienda CSV)



17. Mensaje satisfactorio de exportación de archivo.



18. Abrir Software Knime 2.9.2, para realizar minería de datos.



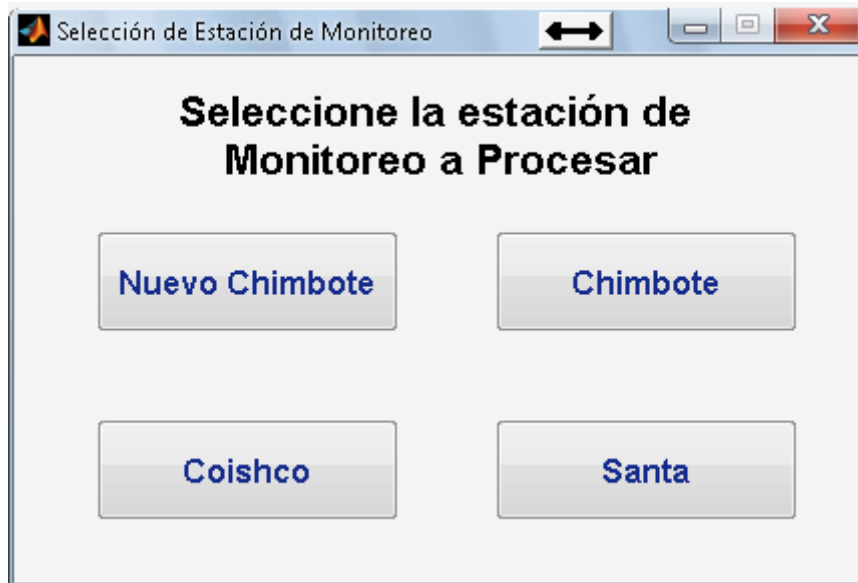
19. Reporte de filtros de data meteorológica registrada, se encuentra diseñado para seleccionar los filtros a través de rangos de acuerdo a cada variable.

The screenshot shows a dialog box titled 'Filtro de Data registrada'. It contains a 'Seleccionar Filtros' section with several dropdown menus for filtering data based on various meteorological variables. Below the filters are three buttons: 'Consultar con filtros establecidos', 'Nueva Consulta', and 'Salir'. At the bottom of the dialog is a table displaying the filtered data.

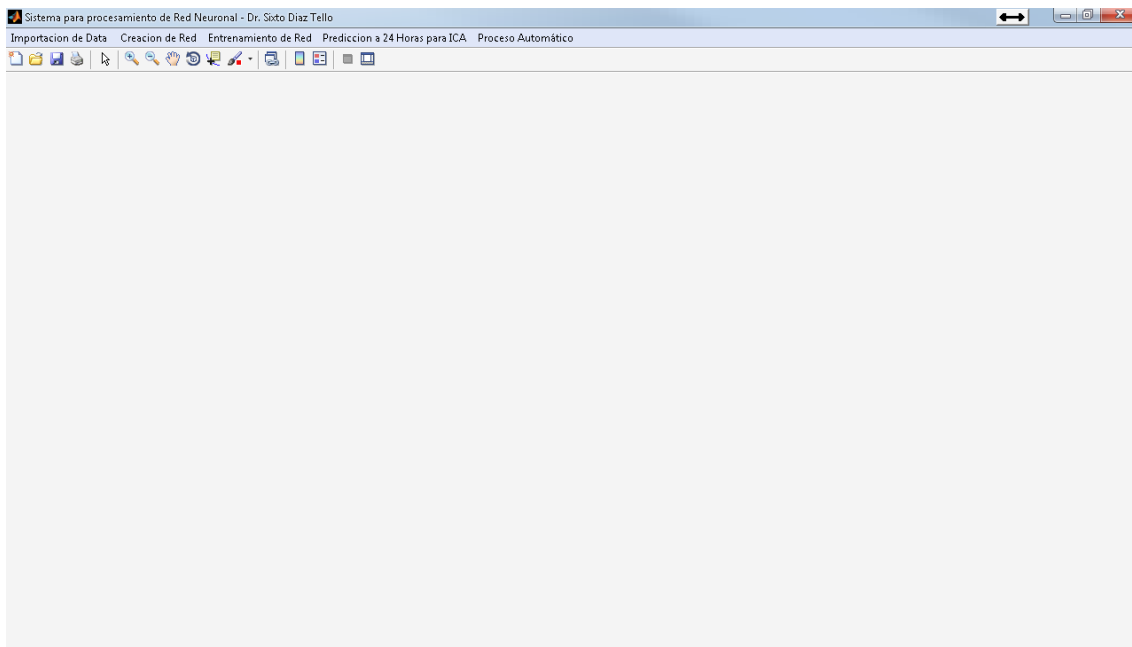
Id Data	Direccion Viento	Velocidad Predominante	Velocidad	Velocidad Maxima	Temperatura	Presion	Humedad	Fecha
9	50.00	60.00	50.00	20.00	50.00	10.00	20.00	01/01/2014
38	10.00	10.00	10.00	10.00	10.00	10.00	10.00	01/07/2014
13	30.00	60.00	50.00	10.00	20.00	1020.00	10.00	26/02/2014
40	10.00	10.00	10.00	10.00	10.00	10.00	10.00	02/07/2014
44	10.00	20.00	30.00	10.00	20.00	30.00	10.00	02/07/2014

B. SISTEMA DE PROCESAMIENTO DE RED NEURONAL (MATLAB)

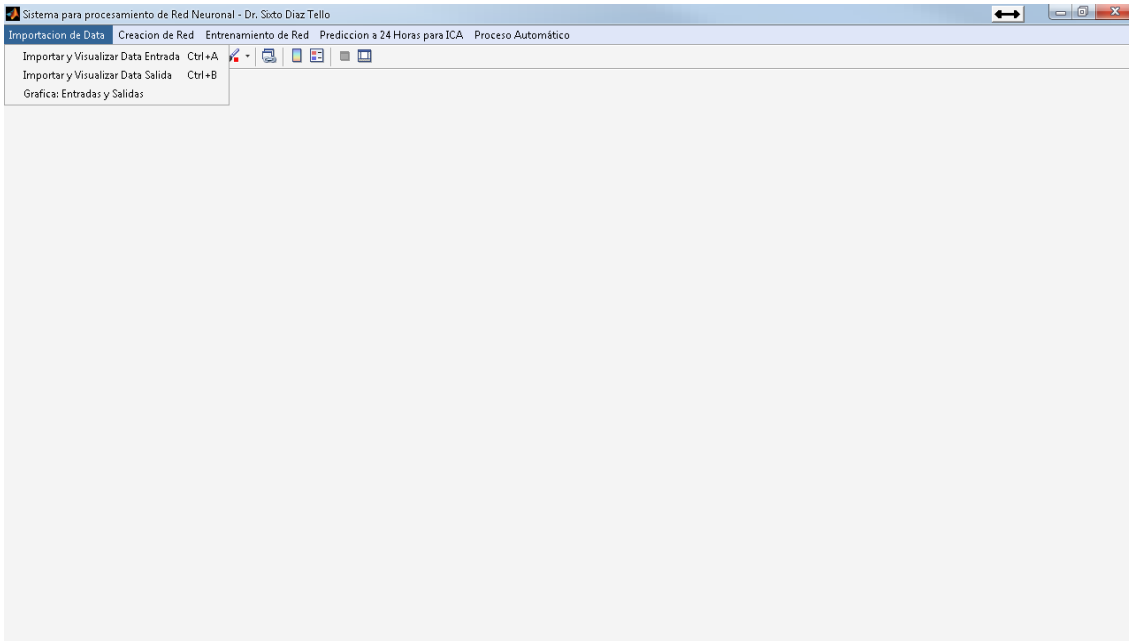
1. Selección de estación de monitoreo.- se selecciona la estación de trabajo, con la cual se procesará toda la red neuronal. Para poder hacer uso del sistema desarrollado en Matlab.



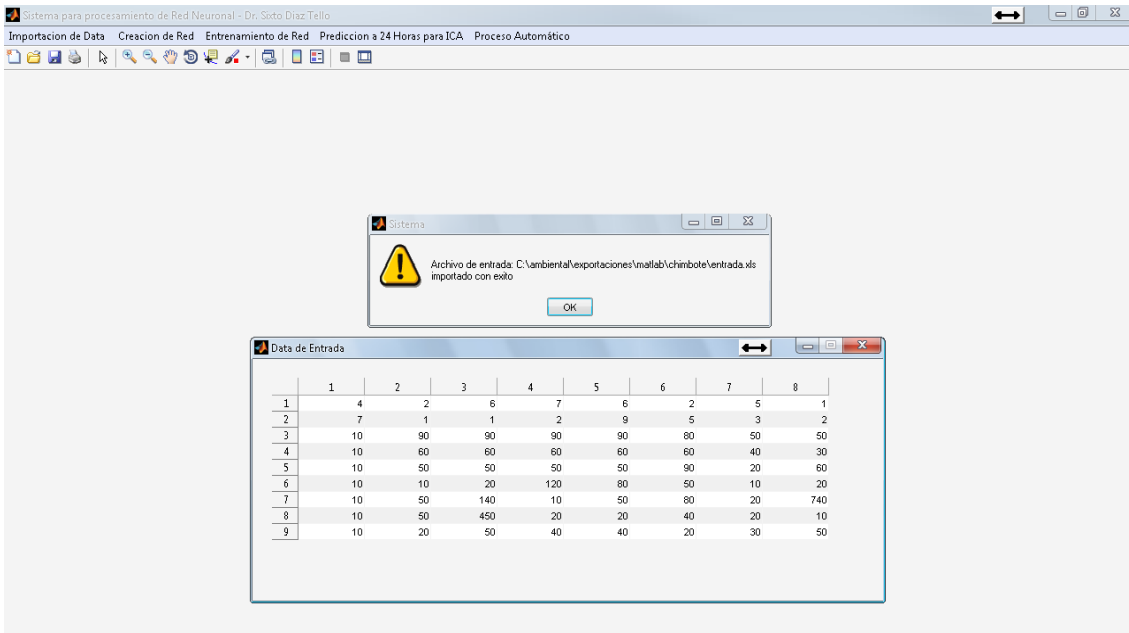
2. Pantalla principal del sistema en Matlab.



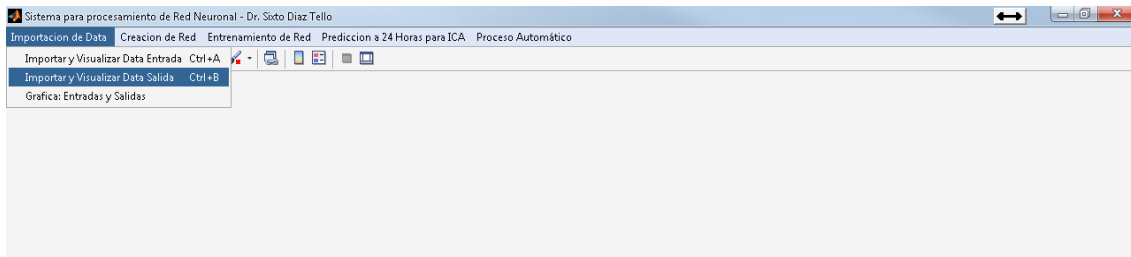
3. Importar y visualizar data de entrada.- importa la data input de entrada para que sea usada en el entrenamiento de la red neuronal, la cual ha sido exportado por el sistema de Power Builder.



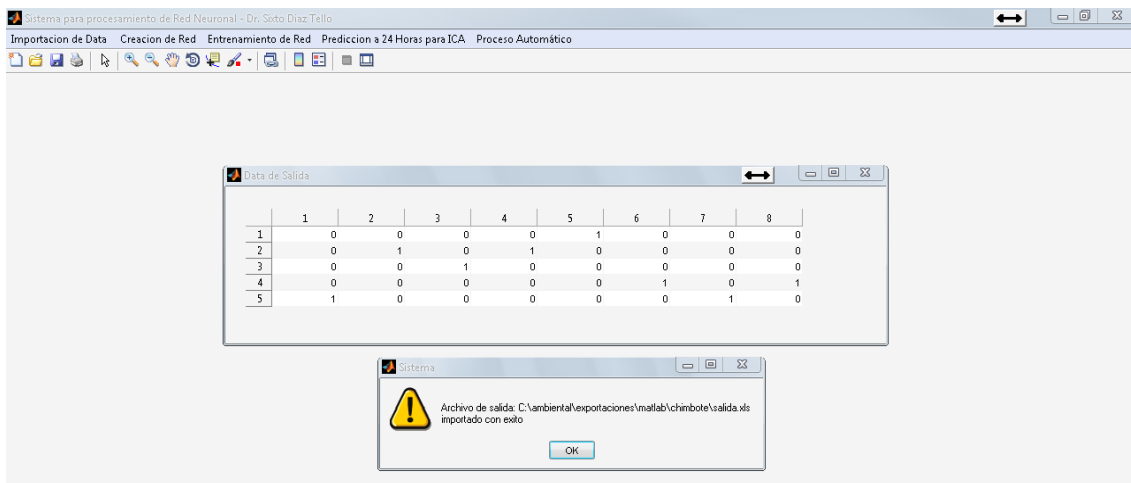
4. Resultado de la importación data entrada.



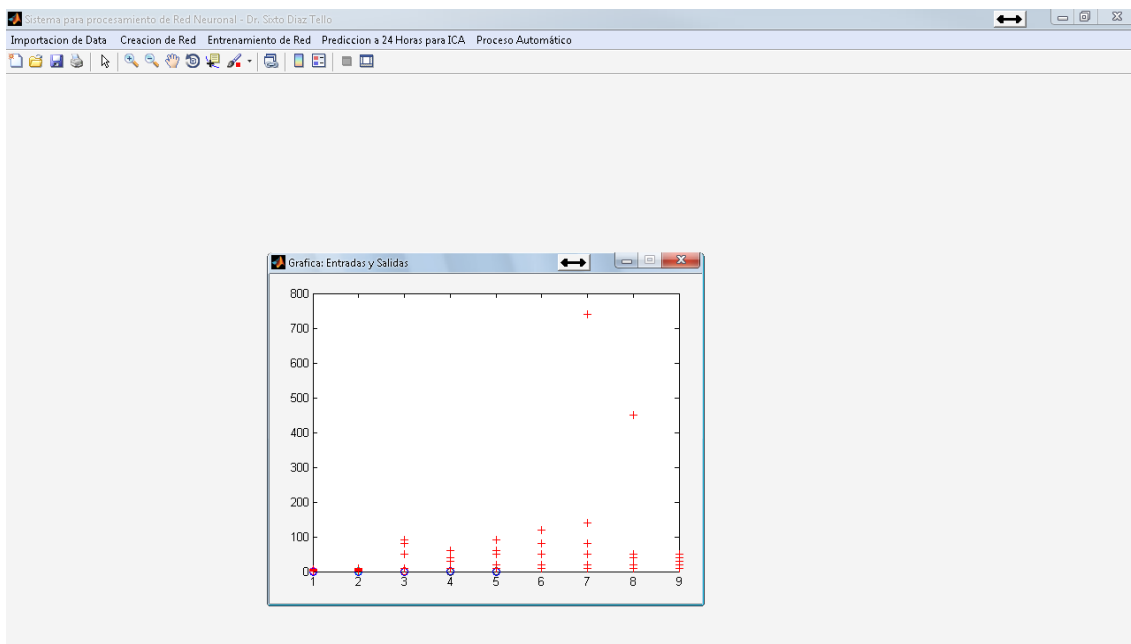
5. Importación y visualización de data salida.- importa la data input de salida para que sea usada en el entrenamiento de la red neuronal, la cual ha sido exportado por el sistema de Power Builder.



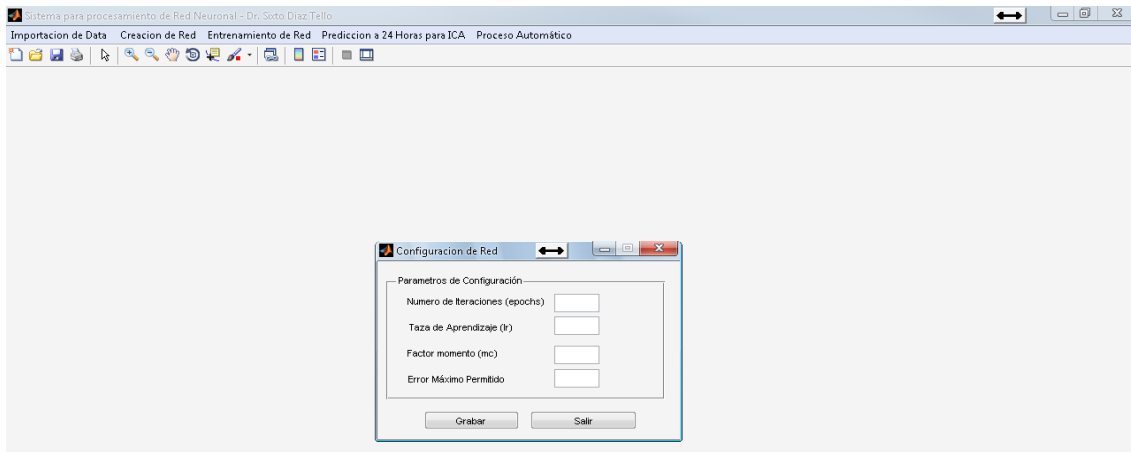
6. Resultado de la importación de data salida



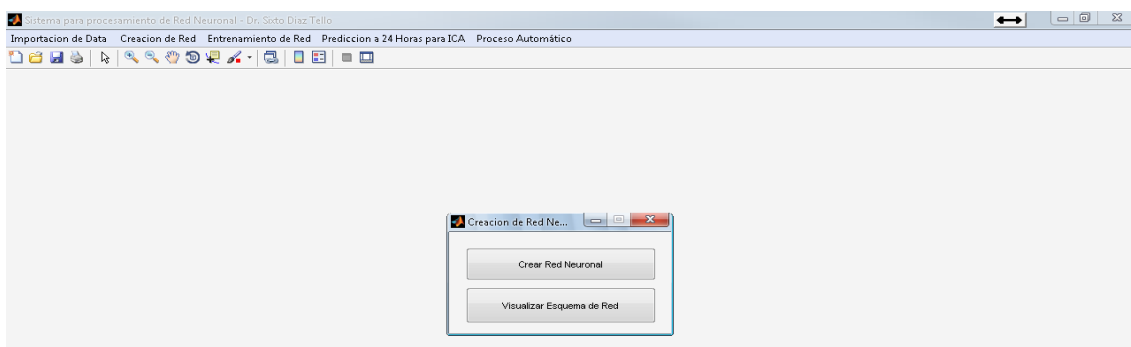
7. Gráfica de entradas y salidas.



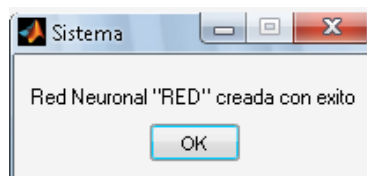
8. Configuración de red: sirve para que en el caso, el administrador desee, configure de la manera que desee la red neuronal, con los parámetros óptimos que se crea conveniente. Cabe mencionar que la red, ya está inicializada con los parámetros óptimos, para su correcto funcionamiento.



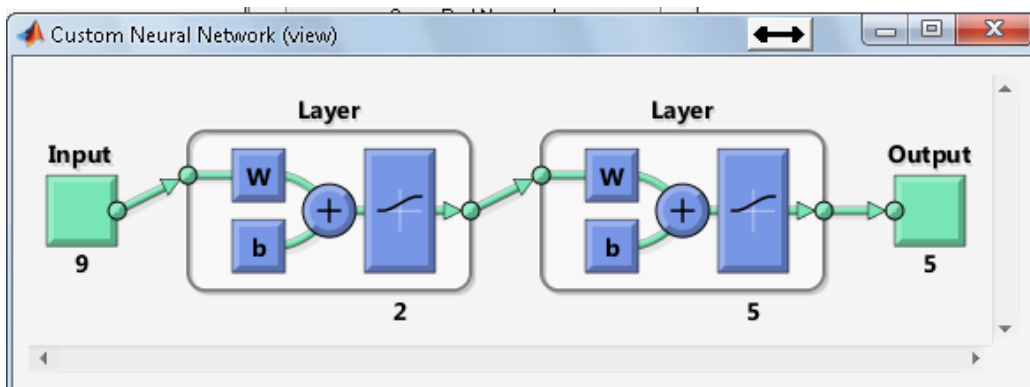
9. Creación de red.



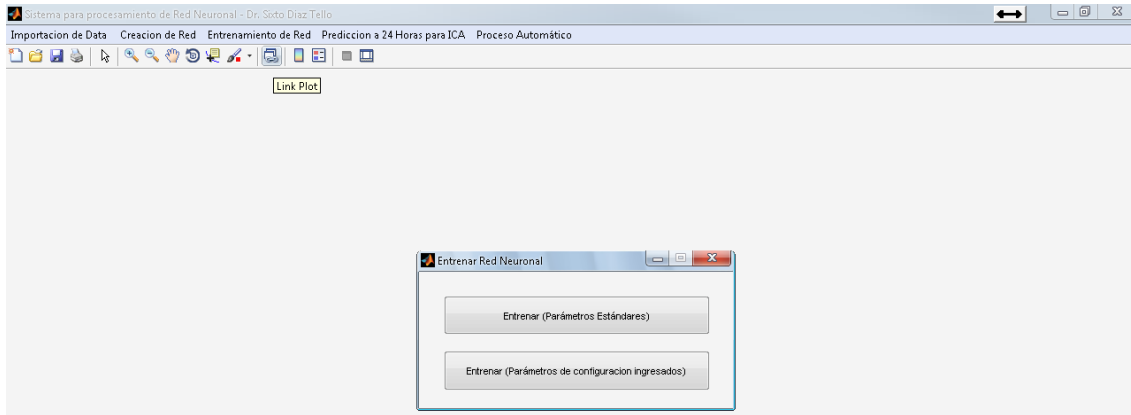
10. Mensaje satisfactorio de creación de red



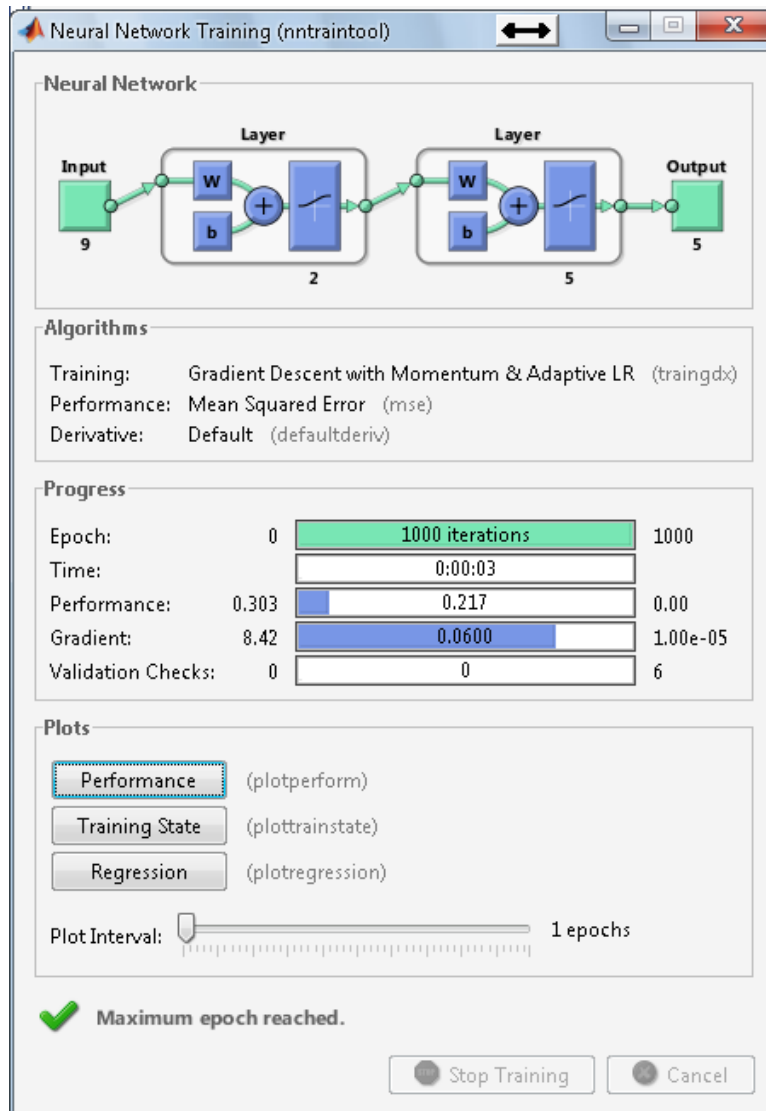
11. Visualización del esquema de red



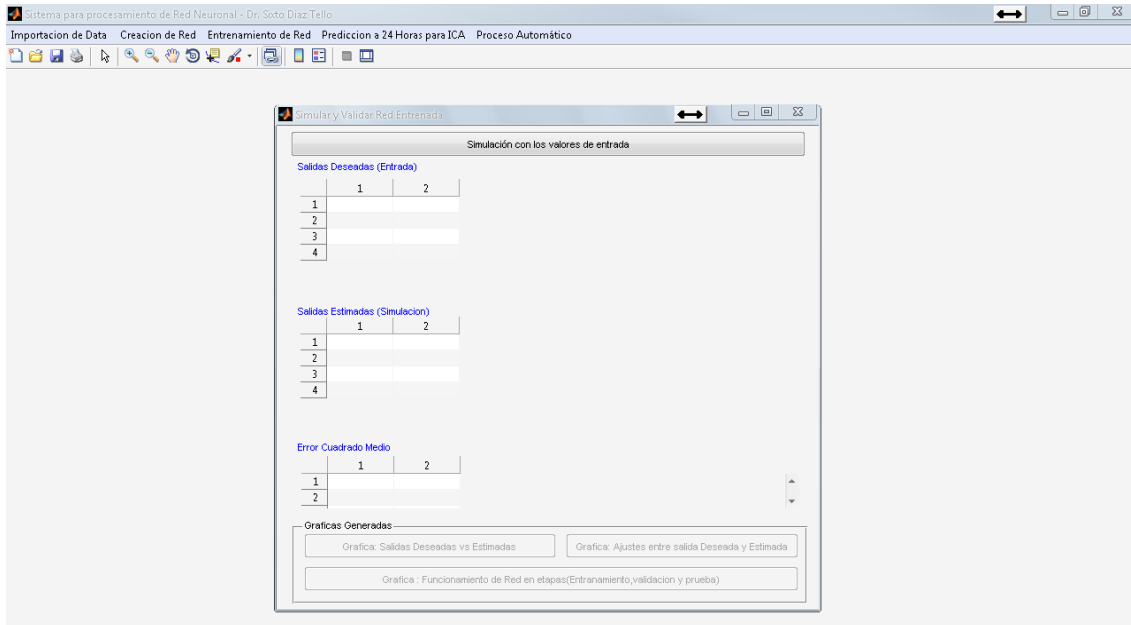
12. Entrenar red.- tiene dos opciones la primera realiza el entrenamiento con los parámetros por defecto que tiene asignado ya el sistema, y la segunda genera el entrenamiento con la configuración que asignó el administrador.



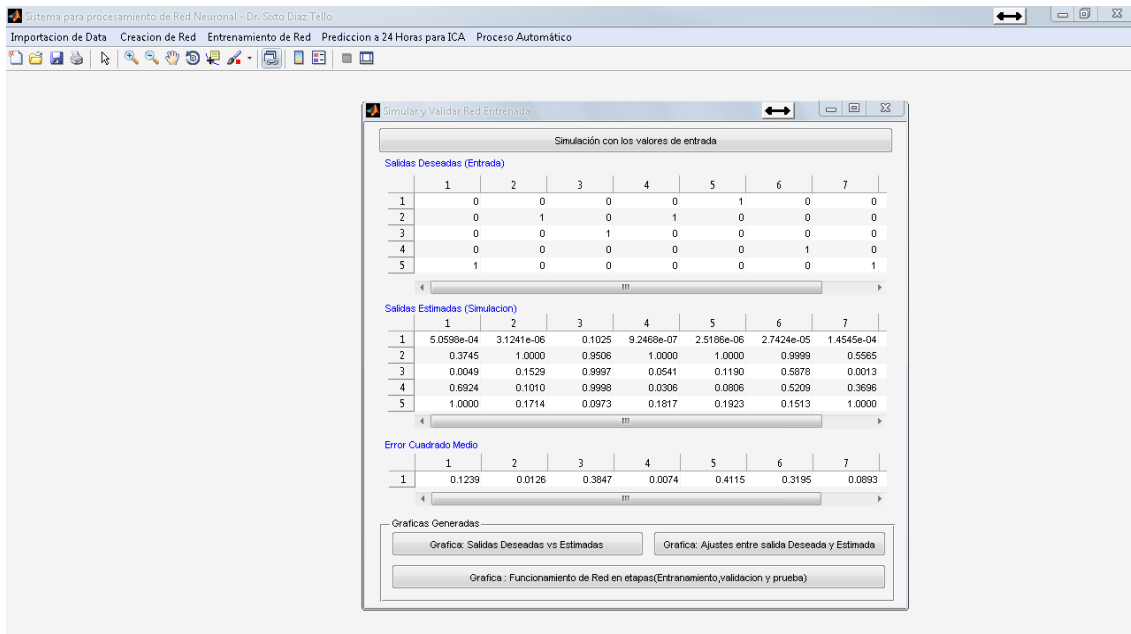
13. Visualización de esquema de red (nntaintool)



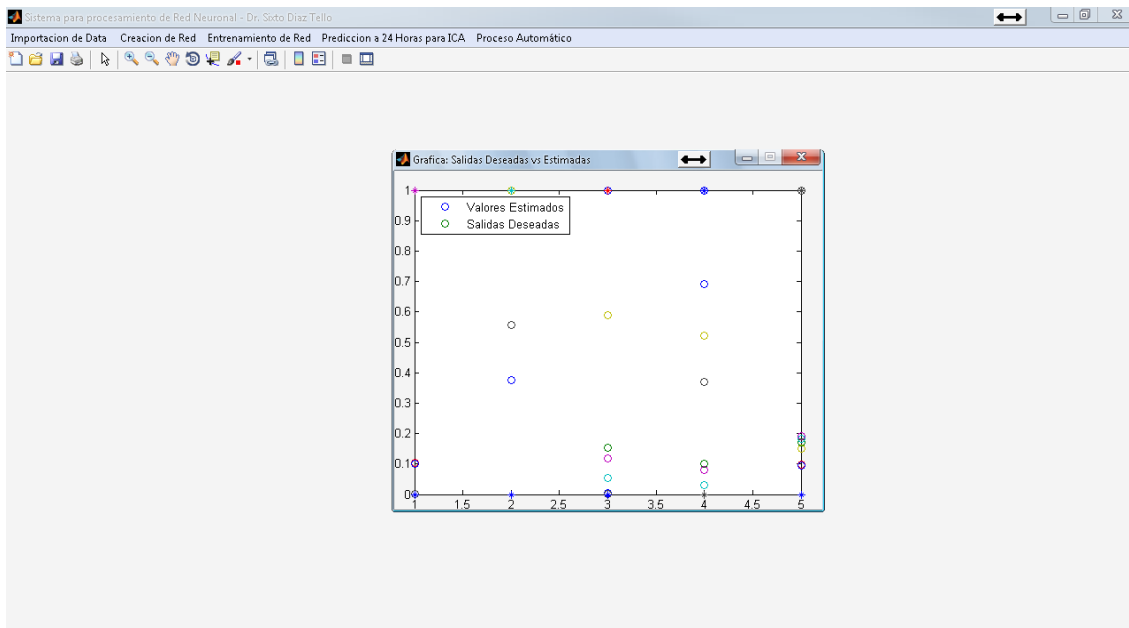
14. Simulación y validación de red con datos de entrada: sirve para realizar una primera simulación de la red que ya ha sido entrenada, pero como dato de entrada se tendrá a los mismo valores input entrada, a fin de poder verificar que las salidas estimadas, sean semejantes a las salidas deseadas.



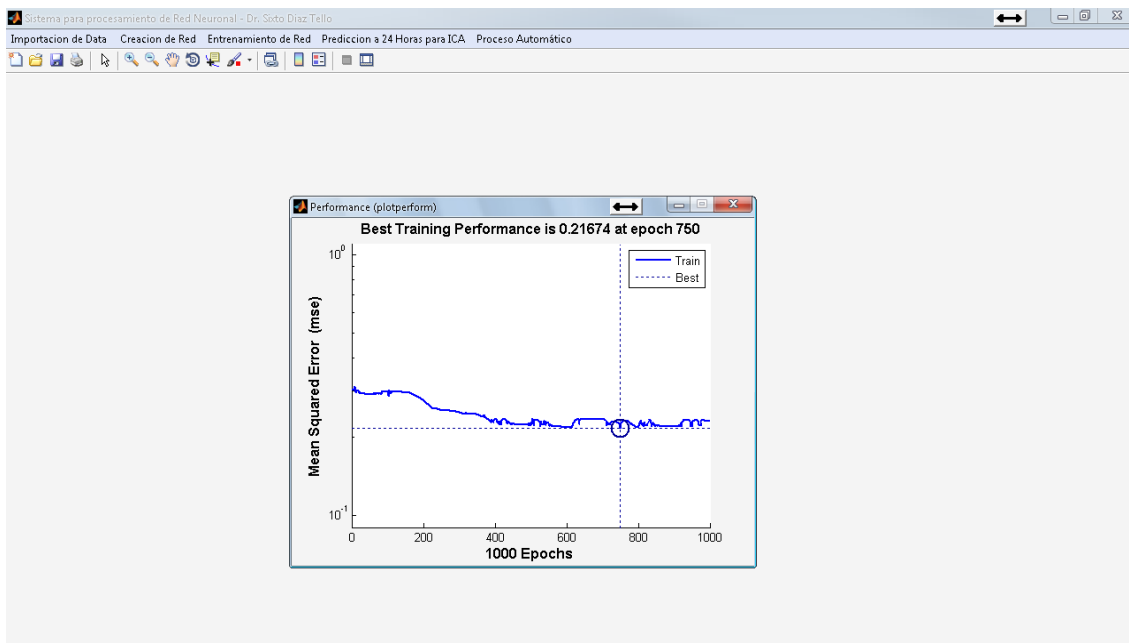
15. Resultado de la simulación y validación



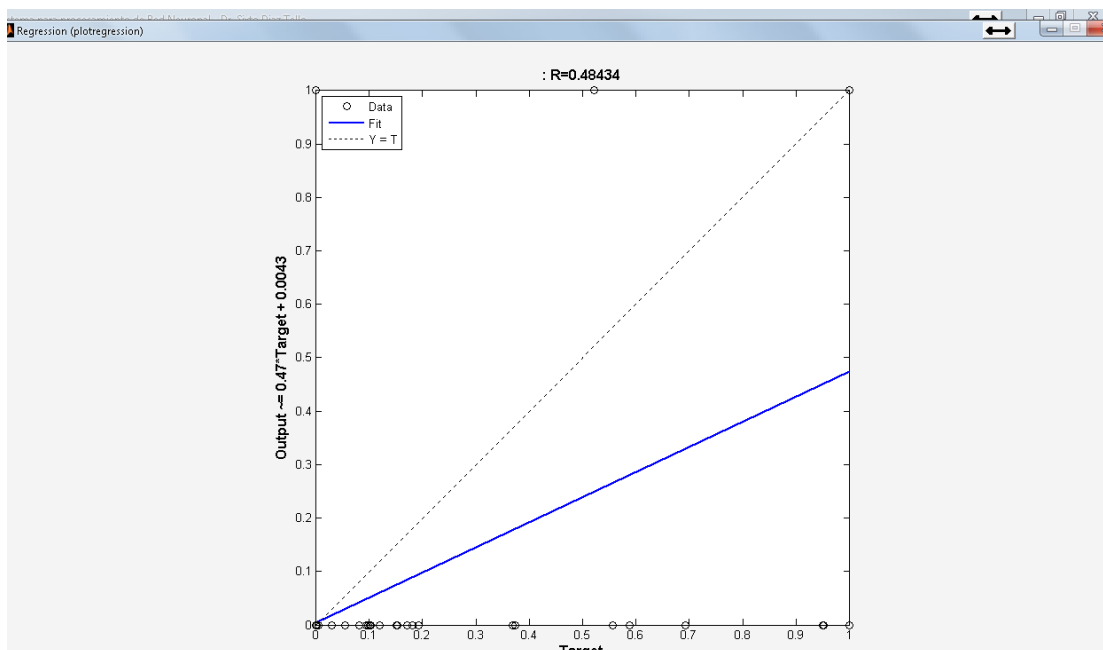
16. Resultado: gráfica Salidas deseadas vs estimadas



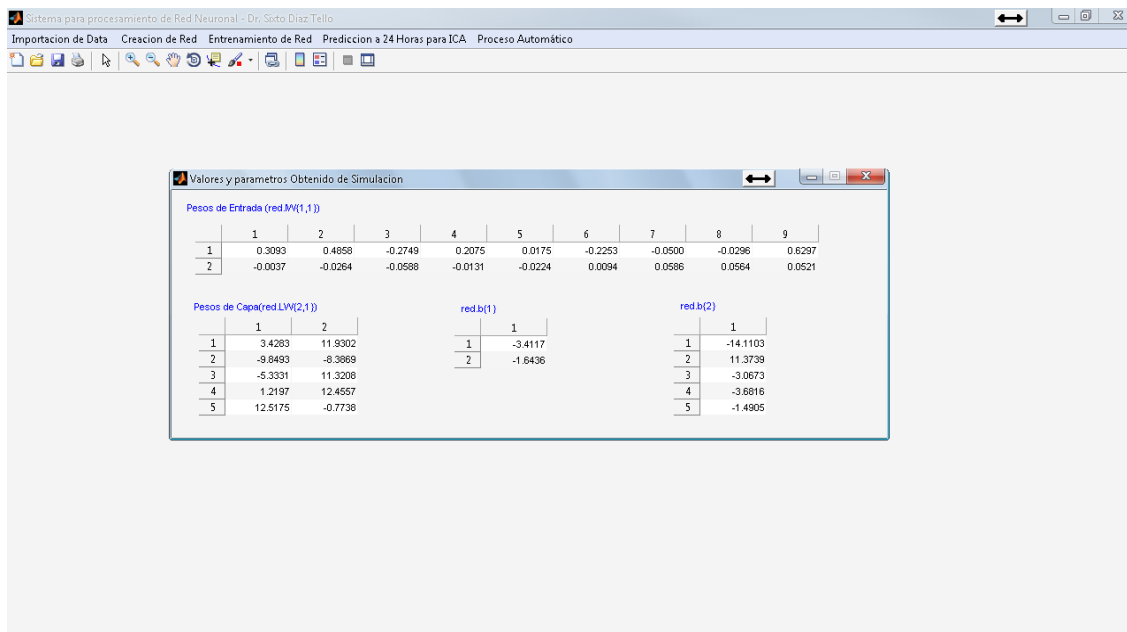
17. Resultado: gráfica Funcionamiento de red en etapas



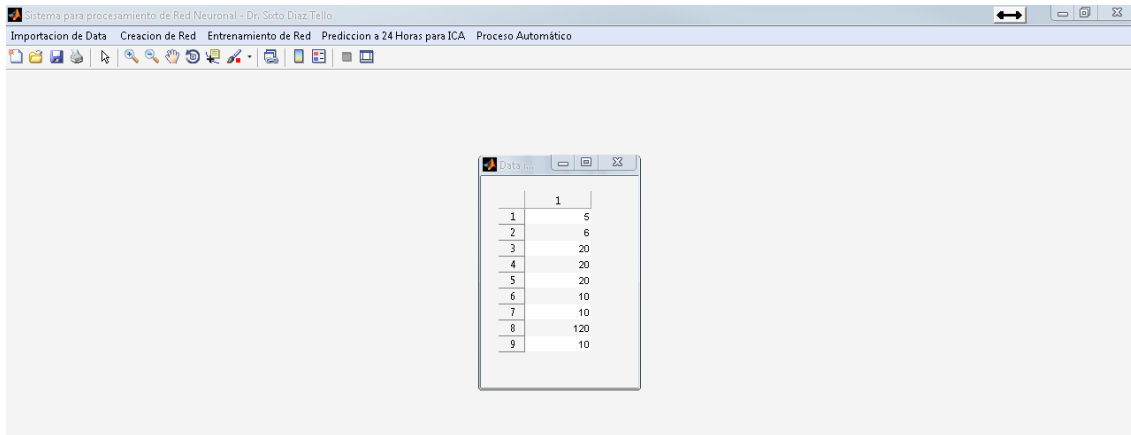
18. Resultado : gráfica Ajustes entre salida deseada y estimada



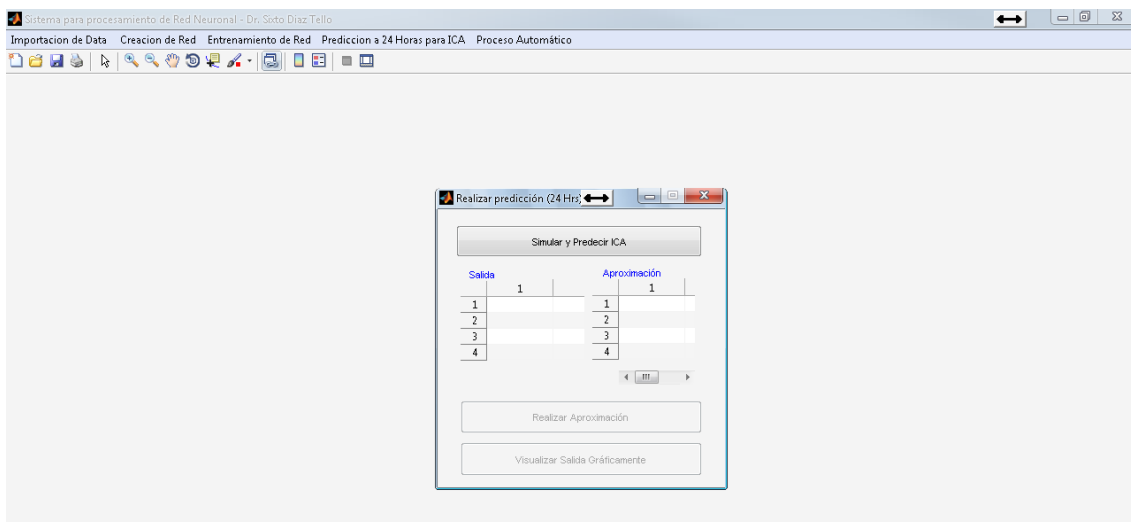
19. Valores y parámetros obtenidos de simulación



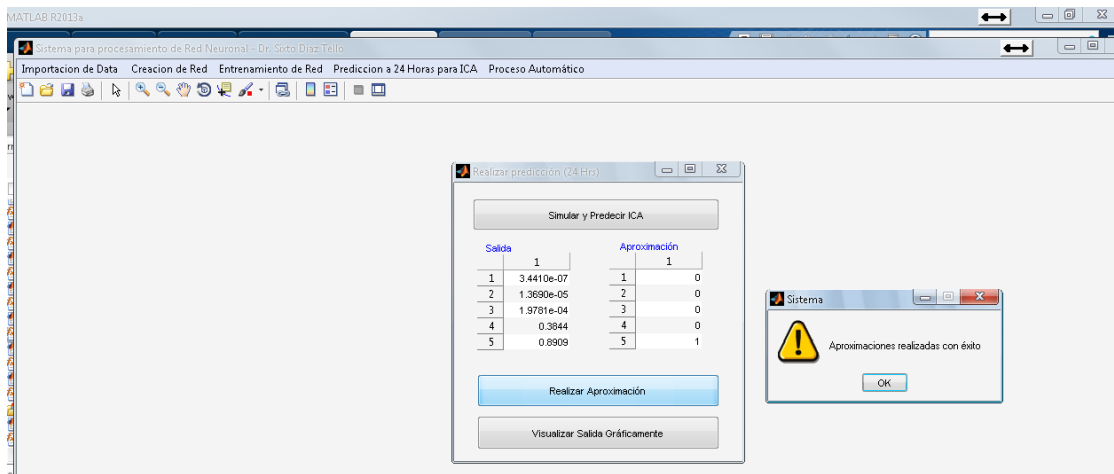
20. Importación de data para predicción: Sirve para importar la data que contiene las entradas para realizar la simulación y predicción a 24 horas.



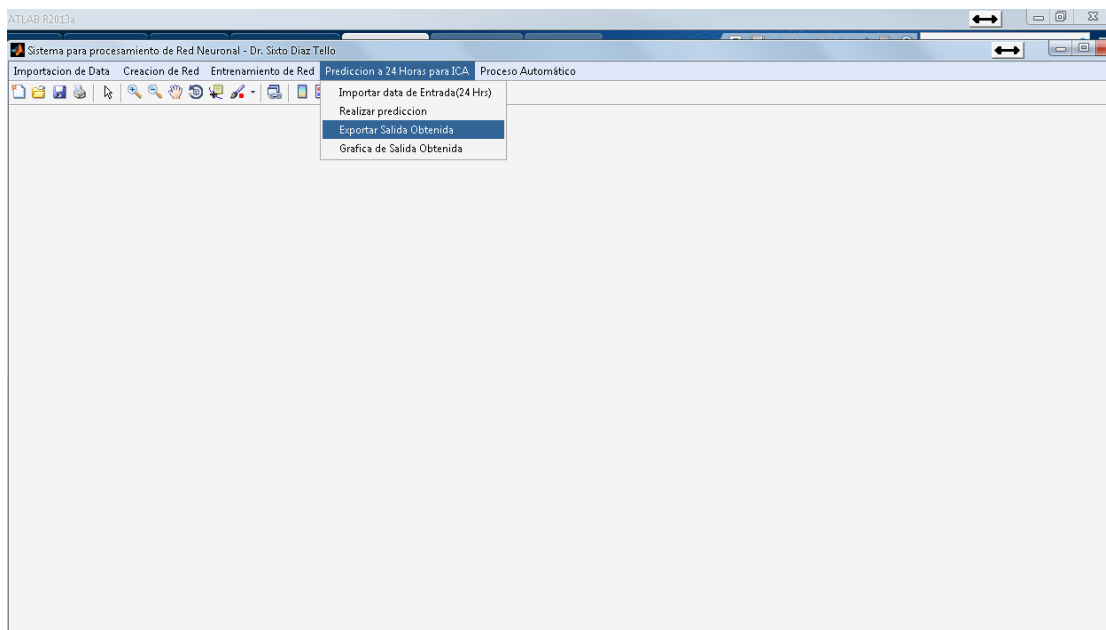
21. Simulación y predicción de ICA:



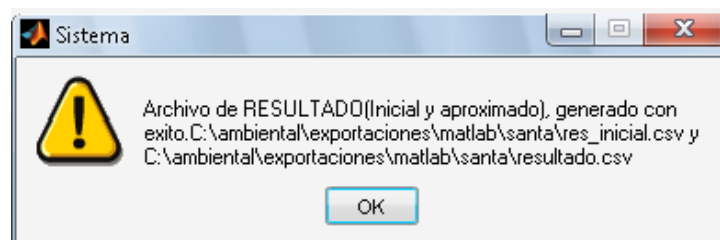
22. Resultados de simulación y realización de aproximación



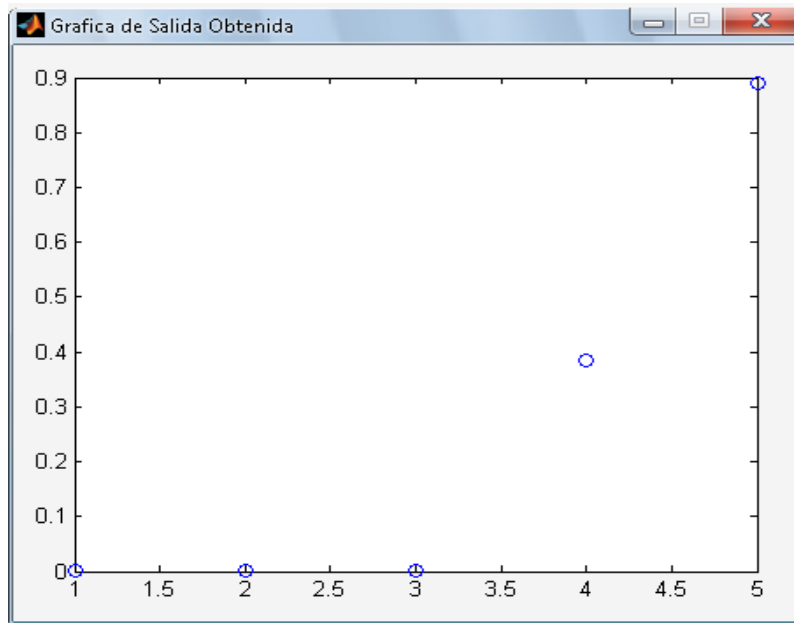
23. Exportación de salida obtenida (resultado y aproximación)



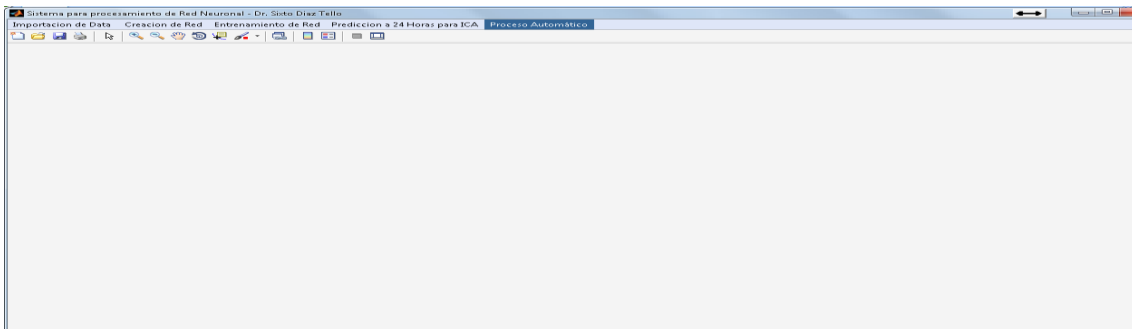
24. Mensaje satisfactorio



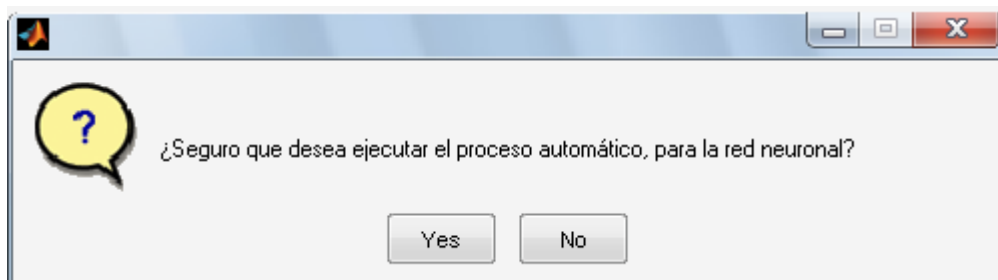
25. Resultado: gráfica salida obtenida



26. Procesamiento automático: Esta opción sirve para que realice la importación, procesamiento, creación, entrenamiento, simulación, predicción, hasta los resultados de la red neuronal en un solo click. Con la finalidad de abreviar tiempo, pero ya no se verán los resultados que van emitiendo en cada paso, sino solo el resultado final.



27. Mensaje de confirmación de procesamiento automático

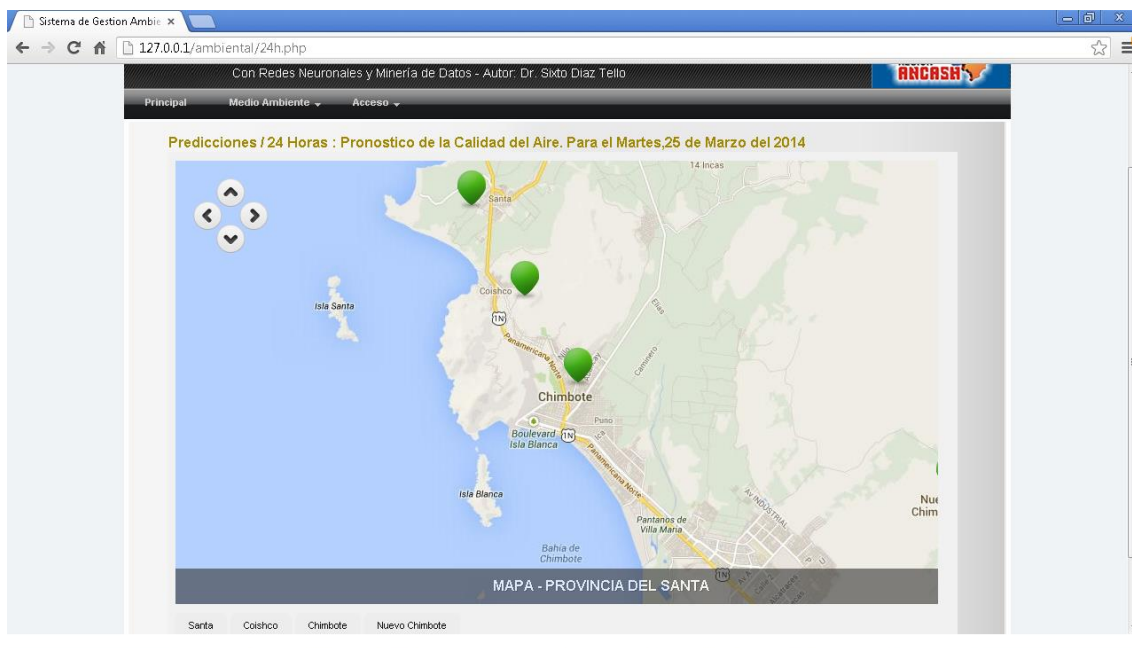


C. Sistema de visualización de predicciones(Sistema WEB)

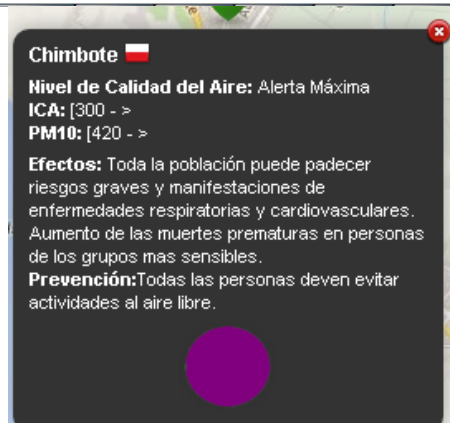
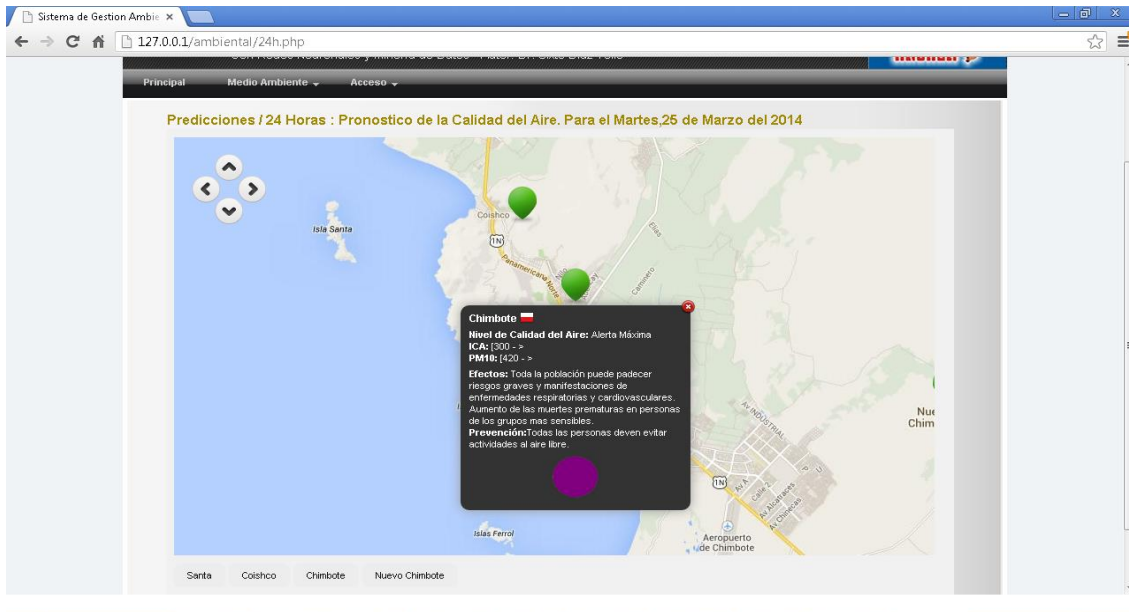
1. Portal principal del sistema web



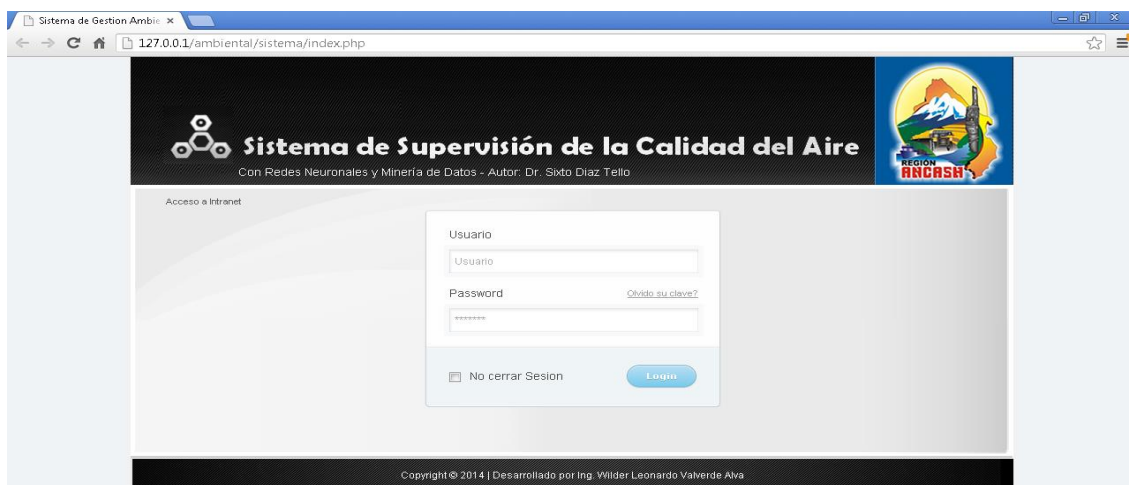
2. Mapa geográfico para la consulta de predicciones del ICA: Se muestra un mapa geográfico interactivo de la provincia del Santa, en la cual servirá para seleccionar la estación que se desee consultar el ICA a 24 horas siguientes.



3. Resultado de la consulta del ICA de una Estación (Ejem. Chimbote).- Al hacer click en cada estación, se mostrará la información correspondiente al ICA y PM10.

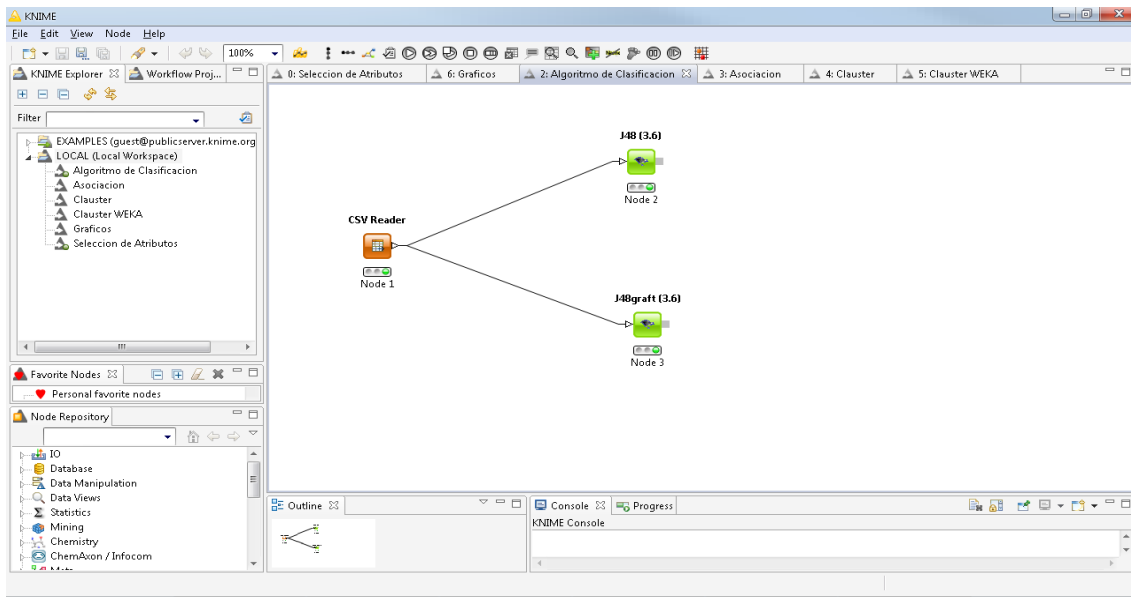


4. Acceso de administrador al sistema web

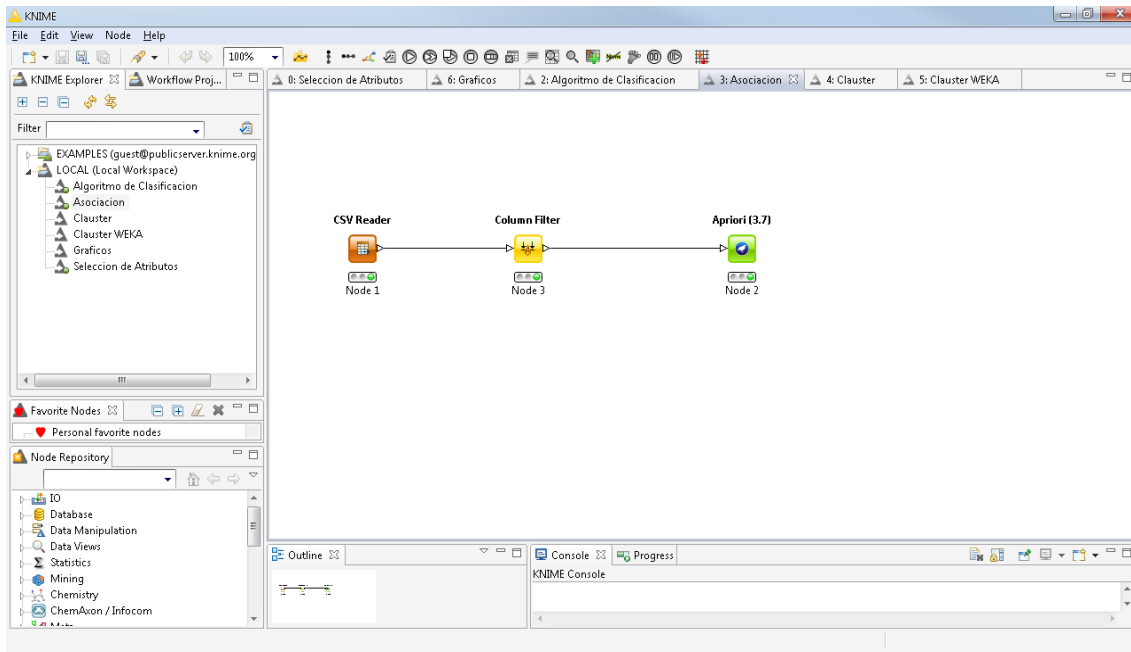


D.SISTEMA DE PROCESAMIENTO DE MINERÍA DE DATOS (KNIME 2.9.2)

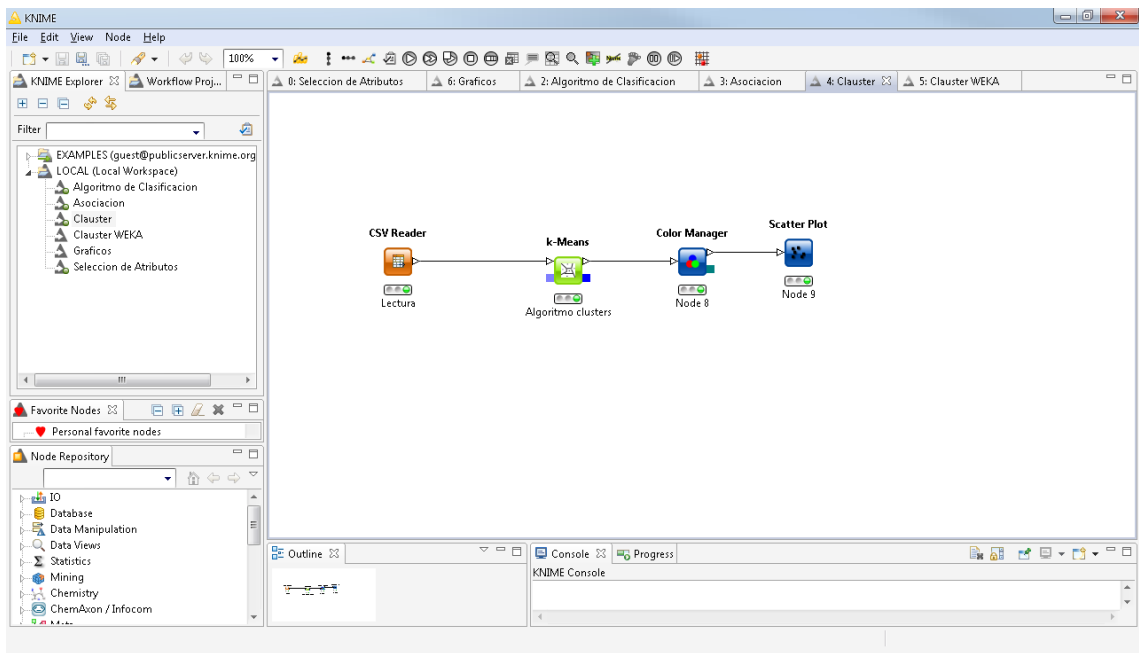
1. Algoritmo de clasificación



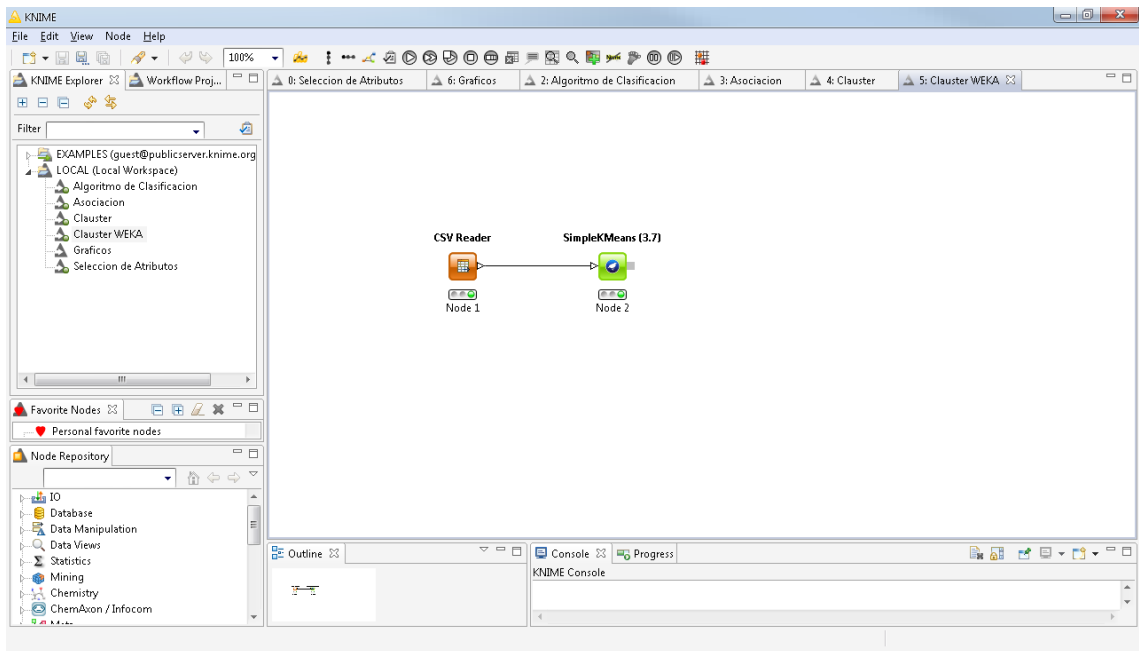
2. Asociación



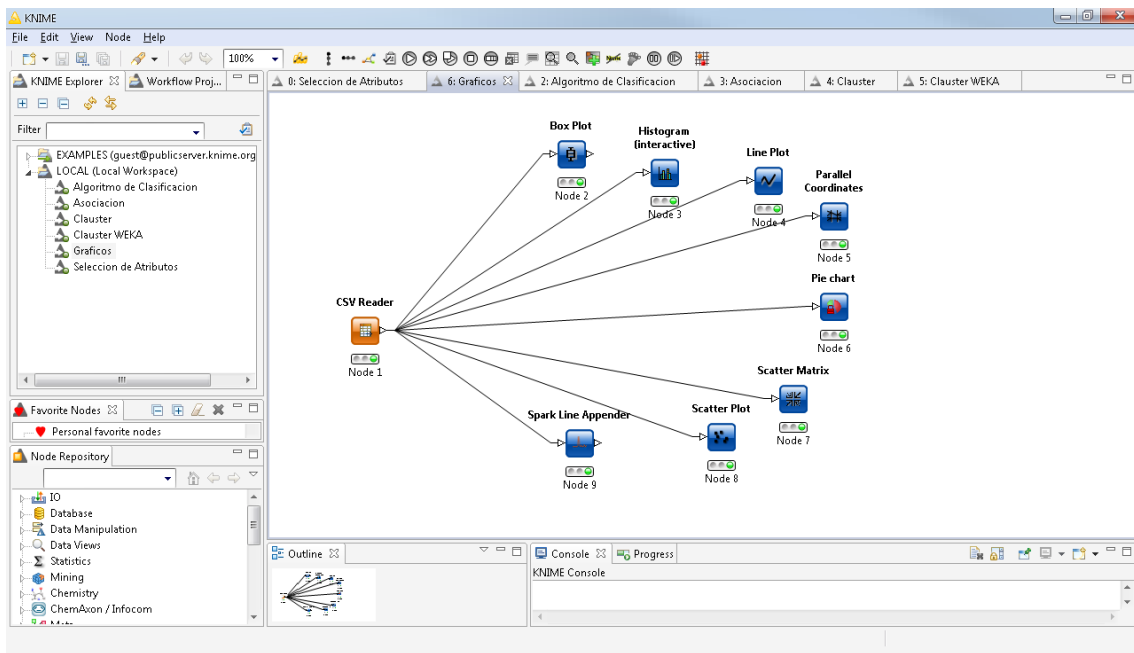
3. Cluster K-means



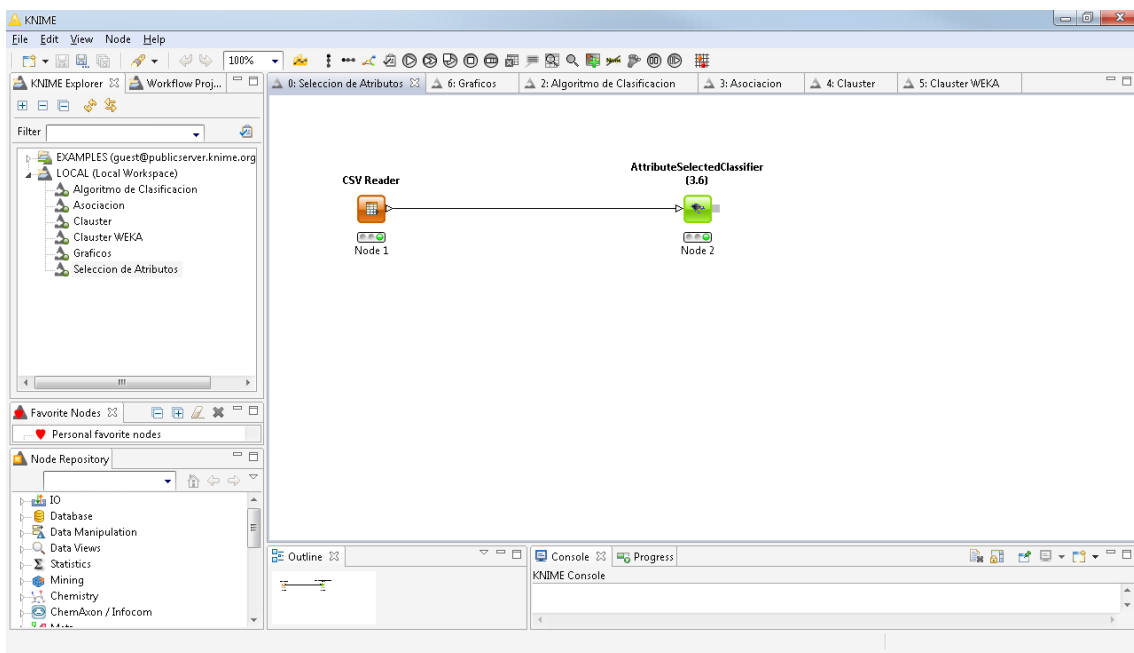
4. Cluster WEKA



5. Gráficos de interpretación



6. Selección de atributos



4.2.3.2. CODIFICACION DEL SISTEMA

A continuación se presenta, líneas de código de los principales procesos que involucra el sistema:

1. Evento Clicked: Registrar o modificar Parámetro ICA

```
boolean validacion
integer nfila

validacion=true

id_parametros=integer(sle_id_parametros.text)
ica_valor_minimo=integer(sle_ica_valor_minimo.text)
ica_valor_maximo=integer(sle_ica_valor_maximo.text)
estado=sle_estado.text

pm_valor_minimo=double(sle_pm_valor_minimo.text)
pm_valor_maximo=double(sle_pm_valor_maximo.text)
efectos=sle_efectos.text
acciones_preventivas=sle_acciones_preventivas.text

if (TRIM(efectos)<>"" and ica_valor_maximo<>0 and pm_valor_maximo<>0) then
    w_parametros.dw_1.ScrollToRow(fila_modifica)
    nfila=fila_modifica

    w_parametros.dw_1.object.id_parametros[nfila]=id_parametros
    w_parametros.dw_1.object.ica_valor_minimo[nfila]=ica_valor_minimo
    w_parametros.dw_1.object.ica_valor_maximo[nfila]=ica_valor_maximo
    w_parametros.dw_1.object.estado[nfila]=estado
    w_parametros.dw_1.object.pm_valor_minimo[nfila]=pm_valor_minimo
    w_parametros.dw_1.object.pm_valor_maximo[nfila]=pm_valor_maximo
    w_parametros.dw_1.object.efectos[nfila]=efectos
    w_parametros.dw_1.object.acciones_preventivas[nfila]=acciones_preventivas

Integer nRet
nRet=w_parametros.dw_1.Update()

If nRet > 0 Then
    Commit;
    If sqlca.SQLCode < 0 Then
        MessageBox("Error en la base de datos ", "Error # "+
String(sqlca.SQLDBCode) + ":"+ sqlca.SQLErrMsg)
        Halt Close
    End If
Else
    Rollback;
    If sqlca.SQLCode < 0 Then
        MessageBox("Error en la base de datos ", "Error # "+
String(sqlca.SQLDBCode) + ":"+ sqlca.SQLErrMsg)
        Halt Close
    End If
End If

modificar_parametros=false

w_parametros.cb_1.enabled=true
// w_parametros.dw_1.settransobject( sqlca)
// w_n_producto.dw_1.retrieve( )
close(w_parametros_n)
else
    MessageBox("Error", "Tiene que completar todos los datos")
    sle_ica_valor_minimo.setfocus( )
```

end if

2. Evento Clicked: Registrar de nueva data meteorológica diaria

```
boolean validacion
integer nfila
validacion=true

id_data=integer(sle_id_data.text)

direccion_viento=double(sle_direccion_viento.text)
velocidad_predominante=double(sle_velocidad_predominante.text)
velocidad=double(sle_velocidad.text)
velocidad_maxima=double(sle_velocidad_maxima.text)
temperatura=double(sle_temperatura.text)
presion=double(sle_presion.text)
humedad=double(sle_humedad.text)
fecha=date(sle_fecha.text)

dia=integer(sle_dia.text)
mes=integer(sle_mes.text)

pm10=double(sle_pm10.text)
if (pm10<50) then

indice_calidad=pm10
end if
if (pm10>=50 and pm10<150 ) then
    indice_calidad=(pm10 - 50) *0.50 + 50
end if
if (pm10>=150 and pm10<250 ) then
    indice_calidad=(pm10 - 150) *1.00 + 100
end if
if (pm10>=250 and pm10<420 ) then
    indice_calidad=(pm10 - 250) *0.5882 + 200
end if
if (pm10>=420) then
    indice_calidad=(pm10 - 420) *1.2069 + 300
end if

indice_calidad_g=sle_indice_calidad.text
estado=sle_estado.text
estado_red=sle_estado_red.text

estacion=sle_estacion.text

if mapa="SI" THEN
    open(w_datos)
end if

if (direccion_viento<>0.00 and velocidad_predominante<>0 and velocidad<>0 and
velocidad_maxima<>0 and temperatura<>0 and presion<>0 and humedad<>0 and pm10<>0
AND estacion<>"" ) then
    if grabar_data=true then
        nFila=w_datos.dw_1.InsertRow(0)
        w_datos.dw_1.ScrollToRow(nFila)
    else
        w_datos.dw_1.ScrollToRow(fila_modifica)
        nfila=fila_modifica
    end if

    w_datos.dw_1.object.id_data[nfila]=id_data
    w_datos.dw_1.object.direccion_viento[nfila]=direccion_viento

w_datos.dw_1.object.velocidad_predominante[nfila]=velocidad_predominante
w_datos.dw_1.object.velocidad[nfila]=velocidad
w_datos.dw_1.object.velocidad_maxima[nfila]=velocidad_maxima
w_datos.dw_1.object.temperatura[nfila]=temperatura
w_datos.dw_1.object.presion[nfila]=presion
w_datos.dw_1.object.humedad[nfila]=humedad
```



```

w_datos.dw_1.object.fecha[nfila]=fecha
w_datos.dw_1.object.dia[nfila]=integer(numero_d)
w_datos.dw_1.object.mes[nfila]=integer(numero_m)
w_datos.dw_1.object.pm10[nfila]=pm10
w_datos.dw_1.object.indice_calidad[nfila]=indice_calidad
w_datos.dw_1.object.indice_calidad_g[nfila]=indice_calidad_g
w_datos.dw_1.object.estado[nfila]=estado
w_datos.dw_1.object.estado_red[nfila]=estado_red

        if (direccion_viento>=0 and direccion_viento<20) then
            direccion_viento_g="[0-20>"
        end if
        if (direccion_viento>=20 and direccion_viento<40) then
            direccion_viento_g="[20-40>"
        end if
        if (direccion_viento>=40 and direccion_viento<60) then
            direccion_viento_g="[40-60>"
        end if
        if (direccion_viento>=60 and direccion_viento<80) then
            direccion_viento_g="[60-80>"
        end if
        if (direccion_viento>=80 and direccion_viento<100) then
            direccion_viento_g="[80-100]"
        end if
w_datos.dw_1.object.direccion_viento_g[nfila]=direccion_viento_g
        if (velocidad_predominante>=0 and velocidad_predominante<5)
            then
                velocidad_predominante_g="[0-5>"
            end if
            if (velocidad_predominante>=5 and velocidad_predominante<10)
                then
                    velocidad_predominante_g="[5-10>"
                end if
            end if
        if (velocidad_predominante>=10 and velocidad_predominante<15) then
            velocidad_predominante_g="[10-15>"
        end if
        if (velocidad_predominante>=15 and velocidad_predominante<20)
            then
                velocidad_predominante_g="[15-20>"
            end if
        if (velocidad_predominante>=20 and velocidad_predominante<25)
            then
                velocidad_predominante_g="[20-25>"
            end if
        if (velocidad_predominante>=25 and velocidad_predominante<30)
            then
                velocidad_predominante_g="[25-30>"
            end if
        if (velocidad_predominante>=30) then
            velocidad_predominante_g="[30-..>"
        end if
w_datos.dw_1.object.velocidad_predominante_g[nfila]=velocidad_predominante_g
        if (velocidad>=0 and velocidad<5) then
            velocidad_g="[0-5>"
        end if
        if (velocidad>=5 and velocidad<10) then
            velocidad_g="[5-10>"
        end if
        if (velocidad>=10 and velocidad<15) then
            velocidad_g="[10-15>"
        end if
        if (velocidad>=15 and velocidad<20) then
            velocidad_g="[15-20>"
        end if
        if (velocidad>=20 and velocidad<25) then
            velocidad_g="[20-25>"
        end if

```

```

if (velocidad>=25 and velocidad<30) then
    velocidad_g="[25-30>"
end if
if (velocidad>=30) then
    velocidad_g="[30-..>"
end if
w_datos.dw_1.object.velocidad_g[nfila]=velocidad_g
if (velocidad_maxima>=0 and velocidad_maxima<10) then

    velocidad_maxima_g="[0-10>"
end if
if (velocidad_maxima>=10 and velocidad_maxima<20) then

    velocidad_maxima_g="[10-20>"
end if
if (velocidad_maxima>=20 and velocidad_maxima<30) then

    velocidad_maxima_g="[20-30>"
end if
if (velocidad_maxima>=30 and velocidad_maxima<40) then

    velocidad_maxima_g="[30-40>"
end if
if (velocidad_maxima>=40 and velocidad_maxima<50) then

    velocidad_maxima_g="[40-50>"
end if
if (velocidad_maxima>=50 and velocidad_maxima<60) then

    velocidad_maxima_g="[50-60>"
end if
if (velocidad_maxima>=60) then
    velocidad_maxima_g="[60-..>"
end if
w_datos.dw_1.object.velocidad_maxima_g[nfila]=velocidad_maxima_g
if (temperatura>=0 and temperatura<5) then
    temperatura_g="[0-5>"
end if
if (temperatura>=5 and temperatura<10) then
    temperatura_g="[5-10>"
end if
if (temperatura>=10 and temperatura<15) then
    temperatura_g="[10-15>"
end if
if (temperatura>=15 and temperatura<20) then
    temperatura_g="[15-20>"
end if
if (temperatura>=20 and temperatura<25) then
    temperatura_g="[20-25>"
end if
if (temperatura>=25 and temperatura<30) then
    temperatura_g="[25-30>"
end if
if (temperatura>=30 and temperatura<35) then
    temperatura_g="[30-35>"
end if
if (temperatura>=35) then
    temperatura_g="[35-..>"
end if
w_datos.dw_1.object.temperatura_g[nfila]=temperatura_g
if ( presion<1000) then
    presion_g="[..-1000>"
end if
if (presion>=1000 and presion<1010) then
    presion_g="[1000-1010>"
end if
if (presion>=1010 and presion<1020) then

```

```

        presion_g="[1010-1020>"
    end if
    if (presion>=1020 and presion<1030) then
        presion_g="[1020-1030>"
    end if
    if (presion>=1030 and presion<1040) then
        presion_g="[1030-1040>"
    end if
    if (presion>=1040 and presion<1050) then
        presion_g="[1040-1050>"
    end if
    if (presion>=1050) then
        presion_g="[1050-..>"
    end if
w_datos.dw_1.object.presion_g[nfila]=presion_g
    if ( humedad<1000) then
        humedad_g="[..-1000>"
    end if
    if (humedad>=1000 and humedad<1010) then
        humedad_g="[1000-1010>"
    end if
    if (humedad>=1010 and humedad<1020) then
        humedad_g="[1010-1020>"
    end if
    if (humedad>=1020 and humedad<1030) then
        humedad_g="[1020-1030>"
    end if
    if (humedad>=1030 and humedad<1040) then
        humedad_g="[1030-1040>"
    end if
    if (humedad>=1040 and humedad<1050) then
        humedad_g="[1040-1050>"
    end if
    if (humedad>=1050) then
        humedad_g="[1050-..>"
    end if
    w_datos.dw_1.object.humedad_g[nfila]=humedad_g
    w_datos.dw_1.object.dia_g[nfila]=nombre_d
    w_datos.dw_1.object.mes_g[nfila]=nombre_m
    w_datos.dw_1.object.estacion[nfila]=estacion
Integer nRet
nRet=w_datos.dw_1.Update()
    If nRet > 0 Then
        Commit;
        If sqlca.SQLCode < 0 Then
            MessageBox("Error en la base de datos ", "Error # "+
String(sqlca.SQLDBCode) + ":" + sqlca.SQLErrMsg)
            Halt Close
        End If
    Else
        Rollback;
        If sqlca.SQLCode < 0 Then
            MessageBox("Error en la base de datos ", "Error # "+
String(sqlca.SQLDBCode) + ":" + sqlca.SQLErrMsg)
            Halt Close
        End If
    End If
    modificar_data=false
    grabar_data=false
    w_datos.cb_1.enabled=true
    w_datos.cb_2.enabled=true
    w_datos.cb_3.enabled=true

    if mapa="SI" THEN
        close(w_datos)
    end if
close(w_datos_n)

```

```

else
    MessageBox("Error","Tiene que completar todos los datos")
    sle_direccion_viento.setFocus( )
    if mapa="SI" THEN
        close(w_datos)
    end if
end if

```

3. Evento modified: Cálculo de resultados PM10

```

    int g
    for g=1 to dw_1.rowcount( )
        if (dw_1.object.pm_valor_minimo[g]<=double(sle_pm10.text) and
dw_1.object.pm_valor_maximo[g]>double(sle_pm10.text) ) then
            sle_indice_calidad.text="["+string(dw_1.object.ica_valor_minimo[g])+ " - "+
string(dw_1.object.ica_valor_maximo[g])+"]"
            sle_estado.text=dw_1.object.estado[g]
            if (sle_estado.text="Bueno") then
                r_2.visible=true
                r_4.visible=false
                r_5.visible=false
                r_6.visible=false
                r_7.visible=false
                sle_estado_red.text="00001"
            end if

            if (sle_estado.text="Moderado") then
                r_2.visible=false
                r_4.visible=true
                r_5.visible=false
                r_6.visible=false
                r_7.visible=false

                sle_estado_red.text="00010"
            end if

            if (sle_estado.text="Malo") then
                r_2.visible=false
                r_4.visible=false
                r_5.visible=true
                r_6.visible=false
                r_7.visible=false
                sle_estado_red.text="00100"
            end if

            if (sle_estado.text="Muy Malo") then
                r_2.visible=false
                r_4.visible=false
                r_5.visible=false
                r_6.visible=true
                r_7.visible=false
                sle_estado_red.text="01000"
            end if

            if (sle_estado.text="Alerta Máxima") then
                r_2.visible=false
                r_4.visible=false
                r_5.visible=false
                r_6.visible=false
                r_7.visible=true
                sle_estado_red.text="10000"
            end if
        end if
    next

```

4. Evento clicked: Procesamiento de input para Matlab

```

dw_1.settransobject(sqlca)
dw_1.retrieve()

```

```

dw_1.setfilter("")
dw_1.filter()
dw_1.setfilter("estacion="+sle_estacion.text+"")
dw_1.filter()
dw_2.settransobject(sqlca)
dw_2.retrieve()
dw_3.settransobject(sqlca)
dw_3.retrieve()
dw_3.setfilter("")
dw_3.filter()
dw_3.setfilter("estacion="+sle_estacion.text+"")
dw_3.filter()

ancho=0
integer i,nfila
if (dw_1.rowcount()=0) then
    messagebox("Sistema","No se encontro data para procesar")
else
    for i=1 to dw_3.rowcount()
        nFila=dw_2.InsertRow(0)
        dw_2.ScrollToRow(nFila)
        if dw_3.object.estado[i]="Bueno" then
            dw_2.object.a[nFila]=0
            dw_2.object.b[nFila]=0
            dw_2.object.c[nFila]=0
            dw_2.object.d[nFila]=0
            dw_2.object.e[nFila]=1
        end if
        if dw_3.object.estado[i]="Moderado" then
            dw_2.object.a[nFila]=0
            dw_2.object.b[nFila]=0
            dw_2.object.c[nFila]=0
            dw_2.object.d[nFila]=1
            dw_2.object.e[nFila]=0
        end if
        if dw_3.object.estado[i]="Malo" then
            dw_2.object.a[nFila]=0
            dw_2.object.b[nFila]=0
            dw_2.object.c[nFila]=1
            dw_2.object.d[nFila]=0
            dw_2.object.e[nFila]=0
        end if
        if dw_3.object.estado[i]="Muy Malo" then
            dw_2.object.a[nFila]=0
            dw_2.object.b[nFila]=1
            dw_2.object.c[nFila]=0
            dw_2.object.d[nFila]=0
            dw_2.object.e[nFila]=0
        end if
        if dw_3.object.estado[i]="Alerta Máxima" then
            dw_2.object.a[nFila]=1
            dw_2.object.b[nFila]=0
            dw_2.object.c[nFila]=0
            dw_2.object.d[nFila]=0
            dw_2.object.e[nFila]=0
        end if

    next
    cb_2.enabled=true
    cb_1.enabled=FALSE

    sle_estacion.enabled=false
    cb_4.enabled=true
    timer(0.04)
end if

```

4.2.3.3. PRUEBAS DEL SISTEMA

Las pruebas del software se realizaron a través de la aplicación de un cuestionario y también se diseñó los casos de prueba requeridos (prueba de Caja Negra) para los procesos principales, así como su interoperabilidad a través de la Base de Datos.

4.2.3.3.1. Cuestionario

El cuestionario está orientado a evaluar la amigabilidad del sistema, el cumplimiento de requerimientos y la exactitud de la información contenida en los reportes e informes. Los resultados del mismo se muestran a continuación:

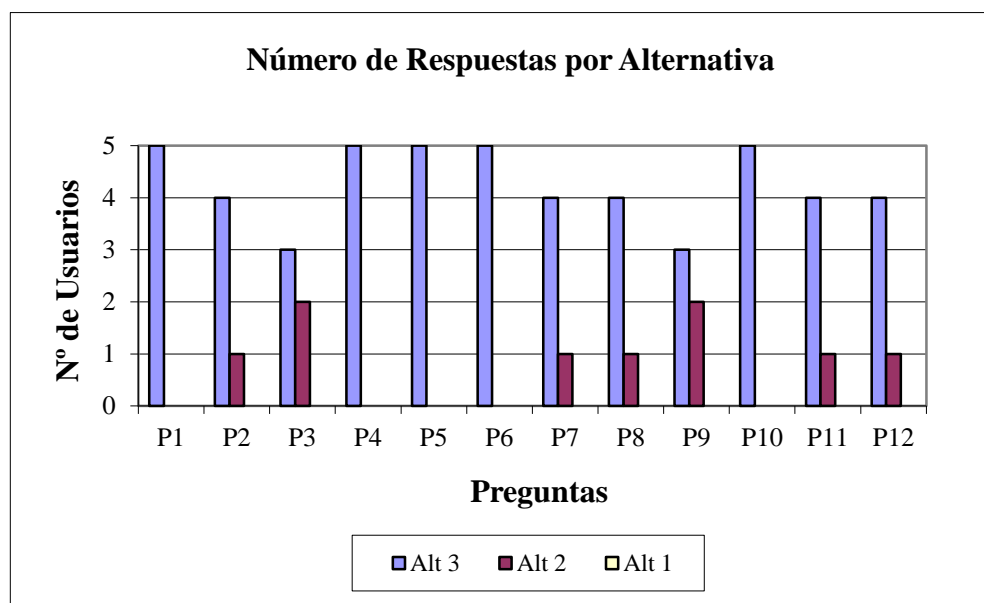


Figura 4.95. Resultados de cuestionario

Estos resultados nos demuestran que el sistema es fácil de usar, cumple los requerimientos de usuario y la información que proporciona es exacta.

4.2.3.3.2. Realizar pruebas del sistema.

Se realizaron las pruebas planificadas, interviniendo en este proceso, personal de las áreas de ingeniería, químicos, entre otros profesionales entendidos en el tema. Las realizaciones se dieron de la siguiente forma:

- ✓ Se registraron 05 parámetros de ICA
- ✓ Se registraron un promedio de 500 registros de data meteorológica diaria.
- ✓ Se validaron aproximadamente 50 veces las 09 variables para la red neuronal.
- ✓ Se exportaron los input de entrada y salida (40), data para predicción (80), data para knime (10).
- ✓ Se procesaron 45 veces la red neuronal, a través del sistema desarrollado en Matlab.
- ✓ Se realizó aproximadamente 240 entrenamientos de la red.
- ✓ Se realizaron alrededor de 40 predicciones a 24 horas, con sus respectivas aproximaciones en el sistema desarrollado en Matlab.
- ✓ Se realizaron 200 consultas a través del sistema web, para visualizar las predicciones de las 4 estaciones.
- ✓ Se generaron 25 reportes gráficos de Matlab

4.2.4. FASE DE TRANSICIÓN

4.2.4.1. PREPARACIÓN DE LA VERSIÓN BETA

Se terminó de integrar el Sistema (Power Builder, Matlab, Web, Knime), tanto con los referencias a las librerías, como a los gráficos (presentación de ventanas), animaciones, en una sola carpeta.

Se configuró la conexión con la base datos.

4.2.4.2. INSTALACIÓN DE LA VERSIÓN BETA

Para la instalación de la versión beta, se hace uso del programa FileZilla, para subir los archivos al servidor Web, solamente se copió la carpeta que contenía los archivos en el cliente FileZilla. Para el caso del resto de sistemas, se instalaron con sus respectivas librerías en la PC del administrador.

4.2.4.3. REACCIÓN A LOS RESULTADOS DE PRUEBA

- ✓ Incredulidad a cuanto a los resultados de reportes.
- ✓ Aparición de nuevos requerimientos para implementar

4.2.4.4. FINALIZACIÓN DEL ANÁLISIS DE NEGOCIO

Se aceptó la implementación de la Versión Beta la cual pasará por un periodo de prueba de dos semanas, que servirá para el ingreso de toda la data histórica, que servirá para arrojar mejores resultados a través del entrenamiento de la red. Se realizaron algunas correcciones y personalizaciones de tal manera que se presenten interfaces amigables para el usuario.

CONCLUSIÓN

Se culminó con la integración y la instalación de la versión Beta, lo cual se concretó con la capacitación, la elaboración del manual y la puesta en marcha de la Versión Beta.

4.3. RESULTADOS PROCESAMIENTO DE MINERÍA DE DATOS

4.3.1. Descripción de la herramienta.

KNIME es un entorno totalmente gratuito para el desarrollo y ejecución de técnicas de minería de datos. KNIME fue desarrollado originalmente en el departamento de bioinformática y minería de datos de la Universidad de

Constanza, Alemania, bajo la supervisión del profesor Michael Berthold. En la actualidad, la empresa KNIME.com, GmbH, radicada en Zúrich, Suiza, continúa su desarrollo, además de prestar servicios de formación y consultoría.

KNIME está desarrollado sobre la plataforma Eclipse y programado, esencialmente, en Java. Como otros entornos de este tipo, algunos de los cuales aparecen referenciados al final de este documento, su uso se basa en el diseño de un flujo de ejecución que plasme las distintas etapas de un proyecto de minería de datos

Para ello, KNIME proporciona distintos nodos agrupados en fichas:

- a) *Entrada de datos [IO > Read].*
- b) *Salida de datos [IO > Write].*
- c) *Preprocesamiento [Data Manipulation], para filtrar, discretizar, normalizar, filtrar, seleccionar variables...*
- d) *Minería de datos [Mining], para construir modelos (reglas de asociación, clustering, clasificación, MDS, PCA...).*
- e) *Salida de resultados [Data Views] para mostrar resultados en pantalla (ya sea de forma textual o gráfica).*

KNIME incluye otras herramientas además de la explicada anteriormente, que se pueden identificar fácilmente:

- ✓ **La ventana de Editor:** facilita la visualización de los diagramas de ejecución construidos.
- ✓ **Repositorio de Nodos:** se encuentran los nodos que pueden incorporarse a un flujo.

- ✓ **Consola de eventos:** visualiza los mensajes de warning y error. Estos mensajes se almacenan en un archivo, llamado knime.log (stack trace), ubicado en la carpeta .metada del proyecto.
- ✓ **Descripción de Nodos:** provee información del nodo seleccionado. La misma se encuentra vacía si no se ha seleccionado ningún nodo. En el caso de seleccionar una categoría que incluye más de un nodo, muestra los nodos que forman parte de esa categoría.
- ✓ **Explorador de Proyectos:** muestra todos los proyectos existentes en el workspace, esta herramienta permite la creación de proyectos, importar proyectos existentes o exportar.

4.3.2. Preparación de data para realización de minería de datos.

Knime, necesita de datos, para poder realizar sus procesos, para esto el sistema realizado en Power Builder da la opción de poder procesar y exportar la información en los formatos aceptados por Knime, de tal manera que se vinculen directamente para poder procesar los modelos ya diseñados en esta herramienta.

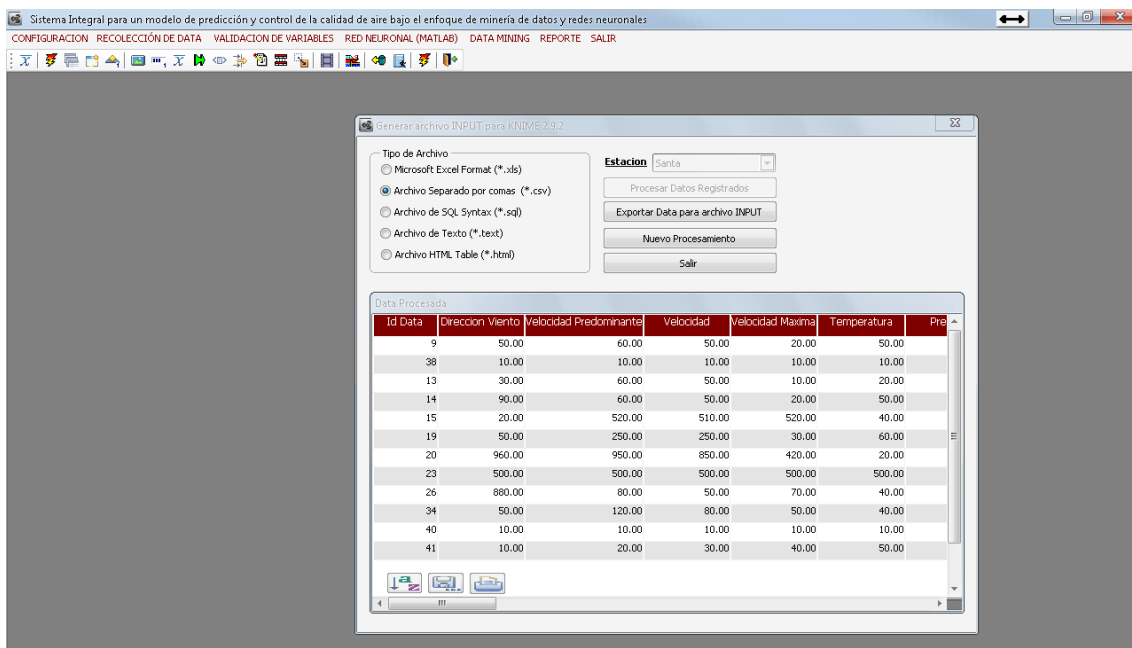


Figura N° 4.96: Preparación de data para knime

4.3.3. Modelos implementados en knime

A. Modelo: Árbol de clasificación

Es uno de los métodos de aprendizaje inductivo supervisado no paramétrico más utilizado. Como forma de representación del conocimiento, los árboles de clasificación destacan por su sencillez. A pesar de que carecen de la expresividad de las redes semánticas o de la lógica de primer orden

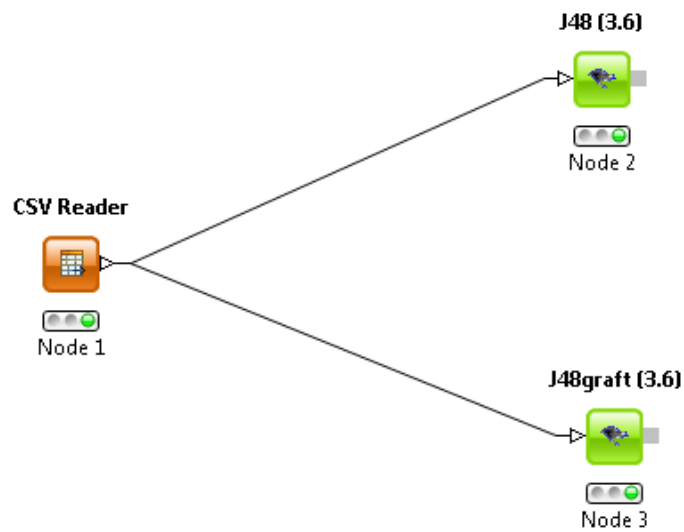


Figura N° 4.97: árbol de clasificación

B. Modelo: Algoritmo de asociación

El **algoritmo a priori** se usa en minería de datos para encontrar Reglas de asociación en un conjunto de datos. Este algoritmo se basa en el conocimiento previo o “a priori” de los conjuntos frecuentes, esto sirve para reducir el espacio de búsqueda y aumentar la eficiencia.

El objetivo de las reglas de asociación es encontrar asociaciones o correlaciones entre los elementos u objetos de bases de datos transaccionales, relacionales o data warehouses

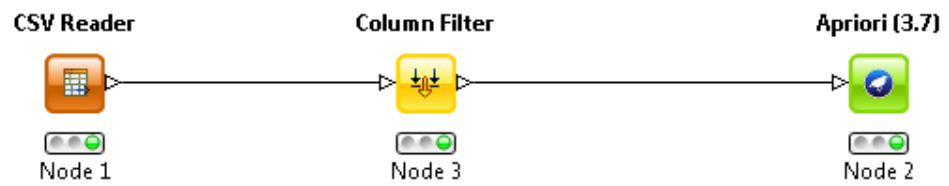


Figura N° 4.98: Algoritmo de asociación

C. Modelo: Cluster k-means:

Permiten clasificar un conjunto de elementos de la data que se está procesando, en un determinado número de grupos basándose en las semejanzas y diferencias existentes entre los componentes de la muestra.

Para este caso se hace uso del algoritmo k-means

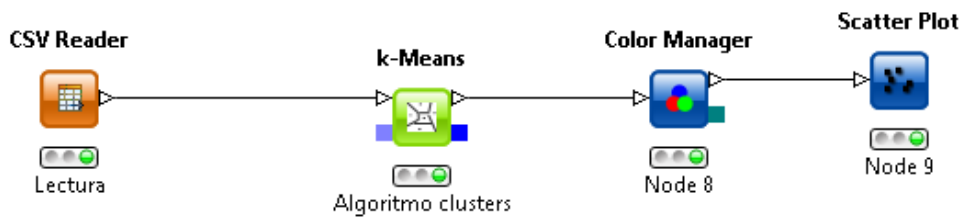


Figura N° 4.99: Cluster k-means

D. Modelo: Cluster weka

Permiten clasificar un conjunto de elementos de la data que se está procesando, en un determinado número de grupos basándose en las semejanzas y diferencias existentes entre los componentes de la muestra.

Para este caso se hace uso del algoritmo Simple k-Means

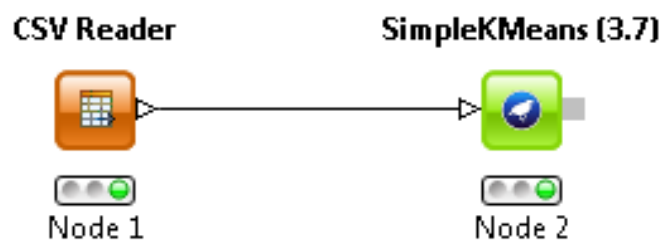


Figura N° 4.100: Cluster weka

E. Modelo: Gráficos de minería de datos

Se presentan diversos tipos de gráficos que usan la técnica de minería de datos, para tener un mejor entendimiento, de acuerdo a los parámetros configurados en cada uno de estos.

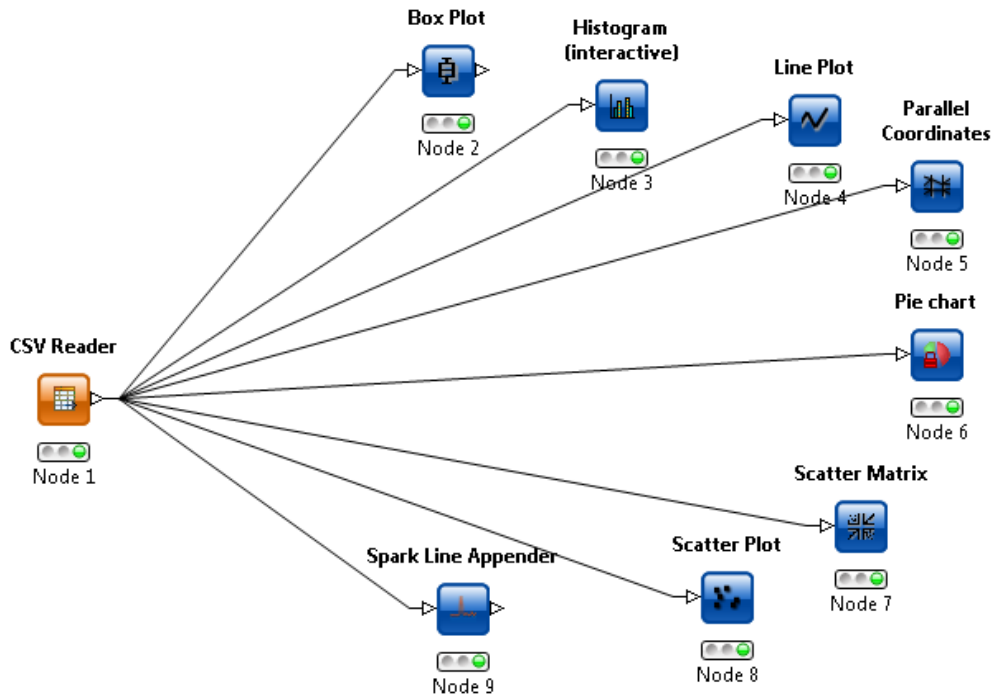


Figura N° 4.101: Gráficos minería de datos

F. Modelo: Selección de atributos.

Esta etapa es crucial para este proyecto, es fundamental la elección adecuada de cuáles van a ser las variables consideradas para la generación de patrones de comportamiento. Este trabajo pretende focalizarse en algunas variables que se encuentran vinculadas directamente a los factores de contaminación atmosférica, con el objetivo de descubrir relaciones y aspectos que influyen en los reportes de toma de decisiones.



Figura N° 4.102: Selección de atributos

4.3.4. Procesamiento y resultados de los modelos

A. Modelo: Árbol de clasificación

- Configuración de Algoritmo

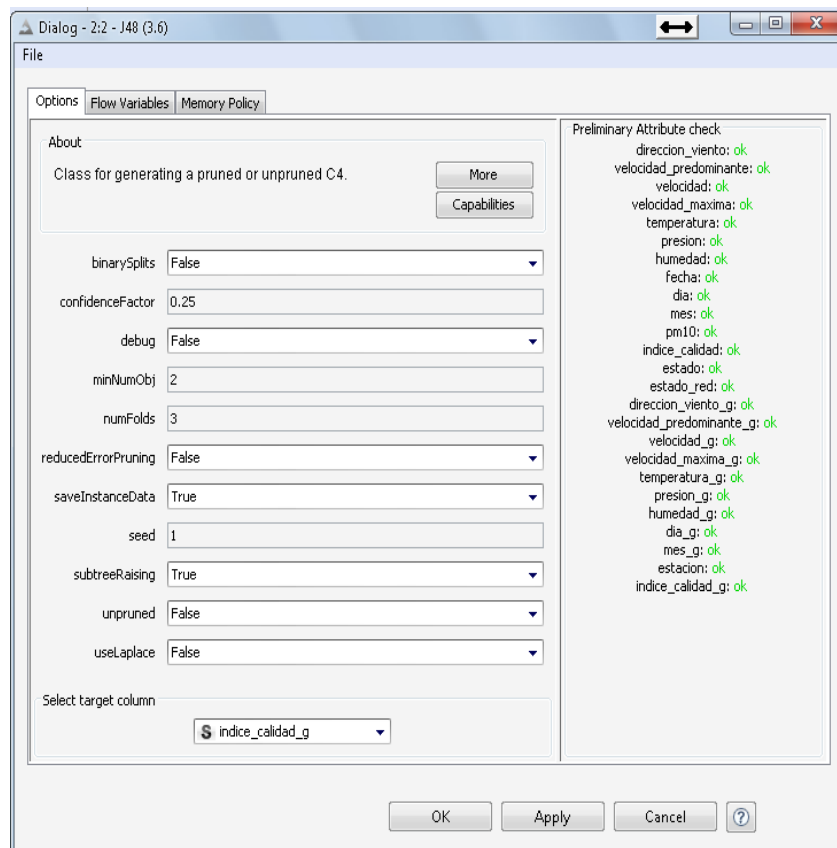


Figura N° 4.103: Configuración de algoritmos

- Resultados de algoritmo J48

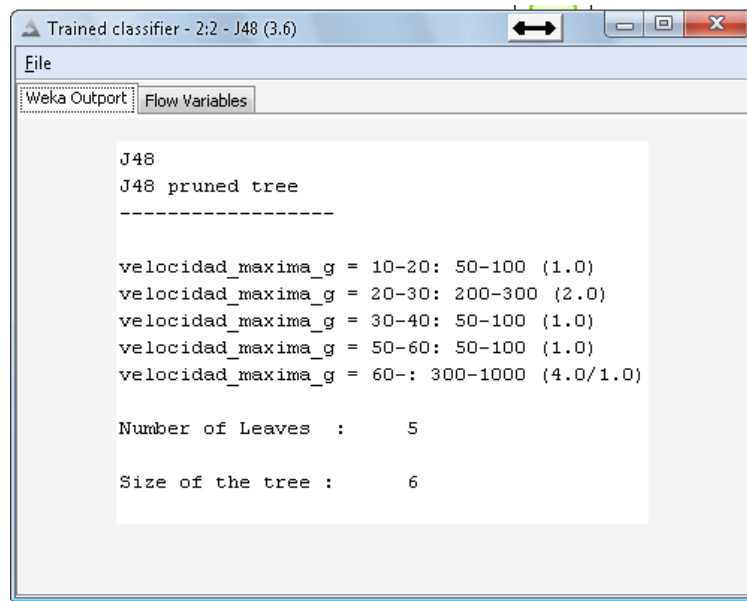


Figura N° 4.104: Resultado Algoritmo

- Resultado árbol

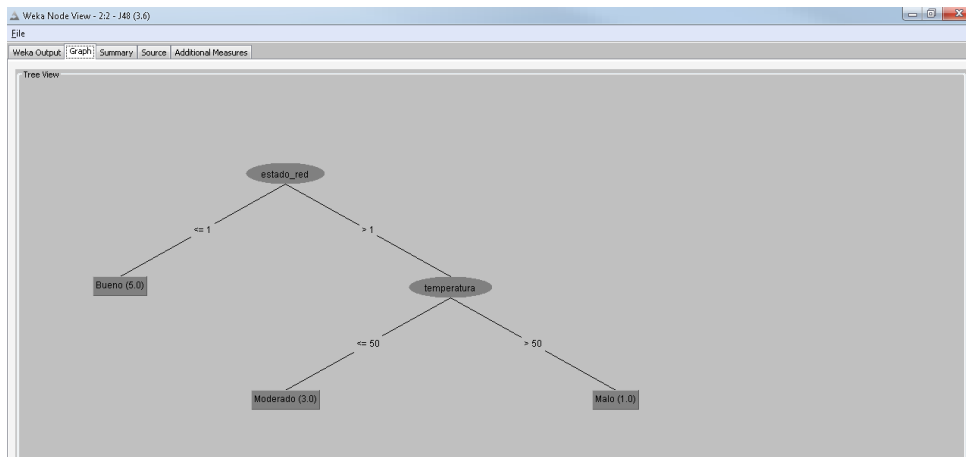


Figura N° 4.105: Resultado árbol

- Resultados algoritmo J48graft

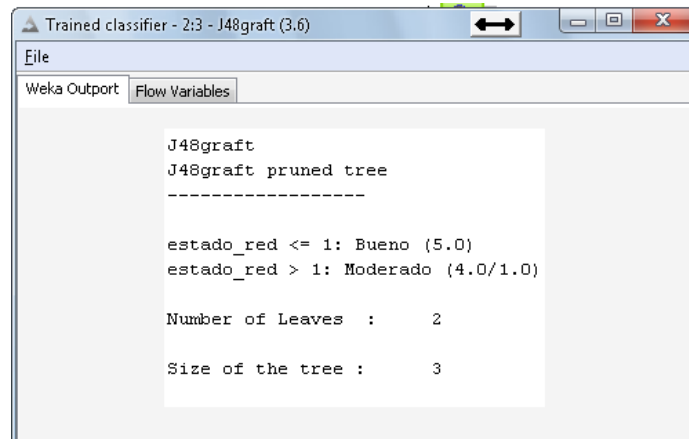


Figura N° 4.106: Algoritmo J48graft

B. Modelo: Algoritmo de asociación

- Configuración algoritmo A priori

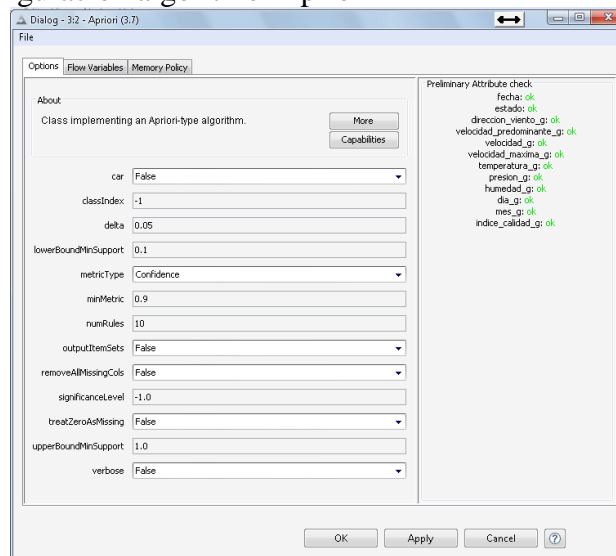


Figura N° 4.107: Configuración algoritmo a priori

- Filtro de columnas

Row ID	S fecha	S estado	S direccio...	S velocid...	S velocid...	S velocid...	S temper...	S presion_g	S humed.
9	01/01/2014	Bueno	[40-60>	[30-..>	[30-..>	[20-30>	[35-..>	[...-1000>	[...-1000>
13	26/02/2014	Bueno	[20-40>	[30-..>	[30-..>	[10-20>	[20-25>	[1020-1030>	[...-1000>
14	24/02/2014	Bueno	[80-100]	[30-..>	[30-..>	[20-30>	[35-..>	[1050-..>	[1020-1030]
15	03/03/2014	Moderado	[20-40>	[30-..>	[30-..>	[60-..>	[35-..>	[...-1000>	[1030-1040]
19	05/04/2014	Malo	[40-60>	[30-..>	[30-..>	[30-40>	[35-..>	[...-1000>	[1020-1030]
20	09/03/2014	Bueno	[40-60>	[30-..>	[30-..>	[60-..>	[20-25>	[1050-..>	[1050-..>
23	04/03/2014	Bueno	?	[30-..>	[30-..>	[60-..>	[35-..>	[...-1000>	[1050-..>
26	09/04/2014	Moderado	?	[30-..>	[30-..>	[60-..>	[35-..>	[...-1000>	[1050-..>
34	08/04/2014	Moderado	[40-60>	[30-..>	[30-..>	[50-60>	[35-..>	[...-1000>	[1050-..>

Figura N° 4.108: Filtro de columnas

- Visualización de nodos weka

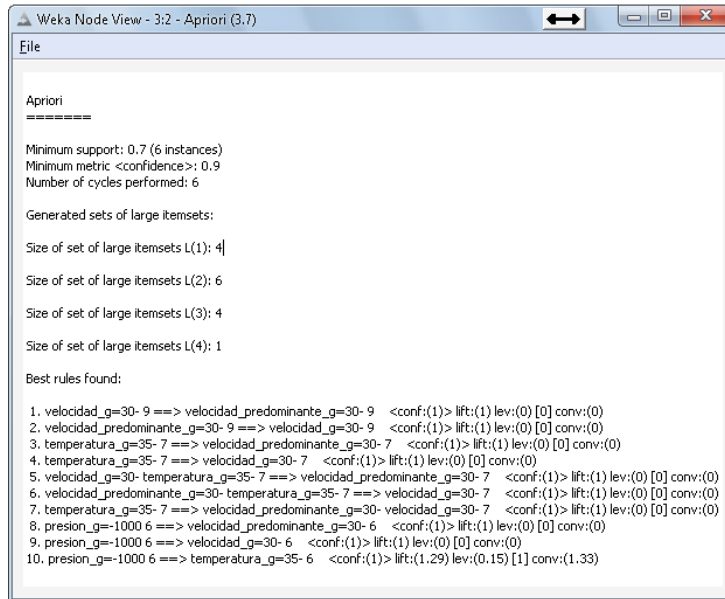


Figura N° 4.109: Visualización de nodos weka

C. Modelo: Cluster k-means:

- Configuración de cluster

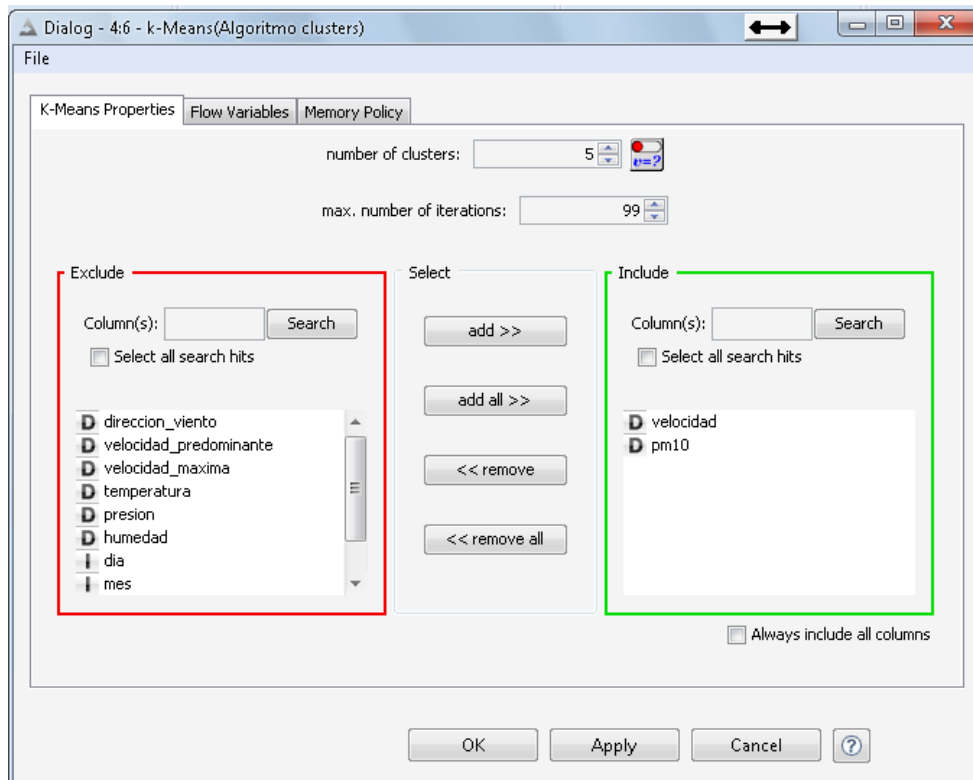


Figura N° 4.110: Configuración de cluster

- Etiquetas de entrada

Row ID	direccion	velocidad	velocidad	velocidad	temper...	presion	humedad	fecha	dia	mes	
1	50	60	10	20	50	10	20	01/01/2014	1	1	17
2	10	10	20	10	10	10	10	01/07/2014	2	7	10
3	30	60	30	10	20	1020	10	26/02/2014	3	2	15
4	90	60	40	20	50	5520	50	24/02/2014	4	2	20
5	20	520	50	520	40	50	60	03/03/2014	5	3	10
6	50	250	60	30	60	20	50	05/04/2014	6	4	15
7	960	950	70	420	20	6550	690	09/03/2014	7	3	40
8	34	67	23	32	50	10	20	01/01/2014	20	1	17
9	23	86	44	10	10	10	10	01/07/2014	21	7	10
10	30	60	56	10	20	1020	10	26/02/2014	22	2	15
11	32	60	34	20	50	5520	50	24/02/2014	23	2	20
12	20	23	43	520	40	50	60	03/03/2014	24	3	10
13	50	250	250	30	60	20	50	05/04/2014	25	4	15
14	960	950	850	420	20	6550	690	09/03/2014	26	3	40
15	500	500	500	500	500	500	500	04/03/2014	27	3	40
16	50	60	50	23	50	10	20	01/01/2014	28	1	17

Figura N° 4.111: Etiqueta de entradas

- Modelo de cluster

```

PMML version="4.1" xmlns="http://www.dmg.org/PMML-4_1"
  Header copyright="ABC"
  DataDictionary numberOfFields="25"
  ClusteringModel modelName="k-means" functionName="clustering" modelClass="centerBased" numberOfClusters="5"
    MiningSchema
      MiningField name="velocidad" invalidValueTreatment="asIs"
      MiningField name="pm10" invalidValueTreatment="asIs"
    ComparisonMeasure kind="distance"
      squaredEuclidean
      ClusteringField field="velocidad" compareFunction="absDiff"
      ClusteringField field="pm10" compareFunction="absDiff"
    Cluster name="cluster_0" size="10"
      Array n="2" type="real"
        44.9 157.0
    Cluster name="cluster_1" size="10"
      Array n="2" type="real"
        25.8 14.0
    Cluster name="cluster_2" size="3"
      Array n="2" type="real"
        250.0 150.0
    Cluster name="cluster_3" size="9"
      Array n="2" type="real"
        51.44444444444444 85.55555555555556
    Cluster name="cluster_4" size="7"
      Array n="2" type="real"
        651.4285714285714 48.57142857142857
  
```

Figura N° 4.112: Modelo de cluster

- Tabla con colores

Row ID	direccion...	velocid...	velocidad	velocid...	temper...	presion	humedad	S fecha	dia	mes
1	50	60	10	20	50	10	20	01/01/2014	1	1
2	10	10	20	10	10	10	10	01/07/2014	2	7
3	30	60	30	10	20	1020	10	26/02/2014	3	2
4	90	60	40	20	50	5520	50	24/02/2014	4	2
5	20	520	50	520	40	50	60	03/03/2014	5	3
6	50	250	60	30	60	20	50	05/04/2014	6	4
7	960	950	70	420	20	6550	690	09/03/2014	7	3
8	34	67	23	32	50	10	20	01/01/2014	20	1
9	23	86	44	10	10	10	10	01/07/2014	21	7
10	30	60	56	10	20	1020	10	26/02/2014	22	2
11	32	60	34	20	50	5520	50	24/02/2014	23	2
12	20	23	43	520	40	50	60	03/03/2014	24	3
13	50	250	250	30	60	20	50	05/04/2014	25	4
14	960	950	850	420	20	6550	690	09/03/2014	26	3
15	500	500	500	500	500	500	500	04/03/2014	27	3
16	50	60	50	23	50	10	20	01/01/2014	28	1
17	10	10	10	10	10	10	10	01/07/2014	29	7
18	30	60	50	10	20	1020	10	26/02/2014	30	2
19	90	60	50	20	120	5520	50	24/02/2014	1	2
20	20	520	510	23	10	50	60	03/03/2014	2	3
21	50	250	250	30	20	20	50	05/04/2014	3	4
22	960	950	850	420	10	6550	690	09/03/2014	4	3
23	500	500	500	500	20	500	500	04/03/2014	5	3
24	880	80	50	70	40	50	850	09/04/2014	6	4
25	50	120	80	50	40	250	240	08/04/2014	7	4

Figura N° 4.113 Tabla con colores

- Visualización de gráfico de dispersión

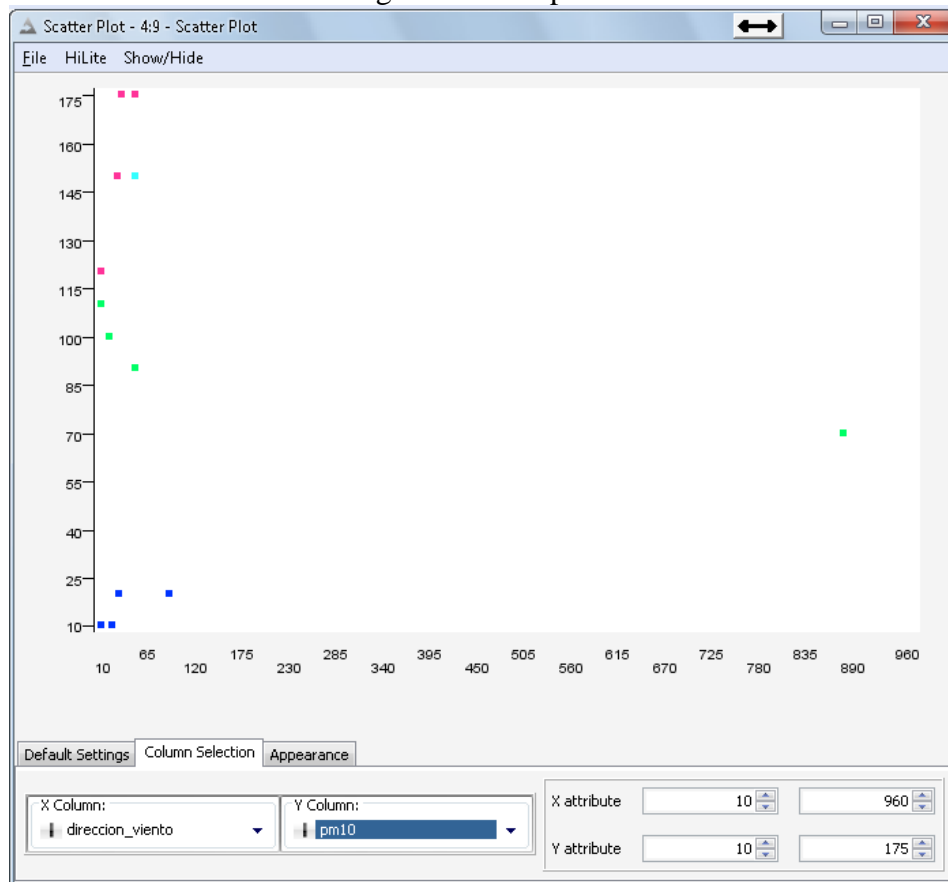


Figura N° 4.114: Visualización gráficos de dispersión

D. Modelo: Cluster weka

- Configuración cluster weka

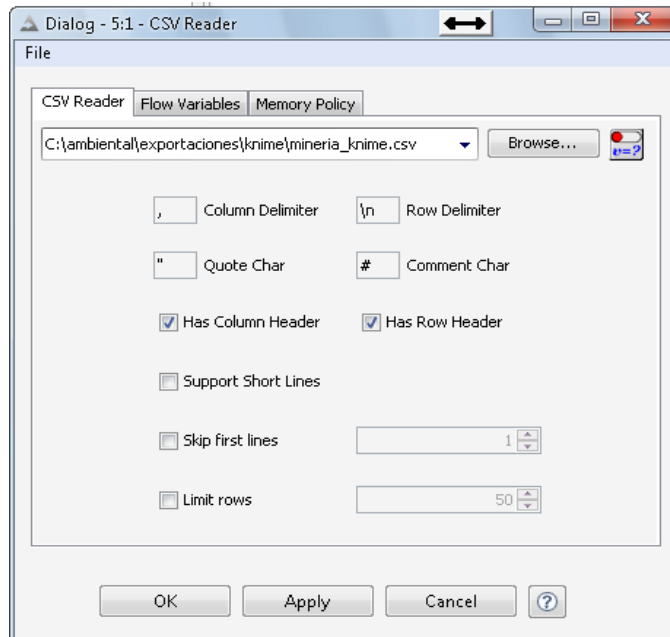


Figura N° 4.115: Configuración cluster weka

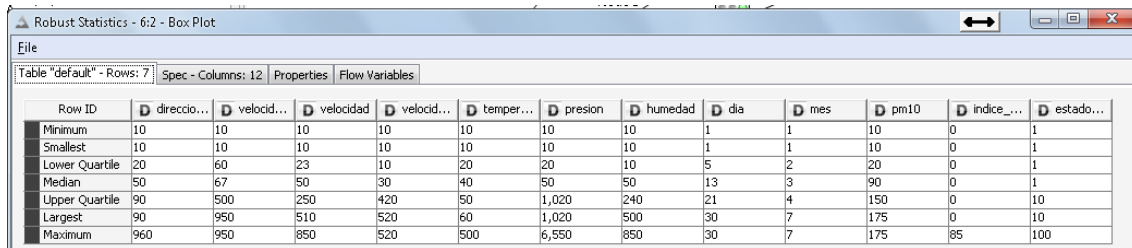
- Resultados cluster

Attribute	Full Data (39)	Cluster#			
		0 (12)	1 (13)	2 (6)	3 (8)
direccion_viento	209.7179	17.75	271.0769	50	517.75
velocidad_predominante	235.5385	34.6667	268.4615	206.6667	505
velocidad	166.1538	27.5	186.6154	161.6667	344.25
velocidad_maxima	140.2051	14.1667	255.2308	36.6667	220
temperatura	59.4615	19.1667	110	46.6667	42.5
presion	1421.7949	395.8333	141.5385	96.6667	6035
humedad	186.9231	10.8333	270.7692	113.3333	370
fecha	01012014	01072014	01012014	05042014	09032014
dia	13.6923	17	13	11.8333	11.25
mes	3.6154	5.3333	2.5385	4	2.5
pm10	83.8462	74.1667	104.6154	130	30
indice_calidad	5.7692	18.75	0	0	0
estado	Bueno	Bueno	Bueno	Malo	Bueno
estado_red	13.4615	2.5	5.1538	70	1
direccion_viento_g	40-60	0-20	40-60	40-60	40-60

Figura N° 4.116: Resultados cluster

E. Modelo: Gráficos de minería de datos

- Procesamiento Box Plot



Row ID	D direccio...	D velocid...	D velocidad	D velocid...	D temper...	D presion	D humedad	D dia	D mes	D pm10	D indice_...	D estado...
Minimum	10	10	10	10	10	10	10	1	1	10	0	1
Smallest	10	10	10	10	10	10	10	1	1	10	0	1
Lower Quartile	20	60	23	10	20	20	10	5	2	20	0	1
Median	50	67	50	30	40	50	50	13	3	90	0	1
Upper Quartile	90	500	250	420	50	1,020	240	21	4	150	0	10
Largest	90	950	510	520	60	1,020	500	30	7	175	0	10
Maximum	960	950	850	520	500	6,550	850	30	7	175	85	100

Figura N° 4.117: Procesamiento Box Plot

- Gráfico Histogram (Intercative)

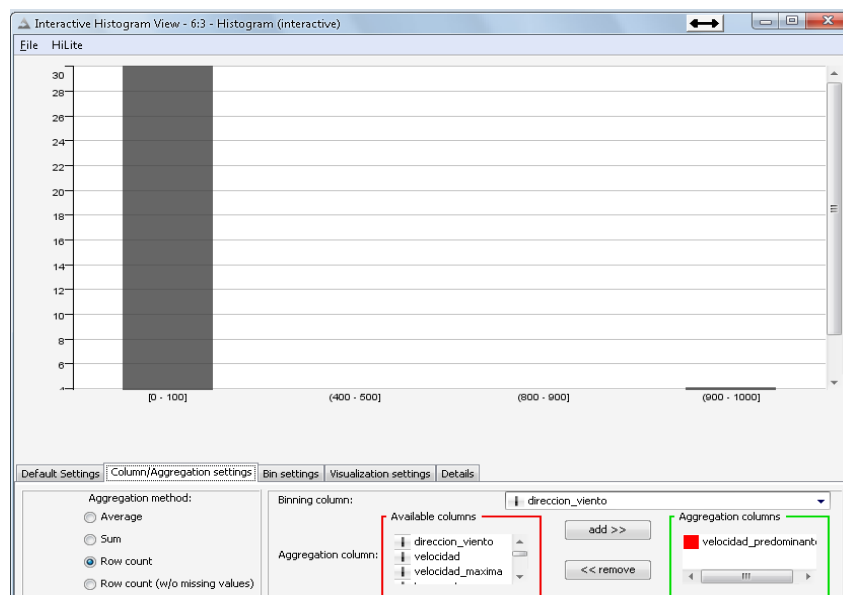


Figura N° 4.118: Gráfico Histograma

- Gráfico Line plot

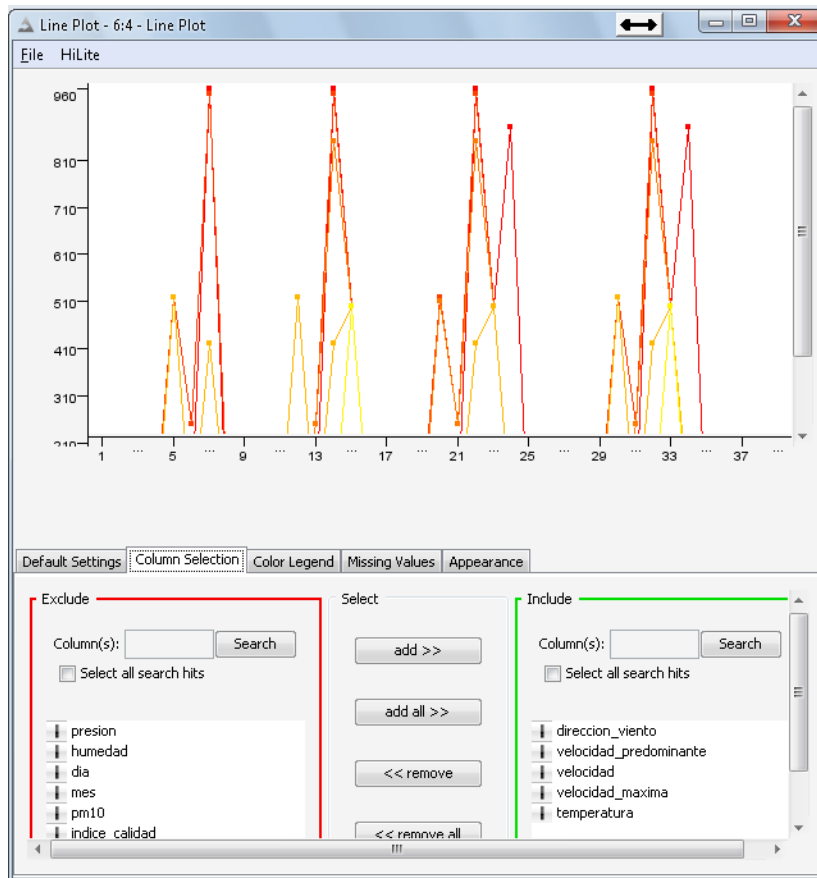


Figura N° 4.119: Gráfico Line Plot

- Gráfico parallel Coordinates

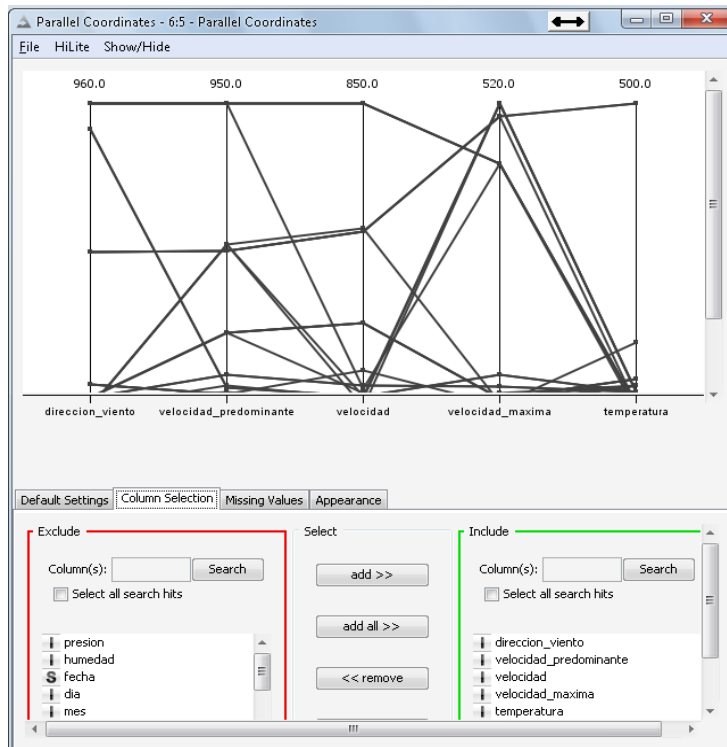


Figura N° 4.120: Gráfico parallel coordinates

- Gráfico Pie Chart

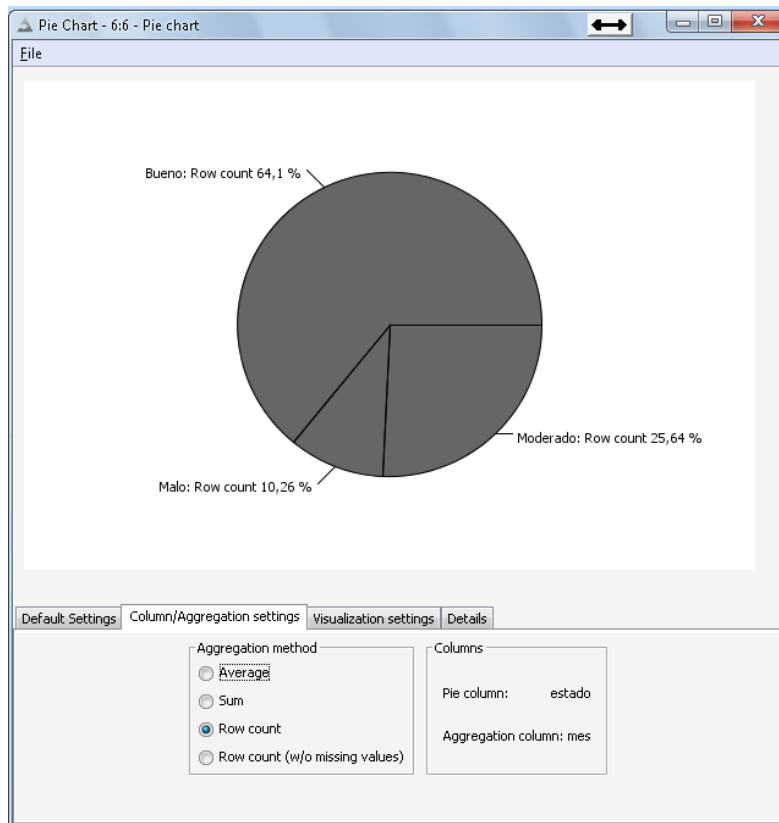


Figura N° 4.121: Gráfico Pie Chart

- Gráfico Scatter Matrix

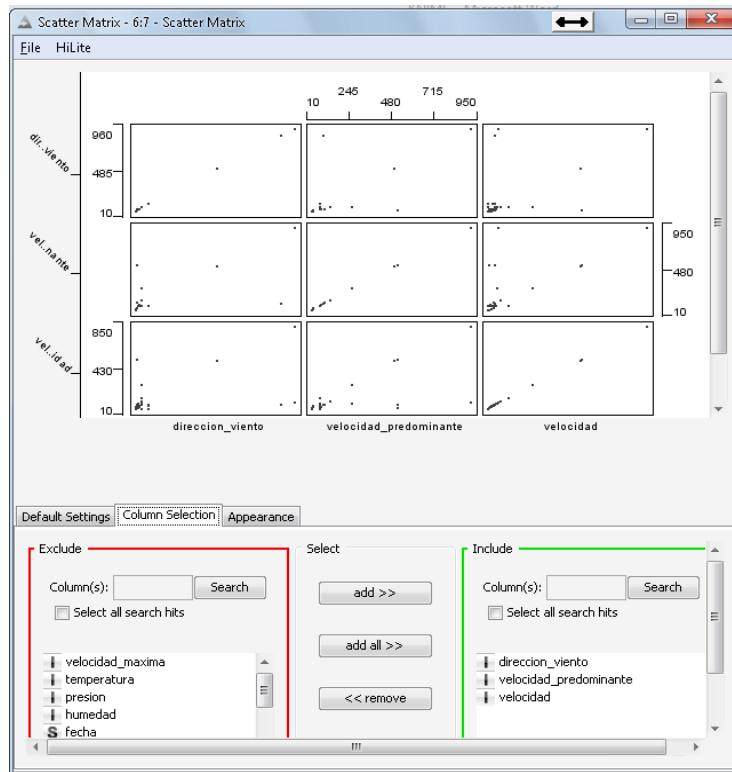


Figura N° 4.122: Gráfico scatter matrix

- Gráfico Scatter Plot

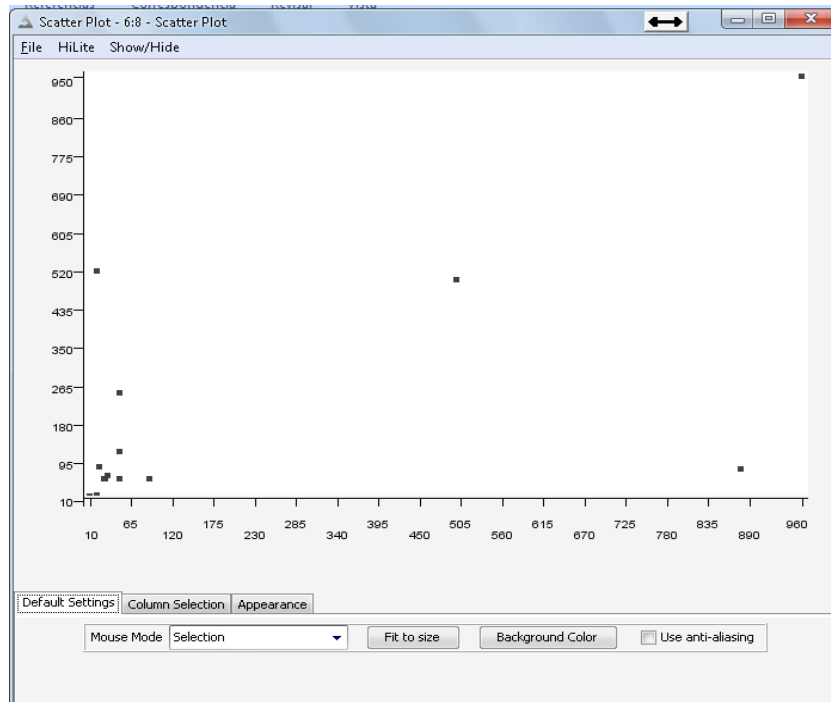


Figura N° 4.123: Gráfico Scatter Plot

- Procesamiento Spark line Appender

Row ID	direccio...	velocid...	velocidad	velocid...	temper...	presion	humedad	S fecha
1	50	60	10	20	50	10	20	01/01/2014
2	10	10	20	10	10	10	10	01/07/2014
3	30	60	30	10	20	1020	10	26/02/2014
4	90	60	40	20	50	5520	50	24/02/2014
5	20	520	50	520	40	50	60	03/03/2014
6	50	250	60	30	60	20	50	05/04/2014
7	960	950	70	420	20	6550	690	09/03/2014
8	34	67	23	32	50	10	20	01/01/2014
9	23	86	44	10	10	10	10	01/07/2014
10	30	60	56	10	20	1020	10	26/02/2014
11	32	60	34	20	50	5520	50	24/02/2014
12	20	23	43	520	40	50	60	03/03/2014
13	50	250	250	30	60	20	50	05/04/2014
14	960	950	850	420	20	6550	690	09/03/2014
15	500	500	500	500	500	500	500	04/03/2014
16	50	60	50	23	50	10	20	01/01/2014
17	10	10	10	10	10	10	10	01/07/2014
18	30	60	50	10	20	1020	10	26/02/2014
19	90	60	50	20	120	5520	50	24/02/2014
20	20	520	510	23	10	50	60	03/03/2014
21	50	250	250	30	60	20	50	05/04/2014

Figura N° 4.124 Procesamiento spark line appender

F. Modelo: Selección de atributos.

- Configuración de Selección de atributos

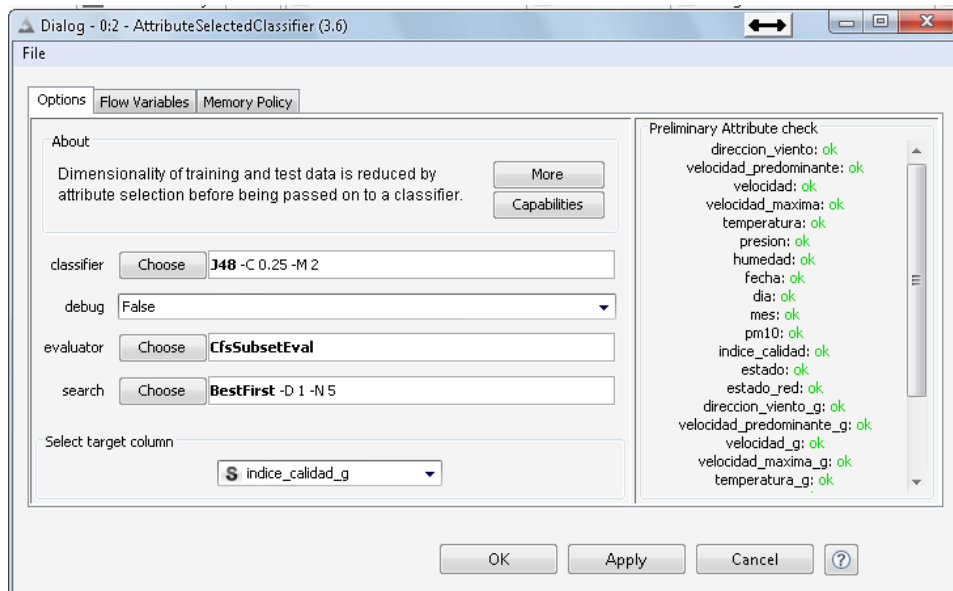
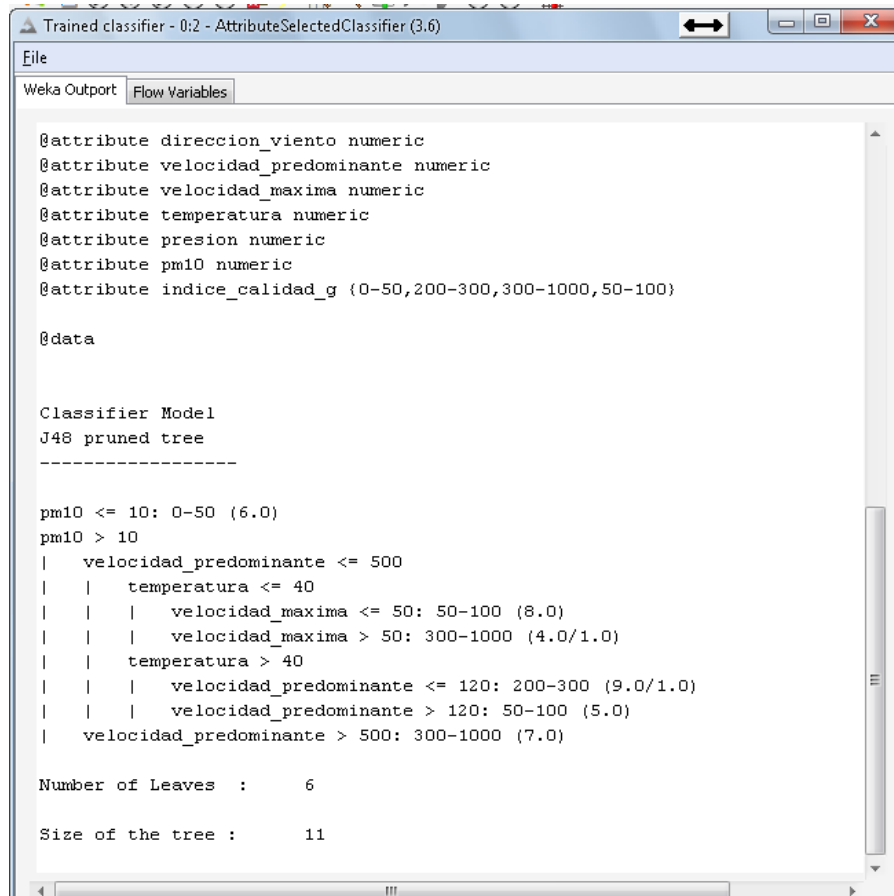


Figura N° 4.125: Selección de atributos

- Resultado de procesamiento



```
Trained classifier - 0:2 - AttributeSelectedClassifier (3.6)
File
Weka Output Flow Variables

@attribute direccion_viento numeric
@attribute velocidad_predominante numeric
@attribute velocidad_maxima numeric
@attribute temperatura numeric
@attribute presion numeric
@attribute pm10 numeric
@attribute indice_calidad_g {0-50,200-300,300-1000,50-100}

@data

Classifier Model
J48 pruned tree
-----

pm10 <= 10: 0-50 (6.0)
pm10 > 10
|  velocidad_predominante <= 500
|  |  temperatura <= 40
|  |  |  velocidad_maxima <= 50: 50-100 (8.0)
|  |  |  velocidad_maxima > 50: 300-1000 (4.0/1.0)
|  |  |  temperatura > 40
|  |  |  velocidad_predominante <= 120: 200-300 (9.0/1.0)
|  |  |  velocidad_predominante > 120: 50-100 (5.0)
|  |  velocidad_predominante > 500: 300-1000 (7.0)

Number of Leaves :      6
Size of the tree :     11
```

Figura N° 4.126: Resultados de procesamiento

CAPÍTULO V: EVALUACIÓN

5.1. DEMOSTRACIÓN DE LA HIPÓTESIS:

Para la demostración de la hipótesis, se realizó la evaluación en función de los indicadores de la variable dependiente (Predicción de la contaminación atmosférica por PM10, en la ciudad de Chimbote) y de la variable independiente (Sistema Integral, bajo el enfoque de minería de datos y redes neuronales); para lo cual en el Capítulo IV, con la metodología RUP, se realizó la demostración de la variable independiente, a través de los indicadores de confiabilidad, seguridad, accesibilidad y flexibilidad del sistema. A continuación se realizará la demostración y validación de la hipótesis en función de los dos principales indicadores, que tienen más relevancia, respecto a la variable dependiente: “satisfacción del usuario” y “tiempo”.

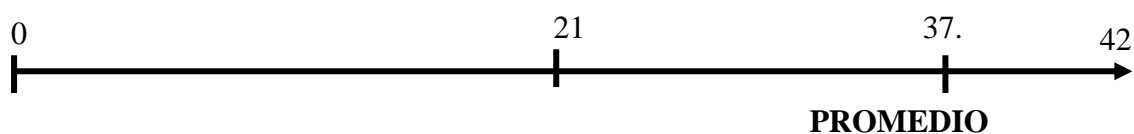
5.1.1. Demostración de la hipótesis en función de usuarios.

Para medir los indicadores de la variable dependiente, se utilizó el cuestionario C (Anexo C), aplicándolo a 05 expertos (usuarios), con conocimientos en contaminación atmosférica y desarrollo de software, se obtiene la tabla 5.1.

Tabla Nº 5.1. Resultados de Cuestionario

Usuarios	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	Puntaje
Usuario 1	3	3	3	2	3	3	3	2	3	2	3	3	3	3	39
Usuario 2	2	3	3	2	2	3	2	2	3	3	3	3	2	3	36
Usuario 3	3	3	3	2	2	3	3	3	3	2	3	3	3	2	38
Usuario 4	3	3	3	2	3	3	2	2	3	3	3	2	3	3	38
Usuario 5	2	3	3	3	2	3	2	2	3	3	3	2	3	3	37
PROMEDIO															37.6

El rango de puntuación se muestra a continuación:



El promedio de las respuestas obtenidas es de 37.6 y nos indica un alto nivel de eficiencia de los procesos y aceptación de los usuarios que manejan el sistema; debemos probar que estadísticamente es significativo a través de la prueba t de student.

Prueba de Hipótesis Estadística:

1. Planteamiento:

Ho: $\mu=21$; El sistema integral, bajo el enfoque de minería de datos y redes neuronales, no permite realizar la predicción de la contaminación atmosférica por PM_{10} en la ciudad de Chimbote.

Ha: $\mu>21$; El sistema integral, bajo el enfoque de minería de datos y redes neuronales, si permite realizar la predicción de la contaminación atmosférica por PM_{10} en la ciudad de Chimbote.

2. Nivel de Confianza (99%)

$\alpha=1\%$ (Margen de Error)

3. Regiones de Aceptación y rechazo

Si: $t_0 \leq t(0.99,5)$ entonces Ho se acepta; para esto $t_0 \leq 3.365$. En caso contrario Ho se rechaza y se acepta Ha, es decir $t_0 > 3.365$.

4. Cálculos:

$$t = \frac{\bar{X} - \mu}{S} \sqrt{N - 1}$$

$$S = \sqrt{\frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N}}$$

Dónde:

$$N=5$$

$$\bar{X} = 37.6$$

$$\mu=3.365$$

Calculamos la desviación Estándar Típica: $S=1.02$, tenemos: $t_0=32.555$

5. Conclusión

Como $t_0 > 3.365$, H_0 se rechaza y se acepta H_a ; por lo tanto el sistema integral, bajo el enfoque de minería de datos y redes neuronales, permite realizar la predicción de la contaminación atmosférica por PM_{10} , en la ciudad de Chimbote; esta afirmación tiene un margen de error del 1%.

La prueba de Hipótesis usando la distribución t de student, ha demostrado que el valor promedio hallado es significativo, con lo cual se ha demostrado que la hipótesis de la presente investigación es válida.

5.1.2. Demostración de la hipótesis en función del tiempo

A) Recolección de datos:

A continuación se presenta la tabla 5.2., con tiempos tomados para los principales procesos, antes de aplicar el sistema integral dinámico, elaborados para distintas pruebas. (Pre prueba: M1).

Tabla N° 5.2. Tiempos de procesos antes de aplicar el sistema

PRUEBAS	PROCESOS (minutos)									TOTAL
	P1	P2	P3	P4	P5	P6	P7	P8	P9	
	PROCESO DE GESTIÓN DE PARÁMETROS ICA	RECOLECCIÓN DE DATA METEOROLÓGICA	PROCESAMIENTO DE DATA METEOROLÓGICA	VALIDACIÓN DE VARIABLES METEOROLÓGICAS	PROCESAMIENTO DE MINERÍA DE DATOS	CONSULTAS Y FILTROS DE INFORMACION	GENERACIÓN DE REPORTE	PROCESAMIENTO DE PREDICCIÓN DEL PM_{10}	RESULTADOS DE PREDICCIÓNES DEL PM_{10}	
Prueba N° 1	8.00	10.00	15.00	50.00	1.80	5.00	5.90	15.00	5.00	367
Prueba N° 2	7.50	12.00	16.00	52.00	2.50	6.20	6.00	20.00	3.00	383
Prueba N° 3	10.00	11.00	20.00	50.00	2.30	4.80	5.00	21.00	4.00	409
Prueba N° 4	9.00	12.30	15.80	30.00	1.90	5.00	7.00	15.60	4.00	382
Prueba N° 5	9.20	12.50	16.00	62.00	2.30	6.00	5.60	14.80	5.00	372
Prueba N° 6	8.00	10.00	14.00	45.00	2.00	7.00	3.00	12.00	3.00	435
Prueba N° 7	7.00	11.50	15.00	60.00	2.00	6.50	4.20	14.00	4.00	438
Prueba N° 8	7.20	12.00	13.50	42.00	2.10	6.20	2.30	12.00	2.50	398
Prueba N° 9	7.60	7.00	10.20	52.00	2.00	4.00	5.20	25.00	5.00	439
Prueba N° 10	5.00	8.00	15.70	42.00	1.80	7.20	4.20	18.00	2.80	415
Promedio	12	66.7	47.7	35.3	64.6	18.4	25.9	113.2	20	403.8

De la tabla 5.2., determinamos la gráfica 5.1., del tiempo que toma la ejecución de los procesos antes de la aplicación del sistema:

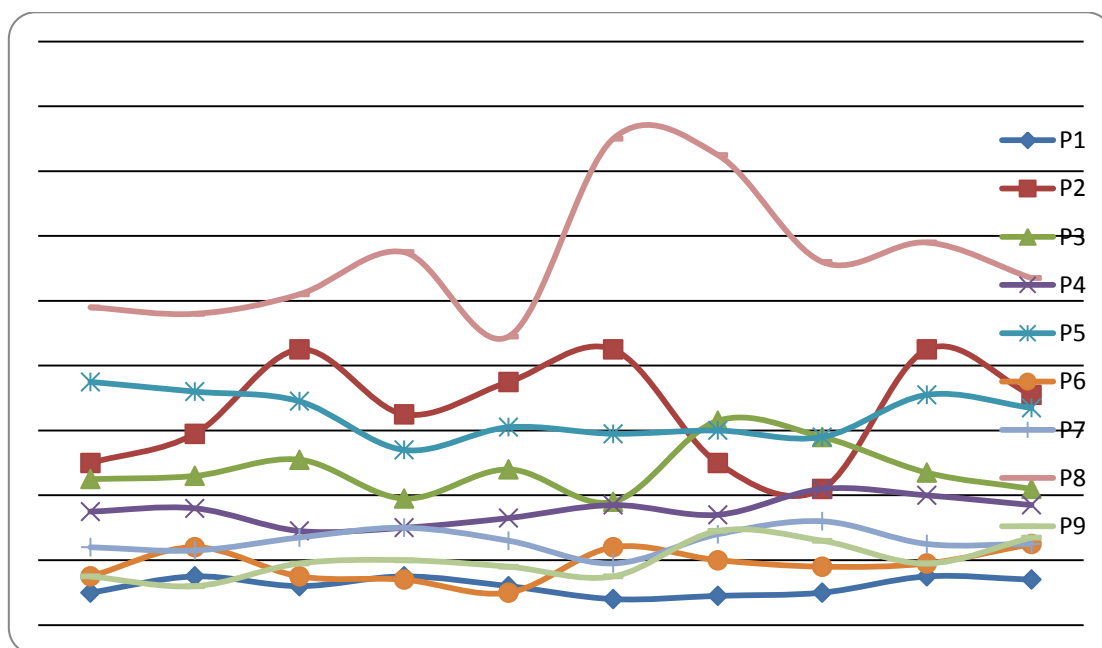


Figura N° 5.1. Gráfica de Tiempos de procesos antes de aplicar el sistema

A continuación se presenta la tabla 5.3., con tiempos tomados para los mismos procesos de la etapa previa, después de aplicar el sistema integral dinámico, elaborados para distintas pruebas. (Post prueba: M2).

Tabla N° 5.3: Tiempos de procesos después de aplicar el sistema

PRUEBAS	PROCESOS (minutos)									TOTAL
	P1	P2	P3	P4	P5	P6	P7	P8	P9	
	PROCESO DE GESTIÓN DE PARÁMETROS SICA	RECOLECCIÓN DE DATA METEOROLÓGICA	PROCESAMIENTO DE DATA METEOROLÓGICA	VALIDACIÓN DE VARIABLES METEOROLÓGICAS	PROCESAMIENTO DE MINERÍA DE DATOS	CONSULTAS Y FILTROS DE INFORMACION	GENERACIÓN DE REPORTES	PROCESAMIENTO DE PREDICCIÓN DEL PM ₁₀	RESULTADOS DE PREDICCIÓNES DEL PM ₁₀	
Prueba N° 1	2	9	8	0.5	15	3	1	8	1.5	48
Prueba N° 2	1	5	7	1	15	2	1	7	2	41
Prueba N° 3	1.8	5	4.5	0.5	17	1	0.9	9	1.8	41.5
Prueba N° 4	1.2	7	3.5	0.9	14	2	0.8	15	0.8	45.2
Prueba N° 5	1.2	4	6	0.8	15	4	1	12	1	45
Prueba N° 6	0.8	12	5.6	0.9	18	2	0.5	8	1.4	49.2
Prueba N° 7	1	8	2.9	1	19	4	1	7	2	45.9
Prueba N° 8	1	9	3.9	1	18	5	1	9	1	48.9
Prueba N° 9	0.8	9	7	1.5	20	3	0.5	10	1.8	53.6
Prueba N° 10	1	8	6.4	1	18	5	0.6	11	2.3	53.3
Promedio	1.18	7.6	5.48	0.91	16.9	3.1	0.83	9.6	1.56	47.16

Datos mostrados en la gráfica 5.2., del tiempo que toma la ejecución de los procesos después de la aplicación del sistema:

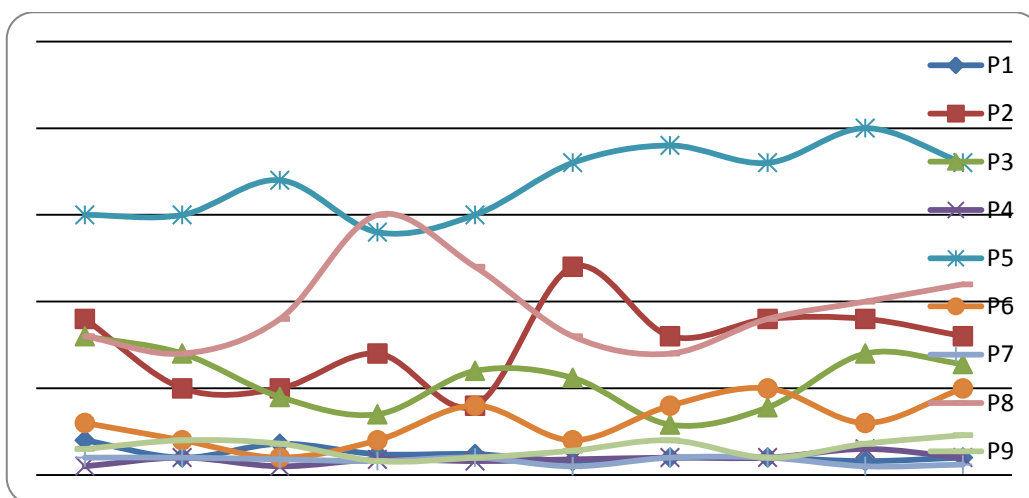


Figura N° 5.2. Gráfica de Tiempos de procesos después de aplicar el sistema

B) Comparación en función de Tiempos

Tabla N° 5.4: Comparación de tiempos

Medición	Medición Previa (M1)	Medición Posterior (M2)
1	367	48
2	383	41
3	409	41.5
4	382	45.2
5	372	45
6	435	49.2
7	438	45.9
8	398	48.9
9	439	53.6
10	415	53.3
Promedio	403.8	47.16
D. Estándar	27.57	4.31
Máximo	439	53.6
Mínimo	367	41

C) Prueba de comparación de varianzas en la medición previa ($\alpha=99\%$)

Tabla N° 5.5. Prueba de comparación de varianzas

Medición		Media y D. Estándar (M1)	Media y D. Estándar (M2)	Nivel de Significancia	Valor Experimental	Resultados de la Prueba
1	n	X1=403.8 S1=27.57	X2=47.16 S2= 4.31	$\alpha=0.01$	Fo=40.92	Significativa

$$n_2=10, \alpha=0.01$$

Valor Tabular (0.01, 9,9)=5.351 (Valor en la tabla de la Distribución

Acumulativa F)

Aplicando la Fórmula:

$$F_o = (\text{Desviación Mayor})^2 / (\text{Desviación Menor})^2$$

$$F_o = (27.57)^2 / (4.31)^2$$

$$F_o = 40.92 \text{ (valor Experimental)}$$

Si F_o (experimental) > Valor Tabular, es significativo ambas varianzas son diferentes.

Interpretación de la Tabla N° 5.5.

La prueba de comparación de varianzas según los datos de la tabla resulta SIGNIFICATIVA con un margen de error del 1%, lo que nos permite concluir que existen diferencias significativas entre las varianzas de ambos grupos.

Tabla N° 5.6. Coeficiente de variación y rango

Tipo de Medición	Media (Min)		D. Estándar		Coeficiente de variación CV=(D.E./Promedio) *100		Rango	
	M2	M1	M2	M1	M2	M1	M2	M1
1	47.16	403.8	4.31	27.57	9.14	6.83	41-53.6	367 - 439

Interpretación de la Tabla N° 5.6.

En los datos de la tabla se observa que los tiempos promedios para el grupo M2 y M1 son 47.16 y 403.8 de los cuales el grupo más homogéneo es el grupo M1 con un CV=6.83%

Tabla N° 5.7. Prueba de comparación de varianzas

Tipo de Medición	Media y D. Estándar (M1)	Media y D. Estándar (M2)	Nivel de Significancia	Valor Experimental	Resultados de la Prueba
1	X1=403.8 S1=27.57	X2=47.16 S2= 4.31	$\alpha=0.01$	Zo=-40.41	Significativa

Pasos:

1. **H₀**= $\mu M1=\mu M2$ (No hay diferencias significativas entre el grupo M1 y M2, por lo tanto se considera promedios iguales)

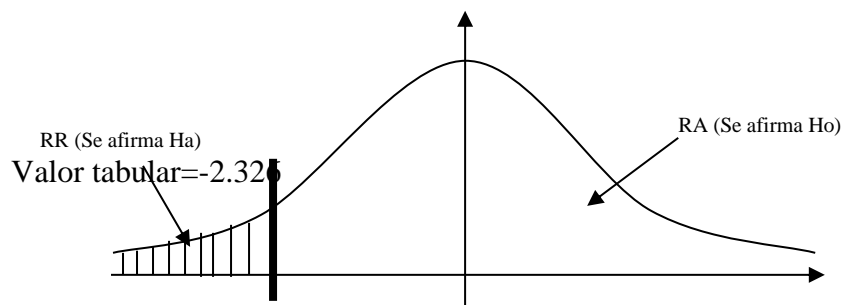
H_a= $\mu M1>\mu M2$ (Existen diferencias significativas entre el grupo M1 y M2, por lo tanto el promedio M2 es menor que al promedio del M1)

2. **Nivel de significancia:** $\alpha=0.01$

3. **Fórmula para Obtener el valor Experimental:**

$$Z_o = (\bar{X}_2 - \bar{X}_1) / \sqrt{S_2^2 / N_2 + S_1^2 / N_1}$$

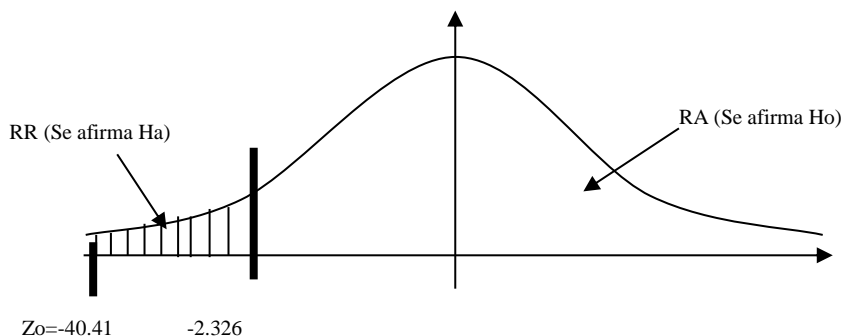
4. **Determinar la RA y RR**



5. Cálculos para hallar Zo (Experimental)

$$Z_o = (47.16 - 403.8) / \sqrt{4.31^2 / 10 + 27.57^2 / 10}$$

$$Z_o = -40.41$$



6. Interpretación:

Según el resultado de la prueba se observa que el tiempo promedio del grupo M2 es menor que el tiempo promedio del Grupo M1; por lo tanto se acepta la hipótesis Alternativa.

5.2. DETERMINACIÓN DEL NIVEL PORCENTUAL:

En el punto anterior, se ha demostrado que la hipótesis de investigación es válida, en función de los dos indicadores (satisfacción del usuario y tiempo). Ahora se procede a determinar la medida porcentual en la que se determina la predicción de la contaminación atmosférica por PM_{10} .

5.2.1. Etapa Previa.:

- **Indicador:** Satisfacción del Usuario

Tabla N° 5.8. Etapa previa – Satisfacción del usuario

Usuarios	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	Puntaje
Usuario 1	1	1	0	2	1	1	2	1	1	0	1	1	0	1	13
Usuario 2	1	1	0	2	0	1	2	1	1	1	1	1	1	0	13
Usuario 3	0	1	0	1	1	1	1	1	1	1	1	1	1	1	12
Usuario 4	0	1	0	1	1	1	1	1	0	2	0	2	0	1	11
Usuario 5	1	1	0	1	0	1	1	1	1	1	1	2	0	0	11
PROMEDIO															12

- **Indicador:** Tiempo

Tabla N° 5.9. Etapa previa – Tiempo

PRUEBAS	PROCESOS (minutos)									TOTAL
	P1	P2	P3	P4	P5	P6	P7	P8	P9	
	PROCESO DE GESTIÓN DE PARÁMETROS	RECOLECCIÓN DE DATA METEOROLÓGICA	PROCESAMIENTO DE DATA METEOROLÓGICA	VALIDACIÓN DE VARIABLES METEOROLÓGICAS	PROCESAMIENTO DE MINERÍA DE DATOS	CONSULTAS Y FILTROS DE INFORMACION	GENERACIÓN DE REPORTES	PROCESAMIENTO DE PREDICCIÓN DEL PM ₁₀	RESULTADOS DE PREDICCIÓNES DEL PM ₁₀	
Prueba N° 1	8.00	10.00	15.00	50.00	1.80	5.00	5.90	15.00	5.00	367
Prueba N° 2	7.50	12.00	16.00	52.00	2.50	6.20	6.00	20.00	3.00	383
Prueba N° 3	10.00	11.00	20.00	50.00	2.30	4.80	5.00	21.00	4.00	409
Prueba N° 4	9.00	12.30	15.80	30.00	1.90	5.00	7.00	15.60	4.00	382
Prueba N° 5	9.20	12.50	16.00	62.00	2.30	6.00	5.60	14.80	5.00	372
Prueba N° 6	8.00	10.00	14.00	45.00	2.00	7.00	3.00	12.00	3.00	435
Prueba N° 7	7.00	11.50	15.00	60.00	2.00	6.50	4.20	14.00	4.00	438
Prueba N° 8	7.20	12.00	13.50	42.00	2.10	6.20	2.30	12.00	2.50	398
Prueba N° 9	7.60	7.00	10.20	52.00	2.00	4.00	5.20	25.00	5.00	439
Prueba N° 10	5.00	8.00	15.70	42.00	1.80	7.20	4.20	18.00	2.80	415
Promedio	12	66.7	47.7	35.3	64.6	18.4	25.9	113.2	20	403.8

5.2.2. Etapa posterior:

- **Indicador:** Satisfacción del Usuario

Tabla N° 5.10. Etapa Posterior – Satisfacción del usuario

Usuarios	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	Puntaje
Usuario 1	3	3	3	2	3	3	3	2	3	2	3	3	3	3	39
Usuario 2	2	3	3	2	2	3	2	2	3	3	3	3	2	3	36
Usuario 3	3	3	3	2	2	3	3	3	3	2	3	3	3	2	38
Usuario 4	3	3	3	2	3	3	2	2	3	3	3	2	3	3	38
Usuario 5	2	3	3	3	2	3	2	2	3	3	3	2	3	3	37
PROMEDIO															37.6

- **Indicador:** Tiempo

Tabla N° 5.11. Etapa Posterior – Tiempo

PRUEBAS	PROCESOS (minutos)									TOTAL
	P1	P2	P3	P4	P5	P6	P7	P8	P9	
	PROCESO DE GESTIÓN DE PARÁMETROSICA	RECOLECCIÓN DE DATA METEOROLÓGICA	PROCESAMIENTO DE DATA METEOROLÓGICA	VALIDACIÓN DE VARIABLES METEOROLÓGICAS	PROCESAMIENTO DE MINERÍA DE DATOS	CONSULTAS Y FILTROS DE INFORMACION	GENERACIÓN DE REPORTES	PROCESAMIENTO DE PREDICCIÓN DEL PM ₁₀	RESULTADOS DE PREDICCIÓNES DEL PM ₁₀	
Prueba N° 1	2	9	8	0.5	15	3	1	8	1.5	48
Prueba N° 2	1	5	7	1	15	2	1	7	2	41
Prueba N° 3	1.8	5	4.5	0.5	17	1	0.9	9	1.8	41.5
Prueba N° 4	1.2	7	3.5	0.9	14	2	0.8	15	0.8	45.2
Prueba N° 5	1.2	4	6	0.8	15	4	1	12	1	45
Prueba N° 6	0.8	12	5.6	0.9	18	2	0.5	8	1.4	49.2
Prueba N° 7	1	8	2.9	1	19	4	1	7	2	45.9
Prueba N° 8	1	9	3.9	1	18	5	1	9	1	48.9
Prueba N° 9	0.8	9	7	1.5	20	3	0.5	10	1.8	53.6
Prueba N° 10	1	8	6.4	1	18	5	0.6	11	2.3	53.3
Promedio	1.18	7.6	5.48	0.91	16.9	3.1	0.83	9.6	1.56	47.16

5.2.3. Comparación de Valores:

- Indicador: Satisfacción del usuario

Valor Previo : 12 puntos

Valor posterior : 37.6 puntos

Valor Optimo : 42 puntos (valor máximo)

- Indicador: Tiempo

Valor Previo: 403.8 min.

Valor posterior: 47.16 min.

Valor Optimo: 45 min. (Tiempo óptimo en la realización de los procesos de estudio)

Tabla N° 5.12. Comparación de Ambas etapas

Indicador	Etapa Previa		Etapa posterior		Optimo	
	Valor	Porcentaje	Valor	Porcentaje	Valor	Porcentaje
Tiempo (Peso 1)	403.8	11.14%	47.16	95.42%	45	100%
Satisfacción de usuario (Peso 3)	12	28.57%	37.6	89.52%	42	100%
Promedio		24.21%		91.00%		

Interpretación de tabla N° 5.12:

En base a los datos obtenidos para ambas variables, se determina los porcentajes, promediados, dándole un ponderado de 3 a la satisfacción del usuario y 1 para el tiempo; esto debido a que se da mayor prioridad a la satisfacción del usuario, debido a que ésta tesis se ve motivada por la necesidad de advertir a la población con anticipación del posible riesgo que corre y proteger a los sectores más vulnerables de la exposición a altas concentraciones de contaminantes; así mismo el nivel de predicción logrado después de la aplicación del sistema es de 91.00%.

5.3. DISCUSIÓN DE RESULTADOS

La problemática que este trabajo de investigación pretende solucionar es la siguiente: ¿En qué medida la implementación de un sistema integral, bajo el enfoque de minería de datos y redes neuronales, permitirá realizar la predicción de la contaminación atmosférica por PM10, en la ciudad de Chimbote?, para dar solución al problema se planteó la siguiente hipótesis: “La implementación de un sistema integral, bajo el enfoque de minería de datos y redes neuronales, permitirá realizar la predicción de la contaminación atmosférica por PM10, en la ciudad de Chimbote, en un 91%, a través de información clara y transparente”.

Para demostrar la hipótesis los indicadores de la variable independiente: confiabilidad, seguridad, accesibilidad y flexibilidad; fueron demostrados a través de la Metodología RUP, en la fase de construcción, a través de las pruebas del sistema, y haciendo uso del cuestionario del Anexo A.

Los indicadores de la variable dependiente: tiempo y satisfacción del usuario se validaron con el cuestionario C (ver anexo C), luego se determinó el nivel porcentual de predicción del agente contaminante PM₁₀, obteniendo el resultado tal como lo muestra el gráfico 5.3.

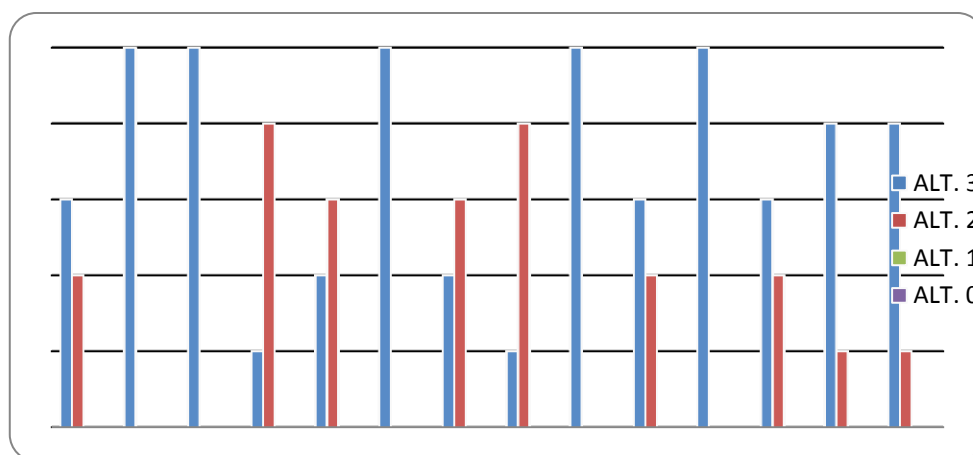


Figura N° 5.3. Resultado de cuestionario

En todas las preguntas se obtuvo respuestas de aceptación significativa, distribuidas entre las alternativas “3” y “2” y ninguno respondió las alternativas “1” y “0” que son los niveles más bajos de aceptación (en realidad se interpretan como niveles de rechazo).

Para lograr un nivel de aceptación en los resultados generales de la encuesta, la puntuación general oscila entre un mínimo de 0 y un máximo de 15, obteniéndose los siguientes intervalos:

- [0-14] : No hay Aceptación**
- [15-28] : Aceptación regular**
- [29-42] : Alta Aceptación**

El resultado obtenido es 37.6, el cual se encuentra en un rango de valores de alta aceptación.

De la aplicación del cuestionario se puede concluir lo siguiente:

- ✓ **Pregunta 1:** Se demuestra que el trabajo del personal en conjunto para consultar y verificar los parámetros del Índice de Calidad del Aire ha disminuido notablemente.
- ✓ **Pregunta 2:** De igual manera el sistema demuestra que el trabajo del personal en conjunto para registrar y/o almacenar la información y data meteorológica ha disminuido notablemente, esto se afirma en base a que todos los usuarios opinaron por esta opción.
- ✓ **Pregunta 3:** Todos los usuarios opinan que la disponibilidad de la información meteorológica es instantánea, se debe a que ahora, a diferencia de su forma anterior de trabajar, las obtienen en el instante y en el lugar en que lo requieran por trabajarse también vía web.
- ✓ **Pregunta 4:** Un usuario opina que la cantidad de trabajo se ha reducido notablemente respecto a la validación de las variables tanto estacionales como meteorológicas; frente a 04 usuarios que observan que dicho trabajo ha disminuido.
- ✓ **Pregunta 5:** Los usuarios que opinan que la velocidad con que se exporta e importa ahora la data meteorológica es muy rápida se debe a que ahora, a diferencia de su forma anterior de trabajar, las generan al instante, tan sólo con haber ingresado la data diaria; el resto de usuario que opinan que solo es Rápida, porque no lo compara con la forma anterior puesto que no participan directamente en la recolección de la información.

- ✓ **Pregunta 6:** Todos los usuarios respondieron que la disponibilidad de las predicciones del agente contaminante PM_{10} es al instante, debido a que su experiencia sin el uso del sistema era muy laborioso, tenían que hacer uso de fórmulas matemáticas; pero ahora con la utilidad que brinda el sistema observan las cuentas en el momento que quieran y en segundos.
- ✓ **Pregunta 7:** La diferencia de las respuestas se deben a una cuestión de apreciación, ya que dos usuarios califica con 3 puntos compara el trabajo anterior con el actual, el cual ha reducido notablemente, debido a que la realización de los reportes ahora se realiza en forma automática en pocos segundos, en cambio los demás usuarios no observan este hecho y sólo se dan cuenta que dicho proceso ha disminuido.
- ✓ **Pregunta 8:** Todos los usuarios opinaron que la velocidad con la que se genera el mapa geográfico para la consultas de las predicciones de las estaciones de trabajo, es muy rápida.
- ✓ **Pregunta 9:** Aquí también la observación en la práctica les demuestra a los usuarios que el tiempo empleado para visualizar las predicciones del agente contaminante PM_{10} es cuestión de segundos, lo cual dicho tiempo se ha reducido notablemente.
- ✓ **Pregunta 10:** La mayoría de usuarios indicaron que la disponibilidad de los reportes es al instante, esto debido a que una vez que se cuenta la data registrada, el sistema procesará automáticamente los reportes que el usuario elija, tanto a nivel del sistema de Power Builder, de Matlab, como el Web.
- ✓ **Pregunta 11:** Todos los usuarios opinan que la velocidad con que se generan los reportes es muy rápida, esto debido a que todos los usuarios en sus diversos módulos; pueden generar varios tipo de reportes que son de utilidad para estos.

- ✓ **Pregunta 12:** El sistema nos demuestra que el tiempo de recuperación de los diversos datos es al instante.
- ✓ **Pregunta 13:** Se puede verificar que el tiempo utilizado en la realización de informes sobre los procesos meteorológicos, ha reducido notablemente, esto debido a la facilidad que tiene el sistema de exportar los diversos datos. Así como también para el procesamiento de reportes, filtros y/o consultas.
- ✓ **Pregunta 14:** Los usuarios que opinan que el trabajo para realizar la minería de datos ha reducido notablemente; el resto de usuario opinan que solo ha disminuido, porque no lo compara con la forma anterior puesto que muchos no realizan este proceso, por no contar con conocimiento en el tema de minería de datos.

Con las demostraciones realizadas a nivel de satisfacción de usuario y tiempo, los resultados caen en la región de aceptación; por lo tanto podemos decir que la hipótesis “La implementación de un sistema integral, bajo el enfoque de minería de datos y redes neuronales, permite realizar la predicción de la contaminación atmosférica por PM10, en la ciudad de Chimbote, en un 91%, a través de información clara y transparente”, esta es válida, demostrando un nivel porcentual de predicción del agente contaminante PM₁₀, en un 91.00%, después de la aplicación del sistema integral y redes neuronales, en la ciudad de Chimbote. Por lo tanto la hipótesis queda demostrada.

CAPÍTULO VI: CONCLUSIONES Y SUGERENCIAS

6.1. CONCLUSIONES

1. Se elaboró e implementó el sistema integral, bajo el enfoque de minería de datos y redes neuronales, haciendo uso del Perceptrón Multicapa, para realizar la predicción de la contaminación atmosférica, por PM_{10} , evaluando de esta manera la calidad del aire, en la ciudad de Chimbote, el cual permite anticiparse a situaciones de riesgo.
2. Se identificaron los requerimientos y los procesos del negocio sirviendo de base para el desarrollo del proyecto.
3. Con base en el trabajo realizado se justificó y validó, el uso de las variables meteorológicas: Dirección predominante del viento, Velocidad del viento predominante, Velocidad promedio Velocidad máxima, Temperatura ambiente promedio, Presión promedio, Humedad promedio y variables estacionales: día y mes del año, como variables que intervienen en la concentración del contaminante: PM_{10} , de la zona de estudio.
4. Se implementó y validó la red neuronal (Perceptrón multicapa), con una capa de entrada de 09 neuronas, 02 capas ocultas y la capa de salida de 05 neuronas; la primera capa oculta tenía 90 neuronas y la segunda 50 neuronas; cuya función de activación fue sigmoideal logarítmica, el cual fue entrenada por medio del método gradiente descendiente con velocidad de aprendizaje variable y momento. Con el modelo propuesto, las estaciones de monitoreo, tienen la capacidad para realizar predicciones de calidad del aire, las cuales fueron implementados en el sistema y permite realizar la predicción en tiempo real y analizar el impacto en la salud de la población.

5. Se aplicó el Proceso Unificado de Rational (RUP) para determinar los principales procesos del negocio, requisitos funcionales, análisis y diseño del sistema integral. Utilizando la Metodología RUP y el Lenguaje de Modelado Unificado como base para la representación de los procesos, realizando los siguientes modelos y diagramas: Modelo de Negocio, Modelo de Dominio, Modelo Conceptual, Diagrama de Casos de Uso, Diagrama de Paquetes, Diagrama de Clases, Diagrama de Secuencias, Diagrama de Colaboración.
6. Se implementó en el Sistema Integral 04 subsistemas: Sistema de Registro y Validación de Data, el cual fue desarrollado en Power Builder; Sistema de Procesamiento de la Red Neuronal, desarrollado en Matlab; Sistema Web de Visualización de Predicciones, que fue desarrollado en Php y el Sistema de Procesamiento de Minería de Datos, para lo cual se hace uso del software Knime.
7. El sistema implementado cumple con la calidad en el software en cuanto a su funcionalidad, facilidad de uso, confiabilidad, desempeño y seguridad.
8. Se demostró a través de pruebas estadísticas que el Sistema Integral realmente cumple el objetivo de predecir los Índices de Calidad del Aire, respecto a la contaminación atmosférica por PM10.

6.2. SUGERENCIAS

1. Se debe de realizar una capacitación a todos los usuarios involucrados en el uso del sistema, para explicarles las funcionalidades y poder levantar todas sus inquietudes.
2. Se debe efectuar la capacitación constante a los nuevos usuarios en el manejo del sistema y en el ingreso o recojo de los datos, ya que de esto depende la

calidad de la información contenida y del entrenamiento realizado por la red neuronal.

3. Es preciso mantener reuniones constantes con los usuarios del sistema, para analizar nuevos requerimientos.
4. Realizar planes de contingencia ante pérdidas de información, realizando periódicamente copias de seguridad, dado el volumen de información que se maneja.
5. Divulgar los resultados obtenidos en la presente investigación entre toda la comunidad de tal manera que sirva como un aporte científico-tecnológico.

6.3. FUTURAS LÍNEAS DE INVESTIGACIÓN

A continuación se presentan las principales líneas de investigación que han quedado abiertas en esta tesis:

- El crecimiento de la población ha traído como consecuencia el aumento de vehículos en la ciudad de Chimbote, debido a esto, se considera en un futuro la incorporación de factores de emisión (CO₂) de los diferentes vehículos, información de flujo vehicular en los modelos de predicción.
- Chimbote es una ciudad industrial cuenta con la industria pesquera y la industria Siderúrgica, es conveniente ampliar la investigación para el agente contaminante del SO₂ y crear un nuevo modelo de predicción y control utilizando Redes Neuronales.
- Chimbote cuenta con la industria pesquera, es uno de los principales puertos pesqueros del mundo, en su litoral se encuentran en promedio 53 fábricas pesqueras. Para la producción de harina de pescado y conservas, utilizan calderos y como combustible petróleo los diferentes gases que emanan de la producción son expulsados por chimineas, es conveniente realizar una investigación para determinar si éstas tienen la altura apropiada para la dispersión de los contaminantes.

- Comparar los resultados de predicción del método propuesto con otro método de computación no convencional como los algoritmos de correspondencia de cadenas, para la predicción de la contaminación atmosférica.

6.4. ARTÍCULOS PUBLICADOS Y EXPOSICIONES.

[Díaz.2010] Díaz Tello, S. Modelo de predicción de la contaminación atmosférica por partículas en suspensión menores de 10 micras y del SO₂ bajo el enfoque de minería de datos y redes neuronales en la ciudad de Chimbote. Publicado en la Revista Científica de la Escuela de Post Grado. Vol. 1 N° 1 . Primera Edición 2011. Universidad Nacional del Santa.

[Díaz.2012] Díaz Tello Sixto. Modelo de Predicción bajo el enfoque de Minería de Datos y Redes Neuronales para el PM₁₀ en la Ciudad de Chimbote. Seminario de Ingeniería de Sistemas e Informática. Escuela de Ingeniería de Sistemas e Informática. UNS. 2012.

[Díaz.2013] Díaz Tello Sixto. Aplicación de la Minería de Datos para Solucionar el Problema de la Contaminación Atmosférica en Chimbote. Colegio de Ingenieros del Perú-Filial Chimbote. 2013.

REFERENCIAS BIBLIOGRÁFICAS

- [Agirre, et.al.2006] Agirre, E., Ibarra, G., Madariaga, I., *Regresión and multilayer perceptron based models to forecast hourly O3 and NO2 levels in the Bilbao area*. Environmental Modelling and Software, 2006.
- [Agirre, et al. 2006] Agirre B., Álvaro A., Barrón L. y Exeberria M. Relevancia de las variables meteorológicas en el diseño de un modelo de predicción de los niveles de ozono, en tiempo real, basado en el uso de redes neuronales. Universidad del País Vasco. Escuela Universitaria de Ingeniería Técnica Industrial. 2006.
- [Araujo, et.al.2005] Araujo-Andrade C., Campos-Cantón I., Martínez J. R., and Ruiz F. Modelo de predicción basado en el análisis multivariante para la determinación de concentración de azúcares en solución. Revista Mexicana de Física, 2005. pág. 63–73.
- [Ballester.2005] Ballester F. Contaminación atmosférica, cambio climático y salud. Unidad de Epidemiología y Estadística-Revista Española de Salud, 2005. pág. 159–175.
- [Barai, et.al.2009] Barai S. V., Gupta A. K., and Kodali J. Air quality forecaster: Moving window based neuro models. In Applications of Soft Computing, Advances in Intelligent and Soft Computing, 2009. pages 137–146. Springer Berlin / Heidelberg.
- [Barchini.1989] Barchini, Graciela E. (1989). *Teorías Presupuestas de la Informática*. Artículo realizado para el "Proyecto de Reformulación Académica de las carreras de Computación de la Universidad Católica de Santiago del Estero", 1989. Chile.

- [Barreras.1999] Barrera, M. *El intelectual y los modelos epistémicos*. Caracas: Sypal, 1999.
- [Barrón.2010] Barrón Adame J.M. (2010). Modelado de un sistema de supervisión de la calidad del aire usando técnicas de fusión de sensores y redes neuronales. PhD. Thesis, Universidad Politécnica de Madrid. España.
- [Bernstein, et.al.2004] Bernstein J. A., Alexis N.and Barnes C., Bernstein I.L., Nel A., Peden D., Diaz-Sanchez D., Tarlo S.M., and Williams P.B. (2004). Health effects of air pollution. *Journal of Allergy and Clinical Immunology*. Pág. 116 – 123.
- [Britos, et al. 2008] Britos, P.; Dieste, O.; García Martínez, R. 2008. Requirements Elicitation in Data Mining for Business intelligence Projects. En: *Advances in Information Systems Research, Education and Practice*. Springer, p. 139–150.
- [Brunet G.2005] The first hundred years of numerical weather prediction. In *Proceedings of the 19th International Symposium on High Performance Computing Systems and Applications, 2005* pages 276–279, Washington, DC, USA. IEEE Computer Society.
- [Burk. 2001] Burk, Ignacio. *Filosofía: Una introducción actualizada*. Venezuela.2001
- [Cabrera. 2012] Cabrera Álvarez, Erick. *Aplicación de cómputo no convencional en la predicción de contaminantes ambientales*. Tesis de Maestro. Instituto Politécnico Nacional. México, 2012.

- [Camaño.2004] Camaño J.L. Sistema Integrado para la Modelización y el Análisis de la Calidad del Aire en Modo Operacional. Tesis Doctoral, Universidad Politécnica de Madrid, 2004.
- [Coman, et.al.2008] Coman A., Ionescu A., and Candau Y. Hourly ozone prediction for a 24- h horizon using neural networks. Environmental Modelling and Software, 2008. pág. 1407–1421.
- [Cortina.2012] Cortina Januchs M. Aplicación de Técnicas de Inteligencia Artificial a la Predicción de Contaminantes Atmosféricos. Tesis Doctoral. Universidad Politécnica de Madrid, 2012.
- [Castellanos. 2013] Castellanos Pañuela, A. Algoritmos para minería de datos con redes neuronales. Tesis Doctoral. Facultad de Informática (UPM). 2013.
- [CIEMET.2014] Centro de Investigaciones Energéticas Medioambientales y Tecnológicas. Modelización de la contaminación atmosférica. <http://www.ciemat.es> (Último acceso: Abril 2014).
- [Chu and Armstrong 2005] Chu P. C. and Armstrong, A. E. Instantaneous current prediction for naval operations. In OCEANS. Proceedings of MTS/IEEE, volumen 1, 2005. pages 884–890.
- [OMS.2010] Departamento de Protección del Medio Humano. Guías para la Calidad del Aire. Organización Mundial de la Salud, Ginebra, 2010.
- [De la Paz Cortez. 2013] De la Paz Cortés Araujo J. Evaluación de la contaminación atmosférica en el aire ambiente de Manizales por COPs y

PM10. Tesis de Maestría. Universidad Nacional de Colombia.

- [Díaz-Robles, et al. 2008] Díaz-Robles, L. et al. A Hybrid ARIMA and Artificial Neural Networks Model to Forecast Particulate Matter in Urban Areas: The case of Temuco, Chile. Atmospheric Environment, 2008.
- [DIGESA.2010] Dirección General de Salud Ambiental. Evaluación de la Calidad del Aire en Chimbote. Informe 2008. Lima, 2010.
- [DIGESA.2012] Dirección General de Salud Ambiental. Evaluación de la Calidad del Aire en Chimbote. Informe N° 1062-2012.Lima.2012.
- [DGP.2014] Dirección General Parlamentaria. Oficina de la Gestión de la Información y Estadística. Carpeta Georeferencial Ancash. Perú.2014.
- [EPA.2011] Environmental Protection Agency EPA. A Guide to Air Quality and Your Health, 20 de enero 2011.
- [García Martínez, 2007] García-Martínez, R., Britos, P., Diez, E., Ochoa, M. Merlino, H., Fernández, E., Rancán, C. Minería de datos en inteligencia de negocios. Centro de Ingeniería del Software e Ingeniería del Conocimiento Escuela de Postgrado. Instituto Tecnológico de Buenos Aires, 2007.
- [Grajales. 2002] Grajales Guerra, T. La metodología de la investigación histórica: una crisis compartida, por Tevni Grajales Guerra Revista: Enfoques 2002.

- [Grivas and Choulakou 2006] Grivas G. and Chaloulakou A. Artificial neural network models for prediction of P M10 hourly concentrations, in the greater area of Athens, Greece. Atmospheric Environment, 2006. pages 1216–1229,
- [Ellwood, et.al.2005] Ellwood P., Asher M.I., Beasley R., and Clayton T.O. and Stewart A.W. The international study of asthma and allergies in childhood (ISAAC): phase three rationale and methods. The International Journal of tuberculosis and lung disease, 2005. Pages 10–16.
- [Hamilton and Crabbe 2009] Hamilton R., and Crabbe H. Environment, pollution and effects. The Effects of air pollution on cultural heritage, 2009. .pages 1–27.
- [Hanumant y Prasada.2013] Hanumanth Sastry. S y Prasada Babu, M. Implementation of CRISP Methodology for ERP System. International Journal of Computer Science Engineering (IJCSE). 2013
- [Harri, et.al.2006] Harri S., Wolfson O., Mena E., Illarramendi A., and Rische N. (2006). Processing of data streams with prediction functions. Proceedings of the 39th Hawaii International Conference in Systems Sciences, 2006.
- [INE. 2011] Instituto Nacional de Ecología (INE). Dirección General de Investigación sobre la Contaminación Urbana Regional. Investigación sobre la calidad del aire. 2011.
- [INEI. 2008] Instituto Nacional de Estadística e Informática. Censos Nacionales 2007. XI de Población y VI de Vivienda. Informe 2008. Lima. 2008

- [Itzama y Yañez 2011] Ytzama López, Y. y Yañez Marquez, C. Predicción de la concentración de contaminantes atmosféricos basados en un clasificador asociativo de patrones. Revista de la Facultad de Ingeniería de la Universidad de Antioquia. 2011.
- [Jiménez, et.al.2006] Jiménez Guerrero J.F., Gázquez Abad J.C., and Sánchez Fernández R. (2006). La capacidad predictiva en los métodos box-jenkins y holt-winters: Una aplicación al sector turístico. Revista Europea de Dirección y economía de la empresa, pages 185–198.
- [King's.2011] King's College London. London air quality network. 2011 <http://www.londonair.org.uk/London/asp/default.asp>. (Último acceso: marzo 2011).
- [Kurt and Oktay 2010] Kurt A. and Oktay A. B. Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks. Expert Systems with Applications, 2010.
- [Londoño-Linares.2007] Londoño Linares Juan Pablo. Modelización de problemas ambientales en entornos urbanos utilizando sistemas de información geográfica y métodos multivariantes. Revista Internacional de sostenibilidad, tecnología y humanismo. 2007.
- [Manahan.2007] Manahan S. E. Introducción a la Química Ambiental. Editorial Reverté, 2007.
- [Manzur, et al. 2014] Manzur E., Benzal G., y Gonzáles S. Modelo de dispersión de contaminantes atmosféricos. Séptimo congreso de medio ambiente, AUGM . Mayo 2012 UNLP. La Plata Argentina.

- [MINAM.2014] Ministerio del Medio Ambiente. Compendio de la Legislación Ambiental Peruana. Volumen V. Calidad Ambiental. 2013.
- [Rodríguez y Domínguez 2011] Rodríguez Méndez, R. y Domínguez García M. Cambio climático, turismo y políticas reguladoras. Revista de Análisis Turístico. 2011. Asociación Española de Expertos Científicos en turismo (AECIT).
- [Mesin, et.al.2010] Mesin L., Orione F., Taormina R., and Pasero E. A feature selection method for air quality forecasting. In Artificial Neural Networks ICANN 2010, volume 6354, pages 489–494. Springer Berlin / Heidelberg, 2010.
- [Molina.2014] Molina Jiménez, Pedro. Simulación Numérica de un problema de contaminación numérica. Tesis Doctoral. Universidad Politécnica de Madrid. España, 2014.
- [Moses.1968] Moses, H., Anderson; J.B.; and Gatz, D.F. The tabulation technique for forecasting concentrations of urban air pollutants. ANL-7615. ANL.[reports]. U.S. Atomic Energy Commission, pages 167-179, 1968.
- [Nagendra and Mukesh 2004] Nagendra S. M. and Mukesh Khare. Artificial neural network based line source models for vehicular exhaust emission predictions of an urban roadway. Transportation Research Part D: Transport and Environment, 2004. pages 69, 70.

- [Rivera. 2012] Rivera Poma. J. Modelo de identificación de factores contaminantes atmosféricos críticos en Lima-Callo. Tesis de Maestría. Universidad Nacional Mayor de San Marcos. 2012.
- [Ordieres, et.al.2005] Ordieres J. B., Vergara E. P., Capuz R. S., and Salazar R. E. Neural network prediction model for fine particulate matter (P M_{2,5}) on the US-México border in El Paso (Texas) and Ciudad Juarez (Chihuahua). Environmental Modelling and Software, 2005.
- [Ozbay, et.al,2011] Ozbay, B., Keskin, G., Dogruparmak, S. and Ayberk, S. Predicting tropospheric ozone concentrations in different temporal scales by using multilayer perceptron models. Ecological Informatics, 6(3-4):242{247,
- [OMM.2009] Organización meteorológica mundial. El tiempo, el clima y el aire que respiramos. 7 bis avenue de la Paix, Ginebra, Suiza, 2009.
- [ONU.2010] Organización de las Naciones Unidas. Objetivos de desarrollo del milenio, Informe 2010. New York, 2010.
- [Pita y Petergas 2002] Pita Fernández, S., Pértegas Díaz, S. “Investigación cuantitativa y cualitativa”. Unidadde Epidemiología Clínica y Bioestadística. Complejo Hospitalario Juan Canalejo. La Coruña (España).CAD ATEN PRIMARIA 2002; 9: 76-78.
- [Pascual et.al.1989] Pascual, D., Sánchez, M., Ramos, M. and Perez, I. A stochastic model to forecast lead pollutant. Nuovo Cimento C Geophys. Space Phys., 12(4):415-425, 1989.

- [Salazar.2008] Salazar Ruiz E. Desarrollo de modelos predictivos de contaminantes atmosféricos. Tesis Doctoral. Universidad Politécnica de Valencia. España. 2008.
- [Salazar y Álvarez 2013] Salazar Ceballos, A. y Álvarez Miño, L. Asociación de síntomas respiratorios con factores atmosféricos y climáticos en adultos en Santa Marta, Colombia. Revista Costarricense de Salud Pública. 2013. vol.22 n.1, San José.
- [Salini y Pérez 2006] Salini Calderón G. and Pérez Jata P. Estudio de series temporales de contaminación ambiental mediante técnicas de redes neuronales artificiales. Revista chilena de ingeniería, 2006. pages: 284–290.
- [Salini. 2009] Salini Calderón, A. Desarrollo de un modelo para pronosticar concentraciones extremas de PM2.5 en Santiago. Tesis Doctoral. 2009.
- [Sánchez.2012] Sánchez, P. Una nueva metodología de entrenamiento de redes neuronales y sus implicaciones en la selección de modelos. Tesis Doctoral. Universidad Nacional de Colombia. 2012.
- [Seetharam and Udaya 2009] Seetharam A. L. and Udaya Simha B. L. Urban air pollution trend and forecasting of major pollutants by timeseries analysis. International Journal of Civil and Environmental Engineering, pages 71–74, 2009.
- [SENAMHI.2013] Dirección General de Investigación y Asuntos Ambientales del Servicio Nacional de Meteorología e Hidrología.

(SENAMHI). Evaluación de la calidad del aire en Lima 2011.
Lima 2013.

[Sfetsos and Vlachogiannis 2010] Sfetsos A. and Vlachogiannis D. A new methodology development for the regulatory forecasting of P M10. Application in the Greater Athens Area, Greece. Atmospheric Environment, pages 3159 – 3172, 2010.

[Sheih and Moroz, 1973] Sheih, C.M. and Moroz, W.J. A Lagrangian puff diffusion model for the prediction of pollutant concentrations over urban areas. Proc. Third Inter. Clean Air Congress, VDI Verlag GMBH, pages B43-52, 1973.

[Tokumar and Habata] Tokumar, H. and Habata, O. Prediction of pollution levels by mixed order multi-variable AR scheme. Automatica, 14(6):597-601, 1978.

[Williams and Yamada, 1990] M. Williams and T. Yamada. A microcomputer-based forecasting model: Potential applications for emergency response plans and air quality studies. Journal of the Air and Waste Management Association, 40(9):1266-1274, 1990.

ANEXOS

ANEXO A: DISEÑO DEL CUESTIONARIO DE PRUEBAS DE SOFTWARE

1. El manejo de la pantalla “Registrar Parámetros ICA” es:
[3] Muy Fácil [2] Entendible [1] No se Entiende
2. ¿El Reporte “Filtros de data” es emitido por el sistema?
[3] Sí, Completamente [2] Si, Parcialmente [1] No lo Genera
3. La información mostrada en el reporte “Filtros de Data” es:
[3] Precisa [2] Inexacta [1] Totalmente Errada
4. El manejo de la ventana “Registrar Data meteorológica diaria”
[3] Muy Fácil [2] Entendible [1] No se Entiende
5. ¿ La exportación de los reportes es realizado por el Sistema ?
[3] Sí, Completamente [2] Si, Parcialmente [1] No lo Genera
6. La información mostrada en la “Validación de variables” es:
[3] Precisa [2] Inexacta [1] Totalmente Errada
7. El manejo de la pantalla “Generar y exportar archivos input” es:
[3] Muy Fácil [2] Entendible [1] No se Entiende
8. ¿ El procesamiento de la red neuronal es emitido por el Sistema ?
[3] Sí, Completamente [2] Si, Parcialmente [1] No lo Genera
9. La información mostrada en los reportes de matlab es:
[3] Precisa [2] Inexacta [1] Totalmente Errada
10. El manejo de la pantalla “Predecir ICA” es:
[3] Muy Fácil [2] Entendible [1] No se Entiende
11. ¿ La validación de la red neuronal es realizada por el Sistema ?
[3] Sí, Completamente [2] Si, Parcialmente [1] No lo Genera
12. La información mostrada en los resultados del sistema web es:
[3] Precisa [2] Inexacta [1] Totalmente Errada

ANEXO B: RANGO DEL ÍNDICE DE CALIDAD DE AIRE

Id	ICA		Estado	PM10		Efectos	Acciones Preventivas
	Mínimo	Máximo		Mínimo	Máximo		
1	0	50	Buena	0	50.00	No hay riesgos en la salud	
2	50	100	Moderado	50.00	150.00	Niños, ancianos y personas con enfermedades respiratorias y cardíacas, pueden presentar síntomas tales como tos y cansancio. La población en general no se ve afectada.	
3	100	200	Malo	150.00	250.00	Toda la población puede padecer de síntomas tales como tos seca, ojos cansados, ardor en la nariz y garganta, niños, ancianos y personas con enfermedades respiratorias y cardíacas, los efectos podrían ser más graves para la salud.	Reducir el esfuerzo físico y pesado al aire principalmente de personas con dolencias cardíacas y pulmonares niños y adultos.
4	200	300	Muy Malo	250.00	420.00	Toda la población puede presentar agravamiento de los síntomas tales como tos seca, ojos cansados, ardor en nariz y garganta. Efectos aun más graves en la salud de los niños ancianos y personas con problemas cardiovasculares.	Evitar cualquier esfuerzo físico al aire libre, principalmente personas con dolencias cardíacas o pulmonares niños y adultos.
5	300	1000	Alerta Máxima	420.00	1000.00	Toda la población puede padecer riesgos graves y manifestaciones de enfermedades respiratorias y cardiovasculares. Aumento de las muertes prematuras en personas de los grupos más sensibles.	Todas las personas deben evitar actividades al aire libre.

ANEXO C

DISEÑO DEL CUESTIONARIO PARA MEDIR LA SATISFACCIÓN DEL USUARIO

1. La cantidad de trabajo realizado para verificar los parámetros de ICA ha:
[3] Reducido Notablemente [2] Disminuido
[1] Aumentado [0] Incrementado Mucho
2. La cantidad de trabajo realizado para registrar y/o almacenar la data meteorológica ha:
[3] Reducido Notablemente [2] Disminuido
[1] Aumentado [0] Incrementado Mucho
3. La disponibilidad de la información meteorológica es:
[3] Al Instante [2] A tiempo
[1] Con Retrazo [0] A Destiempo
4. La cantidad de trabajo realizado para validar las variables que afecten el comportamiento del agente contaminante PM10:
[3] Reducido Notablemente [2] Disminuido
[1] Aumentado [0] Incrementado Mucho
5. La velocidad con la que se exporta e importa la data meteorológica ha:
[3] Muy Rápida [2] Rápida
[1] Lenta [0] Muy Lenta
6. La disponibilidad de las predicciones del PM10, a 24 horas, en las diversas estaciones es:
[3] Al Instante [2] A tiempo
[1] Con Retrazo [0] A Destiempo
7. La cantidad de trabajo empleado para realizar reportes gráficos meteorológicos ha:
[3] Reducido Notablemente [2] Disminuido
[1] Aumentado [0] Incrementado Mucho
8. La velocidad con que la que se genera el mapa geográfico es:
[3] Muy Rápida [2] Rápida
[1] Lenta [0] Muy Lenta

9. El tiempo que demora para visualizar las predicciones del agente contaminante PM10 ha:
- [3] Reducido Notablemente [2] Disminuido
 - [1] Aumentado [0] Incrementado Mucho
10. La disponibilidad de los reportes es:
- [3] Al Instante [2] A tiempo
 - [1] Con Retrazo [0] A Destiempo
11. La velocidad con que se genera los reportes es:
- [3] Muy Rápida [2] Rápida
 - [1] Lenta [0] Muy Lenta
12. El tiempo de recuperación de información de los diversos datos es:
- [3] Al Instante [2] A tiempo
 - [1] Con Retrazo [0] A Destiempo
13. El tiempo utilizado en la realización de informes sobre los procesos meteorológicos
- [3] Reducido Notablemente [2] Disminuido
 - [1] Aumentado [0] Incrementado Mucho
14. El trabajo para realizar la minería de datos, a través de los modelos implementados ha:
- [3] Reducido Notablemente [2] Disminuido
 - [1] Aumentado [0] Incrementado Mucho

ANEXO D: CARTAS DE VALIDACIÓN POR LOS EXPERTOS

Trujillo, 09 de octubre del 2014

Escuela de Post Grado
Universidad Nacional del Santa

Por la presente, doy constancia que he participado como experto en la evaluación y validación del Sistema Integral, para la predicción y control del contaminante PM10 en la ciudad de Chimbote, realizado por Sixto Díaz Tello.

Debo indicar que el sistema cumple con los estándares de calidad en desarrollo de software y permite hacer predicciones sobre la contaminación atmosférica por PM10 en la ciudad de Chimbote

Por lo expuesto, recomiendo el trabajo del doctorando como requisito previo a la obtención del título de doctor por la Universidad Nacional del Santa.



Carlos Chávez Monzón

Carlos Chávez Monzón
Universidad Nacional de Trujillo
Universidad César Vallejo

Lima, 08 de octubre del 2014

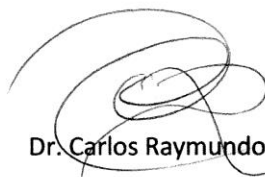
Escuela de Post Grado
Universidad Nacional del Santa

Por la presente, doy constancia que he participado como experto en la evaluación y validación del Sistema Integral, para la predicción y control del contaminante PM10 en la ciudad de Chimbote, realizado por Sixto Díaz Tello.

Debo indicar que se implementó y validó la red neuronal, para la predicción del contaminante PM10, donde se demostró que el sistema cumple con el objetivo de predecir los índices de calidad del aire de Chimbote.

El sistema cumple con la calidad en el software en cuanto a su funcionalidad, facilidad de uso, confiabilidad, desempeño y seguridad-

Por lo expuesto, recomiendo el trabajo del doctorando como requisito previo a la obtención del título de doctor por la Universidad Nacional del Santa.



Dr. Carlos Raymundo Ibañez

Dr. Carlos Raymundo Ibañez
Universidad Peruana de Ciencias Aplicadas
Investigador Asociado en Programa Internacional STIC AMSUD
Miembro Titular del Comité Técnico del Programa Nacional de Fortalecimiento a la Innovación en Consejo Nacional de Ciencia Tecnología, e Innovación Tecnológica (Concytec)

Lima, 08 de octubre del 2014

Escuela de Post Grado
Universidad Nacional del Santa

Por la presente, doy constancia que he participado como experto en la evaluación y validación del Sistema Integral, para la predicción y control del contaminante PM10 en la ciudad de Chimbote, realizado por Sixto Díaz Tello.

Debo indicar que el sistema cumple con las especificaciones y requerimientos para el desarrollo de software.

Por lo tanto, el sistema integral, bajo el enfoque de minería de datos y redes neuronales, si permite realizar la predicción de la contaminación atmosférica por PM10 en la ciudad de Chimbote.

Por lo expuesto, recomiendo el trabajo del doctorando como requisito previo a la obtención del título de doctor por la Universidad Nacional del Santa.



Dra. Diana Muñoz Casanova

Dr. Diana Muñoz Casanova
Universidad Nacional del Santa
Presidente del Capítulo de Ingeniería de Sistemas-CIP-Chimbote
Especialista en Sistemas Inteligentes

Trujillo, 09 de octubre del 2014

Escuela de Post Grado
Universidad Nacional del Santa

Por la presente, doy constancia que he participado como experto en la evaluación y validación del Sistema Integral, para la predicción y control del contaminante PM10 en la ciudad de Chimbote, realizado por Sixto Díaz Tello.

Debo hacerle la recomendación de presentarlo a las instituciones que tienen que ver con el control del medio ambiente en Chimbote como DIGESA o la OEFA, ya que es un trabajo que tiene impacto social.

El sistema cumple con hacer las predicciones del contaminante PM10 en la ciudad de Chimbote

Por lo expuesto, recomiendo el trabajo del doctorando como requisito previo a la obtención del título de doctor por la Universidad Nacional del Santa.



Dr. Luis Fernando Torres Cabrera

Dr. Luis Fernando Torres Cabrera
Universidad Nacional del Santa
Universidad Nacional de Trujillo

Nuevo Chimbote, 10 de octubre del 2014

Escuela de Post Grado
Universidad Nacional del Santa

Por la presente, doy constancia que he participado como experto en la evaluación y validación del Sistema Integral, para la predicción y control del contaminante PM10 en la ciudad de Chimbote, realizado por Sixto Díaz Tello.

Debo indicar que el sistema cumple con los estándares de calidad en desarrollo de software, haciéndole la recomendación para una línea de investigación futura que el sistema de Minería de Datos pueda enlazarse con el sistema web y reportar patrones de comportamiento sobre contaminantes.

Por lo expuesto, recomiendo el trabajo del doctorando como requisito previo a la obtención del título de doctor por la Universidad Nacional del Santa.



Dr. Guillermo Gil Albarrán

Dr. Guillermo Gil Albarrán
Universidad Nacional del Santa
Especialista en desarrollo de software