Propuesta de Proyecto

Seminario de Analítica y Ciencia de Datos

Presentación del Proyecto

Medellín en los últimos años ha ganado una gran relevancia internacional, dada a distinciones como: ciudad más innovadora en el 2013 ([Medellín Cómo Vamos, 2013](https://www.medellincomovamos.org/node/25121)). Además la ciudad ha recibido una gran cantidad de turistas en los últimos años, por ejemplo, en el año 2022 ingresaron 49% más pasajeros a Medellín en comparación con el 2019 y la ocupación hotelera fue del 77%, la más alta de la historia hasta el momento ([Secretaría Desarrollo Económico, 2023](https://www.medellin.gov.co/es/sala-de-prensa/noticias/con-una-cifra-historica-de-pasajeros-que-visitaron-medellin-en-2022-la-ciudad-vive-el-mejor-momento-turistico-de-los-ultimos-anos/)). Una de las preocupaciones más relevantes tanto a nivel gubernamental como individual es la seguridad, y dentro de ésta la problemática de los robos, la cual puede ser de interés tanto para locales como para visitantes. Es importante notar que, anteriormente ya se han realizado distintos análisis del comportamiento de los robos en la ciudad, sin embargo, con este proyecto se pretende ir más allá, ya que incluye tanto una predicción del comportamiento de los robos en futuros periodos; inclusión de información sobre hurtos a establecimientos comerciales y hurtos de motos en el conjunto de datos; también se pretenden realizar análisis de tendencias, patrones en el tiempo y la presentación de una herramienta de visualización para que las personas del común, empresarios y agentes gubernamentales tengan acceso a información geográfica y descriptiva clara y fácil de interpretar, para que estos la puedan tomar en cuenta en su vida cotidiana, planes de negocio y planes de desarrollo territorial.

Objetivo del Proyecto

Desarrollar un pronóstico y análisis del comportamiento de los hurtos a personas, hurtos a establecimientos comerciales y hurtos de motos en la ciudad de Medellín, de tal forma que se obtenga un producto visible y útil para el público en general y los distintos grupos de interés, que posibilite la comprensión intuitiva de las tendencias y distribución geográfica de los hurtos en la ciudad.

Contexto del Problema

Durante el periodo comprendido entre 2016 y 2019, ha habido un crecimiento de las denuncias por hurto a personas, comenzando con 13.326 en el 2016, alcanzando un pico de 26.700 en 2019 y reduciéndose considerablemente en el 2020. En cuanto al hurto a establecimientos comerciales, este tuvo un crecimiento significativo entre el 2016 (1.831) a 2017 (4.386) y aunque tuvo una reducción relativamente pequeña en el 2018 y el 2020, aumentó el valor que se declara en las denuncias. Seguidamente, el comportamiento de los robos de motocicletas fue relativamente variado, pero con una tendencia a la baja ([Medellín Cómo Vamos, 2021](https://www.medellincomovamos.org/system/files/2021-09/docuprivados/Seguridad%20Informe%20de%20Calidad%20de%20Vida%20de%20Medell%C3%ADn%2C%202020.pdf)). Luego, según [Medellín Cómo Vamos (2022)](https://www.medellincomovamos.org/hurtos-en-medellin) en agosto del 2022 se presentó la cifra más alta en los registros de denuncias de hurto a personas, sin embargo, es necesario que se tome en cuenta a su vez el crecimiento de la población: en cuanto a la tasa de hurtos (denuncias por cada 100 mil habitantes) el mayor valor se dió en el 2019 con 665. Además, en el comunicado se presenta la siguiente pregunta: ¿La mayor cantidad de denuncias tiene que ver con el cambio de mecanismos de denuncias o tiene que ver con la victimización, es decir, la percepción de los ciudadanos sobre la inseguridad? En donde entra en relevancia que en los últimos años (2019 a 2021) ambos indicadores (índice de victimización y denuncias por hurto) han aumentado.

En una noticia del periodico el [Colombiano (2023)](https://www.elcolombiano.com/medellin/medellin-es-la-segunda-ciudad-del-pais-en-hurtos-tambien-han-crecido-extorsiones-y-desplazamiento-intraurbano-JH22075804), donde se narran las declaraciones del personero de Bogotá y presidente de Persocapitales, Julián Enrique Pinilla, éste comenta que en Colombia el 80% de los hurtos sucede en sus cinco principales ciudades: Bogotá ocupa el primer lugar en denuncias al día (125), seguido por Medellín (85), Cali (70), Barranquilla (38) y Bucaramanga (25). En suma, el personero menciona que la forma de contrarrestar la inseguridad debe tomar en cuenta tanto el incremento de la fuerza pública como del uso de tecnología, la generación de un entorno seguro y una evaluación constante de las políticas públicas dada la naturaleza cambiante del fenómeno.

Tomando en cuenta lo anterior, aunque han habido ciertas mejoras en los indicadores, las cifras demuestran que la inseguridad sigue siendo una problemática importante en la ciudad y que ésta es concerniente a todos los miembros de la sociedad. En cuanto a las posibles soluciones, existen distintas formas de abordarlas, ya sea desde un enfoque reactivo como desde uno proactivo. Desde esta última perspectiva, es relevante que los ciudadanos y los hacedores de política tengan información clara y concisa que les sea útil para la toma de decisiones y para llevar un accionar desde la prevención y la preparación.

Propuesta de Solución y Alcance

Para conocer a detalle y brindar una herramienta que ayude en la problemática de los hurtos en Medellín, se propone desarrollar un sistema interactivo de visualización de datos que permita entender el comportamiento histórico y futuro de los hurtos a personas, establecimientos comerciales y de motos. La solución combinará técnicas de machine learning y visualización de datos para predecir y analizar la ocurrencia de hurtos en diferentes zonas de la ciudad, y estos resultados se presentarán de forma accesible y comprensible.

El alcance de la solución se abordará de la siguiente forma:

1. **Recopilación y Preparación de Datos:** Se utilizarán datos históricos provenientes de Datos Abiertos, página oficial del gobierno nacional para la publicación de datos de interés para la comunidad, con un rango de tiempo de 2017 a 2022. Estos datos incluyen información detallada como: Lugar del hurto, características del afectado, fecha, tipo de hurto, entre otras. Se asegurará la limpieza y estandarización de los datos, identificando posibles valores erróneos o faltantes y transformando variables necesarias para posibilitar su análisis y uso para generar predicciones.
2. **Análisis Descriptivo y Espacial:** Es importante realizar un análisis descriptivo para encontrar patrones históricos en los hurtos, explorando correlaciones entre las variables que los datos nos proveen y entender un poco más la información disponible en términos de visualizaciones de mapas de calor y gráficos de tendencia temporal. Esto ayudará a la comprensión de la distribución geográfica de los hurtos y el cómo estos han evolucionado a lo largo del tiempo. Este análisis proporcionará información necesaria para mejorar la interpretación de los resultados predictivos.
3. **Desarrollo del Modelo Predictivo:** Se utilizarán algoritmos de machine learning como regresión logística, bosques aleatorios y redes neuronales artificiales. Después, se seleccionará el modelo de predicción o clasificación más pertinente para la problemática y los datos, para concluir evaluando su rendimiento y utilidad con ayuda de métricas de precisión.
4. **Despliegue de visualización de datos Interactiva:** Los resultados obtenidos se integrarán en visualizaciones interactivas accesibles a través de un dashboard que permitirá a los usuarios explorar las tendencias y predicción de los hurtos en los diferentes barrios de Medellín. Esto posibilita realizar un análisis de una manera ágil y aportar una visión holística de la problemática para tomar planes de acción para su mitigación.

Definición del Alcance

**Logros que el proyecto pretende:**

* Identificar patrones temporales y espaciales de la frecuencia de hurtos en la ciudad de Medellín que posibiliten el análisis de los mismos.
* Predecir el comportamiento de los hurtos en futuros meses utilizando técnicas de machine learning.
* Proveer visualizaciones que ayuden a comprender el comportamiento de los hurtos junto con las predicciones realizadas, permitiendo una toma de decisiones informada.

**Limitaciones del proyecto:**

El proyecto contará con los análisis y predicciones que sean posibles de realizar con los datos históricos disponibles. La calidad y precisión del modelo dependerá exclusivamente de la calidad y cantidad de los datos proporcionados para los años de 2017 a 2022.

Los resultados del proyecto están pensados para ser una herramienta de análisis y para la toma de decisiones que puedan impactar la problemática, sin embargo, no pretende ser una herramienta de monitoreo de los hurtos ni para la realización de alertas en tiempo real.

Riesgos e Impacto del Negocio

**Impacto Positivo**

Los resultados del proyecto permitirán una comprensión más profunda de la problemática basándose en datos reales de la ciudad de Medellín. Esto facilitará la toma de decisiones informadas para residentes, turistas, empresarios y autoridades.

Este enfoque ha sido exitoso en otros contextos. Un ejemplo es el proyecto realizado en la ciudad de Bogotá por Pedro Bernal, quien logró mediante modelos predictivos brindar una herramienta a la policía nacional para tener un referente de la cantidad de delitos que se pueden llegar a presentar en la semana próxima. Esta herramienta es un recurso valioso para la planificación operativa y la asignación de recursos (Bernal-Urrutia, 2023).

De manera similar, Carlos Agredo desarrolló un proyecto en donde implementa un interfaz para visualizar la información sobre la problemática de delincuencia. Este, utilizó mapas que describen el movimiento de la criminalidad en la ciudad de Bogotá junto con la probabilidad de robo a partir de variables que el usuario puede proveer (Campo Agredo, 2017).

Se pretende que el resultado del proyecto brinde información accesible y útil para un amplio espectro de usuarios, contribuyendo a fomentar un accionar desde la prevención y preparación.

**Riesgos Potenciales:**

* **Calidad de Datos:** hay una alta dependencia a los datos disponibles, por ejemplo, posibilidad de datos incompletos o incorrectos que pueden afectar negativamente los resultados.
* **Ajuste de los modelos:** En este caso, y en concordancia con lo anterior, la capacidad de ajuste y la calidad de los modelos depende de lo que los datos proveen.
* **Adopción de la Herramienta de visualización**: La utilidad de las visualizaciones depende de su aceptación por parte de los usuarios finales, lo cual puede verse afectado por su diseño, facilidad de uso y capacidad de captar el interés de la población en el tiempo.

**Referencias**

Bernal-Urrutia, P. A. (2023). Modelos estocásticos para la predicción y visualización espacial del delito de hurto a personas en la modalidad de atraco en la localidad de Kennedy (Bogotá D.C, Colombia). Recuperado de: <http://hdl.handle.net/20.500.12010/31909>.

Campo Agredo, C. (2017). Desarrollo de modelos de predicción sobre el hurto a personas en Bogotá a partir del análisis y aplicación de algoritmos en Base de datos. Universidad de los Andes. Disponible en: <http://hdl.handle.net/1992/50785>

Rúbricas

R01: Completitud y consistencia: el estudiante realiza una entrega del momento evaluativo con todos los elementos desarrollados, manteniendo una relación coherente entre los elementos solicitados. El estudiante usa una consistencia argumentativa que incluye el uso de referencias cuando hay mérito de autoría entre otros.

R02: Pertinencia: el proyecto del estudiante está dentro del ámbito de la analítica y la ciencia de datos y es evidente en su descripción, objetivos y contexto

R03: Solución: el estudiante es capaz de relacionar herramientas tecnologico-metodológicas con el contexto de un problema de analítica y ciencia de datos.

R04: Negocio: el estudiante desarrolla teniendo en cuenta un problema real, de un contexto claro y con una solución factible dentro del ámbito que seleccionó.

R05: Delimitación: dado un contexto de solución, el estudiante define claramente los límites de la solución que propone y el impacto, consecuencias y riesgos asociados a su implementación.