



Universidad Internacional de la Rioja México

**Máster Universitario en Análisis y Visualización de Datos Masivos / Visual Analytics & Big
Data**

Nombre del Maestrante

Juan Salvador De la Rosa Barajas - ID 320381

Materia

Proyecto Final de Maestría

Nombre del director del Proyecto

Dr. Antonio Castañeda Solís

Título del Proyecto

Visualización Interactiva y Efectiva de la probabilidad de incidencia criminal en tramos
carreteros federales con riesgo severo.

Fecha

Agosto de 2020

Índice

Presentación General	3
Resumen	3
Diagnóstico	5
Justificación	7
Planteamiento del Problema	7
Objetivos y Metas	
Objetivo General	9
Objetivo Particular	10
Metodología	10
Fundamentación	12
Plan de trabajo y Cronograma	14
Marco Teórico.....	17
Recursos.....	17
Fuentes de Información	19
Referencias Bibliográficas	19

Resumen

La criminalidad prevalece como una severa amenaza al desarrollo de espacios urbanos vulnerando la integridad, seguridad y el libre tránsito de los ciudadanos. La analítica de datos, un proceso de descubrir información oculta a través de técnicas de Big Data es ahora una herramienta importante en la investigación, entendimiento y prevención del crimen.

Gran cantidad de esfuerzos han sido desarrollados por las fuerzas de seguridad pública en operativos para disminuir la ola de crimen, violencia, conflicto e inseguridad en los tramos carreteros Federales, pero los esfuerzos han fallado.

El eje primario del presente proyecto se basa en explorar la naturaleza dinámica de los sucesos criminales con el potencial de obtener un mejor entendimiento y avanzar en su predicción, para ello, se utilizan técnicas de analítica de datos para explorar la información criminal recolectada y clasificada en una base de datos abierta, centrando la atención en una visualización efectiva que permite integrar el modelo espacio-temporal en el que se desarrolla el análisis.

Palabras clave: Visualización de datos, Analítica de datos, Crimen, patrones espacio-temporales

Presentación General

En las zonas urbanas, los niveles municipal, estatal y federal de las fuerzas de seguridad pública necesitan ofrecer una cobertura efectiva y de rápida respuesta a sus ciudadanos. Tal hecho representa en sí mismo un enorme reto por la logística y operaciones necesarias para coordinar y cubrir las extensiones territoriales pertinentes, determinando cuales áreas requieren mayor atención de recursos y personal en función de la probabilidad de ocurrencia de sucesos de naturaleza criminal.

La necesidad de aprovechar las técnicas modernas de estadística y ciencias de la computación crean una oportunidad en los campos de la analítica de datos conocido como “Policía predictiva” o “Vigilancia predictiva”.

El concepto “Vigilancia predictiva” se basa, concretamente, en programas de análisis de gran cantidad de datos, revelando facetas, patrones o relaciones y que no se pueden detectar con los métodos tradicionales de vigilancia (Castellanos, 2019)

En México, tal actividad ya tiene sus primeros pasos a través de modelos predictivos en colaboraciones entre Microsoft y la Dirección General de Seguridad Ciudadana Nezahualcóyotl (Microsoft, 2018)

Tal colaboración ha creado un lazo estrecho entre los sistemas computarizados para seguimiento a crímenes, maximizando el uso de la información y optimizando la toma de decisiones. El análisis de datos computacionales ha comenzado a ayudar a las fuerzas de seguridad pública no sólo a acelerar el proceso de resolución de crímenes, sino que además a predecir los crímenes con antelación.

Sin embargo, las actividades criminales han dejado de estar limitadas a calles y callejones en ciudades, lo cual reitera que la necesidad de nuevos y efectivos métodos de prevención criminal es significativamente mayor fuera de los centros urbanos.

En México, existen otras zonas fuera de ciudades y centros urbanos en las que la tasa de criminalidad representa graves riesgos para el libre tránsito y para los cuales las fuerzas policiacas carecen de herramientas para el combate. Para éste proyecto, se propone analizar de zonas carreteras federales consideradas por su severa incidencia delictiva y mismo nivel de riesgo.

Las aseguradoras automotrices en México poseen niveles de evaluación respecto al riesgo que representa transitar de las principales zonas carreteras del país. Para el primer cuatrimestre de 2020 (SafeLink, 2020) las zonas mostradas en la tabla 1 representan las tasas de riesgo más altas:

Tabla 1. Tramos carreteros con mayor nivel de riesgo en México en 2020

Tramo Carretero	Nivel de Riesgo
México – Veracruz (MEX – 150D)	SEVERO
Circuito Exterior Mexiquense (CEM)	ALTO
México-Saltillo (MEX-57/D)	ALTO

Extraída de SafeLink, 2020

La carretera México – Veracruz (MEX 150D) es la única de las expuestas que posee el nivel de riesgo máximo. Es importante señalar que tal tramo carretero a su vez puede ser subdividido en regiones que permitan analizar la incidencia delictiva de una forma más puntual.

Se propone la división de 3 zonas geográficas para el tramo expuesto, abarcando de la siguiente manera:

- Ciudad de México – Acatzingo (MEX 150D),
- Acatzingo – Córdoba (MEX 150D), y
- Córdoba – Veracruz (MEX 150D).

De acuerdo con SafeLink (2020), el tramo carretero Acatzingo – Córdoba representó, en el primer cuatrimestre de 2020, el 69% de los incidentes criminales registrados en la carretera México – Veracruz (México 150D). Por lo que tal zona puede caracterizarse como el punto medular de la criminalidad carretera en el país, surgiendo de ahí la necesidad de su análisis.

Tal zona de conflictiva, a su vez, se subdivide geográficamente el tramo carretero Federal “Acatzingo - Ciudad Mendoza” y “Ciudad Mendoza – Córdoba” correspondiente a la carretera MEX 150D.

Geográficamente, la zona constituye un tramo carretero en 2 sentidos de aproximadamente 120 km de longitud para el sentido Acatzingo – Córdoba y 130km para el sentido Córdoba – Acatzingo, sumando cerca de 250 km en una vía de comunicación de gran importancia para conectar el centro del país con el Puerto de Veracruz, pasando por las ciudades de Orizaba y Córdoba, Veracruz.

Así mismo, la delimitación temporal se da por los eventos reportados entre el 1 de enero de 2016 y el 31 de diciembre de 2019.

Diagnóstico

En las zonas con alto y severo nivel de riesgo de tránsito en México existen variables que impiden a las fuerzas del orden tener respuestas rápidas y puntuales sobre los incidentes criminales ya

que diversos factores, entre los que se destacan las condiciones geográficas y de coordinación entre diferentes niveles de gobierno interestatal.

Según SensiTech (2019) México prevalece como uno de los principales puntos para el robo de autotransporte en carretera, con una tendencia al alta y un promedio de 55 crímenes por día, lo que las estrategias implementadas hasta ahora no han mostrado resultados positivos.

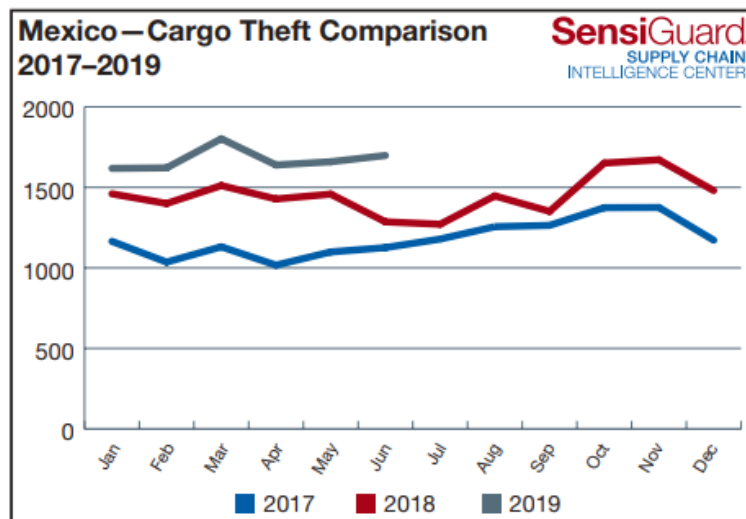


Ilustración 1. México – Comparación anual de robo al transporte de carga

Para el primer cuatrimestre de 202., la carretera federal México – Veracruz (MEX 150D) concentra el 41% de los incidentes criminales en carretera en el país. Tales actividades son realizadas por grupos criminales con extensas redes de distribución que generan violencia, inhibe el desarrollo económico regional, desalienta el uso de la infraestructura carretera federal, desata crisis sociales, afecta a los vehículos que circulan y constituye un latente foco de emergencia para las instituciones de gobierno encargadas de salvaguardar el libre tránsito.

La orografía del tramo carretero propuesto es un factor de especial interés, al encontrarse geográficamente próxima a la “Zona de las altas montañas” de Veracruz. Tales características, hacen que el terreno posea un cambio de inclinación de aproximadamente +/- 1600m en un tramo de 40 km.

Dicha inclinación favorece las actividades criminales ya que crea zonas, e las que, por el exceso de pendiente positiva o negativa los vehículos se ven forzados a disminuir su velocidad, convirtiéndolos en objetivos fáciles para interceptar. Cabe señalar que según SensiTech (2019), las bandas criminales identificadas que operan en la zona poseen un amplio conocimiento del terreno y capacidades logísticas para desplegar y replegarse fácilmente.

Así mismo, la inclinación geográfica mencionada genera grandes zonas en las que las telecomunicaciones celulares no tienen cobertura ya que las ondas radiales poseen rangos limitados por la orografía y carecen de potencia para cubrir de manera eficiente todo el tramo carretero.

La suma de factores mencionados, adicionales a los sociales y políticos, generan incertidumbre en la libre circulación y resultan los vehículos de autotransporte los principales afectados al ser despojados de la mercancía/productos/bienes que desplazan. De acuerdo con el Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública (2019) los atracos en las carreteras de México suman pérdidas \$6 mil Millones de dólares entre empresas y gobiernos.

Según el delegado de la Cámara Nacional del Autotransporte de Carga (CANACAR), la principal estrategia de seguridad impuesta por las fuerzas policiacas federales división caminos se basa en el “operativo escalón”, con el cuál, miembros de la Policía Federal/Guardia Nacional escoltan en horarios establecidos a convoyes de vehículos para ser protegidos mientras transitan por las zonas de mayor riesgo. Si embargo, los horarios y zonas son conocidos públicamente, razón por la que las bandas delictivas simplemente evitan tales horarios.

De acuerdo con los reportes de SensiTech (2019), es imperativo conocer que las bandas identificadas que operan en los tramos carreteros mencionados poseen amplios conocimientos logísticos, de tecnología y el personal suficiente para evitar enfrentamientos con las fuerzas del orden y poder operar impunemente. Es por esta razón, que el dotar a los organismos encargados de combatirlos, a los transportistas, y a todo sujeto de interés sobre estadísticas e información

puntual sobre la probabilidad de criminalidad resulta no solo de importancia sino de alta prioridad.

Planteamiento del Problema

La zona conocida como el “Triángulo Rojo” en los límites geográficos entre Puebla y Veracruz, México representa una alta tasa de criminalidad la cual ha crecido de manera exponencial desde 2012 (Islas, 2017) derivado al robo de combustibles.

Tras las medidas contra el robo de combustibles impuestas por el Gobierno Federal en enero de 2019, y la subsecuente disminución dramática en los índices de criminalidad derivados a tal rubro, otros como el asalto a vehículos ha ido en aumento en el tramo de la autopista Puebla-Veracruz (Morales, 2019).

Ante tal situación, se propone el uso de la analítica de Big Data. La cuál representa una aproximación sistemática para el análisis de patrones, relaciones y tendencias en un gran volumen de datos (M. Feng et al, 2019). Bajo dicha premisa, se pretende realizar análisis sustanciales que permitan a las autoridades y a los transportistas particulares, entender mejor el problema de criminalidad y predecir la posibilidad de incidentes para el uso eficiente de recursos. Así mismo, se propone el uso de técnicas del estado del arte para visualizar los acontecimientos delictivos registrados de manera interactiva, los cuáles pueden ser explorados por el público en general, para un análisis de datos abiertos en una página web específica para tal propósito.

Pregunta de Investigación: ¿Cuáles son las características de una solución tecnológica basada en visualización de datos que incida en disminuir los índices de ocurrencia delictiva en el tramo carretero Federal Acatzingo – Córdoba (MEX 150D)?

Justificación

El desarrollo de la sociedad ha conllevado en continua y creciente urbanización, lo cual ha generado rutas que conecten tales puntos urbanos con carreteras.

Tales desarrollos también han sido acompañados por el incremento en crímenes violentos y accidentes; Para atacar tal problemática diferentes grupos de académicos e investigadores han realizado esfuerzos en el estudio del potencial de la minería de datos y el desarrollo de patrones (Hassani et al, 2016).

El análisis de tales incidentes permite tener un entendimiento del problema y, al mismo tiempo, permite identificar similitudes, manejar recursos y tomar decisiones adecuadas. El entendimiento de situaciones del pasado amplía el conocimiento de los sucesos históricos actuales y sus antecedentes.

El rápido crecimiento de la adquisición de información y su enorme almacenamiento crean una enorme nube de posibilidades para el análisis, sin embargo, se ha vuelto vital el ordenamiento, clasificación y análisis de tal información para su entendimiento.

El Análisis de estadístico entiende los retos sobre la estructura, magnitud y rapidez con que puede ser producido tal volumen de datos es por ello que, como primera aproximación, se realiza la conformación de una base de datos adecuada y su subsecuente exploración. Procesos importantes para la creación de paradigmas y técnicas en diversos campos que permitan encontrar patrones escondidos en la información (Thongsatapornwatana, 2016).

Recapitulando, en el presente proyecto, se aplicarán técnicas de visualización de datos, para realizar minería en los registros criminales reportados por el diario El Mundo de Orizaba en el tramo carretero Federal “Acatzingo - Ciudad Mendoza” y “Ciudad Mendoza – Córdoba” correspondiente a la carretera México 150D, señalado en la Ilustración 2, siendo primordial centrar la atención en recolectar, analizar, comprender e intentar predecir futuras ocurrencias de

los fenómenos para dar a las fuerzas de seguridad pública herramientas que les permitan contar con más elementos para la toma de decisiones.

Las contribuciones principales del presente proyecto serán:

- Base de datos abierta (Open Data) sobre criminalidad reportada en un periodo de tiempo de 4 años (01/01/2016 – 31/12/2019)
- Investigación explorativa para explicar los datos criminales en el tramo carretero mencionado
- Visualización efectiva e interactiva accesible para análisis y exploración

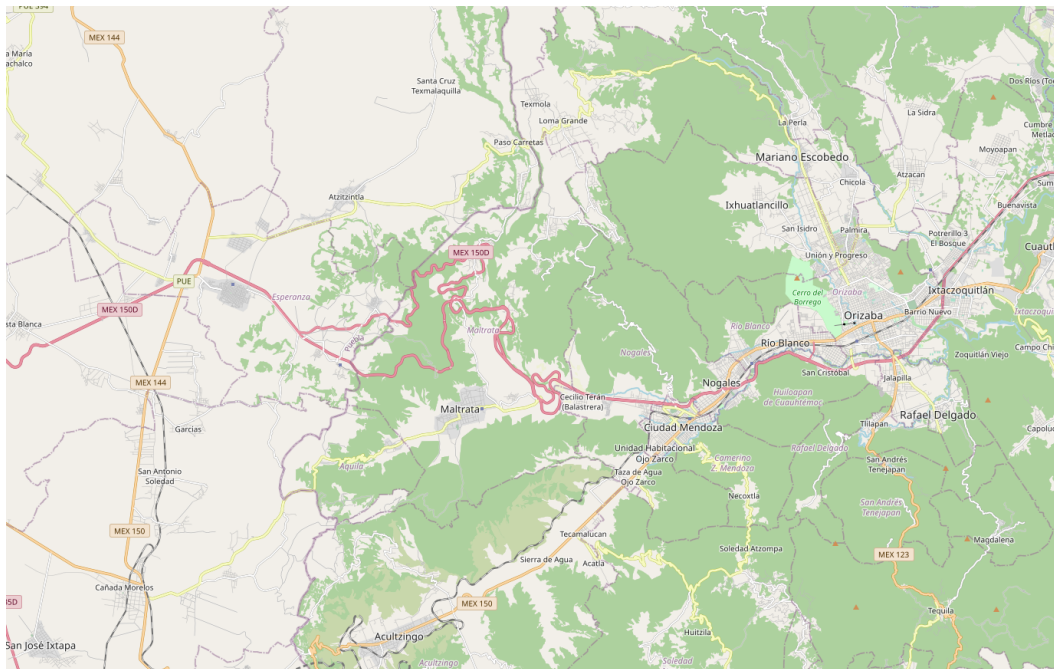


Ilustración 2. Tramo Carretero Federal "Acatzingo - Córdoba" parte de la carretera México 150D

Objetivos

Objetivo General

- Integrar una visualización interactiva y efectiva del comportamiento y tendencia sobre la criminalidad sobre la carretera Federal México 150D entre los límites de Puebla y Veracruz entre 2016 y 2019.

Objetivo Particular

- Crear un producto mínimo viable capaz de establecer relaciones entre zonas calientes a la criminalidad sobre la carretera Acatzingo – Córdoba (MEX 150D) y la probabilidad de ocurrencia en tales zonas.
- Manipular la información obtenida para observar diferentes patrones de comportamiento

Metodología

Tal como se mencionó con anterioridad, los eventos criminales recolectados serán tratados como muestras y para cada incidente criminal registrado en el *dataset* se incluyen los siguientes 13 atributos:

- 1) *ID* – Número del incidente
- 2) *FechaPublicacion* – Fecha en que fue publicado el incidente
- 3) *FechaIncidente* – Fecha en que ocurrió el Incidente reportado
- 4) *Categoria* – Tipo de crimen
- 5) *DiaDeLaSemanaPublicacion* – Día de la semana en que fue publicado el incidente
- 6) *DistritoPoliciaco* – Distrito al que pertenece el incidente reportado
- 7) *Dirección* – Kilómetro carretero aproximado en el que sucedió el incidente
- 8) *HoraDelIncidente* – Hora aproximada en la que ocurrió el incidente
- 9) *X* – Longitud de la localización del Incidente
- 10) *Y* – Latitud de la localización del Incidente
- 11) *Coordenadas* – Par de coordenadas “X” y “Y”
- 12) *Arrestos* – Número de arrestos que ocurrieron

13) *Descripcion* – Breve descripción del incidente

Considerando la naturaleza geográfica de los incidentes delictivos, la variedad de gráficas útiles para la interpretación y la característica de interactividad propuestas, se presenta el desarrollo de una aplicación web con el framework *Streamlit* y el lenguaje de programación Python.

Así mismo, el análisis del *dataset* pretende ser analizado con la librería *pandas*, los cálculos sugeridos con *numpy*, los gráficos espaciales con *pydeck* y demás gráficas integradas e interactivas con *plotly* para la visualización de tendencias.

Al utilizar como recurso un mapa interactivo, es posible filtrar incidentes criminales en lugares y fechas específicas, para su exploración. Por lo que la temporalidad apuntada es en periodos de días y horas.

Finalmente, si el alcance lo permite, resulta de interés el desarrollo del modelo predictivo propuesto, para lo cual se propone el uso del algoritmo C4.5 en la herramienta Weka.

Hipótesis: La visualización interactiva de la información será vital para la contestación de la pregunta de investigación ya que, si refleja tendencias claras y patrones de comportamiento útiles con la dinámica criminal, el explorar otras herramientas puede resultar en overfitting.

Por su naturaleza, el proyecto tiene una naturaleza explicativa en la que se busca entender suficientemente los impactos y relaciones causales para explicar el éxito o fracaso del modelo de visualización propuesto, así mismo tales resultados requieren un enfoque de investigación mixto en el que se encuentre una correlación entre la descripción de los sucesos criminales y el número de eventos registrados por lo que es evidente el uso de un diseño cuasiexperimental al tener un control parcial de las variables estudiadas.

Así mismo, al tratarse de un periodo temporal limitado es necesario considerar todos los sucesos registrados en una recolección continua y permanente que da lugar a un alcance longitudinal, pero cuyo resultado será en un campo teórico ya que, una vez finalizado el modelo propuesto, estará sujeto a arduas verificaciones que permitan asegurar su utilidad en el campo.

Fundamentación

De acuerdo con Castelvechi (2020), en los últimos años, matemáticos, estadistas y especialistas en ciencias de la computación han desarrollado algoritmos que sintetizan grandes volúmenes de datos y generando modelos que claman ayudar a las fuerzas policíacas a reducir la incidencia criminal. La efectividad de las predicciones es un elemento desconocido para el público, ya que se clasifica como información confidencial.

Con la introducción del Big Data, diversos campos de acción en diversas disciplinas han adoptado y adaptado diversas técnicas derivadas de su análisis con la finalidad de obtener mayor precisión en la extracción de información de grandes volúmenes. Según Gupta (2006), las conspiraciones y organizaciones complejas son, comúnmente, difíciles de descifrar debido a que la información sobre los sospechosos y los hechos puede ser geográficamente difusa en largos periodos de tiempo. Según Hosseinkhani (2014) las principales técnicas de minería de datos para la analizar la criminalidad son los Análisis de grupos, la minería de reglas de asociación y, la Clasificación y minería de secuencias.

Complementando las técnicas presentadas, Hassani (2016) ahonda en las técnicas de minería de datos señala que las siguientes son las más utilizadas:

- Reconocimiento de entidades nombradas,
- Análisis de grupos (Clustering),
- Minería de reglas de asociación,
- Árboles de decisiones,
- Máquinas de vectores de soporte,
- Clasificador bayesiano ingenuo,
- Redes Neuronales,
- Análisis de Redes Sociales

Tal aproximación tiene fundamento en el desarrollo analítico, con miras a desarrollar un modelo descriptivo.

Tras la ola de protestas en Estados Unidos por el movimiento “*Black Lives Matter*”, 1400 matemáticos pertenecientes a la *Notices of the American Mathematical Society*, llamaron a sus colegas a esclarecer las metodologías empleadas, ya que, de acuerdo con sus inferencias, los modelos centran sus datos de entrenamiento en las características étnicas de la sociedad. Tal técnica corresponde al Reconocimiento de entidades nombradas.

Por otro lado, es importante señalar que los modelos sugeridos por Hosseinkhani y Hassani basan sus contribuciones desde el modelado clásico y la minería a través de técnicas que ignoran la naturaleza dinámica de la criminalidad en la que el comportamiento humano, la relevancia de los periodos temporales, factores políticos, sociales, climatológicos y geográficos son solo algunos de los elementos que pueden resultar influyentes para el análisis.

Es por lo que la aproximación primordial del presente proyecto centra su atención en el concepto de la Visualización Interactiva y efectiva, siendo los pilares que pretenden dar una cosmovisión más apegada a la realidad sobre el comportamiento particular por periodos de tiempo. Dando lugar al concepto de modelo espacio-temporal.

Una vez establecido el modelo espacio temporal con su respectiva visualización, será de gran utilidad el desarrollo analítico propuesto de Análisis de grupos, para contrastar la efectividad de resultados en un campo de aplicación real.