

Análisis de Crimen en el Oriente del Valle de México usando Data Mining

Trabajo Terminal No. 2021-B106

Alumnos: Mendoza Cuellar José Oscar, *Morales Zaranda Víctor Hugo

Director: Dr. Roberto Eswart Zagal Flores

*e-mail: vmoralesz1400@alumno.ipn.mx

Resumen – El comportamiento en tiempo y espacio de crímenes en zonas urbanas como el Valle de México es tema vigente y que aún no se ha estudiado en extenso desde la perspectiva de la integración de datos abiertos, en este sentido es de principal utilidad para tomadores de decisiones en ámbito de política pública, para conocer las regiones donde los crímenes tienen una mayor concentración en el tiempo, caracterizando la frecuencia de ocurrencia y ubicación, con el objetivo de que estos generen nuevas estrategias en apoyo a la comunidad, así como para los habitantes del Valle de México interesados en mantenerse informados acerca del índice de delincuencia en su localidad y como este evoluciona a través del tiempo. El oriente del Valle de México es una región con una gran densidad poblacional en el país, como son las alcaldías y municipios de Nezahualcóyotl, Ecatepec, Gustavo A. Madero y Azcapotzalco. En este trabajo se propone el análisis de crimen en dicha región del valle de México desde una perspectiva espacio temporal, considerando análisis de concentración de crímenes usando minería de datos y Sistemas de Información Geográfica. El objetivo es descubrir y comprender cómo se comporta la concentración y densidad de crímenes de alto impacto en estos municipios y alcaldías, para mostrar patrones tanto geográficos como temporales.

Palabras clave: Hot Spots, Sistemas de Información Geográfica, Delitos de alto impacto, Minería de datos espaciales.

1. Introducción

La violencia que sufre actualmente la población mexicana se traduce en un problema de seguridad pública, tanto por la dimensión que ha adquirido la muerte por dichas causas, como por los efectos materiales y emocionales que ocasiona, y cuyo origen se encuentra en factores históricos, demográficos, psicológicos, económicos, biológicos, sociales, entre otros. [1] Es sustancial mencionar que México es un país de altos contrastes y que nuestras raíces desde la colonización española han tenido cierta presencia de sufrimiento que después de varias luchas sociales iniciaron un proceso de transformación nacional, como lo es la lucha contra el crimen y la violencia que se presenta en nuestro día a día.

La delincuencia posee componentes geográficos inigualables ya que la mayor parte de los delitos ocurren en lugares concretos y los perpetran personas que viven o viajan a dicho lugar. [2] Si tenemos en cuenta este componente, el hecho delictivo puede ser interpretado de mejor manera y con más facilidad, partiendo del supuesto teórico de que los fenómenos sociales son dependientes del espacio donde suceden.

Cabe vislumbrar que analizar de manera precisa y oportuna la información criminal es un proceso clave para comprender los factores que impulsan las actividades delictivas, así como intuir posibles formas de operación de diversos grupos criminales o encontrar patrones que identifiquen el funcionamiento criminal en diversas zonas.

Como ya se ha mencionado, la distribución espacial de delitos puede proveer información muy valiosa, existe una gran cantidad de técnicas para el análisis de datos espaciales y geográficos, la técnica de hotspots (puntos críticos) describe áreas que tienen un alto índice de intensidad delictiva, los cuales suelen visualizarse mediante mapas, lo que facilita el análisis de áreas geográficas con relación a la delincuencia. Estas herramientas se están convirtiendo en ayuda fundamental para la vigilancia y para la comprensión y conocimiento de las diferentes zonas de densidad donde es más probable que ocurran crímenes.

De acuerdo con la Procuraduría General de Justicia de la CDMX y la Secretaría de Seguridad Pública de la CDMX, las alcaldías con mayores índices de crímenes reportados desde 2014 son: Iztapalapa, Cuauhtémoc, Gustavo A. Madero, Benito Juárez, Coyoacán y Tlalpan. [3] Debido a la gran segmentación de estas zonas y el tamaño de área que abarcan, los informes policiales suelen subestimar el crimen de manera sustancial, por lo que pueden llegar a ser engañosos. Por estos motivos la delincuencia es muy difícil de medir y los datos obtenidos deben tratarse con precaución.

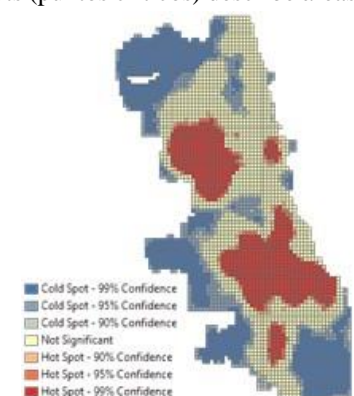


Fig. 1 Ejemplo de análisis de concentración de delitos

Asimismo, en el Estado de México, tenemos a Ecatepec con sus casi 1.7 millones de habitantes siendo el municipio más poblado del país, por lo que no sorprende que también sea el municipio más violento para las mujeres, registrando un 17% de los feminicidios en dicha entidad. [4] Según informa la Dirección General de Seguridad Ciudadana de Nezahualcóyotl una gran parte de los homicidios registrados se vincula con el robo de vehículo con violencia, pues el 17% de los 805 homicidios captados entre 2013 y 2018 (143) están vinculados con este delito. [5] Finalmente la extorsión aumento un 105% en la alcaldía Gustavo A. Madero entre enero y septiembre de 2019 con respecto a 2018 y el robo a negocio con violencia también creció en un 64% en relación con el mismo lapso. [6]

Se han estudiado 2 trabajos relacionados con el presente plantado, el primero es usado para nutrir el sistema de justicia de Estados Unidos, analizar datos del crimen y encontrar comportamientos que a simple vista o con un análisis superficial no se pueden encontrar, siendo necesario hacer uso de técnicas de minería de datos y de análisis espacial con el fin de visualizar la información recopilada. El otro es un trabajo desarrollado por la Universidad Autónoma del Estado de México en el que de igual forma se hizo uso de estas técnicas para representar los hallazgos encontrados en la incidencia del delito callejero en la Ciudad de México.

Nombre	Características	Lugar de desarrollo	Año
CrimeStat: Spatial Statistics Program for the Analysis of Crime Incident Locations [7]	Un sistema de software de estadísticas espaciales para el análisis de la ubicación de los incidentes delictivos. Interactúa con la mayoría de los programas GIS de escritorio. Incluye más de 100 rutinas estadísticas para el análisis espacial del crimen.	Houston Texas, National Institute of Justice, Agency of the United States Department of Justice.	2015
Spatial analysis of street crime in Mexico City [8]	Trabajo de investigación de patrones espaciotemporales de incidencia del delito callejero en la Ciudad de México durante 2018. Hacen uso de técnicas como Hot Spots.	México, Universidad Autónoma Del Estado De México	2018

Tabla 1 Comparativa de trabajos relacionados

Concluyendo con lo ya expuesto, tenemos que la zona oriente del Valle de México cuenta con un gran índice delictivo, pero a su vez se tiene información que puede ser de ayuda si esta se analiza de manera correcta para encontrar patrones o identificar zonas de riesgo (alta concentración de crímenes). Lo que busca el presente trabajo terminal es desarrollar este análisis con el fin de generar conocimientos y herramientas de consulta general como ya ocurre en países como Estados Unidos con CrimeStat. [7] Además de generar literatura cerca de este problema que tanto agobia a la sociedad actual.

2. Objetivo

Desarrollar un prototipo de software para procesar datos de delitos de alto impacto del oriente del Valle de México (Nezahualcóyotl, Ecatepec, Gustavo A. Madero y Azcapotzalco) para identificar zonas de concentración de crímenes, patrones de comportamiento, así como tendencias en tiempo y en espacio, empleando técnicas data mining y de sistemas de información geográfica.

2.1. Objetivos específicos

- Seleccionar y comprender las fuentes de datos de crímenes en las zonas de estudio
- Definir una metodología de data mining para caracterizar en el tiempo y el espacio los conjuntos de datos de crimen de las zonas de estudio
- Acoplar a la metodología mecanismos de análisis espacio temporal que apliquen para el procesamiento de datos de crímenes
- Estudiar y definir posibles algoritmos específicos para el análisis de datos de crímenes
- Explorar mecanismos para identificar y visualizar zonas de concentración o densidad de crímenes de alto impacto de acuerdo con la región de estudio.
- Diseñar una arquitectura de data mining que permite el descubrimiento no trivial de conocimiento en crimen, procesamientos de datos y de visualización de resultados.

3. Justificación

El reto es definir mecanismos de integración de datos de delitos ya que las estructuras pueden cambiar según la región, pero sobre todo identificar zonas de concentración de delitos con claridad y útiles es una tarea que depende de la densidad de los diferentes tipos de delitos en ciertas zonas y periodos de tiempo esto es necesario para poder analizar el comportamiento delictivo y conocer patrones de comportamiento de delitos, la complejidad aumenta si consideramos otras fuentes de datos relacionadas como densidad de población o niveles socioeconómicos.

La principal motivación para este presente estudio es analizar el clima de violencia y crímenes en relación con las zonas de interés, las cuales consideramos que son relevantes pues son de gran afluencia no solo por los habitantes si no por gran parte de personas de la Zona Metropolitana, ya sea por cruzarlas para llegar al trabajo o centro de estudios.

Este análisis nos puede ayudar a mejorar la seguridad pública, identificar tendencias emergentes y planificar estrategias de prevención del delito. En estados unidos este análisis de zonas de concentración del delito proporciona herramientas estadísticas para ayudar a los organismos encargados de hacer cumplir la ley y a los investigadores de justicia penal en sus esfuerzos de mapeo del crimen. [7].

A lo largo de este trabajo terminal, se busca la realización de un dashboard que muestre los conocimientos adquiridos sobre la concentración de delitos en el oriente del Valle de México haciendo uso de las técnicas de minería de datos y análisis espacial, generando así una fuente de información confiable, siendo esta de acceso público para informar, concientizar y funcionar como un apoyo en la medida de lo posible a los tomadores de decisiones en el ámbito político.

4. Productos o resultados esperados

Dado que nuestra propuesta se basa en el desarrollo de un dashboard e histórico, se espera un sistema que permita llevar a cabo todas las actividades ya mencionadas en este documento, así como:

Productos esperados

- Código fuente del sistema terminado.
- Base de datos integrada con toda la información recopilada.
- Módulos operativos del sistema.
- Sistema funcional.
- Manual de usuario.
- Documentación técnica del sistema.
- Dashboard.

Arquitectura del sistema

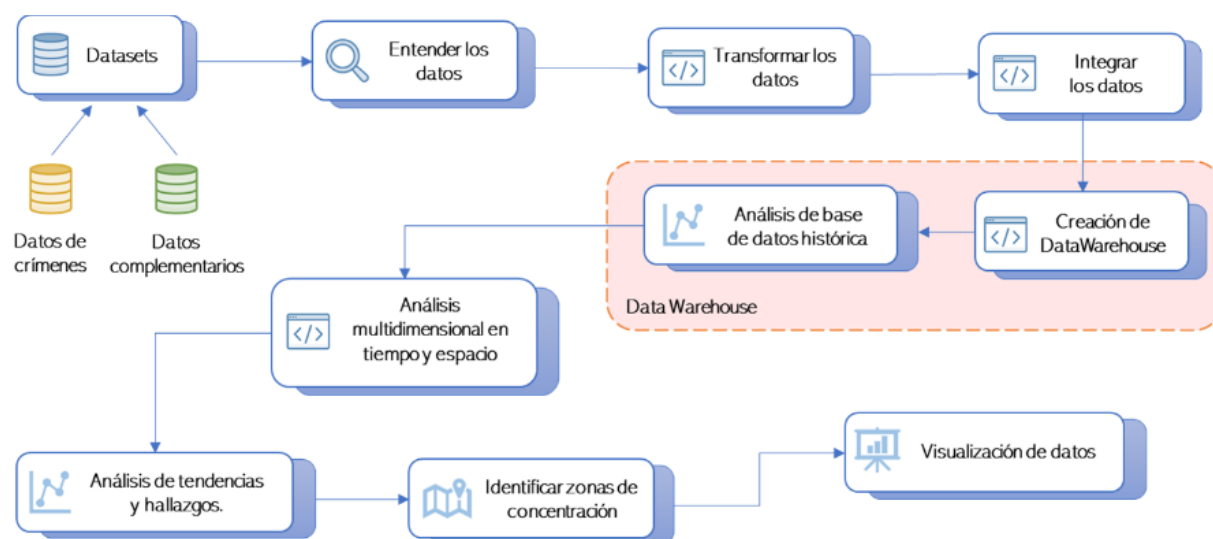


Fig. 2 Arquitectura del sistema

5. Metodología

Para la elaboración de la aplicación dashboard se pretende utilizar el modelo de desarrollo en espiral, el cual al ser del tipo evolutivo y con una naturaleza iterativa se acopla fácilmente al desarrollo de prototipos. [9] Se escogió este modelo de desarrollo debido a que uno de sus principales componentes es la comunicación entre los integrantes del equipo, así como su potencial para hacer desarrollo rápido de versiones cada vez más completas debido a la escalabilidad del prototipo base.

Con el empleo del modelo en espiral, el software se desarrolla en una serie de entregas evolutivas. Durante las primeras iteraciones, lo que se entrega puede ser un modelo o prototipo. En las iteraciones posteriores se producen versiones cada vez más completas del sistema cuya ingeniería se está haciendo.

Durante cada iteración de desarrollo se hará uso de la metodología CRISP-DM la cual es la guía de referencia más utilizada en el desarrollo de proyectos de data mining [10] y está conformada por las siguientes fases:

- **Comprensión del negocio.** En esta fase se centra en entender los objetivos y requerimientos del proyecto y se define el problema de minería de datos.
- **Estudio y comprensión de los datos.** Se identifican, colectan y analizan los data sets con los que se planea alcanzar el objetivo del proyecto.
- **Preparación de datos.** Consta de preparar los o el data set final antes del modelado, para lo que se necesitan realizar las siguientes cinco tareas a el conjunto de datos: selección de datos, limpieza de datos, construcción de datos, integración de datos y por último dar formato a los datos.
- **Modelado.** Durante la cuarta fase se seleccionan las técnicas de modelado más apropiadas para el proyecto de data mining, realizando las siguientes cuatro tareas: seleccionar la técnica de modelado, generar el diseño de prueba, construcción del modelo y la evaluación del modelo. Esta fase se repetirá hasta que se encuentre un modelo lo suficientemente bueno para continuar con el modelo CRISP-DM y en el futuro mejorar el modelado.
- **Evaluación.** En esta fase se analiza de una manera más amplia que modelo se adapta mejor al negocio y que hacer al final, se evalúan los resultados, se revisa el proceso y se determinan los siguientes pasos.
- **Despliegue.** La última fase del modelo se despliega el plan de proyecto, se monitorea y mantiene y se hace una revisión del proyecto

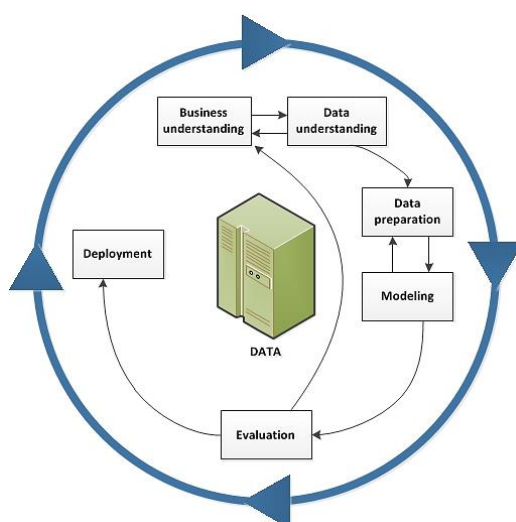


Fig. 3 Diagrama del modelo CRISP-DM

6. Cronograma de actividades

Ver anexo 1.

7. Referencias

- [1] R. A. Jiménez Ornelas, «La cifra negra de la delincuencia en México: sistema de encuestas sobre victimización.,» de *Proyectos legislativos y otros temas penales.*, Ciudad dde México, UNAM, Instituto de Investigaciones Jurídicas, 2003, pp. 167 - 190.
- [2] G. P. J. M. Dávila, «La distribucion espacial de la delincuencia en el Pais Valenciano y su relacioncon algunas variables socioeconomicas,» *Investigaciones Geograficas*, vol. 6, pp. 187-205, 1988.
- [3] C. A. P. G. & L. R. Ramirez, «Exploring crime patterns in México City,» *Journal of Big Data*, vol. 6, nº 1, 2019.
- [4] A. G. Rojas, «"Monstruo de Ecatepec": ¿por qué este municipio de México es el más peligroso para ser mujer?,» *BBC News*, 11 Octubre 2018.
- [5] E. JAIME, Hot Spot Neza 10,000 cuadradas resguardadas por vecinos, México: México Evalúa, 2020.
- [6] Hallazgos Índice GLAC, «Extorsión y robo a negocio aumentaron en Gustavo A. Madero en 2019,» *Animal Politico*, 13 Noviembre 2019. [En línea]. Available: <https://tinyurl.com/pdjnbre8>. [Último acceso: 9 Noviembre 2021].
- [7] N. Levine, «CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations,» National Institute of Justice, Washington, D.C, 2015.
- [8] F. G. Socorro, Spatial analysis of street crime in Mexico City, México: UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MÉXICO, 2018.
- [9] R. S. Pressman, Ingeniería del Software un enfoque práctico, México, D. F: McGRAW-HILL INTERAMERICANA EDITORES, S.A. DE C.V, 2010.
- [10] J. A. M. G. Gallardo Arancibia, Metodología para la definición de requisitos en proyectos de data mining, Madrid: Facultad de Informática (UPM), 2009.
- [11] E. Moto, «CityCop, una app que te ayuda a combatir la delincuencia en tu comunidad,» *QORE*, 11 Mayo 2018.

8. Alumnos y directores

CARÁCTER: Confidencial
FUNDAMENTO LEGAL: Artículo 11 Fracc. V y Artículos
108, 113 y 117 de la Ley Federal de Transparencia y Acceso
a la Información Pública.
PARTES CONFIDENCIALES: Número de boleta y teléfono.

Mendoza Cuellar José Oscar. - Alumno de la carrera de Ing. en Sistemas Computacionales en ESCOM, Especialidad Sistemas, Boleta: 2015010680, Tel. 5537099474, email: jmendozac1400@alumno.ipn.mx.

Firma:

Morales Zaranda Víctor Hugo. - Alumno de la carrera de Ing. en Sistemas Computacionales en ESCOM, Especialidad Sistemas, Boleta: 2015010727, Tel. 5623883231, email: vmoralesz1400@alumno.ipn.mx.

Firma:

Dr. Roberto Zagal Flores. Es egresado de la Ingeniería en Sistemas Computacionales de la Escuela Superior de Cómputo del IPN, culminó sus estudios de Maestría en Ciencias de la Computación en el Centro de Investigación en Computación del IPN (No. Cedula 11050111). Tiene un doctorado en tecnología avanzada de la Sección de Estudios de Posgrado de la UPIITA IPN. Actualmente es profesor de la escuela Superior de Cómputo y sus áreas de interés son Data Mining, Spatial Data Mining, GIS, Web Semántica, Data Integration, IoT y Arquitecturas de Sistemas de Información. Ha trabajado en proyectos de tecnología en la iniciativa privada y en el sector público. Email: zagalmmx@hotmail.com, Tel. 57296000, Ext. 52032.

Firma:

Anexo 1: Cronogramas Individuales

Alumno 1

Nombre del alumno: Mendoza Cuellar José Oscar

Título de TT: Análisis de crimen en el oriente en el Valle de México

No.	Actividad	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SEP	OCT	NOV	DIC
1.	Análisis y entendimiento del comportamiento de crímenes de alto impacto												
2.	Estudio del marco teórico: de análisis del comportamiento del crimen												
3.	Estudio de técnicas de minería de datos y análisis espacial												
4.	Análisis de necesidades y requisitos de software de minería de datos												
5.	Definición de preguntas de minería												
6.	Diseño de la arquitectura del sistema y base de datos												
7.	Avance en la Implementación del módulo de comunicación y dashboard (prototipo)												
8.	Presentación de los primeros hallazgos												
9.	Elaboración del documento de análisis y diseño												
10.	Evaluación TT1												
11.	Pruebas preliminares del módulo												
12.	Desarrollo del módulo de Dashboard												
13.	Pruebas preliminares del módulo												
14.	Desarrollo del módulo de exploración y analítica de datos recolectados												
15.	Integración de los módulos												
16.	Pruebas del sistema integrado												
17.	Elaboración del documento de análisis, diseño y pruebas.												
18.	Elaboración de la presentación de TT 2.												
19.	Evaluación TT2												

Alumno 2

Nombre del alumno: Morales Zaranda Víctor Hugo

Título de TT: Análisis de crimen en el oriente en el Valle de México

No.	Actividad	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SEP	OCT	NOV	DIC
20.	Análisis y entendimiento del comportamiento de crímenes de alto impacto												
21.	Estudio del marco teórico: de análisis del comportamiento del crimen												
22.	Estudio de técnicas de minería de datos y análisis espacial												
23.	Análisis de necesidades y requisitos de software de minería de datos												
24.	Diseño del comportamiento de los módulos del sistema												
25.	Definición de preguntas de minería												
26.	avance en la implementación del repositorio de datos pre procesados y del análisis exploratorio de datos												
27.	Presentación de los primeros hallazgos												
28.	Elaboración del documento de análisis y diseño												
29.	Evaluación TT1												
30.	Desarrollo del módulo de recolección de datos												
31.	Pruebas preliminares del módulo												
32.	Desarrollo del módulo de alertamiento de datos												
33.	Pruebas preliminares del módulo												
34.	Integración de los módulos												
35.	Pruebas del sistema integrado												
36.	Elaboración del documento de análisis, diseño y pruebas.												
37.	Elaboración de la presentación de TT 2.												
38.	Evaluación TT2												

Anexo 2: Acuses de aceptación por alumnos y director

- Director de Trabajo Terminal: Dr. Roberto Eswart Zagal Flores

RE: Protocolo



Roberto Eswart Zagal Flores <rzagalf@ipn.mx>

10:17 p. m.



Para: Víctor Hugo Morales Zaranda; Roberto Zagal

Estoy de acuerdo en ser el director del Trabajo Terminal
"Análisis de crimen en el oriente del Valle de México" de
los alumnos Mendoza Cuellar Jose Oscar y Morales Zaranda
Victor Hugo.

- Alumno 1: Mendoza Cuellar Jose Oscar

Protocolo



Jose Oscar Mendoza Cuellar <jmendozac1400@alumno.ipn.mx>

02:07 p. m.



Para: Víctor Hugo Morales Zaranda

Para: Morales Zaranda Victor Hugo

Soy Mendoza Cuellar José Oscar, alumno de la ESCOM-IPN
con boleta 2015010680 y estoy de acuerdo en participar en
el proyecto de TT "Análisis de crimen en el oriente del Valle
de México"

- Alumno 2: Morales Zaranda Victor Hugo

Protocolo



Victor Hugo Morales Zaranda <vmoralesz1400@alumno.ipn.mx>

02:07 p. m.



Para: Víctor Hugo Morales Zaranda

Para: Morales Zaranda Victor Hugo

Soy Morales Zaranda Victor Hugo, alumno de la ESCOM-IPN
con boleta 2015010727 y estoy de acuerdo en participar en
el proyecto de TT "Análisis de crimen en el oriente del Valle
de México"