

**UNIVERSIDAD MAYOR DE SAN ANDRÉS
FACULTAD DE CIENCIAS PURAS Y NATURALES
CARRERA DE INFORMÁTICA**



TESIS DE GRADO

**“MODELO DE PREDICCIÓN DE PREVALENCIA DE
ENFERMEDADES PARA CENTROS DE SALUD BASADO EN
MINERÍA DE DATOS”**

PARA OPTAR AL TÍTULO DE LICENCIATURA EN INFORMÁTICA
MENCIÓN INGENIERÍA DE SISTEMAS INFORMÁTICOS

Postulante: Boris Rodrigo Apaza Quevedo

Tutor: Lic. Efraín Silva Sánchez

Revisor: Msc. Lic. Lucio Torrico Diaz

**LA PAZ – BOLIVIA
2009**



**UNIVERSIDAD MAYOR DE SAN ANDRÉS
FACULTAD DE CIENCIAS PURAS Y NATURALES
CARRERA DE INFORMÁTICA**



LA CARRERA DE INFORMÁTICA DE LA FACULTAD DE CIENCIAS PURAS Y NATURALES PERTENECIENTE A LA UNIVERSIDAD MAYOR DE SAN ANDRÉS AUTORIZA EL USO DE LA INFORMACIÓN CONTENIDA EN ESTE DOCUMENTO SI LOS PROPÓSITOS SON ESTRICTAMENTE ACADÉMICOS.

LICENCIA DE USO

El usuario está autorizado a:

- a) visualizar el documento mediante el uso de un ordenador o dispositivo móvil.
- b) copiar, almacenar o imprimir si ha de ser de uso exclusivamente personal y privado.
- c) copiar textualmente parte(s) de su contenido mencionando la fuente y/o haciendo la referencia correspondiente respetando normas de redacción e investigación.

El usuario no puede publicar, distribuir o realizar emisión o exhibición alguna de este material, sin la autorización correspondiente.

TODOS LOS DERECHOS RESERVADOS. EL USO NO AUTORIZADO DE LOS CONTENIDOS PUBLICADOS EN ESTE SITIO DERIVARA EN EL INICIO DE ACCIONES LEGALES CONTEMPLADOS EN LA LEY DE DERECHOS DE AUTOR.

DEDICATORIA

A mis papás, Gregorio y Elvira

AGRADECIMIENTOS

A Dios Padre, Hijo, Espíritu Santo y la virgen María por darle un sentido a mi vida.

A mis papás Gregorio y Elvira por su apoyo incondicional en momentos de alegría y adversos, a mis hermanos Henry, Aleida, Amira, Constancio, Fabiola y mis sobrinos Daniel y Gabriela que me alegran cada día en esta tierra.

A Walter, Marcelo, Walex, Hugo quienes con muchos más se constituyeron en mi familia.

Al lic. Lucio Torrico Díaz por su paciencia y comprensión y al lic. Efraín Silva por sus correcciones e interés, sin el apoyo de los mismos no hubiera sido posible este trabajo.

A todos los profesores y docentes que se encargaron de mi formación académica y personal comenzando de la escuela Pablo Iturri Jurado, Colegio San Simón de Ayacucho, Colegio San Juan Bautista de La Salle y finalmente la carrera de Informática de la Universidad Mayor de San Andrés.

Y compañeros de los cuales aprendí mucho.

GRACIAS.

RESUMEN

La epidemiología es el estudio de "enfermedades que afectan transitoriamente a muchas personas en un sitio determinado". Se considera vigilancia epidemiológica al seguimiento de la ocurrencia de las enfermedades en las poblaciones. El Ministerio de Salud y Deportes de la República de Bolivia ha seleccionado un grupo de enfermedades que están sujetas a notificación obligatoria. Esta selección se ha realizado en base a las normas internacionales, que permiten priorizar las enfermedades en base a diferentes criterios, uno de ellos es la magnitud, es decir si la enfermedad tiene alta incidencia o prevalencia. La prevalencia es la frecuencia estadística (absoluta o relativa) con la que se presenta un determinado fenómeno en una población, entendido este término en su sentido estadístico es un real entre 0 y 1 denotando la ausencia o presencia de una enfermedad.

El presente trabajo de tesis tiene como fin desarrollar un modelo de predicción de este valor de prevalencia para una enfermedad determinada en un establecimiento de salud en un tiempo determinado. Para ello es importante identificar variables o atributos que sean importantes para la predicción de este valor. Gracias a la teoría de conjuntos aproximados es posible este hecho, pues básicamente la aplicación de dicha teoría es la identificación de atributos relevantes de un conjunto de ellos conocidos como "reducto". Ubicados dichos atributos es posible generar datos de entrenamiento y datos de prueba para una red neuronal artificial con el fin de desarrollar un modelo de predicción. Este hecho se realizó en los capítulos 3 y 4 del presente trabajo.

Para la realización de pruebas se desarrolló un prototipo software con principios de la ingeniería de software. Las pruebas se realizaron con datos de las 10 enfermedades con mayor índice de prevalencia en centros de salud de Bolivia, información otorgada por la ONG Medicus Mundi. Obteniendo primeramente los atributos relevantes y posteriormente una red neuronal que como datos de entrada recibe valores de los atributos relevantes que forman un reducto y como dato de salida obtiene la prevalencia. Los datos esperados y los datos predichos por el modelo para dichas enfermedades no mostraron importante diferencia a un nivel de 95% de confianza para la prueba de hipótesis que se realizó.

Llegando a la conclusión que es posible construir e implementar un modelo de predicción de prevalencia para enfermedades seleccionando adecuadamente variables relevantes que influyen en las mismas.

ÍNDICE

CAPÍTULO 1 MARCO REFERENCIAL	1
1.1 Introducción	2
1.2 Antecedentes	4
1.2.1 Descubrimiento de conocimiento en bases de datos	4
1.2.2 Minería de datos	4
1.2.3 Teoría de conjuntos aproximados	6
1.2.4 Redes neuronales artificiales	7
1.2.5 Trabajos de minería de datos en la Carrera de Informática	7
1.3 Planteamiento del problema	8
1.3.1 Problema general	8
1.3.2 Problemas específicos	8
1.4 Planteamiento de la hipótesis	9
1.5 Objetivos	9
1.5.1 Objetivo general	9
1.5.2 Objetivos específicos	9
1.6 Justificación	10
1.6.1 Justificación científica	10
1.6.2 Justificación técnica	10
1.6.3 Justificación social	10
1.7 Métodos y técnicas	11
1.8 Alcance y aporte	12
CAPÍTULO 2 MARCO TEÓRICO	13
2.1 Epidemiología	14
2.2 Morbilidad	14
2.3 Prevalencia	14
2.3.1 Prevalencia puntual o instantánea	15
2.3.2 Prevalencia de periodo	15
2.3.3 Dimensiones de la prevalencia	16
2.4 Vigilancia Epidemiológica	16
2.4.1 Enfermedades sujetas a notificación	16
2.4.2 Enfermedades de notificación inmediata	17
2.4.3 Enfermedades de notificación semanal	18

2.5 Fuentes de notificación	18
2.5.1 Puestos normales	19
2.5.2 Centros de salud	19
2.5.3 Hospitales de segundo y tercer nivel	20
2.5.4 Gerencia de red	20
2.5.5 SNIS (Servicio nacional de Información en salud) – SEDES (Servicio departamental de salud)	21
2.5.6 SNIS nacional (Sistema Nacional de Información en Salud)	21
2.6 Clasificación Internacional de Enfermedades 10 (CIE 10)	22
2.6.1 Definición	22
2.6.2 Propósito	22
2.6.3 Utilidad	22
2.6.4 Estructura básica de la CIE 10ª Revisión	23
2.7 Minería de datos	23
2.8 Conjuntos Aproximados	24
2.8.1 Sistema de Información	24
2.8.2 Sistema de decisión (SI)	24
2.8.3 Función de información	24
2.8.4 Indiscernibilidad	24
2.8.4.1 Relación de equivalencia	25
2.8.4.2 Clase de equivalencia	25
2.8.5 Inseparabilidad	25
2.8.6 Aproximación superior e inferior a un conjunto	25
2.8.7 Región límite	26
2.8.8 Región exterior	26
2.8.9 Precisión de la aproximación	27
2.8.10 Calidad de la aproximación	27
2.8.11 Calidad de la clasificación	27
2.8.12 Atributos dispensables e indispensables	28
2.8.13 Región positiva	28
2.8.14 Independencia	28
2.8.15 Reducto	28
2.8.16 Núcleo	28
2.9 Redes Neuronales Artificiales	29

2.9.1	Perceptrón	29
2.9.2	Retropropagación	30
2.10	Proceso unificado de desarrollo de software	31
CAPÍTULO 3 MARCO DE MODELADO		33
3.1	Fases del modelo	34
3.2	Estudio de atributos con relación a una enfermedad en una tabla de base de datos médica	34
3.2.1	Estudio de dominio de atributos de una enfermedad en una tabla médica	36
3.2.2	Refinación de tablas médicas	37
3.2.2.1	Tablas de análisis	37
3.2.2.2	Atributos condición de una tabla de análisis	37
3.2.2.3	Proporción	37
3.2.2.4	Atributo decisión de una tabla de análisis	39
3.2.2.5	Prevalencia de enfermedades	39
3.2.2.6	Valores propuestos para prevalencia	39
3.2.2.7	Construcción de tablas de análisis	41
3.3	Modelado para conjuntos aproximados	43
3.3.1	Sistema de Información	43
3.3.2	Algoritmo de reducción	43
3.3.3	Reglas de decisión	44
3.4	Red Neuronal	45
3.4.1	Arquitectura	45
3.4.1.1	Capa de entrada	46
3.4.1.2	Capas ocultas	46
3.4.1.3	Capa de salida	47
3.5	Pruebas	47
3.5.1.1	Datos de entrenamiento	48
3.5.1.2	Datos entrenamiento de atributos de condición para un reducto obtenido	48
3.5.1.3	Datos de entrenamiento del atributo decisión	49
3.5.1.4	Datos de prueba	49

CAPÍTULO 4 MARCO APLICATIVO	50
4.1 Desarrollo de la herramienta de predicción	51
4.1.1 Captura de requisitos como casos de uso	51
4.1.1.1 Descripción de los casos de uso	52
4.1.1.1.1 Configurar la conexión de origen y salida	52
4.1.1.1.2 Preparar atributos y límites de tiempo, enfermedad y centro de salud	52
4.1.1.1.3 Configurar la tabla de análisis (sistema de información) para conjuntos aproximados	53
4.1.1.1.4 Generar red neuronal de predicción para un reducto	53
4.1.1.1.5 Probar red neuronal para un reducto	53
4.1.1.1.6 Predecir	53
4.1.2 Análisis	54
4.1.2.1 Configurar la conexión de origen y salida	54
4.1.2.2 Preparar atributos y límites de tiempo, enfermedad y centro de salud	55
4.1.2.3 Configurar la tabla de análisis (sistema de información) para conjuntos aproximados	56
4.1.2.4 Generar red neuronal de predicción para un reducto	56
4.1.2.5 Probar red neuronal para un reducto	57
4.1.2.6 Predecir	58
4.1.2.7 Identificación de subsistemas y sus dependencias	58
4.1.3 Diseño	59
4.1.3.1 Paquete de diseño Conexión	59
4.1.3.2 Paquete de diseño de HerramientasIA	60
4.1.3.3 Paquete de diseño Preparación de datos	61
4.1.3.4 Paquete de diseño de Sistema de Predicción de Prevalencia de Enfermedades para Centros de Salud (GUI)	62
4.1.4 Implementación	62
4.1.4.1 Solución Conexión	63
4.1.4.2 Solución HerramientasIA	63
4.1.4.3 Solución Preparación de datos	64
4.1.4.4 Solución Sistema de Predicción de Prevalencia de Enfermedades para Centros de Salud (GUI)	64
4.2 Preparación de tablas de análisis	65
4.3 Solución basada en la teoría conjuntos aproximados	65

4.3.1	Selección de reductos	65
4.4	Red neuronal	67
4.5	Características para datos de análisis	68
4.5.1	Asignación de tiempo	68
4.5.2	Elección de columnas de análisis	69
4.5.3	Elección de enfermedades	69
4.5.4	Elección de centros de salud	70
4.6	Variables	71
4.6.1	Identificación de variables	71
4.7	Error de predicción	71
4.8	Prueba de hipótesis	72
4.9	Pruebas	73
4.9.1	Prueba 1	73
4.9.2	Prueba 2	77
4.9.3	Prueba 3	81
4.9.4	Prueba 4	85
4.9.5	Prueba 5	88
4.9.6	Prueba 6	92
4.9.7	Prueba 7	96
4.9.8	Prueba 8	100
4.9.9	Prueba 9	104
4.9.10	Prueba 10	108
4.10	Conclusiones de las pruebas	111
4.11	Ejemplo de predicción de prevalencia	111
CAPÍTULO 5 CONCLUSIONES		115
5.1	Conclusiones generales	116
5.2	Estado de los objetivos	117
5.2.1	Estado de los objetivos específicos	117
5.2.2	Estado del objetivo general	118
5.3	Estado de la hipótesis	118
5.4	Recomendaciones	118
Bibliografía		120
ANEXOS		

Anexo 1 Categorías de la clasificación Internacional de Enfermedades CIE

Anexo 2 Árbol de problemas y objetivos

ÍNDICE DE FIGURAS

Figura 1. Representación del término medio en el modelamiento predictivo de minería de datos	3
Figura 2. Relación entre dato, información y conocimiento	4
Figura 3. Definición de un conjunto aproximado	26
Figura 4. Perceptrón	29
Figura 5. Fases e iteraciones del proceso unificado de desarrollo de software	32
Figura 6. Extracción de la base de datos Morbilidad de Medicus Mundi	35
Figura 7. Principales causas de morbilidad en consulta externa área urbana	36
Figura 8: Subconjuntos del dominio d	38
Figura 9: Intervalos de valores de prevalencia propuestos	40
Figura 10. Esquema general de las conexiones de una red neuronal	46
Figura 11. Función de transferencia ‘Tansig’	47
Figura 12. Función de transferencia ‘Purelin’	47
Figura 13. Diagrama de casos de uso	52
Figura 14. Diagrama de colaboración para el caso de uso: configurar la conexión de origen y salida	54
Figura 15. Diagrama de colaboración para el caso de uso: Preparar Atributos y límites de tiempo, enfermedad y centro de salud	55
Figura 16. Diagrama de colaboración para el caso de uso: Configurar la tabla de análisis (sistema de información) para conjuntos aproximados	56
Figura 17. Diagrama de colaboración para el caso de uso: Generar red neuronal de predicción para un reducto	57
Figura 18. Diagrama de colaboración para el caso de uso: Probar red neuronal para un reducto	57
Figura 19. Diagrama de colaboración para el caso de uso: Predecir	58
Figura 20. Diagrama de paquetes y dependencias	58
Figura 21. Diagrama de clases para el paquete de diseño: Conexión	59
Figura 22. Diagrama de clases para el paquete de diseño: HerramientasIA	60
Figura 23. Diagrama de clases para el paquete de diseño: Preparación de datos	61
Figura 24. Diagrama de clases para el paquete de diseño: Paquete de diseño de Sistema De Predicción de Prevalencia de Enfermedades para Centros de Salud (GUI)	62
Figura 25. Solución: Conexión	63
Figura 26. Solución: HerramientasIA	63

Figura 27. Solución: Preparación de Datos	64
Figura 28. Solución: Sistema De Predicción de Prevalencia de Enfermedades para Centros de Salud (GUI)	64
Figura 29. Código del algoritmo propuesto para encontrar un reducto	65
Figura 30. Heurístico propuesto para encontrar varios reductos	66
Figura 31. Código del heurístico propuesto para encontrar varios reductos	67
Figura 32. Código fuente de la red neuronal artificial	68
Figura 33. Las diez enfermedades con mayor índice de prevalencia	69
Figura 34. Información provista por el hospital Boliviano Holandés para “Diarrea y gastroenteritis de presunto origen infeccioso”	71
Figura 35. Resultados de predicción de reducto 2 en prueba 1	76
Figura 36. Resultados de predicción de reducto 2 en prueba 2	80
Figura 37. Resultados de predicción de reducto 2 en prueba 3	84
Figura 38. Resultados de predicción de reducto 1 en prueba 4	87
Figura 39. Resultados de predicción de reducto 3 en prueba 5	91
Figura 40. Resultados de predicción de reducto 3 en prueba 6	95
Figura 41. Resultados de predicción de reducto 6 en prueba 7	99
Figura 42. Resultados de predicción de reducto 5 en prueba 8	102
Figura 43. Resultados de predicción de reducto 1 en prueba 9	107
Figura 44. Resultados de predicción de reducto 4 en prueba 10	110
Figura 45. Promedio de pacientes que padecen diarrea y gastroenteritis de presunto origen infeccioso por meses en el Hospital Boliviano Holandés	113
Figura 46. Pacientes que padecen diarrea y gastroenteritis de presunto origen infeccioso clasificados en menores a 1 años, de 1 a 15 años y masculino por promedio de meses en el Hospital Boliviano Holandés	113
Figura 47. Predicción de prevalencia de diarrea y gastroenteritis de presunto origen infeccioso basada en promedio de pacientes por meses en el Hospital Boliviano Holandés	114

ÍNDICE DE TABLAS

Tabla 1. Tabla de datos médica tradicional	41
Tabla 2. Recorte de tabla de datos médica tradicional para la enfermedad "A"	41
Tablas 3 y 4. Atributos de condición de la enfermedad "A"	42
Tabla 5. Tabla de análisis refinada para la enfermedad "A"	43
Tabla 6. Reducto de una tabla para análisis según los subconjuntos	44
Tabla 7. Esquema general de una tabla de datos de entrenamiento para atributos de condición en un reducto	48
Tabla 8. Esquema general de datos de entrenamiento para prevalencia	49
Tabla 9. Tabla de análisis para diarrea y gastroenteritis de presunto origen infeccioso en el hospital Boliviano Holandés de la ciudad de El Alto	74
Tabla 10. Meses y prevalencia propuesta de la tabla de análisis para diarrea y gastroenteritis de presunto origen infeccioso en el hospital Boliviano Holandés de la ciudad de El Alto	75
Tabla 11. Reductos de la tabla de análisis para diarrea y gastroenteritis de presunto origen infeccioso en el hospital Boliviano Holandés de la ciudad de El Alto	76
Tabla 12. Tabla de análisis para Rinofaringitis aguda en el hospital Boliviano Holandés de la ciudad de El Alto	78
Tabla 13. Meses y prevalencia propuesta de la tabla de análisis para Rinofaringitis aguda en el hospital Boliviano Holandés de la ciudad de El Alto	79
Tabla 14. Reductos obtenidos de la tabla de análisis para Rinofaringitis aguda en el hospital Boliviano Holandés de la ciudad de El Alto	80
Tabla 15. Tabla de análisis para Hipertensión en el Hospital General Universitario Japonés	82
Tabla 16. Meses y prevalencia de la tabla de análisis para Hipertensión en el Hospital General Universitario Japonés	83
Tabla 17. Reductos obtenidos de la tabla de análisis para Hipertensión en el Hospital General Universitario Japonés	83
Tabla 18. Tabla de análisis para Conjuntivitis en el Instituto Nacional de Oftalmología – La Paz	85
Tabla 19. Meses y prevalencia de la tabla de análisis para Conjuntivitis en el Instituto Nacional de Oftalmología – La Paz	86
Tabla 20. Reductos obtenidos de la tabla de análisis para Conjuntivitis en el Instituto Nacional de Oftalmología – La Paz	87

Tabla 21. Tabla de análisis para Infección de vías urinarias en el Hospital General Universitario Japonés	88
Tabla 22. Meses y prevalencia de la tabla de análisis para Infección de vías urinarias en el Hospital General Universitario Japonés	90
Tabla 23. Reductos obtenidos de la tabla de análisis para Infección de vías urinarias en el Hospital General Universitario Japonés	90
Tabla 24. Tabla de cefalea en el Hospital Santa Bárbara - Sucre	92
Tabla 25. Meses y prevalencia de la tabla de cefalea en el Hospital Santa Bárbara - Sucre	94
Tabla 26. Tabla de cefalea en el Hospital Santa Bárbara - Sucre	94
Tabla 27. Tabla de análisis de Faringitis aguda en el Hospital Santa Bárbara - Sucre	96
Tabla 28. Tabla de mes y prevalencia de la tabla de análisis de Faringitis aguda en el Hospital Santa Bárbara - Sucre	98
Tabla 29. Reductos obtenidos de la Tabla de análisis de Faringitis aguda en el Hospital Santa Bárbara - Sucre	98
Tabla 30. Tabla de análisis de bronquitis aguda en el instituto nacional del tórax	100
Tabla 31. Meses y prevalencia de la tabla de análisis de bronquitis aguda en el instituto nacional del tórax	101
Tabla 32. Reductos obtenidos de la tabla de análisis de bronquitis aguda en el instituto nacional del tórax	102
Tabla 33. Tabla de análisis de dorsalgia en el Hospital Boliviano Holandés	104
Tabla 34. Meses y prevalencia de la tabla de análisis de dorsalgia en el Hospital Boliviano Holandés	106
Tabla 35. Tabla de análisis de dorsalgia en el Hospital Boliviano Holandés	106
Tabla 36. Tabla de análisis fiebre de origen desconocido en el Hospital de niños	108
Tabla 37. Mes y prevalencia de la tabla de análisis fiebre de origen desconocido en el Hospital de niños	109
Tabla 38. Reductos obtenidos de la tabla de análisis fiebre de origen desconocido en el Hospital de niños	110
Tabla 39. Número de pacientes con relación al tiempo en el Hospital Boliviano Holandés de la ciudad de El Alto	112

CAPÍTULO 1

MARCO REFERENCIAL

Jaws PDF Creator

EVALUATION

VALUTAZIONE

EVALUATION

EVALUACIÓN

Resumen

Donde se presenta la temática de la presente tesis mediante la introducción, la identificación del problema, planteamiento de objetivos, formulación de la hipótesis, justificación, metodología y alcances.

1.1 Introducción

Desde finales de los años 1950 las organizaciones han almacenado incrementalmente en bases de datos información referente a sus actividades. Esto se debe a que los costos de almacenamiento y la tecnología cada vez son más disponibles. La capacidad de generar y almacenar información creció considerablemente en los últimos tiempos. En las últimas dos décadas se ha visto un dramático aumento en la cantidad de datos almacenados en formato electrónico. Moore (1965) estimó que la cantidad de información en el mundo se duplica cada 20 meses, los datos históricos¹ revelan que la cantidad de información crece aún más rápido.

Gracias al lenguaje estructurado de consultas (SQL)² las organizaciones en un primer análisis obtienen aproximadamente el 8 % de la información de sus bases de datos. El 20 % restante, que la mayoría de las veces contiene la información más importante, requiere la utilización de técnicas más avanzadas. La minería de datos engloba técnicas para este fin. La minería de datos es un proceso no trivial de identificación válida, novedosa, potencialmente útil y entendible de patrones comprensibles que se encuentran ocultos en los datos según Fayyad, Piatetsky-Shapira y Smyth (1993).

Molina (2000) señala que para realizar minería de datos se utilizan algoritmos basados principalmente en árboles y reglas de decisión, reglas de asociación, redes neuronales, redes bayesianas, teoría de conjuntos aproximados, máquinas de soporte vectorial, algoritmos genéticos y lógica difusa.

La “teoría de conjuntos aproximados” propuesta por Pawlak en 1982 aporta métodos de clasificación y predicción. Una aplicación de esta teoría es la obtención de reducidos. Un reducido es un subconjunto mínimo de atributos que almacenan información relevante sobre otros (Cárdenas, Caballero, Bello, 2007).

Algunas técnicas de minería de datos han sido empleadas para la predicción de enfermedades, probabilidad de respuesta satisfactoria a tratamientos

¹ Más detalles sobre datos históricos en <http://www.intel.com/technology/mooreslaw/index.htm>

² Abreviación del original: “Structured Query Language”

médicos y pronósticos meteorológicos, entre ellas están las redes neuronales, evaluación por puntuación, reglas de decisión y árboles de decisión, de donde se observa que las redes neuronales tienen un alto poder de predicción. (Valcácel, 2004).

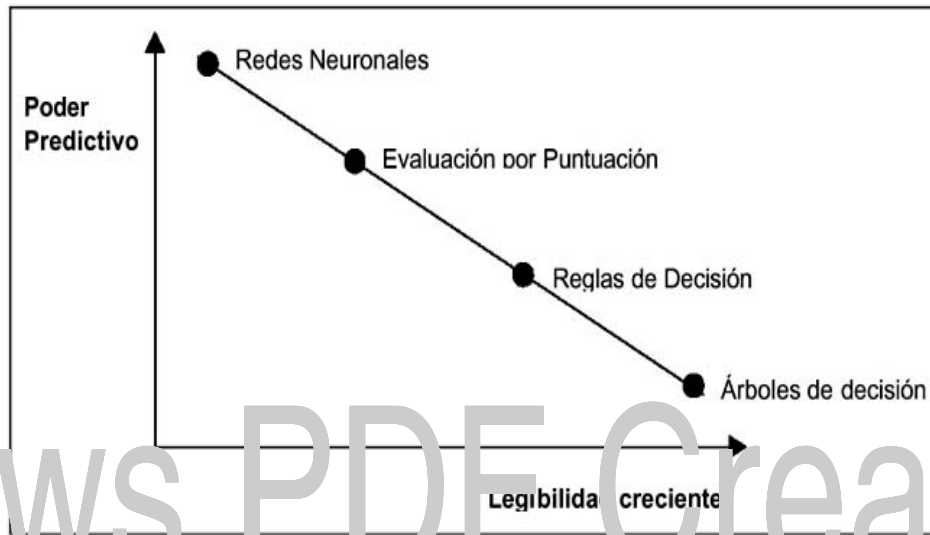


Figura 1. Representación del término medio en el modelamiento predictivo de minería de datos
Fuente: (Valcácel 2004)

Los centros de salud almacenan en bases de datos información de pacientes que acuden al centro, síntomas y tratamiento de enfermedades. Estos datos son importantes para la emisión de reportes sobre enfermedades, como la prevalencia que es la proporción de individuos de una población que padecen una enfermedad en un momento o periodo de tiempo determinado (Fernández, Pértigas y Cañedo, 2004). Medicus Mundi³ es una organización no gubernamental (ONG) de apoyo al desarrollo de centros de salud de Bolivia, especializada en temas de salud que requiere de un modelo de predicción para la prevalencia de enfermedades.

Esta tesis está orientada a la construcción de un modelo predictivo mediante técnicas de minería de datos de conjuntos aproximados y redes neuronales artificiales para la predicción de prevalencia de enfermedades.

³ Más información en: <http://www.medicusmundi.info/>

1.2 Antecedentes

1.2.1 Descubrimiento de conocimiento en bases de datos

El descubrimiento de conocimientos en bases de datos (KDD)⁴ es un proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles a partir de los datos (Fayyad et al., 1996). El objetivo fundamental del KDD es encontrar conocimiento útil, válido, relevante y nuevo sobre un fenómeno o actividad mediante algoritmos eficientes, dado el creciente almacenamiento de datos.

El proceso de descubrimiento del conocimiento en base de datos implica un proceso interactivo e iterativo involucrando la aplicación de varios algoritmos de datos para llegar al conocimiento.

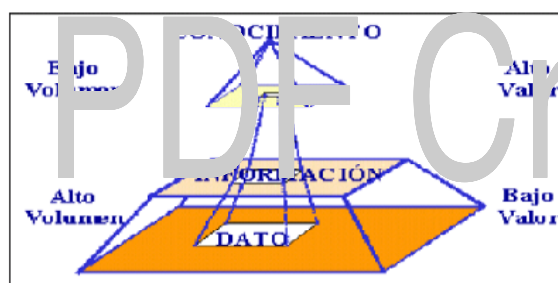


Figura 2 Relación entre dato, información y conocimiento
Fuente: (Molina, 1998)

1.2.2 Minería de datos

La idea de minería de datos no es nueva. Ya desde los años 1960 estadísticos manejaban términos como “pesca de datos”, “minería de datos” o “arqueología de datos” con la idea de encontrar correlaciones sin una hipótesis previa en bases de datos con ruido. A principios de los años 1980, Rakesh Agrawal, Gio Wiederhold, Robert Blum y Gregory Piatetsky-Shapiro⁵, entre otros, empezaron a consolidar los términos de minería de datos y KDD.

La minería de datos es una tecnología que reúne varias áreas, sustentadas principalmente por la estadística, la inteligencia artificial, el reconocimiento de patrones, las bases de datos y la teoría de la información. Otras áreas como la

⁴ Abreviación del original: “Knowledge Discovery in Data Bases”

⁵ Publicaciones de Agrawal, Blum y Piatetsky en <http://www.kdnuggets.com>

computación gráfica, el procesamiento paralelo y la teoría de grafos pueden participar en esto. Su objetivo es la extracción de conocimiento en bases de datos, buscando situaciones novedosas y relevantes (Molina y Ribeiro, 2001).

En el sentido aplicativo la minería de datos es la etapa de descubrimiento en el proceso de KDD: representa el paso consistente en el uso de algoritmos concretos que generan una enumeración de patrones a partir de datos preprocesados (Fayyad et al., 1996).

Diferentes técnicas y algoritmos de minería de datos son utilizados para analizar y descubrir tendencias no solo para extraer información sino para identificar comportamientos; estas tendencias no son muy evidentes.

Según Cárdenas y sus colegas (2007) algunas de las técnicas de minería de datos más utilizadas son las siguientes:

- Árboles de decisión. Un árbol de decisión es una estructura en forma de árbol que visualmente describe una serie de reglas que causan que una decisión sea tomada.
- Algoritmos genéticos. Son métodos de optimización basados en los mecanismos de la reproducción natural que se utilizan para mejorar otros algoritmos de minería de datos.
- Redes neuronales artificiales. Son modelos de predicción no lineales que “aprenden” a través de experiencia, en minería de datos son utilizadas para generar modelos de regresión que puedan predecir comportamientos futuros.
- Técnicas estadísticas. Se utilizan para identificar patrones, los cuales pueden ser entonces utilizados para predecir el futuro. Estas incluyen las regresiones lineales, los modelos aditivos generalizados y las regresiones adaptativas multivariadas por ranuras.
- Reglas de asociación. Se utilizan para la búsqueda de reglas que revelan la naturaleza y frecuencia de las relaciones o asociaciones entre las entidades de los datos.
- Lógica difusa. Proporciona la noción de un conjunto difuso más que una clara demarcación de límites.

- Conjuntos aproximados. La teoría de conjuntos aproximados es útil para determinar la relevancia de los atributos y generar relaciones entre ellos.

1.2.3 Teoría de conjuntos aproximados

La teoría de conjuntos aproximados fue introducida por Pawlak en 1982. Se basa en aproximar un subconjunto del dominio de un atributo por un par de conjuntos exactos, llamados aproximación superior y aproximación inferior del subconjunto. Con esta teoría es posible tratar tanto datos cuantitativos como cualitativos, y no se requiere eliminar inconsistencias previas al análisis, puede ser usada para determinar la relevancia de los atributos, generar relaciones entre ellos (en forma de reglas), entre otras.

El descubrimiento de dependencias entre los atributos que conforman los conjuntos aproximados es importante para la determinación de información redundante y la determinación de atributos más relevantes y de mayor capacidad descriptiva. Según Pawlak (1982) este subconjunto mínimo de atributos recibe la denominación de “reducto” y la intersección de todos los reductos “núcleo”, que constituye la colección de atributos de mayor relevancia.

Una aplicación de conjuntos aproximados se realizó en un ambiente educativo:

- “Un ejercicio de análisis utilizando conjuntos aproximados en un dominio de educación superior mediante el proceso (KLD)”. Estudio sobre los recién titulados de la Carrera de Ingeniería en Sistemas Computacionales del Instituto Tecnológico de Chihuahua II, en México. Mediante la aplicación de conjuntos aproximados se descubrió que existían cuatro variables que determinaban la adecuada inserción laboral, que son citadas de acuerdo con su importancia: zona económica donde habitaba el estudiante, colegio de donde provenía, nota al ingresar y promedio final al salir de la carrera. A partir de estos resultados, la universidad tendrá que hacer un estudio socioeconómico sobre grupos de alumnos que pertenecían a las clases económicas bajas para dar posibles soluciones, debido a que tres de las cuatro variables no dependían de la universidad (Rodas, 2001).

1.2.4 Redes neuronales artificiales

Desde que McCulloch y Pitts (1949) propusieron un modelo simplificado de la actividad nerviosa real en el que cada neurona de una red neuronal podría activarse o desactivarse en función de lo que hicieran las neuronas conectadas a ella, la investigación acerca de redes neuronales artificiales se ha incrementado notablemente.

Se han realizado aplicaciones de las redes neuronales en materia de predicción:

- “Un modelo para la predicción de recidiva de paciente operados de cáncer de mama (CMO) en redes neuronales”, donde se construye un modelo de predicción de recidiva en pacientes que han sido operado de cáncer de mama para el diagnóstico y la planificación del tratamiento que hay que realizar (Gómez Jerez y Muñiz, 2000).
- “Aplicabilidad de redes neuronales artificiales para la predicción de los resultado individuales de la cirugía cardiaca”, donde se estudia la aplicabilidad de redes neuronales artificiales para pronosticar los resultados quirúrgicos individuales y por grupos de riesgo de la cirugía cardiaca (Borracci y Rubio, 2003).
- “Predicción de series temporales con redes neuronales: una aplicación a la inflación colombiana” donde se muestra una aplicación de pronóstico de valores futuros de la serie de inflación colombiana (Santana, 2006).

1.2.5 Trabajos de minería de datos en la Carrera de Informática

“Minería de uso de la Web mediante el análisis de transacciones de acceso a un servidor Web”, tesis de grado por Wendy Laura Sotomayor Segales. Donde se realiza un estudio sobre transacciones de acceso a un servidor Web (Sotomayor, 2006).

“Minería de datos para la predicción de grupos poblacionales”, tesis de grado realizada por Claudia Carola Calderón Yujra. Donde se desarrolla un producto

software basado en un modelo predictivo de minería de datos para el servicio de la educación pública boliviana (Calderón, 2006).

“Desarrollo de un modelo de minería de datos académico”, tesis de grado realizada por Edwin Percy Marqués Granado. Donde se desarrolla un modelo de minería de datos basado en redes neuronales de predicción en un ambiente académico (Marqués, 2006).

Los trabajos citados muestran la aplicación de técnicas de minería de datos como redes neuronales artificiales y conjuntos aproximados en la construcción de modelos de predicción. Los mismos servirán como referencia para un mejor modelado del modelo propuesto.

1.3 Planteamiento del problema

1.3.1 Problema general

La presente investigación centra su atención en la información almacenada en bases de datos de centros de salud en Bolivia a los que la ONG Medicus Mundi presta servicio. Información de pacientes que acuden al centro de salud, síntomas y tratamiento de enfermedades. Estos datos no se utilizan para identificar patrones o información novedosa.

La información sobre enfermedades con alto índice de prevalencia es muy importante para Medicus Mundi pues según esta información (recolectada en centros de salud) se realizan campañas de prevención, asignación de presupuesto y estudios médicos y sociales. Actualmente Medicus Mundi carece de un modelo de predicción de prevalencia de enfermedades para centros de salud de Bolivia.

1.3.2 Problemas específicos

- No se utilizan datos almacenados en bases de datos de centros de salud de Bolivia para identificar patrones e información novedosa.
- Se desconocen variables relevantes que influyan en la presencia de enfermedades.

- Se desconocen variables relevantes que contribuyan al tratamiento de enfermedades.
- Se desconoce el comportamiento de prevalencia de enfermedades en centros de salud de Bolivia.
- No se realiza un estudio de prevalencia analizando la ubicación del centro de salud.

1.4 Planteamiento de la hipótesis

El modelo de predicción obtiene al menos 70% de precisión⁶ para la predicción de prevalencia para las diez enfermedades con mayor índice de prevalencia en centros de salud de Bolivia⁷.

1.5 Objetivos

1.5.1 Objetivo general

Desarrollar un modelo de predicción de prevalencia de enfermedades para centros de salud de Bolivia utilizando técnicas de minería de datos como conjuntos aproximados y redes neuronales artificiales.

1.5.2 Objetivos específicos

- Utilizar la minería de datos para identificar patrones e información novedosa en bases de datos de centros de salud de Bolivia.
- Utilizar la técnica de conjuntos aproximados para identificar variables relevantes que influyen en la presencia de enfermedades.
- Utilizar la técnica de conjuntos aproximados para identificar variables relevantes que contribuyen al tratamiento de enfermedades.
- Estudiar el comportamiento de enfermedades a través de un modelo de redes neuronales artificiales respecto a variables relevantes identificadas para predicción de prevalencia de enfermedades.
- Realizar un estudio de prevalencia de enfermedades tomando en cuenta la ubicación del centro de salud analizando sus tasas de prevalencia.

⁶ Según Olson y Delen (2008) la precisión de un modelo es: $precisión = \frac{TP}{TP + FP}$ donde TP es el número de valores correctamente predichos y FP es el número de valores incorrectamente predichos.

⁷ Se realizará con los datos almacenados en bases de datos de centros de salud de Medicus Mundi.

- Aplicar métodos adecuados de la ingeniería del software para la construcción y prueba de un prototipo software para la simulación del modelo de predicción de prevalencia propuesto.

1.6 Justificación

1.6.1 Justificación científica

En los últimos años la Inteligencia Artificial se ha tornado interesante por los conocimientos que alberga y las soluciones novedosas que propone a problemas que con técnicas, métodos, modelos y algoritmos tradicionales son muy difíciles o no se han podido resolver. La propuesta de construir un modelo de bases de predicción para prevalencia de enfermedades es importante en el sentido de que se lo construye con el aporte de dos ramas de la inteligencia artificial: los conjuntos aproximados, las redes neuronales artificiales aplicadas como funciones y la minería de datos.

Además por la identificación de variables relevantes que influyen en la presencia y variables que contribuyen en el tratamiento de enfermedades se permitirá realizar estudios médicos, sociales y económicos futuros.

1.6.2 Justificación técnica

La construcción del modelo y del prototipo software de aplicación para la predicción de prevalencia de enfermedades es importante pues será útil en instituciones de salud para la planificación de campañas de prevención y asignación de recursos económicos apoyando al desarrollo de las mismas.

1.6.3 Justificación social

El modelo de predicción de prevalencia de enfermedades colaborará para una mejor prevención de enfermedades, asignación de presupuesto en centros de salud y motivará futuros estudios sobre variables relevantes relacionadas con las enfermedades brindando a la sociedad un mejor servicio en los centros de salud de Bolivia.

1.7 Métodos y técnicas

Se utilizará la metodología de investigación científica deductiva, siguiendo los lineamientos de Mario Bunge (1997):

1. Observación. Se identificará los índices de prevalencia de enfermedades en centros de salud de Bolivia observando a priori el comportamiento y algunas variables que influyan la prevalencia de las mismas.
2. Hipótesis. Se plantea la siguiente hipótesis: “El modelo de predicción obtiene al menos 70% de precisión para la predicción de prevalencia en diez enfermedades con mayor índice de prevalencia en centros de salud de Bolivia”.
3. Diseño de la aplicación: Se construirá un modelo predictivo basado en minería de datos con las técnicas de conjuntos aproximados y redes neuronales artificiales en dos fases:
 - Identificación de variables relevantes para la presencia y tratamiento de enfermedades a través de conjuntos aproximados.
 - Construcción de una red neuronal artificial para la predicción de prevalencia de enfermedades respecto a las variables relevantes identificadas.
4. Casos de prueba. Se realizará mediante un subconjunto no utilizado en la construcción del modelo, realizando una prueba de caja negra bajo principios de la ingeniería de software tomando a la precisión de predicción como parámetro.
5. Conclusiones. Se redactará un informe final de investigación acerca de la aplicación de minería de datos para la construcción del modelo propuesto de predicción para la prevalencia de enfermedades de centros de salud de Bolivia.

1.8 Alcance y aporte

El alcance es:

- El desarrollo de un modelo de predicción de prevalencia de enfermedades construido en base a técnicas de minería de datos: conjuntos aproximados y redes neuronales artificiales.

El aporte será:

- Un prototipo software de utilidad computacional basado en minería de datos para la predicción de prevalencia de enfermedades.

Jaws PDF Creator

EVALUATION
VALUTAZIONE
EVALUATION
EVALUACIÓN
EVALUATION

CAPÍTULO 2

MARCO TEÓRICO

Jaws PDF Creator

Resumen

Donde se revisan conceptos de epidemiología, morbilidad, prevalencia y vigilancia epidemiológica en principio, luego aspectos relacionados con el sistema de codificación CIE10, importante para el desarrollo de la tesis ya que las bases de datos que se estudiarán están referidas a ese código de enfermedades internacional. Posteriormente se introduce la teoría de conjuntos aproximados y funcionamiento de redes neuronales. Terminando el capítulo con conceptos del proceso de desarrollo incremental del software.

2.1 Epidemiología

La epidemiología es, en la acepción más común, el "estudio de las epidemias" es decir, de las "enfermedades que afectan transitoriamente a muchas personas en un sitio determinado".

Su significado deriva del griego *Epi* (sobre) *Demos* (pueblo) *Logos* (ciencia). Una definición técnica es la que propone que la epidemiología es "el estudio de la distribución y determinantes de enfermedades en poblaciones humanas".

La epidemiología investiga la distribución, frecuencia y determinantes de las condiciones de salud en las poblaciones humanas así como las modalidades y el impacto de las respuestas sociales necesarias para atender estas necesidades. La distribución, frecuencia y determinantes de la enfermedad. Las consecuencias biológicas, psicológicas y sociales de la enfermedad.

2.2 Morbilidad

Morbilidad se refiere a los efectos de una enfermedad en una población en el sentido de la proporción de personas que la padecen en un sitio y tiempo determinado. En el sentido de la epidemiología se puede ampliar al estudio y cuantificación de la presencia y efectos de una enfermedad en una población.

También es una adaptación mal empleada al español que proviene de la inglesa "morbidity" generalmente usada así en Sudamérica para identificar una condición médica en la que se estudian los padecimientos de una enfermedad y cuyo verdadero significado es 'patológico'. También para definir discapacidad, es el porcentaje de individuos que contrae una cierta enfermedad en una población.

2.3 Prevalencia

La prevalencia es la frecuencia estadística (absoluta o relativa) con la que se presenta un determinado fenómeno en una población (entendido este término en su sentido estadístico). En epidemiología los fenómenos a los que suele aplicarse este concepto son enfermedades y la población es humana con la

peculiaridad y las derivaciones dadas por su aplicación a conjuntos de seres humanos y enfermedades o fenómenos relacionados con la salud.

La prevalencia se define como el número de casos de una enfermedad o evento en una población en un momento dado. Existen dos tipos de prevalencia: prevalencia puntual y prevalencia de periodo.

2.3.1 Prevalencia puntual o instantánea

Denominada también tasa de prevalencia, proporción de prevalencia o casos prevalentes.

Se habla de prevalencia de punto, puntual o instantánea para referirse a dos cosas:

- Al número o a la frecuencia absoluta de personas que presentan una cierta característica (normalmente una enfermedad) en un instante dado.
- A la proporción o a la frecuencia relativa de individuos de la población que presentan dicha característica.

La prevalencia puntual es la frecuencia de una enfermedad o condición en un punto del tiempo. Es una proporción que expresa la probabilidad de que una persona sea un caso en un momento o edad determinados.

La prevalencia puntual se estima con la siguiente fórmula:

$$\text{Prevalencia Puntual} = \frac{\text{Número de casos nuevos y antiguos}}{\text{Población en estudio}}$$

Es la medida estimada en las llamadas encuestas de prevalencia o transversales.

2.3.2 Prevalencia de periodo

Algunos autores la denominan prevalencia periódica o prevalencia lápsica.

La prevalencia de período se refiere al número o la proporción del total de población que habría presentado ese fenómeno o característica en algún

momento durante el período considerado. Expresa la probabilidad de que un individuo sea un caso en cualquier momento de un determinado periodo de tiempo.

$$\text{Prevalencia de periodo} = \frac{\text{Número de casos prevalentes} + \text{Número de casos nuevos}}{\text{Población en estudio}}$$

2.3.3 Dimensiones de la prevalencia

Como proporción normalmente expresada en tanto por ciento o en cantidad de casos (o sea, personas enfermas) la prevalencia es una magnitud adimensional. La prevalencia de período es también una proporción ya que el tiempo no se integra como magnitud en ella y, por tanto, también es adimensional.

2.4 Vigilancia Epidemiológica

El concepto de vigilancia epidemiológica ha evolucionado en el tiempo, inicialmente vigilancia significaba la observación cuidadosa de personas expuestas a enfermedades transmisibles para detectar síntomas tempranos y establecer aislamiento y medidas de control precoces. Luego se introdujo el concepto de vigilancia epidemiológica como seguimiento de la ocurrencia de las enfermedades en las poblaciones.

Posteriormente se adoptó el concepto de vigilancia colectiva y se adicionaron sus tres características básicas: recolección sistemática de datos; consolidación y análisis de los datos recolectados / diseminación de la información por medio de reportes epidemiológicos narrativos.

La nueva generación de la vigilancia epidemiológica, hace énfasis en el hecho de que la vigilancia tenga como objetivo principal realizar el seguimiento de los trazadores de riesgo, entre ellos el comportamiento.

2.4.1 Enfermedades sujetas a notificación

El Ministerio de Salud y Deportes de la república de Bolivia, ha seleccionado un grupo de enfermedades que están sujetas a notificación obligatoria. Esta

selección se ha realizado en base a las normas internacionales, que permiten priorizar las enfermedades en base a los siguientes criterios:

- Magnitud. Tiene alta incidencia o prevalencia.
- Trascendencia. Tienen importante letalidad y/o repercusión social y económica.
- Vulnerabilidad. Pueden modificarse a través de intervenciones de los servicios de salud.
- Costo. Tienen costo para los servicios de salud y la sociedad en su conjunto.
- Vigilancia Internacional. Están sujetas a reglamentos sanitarios internacionales.

La notificación epidemiológica se realiza en base al reporte de un caso sospechoso de una patología bajo vigilancia. El proceso de notificación forma parte importante de la vigilancia epidemiológica y se realiza por el personal de salud a un nivel inmediato superior.

La notificación debe realizarse en forma inmediata o semanal, en dependencia del potencial epidémico de las enfermedades y las condiciones de control que requieran.

Son responsables de la notificación tanto el personal de salud de los establecimientos pertenecientes a los subsectores: público, seguros sociales, iglesias, ONG's y privado, en los distintos niveles de atención, de acuerdo a las funciones establecidas.

2.4.2 Enfermedades de notificación inmediata

Las enfermedades y eventos de notificación inmediata, constituyen enfermedades o daños con alto potencial epidémico, que están sujetas a vigilancia internacional o que requieren de actividades de control epidemiológico rápido. En este grupo de enfermedades o daños se pueden incluir aquellas que no tienen antecedentes de presencia en el país y pueden resultar de riesgo epidémico, o que si bien tiene registro de casos, su aparición no es frecuente.

Todo caso sospechoso de estas enfermedades o daños debe ser notificado inmediatamente a los niveles superiores establecidos en el sistema. Una cifra acumulada de estos casos se reporta en forma semanal.

Además de las enfermedades todo evento agudo o brote debe ser comunicado bajo el mismo procedimiento, para contemplar posteriormente la información recolectada.

Entre las dispuestas por el ministerio se encuentran: sarampión, rubéola, tos ferina, difteria, parálisis flácida aguda, fiebre amarilla, dengue hemorrágico, dengue clásico, fiebre hemorrágica boliviana, peste, meningocócica, cólera, enfermedades por hanta virus, rabia humana.

2.4.3 Enfermedades de notificación semanal

Son patologías o condiciones no incluidas en la categoría anterior, debido a que por su potencial epidémico o carácter endémico, no requieren de una notificación inmediata ya que no se realizan actividades especiales de control epidemiológico rápido, permitiendo obtener información que contribuya a la adecuación de los servicios para el enfrentamiento apropiado durante fases epidémicas.

Entre ellas se encuentran: Tétanos neonatal, tétanos, hepatitis viral, parotiditis, úlcera genital, flujo uretral, VIH/SIDA, enfermedad diarreica aguda, IRA sin neumonía, neumonía, malaria, leishmaniosis, entre otras.

2.5 Fuentes de notificación

Todos los establecimientos que conforman el Sistema de Salud Nacional de Salud, pertenecientes al subsector público, los seguros sociales, iglesias y organizaciones no gubernamentales, se constituyen en fuentes fundamentales de notificación, los mismos se han organizados en espacios geográficos de acuerdo a capacidad resolutiva constituyendo las Redes de Servicios de Salud.

Toda la red de notificación establecida en un municipio o mancomunidad de municipios está coordinada por la gerencia de red correspondiente. Las funciones a realizar para el subsistema de vigilancia epidemiológica en los establecimientos de salud y niveles gerenciales son los siguientes:

2.5.1 Puestos normales

- Reciben notificación sintomática de la comunidad o demanda espontánea de la población, verifican, e incorporan los datos en su reporte inmediato / semanal.
- Si es enfermedad de notificación inmediata, utilizan el medio disponible más rápido para la notificación al nivel inmediatamente superior (Gerencia de red) e inician la investigación con la ficha epidemiológica.
- Realizan toma de muestra y envío al laboratorio de referencia con medidas de control con participación comunitaria, solicitan apoyo al nivel inmediato superior para la dotación de insumos y otras necesidades.
- Realizan análisis de la información con participación comunitaria, para comunicar sobre los problemas detectados y las acciones realizadas.

2.5.2 Centros de salud

- Reciben notificación sintomática de la comunidad o demanda espontánea de la población, verifican síntomas e incorporan el dato en su reporte inmediato/semanal.
- Atienden por demanda espontánea y diagnostican casos por sospecha clínica, incorporan en su notificación inmediata / semanal.
- Si es enfermedad de notificación inmediata, utilizan el medio disponible más rápido para la notificación al nivel inmediatamente superior (Gerencia de Red).
- Realizan la investigación epidemiológica preliminar de casos sospechosos de enfermedades que lo requieren (ficha epidemiológica), incluyendo la adecuada toma de muestras para examen laboratorio.
- Toman medidas de control con participación comunitaria

- Realizan análisis de la información con participación del personal y participación comunitaria para comunicar sobre los problemas detectados y las acciones realizadas.

2.5.3 Hospitales de segundo y tercer nivel

- Notifican inmediata / semanalmente casos sospechosos objeto de notificación obligatoria, detectados en los servicios de su dependencia, por el medio más rápido.
- Realizan la investigación epidemiológica preliminar de casos sospechosos de enfermedades que lo requieran (ficha epidemiológica), incluyendo la adecuada toma de muestras para examen laboratorial.
- De acuerdo a capacidad resolutive confirman en laboratorio los casos sospechosos notificados o refieren a otro laboratorio de acuerdo a lo establecido en la ley.
- Toman medidas de control, eventualmente con apoyo de la Gerencia de Red.
- Realizan análisis de la información del ámbito del establecimiento.
- Participan en las reuniones de análisis de la situación de salud convocados por la Gerencia de Red.

2.5.4 Gerencia de red

- Reciben notificación inmediata / semanal de casos sospechosos atendidos en los establecimientos de salud
- Transmiten información inmediata / semanal al SMS - SEDES
- Brindan apoyo a los establecimientos de la red para realizar la investigación epidemiológica de casos sospechosos que lo requieran.
- Recomiendan y apoyan la implementación de medidas de control.
- Realizan procesamiento, control de calidad, análisis y comunicación de la información.
- Organizan y dirigen las reuniones de análisis de la situación de salud.
- Supervisan los establecimientos notificantes y coordinan con la red de laboratorios
- Monitorean y evalúan las medidas de control aplicadas.

- Organizan y desarrollan programas de capacitación continua.

2.5.5 SNIS (Servicio nacional de Información en salud) – SEDES (Servicio departamental de salud)

- Reciben notificación inmediata /semanal de casos sospechosos
- Transmiten información inmediata /semanal al nivel nacional.
- Brindan información inmediata a los programas del SEDES para el inicio de acciones de control.
- Reciben información de los programas, sobre la implementación de medidas de control.
- Realizan procesamiento, control de calidad, análisis y retroalimentación de la información.
- Organizan y dirigen las reuniones de análisis de la situación de salud con participación de las Gerencias de Red.
- Supervisan las Gerencias de Red y establecimientos de mayor complejidad.
- Organizan y desarrollan programas de capacitación continua.
- Adecúan normas nacionales y elaboran normas regionales para el manejo del subsistema.

2.5.6 SNIS nacional (Sistema Nacional de Información en Salud)

- Recibe información inmediata/semanal de los SEDES y mantiene comunicación permanente.
- Realizan el seguimiento de casos notificados por vía rápida y brinda información permanente a los programas.
- Realiza el control de calidad, análisis de la información semanal y retroalimentación a los programas y SEDES.
- Realiza el análisis de las enfermedades endémicas de interés para las regiones y retroalimenta a los SEDES.
- Elabora informes generales y particulares.
- Elabora y difunde boletines con la información de vigilancia epidemiológica

- Organiza y dirige las reuniones de análisis de situación de salud a nivel nacional.
- Supervisa el SNIS a nivel del SEDES
- Monitorea y evalúa el subsistema de vigilancia epidemiológica.
- Elabora y difunde normas y procedimientos para el subsistema de vigilancia epidemiológica.
- Coordina con los programas a nivel nacional.

Informa a los organismos y países sobre los casos sujetos a vigilancia internacional.

2.6 Clasificación Internacional de Enfermedades 10 (CIE 10)

2.6.1 Definición

Una clasificación de enfermedades se define como un sistema de categorías a las cuales se le asignan entidades morbosas de acuerdo con criterios establecidos.

2.6.2 Propósito

Permitir el registro sistemático, el análisis, la interpretación y la comparación de los datos de mortalidad y morbilidad recolectados en diferentes países o áreas y en diferentes épocas.

2.6.3 Utilidad

Convertir los términos diagnósticos y de otros problemas de salud, de palabras a códigos alfanuméricos que permiten su fácil almacenamiento y posterior recuperación para el análisis de la información. En la práctica se ha convertido en una clasificación diagnóstica estándar internacional para todos los propósitos epidemiológicos generales y muchos otros de administración de salud. El CIE, no es adecuado para indagar entidades clínicas individuales y es adecuada para clasificar enfermedades y otros tipos de problemas de salud, consignados en distintos tipos de registros vitales y de salud.

2.6.4 Estructura básica de la CIE 10ª Revisión

La CIE es un sistema de clasificación de ejes variables cuyo esquema debe servir a todos los propósitos prácticos y epidemiológicos. Este patrón puede ser identificado en los capítulos de la CIE y hasta el momento es considerado como la estructura más útil que cualquiera de las alternativas que se han probado.

La CIE utiliza un código alfanumérico, con una letra en la 1ª posición y números en la 2ª, 3ª, y 4ª posición; el cuarto carácter sigue a un punto decimal, los códigos posibles van por lo tanto de **A00.0 a Z99.9**⁸.

2.7 Minería de datos

La minería de datos es un proceso no trivial de identificación válida, novedosa, potencialmente útil y entendible de relaciones comprensibles que se encuentran ocultos en los datos según (Fayyad et al., 1996).

Según Valcárcel (2004) la aplicación de técnicas de minería de datos persiguen los siguientes resultados:

- Clasificación: Se trata de obtener un modelo que permita asignar un caso de clase desconocida a una clase concreta.
- Regresión: Se persigue la obtención de un modelo que permita predecir el valor numérico de alguna variable.
- Agrupamiento⁹: Hace corresponder cada caso a una clase, con la peculiaridad de que las clases se obtienen directamente de los datos de entrada utilizando medidas de similitud.
- Resumen: Se obtienen representaciones compactas para subconjuntos de los datos de entrada.
- Modelado de dependencias: Se obtienen descripciones de dependencias existentes entre variables.
- Análisis de secuencias: Se intenta modelar la evolución temporal de alguna variable, con fines descriptivos o predictivos.

⁸ Ver anexo I

⁹ "clustering"

2.8 Conjuntos Aproximados

La teoría de conjuntos aproximados al igual que la teoría de conjuntos difusos es una alternativa a la teoría clásica de conjuntos, una aplicación de ésta es la obtención de atributos más relevantes (denominados reductos) y a través de ellos reglas de decisión.

2.8.1 Sistema de Información

Un sistema de información es un par (U, A) . Donde U es un conjunto no vacío de objetos y A es un conjunto no vacío de atributos, donde:

$$a: U \rightarrow V_a$$

Para cada atributo $a \in A$, V_a es llamado conjunto de valores de a .

2.8.2 Sistema de decisión (S)

Está definido como $(U, A \cup \{d\})$. Donde el atributo $d \notin A$ es llamado *atributo de decisión*, en vez de considerar solamente uno se pueden considerar más atributos. Y los elementos de A son llamados *atributos de condición*.

2.8.3 Función de información

A cada atributo a_i se le asocia un dominio v_j .

Se tiene una función $f: U \times A \rightarrow V$ $V = \{v_1, v_2, \dots, v_p\}$ tal que $f(x, a_i) \in v_j$ para cada $a_i \in A, x \in U$, llamada función de información.

2.8.4 Indiscernibilidad

La indiscernibilidad es la característica principal de esta teoría. Los criterios para considerar que dos objetos son iguales o indiscernibles bajo una cantidad finita de atributos son:

2.8.4.1 Relación de equivalencia

Es una relación binaria $R \subseteq X \times X$ la cual es reflexiva (xRx para todo objeto x), simétrica (si xRy entonces yRx), y transitiva (si xRy y yRz entonces xRz).

2.8.4.2 Clase de equivalencia

Las clases de equivalencia $[x]_R$ de un elemento $x \in X$ consiste en todos los objetos $y \in X$ tal que cumplan xRy .

2.8.5 Inseparabilidad

Dado el sistema de información (U, A) y dado un subconjunto $B \subseteq A$ se le asocia una relación binaria de inseparabilidad denotada por R , la cual es el conjunto de pares de objetos que son inseparables unos de otros por esta relación:

$$R = \{(x, y) \in U \times U : f(x, a_i) = f(y, a_i) \forall a_i \in B\}$$

$$IND_S(B) = \{(x, y) \in U^2 \mid \forall a \in B, a(x) = a(y)\}$$

Si $(x, y) \in IND_{SI}(B)$, entonces x y y son indiscernibles el uno del otro con los atributos que pertenecen a B .

Las clases de B -relación de indiscernibilidad son denotadas por $[x]_B$.

Además:

- Una relación de equivalencia siempre induce una partición del universo.
- Las particiones pueden ser usadas para construir nuevos conjuntos del universo.

2.8.6 Aproximación superior e inferior a un conjunto

Dado un sistema de información $T = (U, A)$ y dado $B \subseteq A$ y $X \subseteq U$. Podemos aproximar X usando solo la información contenida en B construyendo una B -aproximación inferior y una B -aproximación superior de X denotadas por $\underline{B}X$ y $\overline{B}X$ respectivamente, donde:

$$\underline{B}X = \{x|[x]_B \subseteq X\}$$

$$\overline{B}X = \{x|[x]_B \cap X \neq \emptyset\}$$

2.8.7 Región límite

La región límite está definida por:

$$BN_B = \overline{B}X - \underline{B}X$$

Que consiste de los objetos que no pueden ser clasificados dentro de X según B .

2.8.8 Región exterior

La región exterior de X según B está definida por

$$RE = U - \overline{B}X$$

Que consiste en todos los objetos que pueden ser clasificados con certeza como no pertenecientes a X .

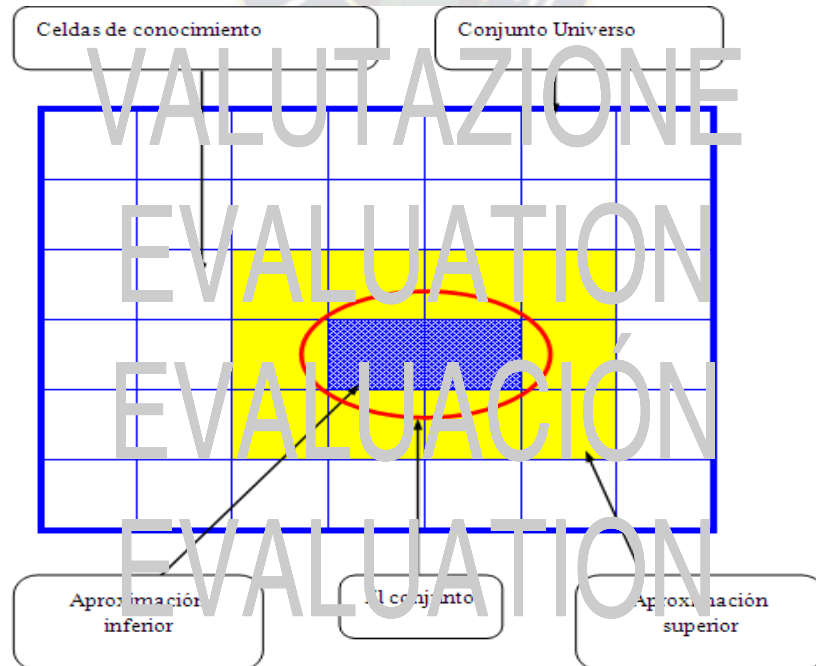


Figura 3. Definición de un conjunto aproximado
Fuente: (Pawlak, 1998)

2.8.9 Precisión de la aproximación

Se puede expresar numéricamente la precisión de la aproximación a un conjunto X , donde $X \neq \phi$:

$$\alpha_B(X) = \frac{|BX|}{|\overline{BX}|}$$

Donde $|X|$ denota la cardinalidad de $X \neq \phi$. Obviamente $0 \leq \alpha_B \leq 1$.

Si $\alpha_B(X) = 1$, X es llamado *duro* con respecto a B .

Si $\alpha_B(X) < 1$, X es llamado *aproximado* con respecto a B .

2.8.10 Calidad de la aproximación

Definido por

$$\gamma(X) = \frac{|BX|}{|X|}$$

Que expresa la proporción de objetos que pueden ser correctamente clasificados en la clase X . Además $0 \leq \alpha_B(X) \leq \gamma(X) \leq 1$ y $\gamma(X) = 0$ si y solo si $\alpha(X) = 1$.

2.8.11 Calidad de la clasificación

Este coeficiente expresa la inexactitud de las clasificaciones aproximadas:

$$\Gamma(DS) = \frac{\sum_{i=1}^l |BX_i|}{|U|}$$

Que expresa la cantidad de objetos que pueden estar correctamente clasificados dentro del sistema. Si es igual a 1, el sistema de decisión es consistente, en otros casos es inconsistente.

2.8.12 Atributos dispensables e indispensables

Dado $c \in C$, el atributo c es dispensable en el sistema de decisión si $POS_c(D) = POS_{C-\{c\}}(D)$, de otro modo el atributo c es indispensable en el sistema de decisión.

2.8.13 Región positiva

La C -región positiva de D está definida por:

$$POS_C(D) = \bigcup_{x \in U/D} \underline{C}x$$

2.8.14 Independencia

Sea T un sistema de decisión $T = (U, C, D)$ e independiente si todos los atributos $c \in C$ son indispensables en T .

2.8.15 Reducto

El conjunto de atributos $R \subseteq C$ es llamado un reducto de C si $T' = (U, R, D)$ es independiente y $POS_R(D) = POS_C(D)$.

2.8.16 Núcleo

El conjunto de todos los atributos indispensables en T es denotado por $CORE(C)$.

$$CORE(C) = \bigcap RED(C)$$

Donde $RED(C)$ es el conjunto de todos los reductos de C .

2.9 Redes Neuronales Artificiales

Una red neuronal artificial, intenta ser la representación matemática de una red neuronal biológica. Se dice intenta ser, porque dada la complejidad todavía no resuelta del funcionamiento del cerebro, todos los trabajos hasta ahora son muy burdos en comparación de lo que realmente es.

Las redes neuronales se caracterizan por:

- Adquirir conocimiento a través de la experiencia.
- Tienen una gran adaptabilidad, pueden cambiar dinámicamente junto al medio en que se encuentran.
- Poseen un alto nivel de tolerancia a fallas.

Los modelos de redes neuronales artificiales son especificados por tres características.

Topología de la red, que hace referencia al lugar donde se describe la estructura y los tipos de enlaces.

- Características de los nodos, que se refiere al modelo de la neurona.
- Reglas de aprendizaje, que hace referencia al método de ajuste de los pesos.

Se mostrarán dos modelos de redes neuronales artificiales, el perceptrón y la retropropagación.

2.9.1 Perceptrón

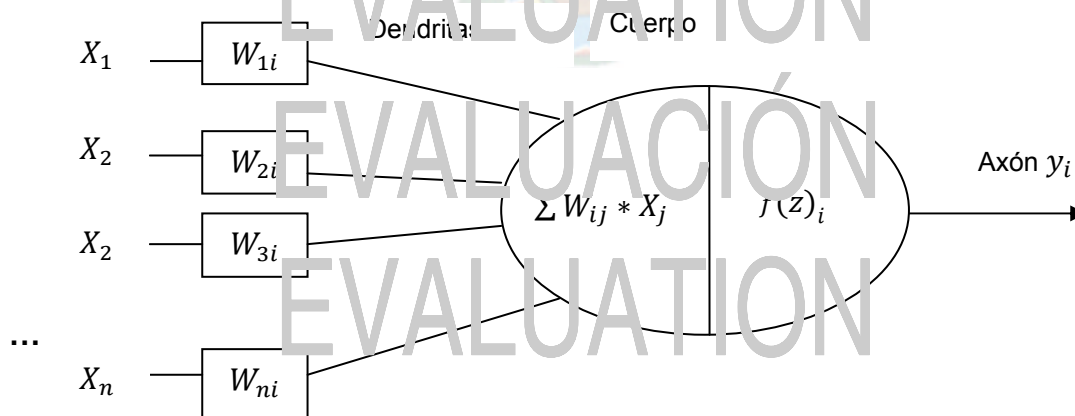


Figura 4. Perceptrón
Fuente: (Choque, 2002)

El cuerpo de la neurona se representa como un sumador lineal de los estímulos externos z_j , seguida de una función no lineal $y_j = f(z)_j$. La función $f(z)_j$ es llamada función de activación, y es la función que utiliza la suma de estímulos para determinar la actividad de salida de la neurona.

El entrenamiento de un perceptrón es por medio de la regla de aprendizaje delta: para cada peso w se realiza un ajuste dW según la regla: $dW = LR (T - Y)X$. Donde LR es la razón de aprendizaje, T el valor deseado, Y el valor obtenido y X la entrada aplicada al Perceptrón.

2.9.2 Retropropagación

Si se combinan varios perceptrones en una capa tendremos una red neuronal.

En esta red de interconectar varias unidades de procesamiento en capas, las neuronas de una misma no se interconectan entre sí. Sin embargo cada neurona de una capa proporciona una entrada a las de la siguiente capa.

El algoritmo de aprendizaje proporciona una forma de entrenar una red multicapa con alineación hacia delante de acuerdo a las siguientes ecuaciones:

$$net_i = \sum_{j \in A} O_j W_{ji} \quad \forall i \in B$$

Donde A es el grupo de neuronas en una capa y B en otra. O_j es la activación para la neurona j se toman los valores de salida y alimentan a la siguiente capa a través de los pesos. Esta operación se realiza para cada neurona de la siguiente capa, produciendo un valor en la red. Este valor es la suma de todos los valores de activación en las neuronas de la capa anterior, y cada valor de red es aplicado a la siguiente ecuación, conocida como función de activación, para producir la activación de esa neurona.

$$O_{i \in B} = f(net) = \frac{1}{1 + e^{-net_i}}$$

Los errores de neuronas de salida son propagados hacia atrás, a la capa anterior, esto se calcula por:

$$\delta_i = f'(net) \sum_{j \in E} \delta_j W_{ij} \quad \forall i: i \in D$$

Donde D es el grupo de neuronas en una capa que no es de entrada y E es el grupo de neuronas de la siguiente capa. Este cálculo se repite para cada capa escondida de la red.

Después de que se ha encontrado la activación y el error asociado a cada grupo de neuronas, los pesos se actualizan, primero encontrando el valor con el que cada peso debe modificarse, esto se logra calculando:

$$\Delta W_{ij} = \eta \delta_j \cdot W_{ij} \quad \forall i, j: i \in A, j \in E$$

Donde η , conocida como la razón de aprendizaje, es una constante que controla el valor de cambio de los pesos y W_{ij} es el cambio de los pesos entre la neurona i y j. El peso cambia evaluando:

$$\Delta W_{ij,t+1} = W_{ij,t} + \Delta W_{ij}$$

2.10 Proceso unificado de desarrollo de software

El proceso unificado es un marco de desarrollo iterativo e incremental compuesto de cuatro fases denominadas inicio, elaboración, construcción y transición. Cada una de estas fases es a su vez dividida en una serie de iteraciones (la de inicio sólo consta de varias iteraciones en proyectos grandes). Estas iteraciones ofrecen como resultado un *incremento* del producto desarrollado que a su vez mejora las funcionalidades del sistema en desarrollo.

Cada una de estas iteraciones se divide a su vez en una serie de disciplinas que recuerdan a las definidas en el ciclo de vida clásico o en cascada: análisis de requisitos, diseño, implementación y prueba. Aunque todas las iteraciones suelen incluir trabajo en casi todas las disciplinas, el grado de esfuerzo dentro de cada una de ellas varía a lo largo del proyecto.

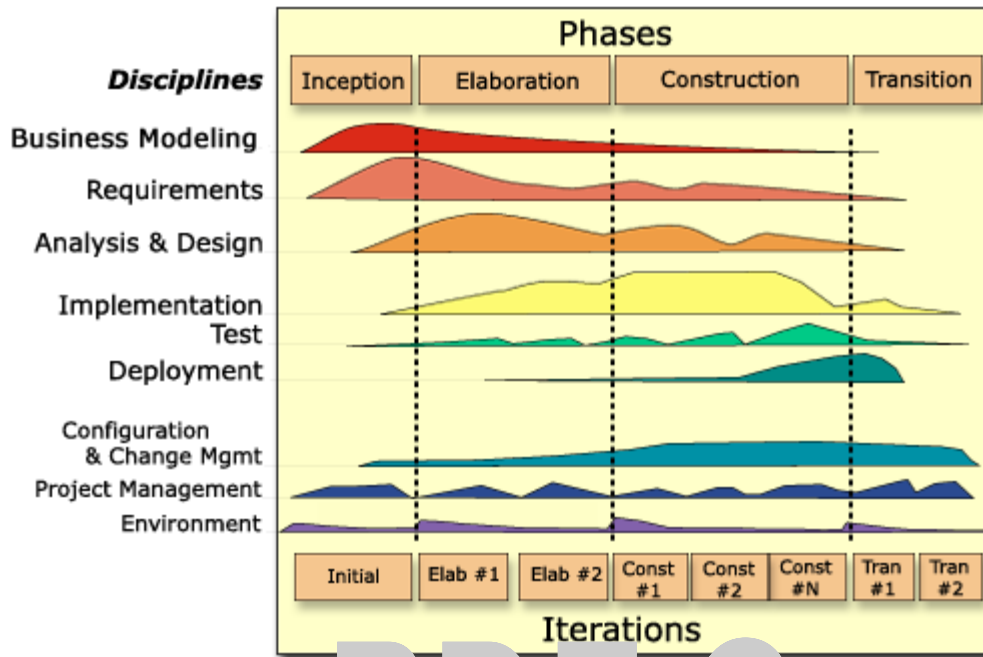


Figura 1. Fases e iteraciones del proceso unificado de desarrollo de software
Fuente: (Jacobson, Booch y Rumbaugh, 2000)

En el proceso unificado los casos de uso se utilizan para capturar los requisitos funcionales y para definir los contenidos de las iteraciones.

El proceso unificado asume que no existe un modelo único que cubra todos los aspectos del sistema. Por dicho motivo existen múltiples modelos y vistas que definen la arquitectura de software de un sistema. La analogía con la construcción es clara, cuando construyes un edificio existen diversos planos que incluyen los distintos servicios del mismo: electricidad, fontanería, etc.

El proceso unificado requiere que el equipo del proyecto se centre en identificar los riesgos críticos en una etapa temprana del ciclo de vida. Los resultados de cada iteración, en especial los de la fase de elaboración, deben ser seleccionados en un orden que asegure que los riesgos principales son considerados primero.

CAPÍTULO 3

MARCO DE MODELADO

Jaws PDF Creator

EVALUATION

Resumen

VALUTAZIONE

Donde se presenta la unificación en un modelo de predicción de los aportes de la teoría de conjuntos aproximados y las redes neuronales artificiales.

EVALUATION

Además se presenta la definición y el proceso de construcción de tablas de análisis que son el centro de estudio de la prevalencia, un algoritmo y un heurístico para hallar reductos y la descripción general de la estructura que tendrán las redes neuronales artificiales.

EVALUACIÓN

EVALUATION

El presente modelo se desarrolla en dos fases:

3.1 Fases del modelo

- La primera fase es estudiar los atributos más importantes que tienen relación con la prevalencia de una enfermedad mediante la obtención de reductos aplicando la teoría de conjuntos aproximados.
- Una vez conocidas estas variables y sus valores posibles, se construirá un modelo de predicción para prevalencia en dicha enfermedad mediante redes neuronales artificiales.

3.2 Estudio de atributos con relación a una enfermedad en una tabla de base de datos médica

Para el almacenamiento en tablas de bases de datos de información médica se tiene un conjunto de atributos y su dominio (es decir los posibles valores que puede tomar el atributo para un determinado objeto o fila). Esos atributos pueden brindar información importante acerca de una enfermedad pero procesados facilitarán el estudio de las características de la misma.

En la figura se muestra una tabla de información médica y atributos de sexo de la persona (`im_sx_codigo`), la especialidad que se requiere (`im_es_codigo`), la ocupación (`im_oc_codigo`) y el diagnóstico (`im_die_codigo1`)

im_sx_codigo	im_es_codigo	im_oc_codigo	im_cie_codigo1
Masculino	PSICOLOGIA	ORFEBRE	Trastorno de ansiedad, no especificado
Femenino	MEDICINA GENERAL	SECRETARIA	Infección de vías urinarias, sitio no especificado
Masculino	REUMATOLOGIA	SECRETARIA	Reumatismo, no especificado
Masculino	OFTALMOLOGIA	94	Examen de ojos y de la visión
Masculino	ENDOCRINOLOGIA	No Especifica	Nódulo tiroideo solitario no tóxico
Masculino	CIRUGIA	115	Causas de morbilidad desconocidas y no especificadas
Masculino	CIRUGIA MAXILO FACIAL	ARTESANO	Cuidado posterior a la cirugía, no especificado
Masculino	OTORRINOLARINGOLOGIA	No Especifica	Cerumen impactado
Masculino	OTORRINOLARINGOLOGIA	92	Cerumen impactado
Masculino	TRAUMATOLOGIA	AGRICULTOR	Traumatismo por aplastamiento de otras partes y de las no especificadas de la m...
Masculino	CIRUGIA	119	Embolia y trombosis de vena no especificada
Femenino	ENDOCRINOLOGIA	ARTESANO	Diabetes mellitus no insulino dependiente, sin mención de complicación
Masculino	PSIQUIATRIA	SECRETARIA	
Masculino	NEUMOLOGIA	RELIGIOSA	Paquipleuritis sin asbestosis
Femenino	NEFROLOGIA	MECANICO DENTAL	Lumbago no especificado
Masculino	DERMATOLOGIA	No Especifica	Fibrosis y afecciones cicatriciales de la piel
Femenino	NEUROCIURUGIA	TOPOGRAFO	
Femenino	NEUMOLOGIA	TOPOGRAFO	Neumonía, no especificada
Femenino	CARDIOLOGIA	TOPOGRAFO	
Masculino	OTORRINOLARINGOLOGIA	ORFEBRE	Traumatismo de la cabeza, no especificado
Femenino	NEUROCIURUGIA	TOPOGRAFO	Epilepsia, tipo no especificado
Masculino	NEUROCIURUGIA	No Especifica	Lumbago no especificado
Femenino	DERMATOLOGIA	SECRETARIA	Liquen simple crónico
Femenino	CARDIOLOGIA	No Especifica	Asistencia y ajuste de marcapaso cardíaco
Masculino	URLOGIA	No Especifica	Trastornos de la vejiga y de la uretra, no especificados
Masculino	OFTALMOLOGIA	No Especifica	Hemorragia conjuntival

Figura 6. Extracción de la base de datos Morbilidad de Medicus Mundi
Fuente: (Medicus Mundi, 2008)

Una ventaja de tener solo dos valores para un dominio es que pueden convertirse a números fácilmente, como 0 ó 1 y procesarlos. Pero si el dominio tiene un gran número de atributos el análisis se hace difícil. Tomando el caso del atributo sexo (im_sx_codigo) el dominio es favorable, ya que solamente es {masculino, femenino} con cardinalidad 2. Para el atributo especialidad (im_es_codigo) el dominio sería {Ginecología, Pediatría,..., Cirugía} con cardinalidad 80.

Las redes neuronales artificiales de retropropagación utilizan datos numéricos como valores de entrada, para ello es necesario refinar la información que tenemos de una tabla médica con dominios como los presentados a una con dominios numéricos.

El conocimiento que brinda cada tabla médica es importante pero se puede extender a uno más profundo realizando un estudio de dominio de atributos.

3.2.1 Estudio de dominio de atributos de una enfermedad en una tabla médica

Las características asociadas a un atributo muchas veces no están incluidas en la tabla original de estudio. Es más, pueden ser parte de otras bases de datos y contener información relevante que no se aprecia cuando se trabaja con solo los dominios de un atributo.

Una observación importante es la posibilidad de generar nuevos subconjuntos del dominio de un atributo y obtener datos numéricos de los mismos. Por ejemplo para el atributo edad se puede generar subconjuntos y obtener información importante acerca de la cardinalidad de dichos subconjuntos: menores de 4 años, de 5 a 14 años, de 15 a 49 años y la última de 50 años y más. La presente figura es una tabla de enfermedades tomando en cuenta este criterio.

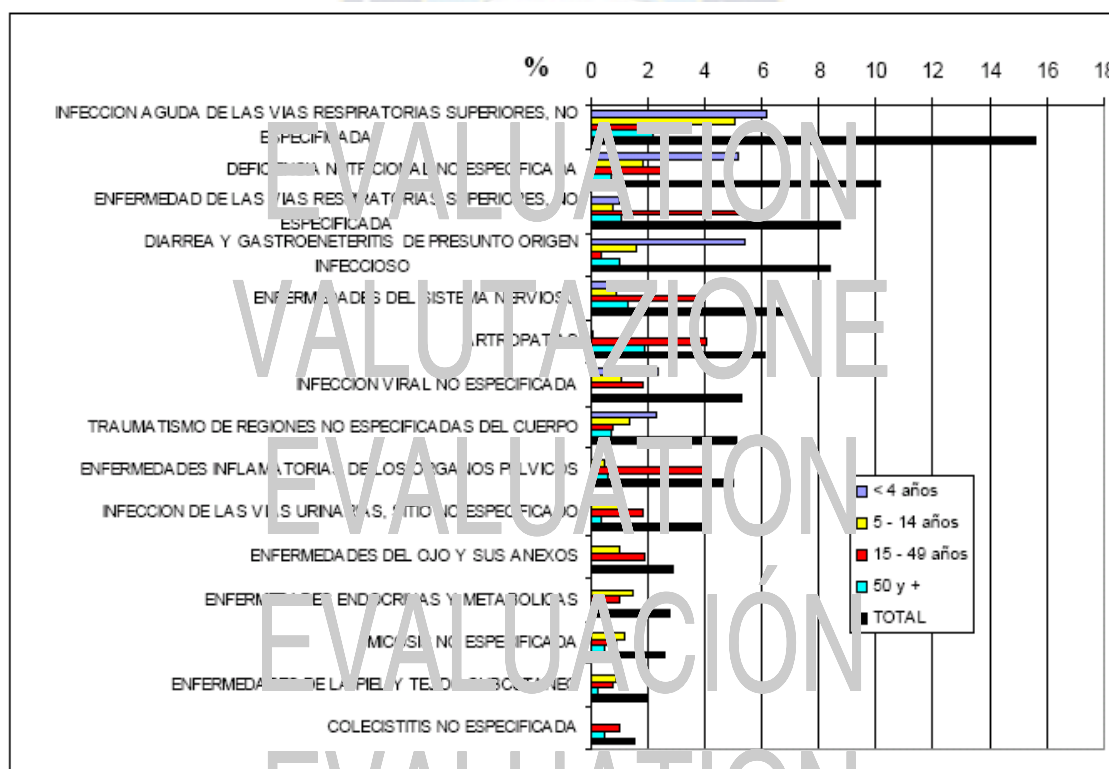


Figura 7. Principales causas de morbilidad en consulta externa área urbana
Fuente: (Ministerio de Salud y Deportes, 2004)

Datos como éstos facilitarían el trabajo con redes neuronales artificiales pues son datos numéricos.

3.2.2 Refinación de tablas médicas

3.2.2.1 Tablas de análisis

Para aplicar la Teoría de Conjuntos Aproximados es necesario un sistema de información que viene a ser la tabla que contiene atributos condición y un atributo decisión.

Los datos de dicha tabla en su mayoría son resultados de consultas, que no son almacenadas en tablas de bases de datos, volviendo a los datos de la figura 7 la consulta: ¿cuántas personas menores a un año tienen una enfermedad? tiene como resultado un número, pero dicho valor no se almacena físicamente en la base de datos. Por lo tanto es necesario almacenar datos procesados en una tabla que vendrá a ser el sistema de información requerido.

Se denominará a una tabla como la descrita tabla de análisis.

3.2.2.2 Atributos condición de una tabla de análisis

Los atributos de condición tendrán las siguientes características:

- Están relacionados con subconjuntos de un dominio de un atributo existente en una tabla de la base de datos. Es decir si un determinado dominio se puede clasificar en V subconjuntos existirán N atributos de condición para ese dominio.
- Cada atributo ' a ' de la tabla de análisis será elegido de acuerdo a un criterio proporcionado por la investigación.
- Su dominio estará determinado por números dentro del intervalo $[0,1]$.

3.2.2.3 Proporción

La relación entre un atributo de condición y un subconjunto de un determinado dominio es la proporción de individuos que presentan la característica del atributo condición a sobre el total de individuos estudiados en el dominio d en un determinado lapso de tiempo t .

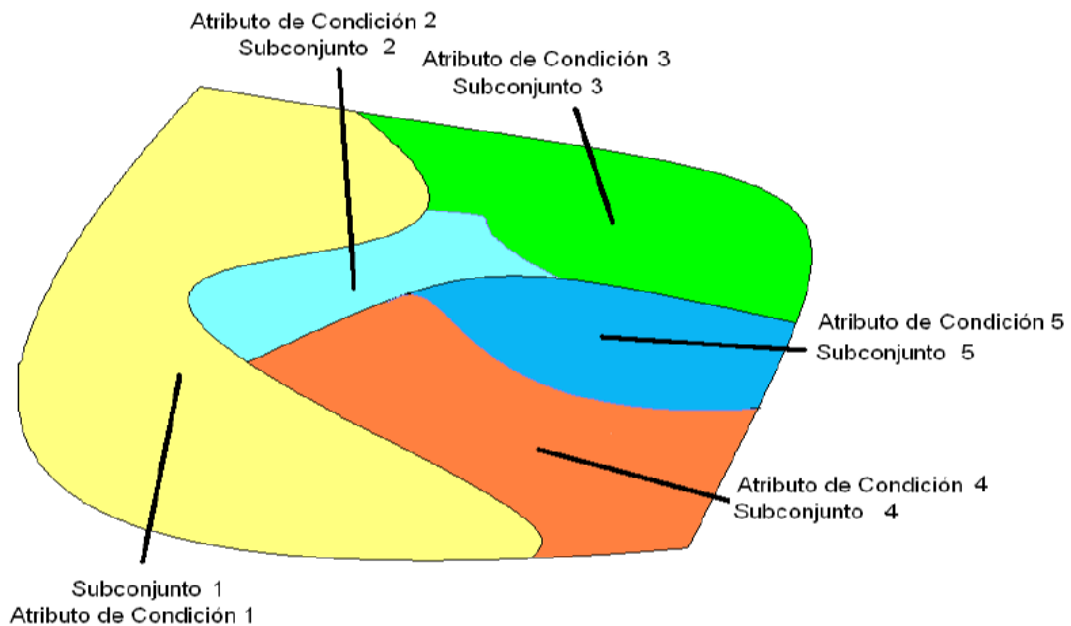


Figura 1: Subconjuntos del dominio 'd'
 Fuente: (Elaboración propia)

$$pp(a)_t = \frac{\text{Número de individuos que tienen la propiedad del atributo condición 'a'}}{\text{Número de individuos que pertenecen al dominio 'd'}}$$

Para el procesamiento de datos con conjuntos aproximados es importante tener un dominio de atributo no muy extenso. Si estudiamos a la proporción como un número real tendremos infinitos valores dentro del intervalo 0 a 1 por lo tanto es necesario expresarla en un dominio más estricto. El dominio propuesto está comprendido en un intervalo cerrado de 0 a 1 con los siguientes posibles valores:

$$dom(a) = \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$$

Para asignar a las celdas valores como los señalados se utiliza una función de redondeo para un determinado lapso de tiempo t , se asigna a t el valor de un mes:

$$f(pp(a)_t) = \text{redondear}(pp(a)_t) \text{ con precisión de un decimal}$$

3.2.2.4 Atributo decisión de una tabla de análisis

Se denominará PREVALENCIA al atributo decisión que consta de los valores de prevalencia en un determinado hospital para una enfermedad en el lapso de tiempo de un mes.

3.2.2.5 Prevalencia de enfermedades

Para el desarrollo del modelo se utilizará la prevalencia por periodo, donde el lapso de tiempo por observación será un mes. El fenómeno observado será la prevalencia de una enfermedad en pacientes de un establecimiento de salud determinado para un determinado mes.

$$prevalencia(E)_c = \frac{\text{Pacientes que padecen 'E'}}{\text{Pacientes totales dentro de 'c'}}$$

Donde: E denota la enfermedad 'c' el establecimiento de salud. $prevalencia(E)_c \leq 1$ la prevalencia que tiene la enfermedad de acuerdo al número de pacientes estudiados en un determinado establecimiento de salud, o sea c.

3.2.2.6 Valores propuestos para prevalencia

Pese a que la prevalencia de una enfermedad se encuentra contenida en el intervalo [0,1] sus valores son infinitos ya que es un número real, por lo cual son necesarios nuevos valores para facilitar el análisis de prevalencia.

Sea un conjunto P de prevalencias correspondientes a distintos meses en un lapso de tiempo. Se propone el siguiente cociente para la asignación de nuevos valores para la prevalencia. Un cociente cercano a 1 indica una prevalencia muy alta y un cociente cercano a 0 indica una prevalencia muy baja en dicho lapso de tiempo.

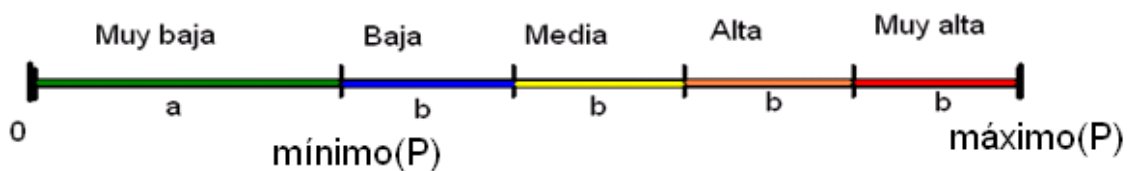


Figura 9: Intervalos de valores de prevalencia propuestos
Fuente: (Elaboración propia)

Los valores propuestos para la distancia a y b son:

$$a = \text{mínimo}(P) - 0,$$

$$b = \frac{\text{máximo}(P) - \text{mínimo}(P)}{4}$$

Los intervalos en una vista sencilla son:

$$[0, a], (a, a + b], (a + b, a + 2 * b], (a + 2 * b, a + 3 * b], (a + 3 * b, 1]$$

O bien de manera desarrollada:

Intervalo	Prevalencia
$0 \leq \text{prevalencia}(E)_c \leq \text{mínimo}(P)$	→ Muy Baja
$\text{mínimo}(P) < \text{prevalencia}(E)_c \leq \text{mínimo}(P) + \frac{\text{máximo}(P) - \text{mínimo}(P)}{4}$	→ Baja
$\text{mínimo}(P) + \frac{\text{máximo}(P) - \text{mínimo}(P)}{4} < \text{prevalencia}(E)_c \leq \text{mínimo}(P) + 2 * \left(\frac{\text{máximo}(P) - \text{mínimo}(P)}{4} \right)$	→ Media
$\text{mínimo}(P) + 2 * \left(\frac{\text{máximo}(P) - \text{mínimo}(P)}{4} \right) < \text{prevalencia}(E)_c \leq \text{mínimo}(P) + 3 * \left(\frac{\text{máximo}(P) - \text{mínimo}(P)}{4} \right)$	→ Alta
$\text{mínimo}(P) + 3 * \left(\frac{\text{máximo}(P) - \text{mínimo}(P)}{4} \right) < \text{prevalencia}(E)_c \leq 1$	→ Muy alta

3.2.2.7 Construcción de tablas de análisis

Sea la siguiente tabla como universo total de donde se construirá una tabla de análisis T :

Individuo	Edad	Sexo	Enfermedad	Mes
I1	12	Masculino	A	Enero
I2	45	Femenino	B	Febrero
I3	65	Masculino	B	Febrero
I4	4	Femenino	A	Febrero
I5	33	Femenino	A	Marzo
I6	1	Femenino	B	Abril
I7	9	Masculino	B	Abril
I8	17	Masculino	A	Mayo
I9	36	Masculino	F	Mayo
I10	34	Femenino	A	Mayo
I11	10	Femenino	A	Junio
I12	5	Masculino	A	Junio
I13	62	Femenino	B	Junio
I14	1	Femenino	A	Julio

*Tabla 1. Tabla de casos medicina tradicional
Fuente: (Elaboración propia)*

El estudio se hace para una enfermedad J , en ese caso para A .

Individuo	Edad	Sexo	Enfermedad	Mes
I1	12	Masculino	A	Enero
I4	4	Femenino	A	Febrero
I5	33	Femenino	A	Marzo
I8	17	Masculino	A	Mayo
I10	34	Femenino	A	Mayo
I11	10	Femenino	A	Junio
I12	5	Masculino	A	Junio
I14	1	Femenino	A	Julio

*Tabla 2. Recorte de tabla de datos médica tradicional para la enfermedad "A"
Fuente: (Elaboración propia)*

El procesamiento de datos se realizará tomando en cuenta el atributo “mes” para construir los intervalos de tiempo dentro de c^{10} . Para el ejemplo los valores que puede tomar “mes” son {Enero, Febrero, Marzo, Mayo, Junio y Julio}. La nueva tabla de análisis T tendrá 6 filas, una por mes.

Para generar una tabla de análisis se debe definir atributos de condición. El primer atributo a estudiar es edad. El Atributo “edad” debe estar dividido en subconjuntos, donde cada uno de ellos está relacionado con un atributo condición. Se escogió edad1: personas menores de 18 años que tienen la enfermedad y edad2: personas mayores a 18 que tienen la enfermedad. Por lo tanto tenemos los primeros atributos de condición: {edad1, edad2}.

El segundo atributo a estudiar es sexo. Con la cual se puede crear dos atributos de condición, masculino: porcentaje de varones que tiene la enfermedad y femenino: porcentaje de mujeres que tiene la enfermedad. Estos son los siguientes atributos de condición: {masculino, femenino}.

	Edad1: Menores a 18 años	Edad2: Mayores a 18 años		Masculino	Femenino
Enero	1/1=1	0	Enero	1/1=1	0
Febrero	1/1=1	0	Febrero	0	1/1
Marzo	0	1/1	Marzo	0	1/1
Mayo	$\frac{1}{2} = 0.5$	$\frac{1}{2} = 0.5$	Mayo	$\frac{1}{2} = 0.5$	$\frac{1}{2} = 0.5$
Junio	2/2=1	0	Junio	$\frac{1}{2} = 0.5$	$\frac{1}{2} = 0.5$
Julio	1/1=1	0	Julio	0	1/1

Tablas 3 y 4. Atributos de condición de la enfermedad "A"
Fuente: (Elaboración propia)

La última columna será el atributo de decisión, que indica el valor de la prevalencia. Que es el total de personas que padecen la enfermedad 'A' dividido entre el total de pacientes del establecimiento para un determinado mes.

¹⁰ c denota el centro de salud determinado

	Edad1: Menores a 18 años	Edad2: Mayores a 18 años	Masculino	Femenino	Prevalencia
Enero	1/1=1	0	1/1=1	0	1 = muy alta
Febrero	1/1=1	0	0	1/1	0.33=muy baja
Marzo	0	1/1	0	1/1	1 = muy alta
Mayo	½ = 0.5	½ = 0.5	½ = 0.5	½ = 0.5	0.66 = media
Junio	2/2=1	0	½ = 0.5	½ = 0.5	0.66 = media
Julio	1/1=1	0	0	1/1	1 = muy alta

Tabla 5. Tabla de análisis refinada para la enfermedad "A"

Fuente: (Elaboración propia)

3.3 Modelado para conjuntos aproximados

3.3.1 Sistema de Información

El sistema de información es una tabla de análisis T definida por $(U, A, \{d\})$. Donde el atributo d es el atributo de decisión, prevalencia y los atributos de condición A son extensiones de los conjuntos de atributos existentes en la base de datos. U es el conjunto de objetos o filas de T que corresponden a observaciones durante el lapso de tiempo de un mes en un determinado centro de salud c .

3.3.2 Algoritmo de reducción

Aplicando las definiciones de la teoría de conjuntos aproximados se diseñó un algoritmo capaz de encontrar reductos:

Sea C el conjunto de atributos Condición

Sea D el conjunto de atributos Prevalencia

Sea $cola$ tipo queue

$cola \leftarrow C$

$regionPositiva \leftarrow RegionPositiva(cola, D)$

for ($i \leftarrow 1, |C|$) {

$atributoX \leftarrow C.eliminar(i)$

$regionPositivaSinX \leftarrow RegionPositiva(cola, D)$

if ($regionPositivaSinX < regionPositiva$)

//El atributo es indispensable, eliminándolo la calidad de clasificación es menor

```

cola.adicionar(atributoX)

else{
//El atributo es dispensable, eliminándolo no perdemos calidad de clasificación
}
}
reducto ← cola

```

Algoritmo 1. Algoritmo propuesto para hallar reductos
Fuente: (Elaboración Propia)

El presente algoritmo permite hallar un reducto tomando en un principio todos los atributos como reducto, continúa eliminando atributos dispensables y al final obtiene un reducto mínimo.

Encontrar un reducto es de vital importancia porque a partir de él se extraen reglas de decisión para construir un modelo de predicción.

3.2.3. Reglas de decisión

Si r es un reducto de C (atributos condición de una tabla para análisis) y $\{E\}$ la enfermedad estudiada, las reglas seán de la siguiente forma:

Si:

$$C = \{a_1, a_2, a_3, a_4, a_5\}, D = \{E\}, \text{ y } r = \{a_1, a_2, a_3\} \text{ un reducto de } C$$

Observaciones	a1	a2	a3	E
o1	0.5	0.2	0.3	Alta
o2	0.1	0.3	0.6	Baja
...
op	0.8	0.2	0.1	Muy alta

Tabla 6. Reducto de una tabla para análisis según los subconjuntos u

Fuente: (Elaboración propia)

Entonces se construirán las siguientes reglas de decisión para el centro de salud por observaciones $\{o_1, o_2, \dots, o_p\}$

1) Para el subconjunto o1:

Si a_1 se presenta en un 0.5 y a_2 en un 0.2 y a_3 en un 0.3 en la observación 'o1' entonces la prevalencia para E es Alta.

2) Para el subconjunto c2:

Si a_1 se presenta en un 0.5 y a_2 en un 0.2 y a_3 en un 0.3 en la observación 'o2' entonces la prevalencia para E es Baja.

3) Para el subconjunto cm:

Si a_1 se presenta en un 0.8 y a_2 en un 0.2 y a_3 en un 0.0 en la observación 'op' entonces la prevalencia para E es Muy alta.

3.4 Red Neuronal

Una red neuronal artificial es un entrelazado o estructura formada por nodos o neuronas, que se conectan entre sí y tienen la capacidad de recibir datos de entrada y producir una salida.

Es necesario encontrar un reducto r ya que los atributos que se encuentran en el reducto serán las entradas a la red neuronal. No obstante un modelo de reglas de decisión ya está construido en base a las reglas obtenidas de r , datos sobre atributos del reducto servirán para la predicción propiamente dicha.

3.4.1 Arquitectura

La red neuronal propuesta en el presente trabajo presenta tres capas de neuronas con propagación de la información hacia adelante, están compuestas por una capa de entrada cuyas neuronas son E_i , una capa de neuronas ocultas O_i (capa de unidades de procesamiento) cuyo número de neuronas no necesariamente es igual al de la capa de entrada, y una capa de neuronas de salida, S_i , que codifican el resultado del cálculo (en esta aplicación conformada por una sola neurona que indica la prevalencia resultante de la enfermedad).

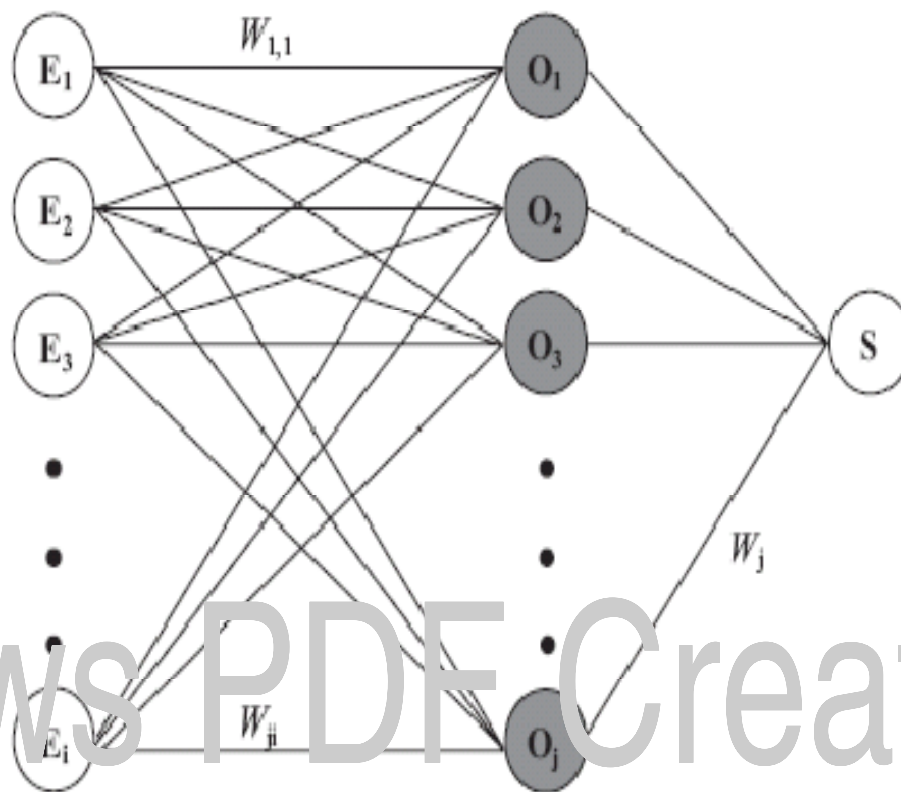


Figura 10. Esquema general de las conexiones de una red neuronal
Fuente: (Ovando, Bolo y Saizago, 2003)

3.4.1.1 Capa de entrada

El número de neuronas E_i en la capa de entrada se es $|r|$, es decir la cardinalidad del reducto r . Para ello se utilizará una tabla donde cada fila estará definida por un subconjunto u del universo y los atributos del reducto r :

3.4.1.2 Capas ocultas

No hay ninguna regla que señale el número óptimo de neuronas en la capa oculta pero el número de dichas neuronas interviene en la eficacia de aprendizaje (Hilera, 1995). Pero la idea más adecuada es utilizar la capa con menor número de neuronas. Como valor inicial para esta capa oculta se tomarán nueve neuronas, adicionando posteriormente más neuronas hasta llegar a $|r|$. En caso de ser necesarias más neuronas se adiciona otra capa oculta con 9 neuronas por principio. Este procedimiento se realizará hasta encontrar un error apropiado.

3.4.1.3 Capa de salida

Consta solamente de una neurona la cual muestra el resultado, la prevalencia de dicha enfermedad. Para esta capa sólo se dispone de una neurona que muestra el valor de la prevalencia para la enfermedad estudiada.

Y para el atributo decisión:

$$p(e)_u = \frac{\text{Pacientes que padecen 'e'}}{\text{Pacientes totales dentro de 'u'}}$$

Para la función de aprendizaje se utilizará la función *Tansig*, ya que su rango de resultado está entre -1 y 1.

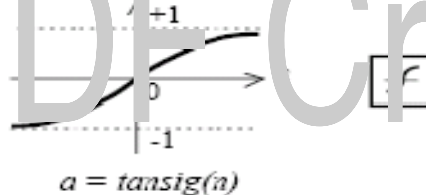


Figura 11. Función de transferencia 'Tansig'
Fuente: (Demuth y Beale, 2000)

Para la función de aprendizaje se utilizara la función *purelin*, pues el dominio de la prevalencia se ha desarrollado en un intervalo abierto a la derecha.

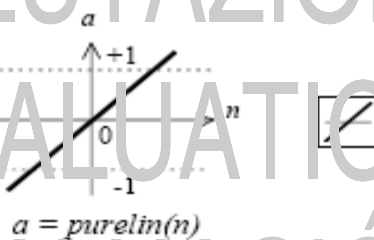


Figura 12. Función de transferencia 'Purelin'
Fuente: (Demuth y Beale, 2000)

3.5 Pruebas

Dado que del universo se pueden obtener subconjuntos que son denominados *u* y que en base a cada uno de éstos se crea una fila para la tabla de análisis *T*, se seleccionará un 70% de estos subconjuntos para la construcción del

modelo, obtención del reducto y el entrenamiento de la red neuronal y el restante 30% será para efectuar las pruebas de predicción en la enfermedad de estudio.

3.5.1.1 Datos de entrenamiento

3.5.1.2 Datos entrenamiento de atributos de condición para un reducto obtenido

Tomando en cuenta los atributos decisión pertenecientes al reducto obtenido se prepararán datos de entrenamiento, tanto para los atributos de condición como para el atributo de decisión prevalencia.

Los atributos de condición de una tabla de entrenamiento tienen como dominio solamente números enteros no negativos. Para términos sencillos las celdas muestran el número de individuos que al mismo tiempo presentan la característica c y pertenecen al subconjunto e de condición los individuos que presentan la característica del atributo de condición a y su ocurrencia fue en el establecimiento de salud c durante un determinado mes.

$$l_{a,e} = \text{Número de sujetos que tienen la propiedad } c \text{ del atributo 'a' dentro de 'e'}$$

Cuyo dominio son los enteros no negativos:

$$\text{dom}(c_i) = \{0, 1, 2, \dots\}$$

meses "m"	Atributo1: a1	...	Atributo n: an
m_1	l_{a_1,e_1}	...	l_{a_n,e_1}
m_2	l_{a_1,e_2}	...	l_{a_n,e_2}
...
m_p	l_{a_1,e_p}	...	l_{a_n,e_p}

Tabla 7. Esquema general de una tabla de datos de entrenamiento para atributos de condición en un reducto

Fuente: (Elaboración propia)

3.5.1.3 Datos de entrenamiento del atributo decisión

En este caso será la prevalencia de enfermedades que corresponde a la prevalencia de la enfermedad E sobre un conjunto c para un determinado mes (expresado en porcentaje).

$$prevalencia(E)_c = \frac{\text{Pacientes que padecen 'E'}}{\text{Pacientes totales dentro de 'c'}}$$

meses "m"	$prevalencia(E)_c$
m_1	$ E_1 / c_1 * 100$
m_2	$ E_2 / c_2 * 100$
...	.
m_p	$ E_p / c_p * 100$

Tabla 8. Esquema general de datos de entrenamiento para prevalencia
Fuente: (Elaboración propia)

El dominio de $prevalencia(E)_c$ está dado por:

$$dom(E) = \{0, 100\}$$

Se adicionará un atributo como decisión $\{E\}$ donde E está relacionado con la prevalencia de la enfermedad a ser estudiada para un determinado conjunto de pacientes c .

3.5.1.4 Datos de prueba

Al igual que los datos de entrenamiento para atributos de condición y atributo de decisión fueron definidos en el subtítulo anterior se tomarán datos de prueba en un determinado lapso de tiempo posterior al entrenamiento para probar la efectividad de la predicción del modelo.

CAPÍTULO 4

MARCO APLICATIVO

Jaws PDF Creator

EVALUATION

VALUTAZIONE

EVALUATION

EVALUACION

EVALUATION

Resumen

Donde se muestra el desarrollo de la aplicación software mediante el proceso verificable de desarrollo del software, la implementación del algoritmo y el heurístico propuesto para hallar reductos y la red neuronal artificial. Posteriormente se muestra el ajuste de datos para realizar pruebas y las 10 enfermedades con mayor prevalencia.

4.1 Desarrollo de la herramienta de predicción

Para realizar pruebas, y contrastar la hipótesis es necesario desarrollar una herramienta software. La presente solución muestra el desarrollo del sistema de predicción de prevalencia de enfermedades para establecimientos de salud tomando el modelo basado en la teoría de conjuntos aproximados y redes neuronales presentado en el capítulo anterior.

La presente herramienta se desarrolló bajo los principios del proceso unificado de desarrollo de software.

El proceso unificado de desarrollo de software consta de cuatro fases: inicio, elaboración, construcción e implementación y en cada fase se realizan tareas de: análisis de requisitos, diseño, implementación y prueba. Existen varias iteraciones hasta la obtención de un producto final. Para la realización de la presente herramienta se realizaron cinco iteraciones. En esta sección se muestra el resultado de la última iteración.

4.1.1 Captura de requisitos como casos de uso

Los casos de uso proporcionan un medio intuitivo y sistemático para capturar los requisitos funcionales con énfasis especial en el valor añadido para cada usuario individual y para cada sistema externo (Jacobson et al., 2000).

Se encontraron los siguientes requisitos representados en el siguiente diagrama de casos de uso.

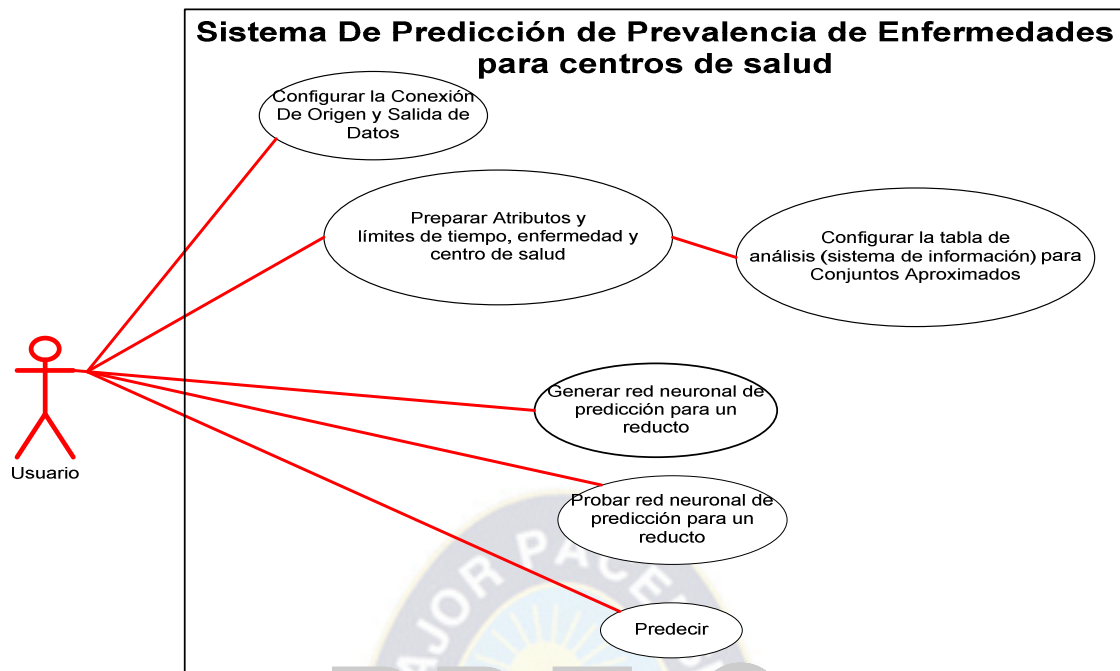


Figura 3. Diagrama de casos de uso
Fuente: Basado en (Jacobson et al., 2001)

4.1.1 Descripción de los casos de uso

4.1.1.1 Configurar la conexión de origen y salida

Un requisito importante es el almacén de datos, en nuestro caso es una base de datos para eso es importante configurar una cadena de conexión que indique el origen de los datos a procesar.

Como se ha mostrado en la sección anterior, se requiere además una tabla de análisis¹¹, por lo que es importante configurar su ubicación, es decir la conexión que indique el destino de los datos para construir dicha tabla.

Mediante este caso de uso el usuario configura la conexión de la base de datos y tabla origen y destino.

4.1.1.1.2 Preparar atributos y límites de tiempo, enfermedad y centro de salud

Los atributos son las columnas de la tabla de análisis, propiamente dicha. Es necesario preparar los atributos mediante un nombre y una condición de selección.

¹¹ Lo que según la teoría de conjuntos aproximados se conoce como “sistema de información”

El periodo de tiempo, que es un lapso de tiempo tomado en cuenta para el análisis de datos, es importante pues no todos los centros de salud tienen el mismo tiempo de funcionamiento por lo tanto no se los puede analizar bajo un límite general.

Tomando en cuenta atributos y tiempo el usuario debe seleccionar el centro de salud y la enfermedad para el análisis y predicción de prevalencia.

4.1.1.1.3 Configurar la tabla de análisis (sistema de información) para conjuntos aproximados

Con la conexión de destino establecida y los atributos y límites configurados el usuario puede crear la tabla de análisis para posteriormente encontrar reductos mediante este caso de uso.

4.1.1.1.4 Generar red neuronal de predicción para un reducto

Donde el usuario puede obtener reductos, escoger uno de ellos y generar una red neuronal de predicción para dicho reducto. Este proceso incluye entrenarla y almacenarla para su posterior utilización.

4.1.1.1.5 Probar red neuronal para un reducto

Se recupera una red neuronal para un reducto anteriormente almacenada y se generan datos de prueba. El error emitido será el resultado de la prueba con los mencionados datos y es el parámetro para medir si una red neuronal es mejor a otra. Este caso de uso aporta estas acciones.

4.1.1.1.6 Predecir

Una vez realizada la prueba de la red neuronal basada en un reducto (se denominará predictor de aquí en adelante) está lista para predecir, dados ciertos valores de entrada. Este caso de uso se encarga de la interacción con el usuario para realizar predicciones dado un predictor.

4.1.2 Análisis

El objetivo del análisis es conseguir una comprensión más precisa de los requisitos y una descripción de los mismos que sea fácil de mantener y que nos ayude a estructurar el sistema entero (Jacobson et al., 2000).

Se describirá a continuación mediante un diagrama de colaboración y una descripción de la realización de los casos de uso.

4.1.2.1 Configurar la conexión de origen y salida

La siguiente realización muestra cómo se realiza una configuración de conexión ya sea para una conexión de origen ó destino. La interfaz IConfiguración muestra la configuración actual de la cadena (los valores de la cadena conexión, base de datos y tabla están almacenados en un archivo XML denominado configuración.xml) pudiendo el usuario verificar si ésta es correcta o presenta fallas y posteriormente actualizarla.

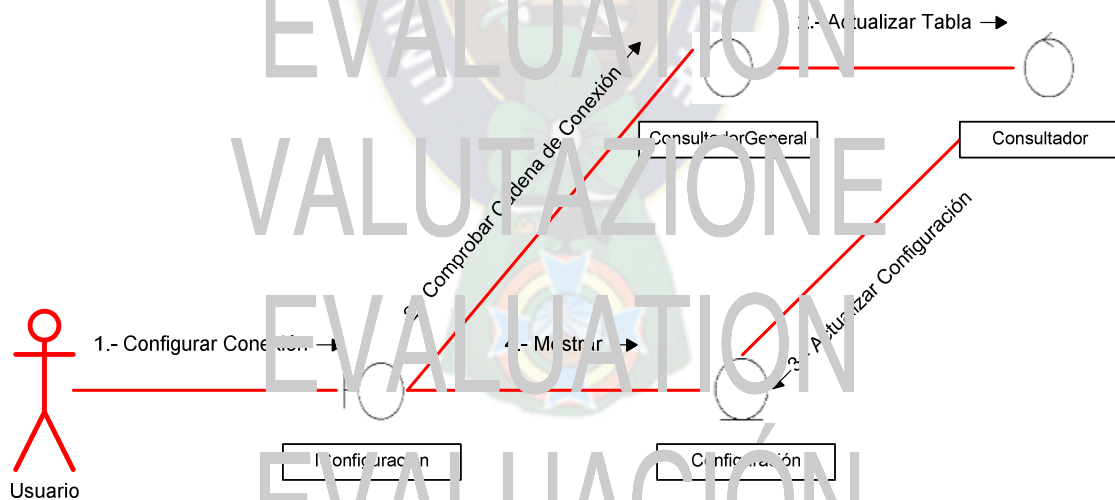


Figura 14. Diagrama de colaboración para el caso de uso: configurar la conexión de origen y salida.
Fuente: Pasado en Jacobson et al., 2000)

4.1.2.2 Preparar atributos y límites de tiempo, enfermedad y centro de salud

El usuario podrá configurar a través de la interfaz *IPreparaciónDeDatos*:

- Los atributos. Éstos se almacenarán de manera física en un archivo XML denominado *nombredelAtributo.xml*.
- El centro de salud. Almacenado en *hospital.xml*.
- Enfermedad. Que denota la enfermedad de estudio y al igual que el anterior se puede almacenar en el archivo *enfermedad.xml*
- Los límites. Que comprenden año y mes de inicio, y año y mes de finalización del estudio.

Una vez que el usuario realiza las modificaciones (1, 2) los valores se almacenan mediante métodos de la clase de control *PreparadorDeDatos* en sus archivos xml correspondientes (3, 4, 5, 6). Posteriormente la interfaz *IPreparaciónDeDatos* muestra las modificaciones al usuario (7).

Con estos datos el usuario puede crear la tabla para análisis de manera física en la dirección establecida.

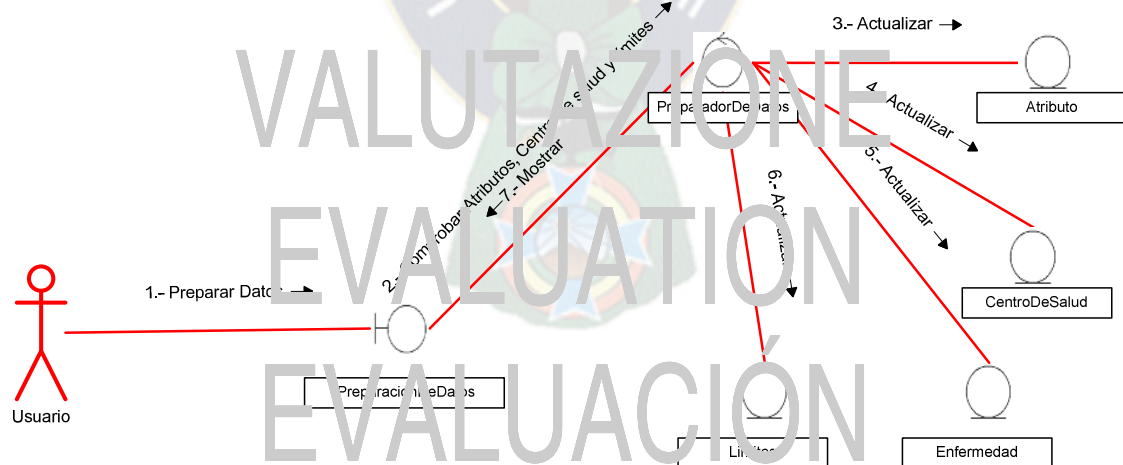


Figura 15. Diagrama de colaboración para el caso de uso: Preparar Atributos y límites de tiempo, enfermedad y centro de salud
Fuente: Basado en (Jacobson et al., 2000)

4.1.2.3 Configurar la tabla de análisis (sistema de información) para conjuntos aproximados

El usuario puede verificar la existencia de la tabla de análisis (1,2). Si esta existe entonces se procede a llenarla mediante la interfaz *IPreparacionDeDatos* con métodos de la clase de control *PreparadorDeDatos* (3).



Figura 16. Diagrama de colaboración para el caso de uso: Configurar la tabla de análisis (sistema de información) para conjuntos aproximados
Fuente: Basado en (Jacobson et al., 2000)

4.1.4 Generar red neuronal de predicción para un reducto

Generar una red neuronal de predicción para un reducto equivale a generar un predictor y comienza con una solicitud del usuario (1):

- Primero se debe encontrar los reductos de la tabla de análisis. Esta tarea está a cargo del *GestorDeConjuntos Aproximados* (2) y se almacenan los reductos hallados (3).
- Posteriormente escoger uno (4, 5) y con los datos de las columnas que pertenecen al reducto escogido crear datos de entrenamiento (6) y una red neuronal (7); después entrenarla (8). Esta tarea está a cargo del *GestorDeRedNeuronal* obteniendo como resultado un predictor.

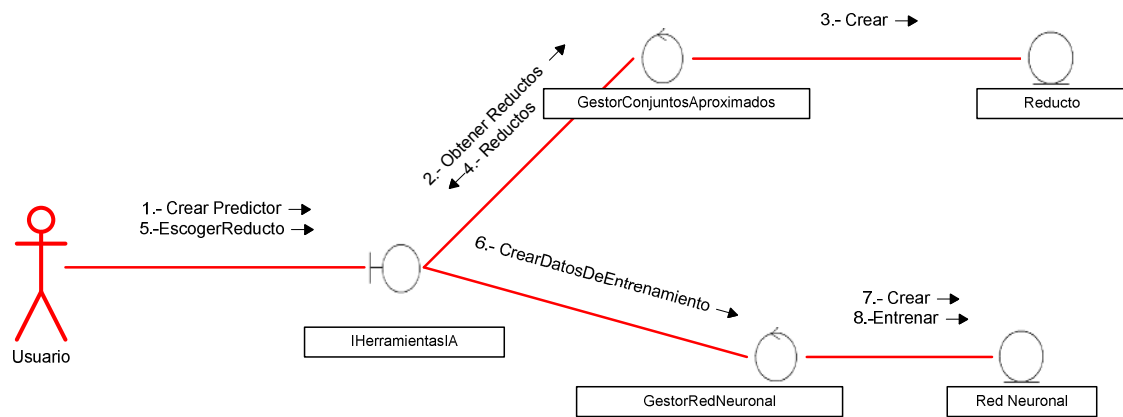


Figura 17. Diagrama de colaboración para el caso de uso: Generar red neuronal de predicción para un reducto
Fuente: Basado en (Jacobson et al., 2000)

4.1.2.5 Probar red neuronal para un reducto

El usuario hace la petición de probar un predictor (1), posteriormente mediante la interfaz de herramienta IA crea datos de prueba (2), y se realiza la prueba (3) generando un error (4).

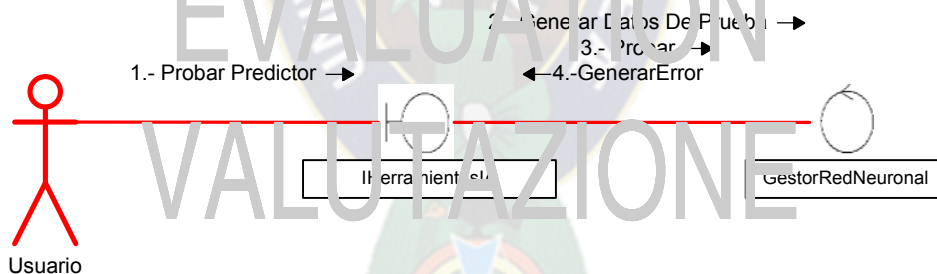


Figura 18. Diagrama de colaboración para el caso de uso: Probar red neuronal para un reducto
Fuente: Basado en (Jacobson et al., 2000)

4.1.2.6 Predecir

Una vez creados varios predictores el usuario puede escoger uno, introducir datos (1) y realizar la predicción de la prevalencia (2, 3) a través de la interfaz IHerramientasIA.

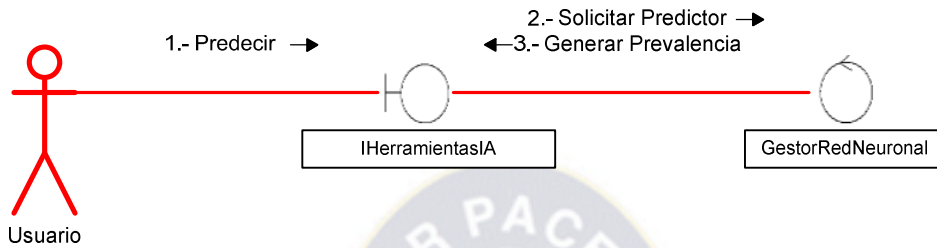


Figura 19. Diagrama de colaboración para el caso de uso: Predecir
Fuente: Basado en (Jacobson et al., 2000)

4.1.2.7 Identificación de subsistemas y sus dependencias

Asociando los casos de uso y sus realizaciones se ha considerado los siguientes subsistemas para la solución.

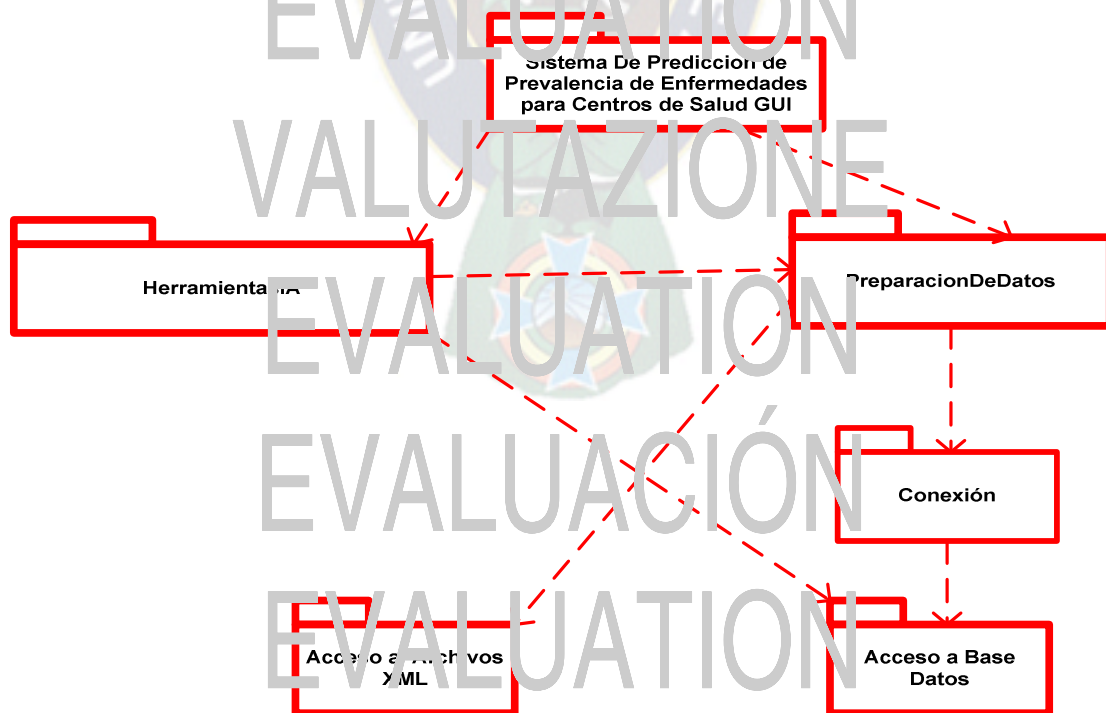


Figura 20. Diagrama de paquetes y dependencias
Fuente: Basado en (Jacobson et al., 2000)

4.1.3 Diseño

En el diseño modelamos el sistema y encontramos su forma (incluida su arquitectura) para que soporte todos los requisitos – incluyendo los funcionales y otras restricciones – que se suponen (Jacobson et al., 2000).

De acuerdo al diagrama de subsistemas y dependencias se realizó el siguiente diagrama de clases por subsistema.

4.1.3.1 Paquete de diseño Conexión

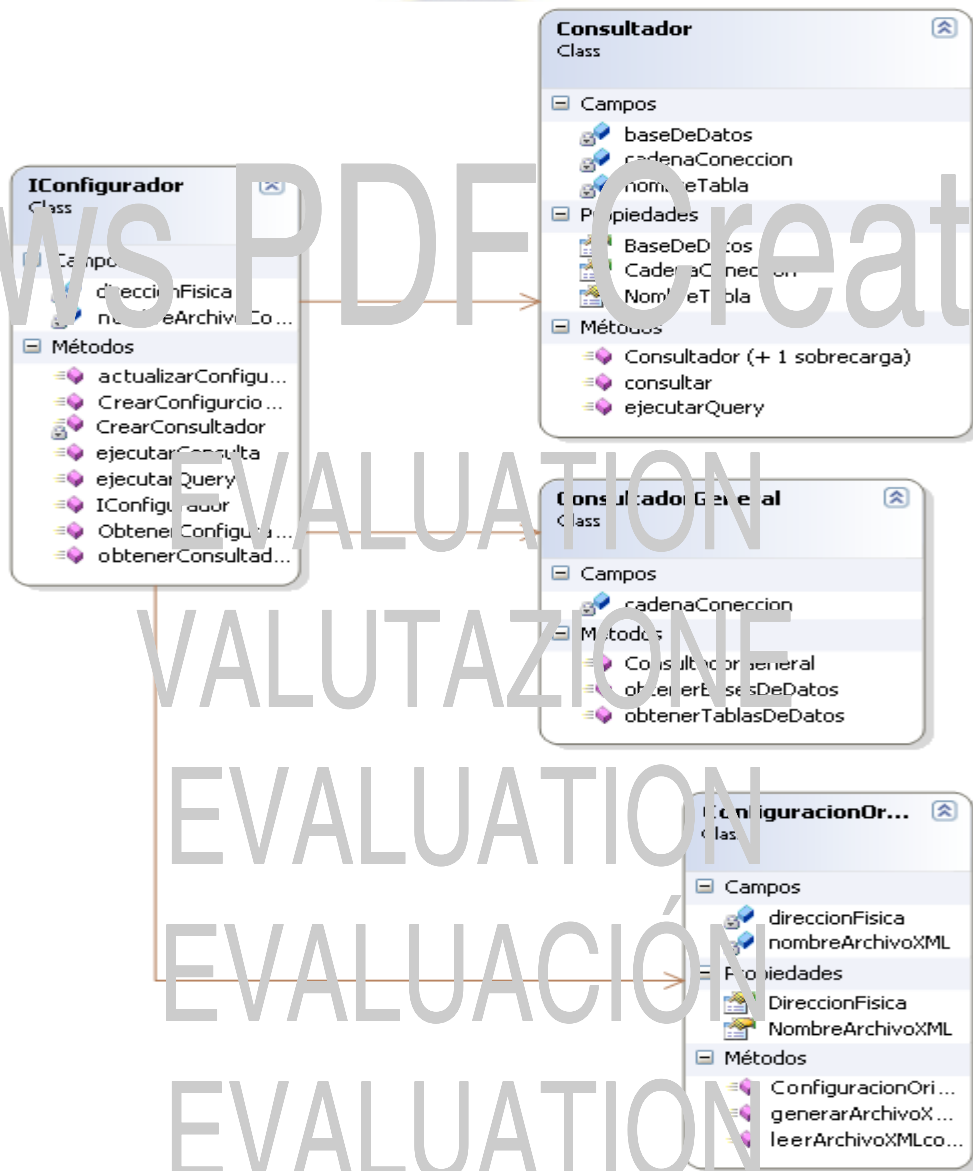


Figura 21. Diagrama de clases para el paquete de diseño: Conexión
Fuente: Basado en (Jacobson et al., 2000)

4.1.3.2 Paquete de diseño de HerramientasIA

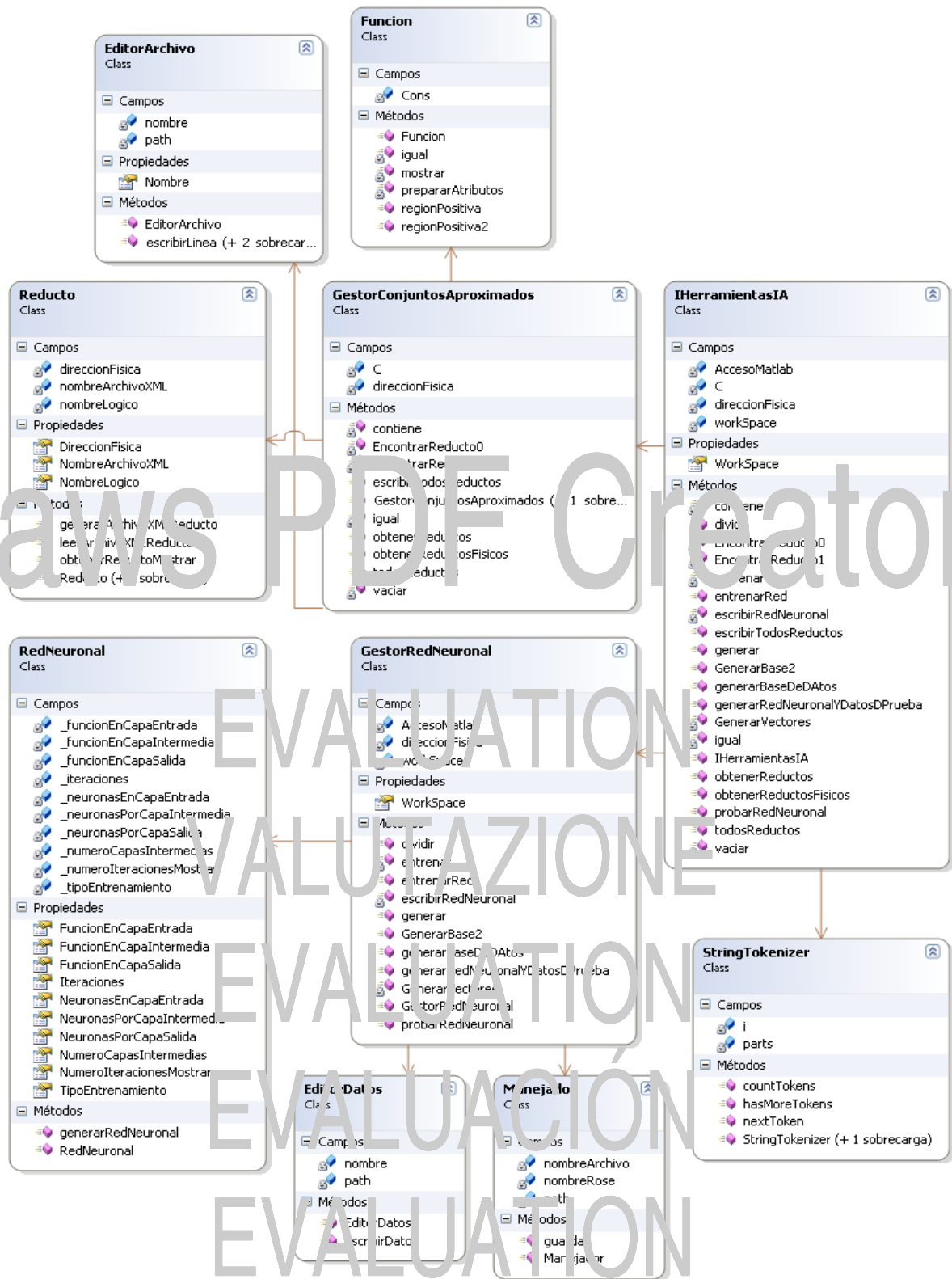


Figura 22. Diagrama de clases para el paquete de diseño: HerramientasIA
Fuente: Basado en (Jacobson et al., 2000)

4.1.3.3 Paquete de diseño Preparación de datos

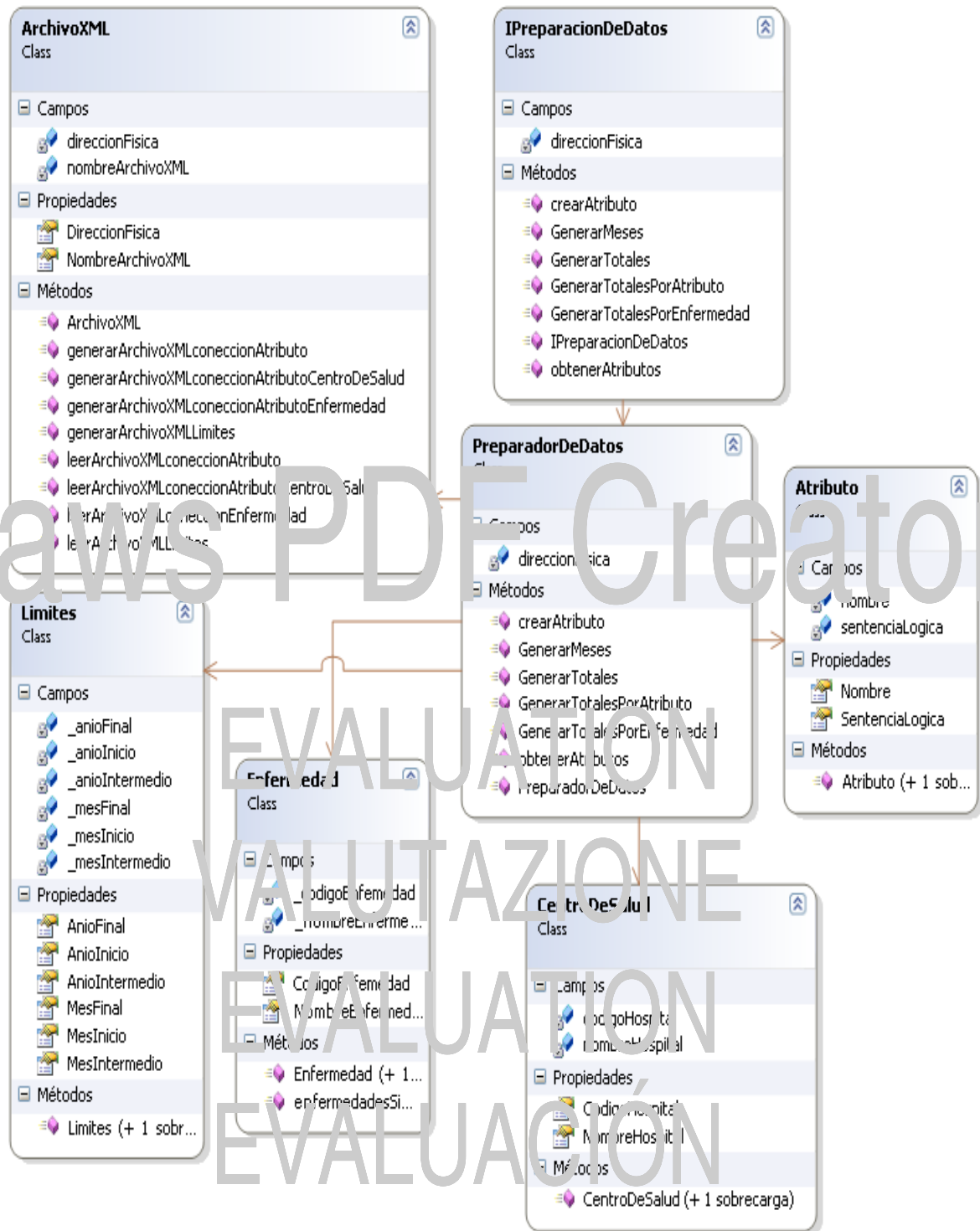


Figura 23. Diagrama de clases para el paquete de diseño: Preparación de datos

Fuente: Basado en (Jacobson et al., 2000)

4.1.3.4 Paquete de diseño de Sistema de Predicción de Prevalencia de Enfermedades para Centros de Salud (GUI)

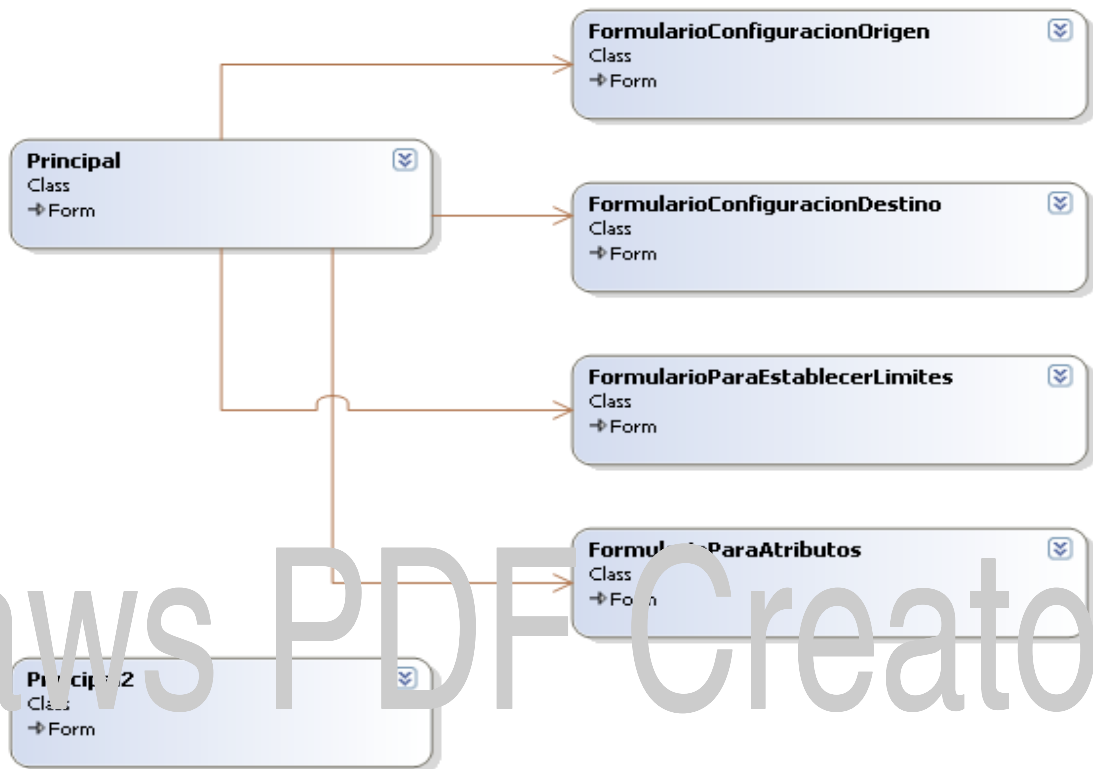


Figura 24. Diagrama de clases para el paquete de diseño. Paquete de diseño de Sistema De Preuicción de Prevaencia de Enfermedades para Centros de Salud (GUI)

Fuente: Basado en (Jacobson et al., 2000)

4.1.4 Implementación

La implementación comenzó con el resultado del diseño y se implementó el sistema en términos de componentes, es decir, ficheros de código fuente, scripts, ficheros de código binario, ejecutables y similares (Jacobson et al., 2000). Como requisitos tenemos:

- La base de datos de Medicus Mundi se encuentra soportada por el gestor SQL server 2000 o superior.
- Para el manejo de redes neuronales se debe tomar en cuenta la compatibilidad de un software de inteligencia artificial con el lenguaje de desarrollo que elijamos.

Para ello se decidió realizar la implementación en C# con el entorno de desarrollo Visual.NET y el software matemático MATLAB, que puede compatibilizar con el anterior. Por lo tanto los subsistemas identificados en la etapa de diseño reciben el nombre de soluciones y los ficheros código fuente serán de extensiones “cs”.

4.1.4.1 Solución Conexión

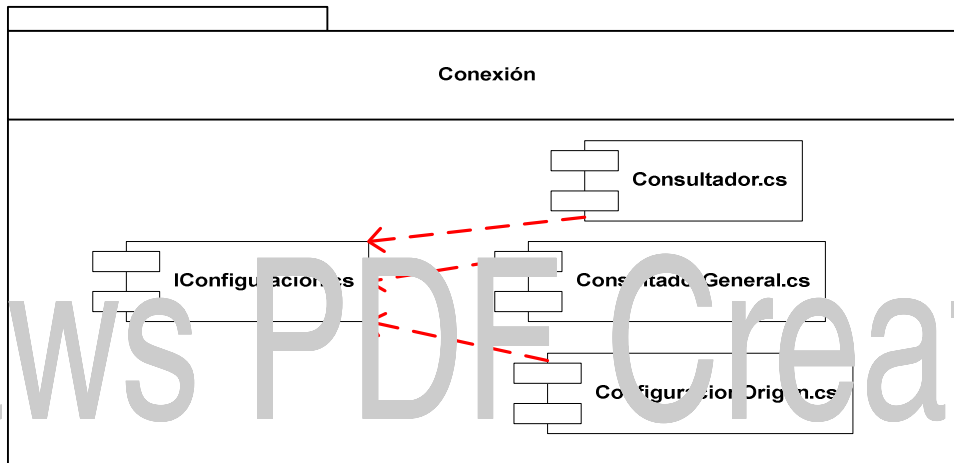


Figura 25. Solución: Conexión
Fuente: Basado en (Jacobson et al., 2000)

4.1.4.2 Solución HerramientasIA

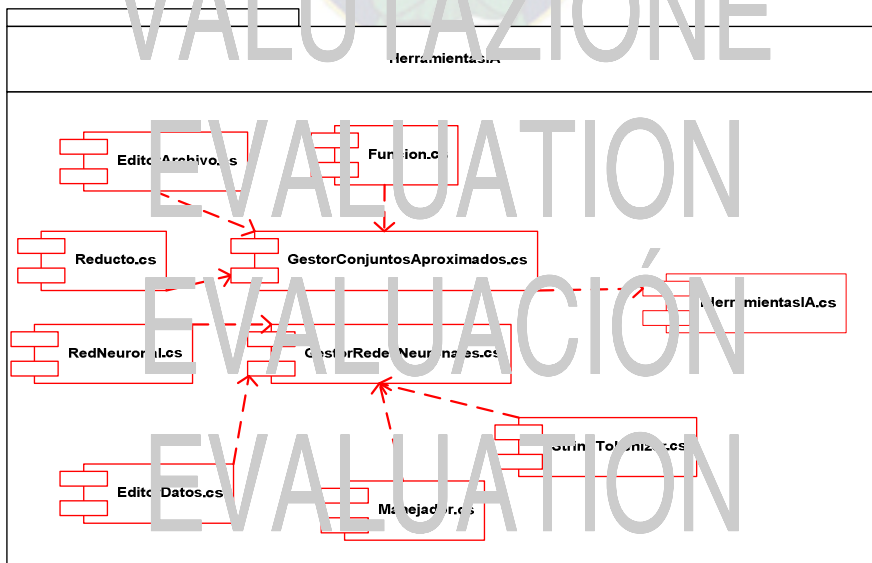


Figura 26. Solución: HerramientasIA
Fuente: Basado en (Jacobson et al., 2000)

4.1.4.3 Solución Preparación de datos

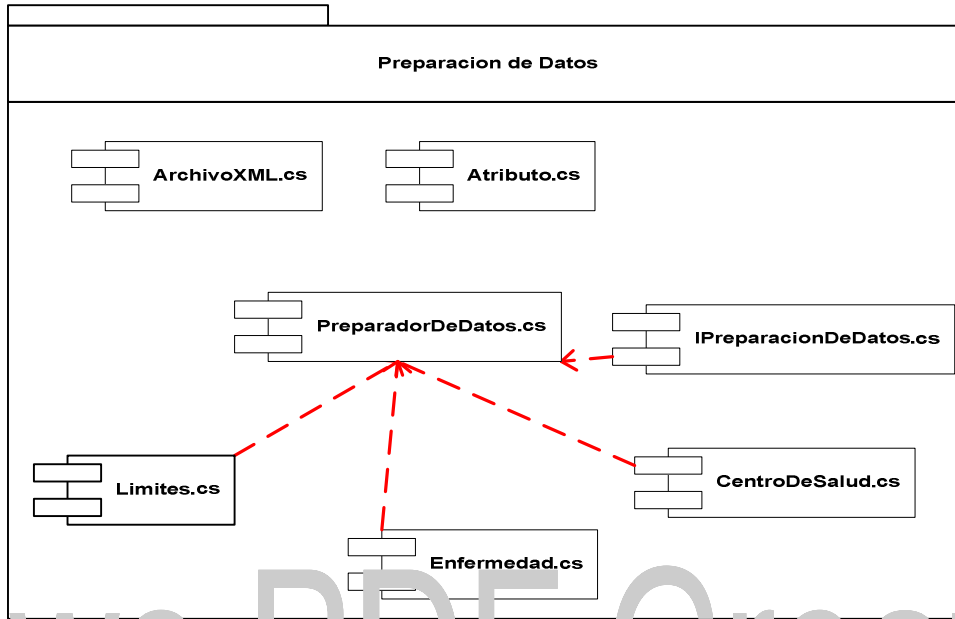


Figura 27. Solución: Preparación de Datos
Fuente: Basado en (Jacobson et al., 2000)

4.1.4.4 Solución Sistema de Predicción de Prevalencia de Enfermedades para Centros de Salud (GUI)

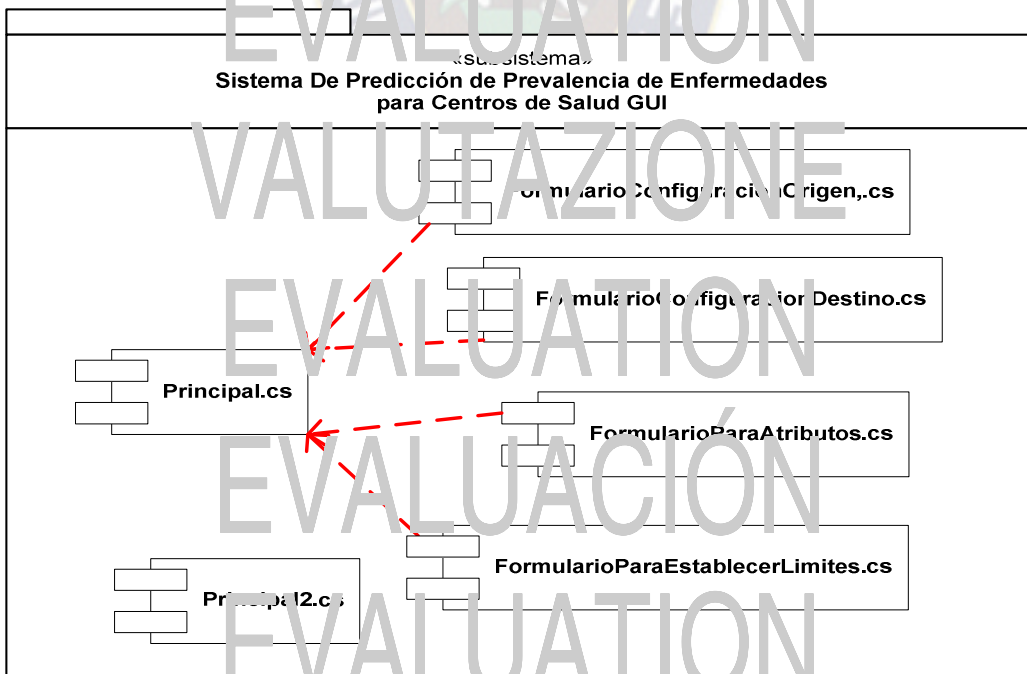


Figura 28. Solución: Sistema De Predicción de Prevalencia de Enfermedades para Centros de Salud (GUI)
Fuente: Basado en (Jacobson et al., 2000)

4.2 Preparación de tablas de análisis

Como se vio en el capítulo anterior es importante generar tablas de análisis para poder extraer de ellas reductos.

El lapso de tiempo juega un papel muy importante. Generalmente en los sistemas de minería de datos el 70% de los datos se utiliza para un entrenamiento y el restante para la prueba. Así que se debe asignar un tiempo de inicio, un tiempo de finalización y un tiempo de prueba.

4.3 Solución basada en la teoría conjuntos aproximados

4.3.1 Selección de reductos

La propuesta para hallar reductos está relacionada con el lenguaje SQL.

```
private Queue<string> encontrarReducto0(Queue<string>
atributosCondicion, Queue<string> atributosDecision)
{
    Funcion metodos = new Funcion(C);
    int rp1 =
metodos.regionPositiva(atributosCondicion, atributosDecision);
    int c1 = atributosCondicion.Count;
    int rp2,
    string x;
    for (int i = 1; i < c1; i++)
    {
        x = atributosCondicion.ElementAt(i);
        rp2 =
metodos.regionPositiva(atributosCondicion, atributosDecision);
        if (rp2 < rp1)
            atributosCondicion.Enqueue(x);
    }
    return atributosCondicion;
}
```

Figura 29. Código del algoritmo propuesto para encontrar un reducto
Fuente. (Elaboración Propia)

Como se ha observado un reducto es un subconjunto de los atributos condición. Contabilizando se llega a un número extremadamente grande. Si se tuviera 10 atributos como condición el número de subconjuntos a analizar para 'verificar' que cada uno sea un reducto sería 1024.

El orden de los atributos influye de manera sustancial en la búsqueda de reductos en el algoritmo propuesto. Si analizamos la disposición de 10 atributos condición ésta se podría dar de $10! = 3628800$ maneras, en un caso general es $n!$ maneras, tomando a n como el número de atributos condición. No es conveniente analizar cada una de estas permutaciones pues no nos interesan todas. Solamente un conjunto pequeño de ellas vendrá a ser un reducto. Por lo tanto es prudente combinar su disposición siguiendo algún heurístico de manera que se pueda encontrar la mayor cantidad de ellos. Se propone un heurístico a continuación:

1. – Sea *Reductos* un conjunto de reductos

2. – *reducto* ← EncontrarReducto(C, D, E)

3. – Si el reducto no existe en *Reductos*:

Reductos.almacenarReducto (*reducto*)

4. – $C \leftarrow invertir \text{ e } intercalar (C)$

5. – volver a 2.

Figura 30. Heurístico propuesto para encontrar varios reductos
Fuente: (Elaboración Propia)

La idea central para hacer la variación es:

```
private void invertir(Queue<string> queue)
{
    int n = queue.Count;
    Random r = new Random();
    int random = (int)Math.Floor((r.NextDouble() * n));
    // Console.WriteLine(" -> " + random);
    for (int i = 1; i <= random; i++)
    {
        queue.Enqueue(queue.Dequeue());
    }
}

private void intercalar(Queue<string> queue)
{
    int n = queue.Count;
    Random r = new Random();
    Queue<string> auxiliar = new Queue<string>();
    for (int i = 1; i <= n / 2; i++)
    {
        auxiliar.Enqueue(queue.Dequeue());
        queue.Enqueue(queue.Dequeue());
    }
    int m = auxiliar.Count;
    for (int j = 1; j <= m; j++)
    {
        queue.Enqueue(auxiliar.Dequeue());
    }
}
```

Figura 31. Código de los dos programas para eliminar varios reductos
Fuente: (Elaboración Propia)

4.4 Red neuronal

Se realizaron pruebas para obtener las siguientes características generales para una red neuronal artificial:

Un número r de neuronas en la capa de entrada, dependiendo del número de columna que tenga el reducto seleccionado. Seis capas con 9 neuronas cada una. Y una capa de salida con una neurona que indica el valor de la prevalencia.

La función de la capa de entrada y las capas intermedias es tansig y para la capa de salida es purelin.

La función de entrenamiento elegida es "traingdx", que es un entrenamiento rápido proporcionado por MATLAB para retropropagación. El número de iteraciones inicialmente será de 10000 o más según sea lo necesario.

El límite de error de entrenamiento esperado es 1e-5.

```
Net = newff(minmax(A),[r, 9,9,9,9,9,1],  
{'tansig', 'tansig', 'tansig', 'tansig', 'tansig', 'tansig', 'tansig', 'purelin'}, 'traingdx');  
net.trainParam.epochs = 10000;  
net.trainParam.goal = 1e-5;  
[net,trE]= train (net, A,B)
```

Figura 32. Código fuente de la red neuronal artificial.
Fuente: elaboración propia.

4.5 Características para datos de análisis

4.5.1 Asignación de tiempo

Es importante contar con información adecuada para la prueba del modelo, por lo cual se debe determinar un lapso de tiempo donde existan datos suficientes para ser tomados en cuenta, es decir donde la prevalencia de enfermedad sea mayor a 0.

Tomando en cuenta lo anterior y realizando observaciones en la base de datos de Medicus Mundi el periodo de análisis y construcción de tablas de análisis y entrenamiento de la red neuronal será de un periodo desde enero 2005 a junio del 2007, pudiendo variar desde enero 2005 a junio 2007. La prueba se realizará con datos tomados desde julio 2007 a junio 2008. Por lo tanto el tiempo de estudio será desde enero 2005 hasta junio 2008.

4.5.2 Elección de columnas de análisis

Para ello se realiza tomando en cuenta tres atributos:

El grupo etéreo. Para un análisis más detallado dividido en cinco columnas: Menores de 1 año, de 1 a 15 años, de 16 a 30 años, de 31 a 45 años y mayores a 45 años.

El género. A partir de este atributo se generan dos columnas: masculino y femenino.

El mes. Que solo genera un valor descriptivo.

En base a estos 8 atributos se generan las tablas de análisis.

4.5.3 Elección de enfermedades

Mediante consultas Sql en la base de datos de medicus mundi se generó la siguiente tabla por una prevalencia general: el número de casos de persona que tuvieron la enfermedad dividido entre el total de casos en todos los centros de salud registrados para el tiempo de estudio mencionado anteriormente.

Número	Código CIE	Nombre de la enfermedad	Prevalencia General
1.-	A09	Diarrea y gastroenteritis de presunto origen infeccioso	0,14499706
2.-	J00	Rinofaringitis aguda [esfria lo común]	0,08135045
3.-	I10	Hipertensión esencial (primaria)	0,07970358
4.-	H10	Conjuntivitis	0,07123916
5.-	N39.0	Inyección de vías urinarias, sitio no especificado	0,05327498
6.-	R51	Cefalea	0,05046561
7.-	J02	Faringitis aguda	0,04711133
8.-	J20	Bronquitis aguda	0,04448966
9.-	M54	Dorsalgia	0,04439279
10.-	R50	Fiebre de origen desconocido	0,03951273

*Figura 33. Las diez enfermedades con mayor índice de prevalencia
Fuente: (Medicus Mundi, 2008)*

4.5.4 Elección de centros de salud

Los centros de salud difieren unos de otros y también la información que proveen, por lo tanto para el estudio de una determinada enfermedad se debe escoger hospitales que provean información importante para ella, por lo menos durante el tiempo de estudio. En otras palabras el número de casos durante este tiempo no debe ser 0 en la mayoría de los años y meses.

Año	Mes	Número de casos
2005	Enero	0
2005	Febrero	0
2005	Marzo	57
2005	Abril	54
2005	Mayo	124
2005	Junio	83
2005	Julio	78
2005	Agosto	98
2005	Septiembre	105
2005	Octubre	11
2005	Noviembre	72
2005	Diciembre	108
2006	Enero	116
2006	Febrero	72
2006	Marzo	91
2006	Abril	74
2006	Mayo	62
2006	Junio	71
2006	Julio	71
2006	Agosto	64
2006	Septiembre	53
2006	Octubre	61
2006	Noviembre	60
2006	Diciembre	73
2007	Enero	53

2007	Febrero	43
2007	Marzo	65
2007	Abril	59
2007	Mayo	64
2007	Junio	72
2007	Julio	53
2007	Agosto	88
2007	Septiembre	51
2007	Octubre	58
2007	Noviembre	53
2007	Diciembre	44
2008	Enero	64
2008	Febrero	42
2008	Marzo	49
2008	Abril	50
2008	Mayo	0
2008	Junio	0

Figura 34. Información provista por el hospital Boliviano Holandés para “Diarrea y gastroenteritis de presunto origen infeccioso”
Fuente: (Medicus Mundi, 2008)

4.6 Variables

4.6.1 Identificación de variables

Variable independiente (VI): Sistema de predicción de prevalencia de enfermedades para centros de salud.

Variable Dependiente (VD): Prevalencia

4.7 Error de predicción

El error propuesto es: $precisión = \frac{TF}{TP + FP}$

El error de precisión será:

$$eprecisión = 1 - precisión$$

Pero este error es difícilmente aplicable a valores reales, por lo que se propone una prueba de hipótesis y un estadístico para asegurar una predicción exitosa.

4.8 Prueba de hipótesis

Se tiene un conjunto de datos de prueba, que es la información de las prevalencias almacenadas en la base de datos desde julio 2007 hasta junio 2008.

$$y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8, y_9, y_{10}, y_{11}, y_{12}$$

Además se tiene un conjunto de datos de valores predichos, que son los que el modelo obtiene después del entrenamiento que ha tenido las prevalencias desde julio 2007 hasta junio 2008.

$$p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8, p_9, p_{10}, p_{11}, p_{12}$$

Como el propósito es verificar que las predicciones se encuentran cerca de las prevalencias almacenadas se realizará este hecho con la prueba t por pares. Generando la variable de diferencia

$$d_i = y_i - p_i$$

Se tiene como hipótesis nula: $H_0: \mu_D = 0$ donde $D = Y - P$ es la diferencia entre la primera y la segunda observación dentro de un par, $\mu_D = \mu_1 - \mu_2$

El valor del estadístico de prueba: $t = \frac{\bar{d} - 0}{S_D / \sqrt{n}}$ donde \bar{d} y S_D son la desviación media y estándar, respectivamente de las d_i .

Como hipótesis alternativa se tiene $H_1: \mu_D \neq 0$ ya sea $t \geq t_{\frac{\alpha}{2}, n-1}$ o $t \leq -t_{\frac{\alpha}{2}, n-1}$ ¹³.

Debre (1998) señala dos tipos de errores el error tipo I consiste en rechazar la hipótesis nula H_0 cuando es verdadera y el error tipo II consiste en no

¹² Donde $\bar{d} = \sum d_i / n$ y $S_D^2 = \frac{\sum d_i^2 - (\sum d_i)^2 / n}{n-1}$

¹³ Hace referencia a la tabla de la distribución t .

rechazar H_0 cuando H_0 falsa. El nivel de significancia consiste en la probabilidad de cometer un error tipo I. El mismo es denotado por α . Se realizará la prueba con un nivel de significancia del 0.05, es decir al 95% de confianza. Por lo tanto $t_{\frac{\alpha}{2}, n-1} = t_{\frac{0.05}{2}, 12-1} = t_{0.025, 11} = 2.201$.

4.9 Pruebas

Tomando en cuenta las anteriores consideraciones se realizaron las pruebas de hipótesis para uno de los reductos obtenidos por la herramienta software desarrollada. Dicho procedimiento se realizó para las diez enfermedades con mayor prevalencia con fuente de la base de datos de Medicus Mundi.

4.9.1 Prueba 1

Enfermedad: Diarrea y gastroenteritis de presunto origen infeccioso

Código de enfermedad: I10 A09

Centro de salud: Hospital Boliviano Holandés – E. M. S.

Código del centro de salud: 200143

Periodo de estudio: 2006 a 2008

Tabla de análisis:

De 16 a 30 años	De 1 a 15 años	De 31 a 45 años	De 45 años en adelante	Femenino	Masculino	Menores a 1 año	Mes	Prevalencia
0	0,5	0	0,2	0,5	0,5	0,2	1	Muy baja
0,5	0	0,5	0	0,5	0,5	0	2	Muy baja
0,1	0,5	0,1	0,1	0,5	0,5	0,2	3	Baja
0,1	0,4	0,1	0,1	0,4	0,6	0,3	4	Baja
0	0,6	0	0	0,5	0,5	0,3	5	Muy alta
0	0,6	0	0	0,5	0,5	0,3	6	Media
0,1	0,6	0,1	0	0,5	0,5	0,3	7	Baja
0	0,6	0,1	0,1	0,5	0,5	0,2	8	Media
0,1	0,5	0	0,1	0,5	0,5	0,2	9	Media
0,1	0,4	0,1	0,1	0,5	0,5	0,3	10	Media
0,1	0,5	0,1	0,2	0,6	0,4	0,2	11	Muy baja
0,1	0,5	0,1	0,1	0,5	0,4	0,1	12	Media
0,1	0,4	0,1	0,2	0,5	0,5	0,1	1	Muy alta
0,2	0,5	0,1	0,1	0,5	0,5	0,2	2	Muy alta
0,1	0,5	0,1	0,1	0,5	0,5	0,3	3	Muy alta
0,1	0,4	0,1	0,2	0,5	0,5	0,2	4	Alta
0,2	0,4	0,1	0,2	0,6	0,4	0,2	5	Baja
0,1	0,5	0,1	0,1	0,6	0,4	0,2	6	Media

0,2	0,3	0,2	0,2	0,5	0,5	0,2	7	Media
0,2	0,3	0,1	0,2	0,6	0,4	0,2	8	Media
0,1	0,4	0,2	0,1	0,7	0,3	0,2	9	Muy baja
0,1	0,4	0,2	0,2	0,5	0,5	0,2	10	Baja
0,2	0,4	0,2	0,2	0,4	0,6	0,2	11	Media
0,1	0,4	0,1	0,1	0,6	0,4	0,3	12	Muy alta
0,1	0,5	0,2	0,1	0,5	0,5	0,1	1	Baja
0	0,5	0	0,2	0,4	0,6	0,2	2	Muy baja
0,1	0,4	0,1	0,1	0,5	0,5	0,3	3	Media
0	0,6	0,1	0,1	0,5	0,5	0,2	4	Media
0,1	0,4	0	0,1	0,4	0,6	0,3	5	Alta
0,1	0,5	0	0,1	0,5	0,5	0,3	6	Muy alta

Tabla 9. Tabla de análisis para diarrea y gastroenteritis de presunto origen infeccioso en el hospital Boliviano Holandés de la ciudad de El Alto
Fuente: (Elaboración Propia)

Observaciones en los atributos condición de la tabla de análisis:

De 6 a 30 años	De la tabla de análisis se puede observar que una pequeña parte de la población es de edad entre 6 a 30 años cuyo porcentaje es 11%.
De 1 a 15 años	El porcentaje de la población que tiene este rango de edad es 45%. Es decir la enfermedad tiene mayor presencia en niños y adolescentes.
De 31 a 45 años	El porcentaje de pacientes registrados en esta categoría de edad fue 11%. Se puede observar los datos no difieren en los meses de manera significativa.
De 45 años en adelante	De todos los datos de pacientes estudiados el 10% de personas en esta edad presentaron la enfermedad.
Femenino	Las mujeres que presentaron la enfermedad tienen como porcentaje 51.33% de la población. Ligeramente más de la mitad de todos los casos totales.
Masculino	El porcentaje de varones fue 48.67%. Se puede observar que no existe mucha diferencia para la enfermedad entre varones y mujeres.
Menores a 1 año	El porcentaje de niños menores a 1 año es de 22.33%.
Mes	El mes de junio denota una similitud en prevalencia ya que es de media a muy alta, asimismo el mes de abril denota una similitud para datos de prevalencia entre baja y alta. Los datos se

presentan en la siguiente tabla:

Mes	Prevalencia
1	Muy baja
1	Muy alta
1	Baja
2	Muy baja
2	Muy alta
2	Muy baja
3	Baja
3	Muy alta
3	Media
4	Media
4	Alta
4	Baja
5	Muy alta
5	Baja
5	Alta
6	Muy alta
6	Media
6	Media
7	Baja
7	Media
8	Media
8	Media
9	Media
9	Muy baja
10	Baja
10	Media
11	Muy baja
11	Media
12	Muy alta
12	Media

Tabla 10. Meses y prevalencia propuesta de la tabla de análisis para diarrea y gastroenteritis de presunto origen infeccioso en el Hospital Boliviano Holandés de la ciudad de El Alto

Fuente: (Elaboración Propia)

El sistema produjo los siguientes reductos:

Reducto	Atributos
Reducto 1	Menores a 1 año, de 1 a 15 años, masculino, mes.
Reducto 2	De 1 a 15 años, menores a 1 año, mes, femenino.
Reducto 3	Menores a 1 año, mes, de 1 a 15 años, de 45 en adelante años.
Reducto 4	Mes, de 1 a 15 años, menores a 1 año, de 31 a 45 años.
Reducto 5	De 1 a 15 años, menores a 1 año, mes, de 16 a 30 años.

Tabla 11. Reductos de la tabla de análisis para diarrea y gastroenteritis de presunto origen infeccioso en el hospital Boliviano Holandés de la ciudad de El Alto

Fuente: (Elaboración Propia)

El núcleo proporcionado por el sistema de información es: mes, menores a 1 año y pacientes de 1 a 15 años. Es decir mes, menores a 1 año y de 1 a 15 años son los atributos más importantes a estudiar.

Se puede utilizar cualquiera de los reductos proporcionados para realizar predicciones. Escogeremos los atributos del reducto 2 para el entrenamiento y predicción.

Resultados

El reducto 2 con un entrenamiento de 3000 iteraciones produjo un estadístico t de -1.6985 .

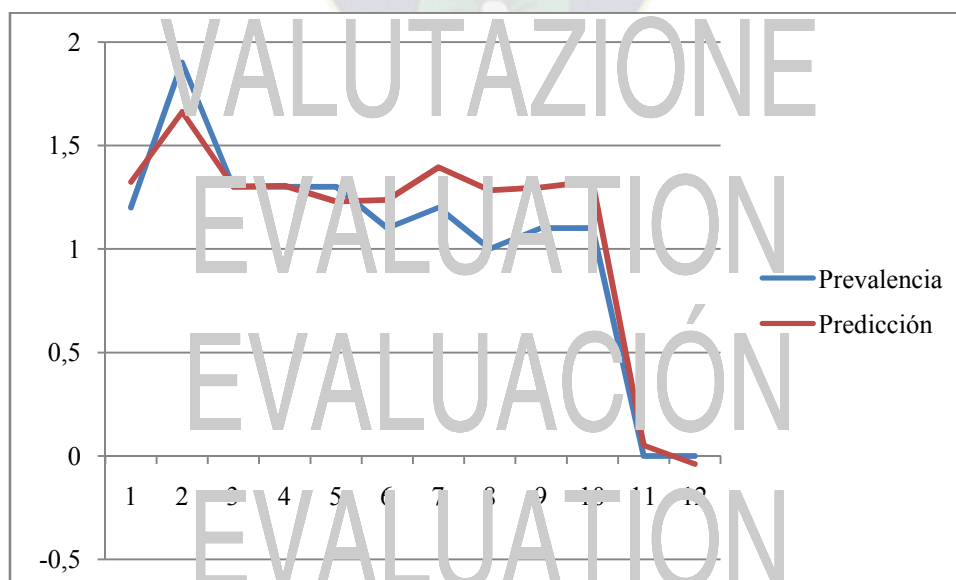


Figura 35. Resultados de predicción de reducto 2 en prueba 1
Fuente: (Elaboración Propia)

Como $t = -1.6985 < t_{0.025,11} = 2.201$ y $t = -1.6985 > t_{0.025,11} = -2.201$ no se puede rechazar la hipótesis que la diferencia sea 0.

El reducto escogido presenta datos y un error aceptable al nivel de confiabilidad propuesto. Los futuros estudios deberán enfocarse en las variables del reducto. Por ejemplo se puede realizar campañas preventivas por el mes de junio enfocado a niños menores de 1 año y personas entre 1 y 15 años, teniendo en cuenta que el sexo no es relevante ya que la enfermedad afecta tanto a varones como a mujeres. La red neuronal creada para este ejemplo puede ser entrenada sucesivas veces mejorando el estadístico t para que sea cercano a 0.

4.9.2 Prueba 2

Información: Rinofaringitis aguda (resfriado común)

Código de enfermedad CIE10: J00

Centro de salud: Hospital Boliviano Holandés – El Alto

Código del centro de salud: 200143

Periodo de estudio: 2006 a 2003

Tabla de análisis:

De 16 a 30 años	De 1 a 15 años	De 1 a 45 años	De 45 años en adelante	Femenino	Vasculino	Menores a 1 año	Mes	Prevalencia
0	0	0	0	0	0	0	1	Muy baja
0	0	0	0	0	0	0	2	Muy baja
0	0,7	0	0	0,5	0,5	0,3	3	Media
0	0,6	0	0	0,5	0,5	0,3	4	Media
0	0,7	0	0	0,5	0,5	0,3	5	Alta
0	0,7	0	0	0,5	0,5	0,3	6	Media
0	0,7	0	0	0,5	0,5	0,2	7	Muy baja
0,1	0,7	0	0	0,5	0,5	0,2	8	Media
0	0,8	0	0	0,5	0,5	0,1	9	Muy alta
0	0,7	0	0	0,5	0,5	0,2	10	Muy alta
0	0,6	0	0	0,5	0,5	0,3	11	Baja
0	0,7	0	0	0,5	0,5	0,2	12	Baja
0	0,6	0	0	0,6	0,4	0,3	1	Baja
0	0,7	0	0	0,5	0,5	0,3	2	Baja
0,1	0,7	0	0	0,5	0,5	0,2	3	Alta
0,1	0,7	0	0	0,5	0,5	0,2	4	Muy alta

0,1	0,6	0	0	0,5	0,5	0,2	5	Muy alta
0	0,8	0	0	0,5	0,5	0,2	6	Media
0	0,8	0	0	0,5	0,5	0,2	7	Baja
0	0,7	0	0	0,6	0,4	0,2	8	Baja
0	0,7	0	0	0,6	0,4	0,2	9	Baja
0	0,7	0	0	0,4	0,6	0,3	10	Media
0	0,7	0	0	0,4	0,6	0,3	11	Media
0	0,7	0	0	0,7	0,3	0,2	12	Muy baja
0,2	0,4	0	0,1	0,6	0,4	0,3	1	Muy baja
0	0,6	0	0	0,4	0,6	0,3	2	Alta
0	0,7	0	0	0,5	0,5	0,3	3	Alta
0	0,8	0	0	0,7	0,3	0,2	4	Muy alta
0	0,7	0	0	0,5	0,5	0,2	5	Muy alta
0	0,7	0	0	0,4	0,6	0,2	6	Muy alta

Tabla 12. Tabla de análisis para Rinofaringitis aguda en el hospital Boliviano

relacionado con la ciudad de Fátima

Fuente: (Elaboración Propia)

Observaciones en los atributos condición de la tabla de análisis:

De 16 a 30 años	De la tabla de análisis se puede observar que una pequeña parte de la población es de edad entre 16 a 30 años, cuyo porcentaje es 2.1%.
De 1 a 15 años	El porcentaje de la población cumple con esta característica es 3.57%. Esta característica se repite en diferentes meses del estudio.
De 31 a 45 años	Según los datos de la tabla de análisis no se presentaron casos registrados en este hospital para personas en este rango de edad.
De 45 años en adelante	Solamente hubo 0.35% de personas en esta edad que presentaron la enfermedad. No es muy relevante.
Femenino	Las mujeres que presentaron la enfermedad tienen como porcentaje 51.42% de la población. Ligeramente más de la mitad de todos los casos presentados.
Masculino	El porcentaje de varones fue 48.57%. Se puede observar que no existe marcada diferencia para la enfermedad entre varones y mujeres.
Menores a 1 año	El porcentaje de niños menores a 1 año es de 23.92%. Es la segunda categoría que más casos de resfriado se almacenó.

Mes	Se puede observar la similitud de datos propuestos de prevalencia para los meses de abril, mayo y junio; con sus respectivos códigos 4, 5 y 6 que muestran una media, alta y muy alta prevalencia. Los datos se presentan en la siguiente tabla:
-----	--

Mes	Prevalencia
1	Muy baja
1	Baja
1	Muy baja
2	Alta
2	Baja
2	Muy baja
3	Media
3	Alta
3	Muy alta
4	Muy alta
4	Muy alta
4	Media
5	Alta
5	Muy alta
5	Muy alta
6	Muy alta
6	Media
6	Media
7	Muy baja
7	Baja
8	Baja
8	Media
9	Muy alta
9	Baja
10	Media
10	Muy alta
11	Baja
11	Media
12	Muy baja
12	Baja

*Tabla 13. Meses y prevalencia propuesta de la tabla de análisis para Rinofaringitis aguda en el Hospital Boliviano Holandés de la ciudad de El Alto
Fuente: (Elaboración Propia)*

El sistema produjo los siguientes reductos:

Reducto	Atributos
Reducto 1	Mes, menores a 1 años, de 1 a 15 años, masculino.
Reducto 2	Mes, menores a a1 año, femenino, de 1 a 15 años.

Tabla 14. Reductos obtenidos de la tabla de análisis para Rinofaringitis aguda en el hospital Boliviano Holandés de la ciudad de El Alto

Fuente: (Elaboración Propia)

El núcleo proporcionado por el sistema de información es: mes, menores a 1 año y pacientes de 1 a 15 años. Es decir mes, menores a 1 año y de 1 a 15 años son los atributos más importantes a estudiar.

Se puede utilizar cualquiera de los reductos proporcionados para realizar predicciones. Escogeremos los atributos del reducto 2 para entrenamiento y predicción.

Resultados

El reducto 2 con un entrenamiento de 20000 iteraciones produjo un estadístico t de -1.0719.

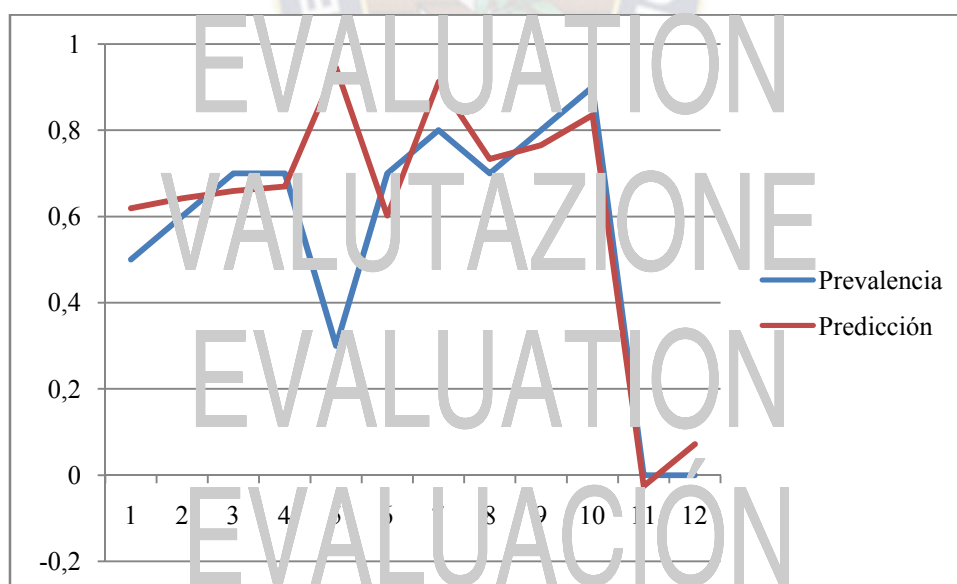


Figura 36. Resultados de predicción de reducto 2 en prueba 2

Fuente: (Elaboración Propia)

Como $t = -1.0719 < t_{0.025,11} = 2.201$ y $t = -1.0719 > t_{0.025,11} = -2.201$ no se puede rechazar la hipótesis que la diferencia sea 0.

El reducto escogido presenta datos y un nivel de confianza aceptable al nivel de confiabilidad propuesto. Los futuros estudios deberán enfocarse en las variables del reducto. Por ejemplo se puede realizar campañas preventivas por el mes de abril, mayo o junio enfocados a niños menores de 1 año y personas entre 1 y 15 años, como se observó la diferencia de porcentaje entre varones y mujeres no es de mayor relevancia. La red neuronal creada para este ejemplo puede ser entrenada sucesivas veces mejorando el estadístico t para que sea cercano a 0.

4.9.3 Prueba 3

Enfermedad: Hipertensión

Código de enfermedad CIE10: I10

Centro de salud: Universitario Japonés – Santa Cruz de la Sierra

Código del centro de salud: 70000

Período de estudio: 2006 a 2012

Tabla de análisis:

De 16 a 30 años	De 1 a 15 años	De 31 a 45 años	De 45 años en adelante	Femenino	Masculino	Menores a 1 año	Mes	Prevalencia
0	0	0	0	0	0	0	1	Muy baja
0	0	0	0	0	0	0	2	Muy baja
0	0	0	0	0	0	0	3	Muy baja
0	0	0	0	0	0	0	4	Muy baja
0	0	0	0	0	0	0	5	Muy baja
0	0	0	0	0	0	0	6	Muy baja
0	0	0,3	0,7	0,5	0,4	0	7	Muy baja
0	0	0,2	0,7	0,7	0,3	0	8	Media
0,1	0	0,1	0,8	0,6	0,4	0	9	Alta
0	0	0,2	0,8	0,7	0,3	0	10	Muy alta
0	0	0,1	0,8	0,5	0,4	0	11	Media
0	0	0,1	0,8	0,5	0,4	0	12	Media
0,1	0	0,1	0,8	0,7	0,3	0	1	Media
0	0	0,1	0,8	0,7	0,3	0	2	Muy baja
0	0	0,2	0,8	0,7	0,3	0	3	Media
0	0	0,2	0,7	0,7	0,3	0	4	Baja
0,1	0	0,2	0,7	0,7	0,3	0	5	Media
0,1	0	0,2	0,8	0,7	0,3	0	6	Muy alta
0	0	0,1	0,8	0,7	0,3	0	7	Muy alta
0	0	0,2	0,8	0,8	0,2	0	8	Baja
0	0	0,1	0,8	0,8	0,2	0	9	Alta

0,1	0	0,1	0,8	0,7	0,3	0	10	Media
0	0	0,1	0,8	0,7	0,3	0	11	Muy baja
0	0	0,1	0,8	0,8	0,2	0	12	Baja
0	0	0,1	0,9	0,7	0,3	0	1	Muy baja
0,1	0	0,1	0,8	0,7	0,3	0	2	Alta
0	0	0,1	0,8	0,7	0,3	0	3	Alta
0	0	0,1	0,9	0,7	0,3	0	4	Alta
0,1	0	0,1	0,8	0,6	0,4	0	5	Media
0	0	0,1	0,8	0,6	0,4	0	6	Muy alta

Tabla 15. Tabla de análisis para Hipertensión en el Hospital General Universitario Japonés

Fuente: (Elaboración Propia)

Observaciones en los atributos condición de la tabla de análisis:

De 16 a 30 años	De la tabla de análisis se puede observar que una pequeña parte de la población se encuentra en la edad de edad entre 16 a 30 años, cuyo porcentaje es 6.42%.
De 1 a 10 años	No se encuentran registrados pacientes de esa edad en el establecimiento de salud.
De 31 a 45 años	El porcentaje de pacientes registrados en ésta edad fue 13.18%.
De 45 años en adelante	80% de personas en esta edad que presentaron la enfermedad. Éste es el porcentaje más alto obtenido, sugiere que la mayor cantidad de personas propensas a tener este mal son mayores de 45 años.
Femenino	Las mujeres que presentaron la enfermedad tienen como porcentaje 69.55% de la población en promedio. Más de la mitad.
Masculino	El porcentaje de varones fue 30.90%. Se puede observar existe diferencia para la enfermedad entre varones y mujeres. Las mujeres son más afectadas con éste mal.
Menores a 1 año	No se encuentran registrados pacientes de esta edad en el establecimiento de salud.
Mes	Los meses varían mucho. Para un mismo mes los valores de prevalencia no son estables, es decir existe meses como junio que presenta valores opuestos de prevalencia. Por lo tanto podemos decir que no existe una relación marcada entre meses y prevalencia. Los datos se presentan la siguiente tabla:

Mes	Prevalencia
1	Muy baja
1	Media
1	Muy baja
2	Alta
2	Muy baja
2	Muy baja
3	Muy baja
3	Media
3	Alta
4	Alta
4	Baja
4	Muy baja
5	Muy baja
5	Media
5	Media
6	Muy alta
6	Muy alta
6	Muy baja
-	Muy baja
7	Muy alta
8	Baja
8	Media
9	Alta
9	Alta
10	Media
10	Muy alta
11	Medi
11	Muy baja
12	Baja
12	Media

Tabla 16. Meses y prevalencia de la tabla de análisis para Hipertensión en el Hospital General Universitario Japonés

Fuente: (Elaboración Propia)

El sistema produjo los siguientes reductos.

Reducto	Atributos
Reducto 1	Masculino, mes, de 16 a 30 años, de 31 a 45 años.
Reducto 2	De 31 a 45 años, De 16 a 30 años, mes, femenino.

Tabla 17. Reductos obtenidos de la tabla de análisis para Hipertensión en el Hospital General Universitario Japonés

Fuente: (Elaboración Propia)

El núcleo proporcionado por el sistema de información es: mes, personas de 16 a 30 años y personas entre la edad de 31 a 45 años.

Se puede utilizar cualquiera de los reductos proporcionados para realizar predicciones. Escogeremos los atributos del reducto 2 para entrenamiento y predicción.

Resultados

El reducto 2 con un entrenamiento de 30000 iteraciones produjo un estadístico t de -0.4439 .

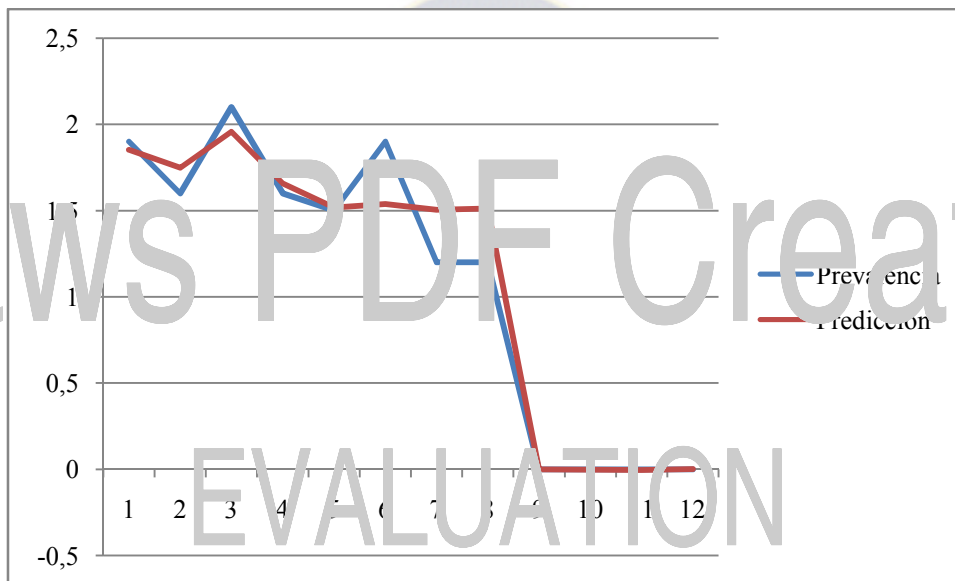


Figura 37. Resultados de predicción de reducto 2 en prueba 3
Fuente: (Elaboración Propia)

Como $t = -0.4439 \leq t_{0.025,11} = 2.201$ y $t = -0.4439 > t_{0.025,11} = -2.201$ no se puede rechazar la hipótesis que la diferencia sea 0.

El reducto escogido presenta datos y un error aceptable al nivel de confianza propuesto. Los futuros estudios deberán enfocarse en las variables del reducto. Por ejemplo se puede realizar campañas preventivas a personas dentro del rango de edad de 31 a 45 años, de 16 a 30 años, mujeres y mes. La red neuronal creada para este ejemplo puede ser entrenada sucesivas veces mejorando el estadístico t para que sea cercano a 0.

4.9.4 Prueba 4

Enfermedad: Conjuntivitis

Código de enfermedad CIE10: H10

Centro de salud: Instituto Nacional de Oftalmología – La Paz

Código del centro de salud: 200080

Periodo de estudio: 2007 a 2008

Tabla de análisis:

De 16 a 30 años	De 1 a 15 años	De 31 a 45 años	De 45 años en adelante	Femenino	Masculino	Menores a 1 año	Mes	Prevalencia
0,1	0	0,2	0,8	0,7	0,3	0	1	Media
0	0	0,1	0,8	0,7	0,3	0	2	Muy baja
0	0	0,2	0,8	0,7	0,3	0	3	Muy baja
0	0	0,2	0,7	0,7	0,3	0	4	Muy baja
0,1	0	0,2	0,7	0,7	0,3	0	5	Baja
0,1	0	0,2	0,8	0,7	0,3	0	6	Media
0,1	0	0,1	0,8	0,7	0,3	0	7	Media
0	0	0,2	0,8	0,8	0,2	0	8	Alta
0	0	0,1	0,8	0,8	0,2	0	9	Muy alta
0,1	0	0,1	0,8	0,7	0,3	0	10	Alta
0	0	0,1	0,8	0,7	0,3	0	11	Muy alta
0	0	0,1	0,3	0,3	0,2	0	12	Muy alta
0	0	0,1	0,9	0,7	0,3	0	1	Muy alta
0,1	0	0,1	0,8	0,7	0,3	0	2	Baja
0	0	0,1	0,8	0,7	0,3	0	3	Baja
0	0	0,1	0,9	0,7	0,3	0	4	Muy baja
0,1	0	0,1	0,8	0,6	0,4	0	5	Muy alta
0	0	0,1	0,8	0,6	0,4	0	6	Muy alta

Tabla 18. Tabla de análisis para Conjuntivitis en el Instituto Nacional de Oftalmología – La Paz

Fuente: (Elaboración Propia)

Observaciones en los atributos condición de la tabla de análisis:

De 16 a 30 años	De la tabla de análisis se puede observar que una considerable parte de la población es de edad entre 16 a 30 años, cuyo porcentaje es 3.125%.
De 1 a 15 años	No se presentaron personas en que categoría,.
De 31 a 45	El porcentaje de pacientes registrados en ésta edad fue

años	13.125%.
De 45 años en adelante	73.8% de personas en esta edad que presentaron la enfermedad. Éste es el porcentaje más alto obtenido.
Femenino	Las mujeres que presentaron la enfermedad tienen como porcentaje 70.50% de la población en promedio. Más de la mitad.
Masculino	El porcentaje de varones fue 29.5 %. Se puede observar existe diferencia para la enfermedad entre varones y mujeres. Las mujeres son mas afectadas con éste mal.
Menores a 1 año	No se registraron casos en esta edad.
Mes	Los meses que más prevalencia tuvieron son octubre, noviembre y diciembre. Sugiere que durante estos meses las personas acuden más a este hospital para el tratamiento de la enfermedad.

Mes	Prevalencia
1	Media
1	Muy alta
2	Baja
2	Muy baja
3	Muy baja
3	Baja
4	Muy baja
4	Muy baja
5	Baja
5	Muy alta
6	Muy alta
6	Media
7	Media
8	Alta
9	Muy alta
10	Alta
11	Muy alta
12	Muy alta

*Tabla 19. Meses y prevalencia de la tabla de análisis para Conjuntivitis en el Instituto Nacional de Oftalmología – La Paz
Fuente: (Elaboración Propia)*

El sistema produjo los siguientes reductos:

Reducto	Atributos
Reducto 1	De 31 a 45 años. De 16 a 30 años, mes.

*Tabla 20. Reductos obtenidos de la tabla de análisis para Conjuntivitis en el Instituto Nacional de Oftalmología – La Paz
Fuente: (Elaboración Propia)*

El núcleo proporcionado por el sistema de información es: mes, personas de 16 a 30 años y personas entre la edad de 31 a 45 años.

Solamente podemos utilizar un reducto proporcionado por el sistema para realizar predicciones.

Resultados

El reducto 1 con un entrenamiento de 1000 iteraciones produjo un estadístico t de 2.029.



*Figura 38. Resultados de predicción de reducto 1 en prueba 4
Fuente: (Elaboración Propia)*

Como $t = 2.029 < t_{0.025,11} = 2.201$ y $t = 2.029 > t_{0.025,11} = -2.201$ no se puede rechazar la hipótesis que la diferencia sea 0.

El reducto escogido presenta datos y un error aceptable al nivel de confiabilidad propuesto. Se debe realizar un especial análisis en personas entre edades de 16 a 45 años para campañas, teniendo a los meses de noviembre y diciembre como relevantes. La red neuronal creada para este ejemplo puede

ser entrenada sucesivas veces mejorando el estadístico t para que sea cercano a 0.

4.9.5 Prueba 5

Enfermedad: Infección de vías urinarias

Código de enfermedad CIE10: N39.0

Centro de salud: Hospital General Universitario Japonés – Santa Cruz de la Sierra

Código del centro de salud: 700006

Periodo de estudio: 2007 a 2008

Tabla de análisis:

De 16 a 30 años	De 1 a 15 años	De 31 a 45 años	De 45 años en adelante	Femenino	Masculino	Menores de 1 año	Mes	Prevalencia
0,3	0,3	0,3	0	1	0	0,7	1	Muy baja
0	0,5	0	0	0,5	0,5	0,5	2	Muy baja
0,3	0,2	0,2	0,2	0,8	0,2	0,1	3	Muy alta
0,4	0,1	0,2	0,1	0,9	0,1	0,1	4	Media
0,3	0,2	0,2	0,2	0,9	0,1	0	5	Media
0,3	0,2	0,2	0,1	0,3	0,2	0,1	6	Alta
0,4	0,2	0,2	0,1	0,8	0,2	0	7	Alta
0,2	0,2	0,2	0,3	0,8	0,2	0,1	8	Muy alta
0,4	0,2	0,1	0,2	0,8	0,2	0,1	9	Muy alta
0,2	0,2	0,2	0,4	0,9	0,1	0,1	10	Muy alta
0,1	0,2	0,2	0,3	0,7	0,3	0,1	11	Alta
0,3	0,3	0,1	0,2	0,8	0,2	0	12	Alta
0,3	0,3	0,1	0,3	0,3	0,7	0	1	Muy alta
0,3	0,2	0,2	0,2	0,3	0,7	0,1	2	Muy baja
0,2	0,2	0,2	0,2	0,3	0,7	0,1	3	Muy baja
0,3	0,2	0,1	0,3	0,8	0,2	0,1	4	Muy baja
0,5	0,2	0	0,2	0,8	0,2	0,1	5	Muy baja
0,3	0,1	0,2	0,3	0,8	0,2	0,1	6	Muy baja

Tabla 21. Tabla de análisis para Infección de vías urinarias en el Hospital General Universitario Japonés
Fuente: (Elaboración Propia)

Observaciones en los atributos condición de la tabla de análisis:

De 16 a 30 años	De la tabla de análisis se puede observar que 26.67% en porcentaje tiene de 16 a 30 años. Este es el porcentaje más alto obtenido.
De 1 a 15 años	Para personas en este rango de edad el promedio expresado en porcentaje es 20.56%.
De 31 a 45 años	El porcentaje 15.56% de pacientes registrados en este salud.
De 45 años en adelante	21.67% de personas en esta edad que presentaron la enfermedad.
Femenino	Las mujeres que presentaron la enfermedad tienen como porcentaje 80.55% de la población en promedio. Este es un considerable número señalando que la enfermedad tiene mayor presencia en mujeres que en varones.
Masculino	El porcentaje de varones fue 19.44%. Existe diferencia para la enfermedad entre varones y mujeres. Las mujeres son más afectadas con este mal.
Menores a 1 año	Sólo el 13.53% de la población afectada con esa enfermedad es menor de 1 año.
Mes	Los meses varían mucho. Solamente en febrero la prevalencia es muy baja, para los demás meses la prevalencia es variable. Los datos se presentan en la siguiente tabla.

Mes	Prevalencia
1	Muy baja
1	Muy alta
2	Muy baja
2	Muy baja
3	Muy alta
3	Muy baja
4	Muy baja
4	Media
5	Media
5	Muy baja
6	Muy baja
6	Alta
7	Alta
8	Muy alta
9	Muy alta
10	Muy alta
11	Alta
12	Alta

Tabla 22. Meses y prevalencia de la tabla de análisis para Infección de vías urinarias en el Hospital General Universitario Japonés

Fuente: (Elaboración Propia)

El sistema produjo los siguientes reducidos:

Reducto	Atributos
Reducto 1	De 45 en adelante años, mes, de 16 a 30 años.
Reducto 2	Mes, de 1 a 45 años.
Reducto 3	Mes, de 1 a 15 años, de 16 a 30 años.

Tabla 23. Reducidos obtenidos de la tabla de análisis para Infección de vías urinarias en el Hospital General Universitario Japonés

Fuente: (Elaboración Propia)

El núcleo proporcionado por el sistema de información es mes. El atributo más importante a estudiar. Se selecciona en reducto 3 porque en porcentaje la mayoría de personas las variables de edad incluyen a la mayor cantidad de personas que en los otros dos.

Resultados

El reducto 3 con un entrenamiento de 10000 iteraciones produjo un estadístico t de -0.9473 .

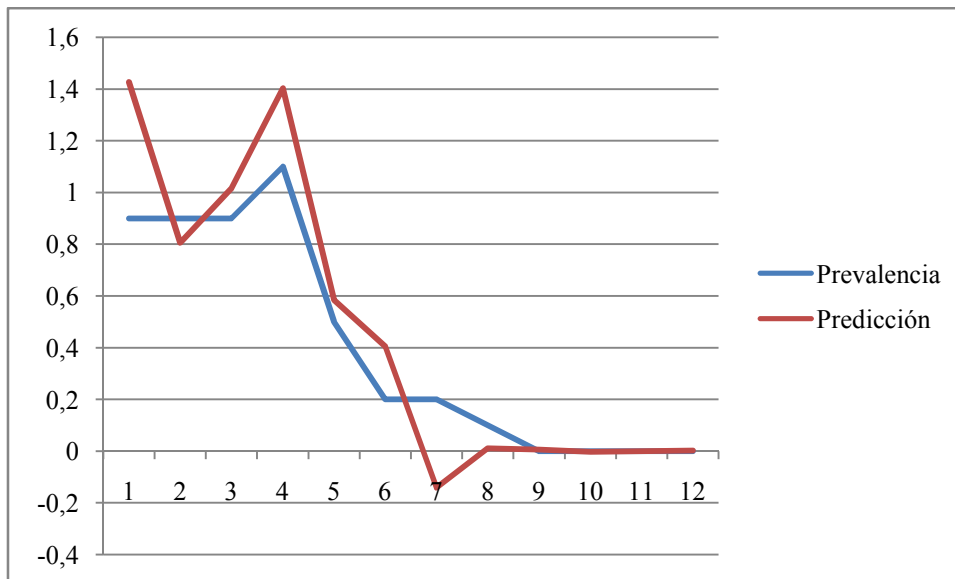


Figura 39. Resultados de predicción de reducto 3 en prueba 5
Fuente: (Elaboración Propia)

Como $t = -0.9473 < t_{0.025, 11} = 2.201$ y $t = -0.9473 > -t_{0.025, 11} = -2.201$ no se puede rechazar la hipótesis que la diferencia sea 0.

El reducto escogido presenta datos y un error aceptable al nivel de confiabilidad propuesto. Los futuros estudios deberán enfocarse en las variables del reducto. Se debe realizar un especial análisis en personas entre las edades de 1 a 30 años para campañas, teniendo un estudio posterior de meses. La red neuronal creada para este ejemplo puede ser entrenada sucesivas veces mejorando el estadístico t para que sea cercano a 0.

4.9.6 Prueba 6

Enfermedad: Cefalea

Código de enfermedad CIE10: R51

Centro de salud: Hospital Santa Bárbara - Sucre

Código del centro de salud: 100030

Periodo de estudio: 2006 a 2008

Tabla de análisis:

De 16 a 30 años	De 1 a 15 años	De 31 a 45 años	De 45 años en adelante	Femenino	Masculino	Menores a 1 año	Mes	Prevalencia
0,4	0	0,5	0,1	0,6	0,4	0	1	Muy baja
0,3	0	0,5	0,2	0,5	0,5	0	2	Muy alta
0,4	0,1	0,3	0,1	0,1	0,1	0	3	Media
0,1	0,1	0,2	0,1	0,1	0,3	0	4	Muy alta
0,1	0,1	0,3	0,3	0,1	0,3	0	5	Baja
0,1	0	0,3	0,3	0,1	0,4	0	6	Media
0,2	0,2	0,3	0,2	0,6	0,4	0	7	Media
0,3	0,1	0,3	0,3	0,7	0,3	0	8	Muy alta
0,3	0,1	0,3	0,3	0,6	0,4	0	9	Alta
0,3	0	0,2	0,3	0,5	0,4	0	10	Muy alta
0,2	0,2	0,2	0,2	0,5	0,4	0	11	Muy baja
0,4	0	0,1	0,3	0,5	0,4	0	12	Muy baja
0,4	0,3	0,2	0,1	0,5	0,5	0	1	Muy baja
0,4	0,1	0,2	0,3	0,8	0,2	0	2	Muy alta
0,5	0,1	0,2	0,2	0,7	0,3	0	3	Muy alta
0,4	0,1	0,2	0,3	0,5	0,5	0	4	Baja
0,4	0	0,2	0,4	0,7	0,3	0	5	Baja
0,4	0	0,3	0,3	0,6	0,4	0	6	Media
0,3	0,1	0,1	0,4	0,7	0,3	0	7	Alta
0,2	0,2	0,1	0,5	0,5	0,4	0	8	Muy alta
0,5	0,1	0,2	0,2	0,5	0,5	0	9	Media
0,5	0	0,3	0,2	0,7	0,3	0	10	Muy alta
0,4	0,1	0,2	0,3	0,7	0,3	0	11	Muy alta
0,3	0	0,2	0,1	0,7	0,3	0	12	Alta
0,3	0,2	0,3	0,2	0,8	0,2	0	1	Baja
0,5	0	0,3	0,1	0,5	0,5	0	2	Muy baja
0,5	0,1	0,1	0,3	0,7	0,3	0	3	Muy alta
0,4	0	0,1	0,2	0,7	0,3	0	4	Media
0,3	0,1	0,4	0,2	0,7	0,3	0	5	Muy baja
0,3	0	0,3	0,4	0,8	0,2	0	6	Media

Tabla 24. Tabla de cefalea en el Hospital Santa Bárbara - Sucre
Fuente: (Elaboración Propia)

Observaciones en los atributos condición de la tabla de análisis

De 16 a 30 años	De la tabla de análisis se puede observar que el 36.3% en porcentaje tiene de 16 a 30 años. Este es el porcentaje más alto obtenido.
De 1 a 15 años	Para personas en este rango de edad el promedio expresado en un porcentaje de 8%.
De 31 a 45 años	El porcentaje de pacientes registrados en ésta edad fue 27.7%.
De 45 años en adelante	26.7% de personas en esta edad que presentaron la enfermedad.
Femenino	Las mujeres que presentaron la enfermedad tienen como porcentaje 61.7% de la población total. Este es un considerable número señalando que la enfermedad tiene una mayor prevalencia en mujeres que en varones.
Masculino	El porcentaje de varones que padecen cefalea fue 35.3 %. Existe diferencia para la enfermedad entre varones y mujeres. Las mujeres son mas afectadas con este mal.
Menores a 1 año	No se presentaron casos de cefalea en menores de 1 año de edad.
Mes	El mes de marzo presenta valores altos.

Mes	Prevalencia
1	Muy baja
1	Muy baja
1	baja
2	Muy baja
2	Muy alta
2	Muy alta
3	Media
3	Muy alta
3	Muy alta
4	Media
4	Baja
4	Muy alta
5	Baja
5	Baja

5	Muy baja
6	Media
6	Media
6	Media
7	Media
7	Alta
8	Muy alta
8	Muy alta
9	Alta
9	Media
10	Muy alta
10	Muy alta
11	Muy baja
11	Muy alta
12	Alta
12	Muy baja

Tabla 5. Masas y prevalencia de la tabla de cefalea en el Hospital Santa Bárbara - Sucre
Fuente: (Elaboración Propia)

El sistema produjo los siguientes reductos:

Reducto	Atributos
Reducto 1	Mes, de 31 a 45 años, femenino.
Reducto 2	Mes, de 16 a 30 años, femenino, de 1 a 15 años.
Reducto 3	Mes, de 31 a 45 años, masculino
Reducto 4	Mes, de 16 a 30 años, masculino, de 1 a 15 años.

Tabla 26. Tabla de cefalea en el Hospital Santa Bárbara - Sucre
Fuente: (Elaboración Propia)

El núcleo proporcionado por el sistema de información es: mes. El atributo más importante a estudiar. Se selecciona el reducto 2 porque en porcentaje la mayoría de personas se encuentra de 16 a 30 años y la mayoría son mujeres.

Resultados

El reducto 2 con un entrenamiento de 10000 iteraciones produjo un estadístico t de -1.9740 .

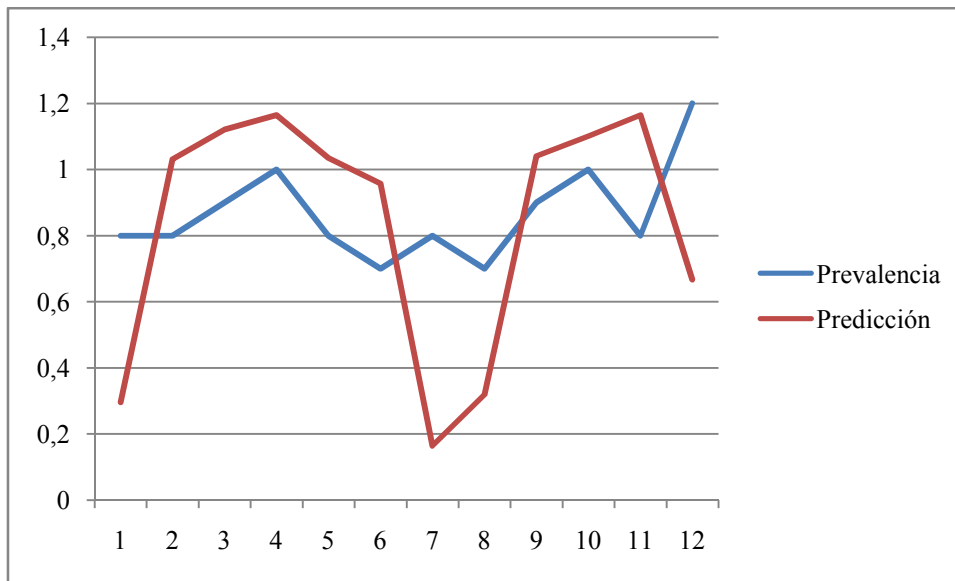


Figura 40. Resultados de predicción de reducto 3 en prueba 6
Fuente: (Elaboración Propia)

Como $t = -1.9740 < t_{0.025, 11} = 2.201$ y $t = -1.9740 > t_{0.975, 11} = -2.201$ no se puede rechazar la hipótesis que la diferencia sea 0.

El reducto escogido presenta datos y un estadístico aceptable al nivel de confiabilidad propuesto. Los futuros estudios deberán enfocarse en las variables del reducto. Se debe realizar un especial análisis en personas entre las edades de 1 a 30 años, teniendo un estudio posterior de meses. La red neuronal creada para este ejemplo puede ser entrenada sucesivas veces mejorando el estadístico t para que sea cercano a 0.

4.9.7 Prueba 7

Enfermedad: Faringitis aguda

Código de enfermedad CIE10: J02

Centro de salud: Hospital Santa Bárbara - Sucre

Código del centro de salud: 100030

Periodo de estudio: 2006 a 2008

Tabla de análisis:

De 16 a 30 años	De 1 a 15 años	De 31 a 45 años	De 45 años en adelante	Femenino	Masculino	Menores a 1 año	Mes	Prevalencia
0,2	0,3	0,2	0,2	0,5	0,5	0	1	Muy baja
0,2	0,3	0,1	0,2	0,4	0,6	0,2	2	Baja
0,1	0,4	0,1	0,1	0,1	0,7	0,1	3	Media
0,1	0,5	0,3	0,1	0,1	0,5	0	4	Media
0,1	0,1	0,2	0,3	0,3	0,7	0,1	5	Media
0,1	0,1	0,2	0,2	0,3	0,4	0	6	Baja
0,2	0,4	0,1	0,2	0,7	0,3	0,1	7	Alta
0,3	0,1	0,3	0,2	0,5	0,5	0	8	Baja
0,3	0,3	0,2	0,1	0,5	0,5	0	9	Muy alta
0,4	0,2	0,1	0,2	0,7	0,3	0	10	Media
0,2	0,4	0,1	0,1	0,4	0,6	0	11	Muy alta
0,4	0,2	0,1	0,1	0,5	0,5	0	12	Media
0,2	0,3	0,2	0,2	0,5	0,5	0	1	Baja
0,2	0,1	0,2	0,5	0,6	0,4	0	2	Muy baja
0,5	0,3	0,1	0,1	0,5	0,5	0	3	Baja
0,5	0,1	0,2	0,2	0,5	0,5	0	4	Media
0,3	0,2	0,3	0,1	0,5	0,5	0	5	Muy alta
0,3	0,2	0,2	0,2	0,5	0,5	0	6	Alta
0,3	0,1	0,1	0,2	0,4	0,6	0	7	Media
0,4	0,3	0,1	0,1	0,5	0,4	0	8	Media
0,3	0,2	0,2	0,3	0,4	0,6	0	9	Media
0,4	0,3	0,2	0,2	0,4	0,6	0	10	Baja
0,3	0,3	0,2	0,2	0,6	0,4	0	11	Media
0,3	0,2	0,1	0,3	0,6	0,4	0	12	Baja
0,3	0,2	0,2	0,2	0,5	0,5	0	1	Baja
0,2	0,2	0,1	0,5	0,6	0,4	0	2	Muy baja
0,4	0,3	0,1	0,1	0,5	0,5	0	3	Media
0,3	0,3	0,1	0,2	0,4	0,6	0	4	Media
0,3	0,2	0,2	0,3	0,5	0,5	0	5	Media
0,4	0,2	0,2	0,2	0,5	0,5	0	6	Muy alta

Tabla 27. Tabla de análisis de Faringitis aguda en el Hospital Santa Bárbara - Sucre

Fuente: (Elaboración Propia)

Observaciones en los atributos condición de la tabla de análisis:

De 16 a 30 años	De la tabla de análisis se puede observar que el 29.7% de pacientes en porcentaje tiene de 16 a 30 años.
De 1 a 15 años	Para personas en este rango de edad el porcentaje expresado en porcentaje es 24.3%.
De 31 a 45 años	El porcentaje de pacientes registrados en ésta edad fue 18.7%.
De 45 años en adelante	21% de personas en esta edad que presentaron la enfermedad.
Femenino	Las mujeres que presentaron la enfermedad tienen como porcentaje 50.3% de la población en promedio. Este hecho no es muy considerable ya que no muestra una relevante diferencia entre hombres y mujeres.
Masculino	El porcentaje de varones que padecen de ella fue 49.7%.
Menores a 1 año	Solamente el 1.6% de la población fue catalogada en este rango de edad.
Mes	El mes de mayo presenta valores de muy alto y medio en tres observaciones. Los demás meses no tienen definido su comportamiento.

Mes	Prevalencia
1	Muy Baja
1	Baja
1	Baja
2	Muy baja
2	Muy baja
2	Baja
3	Media
3	Baja
3	Media
4	Media
4	Media
4	Media
5	Media
5	Muy alta
5	Media
6	Muy alta

6	Alta
6	Baja
7	Alta
7	Media
8	Media
8	Baja
9	Muy alta
9	Media
10	Baja
10	Media
11	Muy alta
11	Media
12	Baja
12	Media

Tabla 28. Tabla de mes y prevalencia de la tabla de análisis de Faringitis aguda en niños del Hospital Santa Bárbara - Sucre

Fuente: (Elaboración Propia)

El sistema produjo los siguientes reductos:

Reducto	Atributos
Reducto 1	De 45 en adelante años, masculino, mes, de 16 a 30 años.
Reducto 2	De 16 a 30 años, de 45 en adelante años, mes femenino.
Reducto 3	De 16 a 30 años, mes, de 15 en adelante años, de 1 a 15 años.
Reducto 4	De 16 a 30 años, mes, de 31 a 45 años, femenino.
Reducto 5	Mes, de 16 a 30 años, de 31 a 45 años, masculino.
Reducto 6	De 31 a 45 años, de 16 a 30 años, de 1 a 15 años, mes.

Tabla 29. Reductos obtenidos de la Tabla de análisis de Faringitis aguda en el Hospital Santa Bárbara - Sucre

Fuente: (Elaboración Propia)

El núcleo proporcionado por el sistema de información es: mes y edades entre 16 y 30 años. Se selecciona el reducto 6 porque en porcentaje la mayoría de personas se encuentra en edades entre 1 y 45 años.

Resultados

El reducto 6 con un tamaño de muestra de 20000 iteraciones produjo un estadístico t de -1.7906 .

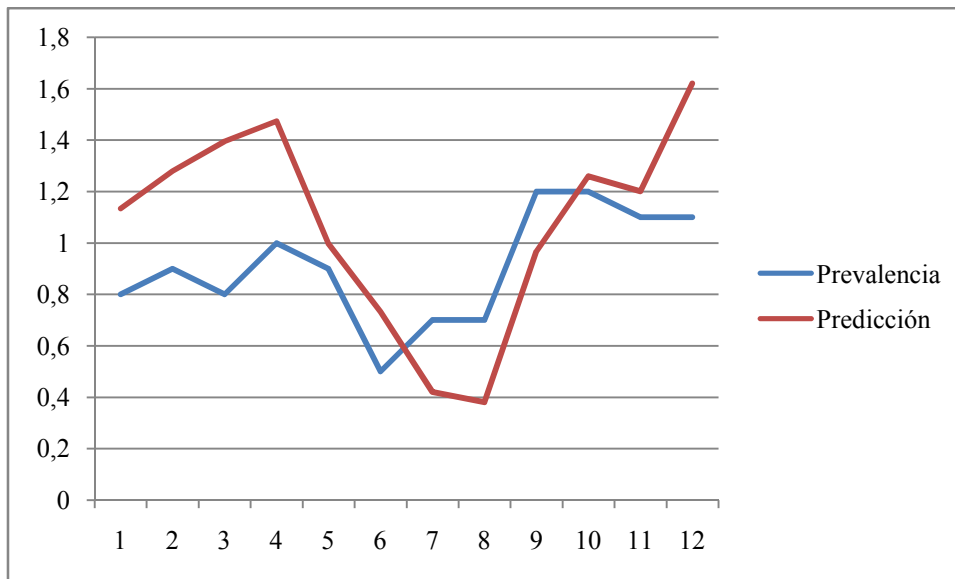


Figura 41. Resultados de predicción de reducto 6 en prueba 7
Fuente: (Elaboración Propia)

Como $t = -1.7906 < t_{0.025, 11} = 2.201$ y $t = -1.7906 > t_{0.975, 11} = -2.201$ no se puede rechazar la hipótesis que la diferencia sea 0.

El reducto escogido presenta datos y un estadístico aceptable al nivel de confiabilidad propuesto. Los futuros estudios deberán enfocarse en las variables del reducto. Se debe realizar un especial análisis en personas entre las edades de 1 a 45 años para cambranas, teniendo un estudio posterior de meses.

VALUTAZIONE
EVALUATION
EVALUACIÓN
EVALUATION

4.9.8 Prueba 8

Enfermedad: Bronquitis aguda

Código de enfermedad CIE10: J20

Centro de salud: Instituto Nacional del Tórax – La Paz

Código del centro de salud: 200081

Periodo de estudio: 2007 a 2008

Tabla de análisis:

De 16 a 30 años	De 1 a 15 años	De 31 a 45 años	De 45 años en adelante	Femenino	Masculino	Menores a 1 año	Mes	Prevalencia
0,1	0,2	0,1	0,2	0,4	0,5	0,5	1	Muy baja
0,1	0,2	0,1	0,5	0,5	0,5	0,1	2	Baja
0,2	0,1	0,2	0,4	0,4	0,5	0,1	3	Media
0,3	0,1	0,2	0,4	0,4	0,5	0,1	4	Alta
0,3	0,1	0,2	0,4	0,4	0,5	0,1	5	Muy alta
0,3	0,1	0,2	0,4	0,4	0,5	0,1	6	Muy alta
0,3	0,1	0,2	0,4	0,4	0,5	0,1	7	Muy alta
0,3	0,1	0,2	0,4	0,4	0,6	0,1	8	Media
0,2	0,1	0,2	0,4	0,3	0,6	0,1	9	Media
0,3	0,1	0,3	0,4	0,4	0,6	0,1	10	Alta
0,2	0,1	0,2	0,5	0,3	0,6	0,1	11	Baja
0,3	0,1	0,2	0,4	0,3	0,6	0,1	12	Muy alta
0,4	0,1	0,2	0,3	0,4	0,6	0,1	1	Muy baja
0,2	0	0,3	0,4	0,4	0,5	0	2	Baja
0,2	0,1	0,3	0,5	0,4	0,6	0	3	Media
0,2	0,1	0,3	0,4	0,4	0,5	0	4	Muy alta
0,2	0,1	0,2	0,5	0,4	0,5	0	5	Media
0,2	0	0,3	0,4	0,5	0,5	0,1	6	Baja

Tabla 30. Tabla de análisis de bronquitis aguda en el instituto nacional del tórax

Fuente: (Elaboración Propia)

Observaciones en los atributos condición de la tabla de análisis:

De 16 a 30 años	De la tabla de análisis se puede observar que el 22.8% en porcentaje tiene de 16 a 30 años. Este es el porcentaje más alto obtenido.
De 1 a 15 años	Para personas en este rango de edad el promedio expresado en porcentaje es 8.9%.
De 31 a 45	El porcentaje promedio de pacientes registrados en ésta edad

años	fue 22.2 %.
De 45 años en adelante	40.5% de personas en esta edad que presentaron la enfermedad.
Femenino	Las mujeres que presentaron la enfermedad tienen como porcentaje 41.7% de la población en promedio.
Masculino	El porcentaje de varones que padecen cefalea fue 53.3 %. Existe diferencia para la enfermedad entre varones y mujeres. Los varones son mas afectadas con éste mal.
Menores a 1 año	7.2 % de todos los pacientes en promedio tiene de 1 a 15 años.
Mes	Los meses que presentan desde media o alta o muy alta prevalencia en las observaciones son abril y mayo.

Mes	Prevalencia
1	Muy baja
1	Muy baja
2	Baja
2	Baja
3	Media
3	Media
4	Muy alta
4	Alta
5	Muy alta
5	Media
6	Baja
6	Muy alta
7	Muy alta
8	Media
9	Media
10	Alta
11	Baja
12	Muy alta

Tabla 31. Meses y prevalencia de la tabla de análisis de bronquitis aguda en el Instituto Nacional de Tóxicos
Fuente: (Elaboración Propia)

El sistema produjo los siguientes reductos:

Reducto	Atributos
Reducto 1	De 31 a 45 años, menores a 1 años, de 1 a 15 años, de 45 en adelante años, masculino.
Reducto 2	De 1 a 15 años, mes, de 31 a 45 años.
Reducto 3	Mes, femenino, de 31 a 45 años.
Reducto 4	Femenino, de 31 a 45 años, de 1 a 15 años, de 45 en adelante años, masculino.
Reducto 5	Mes, de 45 en adelante años, de 31 a 45 años.

Tabla 32. Reductos obtenidos de la tabla de análisis de bronquitis aguda en el instituto nacional del tórax
Fuente: (Elaboración Propia)

El núcleo proporcionado por el sistema de información es: el rango de edad de 31 a 45 años. El atributo más importante a estudiar. Se selecciona el reducto 5 para realizar las prueba .

Resultados:

El reducto 5 con un entrenamiento de 10000 iteraciones produjo un estadístico t de -2.0210 .

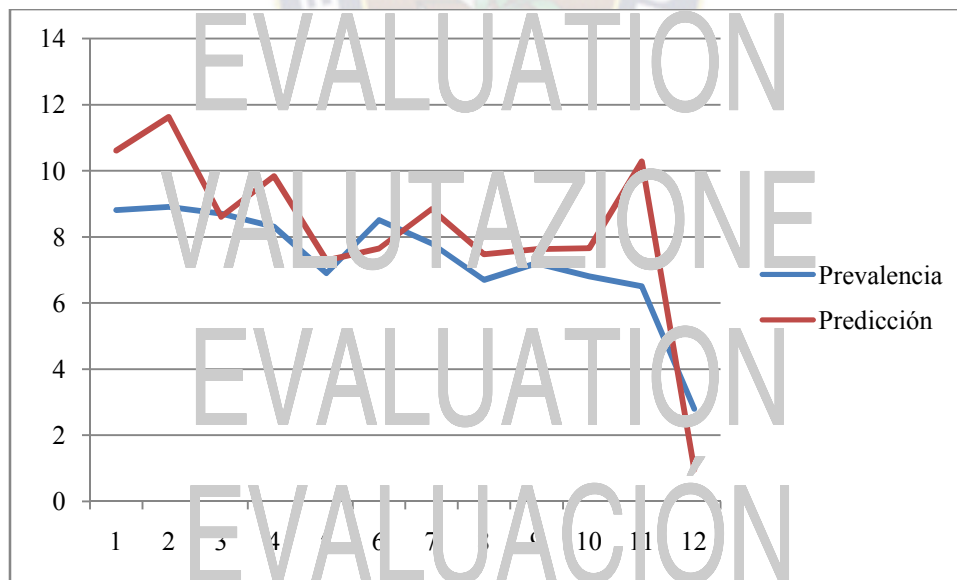


Figura 42. Resultados de predicción de reducto 5 en prueba 8
Fuente: (Elaboración Propia)

Como $t = -2.0210 < t_{0.025, 11} = 2.201$ y $t = -2.0210 > -t_{0.025, 11} = -2.201$ no se puede rechazar la hipótesis que la diferencia sea 0.

El reducto escogido presenta datos y un error aceptable al nivel de confiabilidad propuesto. Los futuros estudios deberán enfocarse en las variables del reducto. Se debe realizar un especial análisis en personas entre las edades de 31 y 45 años y en adelante para futuras campañas, teniendo un estudio posterior de meses.

Jaws PDF Creator

EVALUATION
VALUTAZIONE
EVALUATION
EVALUACIÓN
EVALUATION

4.9.9 Prueba 9

Enfermedad: Dorsalgia

Código de enfermedad CIE10: M54

Centro de salud: Hospital Boliviano Holandés – El Alto

Código del centro de salud: 200143

Periodo de estudio: 2005 a 2008

Tabla de análisis:

De 16 a 30 años	De 1 a 15 años	De 31 a 45 años	De 45 años en adelante	Femenino	Masculino	Menores a 1 año	Mes	Prevalencia
1	0	0	0	0	1	0	1	Muy baja
0	0	0	0	0	0	0	2	Muy baja
0,4	0	0,3	0,3	0,4	0,4	0	3	Muy alta
0,3	0,1	0,3	0,3	0,4	0,4	0	4	Muy alta
0	0	0,3	0,3	0,4	0,4	0	5	Muy alta
0,2	0	0,4	0,3	0,4	0,4	0	6	Muy alta
0,2	0	0,3	0,4	0,4	0,6	0	7	Muy alta
0,3	0	0,4	0,2	0,5	0,5	0	8	Muy alta
0,3	0	0,4	0,3	0,5	0,5	0	9	Muy alta
0,2	0	0,4	0,3	0,6	0,4	0	10	Muy alta
0,2	0	0,3	0,4	0,5	0,5	0	11	Alta
0,3	0	0,3	0,3	0,5	0,5	0	12	Alta
0,3	0	0,4	0,3	0,5	0,5	0	1	Baja
0,3	0	0,4	0,3	0,5	0,5	0	2	Muy baja
0,3	0	0,3	0,4	0,6	0,4	0	3	Alta
0,4	0,1	0,3	0,3	0,4	0,5	0	4	Muy baja
0,3	0	0,3	0,4	0,6	0,4	0	5	Baja
0,3	0,1	0,3	0,3	0,6	0,4	0	6	Muy alta
0,3	0	0,3	0,3	0,5	0,5	0	7	Muy alta
0,3	0,1	0,3	0,3	0,4	0,6	0	8	Media
0,3	0	0,4	0,2	0,5	0,5	0	9	Media
0,3	0	0,2	0,4	0,6	0,4	0	10	Alta
0,2	0	0,3	0,4	0,5	0,5	0	11	Muy alta
0,2	0,1	0,3	0,3	0,5	0,5	0	12	Muy alta
0,2	0	0,3	0,4	0,5	0,5	0	1	Muy alta
0,4	0,1	0,2	0,3	0,4	0,6	0	2	Muy baja
0,4	0	0,4	0,2	0,5	0,5	0	3	Media
0,3	0	0,3	0,2	0,5	0,5	0	4	Media
0,3	0	0,3	0,3	0,4	0,6	0	5	Media
0,2	0	0,4	0,3	0,4	0,6	0	6	Muy alta

*Tabla 33. Tabla de análisis de dorsalgia en el Hospital Boliviano Holandés
Fuente: (Elaboración Propia)*

Observaciones en los atributos condición de la tabla de análisis:

De 16 a 30 años	De la tabla de análisis se puede observar que el 29.4%
De 1 a 15 años	Para personas en este rango de edad el promedio expresado en porcentaje es 2.8%. Lo que denota que la enfermedad incide de manera importante en niños y adolescentes.
De 31 a 45 años	El porcentaje de pacientes registrados en ésta edad fue 33.9%. Que expresa que la mayor cantidad de pacientes que padecen dorsalgia tienen de 31 a 45 años.
De 45 años en adelante	31.1% de personas en esta edad que presentaron dorsalgia. Estos datos muestran que es una enfermedad que se presenta con mayor frecuencia en personas adultas.
Femenino	Las mujeres que presentaron la enfermedad tienen como porcentaje 49.4% de la población en promedio.
Masculino	El porcentaje de varones que padecen dorsalgia fue 50.6%. Como se puede observar no existe una diferencia marcada entre hombres y mujeres que padecen dorsalgia ya que en porcentaje ambos se acercan al 50%.
Menores a 1 año	No se presentaron casos de dorsalgia en menores de 1 año de edad.
Mes	Los meses que guardan similitud con valores altos de prevalencia son: junio, julio, octubre, noviembre y diciembre.

Mes	Prevalencia
1	Muy baja
1	Baja
1	Muy alta
2	Muy baja
2	Muy baja
2	Muy baja
3	Muy alta
3	Alta
3	Media
4	Media
4	Muy baja
4	Muy alta
5	Muy alta

5	Baja
5	Media
6	Muy alta
6	Muy alta
6	Muy alta
7	Muy alta
7	Muy alta
8	Media
8	Muy alta
9	Muy alta
9	Media
10	Alta
10	Muy alta
11	Alta
11	Muy alta
12	Muy alta
12	Alta

Tabla 34. Meses y prevalencia de la tabla de análisis de dorsalgia en el Hospital Boliviano Holandés
Fuente: (Elaboración Propia)

El sistema produjo el siguiente reducto:

Reducto	Atributos
Reducto 1	De 45 adelante años, mes, de 31 a 45 años

Tabla 35. Tabla de análisis de dorsalgia en el Hospital Boliviano Holandés
Fuente: (Elaboración Propia)

El núcleo proporcionado por el sistema de información es: mes, de 45 en adelante años y de 31 a 45 años. Con éste reducto se realizará la prueba.

Resultados

El reducto 1 con un entrenamiento de 20000 iteraciones produjo un estadístico t de 1.9911.

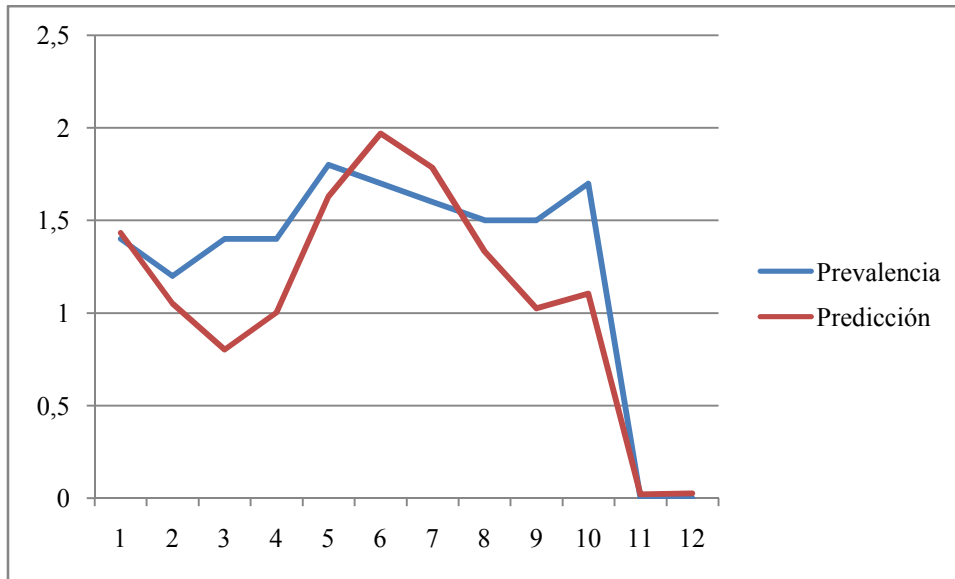


Figura 43. Resultados de predicción de reducto 1 en prueba 9

Fuente: (Elaboración Propia)

Como $t = -1.9911 < t_{0.025,11} = 2.201$ y $t = -1.9911 > t_{0.975,11} = -2.201$ no se puede rechazar la hipótesis que la diferencia sea 0.

El reducto escogido presenta datos y un error aceptable al nivel de confiabilidad propuesto. Los futuros estudios deberán enfocarse en las variables del reducto. Se debe realizar un especial análisis en personas entre las edades de 30 a 45 años y posterior para campañas, teniendo un estudio posterior de meses. La red neuronal creada para este ejemplo puede ser entrenada sucesivas veces mejorando el estadístico t para que sea cercano a 0.

4.9.10 Prueba 10

Enfermedad: Fiebre de origen desconocido

Código de enfermedad CIE10: R50

Centro de salud: Hospital de niños – Santa Cruz de la Sierra

Código del centro de salud: 700001

Periodo de estudio: 2007 a 2008

Tabla de análisis:

De 16 a 30 años	De 1 a 15 años	De 31 a 45 años	De 45 años en adelante	Femenino	Masculino	Menores a 1 año	Mes	Prevalencia
0	0,7	0	0	0,5	0,5	0,3	1	Muy alta
0	0,7	0	0	0,4	0,6	0,3	2	Muy alta
0	0,6	0	0	0,5	0,5	0,4	3	Alta
0	0,7	0	0	0,5	0,5	0,3	4	Muy alta
0	0,8	0	0	0,5	0,5	0,2	5	Media
0	0,7	0	0	0,5	0,5	0,2	6	Muy baja
0	0,7	0	0	0,5	0,5	0,3	7	Media
0	0,7	0	0	0,5	0,5	0,3	8	Media
0	0,6	0	0	0,4	0,6	0,4	9	Media
0	0,6	0	0	0,5	0,5	0,4	10	Alta
0	0,6	0	0	0,5	0,5	0,4	11	Muy alta
0	0,6	0	0	0,5	0,5	0,4	12	Muy alta
0	0,7	0	0	0,4	0,6	0,3	1	Media
0	0,7	0	0	0,5	0,5	0,3	2	Muy alta
0	0,8	0	0	0,5	0,5	0,2	3	Alta
0	0,8	0	0	0,4	0,5	0,2	4	Muy alta
0	0,7	0	0	0,5	0,5	0,3	5	Alta
0	0,7	0	0	0,5	0,5	0,3	6	Muy baja

Tabla 36. Tabla de análisis fiebre de origen desconocido en el Hospital de niños

Fuente: (Elaboración Propia)

Observaciones en los atributos condición de la tabla de análisis

De 16 a 30 años	No se presentaron casos en el Hospital señalado para esta edad.
De 1 a 15 años	Para niños y adolescente en este rango de edad el promedio expresado en porcentaje es 79.4%. Es el mayor porcentaje

	referente a edad observado.
De 31 a 45 años	No se presentaron casos en el Hospital señalado para esta edad
De 45 años en adelante	No se presentaron casos en el Hospital señalado para esta edad
Femenino	Las mujeres que presentaron la enfermedad tienen como porcentaje 47.7% de la población en promedio.
Masculino	El porcentaje de varones que padecen dicha enfermedad fue 52.2 %. Existe una diferencia estrecha entre varones y mujeres en porcentaje, lo cual implica que la enfermedad incide tanto en niñas como en niños sin una marcada diferencia.
Menores a 1 año	El 30% de los casos observado en porcentaje son niños menores a 1 año
Mes	Se observó que por lo general en los meses de junio y julio disminuye la prevalencia y durante todo el año presenta índices altos

Mes	Prevalencia
1	Muy alta
1	Mediana
2	Muy alta
2	Muy alta
3	Alta
3	Alta
4	Muy alta
4	Muy alta
5	Mediana
5	Alta
6	Muy baja
6	Muy baja
7	Mediana
8	Mediana
9	Mediana
10	Alta
11	Muy alta
12	Muy alta

*Tabla 37. Mes y prevalencia de la tabla de análisis fiebre de origen desconocido en el Hospital de niños
Fuente: (Elaboración Propia)*

El sistema produjo los siguientes reductos:

Reducto	Atributos
Reducto 1	De 1 a 15 años, masculino, mes.
Reducto 2	Masculino, menores a 1 año, mes.
Reducto 3	Mes femenino, menores a 1 año.
Reducto 4	Femenino, mes, de 1 a 15 años.

Tabla 38. Reductos obtenidos de la tabla de análisis fiebre de origen desconocido en el Hospital de niños

Fuente: (Elaboración Propia)

El núcleo proporcionado por el sistema de información es: mes. El atributo más importante a estudiar. Se selecciona el reducto 4 porque en porcentaje la mayoría de personas se encuentra de 1 a 15 años y la mayoría son mujeres.

Resultados

El reducto 4 con un entrenamiento de 10000 iteraciones produjo un estadístico t de 2.1279

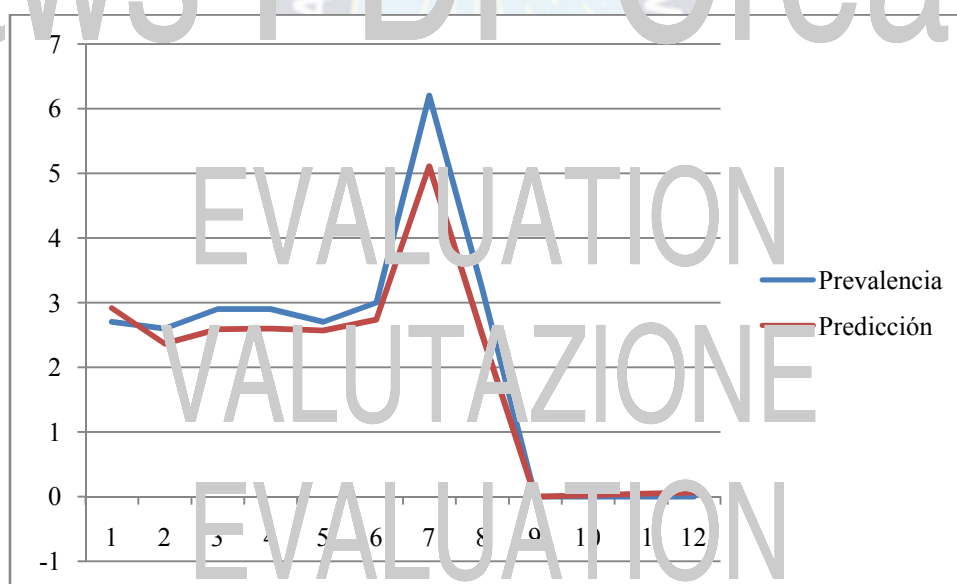


Figura 44. Resultados de predicción de reducto 4 en prueba 10

Fuente: (Elaboración Propia)

Como $t = 2.1279 < t_{0.025,11} = 2.201$ y $t = 2.1279 > t_{0.025,11} = -2.201$ no se puede rechazar la hipótesis que la diferencia sea 0.

El reducto escogido presenta datos y un error aceptable al nivel de confiabilidad propuesto. Los futuros estudios deberán enfocarse en las variables del reducto. Se debe realizar un especial análisis en personas entre las edades de 1 a 15 años para campañas, teniendo un estudio posterior de meses.

4.10 Conclusiones de las pruebas

Acerca del porcentaje de la población en los atributos y los reductos: de las 10 pruebas realizadas se puede observar que en todas existe en la población un sector que está más comprometido con la enfermedad ya sea por edad o por sexo. Es decir existe un sector donde la enfermedad almacena el porcentaje más alto de casos registrados en el establecimiento de salud. En las pruebas 1, 2, 3, 5, 6, 7, 8, 9 y 10 el atributo que almacena el mayor porcentaje de casos registrados pertenece a un reducto, pero en la prueba 4 no, Podemos decir que en su mayoría los atributos de los reductos obtenidos están relacionados con el porcentaje de casos registrados en las columnas de decisión.

Acerca de la predicción: se puede concluir para todas las pruebas que analizando valores en los atributos de reducto propuesto se puede entrenar una red neuronal acorde con la prueba de hipótesis y por lo tanto sea aceptado. Lo cual para las enfermedades analizadas fue posible construir un modelo de predicción de prevalencia con un 95% de confianza.

4.11 Ejemplo de predicción de prevalencia

Con el estudio de prevalencia de enfermedades, respecto al centro de salud y el desarrollo e implementación de un modelo de predicción, es necesario mostrar algunos resultados del mismo. Dichos valores servirán como pauta para realizar futuros estudios.

En esta sección se mostrarán predicciones de prevalencia con número de pacientes como valores de entrada para la primera enfermedad estudiada: diarrea y gastroenteritis de presunto origen infeccioso en el hospital Boliviano Holandés de la ciudad de El Alto.

El número de pacientes y con relación al tiempo se muestra en la siguiente tabla:

Año	Mes	Número de pacientes
2005	1	4
2005	2	2
2005	3	57

2005	4	54
2005	5	124
2005	6	83
2005	7	78
2005	8	98
2005	9	105
2005	10	111
2005	11	72
2005	12	108
2006	1	116
2006	2	72
2006	3	91
2006	4	74
2006	5	62
2006	6	71
2006	7	71
2006	8	64
2006	9	53
2006	10	61
2006	11	60
2006	12	73
2007	1	53
2007	2	43
2007	3	65
2007	4	49
2007	5	64
2007	6	72
2007	7	53
2007	8	88
2007	9	51
2007	10	58
2007	11	53
2007	12	44
2008	1	64
2008	2	42
2008	3	49
2008	4	50

Tabla 39. Número de pacientes con relación al tiempo en el Hospital Boliviano
 H. Francés de la ciudad de El Alto
 Fuente: (Elaboración Propia)

Los datos de predicción serán promedios de los meses, es decir el promedio de todas las observaciones para cada mes.

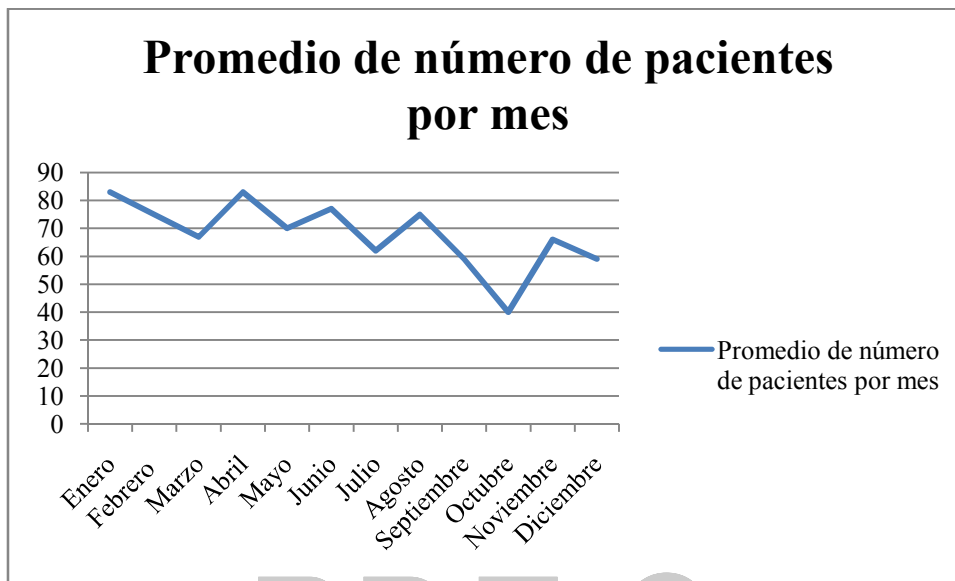


Figura 45. Promedio de pacientes que padecen diarrea y gastroenteritis de presunto origen infeccioso por meses en el Hospital Boliviano Holandés
Fuente: (Elaboración Propia)

Con estos promedios realizaremos la predicción tomando en cuenta los porcentajes hallados en el apartado 4.9.1 y el reducto 2: De 1 a 15 años, menores a 1 año, más femenino; es decir 45% de la población se encuentra entre 1 y 15 años, 22.3% de la población con menores a 1 año y 48.7% de la población son varones.

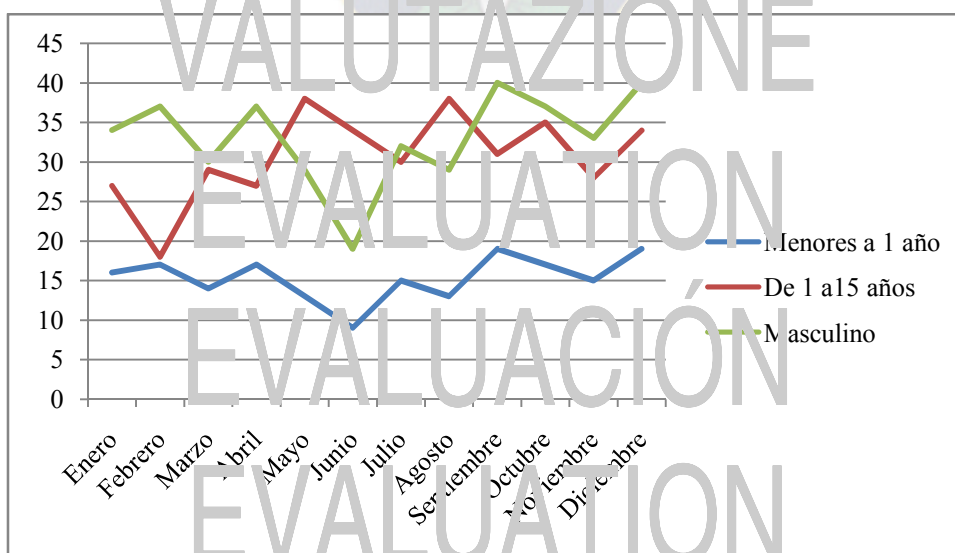


Figura 46. Pacientes que padecen diarrea y gastroenteritis de presunto origen infeccioso clasificados en menores a 1 años, de 1 a 15 años y masculino por promedio de meses en el Hospital Boliviano Holandés
Fuente: (Elaboración Propia)

Estos son los datos de entrada para la red neuronal del reducto 2 y las predicciones se muestran a continuación:

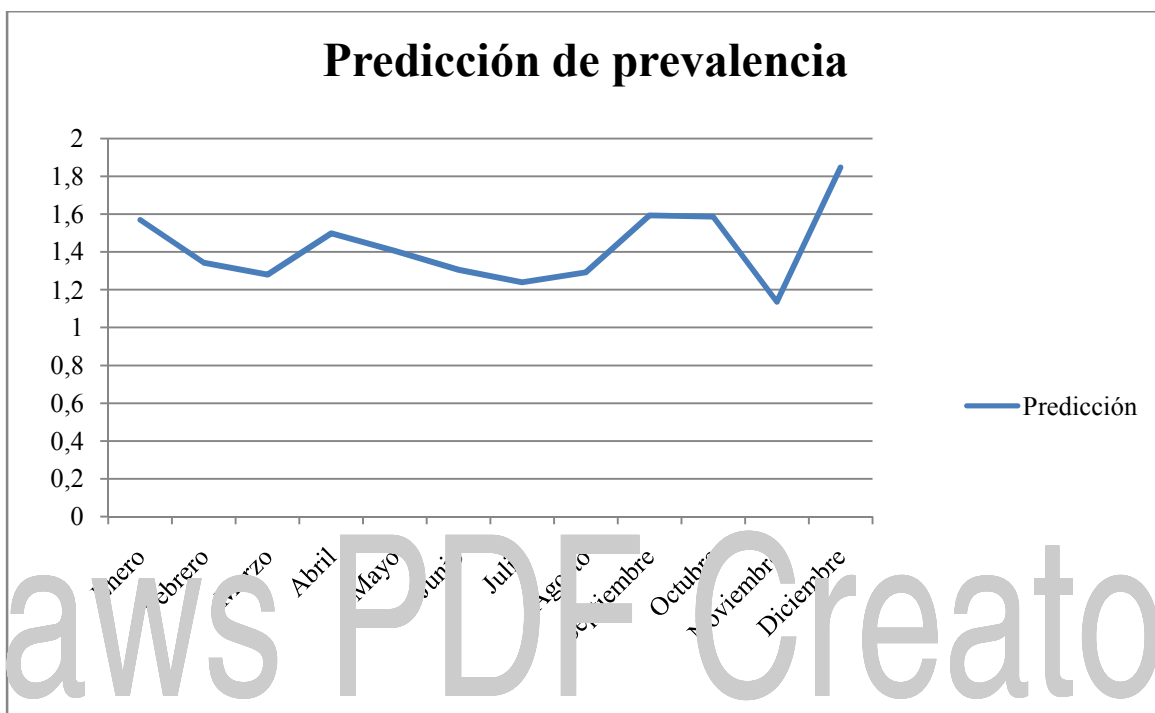


Figura 47. Predicción de prevalencia de diarrea y gastroenteritis de presunto origen infeccioso basada en promedio de pacientes por meses en el Hospital

*Boliviano Horacio
Fuente: (Elaboración Propia)*

Donde se observa que los meses con mayor índice de prevalencia son enero, abril, septiembre, octubre y diciembre además se observa el comportamiento de cada curva, por ejemplo, durante julio existe un descenso en la proporción de varones. Este dato facilitará futuros estudios en las variables del reducto mencionadas y posteriormente campañas preventivas.

CAPÍTULO 5

CONCLUSIONES

Jaws PDF Creator

EVALUATION
VALUTAZIONE
EVALUATION
EVALUACIÓN
Resumen

En este capítulo se presenta el estado de la hipótesis, conclusiones y recomendaciones para futuros trabajos.

5.1 Conclusiones generales

Acerca de los conjuntos aproximados:

- Para el desarrollo de un modelo informático basado en conjuntos aproximados es importante refinar o transformar datos para aplicar ésta técnica ya que los datos posibles de prevalencia son reales y son infinitos. En la presente tesis se propuso una nueva medida para datos de prevalencia con el fin de transformar los datos para un mejor análisis en la sección 3.2.

Acerca de redes neuronales

- Es importante considerar que las redes neuronales tienen un gran poder de predicción pero son algo incomprensibles, esta capacidad de predicción aumenta con el número de neuronas y las capas. En efecto la red neuronal propuesta fue el resultado de experimentos con distintas capas y número de neuronas.

Acerca de los datos de la prevalencia

- El estudio que se realizó fue sobre las 10 enfermedades con mayor índice de prevalencia, pues los datos provistos por Medicus Mundi en algunas enfermedades importantes como el cólera no son sustanciales. Por lo tanto es importante contar con bastante información sobre una enfermedad preferiblemente de 3 o más años para un mejor estudio.

Acerca de los establecimientos de salud

- Cada establecimiento de salud es diferente de otro aunque brinde los mismos servicios. Para el estudio de prevalencia como el que se realizó se observa que es importante seleccionar un establecimiento adecuado, en lo posible el que tenga la mayor cantidad de información.

Acerca de la asignación de variables como atributos decisión

- La propuesta hecha en la presente tesis de subdividir en conjuntos atributos existentes y generar variables de observación mejora de acuerdo con las variables que se estudien. Es decir datos de variables como el acceso a agua potable, energía eléctrica, educación, datos familiares, u otras más descriptivas de la población profundizarían el estudio de prevalencia.

5.2 Estado de los objetivos

5.2.1 Estado de los objetivos específicos

El objetivo “Utilizar la minería de datos para identificar patrones e información novedosa en bases de datos de centros de salud de Bolivia” se ha cumplido:

- Gracias al aporte de la OMS Medicus Mundi con bases de datos de centros de salud de Bolivia.
- Utilizando ramas de la minería de datos como conjuntos aproximados y redes neuronales para la construcción del modelo. Esta construcción se describe en el capítulo 3.

El objetivo “Utilizar la técnica de conjuntos aproximados para identificar variables relevantes que influyen en la presencia de enfermedades” y “Utilizar la técnica de conjuntos aproximados para identificar variables relevantes que contribuyen al tratamiento de enfermedades” se ha cumplido de forma parcial pues se alcanzó identificar variables relevantes con la construcción de tablas de análisis, insertando variables para su estudio como atributos condición en la sección 3.2.2 y el análisis de reducciones en las pruebas de la sección 4.9, dejando el estudio del tratamiento de las mismas para lo posterior.

El objetivo “Estudiar el comportamiento de enfermedades a través de un modelo de redes neuronales artificiales respecto a variables relevantes identificadas para predicción de prevalencia de enfermedades” se ha cumplido con el desarrollo del modelo de predicción en el capítulo 3.

El objetivo “Realizar un estudio de prevalencia de enfermedades tomando en cuenta la ubicación del centro de salud analizando sus tasas de prevalencia.” se ha cumplido en la sección 4.5 con la asignación de tiempo, lugar y enfermedad a estudiar y su posterior prueba.

El objetivo “Aplicar métodos adecuados de la ingeniería del software para la construcción y prueba de un prototipo software para la simulación del modelo de predicción de prevalencia propuesto” se ha cumplido con el desarrollo de la herramienta software mediante el proceso unificado del software en la sección 4.1.

5.2.2 Estado del objetivo general

“Desarrollar un modelo de predicción de prevalencia de enfermedades para centros de salud de Bolivia utilizando técnicas de minería de datos como conjuntos aproximados y redes neuronales artificiales.”

Se ha alcanzado con el desarrollo del capítulo 3 y 4 de la presente tesis.

5.3 Estado de la hipótesis

“El modelo de predicción obtiene al menos 70% de precisión para la predicción de prevalencia para las diez enfermedades con mayor índice de prevalencia en centros de salud de Bolivia.”

Para 10 enfermedades estudiadas y tomando en cuenta el estadístico presentado en la sección 4.8 sugiere que a un nivel de 95% no se puede negar que los valores esperados con los predichos sean diferentes. Por lo tanto podemos decir que la hipótesis se ha comprobado.

5.4 Recomendaciones

Para posteriores análisis es importante contar con la mayor cantidad de datos posibles en especial si éstos son de relevancia y son muy descriptivos.

Es importante que trabajos futuros en el estudio de prevalencia se utilicen los resultados provistos por la presente tesis de factores que influyen de manera importante pues solamente nos centramos en tres atributos grupo étnico, género y mes.

Cabe recalcar también que es importante para el uso de herramientas software de inteligencia artificial facilita en mucho el trabajo. Se recomienda seleccionar de manera minuciosa el producto software con el que se trabaje y tener muy en cuenta la integración del mismo con otros. En la presente tesis se integró Matlab con Visual Net facilitando el estudio de prevalencia.

Jaws PDF Creator

EVALUATION
VALUTAZIONE
EVALUATION
EVALUACIÓN
EVALUATION

Bibliografía

Libros y tesis:

- Choque, G. (2002). *Inteligencia Artificial*. Bolivia (La Paz).
- Castaño, A., Piattini, M. y Martínez, E. (2002). *Diseño de bases de datos relacionales*. México: Alfaomega.
- Molina, L.C. (1998). *Data mining no processo de extracao de conhecimento de bases de datos*. Tesis de máster. Sao Carlos (Brasil): Instituto de Ciencias de Matemáticas y Computación. Universidad de Sao Pablo.
- Sotomayor, W. (2006). *Minería de uso de la Web mediante el análisis de transacciones de acceso a un servidor Web*. Tesis de grado. La Paz (Bolivia), Universidad Mayor de San Andrés.
- Maqués, E. (2002). *Desarrollo de un modelo de minería de datos académico*. Tesis de grado. La Paz (Bolivia), Universidad Mayor de San Andrés.
- Calderón, C. (2006). *Minería de datos para la predicción de grupos poblacionales*. Tesis de grado. La Paz (Bolivia), Universidad Mayor de San Andrés.
- Rodas, J. (2001). *Un ejercicio de análisis utilizando rough sets en un dominio de educación superior mediante el proceso KDD*. Documento interno. Barcelona: Departamento de Lenguajes y Sistemas Informáticos, Universidad Politécnica de Cataluña.
- Chapman, F., Clinton, J., Khabaza, T., Reinartz, T., Rüdiger, W. (1999). *The CRISP-DM*. Ed. AAI/MIT Press.
- Olson D., Delen D. (2008). *Advanced Data Mining Techniques*. Springer-Verlag Berlin Heidelberg.
- Devore, Jay L. (1998). *Probabilidad y estadística para ingeniería y ciencias*. México: International Thomson Editores.

Artículos electrónicos:

- Pettizi, V. (2004). *La inteligencia de Negocios: Revolución en las Pymes*. Grupo Caissa. Consultado el 30 de mayo de 2008. <http://www.grupocaissa.com.ar/articulo-21.html>.
- Valcárcel, A. (2004). *Data Mining y le descubrimiento del conocimiento*. Consultado el 31 de mayo de 2008. www.unmsm.edu.pe.
- Fernández, S., Pértigas, S., Cañedo, F. (2004). *Medidas de frecuencia de enfermedad*. Consultado el 5 de junio de 2008. http://www.fisterra.com/mbe/medidas_frecuencia/med_frec.asp#top.

Artículos de revistas científicas

- Komorowski, J., Pawlak, Z. (1991). *Rough Sets: A tutorial*. Rough Fuzzy Hybridization: A new trend in decision-making. Springer 3-98.
- Slowinski, J. (1999). *New directions in Rough Sets, Data Mining and Granular Soft Computing. 7th International Workshop (RSFDGRC'99)*, Yamaguchi, Japan, Lecture Notes in Artificial Intelligence 1711.
- Grabowski, A. (2003). *Easi: Properties of Rough Sets and Rough Membership Function*. Journal of Formalized Mathematics 15.
- Fayyad, U., Piatesky-Shapiro, G. And Smyth, P. (1996). *From Data Mining to Knowledge Discovery in Databases*. American Association for Artificial Intelligence. 0738-4620. Consultado el 30 de mayo de 2008.
- Moore, G. E. (1965). *Cramming more components onto integrated circuits*. Electronic Magazine. Consultado el 30 de mayo de 2008.
- Molina, L. C., Riveiro, S. (2001). *Descubriendo Conocimiento para el Mejoramiento Gerético Binario usando Técnicas de Data Mining*. Consultado el 30 de mayo de 2008.
- Arco, L., Bello, R. (2006). *On clustering validity measures and the Rough Set Theory*. 7th Mexican International Conference on Artificial Intelligence, IEEE Computer Society Press. Consultado el 1 de mayo de 2008.
- Pawlak, Z. (1982). *Rough Sets*. International journal of Computer and Information Sciences. 11: 341-356. Consultado el 20 de mayo de 2008.

- Skowron. A. (1999). *New directions in Rough Sets, Data Mining, and Granular Soft Computing*. 7th International Workshop (RSFDGRC'99), Yamaguchi (Japón).
- McCulloch, W. S. y Pitts, W. H. (1943). *A logical calculus of the ideas immanent in nervous activity*. *Bulltein of Mathematical Biophysics*.
- Borroacci, R., Rubio, M. (2003). *Aplicabilidad de redes neuronales artificiales para la predicción de los resultados individuales de la cirugía cardíaca*. La Pampa. Consultado el 9 de junio de 2008.
- Santana, J. C. (2006). *Predicción de series temporales con redes neuronales: una aplicación a la inflación colombiana*. *Revista Colombiana de Estadística*. Vol 29 N°1. pp. 77 a 92. Consultado el 5 de junio de 2008.
- Gómez, J. A., Jeréz, C. M., Muñoz, J., Alba E. (2000). *Un modelo para la predicción de recidiva de pacientes operados de cáncer de mama (MII) basado en redes neuronales*. *Inteligencia Artificial*. *Revista Iberoamericana de Inteligencia Artificial*. N° 11, pp 39-45. Consultado el 5 de junio de 2008.
- Molina, L. J. (2000). *Torturarlos hasta que Confiesen*. Departamento de Lenguajes y Sistemas Informáticos, Universidad Politécnica de Cataluña. Barcelona, España.

Fuentes electrónicas

- Hernández, J. (2002). *La disciplina de los Sistemas de Bases de datos*. Historia, Situación Actual y Perspectivas. Obtenida el 30 de mayo de 2008. Página Web de la Universidad Politécnica de Valencia: www.dsic.upv.es.
- Vallejos, S. J. (2006). *Mineía de datos*. Obtenida el 22 de mayo de 2008. Página Web de la Universidad Nacional del Norte-Argentina: unne.edu.ar
- Bunge, M. (1997). *La ciencia, su método, su filosofía*. Obtenida el 1 de junio de 2008. www.dcc.uchile.cl

Lista de enlaces electrónicos:

- <http://www.kdnuggets.com>
- <http://www.medicusmundi.info>

Jaws PDF Creator

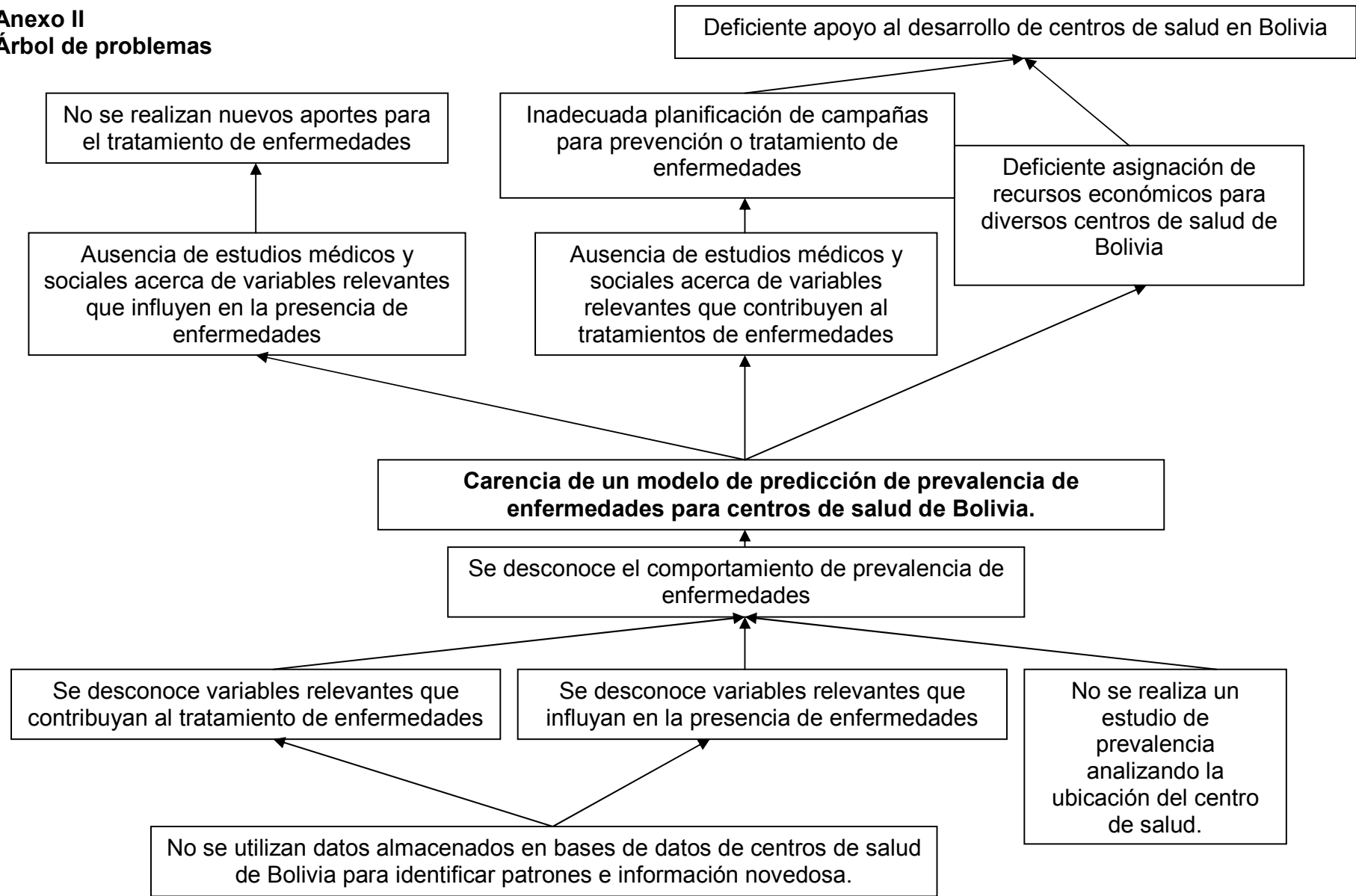
EVALUATION
VALUTAZIONE
EVALUATION
EVALUACIÓN
EVALUATION

ANEXO I

Categorías de la clasificación Internacional de Enfermedades CIE

Capítulo	Códigos	Título
I	A00-B99	Ciertas enfermedades infecciosas y parasitarias
II	C00-D48	Neoplasias
III	D50-D89	Enfermedades de la sangre y de los órganos hematopoyéticos y otros trastornos que afectan el mecanismo de la inmunidad
IV	E00-E90	Enfermedades endocrinas, nutricionales y metabólicas
V	F00-F99	Trastornos mentales y del comportamiento
VI	G00-G99	Enfermedades del sistema nervioso
VII	H00-H59	Enfermedades del ojo y sus anejos
VIII	H60-H95	Enfermedades del oído y de la apófisis mastoides
IX	I00-I99	Enfermedades del sistema circulatorio
X	J00-J99	Enfermedades del sistema respiratorio
XI	K00-K93	Enfermedades del aparato digestivo
XII	L00-L99	Enfermedades de la piel y el tejido subcutáneo
XIII	M00-M99	Enfermedades del sistema osteomuscular y del tejido conectivo
XIV	N00-N99	Enfermedades del aparato genitourinario
XV	O00-O99	Embarazo, parto y puerperio
XVI	P00-P96	Ciertas afecciones originadas en el periodo perinatal
XVII	Q00-Q99	Malformaciones congénitas, deformidades y anomalías cromosómicas
XVIII	R00-R99	Síntomas, signos y hallazgos anormales clínicos y de laboratorio, no clasificados en otra parte
XIX	S00-T98	Traumatismos, envenenamientos y algunas otras consecuencias de causa externa
XX	V01-Y98	Causas extremas de morbilidad y de mortalidad
XXI	Z00-Z99	Factores que influyen en el estado de salud y contacto con los servicios de salud
XXII	U00-U99	Códigos para situaciones especiales

Anexo II
Árbol de problemas



Árbol de Objetivos

