**INSTITUTO POLITECNICNO NACIONAL**

**ESCUELA SUPERIOR DE COMPUTO**

**MINERIA DE DATOS**

**ESCOM**

PROYECTO FINAL

“**CLASIFICACION Y PREDICCION DE ACCIDENTES AUTOMOVILISTICOS EN LA CDMX”**

*PRESENTA*

**MARTINEZ LOPEZ SEBASTIAN**

[smartinezl1301@alumno.ipn.mx](mailto:smartinezl1301@alumno.ipn.mx) | 6CV1

**RAMIREZ RESENDIZ LUIS ROQUE**

[lramirezr1602@alumno.ipn.mx](mailto:lramirezr1602@alumno.ipn.mx) | 6CV1

**RODRIGUEZ YEPEZ JAAZIEL**

[jrodriguezy1600@alumno.ipn.mx](mailto:jrodriguezy1600@alumno.ipn.mx) | 6CV1

*PROFESOR*

**Dr. FABIOLA OCAMPO BOTELLO**

ÍNDICE

[RESUMEN 3](#_Toc123806980)

[PALABRAS CLAVE 3](#_Toc123806981)

[INTRODUCCION 4](#_Toc123806982)

[OBJETIVO 4](#_Toc123806983)

[CONJUNTO DE DATOS 5](#_Toc123806984)

[DICCIONARIO DE DATOS 6](#_Toc123806985)

[PROPUESTAS DE TECNIAS DE MINERIA DE DATOS. 9](#_Toc123806986)

# RESUMEN

México ocupa el 7 lugar en accidentes viales a nivel mundial, esto lo convierte en uno de los países con más índice de mortalidad vial en todo el mundo, en promedio en 2015, se registraron alrededor de 378 mil 504 siniestros de tránsito, que, a su vez, ocasionaron alrededor de 16 mil muertes, es decir, se registraron 44 defunciones diarias a causa de estos siniestros.

Debido a esto el C5, que es un órgano de la secretaria de seguridad publica pudo consolidar que 4 de cada 10 muertos en el país son peatones, 7 de cada 10 muertos, corresponden a usuarios vulnerables (peatones, ciclistas y motociclistas), además el mayor riego de morir en un hecho de transito corresponde a estos mismos y a los conductores de vehículos particulares.

Es por eso que se decide utilizar el dataset propuesto por el C5, para poder analizar en que delegaciones y en que horario es donde se tiene una mayor posibilidad de encontrar un siniestro, para que asi, se puedan crear estrategias preventivas.

# PALABRAS CLAVE

Mineria de datos, Clúster, Regresión lineal, Accidentes viales, Estrategias preventivas, Muertes.

# INTRODUCCION

México es un país donde muchas veces el socio culturalismo se antepone antes las leyes que rigen nuestra sociedad, esto combinado con los actos de corrupción pueden crear consecuencias terribles para salvaguardar la vida de los individuos que viven en esta sociedad. Desde 2015, el C5, que es un órgano dentro de la secretaria de seguridad pública, puso a disposición del publico en general, las estadísticas de accidentes viales registradas en el periodo de 2017 a 2019, esto para crear conciencia de la responsabilidad que conlleva tener un automóvil particular en la ciudad.

Y es que, muchas veces no se puede llegar a dimensionar los estragos que puede ocasionar un accidente vial, que pueden ir desde lesiones leves hasta la muerte, muchas veces, los alcalímetros posicionados no alcanzan a abarcar todo el flujo vial, debido a que los conductores, prefieren saltárselo para evitar una multa, sin preceder que eso puede ser una decisión critica.

Asi mismo, se propone crear diversas estrategias a partir de la informacion compartida por el C5, para que asi se puedan reducir estos accidentes viales y con esto aumentar la posibilidad de salvar una vida en uno de estos siniestros.

# OBJETIVO

* Identificar, clasificar, predecir los tipos de accidentes que pueden ocurrir en las distintas delegaciones de la CDMX.
* Proponer estrategias de posicionamiento de alcoholímetros, puntos de revisión, o puntos de actuación para los elementos de seguridad publica en la CDMX.

# CONJUNTO DE DATOS

Para poder realizar las técnicas de mineria de datos, se propuso el siguiente dataset, el cual, nos habla acerca de los diversos accidentes automovilísticos ocurridos en la CDMX registrados desde 2017 hasta 2019.

|  |  |
| --- | --- |
| **Nombre del conjunto de datos** | México roads accidents |
| **Fecha de publicación** | 05 enero 2020 |
| **Autor(es)** | Eduardo Romero |
| **Objetivo del conjunto de datos** | Este conjunto de datos contiene accidentes registrados por el C4, un sistema mexicano que registra todos los incidentes de tránsito. |
| **Cantidad de atributos** | 18 |
| **Cantidad de registros** | 207,365 |
| **Enlace** | https://www.kaggle.com/datasets/laloromero/mexico-road-accidents-during-2019 |

# DICCIONARIO DE DATOS

Dentro del dataset, podemos observar que se encuentran las siguientes variables, asi mismo, en el siguiente apartado, las desglosamos para la comprensión del lector.

**Tabla 1: Diccionario de los datos**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No. Atributo** | **Nombre** | **Tipo** | **Significado** | **Dominio** |
| **1** | Folio | Nominal | Identificación unica para cada registro | NA |
| **2** | fecha\_creacion | Intervalo | Fecha de creacion de la incidencia | De 15/11/2018 a 30/11/2019 |
| **3** | hora\_creacion | Intervalo | Hora de creacion de la incidencia | De 12:00:00 a. m. a 11:59:59 p. m. |
| **4** | dia\_semana | Nominal | Dia del accidente | De Lunes a Domingo |
| **5** | codigo\_cierre | Nominal | Codigo de clasificación | R: Afirmativo, si el incidente es confirmado por el equipo de emergencias N: Negativo, si el equipo de emergencias no confirma la incidencia en el punto de localización I: Informativo, por si los equipos de atención quieren añadir información extra F: Falso, si el informe inicial no coincide con los hechos reales  D: Duplicados, registros con código de cierre afirmativo, negativo o falso pero los operadores los identifican |
| **6** | fecha\_cierre | Intervalo | Fecha donde se resolvió el accidente | De 01/01/2019 a 30/12/2019 |
| **7** | año\_cierre | Nominal | Año del cierre del accidente | De 2018 a 2019 |
| **8** | mes\_cierre | Nominal | Mes del cierre del accidente | De Enero a Diciembre |
| **9** | hora\_cierre | Intervalo | Hora del cierre del accidente | De 12:00:00 a. m. a 11:59:59 p. m. |
| **10** | delegacion\_inicio | Nominal | Delegación donde se creó el accidente | Álvaro Obregón, Azcapotzalco, Benito Juárez, Coyoacán, Cuajimalpa de Morelos, Cuauhtémoc, Gustavo A. Madero, Iztacalco, Iztapalapa, Magdalena Contreras, Miguel Hidalgo, Milpa Alta, Tláhuac, Tlalpan, Venustiano Carranza, Xochimilco |
| **11** | incidente\_c4 | Nominal | Explicación del accidente | lesionado-atropellado, accidente-choque sin lesionados ,accidente-choque con lesionados ,accidente-motociclista,accidente-otros,accidente-volcadura,accidente-persona atrapada / desbarrancada,accidente-vehiculo atrapado,accidente-choque con prensados,sismo-choque con lesionados,sismo-choque sin lesionados,detenciÃ³n ciudadana-atropellado,cadÃ¡ver-atropellado,cadÃ¡ver-accidente automovilÃ­stico,accidente-vehiculo desbarrancado,detenciÃ³n ciudadana-accidente automovilÃ­stico,sismo-persona atropellada,accidente-ferroviario,sismo-choque con prensados,accidente-ciclista,accidente-vehÃ­culo atrapado-varado, |
| **12** | latitud | Nominal | Latitud del accidente | NA |
| **13** | longitud | Nominal | Longitud del accidente | NA |
| **14** | clas con f\_alarma |  | Código que identifica la gravedad de la situación |  |
| **15** | tipo\_entrada |  | Cómo se reportó el incidente |  |
| **16** | delegacion\_cierre | Nominal | Delegación donde se cerró el accidente | Álvaro Obregón, Azcapotzalco, Benito Juárez, Coyoacán, Cuajimalpa de Morelos, Cuauhtémoc, Gustavo A. Madero, Iztacalco, Iztapalapa, Magdalena Contreras, Miguel Hidalgo, Milpa Alta, Tláhuac, Tlalpan, Venustiano Carranza, Xochimilco |
| **17** | geopunto | Nominal | Columnas de latitud y longitud combinadas | NA |
| **18** | mes | Nominal | Mes en que se reportó el incidente | De Enero a Diciembre |

# PROPUESTAS DE TECNIAS DE MINERIA DE DATOS.

**Tabla 2: Técnicas de mineria de datos**

|  |  |  |
| --- | --- | --- |
| **Técnica** | **Objetivo** | **Atributos** |
| Árboles (Tipo) | NA |  |
| Bagging | NA |  |
| Boosting | NA |  |
| Agrupamiento (#) | Segmentar en Norte, Sur, Poniente y Oriente los accidentes viales en la CDMX, para que asi, se puedan consultar las estadísticas totales de cada región. |  |
| Regresión lineal (&) | Predecir, en que delegación es mas probable que ocurra un sinisestro. |  |
| Reglas de asociación (\*) | NA |  |
| Análisis de componentes | NA |  |