

1. Entender y definir el problema

1.1. Público objetivo y contexto

- Agencia Nacional de Tránsito - ANT
- Policía Nacional
- Secretaria Nacional de Planificación y Desarrollo: Según la constitución de la república en el Art. 280.- El Plan Nacional de Desarrollo es el instrumento al que se sujetarán las políticas, programas y proyectos públicos; la programación y ejecución del presupuesto del Estado; y la inversión y la asignación de los recursos públicos. Su observancia será de carácter obligatorio para el sector público e indicativo para los demás sectores
- Secretaría Técnica de Prevención Integral de Drogas
- Bomberos
- Ministerios Salud
- Servicio Nacional de Gestión de Riesgos y Emergencias
- Cruz Roja Ecuatoriana
- Ministerio de Defensa
- Ministerio Coordinador de Seguridad
- Servicio Integrado de Seguridad ECU 911: Al contar con estadísticas, esto ayuda a conocer la real magnitud de un incidente y enriquecer la toma de decisiones, evaluar el sistema y mejorar acciones para eventos futuros.

1.2. Datos disponibles

Dataset proporcionado por el Portal de Datos Abiertos Ecuador

Estadísticas proporcionadas por el Servicio Integrado de Seguridad ECU 911

1.3. Tipo de problema

Aprendizaje supervisado

1.4. Pregunta de investigación

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

2. Recopilar datos

2.1. Disponibilidad

Dataset utilizado para el aprendizaje de la máquina

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

Estadísticas posteriores para evaluar los resultados

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

3. Preparar los datos

3.1. Limpieza

Dataset original

CENTRO	AÑO	MES	# INCIDENTE POLICIAL (TOTAL)	# INCIDENTE TRÁNSITO (TOTAL)	# INCIDENTE SALUD (TOTAL)	# INCIDENTE SERVICIOS MUNICIPALES (TOTAL)	# INCIDENTE BOMBEROS (TOTAL)	# INCIDENTE GESTIÓN DE RIESGO (TOTAL)	# INCIDENTE MILITAR (TOTAL)	TOTAL DE INCIDENTES RECEPTADOS	# INCIDENTE CLAVES ROJAS (TOTAL)	# DE LLAMADAS MALINTENDIDAS O INDEBIDAS (TOTAL)	CANTIDAD DE TELEFONOS ENVIADOS A SANCIÓN	# DE LLAMADAS GENERADAS POR LOS TELEFONOS SANCIONADOS	# DE LLAMADAS NORMALES (TOTAL)	TOTAL LLAMADAS RECIBIDAS (TOTAL)
SAMBORONDÓN	2012	FEBRERO	17,684	2,602	2,762	3,439	690	327	3	27,507	13,920	116,863	578	15,762	29,609	34,371
SAMBORONDÓN	2012	MARZO	21,836	2,945	3,568	3,171	700	219	18	32,457	16,425	74,615	631	16,405	34,353	37,984
SAMBORONDÓN	2012	ABRIL	20,856	3,161	3,114	2,541	730	34	53	30,489	15,429	312,842	2,922	85,093	33,360	43,711
SAMBORONDÓN	2012	MAYO	24,164	3,736	2,836	2,213	610	15	15	33,589	16,997	360,485	3,362	89,715	40,473	44,838
SAMBORONDÓN	2012	JUNIO	23,639	3,819	2,598	1,148	611	10	12	31,837	16,111	355,838	3,258	84,625	37,254	41,512
SAMBORONDÓN	2012	JULIO	25,994	4,800	2,648	1,379	921	7	13	35,762	18,097	390,133	18,324	172,417	42,981	49,298
SAMBORONDÓN	2012	AGOSTO	23,662	3,128	2,313	767	1,075	1	25	30,971	15,673	351,363	16,762	153,684	37,170	41,242
SAMBORONDÓN	2012	SEPTIEMBRE	21,233	3,399	1,993	562	1,188	5	14	28,394	14,368	340,861	15,756	145,163	34,499	40,258
SAMBORONDÓN	2012	OCTUBRE	19,950	3,754	2,421	701	1,340	3	6	28,175	14,258	353,859	14,705	122,895	33,802	41,507
SAMBORONDÓN	2012	NOVIEMBRE	20,685	3,410	2,289	803	1,465	2	4	28,658	14,502	345,327	14,112	118,931	34,762	40,274
SAMBORONDÓN	2012	DICIEMBRE	25,570	4,604	3,100	2,066	1,518	7	7	36,872	18,659	330,903	14,112	109,746	43,298	40,496
TOTAL			245,273	39,358	29,642	18,790	10,848	630	170	344,711	174,439	3,333,089	104,522	1,114,436	401,561	4,488
SAMBORONDÓN	2013	ENERO	37,598	3,234	2,302	2,285	644	12	16	46,091	21,866	395,192	17,156	141,747	52,813	48,979
SAMBORONDÓN	2013	FEBRERO	37,623	3,660	2,674	2,201	489	45	42	46,734	21,817	374,039	18,364	173,892	54,074	47,148
SAMBORONDÓN	2013	MARZO	42,649	3,840	3,653	4,144	619	480	46	55,431	26,217	387,290	18,828	185,052	66,318	49,146
SAMBORONDÓN	2013	ABRIL	40,124	3,552	3,637	2,673	661	29	22	50,698	24,772	383,076	18,652	182,883	60,978	48,630
SAMBORONDÓN	2013	MAYO	40,168	3,834	3,928	1,669	824	6	24	50,453	25,906	359,244	17,221	158,775	60,049	48,274

Se realiza la eliminación de las observaciones que no pertenecen o que dificultan la aplicación de algún modelo de predicción. Esto se realizó con TurboPrep de RapidMiner.

CENTRO	AÑO	MES	# INCIDENTE ...	# INCIDENTE ...	# INCIDENTE ...	# INCIDENTE ...	# INCIDENTE ...	# INCIDENTE ...
Category	Number	Category	Category	Category	Category	Category	Category	Category
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
SAMBORONDÓN	2012	MAYO	24,164	3,736	2,836	2,213	610	15
SAMBORONDÓN	2012	JUNIO	23,639	3,819	2,598	1,148	611	10
SAMBORONDÓN	2012	JULIO	25,994	4,800	2,648	1,379	921	7
SAMBORONDÓN	2012	AGOSTO	23,662	3,128	2,313	767	1,075	1
SAMBORONDÓN	2012	SEPTIEMBRE	21,233	3,399	1,993	562	1,188	5
SAMBORONDÓN	2012	OCTUBRE	19,950	3,754	2,421	701	1,340	3

3.2.Valores atípicos

En el presente dataset no se detectaron datos atípicos.

3.3.Transformaciones

Según las cifras de seguridad del INEC, los delitos de mayor connotación psicosocial son:

- Homicidios Intencionales
- Robo a Personas
- Robo a Domicilios
- Robo de Carros
- Robo de Motos
- Robo a Unidades Económicas
- Robo de Bienes, Accesorios y Autopartes de Vehículos
- Violaciones
- Fallecidos en sitio por siniestros de tránsito

Relacionando los delitos anteriores, de acuerdo a los incidentes registrados por el ECU 911, se deduce que se utilizarán las siguientes características de nuestro dataset.

INCIDENTE POLICIAL

INCIDENTE TRÁNSITO

INCIDENTE SALUD

INCIDENTE BOMBEROS

INCIDENTE GESTIÓN DE RIESGO

INCIDENTE_POLICIAL		
Valor máximo	59331	
Valor mínimo	239	
Rango	59092	
N = num intervalos	9.78154979	5
Amplitud	11818.4	11819
Clases	Amplitud Intervalo Clase	
Muy Bajo	239	12058
Bajo	12059	23878
Medio	23879	35698
Alto	35699	47518
Muy Alto	47519	59338

Rango= Max – Min

Regla de Sturges

$$N = 1 + 3.322 * \log(440)$$

$$\text{Amplitud} = \frac{\text{Rango}}{N}$$

CENTRO Category	ANIO Category	MES Category	INCIDENTE_TRA... Number	INCIDENTE_SAL... Number	INCIDENTE_BOM... Number	INCIDENTE_GES... Number	INCIDENTE_POL... Category
Samborondón	2012	Febrero	2602	2762	690	327	Bajo
Samborondón	2012	Marzo	2945	3568	700	219	Bajo
Samborondón	2012	Abril	3161	3114	730	34	Bajo
Samborondón	2012	Mayo	3736	2836	610	15	Bajo
Samborondón	2012	Junio	3819	2598	611	10	Bajo
Samborondón	2012	Julio	4800	2648	921	7	Bajo
Samborondón	2012	Agosto	3128	2313	1075	1	Bajo
Samborondón	2012	Septiembre	3399	1993	1188	5	Bajo
Samborondón	2012	Octubre	3754	2421	1340	3	Bajo
Samborondón	2012	Noviembre	3410	2289	1465	2	Baio

CENTRO Category	ANIO Category	MES Category	INCIDENTE_PO... Category	INCIDENTE_TR... Category	INCIDENTE_SA... Category	INCIDENTE_BO... Category	INCIDENTE_GE... Category
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
Samborondón	2012	Mayo	Medio	Medio	Bajo	Muy Bajo	Muy Bajo
Samborondón	2012	Junio	Bajo	Medio	Bajo	Muy Bajo	Muy Bajo
Samborondón	2012	Julio	Medio	Medio	Bajo	Bajo	Muy Bajo
Samborondón	2012	Agosto	Bajo	Bajo	Bajo	Bajo	Muy Bajo
Samborondón	2012	Septiembre	Bajo	Bajo	Muy Bajo	Bajo	Muy Bajo
Samborondón	2012	Octubre	Bajo	Medio	Bajo	Bajo	Muy Bajo
Samborondón	2012	Noviembre	Bajo	Bajo	Bajo	Bajo	Muy Bajo
Samborondón	2012	Diciembre	Medio	Medio	Bajo	Bajo	Muy Bajo
Samborondón	2013	Enero	Alto	Bajo	Bajo	Muy Bajo	Muy Bajo
Samborondón	2013	Febrero	Alto	Medio	Bajo	Muy Bajo	Muy Bajo
Samborondón	2013	Marzo	Alto	Medio	Bajo	Muy Bajo	Alto
Samborondón	2013	Abril	Alto	Bajo	Bajo	Muy Bajo	Muy Bajo
Samborondón	2013	Mayo	Alto	Medio	Bajo	Muy Bajo	Muy Bajo
Samborondón	2013	Junio	Alto	Medio	Bajo	Muy Bajo	Muy Bajo
Samborondón	2013	Julio	Alto	Medio	Medio	Baio	Muv Baio

440 rows - 8 columns (8 nominal)

4. Dividir los datos

No hizo falta la división de los datos debido a que disponemos de otro dataset (datos del 2016) que será utilizado para la validación de los modelos que se apliquen de ML.

4.1. Entrenamiento

CENTRO <i>Category</i>	AÑO <i>Category</i>	MES <i>Category</i>	INCIDENTE_POLICIAL <i>Category</i>
Samborondón	2015	Enero	Muy Alto
Samborondón	2015	Febrero	Muy Alto
Samborondón	2015	Marzo	Muy Alto
Samborondón	2015	Abril	Muy Alto
Samborondón	2015	Mayo	Muy Alto
Samborondón	2015	Junio	Muy Alto
Samborondón	2015	Julio	Muy Alto
Samborondón	2015	Agosto	Muy Alto
Samborondón	2015	Septiembre	Muy Alto
Samborondón	2015	Octubre	Muy Alto
Samborondón	2015	Noviembre	Muy Alto
Samborondón	2015	Diciembre	Muy Alto

4.2. Test

CENTRO <i>Category</i>	AÑO <i>Category</i>	MES <i>Category</i>	INCIDENTES POLICIALES <i>Category</i>
SAMBORONDÓN	2016	ENERO	Muy Bajo
SAMBORONDÓN	2016	FEBRERO	Muy Alto
SAMBORONDÓN	2016	MARZO	Muy Alto
SAMBORONDÓN	2016	ABRIL	Muy Bajo
SAMBORONDÓN	2016	MAYO	Medio
SAMBORONDÓN	2016	JUNIO	Alto
SAMBORONDÓN	2016	JULIO	Muy Alto
SAMBORONDÓN	2016	AGOSTO	Muy Alto
SAMBORONDÓN	2016	SEPTIEMBRE	Muy Alto
SAMBORONDÓN	2016	OCTUBRE	Muy Alto
SAMBORONDÓN	2016	NOVIEMBRE	Muy Alto
SAMBORONDÓN	2016	DICIEMBRE	Muy Alto

5. Entrenar uno o varios modelos

Load Data Select Task Prepare Target Select Inputs Model Types Results

« RESTART < BACK > NEXT

Predict

Want to predict the values of a column?

Clusters

Want to identify groups in your data?

Outliers

Want to detect outliers in your data?

CENTRO Category	ANIO Category	MES Category	INCIDENTE_POLICIAL Category	INCIDENTE_TRANSITO Category	INCIDENTE_SALUD Category	INCIDENTE_BOMBER... Category	INCIDENTE_GESTION... Category
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
Samborondón	2012	Mayo	Medio	Medio	Bajo	Muy Bajo	Muy Bajo
Samborondón	2012	Junio	Bajo	Medio	Bajo	Muy Bajo	Muy Bajo
Samborondón	2012	Julio	Medio	Medio	Bajo	Bajo	Muy Bajo
Samborondón	2012	Agosto	Bajo	Bajo	Bajo	Bajo	Muy Bajo
Samborondón	2012	Septiembre	Bajo	Bajo	Muy Bajo	Bajo	Muy Bajo

440 rows - 8 columns (8 nominal)

Load Data Select Task Prepare Target Select Inputs Model Types Results

« RESTART < BACK > NEXT

Equal settings for all costs and benefits. Define Costs / Benefits...

Class of Highest Interest: Muy Bajo

Define Costs for Wrong Predictions

You can define costs for wrong predictions below. Please make sure to enter costs as negative numbers and benefits or gains as positive numbers.

Cost Matrix	True Muy B...	True Muy ...	True Alto	True Bajo	True Medio
Predicted Muy B...	1	-1	-1	-1	-1
Predicted Muy Alto	-1	1	-1	-1	-1
Predicted Alto	-1	-1	1	-1	-1
Predicted Bajo	-1	-1	-1	1	-1
Predicted Medio	-1	-1	-1	-1	1

Ok Cancel

Load Data Select Task Prepare Target Select Inputs Model Types Results

« RESTART < BACK > NEXT

Selected: 3 / Total: 7

☐ Deselect Red ☐ Deselect Yellow ☒ Select All ☒ Deselect All

Selected	Status ↑	Quality	Name	Correlation	ID-ness	Stability	Missing	Text-ness
<input type="checkbox"/>			INCIDENTE_BOMBEROS	42.09%	1.14%	84.55%	0.00%	32.01%
<input checked="" type="checkbox"/>			CENTRO	38.49%	3.64%	10.91%	0.00%	8.99%
<input checked="" type="checkbox"/>			ANIO	4.12%	1.14%	42.05%	0.00%	2.16%
<input checked="" type="checkbox"/>			MES	0.26%	2.73%	10.00%	0.00%	3.82%



Models

- ☒ Naive Bayes
- ☒ Generalized Linear Model
 - ☒ Use Regularization ☐ Calculate p-Values
- ☒ Logistic Regression
- ☒ Fast Large Margin
 - ☒ Automatically Optimize
- ☒ Deep Learning
- ☒ Decision Tree
 - ☒ Automatically Optimize Maximal Depth: 20

Data Preparation

- ☒ Remove Columns with Too Many Values
 - Maximum Number of Values: 50
- ☐ Extract Date Information
- ☐ Extract Text Information
 - Select Text Columns (0):
- Number of Extracted Features: 1,000
- ☐ Automatic Feature Selection
 - Additional Minutes (Maximum): 60
 - Final Feature Set should be: Accurate

Fast Large Margin - Predictions

Row No.	INCIDENTE_POLICIAL	prediction(INCIDENTE_POLICIAL)	confidence(Bajo)	confidence(Medio)	confidence(Alto)	confidence(Muy Alto)	confidence
10	Alto	Bajo	0.835	0.000	0.000	0	0.165
11	Alto	Alto	0.016	0.000	0.819	0.000	0.165
12	Muy Alto	Alto	0.005	0.000	0.756	0.075	0.165
13	Muy Alto	Muy Alto	0.177	0.000	0	0.658	0.165
14	Muy Alto	Alto	0.004	0.000	0.449	0.382	0.165
15	Muy Alto	Muy Alto	0.011	0.000	0.411	0.413	0.165
16	Muy Alto	Alto	0.009	0.000	0.807	0.020	0.165
17	Muy Alto	Muy Alto	0.007	0.000	0.278	0.550	0.165
18	Muy Alto	Alto	0.013	0.000	0.823	0.000	0.165

Decision Tree - Model



Decision Tree - Predictions

Row No.	INCIDENTE_POLICIAL	prediction(INCIDENTE_POLICIAL)	confidence(Bajo)	confidence(Medio)	confidence(Alto)	confidence(Muy Alto)	confidence(...)
10	Alto	Alto	0	0.000	0.667	0.000	0.333
11	Alto	Alto	0	0.000	0.667	0.000	0.333
12	Muy Alto	Alto	0	0.166	0.667	0.000	0.167
13	Muy Alto	Alto	0	0.000	0.667	0.000	0.333
14	Muy Alto	Muy Alto	0	0.000	0.167	0.500	0.333
15	Muy Alto	Muy Alto	0	0.000	0.167	0.500	0.333
16	Muy Alto	Muy Alto	0.000	0.000	0.167	0.500	0.333
17	Muy Alto	Muy Alto	0	0.166	0.167	0.500	0.167
18	Muy Alto	Alto	0	0.000	0.667	0.000	0.333
19	Medio	Medio	0	0.667	0.000	0	0.333

6. Medir el desempeño

Confusión o error Matrix es una tabla que describe el rendimiento de un modelo supervisado de Machine Learning en los datos de prueba, donde se desconocen los verdaderos valores. Se llama “matriz de confusión” porque hace que sea fácil detectar dónde el sistema está confundiendo dos clases.

Fast Large Margin - Performance

Accuracy	92.0%	± 2.8%
Classification Error	8.0%	± 2.8%

Confusion Matrix

	true Bajo	true Medio	true Alto	true Muy Alto	true Muy Bajo	class precision
pred. Bajo	3	0	0	0	0	100.00%
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	100.00%
pred. Muy Bajo	3	0	0	0	98	97.03%
class recall	50.00%	100.00%	50.00%	40.00%	100.00%	

Decision Tree - Performance

Accuracy	92.8%	± 4.4%
Classification Error	7.2%	± 4.4%

Confusion Matrix

	true Bajo	true Medio	true Alto	true Muy Alto	true Muy Bajo	class precision
pred. Bajo	3	1	0	0	0	75.00%
pred. Medio	0	6	0	0	0	100.00%
pred. Alto	0	4	5	2	0	45.45%
pred. Muy Alto	0	0	0	3	0	100.00%
pred. Muy Bajo	2	0	0	0	99	98.02%
class recall	60.00%	54.55%	100.00%	60.00%	100.00%	