

Resumen

El presente proyecto se enfoca en los siniestros de tránsito en el Ecuador, mismos que se han convertido en un problema incontrolable por las entidades relacionadas con el tránsito, tomando en cuenta este problema se obtuvo del INEC los datos relacionados de los años 2016, 2017 y 2018 para realizar una predicción del día, hora y causa por las que en el Ecuador se producirían este tipo de acontecimientos en el 2020; se procedió a unir y limpiar los dataset, luego a cambiar el tipo de algunas variables a datos numéricos para facilitar su uso al momento de aplicar los algoritmos de regresión lineal y KNN en la herramienta RapidMiner, es importante mencionar que se utilizó muestras de 10, 100 y 500 datos para las respectivas predicciones, al finalizar este proceso y evaluando el desempeño de los algoritmos se evidenció que el algoritmo que genera mejores resultados es el de regresión lineal, ya que identificó que el día viernes, en el intervalo de 12:00 a 13:00 y la impericia e imprudencia del conductor, podrían generar siniestros de tránsito.

1. Problema



Los siniestros de tránsito presentan cifras alarmantes, en los últimos años pese a que se ha realizado campañas de concientización para tratar de disminuir estas cifras.



¿Qué día, hora y por qué causa se podrán registrar la mayoría de siniestros de tránsito en el año 2020 en el Ecuador?



BANCO DE
DATOS ABIERTOS



2. Recopilación de datos

Anuario de Transportes

Año: 2018

Seleccione el período:

ANUAL



Se han realizado 97 descargas para el año seleccionado.

Datos Abiertos CSV	9.45 MB	Descargar
Base de Datos SPSS	9.01 MB	Descargar
Tabulados y series históricas		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 ...

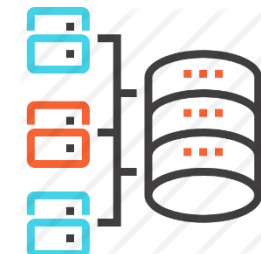
Tabla Metadatos Siniestros de Tránsito 2...

3. Preparación de datos

2016

2017

2018



Siniestros de Tránsito 2016-2018



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	



COD_PROVIN	COD_CANTO	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. División de datos

Predicción

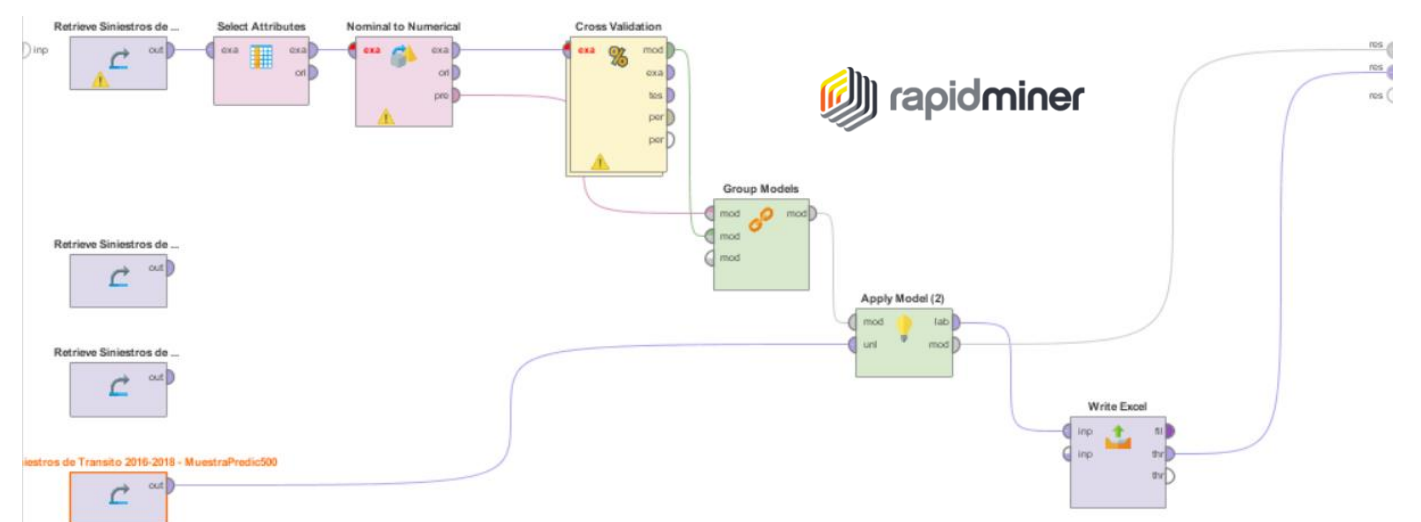


Siniestros de Tránsito 2016-2018 - MuestraPredic10

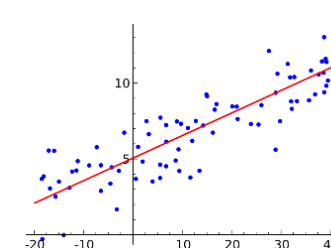
Siniestros de Tránsito 2016-2018 - MuestraPredic100

Siniestros de Tránsito 2016-2018 - MuestraPredic500

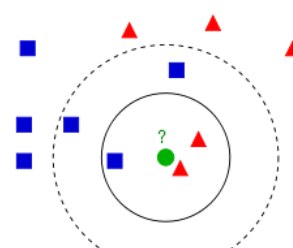
5. Entrenamiento del modelo



Linear Regression



K-Nearest Neighbors



6. Modelo de validación $(\text{Predicción} * 100) / \text{ValorReal}$

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

Repositorio de archivos del proyecto

URL: https://github.com/Jossed94Carpio/AI_Proyecto_SiniestrosT.git