

Predicción del riesgo de un accidente de tránsito en Colombia por medio del software Weka.

Omar Danilo Castrillón, Ph. D¹, Jaime Alberto Giraldo, Ph. D¹, and Santiago Ruiz Herrera, Ph. D¹

¹Universidad Nacional de Colombia - Sede Manizales - Facultad de Ingeniería y Arquitectura – Departamento de Ingeniería Industrial – Grupo de Innovación y Desarrollo Tecnológico - Bloque Q Campus La Nubia, Manizales, 170001 – Colombia. odcastrellong@unal.edu.co, jaiagiraldo@unal.edu.co, sruizhe@unal.edu.co

Resumen— En esta investigación se analizan las bases de datos accidentes de tránsito de Colombia, mediante el programa de inteligencia artificial denominado Weka. Este análisis se realiza sobre 170.854 registros, por medio del algoritmo de clasificación J48, una herramienta de minería de datos del programa Weka, la cual permite construir el respectivo árbol de decisión para los registros analizados. El sistema es entrenado con la mitad de los registros y validado con todos los registros. Como resultado de este proceso se identifican los principales factores influyentes en los accidentes de tránsito, los cuales permiten predecir el riesgo (daños, heridos, muertos) de un accidente de tránsito con una efectividad aproximada del 80%. Como futuras líneas se emplearán otras técnicas de inteligencia artificial con el fin de mejorar el porcentaje de clasificación del sistema.

Palabras Clave—Weka, Inteligencia Artificial, Minería de datos, árboles de decisión, algoritmo J48.

I. INTRODUCCIÓN

En todo el mundo los daños, heridos y muertes ocasionados por los accidentes de tránsito siguen siendo un gran problema, por ende, todos los esfuerzos que se realicen en este sentido siempre serán bien recibidos [1], más aún, si se considera que los accidentes de tránsito pueden suponer un alto riesgo para la vida de las personas [2], llegándose incluso a afirmar que para el 2030 las lesiones producidas por los accidentes de tránsito serán la quinta causa de muerte a nivel mundial [3].

Actualmente los accidentes de tránsito en Colombia cobran un gran número de vidas y heridos cada año, dejando innumerables daños materiales. En este trabajo se analizan todos los registros de los accidentes de tránsito ocurridos en Colombia durante el año 2010 [4] (última base de datos disponible). Considerando aspectos como: Fecha, área, afectación, hora, factores, riesgos: daños, heridos o muertos.

El objetivo fundamental de este artículo es predecir a partir de las variables fecha, área, afectación, hora y factores, el riesgo de un accidente de tránsito (daño, heridos o muertos). Lo anterior permitirá establecer y enfocar programas preventivos viales, los cuales minimicen el impacto de un accidente de tránsito, salvando un gran número de vidas.

En la actualidad, existen diferentes clases de software basados en Técnicas de Inteligencia artificial para el análisis de información y predicción de resultados. En este trabajo

investigativo se emplea el programa de inteligencia artificial denominado Weka [5], el cual facilita el análisis de los datos y construcción de árboles de decisión, mediante el algoritmo de minería de datos J48. Este software es una implementación (open source - Java) del algoritmo C4.5 desarrollado por John Ross Quinlan [6].

Los árboles de decisión generados por este programa establecen una predicción del riesgo de un accidente de tránsito aproximadamente del 80%. Se espera que este trabajo sirva para predecir y reducir el índice de accidentes de tránsito en Colombia. Como futuras líneas de investigación el problema será analizado por otros sistemas de clasificación [7,8], con el fin de lograr una mayor efectividad en el porcentaje de aciertos.

El desarrollo de esta investigación, se estructura en las siguientes secciones: a) Materiales y Métodos. b) Resultados. c) discusiones. d) conclusiones e) bibliografía. Cada una de estas secciones son ilustradas en los apartados siguientes.

II. MATERIALES Y MÉTODOS

Paso 1. Diseño de la base de datos. Con el fin de garantizar que el árbol de decisión generado por el programa Weka, sea lo más conciso y fácil de interpretar, los registros de la base de datos de los accidentes de tránsito de Colombia en el año 2010 [4], deben ser analizados. Este análisis permitirá determinar la estructura de esta base de datos, con el fin de que la misma pueda ser diseñada en Weka.

Paso 2. Diseño del archivo Arff Weka. El resultado del paso anterior, permitirá estructura un archivo Arff el cual pueda ser interpretado por el programa Weka. Este archivo deberá ser conformado por los 170.854 registros disponibles de accidentes de tránsito.

Paso 3. Generación del árbol de decisión. El archivo Arff generado en el paso anterior, deberá ser cargado en el software Weka. Este software permitirá una interpretación de este archivo y estructuración del respectivo árbol de decisión mediante el algoritmo de clasificación J48. Este algoritmo se configura para que el sistema se entrenado con la mitad los datos y validado la totalidad de los mismos.

Paso 4. Efectividad en la predicción del riesgo. El árbol de decisión anterior, permitirá establecer la efectividad del sistema y determinar el porcentaje de clasificación del mismo en la predicción del comportamiento de la variable denominada riesgo, la cual tendrá tres estados: daños, heridos y muertos.

Paso 5. Optimización de la efectividad. Finalmente se realiza un análisis del árbol de decisión establecido por el programa Weka, con el fin de establecer las causas de fallo y mejorar la efectividad en la predicción del comportamiento de la variable denominada riesgo. Igualmente, si el árbol no es legible, los valores de cada uno de los campos del archivo Arff deberán ser analizados, agrupados y/o suprimidos, con el fin de reestructurar el árbol y buscar una mayor legibilidad.

III. RESULTADOS

Paso 1. Diseño de la base de datos. La base de datos inicial contiene 170.854 registros con los siguientes campos: Fecha, departamento, municipio, área, afectación, hora observación y riesgo. Con la siguiente información relevante en cada uno de estos campos:

TABLE I
ESTRUCTURA INICIAL DE LA BASE DE DATOS

Campo	Contenido
Fecha	1/01/2010 – 31/12/2010
Departamento	31 Departamentos de Colombia
Municipio	513 Municipios diferentes
Área	Urbana o Rural
Afectación	Atropello, Caída ocupante, choque, volcamiento, incendio, otro.
Hora	00:00 – 23:59:59
Observación	95 clases diferentes de observaciones.
Riesgos	Con heridos, solo daños, con muertos

Paso 2. Diseño del archivo Arff Weka. Como resultado de este paso se obtuvo el siguiente archivo Arff para ser interpretado por medio del programa Weka:

```
@RELATION relation
@ATTRIBUTE Fecha date yyyy-MM-dd
@ATTRIBUTE Depto {31 Departamento de Colombia}
@ATTRIBUTE Mpio {513 Municipios de Colombia}
@ATTRIBUTE Area {URBANO,RURAL}
@ATTRIBUTE Afectacion {6 afectaciones diferentes}
@ATTRIBUTE hora date HH:mm
@ATTRIBUTE observación {95 diferentes observaciones}
@ATTRIBUTE riesgo {CONHERIDOS, SOLODANOS, CONMUERTOS}
```

@DATA

2010-06-12, NORTEDESANTANDER, ABREGO,RURAL, ATROPELLO, 18:00, FALTADEPRECAUCIONPORNIEBLA, SOLODANOS -> Ejemplo registro

2010-07-18, CUNDINAMARCA, NODEFINE, RURAL, VOLCAMIENTO, 10:35, HUECOS, CONMUERTOS. -> Ejemplo registro.

Total 170.854 registros

Paso 3-4-5. Generación del árbol de decisión - Efectividad en la predicción del riesgo - Optimización de la efectividad. Como resultado de estos pasos se obtuvo un porcentaje de clasificación correcta del 80.0613% y de clasificación incorrecta del 19.9387% Como se ilustra en el siguiente resumen generado por el programa Weka:

TABLE II
RESUMEN DE LA CLASIFICACIÓN GENERADO POR EL PROGRAMA WEKA

Correctly Classified Instances	136.788	80.0613%
Incorrectly Classified Instances	34.066	19.9387%
Total Number of Instances	170.854	

El árbol de clasificación, generado por el programa Weka, presentó un total de 7.098 hojas y 7.659 niveles, lo cual lo hace muy difícil de interpretar. Por ende, se suprimieron los campos municipio y departamento con el fin lograr un árbol más reducido el cual facilitará su interpretación. Como resultado de este proceso se obtuvo un árbol de clasificación de 296 hojas y 401 niveles.

No obstante, el haberse logrado una reducción importante en el número de hojas y niveles, el árbol generado aun presentaba un tamaño considerable para ser analizado e interpretado correctamente. Con el fin de reducir aún más el árbol y hacer más fácil su interpretación, los 95 valores diferentes de la categoría observación fueron agrupados en 13 categorías como se ilustra en la siguiente tabla:

TABLE III
AGRUPAMIENTO DE LA CATEGORÍA OBSERVACIÓN

ANTERIOR	NUEVO
ADELANTARCERRANDO	ADL
ADELANTARENCURVAOPENDIENTES	ADL
ADELANTARENZONAPROHIBIDA	ADL
ADELANTARINVADIENDOCARRILDESENTIDOCON TRARIO	ADL
ADELANTARINVADIENDOCARRILDELMISMOSENTIDOEN ZIG-ZAG	ADL
ADELANTARPORLADERECHA	ADL
APROVISIONAMIENTOINDEBIDO	IMP
ARRANCARSINPRECAUCION	IMP
AUSENCIAODEFICIENCIADEESPEJOSRETROVISOR ES	FME
AUSENCIAODEFICIENCIAENDEMARCAACION	VIA
AUSENCIATOTALOPARCIALDESENALES	SEN
CAMBIODECARRIL SININDICACIONEINADECUADO	IMP
CARGASOBRESALIENTESINAUTORIZACION	IMP
CARGASOBRESALIENTESINSENALES	IMP
CARGUEODESCARGUEENHORASOSITIOSPROHIBIDOS	IMP
CIRC.XCALZ.OCARRILESDESTINADOSABUSES YBU	IMP

ANTERIOR	NUEVO
SETAS	
CONDUCCIONVEHICULOSINADAPTACIONES	IMP
CRUZARENCURVA	CRU
CRUZARENDIAGONAL	CRU
CRUZARENESTADODEEMBRIAGUEZ	ALC
CRUZARSINOBSERVAR	CRU
DEFECTOSFISICOSYPSIQUICOS	IMP
DEJAROMOVILIZARSEMOVIENTEENLAVIA	VIA
DEJARORECOGERPASAJEROSENSITIVOSNODEMARCADOS	SEN
DEJAROBSTACULOENLAVIA	IMP
DESCENDEROSUBIRDELVEHICULOENMARCHA	IMP
DESOBEDECERALAGENTEDETRANSITO	SEN
DESOBEDECERSENALESDETRANSITO	SEN
EMBRIAGUEZAPARENTE	ALC
EMBRIAGUEZODROGA	ALC
EXCESODEPESEO	OTR
EXCESODEVELOCIDAD	SEN
EXCESOENHORASDECONDUCCION	IMP
EXPLOSIVOS	OTR
FALLASENELEXOSTO.GASESENELINTERIOREDELVEHICULO	FME
FALLASENELLIMPIABRISAS	FME
FALLASENELSISTEMAELECTRICO	FSE
FALLASENLADIRECCION	FME
FALLASENLATAPADELMOTOR	FME
FALLASENLASLLANTAS	FME
FALLASENLASPUERTAS	FME
FALLASENLOSRENOS	FME
FALLASENLUCESDIRECCIONALES	FSE
FALLASENLUCESPOSTERIORES	FSE
FALLASENLUCESDELANTERAS	FSE
FALLASENLUCESDEFRENOS	FSE
FALLASENPITO	FSE
FALTADEPRECAUCIONPORNIEBLA	IMP
FALTADESENALESENVEHICULOVARADO	SEN
FALTAMANTENIMIENTO MECANICO	FME
FRENARBRUSCAMENTE	IMP
HUECOS	VIA
IMPARTIRENSENANZAAUTOMOVILISTICASINAUTORIZACION	IMP
INCENDIOPORREPARACIONINDEBIDA	IMP
JUGARENLAVIA	IMP
NINOSENASIENTODELANTERO	IMP
NOCONDUCCIONAHORCAJADAS	IMP
NOHACERUSODESE?ALESREFLECTIVAS	SEN
NOMANTENERDISTANCIADESEGURIDAD	IMP
NOPORTARESPECIOS	FME
NOREPORTO	NRP
NORESPETARPRELACION	IMP
NOSUJETARLOSMANUBRIOS	IMP
OBSTACULOENLAVIA	VIA
OTRA	OTR
PARARSESOBRELA CALZADA	IMP
PASAJEROEMBRIAGADO	LIC
PASAJEROSOBSTRUYENDOALCONDUCTOROSOBRECUPO	IMP
PASARELSEMAFOROENROJO	SEN
PONERENMARCHAUNVEHICULOSINPRECAUCIONES	IMP
REALIZARGIROENU	SEN
REMOLQUESINPRECAUCION	IMP
REPARARVEHICULOENVIAPUBLICA	FME

ANTERIOR	NUEVO
REVERSOIMPRUDENTE	IMP
SALIRPORDELANTEDEUNVEHICULO	IMP
SEMAFOROENROJO	SEN
SUBIRSEALANDENOVIASPEATONALESOCICLORUTAS	IMP
SUJETARSEAOTROVEHICULO	IMP
SUPERFICIEHUMEDA	VIA
SUPERFICIELISA	VIA
TRANSITARCONLASPUERTASABIERTAS	MTR
TRANSITARDISTANTEDELA CERAUORILLADELACALZADA	MTR
TRANSITARENTRAVEHICULOS	MTR
TRANSITARPORLACALZADA	MTR
TRANSITARPORLADERECHAENVIASRURALES	MTR
TRANSITARPORVIASPROHIBIDAS	MTR
TRANSITARSINLOS DISPOSITIVOS LUMINOSOS DE DETENCION	MTR
TRANSITARUNOALLADODELOTRO	MTR
TRANSPORTAROTRAPERSONAO COSAS	SCU
TRANSPORTARPASAJEROS ENVEHICULOS DE CARGA	SCU
TRANSPORTEDECARGASINSEGURIDAD	SCU
VEHICULO MALESTACIONADO	IMP
VIAJARALAZQUIERDADELCONDUCTOR	IMP
VIAJARCOLGADOENLOSESTRIBOS	IMP
VIDRIOSENMALESTADO	FME

Las anteriores, modificaciones permitieron estructurar el siguiente archivo Arff para ser interpretado por el programa Weka:

@RELATION relation
 @ATTRIBUTE Fecha date yyyy-MM-dd
 @ATTRIBUTE Area {URBANO, RURAL}
 @ATTRIBUTE Afectacion {ATR, COC, CHO, OTR, VLC, INC}
 @ATTRIBUTE hora date HH:mm
 @ATTRIBUTE observacion {ADL, IMP, FME, VIA, SEN, CRU, ALC, OTR, FSE, NRP, LIC, MTR, SCU}
 @ATTRIBUTE riesgo {HERIDO, DANOS, MUERE}

@DATA
 2010-06-12, RURAL, ATR, 18:00, IMP, DANOS ->
 Ejemplo registro

2010-11-27, RURAL, ATR, 5:30, CRU, HERIDO ->
 Ejemplo registro

Total 170.854 registros

El nuevo árbol generado por el programa Weka estaba conformado por 88 hojas y 138 niveles. Si bien su porcentaje de efectividad disminuyó levemente al 76.4378%, este nuevo árbol facilitó la interpretación de resultados y análisis de fallos. Este último árbol generado es ilustrado a continuación:

Afectacion = ATR: HERIDO (19346.0/718.0) | | | | | Fecha <= 1284354000000
 Afectacion = COC: HERIDO (6625.0/99.0) | | | | | Fecha <= 1272776400000: HERIDO (3.0/1.0)
 Afectacion = CHO | | | | | Fecha > 1272776400000: DANOS (4.0)
 | observacion = ADL: DANOS (5866.0/1615.0) | | | | | Fecha > 1284354000000
 | observacion = IMP: DANOS (23608.0/5890.0) | | | | | Fecha <= 1287464400000: MUERE (4.0/1.0)
 | observacion = FME: DANOS (933.0/291.0) | | | | | Fecha > 1287464400000: DANOS (3.0/1.0)
 | observacion = VIA: DANOS (1254.0/420.0) | | | | | hora > 61320000: MUERE (2.0)
 | observacion = SEN | | | | | hora > 67920000
 | hora <= 43260000 | | | | | hora <= 80820000
 | | hora <= 42840000: HERIDO (1943.0/908.0) | | | | | hora <= 78000000: DANOS (4.0/1.0)
 | | hora > 42840000 | | | | | hora > 78000000: HERIDO (3.0)
 | | | Fecha <= 1277874000000 | | | | | hora > 80820000: DANOS (5.0)
 | | | | Fecha <= 1276578000000: DANOS (79.0/34.0) | | | | | observacion = VIA
 | | | | Fecha > 1276578000000: HERIDO (7.0) | | | | | Area = URBANO: HERIDO (207.0/29.0)
 | | | | Fecha > 1277874000000: DANOS (197.0/39.0) | | | | | Area = RURAL
 | | hora > 43260000: HERIDO (13063.0/6313.0) | | | | | hora <= 84120000
 | observacion = CRU: HERIDO (87.0/20.0) | | | | | hora <= 39120000: DANOS (5.0)
 | observacion = ALC: HERIDO (2396.0/1056.0) | | | | | hora > 39120000: HERIDO (25.0/10.0)
 | observacion = OTR: DANOS (74813.0/18739.0) | | | | | hora > 84120000: DANOS (9.0/1.0)
 | observacion = FSE | | | | | observacion = SEN: HERIDO (2383.0/67.0)
 | Fecha <= 1265864400000: HERIDO (10.0/4.0) | | | | | observacion = CRU
 | Fecha > 1265864400000: DANOS (24.0/10.0) | | | | | hora <= 76620000: HERIDO (4.0)
 | observacion = NRP | | | | | hora > 76620000: DANOS (2.0)
 | hora <= 93720000: DANOS (5620.0/1754.0) | | | | | observacion = ALC: HERIDO (88.0/5.0)
 | hora > 93720000 | | | | | observacion = OTR
 | | Fecha <= 1266901200000: HERIDO (63.0/20.0) | | | | | Area = URBANO: HERIDO (2148.0/394.0)
 | | Fecha > 1266901200000: DANOS (302.0/120.0) | | | | | Area = RURAL
 | observacion = LIC: HERIDO (1.0) | | | | | Fecha <= 1282626000000
 | observacion = MTR: HERIDO (860.0/337.0) | | | | | Fecha <= 1277355600000: HERIDO (64.0/32.0)
 | observacion = SCU: HERIDO (120.0/61.0) | | | | | Fecha > 1277355600000
 Afectacion = OTR | | | | | Fecha <= 1281157200000
 | observacion = ADL | | | | | hora <= 50700000: MUERE (5.0/1.0)
 | Area = URBANO | | | | | hora > 50700000
 | | Fecha <= 1269061200000: HERIDO (4.0) | | | | | hora <= 79020000: HERIDO (6.0)
 | | Fecha > 1269061200000 | | | | | hora > 79020000
 | | | Fecha <= 1276059600000: DANOS (6.0) | | | | | Fecha <= 1279688400000: HERIDO
 | | | Fecha > 1276059600000: HERIDO (6.0/2.0) | | | | | (3.0/1.0)
 | Area = RURAL: DANOS (4.0/1.0) | | | | | Fecha > 1279688400000: MUERE (3.0)
 | observacion = IMP | | | | | Fecha > 1281157200000: HERIDO (5.0)
 | Area = URBANO: HERIDO (161.0/44.0) | | | | | Fecha > 1282626000000: DANOS (45.0/18.0)
 | Area = RURAL | | | | | observacion = FSE: HERIDO (2.0/1.0)
 | | Fecha <= 1289106000000 | | | | | observacion = NRP: HERIDO (467.0/22.0)
 | | hora <= 41100000 | | | | | observacion = LIC: HERIDO (1.0)
 | | | Fecha <= 1271480400000: HERIDO (3.0/1.0) | | | | | observacion = MTR: HERIDO (19.0/4.0)
 | | | Fecha > 1271480400000: DANOS (3.0/1.0) | | | | | observacion = SCU
 | | | hora > 41100000: HERIDO (23.0/4.0) | | | | | hora <= 72600000: DANOS (7.0/1.0)
 | | | Fecha > 1289106000000: DANOS (7.0/2.0) | | | | | hora > 72600000: HERIDO (3.0)
 | observacion = FME | | | | | Afectacion = VLC
 | Area = URBANO: HERIDO (47.0/14.0) | | | | | Area = URBANO: HERIDO (6131.0/481.0)
 | Area = RURAL | | | | | Area = RURAL
 | | Fecha <= 1265346000000: HERIDO (4.0) | | | | | observacion = ADL
 | | Fecha > 1265346000000 | | | | | hora <= 37020000: DANOS (8.0/1.0)
 | | hora <= 67920000 | | | | | hora > 37020000
 | | | hora <= 61320000 | | | | | hora <= 49860000: HERIDO (5.0/2.0)

```

| | | | hora > 49860000
| | | | hora <= 89100000: DANOS (20.0/7.0)
| | | | hora > 89100000: HERIDO (3.0/1.0)
| | | | observacion = IMP: DANOS (158.0/72.0)
| | | | observacion = FME
| | | | hora <= 32220000: DANOS (13.0/1.0)
| | | | hora > 32220000: HERIDO (165.0/82.0)
| | | | observacion = VIA
| | | | Fecha <= 1268542800000
| | | | hora <= 62400000
| | | | Fecha <= 1264568400000: DANOS (12.0/2.0)
| | | | Fecha > 1264568400000: HERIDO (17.0/5.0)
| | | | hora > 62400000
| | | | hora <= 77100000
| | | | | hora <= 75000000: HERIDO (5.0/1.0)
| | | | | hora > 75000000: MUERE (2.0)
| | | | | hora > 77100000: HERIDO (15.0/1.0)
| | | | Fecha > 1268542800000: DANOS (187.0/75.0)
| | | | observacion = SEN: HERIDO (173.0/78.0)
| | | | observacion = CRU: HERIDO (2.0)
| | | | observacion = ALC: HERIDO (61.0/17.0)
| | | | observacion = OTR: HERIDO (757.0/302.0)
| | | | observacion = FSE: DANOS (2.0)
| | | | observacion = NRP: HERIDO (32.0/10.0)
| | | | observacion = LIC: HERIDO (0.0)
| | | | observacion = MTR
| | | | Fecha <= 1269579600000: DANOS (2.0/1.0)
| | | | Fecha > 1269579600000: HERIDO (2.0)
| | | | observacion = SCU: HERIDO (6.0/1.0)
| | | | Afectacion = INC: DANOS (62.0/10.0)

```

IV. DISCUSIONES

El último árbol generado, permite un porcentaje de clasificación del 76.4378%, en consecuencia, de los 170.854 registros, el programa clasificó de forma incorrecta 40.257 registros equivalentes al 23.5622%. No obstante, lo anterior, es importante resaltar que 91.89% de los errores de clasificación se encuentran en 8 de las 88 ramas del árbol. Estas ramas son resaltadas con rojo en el árbol anterior e ilustradas en la siguiente tabla:

TABLE IV
ANÁLISIS DE LOS PRINCIPALES FALLOS DEL ÁRBOL DE CLASIFICACIÓN

Rama	Fallos	% Fallos
observacion = IMP: DANOS (23608.0/5890.0)	5.890	14.63%
hora > 43260000: HERIDO (13063.0/6313.0)	6.313	15.68%
observacion = OTR: DANOS (74813.0/18739.0)	18.739	46.54%
observacion = ADL: DANOS (5866.0/1615.0)	1.615	4.01%
hora <= 42840000: HERIDO (1943.0/908.0)	908	2.2%
observacion = ALC: HERIDO (2396.0/1056.0)	1.056	2.6%
Afectacion = ATR: HERIDO (19346.0/718.0)	718	1.7%
hora <= 93720000: DANOS (5620.0/1754.0)	1.754	4.35%
Total	36.993	91.89%

Considerando, el anterior análisis, si se obvian las ramas ilustradas en la Tabla IV, se alcanzará un nivel de predicción en la variable riesgo del 1- (36.993-40.257) /170.854 = 98.08%. Se resalta que por razones de espacio no se hace un análisis más profundo de los resultados. Si bien, La Inteligencia artificial está siendo usada en la predicción de accidentes de tránsito [9] el empleo de esta herramienta Weka en accidentes de tránsito no se encuentra referenciada en la literatura y facilita el análisis y predicción de los mismos, disminuyendo considerablemente los costos de predicción.

V. CONCLUSIONES

El análisis de los accidentes de tránsito, por medio de esta herramienta de programación denominada Weka, constituye una excelente forma de predecir los factores que producen un accidente tránsito y establecer el riesgo de los mismos. Con la base de datos original se alcanzaron niveles de predicción del 80%. No obstante, cuando la base de datos es reducida, con el fin de lograr una mejor interpretación del árbol generado por programa Weka, la clasificación bajo el 77% aproximadamente. Sin embargo, evitando las 8 de las 88 ramas del árbol que más producen fallos, se puede lograr una clasificación del 98%.

AGRADECIMIENTOS

Se agradece la colaboración a la Universidad Nacional de Colombia Sede Manizales por la ayuda prestada para el desarrollo de esta investigación.

REFERENCIAS

- [1] R Deb, A Wee-Chung, "Noisy values detection and correction of traffic accident data", *Information Sciences*, vol. 1, no 476, pp 132 -146, February 2019.
- [2] L. Zhongping, C. Lirong and Chen Jianhui, "Traffic accident modelling via self-exciting point processes", *Reliability Engineering and System Safety*, vol 1, no. 180, pp 312–320, December 2018.
- [3] Tzu-Ying Chen, Rong-Chang Jou, "Using HLM to investigate the relationship between traffic accident risk of private vehicles and public transportation", *Transportation Research Part A*, vol 1, no. 119, 148–161, February 2019.
- [4] Datos abiertos. Gobierno digital de Colombia. <https://www.datos.gov.co/Transporte/Registro-nacional-de-accidentes-de-transito/jb4r-tjbv>. Actualizado 16 Agosto de 2017. Revisado febrero de 2019.
- [5] Universidad de Waikato, Hamilton, Nueva Zelandia. <https://www.cs.waikato.ac.nz/ml/weka/downloading.html> (Revisado febrero de 2019).
- [6] J.R. Quinlan, C4.5: Programs for Machine Learning, Elsevier Science & Technology Books, 1992.
- [7] R. Duda, P. E. Hart and D. Stork, Pattern Classification, New York, John Wiley (2001).
- [8] R. Mosquera, O. Castrillón and L. Parra, "Máquinas de Soporte Vectorial, Clasificador Naïve Bayes y Algoritmos Genéticos para la Predicción de Riesgos Psicosociales en Docentes de Colegios Públicos Colombianos." *Información Tecnológica*, vol. 29, no. 6, pp 153 -162, December 2018.
- [9] G. Yasmany, P. Rojas, E. Duque and H. Rojas, "Calibración del Modelo Predictivo de Accidentes de Tránsito del HSM en Carreteras del Cantón Loja (Ecuador)", *Innova Research Journal*. No. 3. 114-123, 2018.