Supervised Learning: Prediction of the ECU911 centers with the greatest number of police incidents in Ecuador for 2020

Facultad de Energía, Carrera de Ingeniería en Sistemas

Raúl Romero-Vega

Abstract

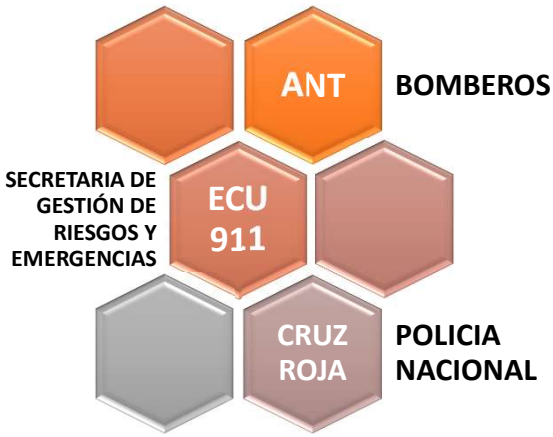
Using artificial intelligence, through machine learning models, it has been predicted which will be the ECU 911 Centres with the greatest number of police incidents by the year 2020, carrying it out through 6 stages carried out sequentially; The first step was to define the problem and collect data from the Ecuadorian Open Data Portal and ECU911, which were then processed one by one to represent them categorically in order to apply classification models, finally, the results have been interpreted to answer the research question, predicting that the centres with the greatest number of police incidents by 2020 will be Quito and Samborondon, and using the confusion matrices, two models with the best performance have been contrasted.

1. Problem

Research question

¿Cuáles serán los Centros del ECU 911 con mayores incidentes policiales cometidos en Ecuador para el 2020, de acuerdo a los registros del ECU 911?

Target public



Type of problem

Supervised Learning

2. Data collection

Data set used for machine learning

http://catalogo.datosabiertos.gob.ec/dataset/bf5970ed-16fb-40af-9e78-910df47ac86a/resource/14980edc-e7e8-4f33-9fd5-da435f3734b9/download/estadisticas-enero.xlsx

Later statistics to evaluate the results

https://www.ecu911.gob.ec/wp-content/uploads/2017/06/Estadisticas.pdf

3. Data preparation

CENTRO	AÑO	MES	# INCIDENTE ...	# INCIDENTE ...	# INCIDENTE ...	# INCIDENTE ...	# INCIDENTE ...	# INCIDENTE ...
SAMBORONDÓN	2012	FEBRERO	17,684	2,602	2,762	3,439	690	327
SAMBORONDÓN	2012	MARZO	21,836	2,945	3,568	3,171	700	219
SAMBORONDÓN	2012	ABRIL	20,856	3,161	3,114	2,541	730	34

Class	Muy Bajo	Bajo	Medio	Alto	Muy Alto

Range	Maximum Value	Minimum Value	Amplitude
59092	59331	239	11819

$$Range = Max - Min$$

Sturges rule

$$N = 1 + 3.322 * \text{Log}(440)$$

$$Amplitude = \frac{Range}{N}$$

CENTRO	ANIO	MES	INCIDENTE_PO...	INCIDENTE_TR...	INCIDENTE_SA...	INCIDENTE_BO...	INCIDENTE_GE...
Samborondón	2012	Febrero	Bajo	Bajo	Bajo	Muy Bajo	Medio
Samborondón	2012	Marzo	Bajo	Bajo	Bajo	Muy Bajo	Bajo
Samborondón	2012	Abril	Bajo	Bajo	Bajo	Muy Bajo	Muy Bajo

4. Data splitting

It was not necessary to divide the data because there is another 2016 data set that will be used for the validation of the models that will be applied during the training.

Training

- ECU911 data available from 2012 to 2015

Test

- ECU911 2016 data available

5. Training a model

6. Model validation

Performance of the Decision Tree model						
Accuracy					92,8%	
Classification error					7,2%	
Confusion matrix						
	True Bajo	True Medio	True Alto	True Muy Alto	True Muy Bajo	Class Precision
Pred. Bajo	3	1	0	0	0	75%
Pred. Medio	0	6	0	0	0	100%
Pred. Alto	0	4	5	2	0	45,45%
Pred. Muy Alto	0	0	0	3	0	100%
Pred. Muy Bajo	2	0	0	0	99	98,02%
Class Recall	60%	54,55%	100%	60%	100%	

Performance of the Fast Large Margin model						
Accuracy					92,0%	
Classification error					8%	
Confusion matrix						
	True Bajo	True Medio	True Alto	True Muy Alto	True Muy Bajo	Class Precision
Pred. Bajo	3	0	0	0	0	100%
Pred. Medio	0	8	4	0	0	66,67%
Pred. Alto	0	0	4	3	0	57,14%
Pred. Muy Alto	0	0	0	2	0	100%
Pred. Muy Bajo	3	0	0	0	98	97,03%
Class Recall	50%	100%	50%	40%	100%	

Repository: https://github.com/raulrrv/MachineLearning-ECU911.git