"Prediction of transit accidents in Ecuador for 2020 applying supervised machine learning"

Universidad Nacional de Loja

Facultad de Energía, Carrera de Ingeniería en Sistemas

Carrera de Ingeniería en Sistemas / Computación

Jossed Carpio-Ortega

Abstract

This project focuses on traffic accidents in Ecuador, which have become an uncontrollable problem for transit related entities, taking into account this problem, the data related to the years 2016, 2017 and 2018 were obtained from INEC to make a prediction of the day, hour and cause for which in Ecuador such events would occur in 2020; the dataset was joined and cleaned, then the type of some variables were changed to numerical data to facilitate its use when applying the linear regression algorithms and KNN in the RapidMiner tool, it is important to mention that samples of 10 were used, 100 and 500 data for the respective predictions, at the end of this process and evaluating the performance of the algorithms it was evidenced that the algorithm that generates better results is that of linear regression, since it identified that on Friday, in the interval of 12:00 at 1:00 pm and the driver's lack of expertise and imprudence could generate traffic accidents.

1. Problem

Transit accidents have alarming figures, in recent years despite awareness campaigns to try to reduce these figures.

What day, hour and cause will the majority of traffic accidents be registered in the year 2020 in Ecuador?

2. Data collection

Agencia Nacional de Tránsito · ECU 911 Servicio Integrado de Seguridad · Cruz Roja Ecuatoriana · Municipio de Loja

BANCO DE DATOS ABIERTOS · INEC instituto nacional de estadística y censos

Anuario de Transportes

Año: 2018

Seleccione el período: ANUAL

Se han realizado 97 descargas para el año seleccionado.

Datos Abiertos CSV	ZIP	★★★	**9.45 MB**	Descargar
Base de Datos SPSS	ZIP	☆☆☆	9.01 MB	Descargar
Tabulados y series históricas	VARIOS	★☆☆		Descargar

URL: http://aplicaciones3.ecuadorencifras.gob.ec/BIINEC-war/index.xhtml

- 2016_Accidentes Tránsito_BDD
  - Diccionario Datos AT 2016
  - Metadatos AT 2016
- 2017_Siniestros_de_Tránsito_BDD
  - Diccionario Datos Siniestros de Tránsito ...
  - Tabla Metadatos Siniestros de Tránsito 2...
- 2018_Siniestros_de_Tránsito_BDD
  - Diccionario Datos Siniestros de Tránsito ...
  - Tabla Metadatos Siniestros de Tránsito 2...

3. Data preparation

2016 · 2017 · 2018 → Siniestros de Tránsito 2016-2018 → CSV

COD_PROVIN	PROVINCIA	COD_CANTO	CANTON	COD_MES	MES	COD_DIA	DIA	COD_HORA	HORA
9	GUAYAS	901	GUAYAQUIL	1	ENERO	1	LUNES	0	00:00 A 00:5
9	GUAYAS	901	GUAYAQUIL	1	ENERO	1	LUNES	3	03:00 A 03:5
9	GUAYAS	901	GUAYAQUIL	1	ENERO	1	LUNES	5	05:00 A 05:5
9	GUAYAS	901	GUAYAQUIL	1	ENERO	1	LUNES	8	08:00 A 08:5

COD_CLASE	CLASE	COD_CAUSA	CAUSA	COD_ZONA	ZONA	NUM_FALLEC	NUM_LESION	TOTAL_VICTIMAS
1	ATROPELLOS	3	EXCESO VELC	1	URBANA	0	1	1
4	ESTRELLAMIE	1	EMBRIAGUEZ	1	URBANA	0	3	3
7	PÉRDIDA DE	3	EXCESO VELC	1	URBANA	0	0	0
7	PÉRDIDA DE	3	EXCESO VELC	1	URBANA	0	1	1

CSV

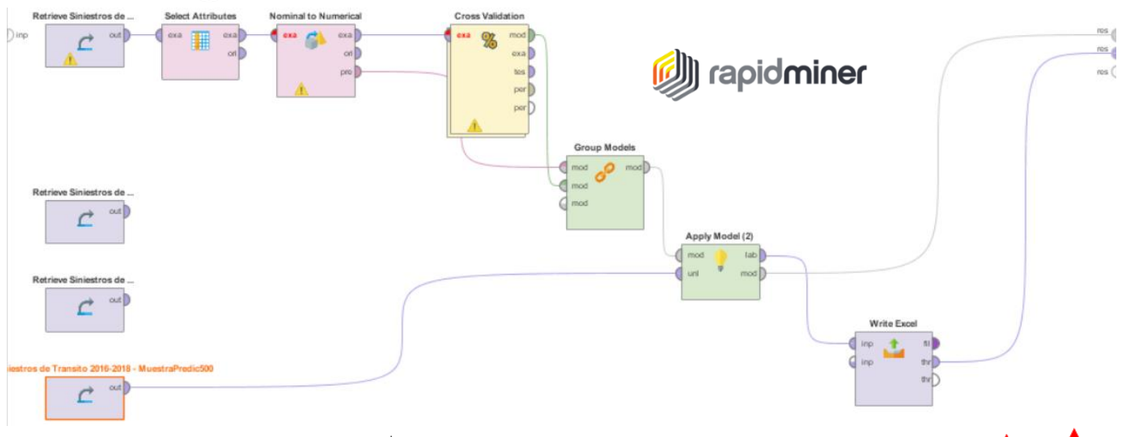
COD_PROVIN	COD_CANTON	COD_HORA	COD_DIA	COD_MES	COD_ANIO	COD_CLASE	COD_CAUSA	COD_ZONA
18	1801	4	5	1	1	4	1	1
18	1801	14	5	1	1	4	1	1
18	1801	6	5	1	1	3	1	2
18	1801	11	5	1	1	4	1	1

4. Data splitting

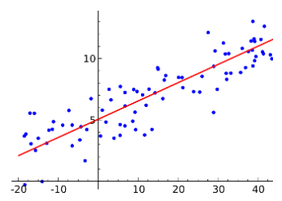
Predicción > 31

- Siniestros de Tránsito 2016-2018 - MuestraPredic10
- Siniestros de Tránsito 2016-2018 - MuestraPredic100
- Siniestros de Tránsito 2016-2018 - MuestraPredic500

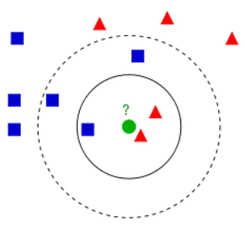
5. Training a model



Linear Regression



K-Nearest Neighbors



6. Model validation

$(Prediction * 00)/RealValue$

Linear Regression

prediction(COD_DIA)	Desempeño
4.736	94.71
4.474	89.47
4.683	93.66
4.552	91.05
3.597	71.95
prediction(COD_HORA)	
13.142	87.62
13.116	69.03
12.875	71.53
12.608	210.13
12.608	66.36
prediction(COD_CAUSA)	
3.999	99.97
3.947	65.78
3.952	395.21
3.941	98.52
3.998	399.81

5 = Viernes

12 = 12:00 A 12:59
13 = 13:00 A 13:59

4 = Impericia o imprudencia del conductor

K-Nearest Neighbors

prediction(COD_DIA)	Desempeño
6.600	132.00
5.600	112.00
6.400	128.00
6.400	128.00
4.200	84.00
prediction(COD_HORA)	
7.200	180.00
7.200	51.43
7.200	120.00
7.200	65.45
13.400	60.91
prediction(COD_CAUSA)	
2.800	280.00
3.400	340.00
3.200	320.00
3.600	360.00
7.000	70.00

Project File Repository
URL: https://github.com/Jossed94Carpio/AI_Proyecto_SiniestrosT.git