



UNIVERSIDAD CÉSAR VALLEJO

FACULTAD DE INGENIERÍA

ESCUELA ACADÉMICO PROFESIONAL DE INGENIERÍA DE SISTEMAS

Uso de la herramienta de minería de datos Weka para el análisis de accidentes de tránsito en Lima Metropolitana. Período: Enero de 2010 a Diciembre de 2013.

**TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:
INGENIERO DE SISTEMAS**

AUTOR:

GONZALES SALDAÑA SERGIO

ASESOR:

DR. ESCOBEDO BAILON FRANK

LÍNEA DE INVESTIGACIÓN:

SISTEMA DE INFORMACIÓN ESTRATEGICO Y DE TOMA DE DECISIONES

LIMA – PERÚ

2014

PÁGINA DEL JURADO.

DEDICATORIA

Este proyecto de investigación va dedicado a la persona más importante en mi vida, mi madre, que es el motor y la fuerza para poder seguir adelante cumpliendo mis objetivos trazados.

AGREDECIMIENTO

A Dios que siempre me acompaña en mi camino.

A mis asesores Dr. Frank Escobedo, Ing. Nancy Bernardo y Mg. Alfredo Daza, por incentivar me con el desarrollo de mi trabajo de investigación y por el tiempo que nos dedican en el proceso de nuestro aprendizaje y desarrollo profesional.

A mi familia que son la inspiración y el motivo para poder culminar mis metas trazadas.

DECLARATORIA DE AUTENTICIDAD

Yo, GONZALES SALDAÑA, SERGIO con DNI N° 46392128, 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 Académico Profesional 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, julio del 2014

Sergio Gonzales Saldaña

PRESENTACIÓN

Señores miembros del jurado, presento a ustedes la Tesis titulada “Uso de la herramienta de minería de datos Weka para el análisis de accidentes de tránsito en Lima Metropolitana. Período: Enero de 2010 a Diciembre de 2013.”, con la finalidad de Usar la herramienta de minería de datos Weka, para analizar los accidentes de tránsito y plantear alternativas de solución con el fin de poder reducir los accidentes de tránsito en los distritos de Lima Metropolitana, 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.

Sergio Gonzales Saldaña

ÍNDICE

PÁGINA DEL JURADO.....	ii
DEDICATORIA.....	v
AGREDECIMIENTO.....	vi
DECLARATORIA DE AUTENTICIDAD.....	vii
PRESENTACIÓN.....	viii
ÍNDICE.....	ix
ÍNDICE DE TABLAS.....	x
ÍNDICE DE GRÁFICOS.....	xi
ÍNDICE DE FIGURAS.....	xii
RESUMEN.....	xiii
ABSTRACT.....	xiv
I. INTRODUCCIÓN.....	13
1.1. Formulación del problema y los problemas específicos.....	32
1.2. Objetivo general y objetivos específicos.....	32
II. MARCO METODOLÓGICO.....	33
2.1. Hipótesis general e hipótesis específicas.....	33
2.2. Variables.....	33
2.3. Operacionalización de variables.....	34
2.4. Metodología.....	36
2.5. Tipos de estudio.....	36
2.6. Diseño.....	37
2.7. Población, muestra y muestreo.....	37
2.8. Técnicas e instrumentos de recolección de datos.....	38
2.9. Métodos de análisis de datos.....	38
III. RESULTADOS.....	39
3.1. Análisis Alfa de Cronbach Spss Statistics.....	50
3.2. Análisis Con La Herramienta De Minería De Datos WEKA.....	54
IV. DISCUSIÓN.....	77
V. CONCLUSIONES.....	78
VI. RECOMENDACIONES.....	79
VII. REFERENCIAS BIBLIOGRÁFICAS.....	80
VIII. ANEXOS.....	82

ÍNDICE DE TABLAS

Tabla N° 1: Cantidad de Accidentes de Tránsito: Años 2010 - 2013 en Lima Metropolitana.....	40
Tabla N° 2: Prueba T: Cantidad de Accidentes 2010-2011.....	41
Tabla N° 3: Prueba T: Cantidad de Accidentes 2011-2012.....	41
Tabla N° 2: Prueba T: Cantidad de Accidentes 2012-2013.....	41
Tabla N° 5: Resultado de la H1.....	42
Tabla N° 6: Factores de Accidentes de tránsito.....	43
Tabla N° 7: Prueba T: Influencia de los factores 2010-2011.....	44
Tabla N° 8: Prueba T: Influencia de los factores 2011-2012.....	44
Tabla N° 9: Prueba T: Influencia de los factores 2012-2013.....	44
Tabla N° 10: Resultados H2.....	45
Tabla N° 11: Accidentes, Muertes y Lesionados 2010 – 2013.....	46
Tabla N° 12: Prueba T: Cantidad de muertes 2010-2011.....	48
Tabla N° 13: Prueba T: Cantidad de muertes 2011-2012.....	48
Tabla N° 14: Prueba T: Cantidad de muertes 2012-2013.....	48
Tabla N° 15: Resultados H3.....	49
Tabla N° 16: Matriz de Correlaciones.....	51
Tabla N° 17: Suma de Correlaciones.....	53
Tabla N° 18: Cantidades de accidentes, muertes y lesiones en los últimos años.....	54

ÍNDICE DE GRÁFICOS

Gráfico N° 1: Resultado H1.....	42
Gráfico N° 2: Resultados H2.....	45
Gráfico N° 3: Cantidad de accidentes por año.....	47
Gráfico N° 4: Cantidad de accidentes, muertes y lesiones por año.....	47
Gráfico N° 5: Resultados H3.....	49
Gráfico N° 6: Cantidades de accidentes, muertes y lesiones en WEKA.....	59
Gráfico N° 7: Asignación de los Cluster.....	65
Gráfico N° 8: Matriz grafica en WEKA.....	68
Gráfico N° 9: Desglose por gráfica.....	69
Gráfico N° 10: Matriz Gráfica en WEKA.....	76

ÍNDICE DE FIGURAS

Figura N° 1: Esquema de experimento y variable.....	36
Figura N° 2: Interpretación de un coeficiente de confiabilidad.....	50
Figura N° 3: Interfaz de inicio WEKA.....	55
Figura N° 4: Interfaz Explorer de WEKA.....	55
Figura N° 5: Carga de datos a WEKA.....	56
Figura N° 6: Mostrando los datos filtrados en WEKA.....	57
Figura N° 7: Detallando los atributos.....	58
Figura N° 8: Interfaz de clasificación en WEKA.....	60
Figura N° 9: Análisis de resultados en WEKA.....	61
Figura N° 10: Análisis de resultados en WEKA.....	62
Figura N° 11: Interface Cluster en WEKA.....	63
Figura N° 12: Resultado del Cluster en WEKA.....	64
Figura N° 13: Interfaz Associate en WEKA.....	66
Figura N° 14: Resultados del Algoritmo A priori.....	67
Figura N° 15: Interfaz Classify de WEKA.....	70
Figura N° 16: Aplicando Algoritmo J48.....	71
Figura N° 17: Árbol de decisión J48.....	71
Figura N° 18: Matriz de Confusión.....	72
Figura N° 19: Árbol de Decisión.....	73
Figura N° 20: Aplicando Algoritmo KMeans.....	74
Figura N° 21: Algoritmo Apriori de WEKA.....	75

RESUMEN

El objetivo de esta investigación es usar la minería de datos para analizar los accidentes de tránsito y plantear un plan de prevención y alternativas de solución para poder reducir los accidentes de tránsito en los distritos de Lima Metropolitana.

Este estudio se basará en la aplicación de la técnica de minería de datos (árboles de decisión, clasificación), orientados a la identificación de los factores que más influyen en la definición de un accidente de tránsito. Por otro lado, se planteó el análisis del problema a través de la información recolectada, lo que nos brindará un mayor conocimiento sobre los accidentes de tránsito, el análisis se realizó con la herramienta de minería de datos WEKA, esta misma nos ayudó a poder obtener información que se desconocía sobre los accidentes de tránsito, con lo cual se pueda plantear alternativas de solución a este creciente problema. Para un correcto desarrollo del proyecto se trabajó con la metodología de investigación CRISP – DM, que es la más recomendable para este tipo de proyectos, también se hizo un estudio estadístico con la herramienta de cálculo MS-Excel y SPSS Statistics 22.

Se muestra también los gráficos estadísticos que respaldan las hipótesis planteadas en el trabajo de investigación y resaltan los resultados estadísticos que se realizaron.

Palabras Clave: Accidentes de tránsito, Minería de datos, CRISP – DM, Arboles de decisión.

ABSTRACT

The objective of this research is to use data mining to analyze traffic accidents and propose a plan of prevention and alternative solutions to reduce traffic accidents in the districts of Lima.

This study is based on the application of data mining technique (decision trees, classification), aimed at identifying the factors that influence the definition of a traffic accident. On the other hand the analysis of the problem through the information collected was raised, which will give us a greater understanding of traffic accidents, the analysis was performed with the tool WEKA data mining, this same helped us to obtain information that was unknown about traffic accidents, which can suggest alternative solutions to this growing problem. For proper development of the project worked with the research methodology CRISP - DM, which is the most suitable for this type of project, a statistical study was also made with the tool of MS-Excel and Spss Statistics 22 calculation.

Statistical graphics that support the hypotheses in the research and highlight the statistical results were performed is also shown.

Keywords: Traffic accidents, Data Mining, CRISP - DM, Decision trees.