



UNIVERSIDAD CÉSAR VALLEJO

FACULTAD DE INGENIERÍA

ESCUELA ACADÉMICO PROFESIONAL DE INGENIERÍA DE
SISTEMAS

Uso de la técnicas de Minería de Datos en la herramienta Weka para la
predicción más adecuada de la enfermedad Diabetes Mellitus en la clínica
San Juan Bautista

TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:
INGENIERO DE SISTEMAS

AUTOR:

SALAZAR ZEA FRANKLIN EDUARDO

ASESOR:

DR. ESCOBEDO BAILÓN FRANK

LÍNEA DE INVESTIGACIÓN:

SISTEMA DE INFORMACIÓN ESTRATÉGICO Y DE TOMA DE DECISIONES

LIMA – PERÚ

2014

Página del Jurado

DEDICATORIA

Le dedico en primer lugar mi trabajo a Dios, a mis Padres, a quien le debo toda mi vida, les agradezco el cariño y su comprensión, a ustedes quienes han sabido formarme con buenos sentimientos, hábitos y valores, lo cual me ha ayudado a salir adelante buscando siempre el mejor camino.

A mis docentes, gracias por su tiempo, por su apoyo así como por la sabiduría que me transmitieron en el desarrollo de mi formación profesional, por haber guiado el desarrollo de este trabajo y llegar a la culminación del mismo.

AGRADECIMIENTO

Primero a Dios, por estar conmigo en cada paso que doy, por fortalecer mi corazón e iluminar mi mente y por haber puesto en mi camino a aquellas personas que han sido mi soporte y compañía durante todo el periodo de estudio.

Agradecer hoy y siempre a mis padres por el esfuerzo realizado por ellos, el apoyo, la alegría y la fortaleza necesaria que me dan para seguir adelante. Al Dr. Frank Escobedo Bailón y Dr. Alfredo Daza Vergaray por tener la paciencia de responder mis dudas y brindarme sus conocimientos.

DECLARACIÓN DE AUTENTICIDAD

Yo, SALAZAR ZEA FRANKLIN EDUARDO con DNI N° 47138784, a efecto de cumplir con las disposiciones vigentes consideradas en el Reglamento de Grados y Títulos de la Universidad César Vallejo, Facultad de Ingeniería, Escuela de Ingeniería de Sistemas, declaro bajo juramento que toda la documentación que acompaño es veraz y auténtica.

Así mismo, declaro también bajo juramento que todos los datos e información que se presenta en la presente tesis son auténticos y veraces.

En tal sentido asumo la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de información aportada por lo cual me someto a lo dispuesto en las normas académicas de la Universidad César Vallejo.

Lima, 28 de Noviembre de 2014.

Franklin Eduardo Salazar Zea

Presentación:

Señores miembros del Jurado, presento ante ustedes la Tesis titulada “Uso de las técnicas de minería de datos en la herramienta weka para la predicción más adecuada de la enfermedad Diabetes Mellitus en la clínica San Juan Bautista”, con la finalidad de determinar cuál es la técnica de minería de datos para la predicción más adecuada de la enfermedad Diabetes Mellitus en la clínica San Juan Bautista, en cumplimiento del Reglamento de Grados y Títulos de la Universidad César Vallejo para obtener el Título Profesional de Ingeniero de Sistemas.

Esperando cumplir con los requisitos de aprobación.

SALAZAR ZEA Franklin Eduardo.

ÍNDICE

PÁGINA DEL JURADO	ii
DEDICATORIA.....	iii
AGRADECIMIENTO.....	vi
DECLARACIÓN DE AUTENTICIDAD	v
PRESENTACIÓN.....	vi
ÍNDICE.....	vii
ÍNDICE DE FIGURAS.....	viii
ÍNDICE DE GRÁFICOS	ix
ÍNDICE DE TABLAS.....	x
RESUMEN	xi
ABSTRACT	xii
I. INTRODUCCIÓN	1
1.1. Problemas.....	35
1.2. Objetivos.....	35
II. MARCO METODOLÓGICO	36
2.1. Hipótesis.....	36
2.2. Variables.....	36
2.3. Operacionalización de variables.....	37
2.4. Metodología	40
2.5. Tipos de estudio	41
2.6. Diseño.....	41
2.7. Población, muestra y muestreo.....	68
2.8. Técnicas e instrumentos de recolección de datos	69
2.9. Métodos de análisis de datos	69
III. RESULTADOS.....	70
IV. DISCUSIÓN	79
V. CONCLUSIONES	80
VI. RECOMENDACIONES.....	81
VII. REFERENCIAS BIBLIOGRÁFICAS	82
ANEXOS.....	85

ÍNDICE DE FIGURAS

Figura 01: Esquema de una investigación experimental	40
Figura 02: Diagrama del diseño pre experimental	41
Figura 03: Archivo .ARFF.....	44
Figura 04: Estructura del archivo .ARFF	45
Figura 05: Iniciar Weka	46
Figura 06: Adjuntar la BD en Weka	46
Figura 07: Datos cargados en Weka	47
Figura 08: Pre-procesado de datos en Weka	48
Figura 09: Aplicación del Algoritmo J48	51
Figura 10: Datos de ejecución	52
Figura 11: Modelo clasificador (conjunto de entrenamiento completo)	52
Figura 12: Evaluación de los datos de entrenamiento	53
Figura 13: Precisión detallada por clase	53
Figura 14: Árbol de Decisión J48	55
Figura 15: Aplicación de la red Neuronal Perceptrón Multicapa	56
Figura 16: Datos de ejecución con Perceptrón Multicapa	56
Figura 17: Modelo clasificador con Perceptrón Multicapa	57
Figura 18: Evaluación de los datos de entrenamiento con Perceptrón Multicapa	60
Figura 19: Precisión detallada por clase con Perceptrón Multicapa	60
Figura 20: Red Neuronal Perceptrón Multicapa	61
Figura 21: Interfaz del Clúster	62
Figura 22: Resultado del agrupamiento con el método kMeans	63
Figura 23: Modelo y evaluación del agrupamiento	63
Figura 24: Interfaz de la Asociación	64
Figura 25: Resultados de la Asociación mediante el algoritmo A priori	65
Figura 26: Visualización de los atributos	66
Figura 27: Interpretación de un coeficiente de confiabilidad	71

ÍNDICE DE GRÁFICOS

Gráfico 01: Discretización de datos en Weka	48
Gráfico 02: Datos estadísticos de cada atributos	50
Gráfico 03: Factores Influyentes	75
Gráfico 04: Factores más influyentes (%)	75
Gráfico 05: Rendimiento de las Red Neuronal	76
Gráfico 06: Rendimiento del Árbol de Decisión	77
Gráfico 07: Comparación del Rendimiento AD y RN	78

ÍNDICE DE TABLAS

Tabla 01: Base de Datos Pima Indians Diabetes	43
Tabla 02: Datos estadísticos de cada variable	70
Tabla 03: Matriz de Correlaciones	72
Tabla 04: Suma de Correlaciones	74

RESUMEN

En los últimos años la Diabetes Mellitus se ha convertido en un problema global como consecuencia del mal hábito de vida, la obesidad y el envejecimiento.

La Minería de Datos actualmente se aplica en diferentes áreas entre ellos la Medicina teniendo como objetivo el diagnóstico médico. En el presente trabajo se aplican las Técnicas de Minería de Datos con el objetivo de poder determinar con una mejor precisión el resultado del diagnóstico.

La presente investigación tiene como finalidad que los centros hospitalarios y los médicos pudieran contar con el apoyo de una herramienta y una técnica adecuada que les permita analizar los datos de cada uno de sus pacientes para poder determinar el diagnóstico más acertado así como el tratamiento óptimo a seguir, lo cual representaría un soporte y ayuda para el médico. Entre las técnicas más utilizadas se encuentran las Redes Neuronales y los Árboles de Decisión.

Dentro de las técnicas de Minería de Datos se usaran las Redes Neuronales y los Árboles de Decisión ya que son las más usadas en el campo de la Medicina y están orientadas a la predicción.

Palabras Clave: Minería de Datos, Redes Neuronales, Árboles de Decisión, Diabetes Mellitus.

ABSTRACT

In recent years, Diabetes Mellitus has become a global problem due to the bad habit of life, obesity and aging.

Data Mining is currently applied in different areas including Medicine aiming medical diagnosis. In this paper, the data mining techniques are applied in order to be able to determine with greater accuracy the outcome of diagnosis.

This research aims to hospital and a tool and a suitable technique that allows them to analyze the data for each of their patients in order to determine the most accurate diagnosis and optimal treatment follow, which would represent a support and assist clinicians, could support medical centers. Among the techniques used are Neural Networks and Decision Trees.

Within the Data Mining Techniques Neural Networks and Decision Trees were used because they are the most used in the medical field and are aimed at the prediction.

Keywords: *Data Mining, Neural Networks, Decision Trees, Diabetes Mellitus.*