

Resumen

Últimamente los bumanguenses han manifestado por diferentes medios, como las redes sociales, que la delincuencia ha aumentado casi por toda la ciudad y expresan que el servicio policial es incompetente. Es por ello, que el siguiente proyecto desarrolla un estudio para indagar cuáles son los sectores en los que se puede considerar son los más peligrosos de la ciudad, en base a la cantidad de delitos infringidos y el número de habitantes en el respectivo sector. Dicho lo anterior, se aplicó el aprendizaje no supervisado en machine learning, de modo que, se hizo un análisis sobre la estructura comúnmente oculta de los datos.

Introducción

La intensión general por la cual se realiza este proyecto es el fenómeno de la inseguridad que se ha presenciado en los últimos años en la ciudad de Bucaramanga. Donde los mismos habitantes manifiestan el aumento de diferentes delitos en varios sectores de la ciudad, como hurtos, homicidios, extorsiones, acoso sexual, lesiones personales, riñas, entre otros. Además, este proyecto también se hace con el fin de probar e implementar los conocimientos aprendidos a lo largo del curso de y de paso fomentar la importancia del uso de la inteligencia artificial en varios ámbitos sociales, económicos, tecnológicos, de salud u otros.

Teniendo en cuenta lo anterior, se realizó la consulta de un dataset que contiene la información de cada delito que se presentó y denunció en Bucaramanga entre las fechas de enero del año 2010 hasta el mes de julio del año en curso. El dataset cuenta con aproximadamente 130.000 registros y 20 características para cada registro.

El enfoque principal de este proyecto es el aprendizaje no supervisado, ya que gracias a este método podemos determinar puntos en común dentro de las características, lo cual es útil para predecir nuevos datos que no se presentan en el dataset.

Proceso y método

El estudio se realizó a partir de la problemática expuesta que es la inseguridad y aumento en la delincuencia en la ciudad de Bucaramanga. Para empezar, se hicieron arreglos al dataset debido a que algunos valores estaban mal escritos. Se debe tener en cuenta la siguiente fórmula para determinar la tasa de delincuencia de algún lugar por su cantidad de habitantes y su número de delitos:

$$\frac{\#Delitos}{Total\ población} \cdot 100 = Tasa\ del\ delito$$

- El enfoque del estudio es determinar con el aprendizaje no supervisado en machine learning, las zonas o sectores más inseguros de Bucaramanga, que son datos específicos que no están en el dataset, realizado en lenguaje python, con gráficos referentes a los datos del dataset e información relevante como lo son los delitos y las zonas registradas en ese tipo de delito.
- El estudio realizado es para los habitantes de la ciudad de Bucaramanga.
- EL análisis estadístico realizado en el proyecto se realizó a partir de los datos arrojados por el dataset, como la cantidad de delitos realizados según el día del mes y año; se analizó la cantidad de delitos por barrios de la ciudad de Bucaramanga, se estudiaron los tipos de delitos, tipos de objetos registrados en el delito y qué tipo de personas eran las más afectadas.

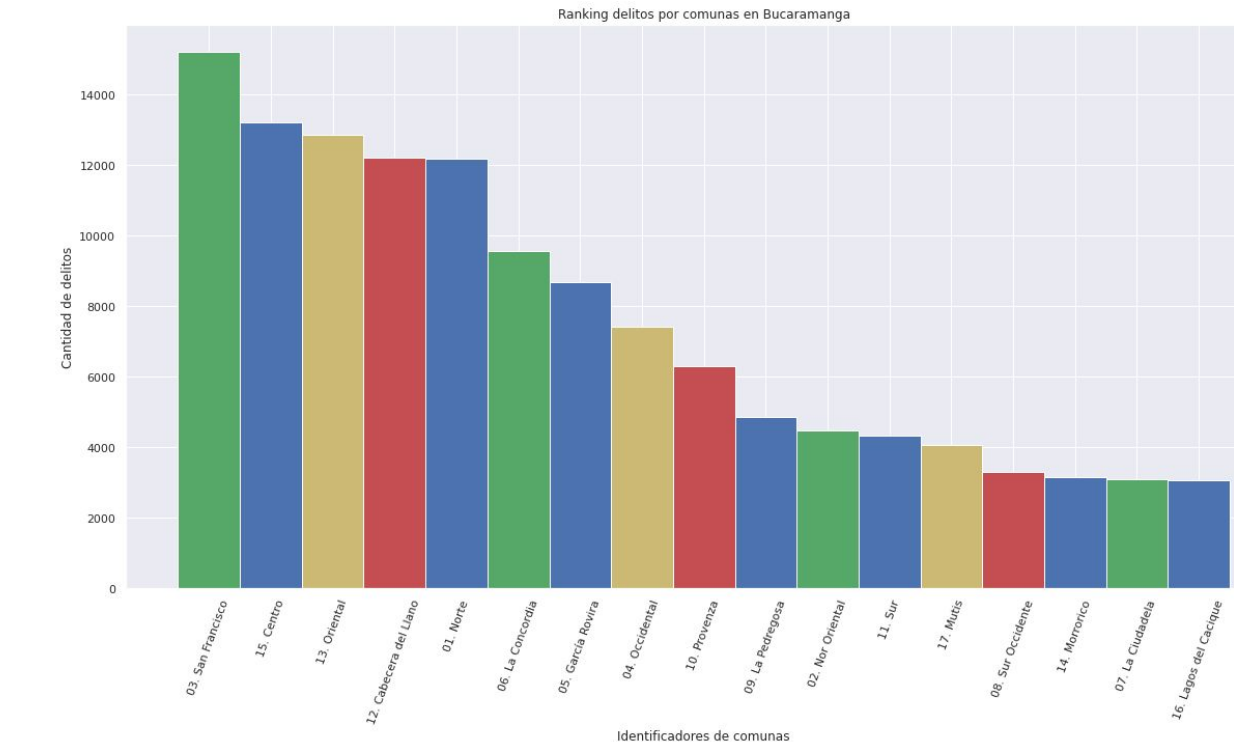


Figura 1. Ranking de delitos.

Resultados

En el estudio pudimos predecir por medio del agrupamiento del aprendizaje no supervisado, tomando en cuenta características como la cantidad de delitos y cantidad de habitantes. Se evidencia que en los sectores de menor número de habitantes, coinciden con la cantidad de delitos que se han realizado en los diferentes sectores de la ciudad.

Como primer resultado obtuvimos la cantidad de delito por día del mes, por año y un porcentaje de delitos por cada año en la ciudad de Bucaramanga. De ahí logramos obtener los valores individuales de cada columna, como por ejemplo obtener el nombre de las 17 columnas de la ciudad de Bucaramanga, tipo de arma implementada en el delito y conductas (nombre de cada delito), con el fin de poder etiquetar cada valor único para su posterior procesamiento de machine learning.

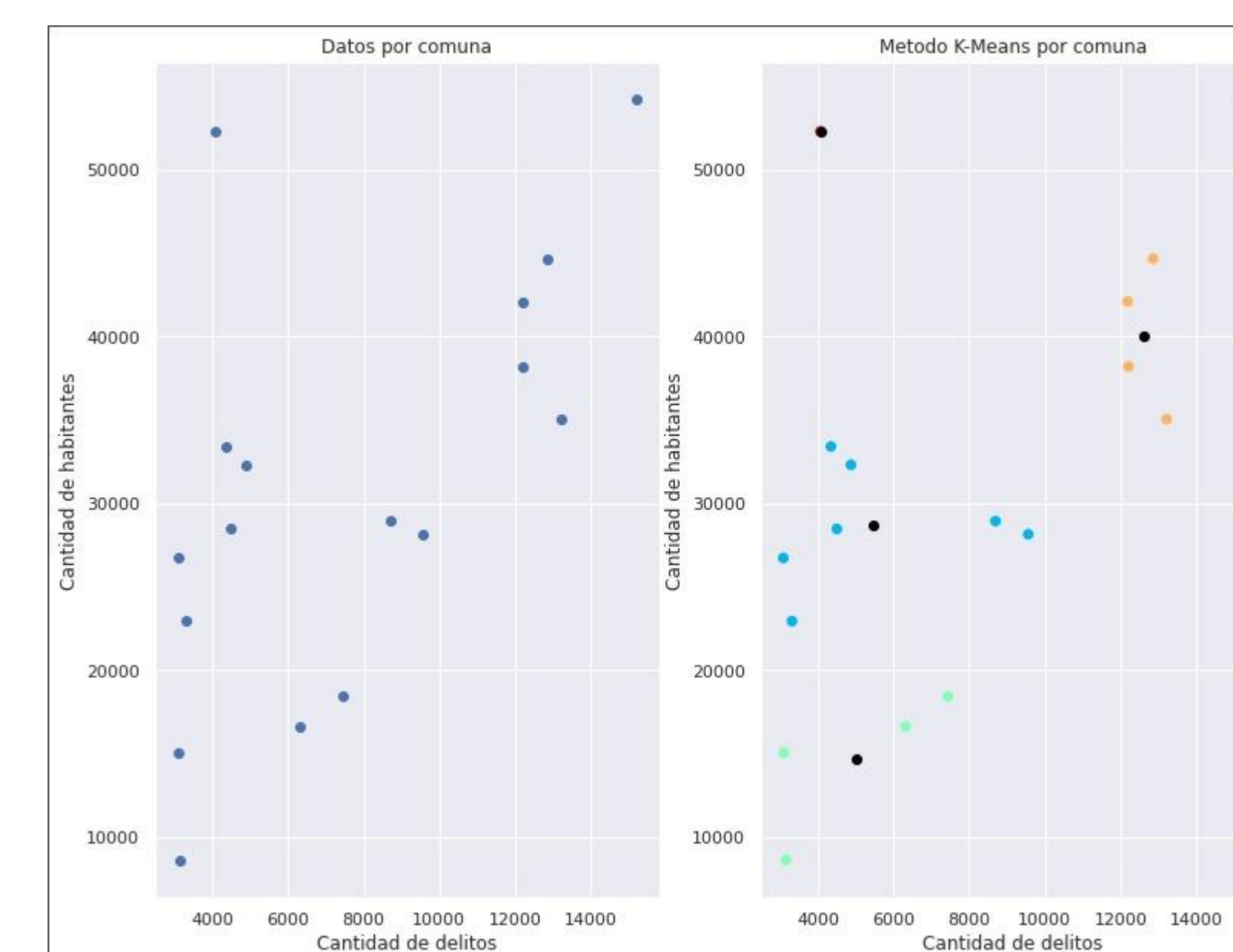
Como podemos observar en la Figura 2. al usar la fórmula para determinar la tasa de delitos en cada comuna, los resultados obtenidos demuestran que la comuna 04.Occidental es la más peligrosa en la ciudad, mientras que la comuna 03.San francisco es la que cuenta con mayor número de delitos a lo largo de más de 10 años (ver Figura 1).

En la Figura 3. observamos la relación que existe en cada comuna entre la cantidad de delitos totales y su respectiva cantidad de habitantes hasta la fecha de hoy utilizando el método de K-means.

Figura 2. Tasa de delitos por comuna

NOI_COMUNAS	N_DELITOS	POB_COMUNAS	TASA_DELITOS
04. Occidental	7430	18411	40.356309
10. Provenza	6316	16623	37.995548
15. Centro	13222	35035	37.739403
14. Moricono	3152	8627	36.536455
06. La Concordia	9561	28143	33.972924
12. Cabecera del Llano	12215	38191	31.983975
05. García Rovira	8696	28933	30.055646
01. Norte	12193	42092	28.967500
13. Oriental	12864	44662	28.803009
03. San Francisco	15219	54183	28.088146
07. La Ciudadela	3092	15045	20.551678
02. Nor Oriental	4494	28459	15.791138
09. La Pedregosa	4868	32299	15.071674
08. Sur Occidente	3307	22936	14.418382
11. Sur	4339	33404	12.989462
16. Lagos del Cacique	3077	26723	11.514426
17. Mutis	4063	52290	7.770128

Figura 3. Aprendizaje no supervisado.



Conclusiones

En definitiva, según la tasa del delito la comuna más peligrosa al día de hoy es la comuna número 4 Occidental y la comuna número 17 Mutis es la más segura de todas en la ciudad. Sin embargo, la comuna 03. San Francisco en la comuna que más delitos ha presentado en los últimos 10 años, por otra parte la comuna 16 Lagos del cacique es la que menos delitos a presentado en los últimos años.

Se determinó que la primera y la última semana del mes, son los días donde hay más frecuencia de delitos, esto es debido por motivos de pago salarial al final de mes de la mayoría de habitantes de la ciudad.

También podemos concluir que en los sectores o comunas donde hay más cantidad de habitantes, es donde se presentan mayor cantidad de delitos. Con esto podemos concluir que la cantidad de delitos incrementa según la cantidad de habitantes del sector.

Trabajo Futuro

Con este proyecto podemos realizar un plan de seguridad para evitar que más casos de delitos se repitan, enfocándose en los sectores que más se presentan.

Utilizando regresión lineal podemos estudiar los casos diarios, interesándose más por el comportamiento de los delitos para así generar alertas en las comunas más peligrosas.

Se puede reunir con los responsables de la seguridad en la ciudad para indicarles los focos más preocupantes en la ciudad de Bucaramanga con el fin de que se tomen cartas en el asunto.

Información de contacto

Juan Felipe Vásquez, juan2172013@correo.uis.edu.co
Marly Maldonado Suárez, marly2172712@correo.uis.edu.co
María Fernanda Rodríguez Pulido, maria2172711@correo.uis.edu.co
Docente:
Fabio Martínez Carrillo, famarcas@saber.uis.edu.co

Referencias Bibliográficas

- kmeans. (s. f.). Unioviado. Recuperado 14 de octubre de 2021, de https://www.unioviado.es/comprum/laboratorios_py/kmeans/kmeans.html
- Google. (s. f.). Google Colaboratory. <https://colab.research.google.com/notebooks/welcome.ipynb?hl=es-419#scrollTo=-Rh3-Vt9Nev9>
- Diagnóstico asistido por computadora - Computer-aided diagnosis - qaz.wiki. (2020, 7 diciembre). Wikipedia. Recuperado 28 de enero de 2021, de https://es.qaz.wiki/wiki/Computer-aided_diagnosis#Methodology
- Frank, M. (2000). Engineering systems thinking and systems thinking. Systems Engineering, 3(3), 163–168. <https://core.ac.uk/download/pdf/266980165.pdf>
- N. (s. f.). Soluciones de deep learning. Nvidia. Recuperado 16 de abril de 2021, de https://www.nvidia.com/es-la/deep-learning-ai/solutions/?ncid=afm-chs-44270&ranMID=44270&ranEAID=a1LgFw09t88&ranSiteID=a1LgFw09t88-xYag_UQsxFhpVpkdH3S.fq