

# ANÁLISIS DE UN MODELO DE **PREDICCIÓN** DE **GRAVEDAD** DE ACCIDENTES VIALES

Yenny Betancur Torres  
José Pinto Rodríguez  
Gustavo Flórez Ortiz  
Edgar Mauricio Flórez



# INTRODUCCIÓN

Analizamos el comportamiento de los siniestros viales registrados en los informes policiales de **accidentes** de tránsito ocurridos en la ciudad de Bogotá durante el año 2018, proporcionados por la página de Datos Abiertos Colombia, con el fin de crear un modelo que logre predecir el tipo de **gravedad** a partir de las diferentes variables seleccionadas, utilizando las técnicas y conceptos de machine learning

# CRISP-DM

Se aplicó la metodología CRISP-DM teniendo en cuenta una comparación entre los resultados obtenidos al aplicar un subconjunto de técnicas algorítmicas de clasificación mediante el uso de las librerías disponibles en Python en el entorno de ejecución Google Colab



# COMPRENSIÓN Y EXPLORACIÓN DE DATOS

- Sobre los datos obtenidos, se realizó inicialmente una exploración de las variables existentes y sus tipos de datos
- Se crearon tablas de frecuencia con el fin de realizar una exploración inicial de los datos

Codigo	Nombre	Cantidad
1	Choque	31,601
2	Atropello	3,759
3	Volcamiento	537
4	Caída de ocupante	920
5	Incendio	6
6	Otro	130
	<b>TOTAL</b>	<b>36,953</b>

# PREPARACION DE LOS DATOS

- 1 Se decidió trabajar solo con accidentes tipo choque (31601 registros, más del 95% de la totalidad)
- 2 Se eliminaron variables redundantes
- 3 Se eliminaron variables que no aportan información
- 4 Se enfocó análisis en Localidad
- 5 Se eliminaron datos atípicos
- 6 Encodificación de campos
- 7 Análisis de correlaciones



# SELECCIÓN DE MODELO

Se toma como variable objetivo GravedadCod y se establece como tamaño de los conjuntos de prueba y entrenamiento el 30% del total de datos. Se validaron los mejores modelos para clasificación para realizar el entrenamiento y validar su comportamiento, por lo cual ingresarán en nuestra validación:

- XGBoost
- DecisionTreeClassifier
- Naive\_bayes
- LinearDiscriminantAnalysis
- RandomForestClassifier

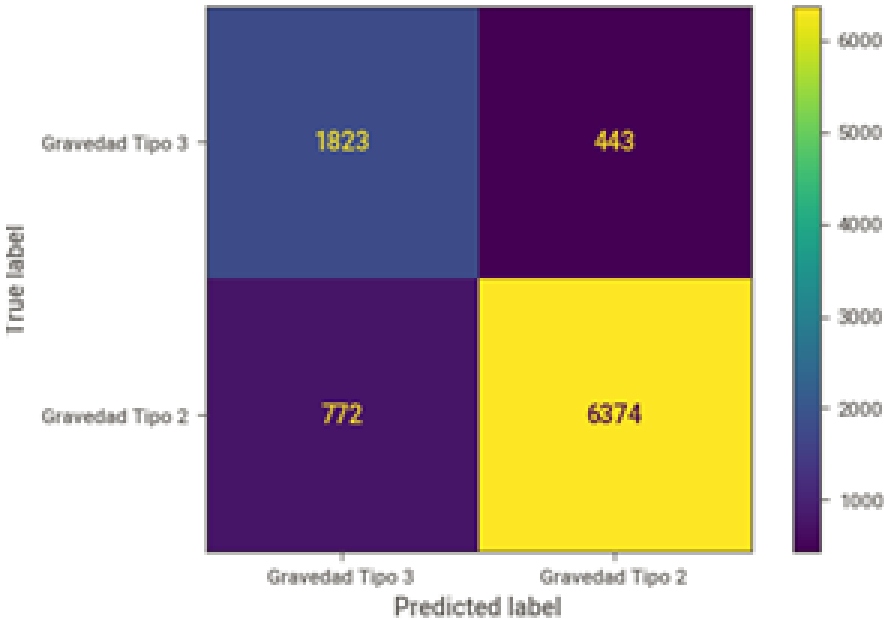


# SELECCIÓN DE MODELO: RESULTADOS

MODELO	MEDIA	DESVIACIÓN
XGBoost	0.835	0.002
DecisionTreeClassifier	0.836	0.003
Naive_bayes	0.805	0.005
LinearDiscriminantAnalysis	0.826	0.002
RandomForestClassifier	0.830	0.002

# SELECCIÓN DE MODELO: MATRIZ DE CONFUSIÓN

MODELO	VALOR	PRECISIÓN	RECALL	F1-SCORE	SUPPORT
Naive Bayes	Gravedad 3	0.7	0.81	0.75	2,266
	Gravedad 2	0.94	0.89	0.91	7,146
	Accuracy			0.87	9,412
	Macro Avg	0.82	0.85	0.83	9,412
	Weighted Avg	0.88	0.87	0.87	9,412
LDA	Gravedad 3	0.70	0.81	0.75	2,266
	Gravedad 2	0.94	0.89	0.91	7,146
	Accuracy			0.87	9,412
	Macro Avg	0.82	0.85	0.83	9,412
	Weighted Avg	0.88	0.87	0.87	9,412





# CONCLUSIONES

- El conocimiento del dataset permite entender las variables más importantes que pueden apoyar el desarrollo del modelo, pero también las que realmente no aportan o simplemente generan ruido dentro de los datos
- Es importante la exploración de los datos ya que podemos encontrar muchos registros/datos atípicos que logran sesgar la información
- La encodificación de los datos permite un mejor control y para los modelos permite un mejor procesamiento
- El balanceo de datos permite mejorar la distribución y clasificación de los datos
- La optimización y mejoramiento de los hiperparametros logran ajustar los modelos y reducen posibles sobre ajustes

# BIBLIOGRAFÍA

- Lorena Fernanda Sacristán Celis, "Análisis estadístico sobre las influencias del entorno en los accidentes de tránsito en la ciudad de Bogotá," 2016. Accessed: May 13, 2022. [Online]. Available: <https://repositorio.uniandes.edu.co/bitstream/handle/1992/18957/u729492.pdf?sequence=1>
- J. S. Edwin Urbano, "Modelo de predicción de gravedad de accidentes de tránsito: un análisis de los siniestros en Bogotá, Colombia," Bogotá, 2018. Accessed: May 13, 2022. [Online]. Available: [https://vicisev.institutoivia.org/wp-content/uploads/2018/11/edwin-urbano-CISEV-Articulo\\_gravedad\\_accidentes-VRFNL.pdf](https://vicisev.institutoivia.org/wp-content/uploads/2018/11/edwin-urbano-CISEV-Articulo_gravedad_accidentes-VRFNL.pdf)
- NICK HOTZ, "What is the Data Science Process? - Data Science Process Alliance." <https://www.datascience-pm.com/data-science-process/> (accessed May 13, 2022).