

Aplicación de minería de datos mediante la metodología CRISP-DM para defunciones accidentales y violentas en Nezahualcóyotl

^aMiguel Rubén Marcial Rodríguez ^bArgenis Gabriel López Montaña, ^cLiliana Rodríguez Páez, ^dRicardo Rico Molina, ^eDoricela Gutiérrez Cruz

^aCentro Universitario UAEM Nezahualcóyotl, Edo. México, mmarcialr001@alumno.uaemex.mx

^bCentro Universitario UAEM Nezahualcóyotl, Edo. México, alopezm024@alumno.uaemex.mx

^cCentro Universitario UAEM Nezahualcóyotl, Edo. México, clodriguezp@uaemex.mx

^dCentro Universitario UAEM Nezahualcóyotl, Edo. México, rricom@uaemex.mx

^eCentro Universitario UAEM Nezahualcóyotl, Edo. México, dgutierrezcr@uaemex.mx

Resumen

La minería de datos es una técnica que consiste en la aplicación de algoritmos específicos que generan una enumeración de patrones a partir de grandes volúmenes de información. Para realizar estos análisis existen diversas metodologías, una de ellas es la CRISP MD. En este artículo se presenta la aplicación de la metodología CRISP MD (Cross Industry Standar Process for Data Mining) la cual es una metodología imparcial o neutra respecto a la herramienta que se utilice Data Warehouse o Minería de Datos. Con el objetivo de clasificar las defunciones accidentales y violentas en el municipio de Nezahualcóyotl, identificando, el tipo de defunción, de acuerdo al género, escolaridad, edad, en un periodo de enero-diciembre entre los años de 1990 a 2018, utilizando el clasificador Naive Bayes y la técnica de árboles de decisión J48 mediante el software de WEKA. Después del análisis de los datos fue posible identificar que la variable sexo, nivel de escolaridad y el rango de edad fueron claves para detectar que defunción fue la más causada en el municipio de Nezahualcóyotl. Estos resultados muestran que con el clasificador Naive Bayes en combinación con otros algoritmos de clasificación podrían ser útil para aplicarlos en futuros análisis en una amplia variedad de datos para identificar patrones de comportamiento.

Palabras clave

Análisis de datos, clasificador, mortalidad, Naive Bayes, WEKA

Abstract

Data mining is a technique that consists of the application of specific algorithms that generate an enumeration of patterns from large volumes of information. To perform these analyses there are several methodologies, one of them is CRISP MD. This article presents the application of the CRISP MD (Cross Industry Standard Process for Data Mining) methodology, which is an impartial or neutral methodology with respect to the Data Warehouse or Data Mining tool used. With the objective of classifying accidental and violent deaths in the municipality of Nezahualcóyotl, identifying, the type of death,

according to gender, schooling, age, in a period from January-December between the years 1990 to 2018, using the Naive Bayes classifier and the technique of decision trees J48 using WEKA software. After the analysis of the data it was possible to identify that the variable sex, level of schooling and age range were key to detect which death was the most caused in the municipality of Nezahualcóyotl. These results show that the Naive Bayes classifier in combination with other classification algorithms could be useful to apply in future analyses on a wide variety of data to identify patterns of behavior.

Keywords

Data analysis, classifier, mortality, Naive Bayes, WEKA

1. INTRODUCCIÓN

De acuerdo con la Organización Mundial de la Salud (OMS) los datos de mortalidad al 28 de febrero de 2022 reflejan las defunciones con causas básicas codificadas en los sistemas nacionales del Registro Civil. La causa básica de defunción se define como "la enfermedad o lesión que desencadenó la sucesión de eventos patológicos que condujeron directamente a la muerte, o las circunstancias del accidente o acto de violencia que produjeron la lesión mortal", según lo expuesto en la Clasificación Internacional de Enfermedades [1].

En México, el Instituto Nacional de Estadística y Geografía (INEGI) [2,3], empezó a contar de manera sistemática las defunciones accidentales y violentas del Ministerio Público. Donde se captan las defunciones registradas en todo el país mediante el Sistema de Registro Civil, así como las defunciones accidentales y violentas en las que intervienen el Servicio Médico Forense y el Ministerio Público. [3].

El reciente informe sobre defunciones publicado por el INEGI destaca que el año 2019 fue el que más fallecimientos por homicidios registró en la entidad desde el año 2015, con 3242. En 2015 la entidad reportó 2693 homicidios; en 2016 un total de dos mil 768; para el siguiente año 3076 y en 2018 acumuló 3170 [2,4].

Uno de los estados con más registros fue el Estado de México, donde se encuentra el municipio de Nezahualcóyotl, el cual durante el año del 2019 registró una tasa de 15.3 casos de homicidio en el municipio. Como se observa la tendencia al alza viene desde 2016, año en el que se tenía una tasa de 11.6 [4,5].

En este municipio el equipo del centro de investigación revisó 805 reportes policiales de homicidios ocurridos entre 2013 y 2018 [5]. En estos encontraron:

- Las víctimas de homicidio fueron mayoritariamente hombres.
- Cuatro de cada diez víctimas eran jóvenes entre 15 y 29 años.
- Los homicidios tienen un patrón temporal los cuales se cometieron entre el viernes y el domingo en

horario más común entre las 19:00 p.m. y las 2:00 a.m.

- Casi la mitad de los homicidios se cometió en la vía pública.
- Siete de cada diez homicidios en sucedieron sobre grandes avenidas y el 60% ocurrió en la frontera con los municipios de Iztapalapa y Texcoco [2].

En ese sentido el gobierno municipal tiene el reto de garantizar la seguridad de sus habitantes y de los miles de ciudadanos que cruzan los límites entre el Estado de México y la Ciudad de México día con día [2,5].

Lo cual ha generado gran cantidad de datos, que deben ser analizados para encontrar patrones de comportamiento delictivo en este ámbito. Entre estas técnicas se encuentra la minería de datos. La minería de datos permite descubrir patrones de comportamiento consistentes, o similares y modelos con un alto grado de utilidad sobre las bases de datos de gran tamaño [6,8], que junto a los métodos estadísticos complementan las acciones para establecer políticas de estrategia en la toma de decisiones [7].

En este contexto se han realizado diversas investigaciones de Minería de Datos y Machine Learning en el campo de seguridad para analizar tendencias de homicidios.

En el caso de Colombia, Ordoñez et al. [9] emplearon el algoritmo de aprendizaje supervisado, cual fue el randomforest y minado por etapas de CRIPS-DM pudieron tener una tendencia baja en el número de homicidios con un total de disminución del 60%.

Por otra parte, el estudio de Aplicación de minería de datos en datos abiertos de Ecuador en Delitos, mostraron que hubo homicidios más frecuentes en Guayaquil en los últimos trimestres del año 2019 y 2020, también se logró demostrar que los robos tuvieron un mayor comportamiento en Quito y Guayaquil y que tuvieron los mayores índices en tipo de delito y así poder diseñar planes de prevención [10].

Jaramillo y Arias [11]. Reportaron que el uso de algoritmos de clasificación como J48 y el algoritmo CHAID pueden identificar las causas fundamentales de la violencia de género contra las mujeres, ya que las mujeres representadas en el árbol de decisiones pueden identificar claramente las causas de la violencia.

En el contexto de hechos delictivos Jaulis [18] reportó que apoyados en la minería de datos y en el uso de la metodología CRISP MD se obtuvo como resultado predecir donde será el acto delictivo y el estimado de horas, sin embargo, este modelo tuvo una tasa de acierto del 72,09%.

Lo que evidencia que estas técnicas y algoritmos han tenido buenos resultados para identificar patrones de comportamiento.

Por lo cual el objetivo de esta investigación es identificar variables como sexo, nivel de escolaridad y el rango de edad para detectar que defunción fue la más causada en el municipio de Nezahualcóyotl, mediante la metodología CRISP DM de minería de datos.

La importancia de este artículo es conocer los factores que se generan en la población del municipio de Nezahualcóyotl para el alza en la tasa de las defunciones accidentales y violentas desde 1990 y que estrategias sociales se pueden utilizar para reducir este problema.

2. MATERIALES Y MÉTODOS

Se empleó la metodología CRISP-DM siendo una de las más empleadas actualmente para minería de datos que tiene como característica principal su flexibilidad que permite personalizar fácilmente, teniendo en cuenta las siguientes etapas [12, 15, 16]:

1. Comprensión del problema
2. Comprensión de datos
3. Preparación de datos
4. Modelado
5. Evaluación del modelo
6. Implantación del modelo

2.1 Comprensión del problema

Esta primera etapa permite entender y delimitar la problemática, así como identificar los requisitos, supuestos, restricciones y beneficios del proyecto. En México el INEGI integra a los homicidios en las defunciones accidentales y violentas, esto se debe a que se determina con base en las afectaciones y lesiones [13]. En el 2019 el Estado de México fue uno de los estados con mayor índice de defunciones accidentales y violentas. Siendo el municipio de Nezahualcóyotl el que mayor índice de defunciones de este tipo registrados tuvo entre los años del 2016 y 2019. Determinando que el género masculino era mayor víctima de homicidio y el índice femenino era a causa de asesinatos por parte de sus familiares o parejas.

En el 2017 a nivel mundial se tuvo registrado que alrededor del 81 por ciento de las víctimas de homicidio eran hombres y niños. Existen muchas causas, pero entre los factores principales que impulsan este delito son: el desempleo, la desigualdad, la inestabilidad política y la presencia de la delincuencia organizada [14].

El objetivo que se ha definido es identificar el tipo de defunción accidental y violenta que predomina en el municipio de Nezahualcóyotl de acuerdo al género, escolaridad, edad, en un periodo de enero-diciembre entre los años de 1900 a 2018.

2.2 Comprensión de datos

Esta segunda etapa, comprende principalmente la recolección, descripción y explotación de los datos. Los datos recopilados se obtuvieron del Instituto Nacional de Estadística y Geografía (INEGI) [13]. Los cuales representa

las defunciones accidentales y violentas en el municipio de Nezahualcóyotl con 2120 registros, se encuentran almacenados por: año, mes, escolaridad, sexo, edad y tipo de defunción (Tabla 1).

Estas variables fueron elegidas para conocer el rango de año y mes de ocurrencia de las defunciones, así también que grado de estudios y edad tenían las personas fallecidas en este municipio y conocer si el gobierno municipal ha realizado algo para reducir este índice de defunciones.

Tabla 1. Descripción de los datos.

Atributo	Definición	Descripción
A	Año	Año de en qué se reporta en la defunción
M	Mes	Mes de ocurrencia de la defunción.
ES	Escolaridad	Hace referencia a que grado de estudios tenía el fallecido.
E	Edad	Edad que tenía el fallecido.
S	Sexo	Género del fallecido.
DE	Defunción	Tipo de fallecimiento.

Fuente: Elaboración propia.

2.3 Preparación de datos

En esta etapa, comprende la preparación del conjunto de datos, de tal forma que se puede adaptar las técnicas de Data Mining, que estas permiten buscar relaciones entre las variables u otras medidas para la exploración de los datos. Se incluye tareas de selección y limpieza de datos que directamente se realizaron desde la página del INEGI y obteniendo lo que se muestra en la Tabla 2.

Tabla 2. Descripción de los datos relacionados con las defunciones accidentales y violentas en Nezahualcóyotl.

Variable	Rango
Año	1990 a 2018
Mes	Enero a Diciembre
Escolaridad	Secundaria completa, Secundaria incompleta, Bachillerato completo, Bachillerato incompleto y Profesional
Edad	15 años a 30 años
Sexo	Hombre-Mujer
Defunción	Homicidio, Accidente y Suicidio

Fuente: Elaboración propia.

Para la tarea de cambio de formato a cada una de las variables se optó por crearle un acrónimo que se muestra en las siguientes tablas:

Tabla 3. Acrónimos para el tipo de defunción registrada.

Atributo	Definición
ho	Homicidio
ac	Accidente
su	Suicidio

Fuente: Elaboración propia.

Los atributos analizados para el tipo de defunción registrada fueron de acuerdo a la clasificación de muertes violentas y accidentales del INEGI las cuales se dividen en presuntos homicidios, presuntos accidentes, presuntos suicidios, evento de intención no determinada y otras causas [3].

Tabla 4. Acrónimo para el grado de estudios.

Atributo	Definición
Sc	Secundaria Completa
Si	Secundaria Incompleta
Bc	Bachillerato Completo
Bi	Bachillerato Incompleto
Pr	Profesional

Fuente: Elaboración propia.

Tabla 5. Acrónimo para el género.

Atributo	Definición
H	Hombre
M	Mujer

Fuente: Elaboración propia.

Al concluir con las tareas que se mencionaron anteriormente se obtiene las variables que se consideran relevantes para cumplir con los objetivos.

2.4 Modelado

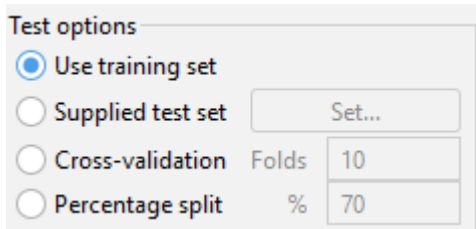
En esta etapa, se selecciona la técnica de modelado de minería de datos a utilizar. Debido a que se emplea el software WEKA [17], para realizar las tareas de minería de datos, el cual utilizan diferentes técnicas de modelado, y entre los modelos que mejor se adaptan a nuestro objetivo son: Naive Bayes y Árboles de decisión puesto que el objetivo de las defunciones accidentales y violentas en el municipio de Nezahualcóyotl es establecido como un problema de clasificación. Naive Bayes es uno de los algoritmos de aprendizaje inductivo más efectivos. Simplifica enormemente el aprendizaje al asumir la independencia de características, pero en la práctica compite con clasificadores más complejos [15].

2.5 Evaluación del modelo

En esta etapa, se evalúa la fiabilidad calculada para los modelos planteados y determinar el mejor, mismo que nos permitirá llegar al objetivo planteado. La mayoría de los trabajos realizados con minería de datos recomienda utilizar un 70% del conjunto de datos como de entrenamiento y un 30% como prueba. Por lo cual se ha decidido que se utilice estos valores para los dos clasificadores.

En WEKA el valor inicial es de 66%, por lo que se modifica los valores por defecto que proporciona WEKA y se establece como en la Fig. 1.

Fig. 1. Valores que permiten evaluar la calidad y validez.



Fuente: Elaboración propia.

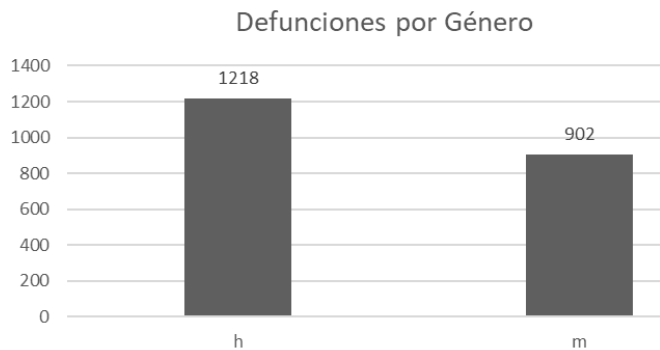
2.6 Implantación del modelo

En esta etapa, se busca explotar los resultados mediante el modelo seleccionado. Esto permite obtener la clasificación del tipo de defunción que mayor afecta al municipio y a que género es el que mayor incide en esto, de acuerdo a la escolaridad y edad en un periodo de enero a diciembre del año 1990- 2018. Teniendo cuenta el estudio antes realizado por el INEGI y así poder conocer si la clasificación es correcta o próxima al estudio realizado.

3. RESULTADOS

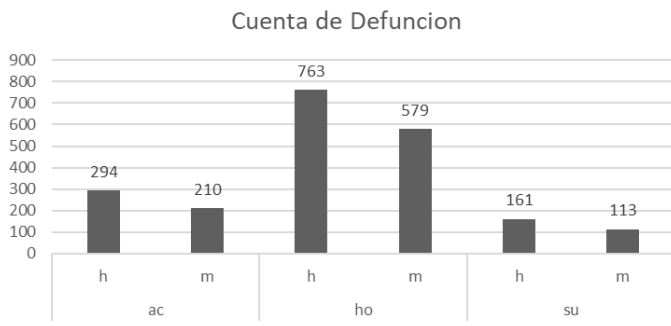
De acuerdo con lo analizado en el municipio de Nezahualcóyotl de 2120 defunciones 1218 fueron de hombres y 902 de mujeres, esto de 1990 a 2018 como se aprecia en la Fig. 2.

Fig. 2. Total de defunciones por género.



Fuente: Elaboración propia.

Fig. 3. Clasificación de defunciones por género.

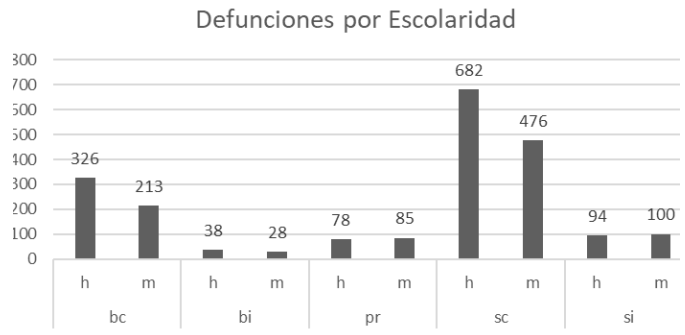


Fuente: Elaboración propia.

A su vez también se puede observar en la Fig. 3. que el homicidio que predomina como la mayor causa de defunción en hombres con 763 en el periodo analizado.

Al contrastar la información de defunciones por homicidio, accidente o suicidio con el grado de escolaridad, se encontró que el número de defunciones de mayor frecuencia (homicidio) se relaciona con la escolaridad de secundaria completa (sc) siendo esta con un total de 682 defunciones (Fig. 4).

Fig. 4. Clasificación de defunciones por escolaridad y género.



Fuente: Elaboración propia.

De acuerdo a lo obtenido con el software WEKA haciendo uso del clasificador Naive Bayes se obtuvo que el año promedio estuvo en 2001 y 2004 en hombre y mujer respectivamente, el mes en ambos casos fue el 6 perteneciente a junio, en escolaridad el factor predominante fue sc (secundaria completa) con 683 casos en hombres y 477 casos en mujeres, la edad promedio estuvo entre 23 para hombre y 22 para mujer, a su vez la defunción con mayor predominancia fue ho (homicidio) con 764 y 580, en hombre y mujer respectivamente (Fig. 5).

Fig. 5. Descripción de los resultados Naive Bayes.

Variable	Hombres	Mujeres
Año	2001	2004
Mes	junio	junio
Escolaridad	sc	sc
Edad	23	22
Defunción	ho	ho

Fuente: Elaboración propia.

El clasificador Naive Bayes al realizar la clasificación del conjunto de entrenamiento, se obtiene las instancias que fueron clasificadas (Fig. 6.) obteniendo 1233 instancias correctas y 887 incorrectas.

Fig. 6. Descripción de instancias clasificadas Naive Bayes.

Clasificación	Cantidad	Porcentaje
Correctas	1233	58.1604 %
Incorrectas	887	41.8396 %

Fuente: Elaboración propia.

A su vez el software Weka, después de procesar los datos realizo la matriz de confusión que en su diagonal se obtuvo 937 instancias correctas para hombres y 296 instancias correctas para mujeres ver Fig. 7.

Fig. 7. Matriz de confusión Naive Bayes.

	h	m
h	937	281
m	606	296

Fuente: Elaboración propia.

Utilizando la técnica de los árboles de decisión, enfocándonos al J48 teniendo un conjunto de entrenamiento del 70% y un 30% para el conjunto de prueba, se obtuvieron los siguientes resultados plasmados en la Fig. 8.

Fig. 8. Descripción de los resultados J48.

Variable	Hombres	Mujeres
Año	2015	2009
Mes	junio	junio
Escolaridad	sc	sc
Edad	23	21
Defunción	ho	ho

Fuente: Elaboración propia.

A su vez y como se plasma en la figura anterior utilizando el J48 se obtuvo que el año promedio estuvo en 2015 y 2009 en hombre y mujer respectivamente, el mes en ambos casos fue el 6 perteneciente a junio, en escolaridad el factor predominante fue sc (secundaria completa), la edad promedio estuvo entre 23 para hombre y 21 para mujer, a su vez la defunción con mayor predominancia fue ho (homicidio) para hombre y mujer descritos en la Fig. 12.

El árbol de decisión J48 desarrollado en WEKA obtuvo un total de 119 número de hojas y con un total de tamaño de 208 descritos en la Fig. 9.

Fig. 9. Descripción de la formación del J48.

	Total
Número de hojas	119
Tamaño del árbol	208

Fuente: Elaboración propia.

La técnica de árboles de decisiones al realizar la clasificación del conjunto de entrenamiento, se obtiene las instancias que fueron clasificadas (Fig. 10.) obteniendo 1750 instancias correctas y 370 incorrectas.

Fig. 10. Descripción instancias clasificadas J48.

Clasificación	Cantidad	Porcentaje
Correctas	1750	82.5472 %
Incorrectas	370	17.4528 %

Fuente: Elaboración propia.

Weka, después de procesar los datos realizó la matriz de confusión que en su diagonal se obtuvo 1118 instancias correctas para hombres y 632 instancias correctas para mujeres ver Fig. 11.

Fig. 11. Matriz de confusión J48.

	h	m
h	1118	100
m	270	632

Fuente: Elaboración propia.

Al realizar el análisis de los resultados de ambos clasificadores y teniendo en cuenta la matriz de confusión, podemos llegar a la conclusión de que el J48 es más certero al realizar la clasificación y obteniendo el resultado que se realizó por el INEGI, esto se debe a que tuvo una mayor clasificación correcta al momento del conjunto de entrenamiento y por el lado del Naive Bayes fue más bajo y por lo tanto la certeza del resultado no fue la correcta al tener un bajo porcentaje al clasificar y por lo tanto no se acercó a los resultados que previamente se habían mencionado (Fig. 12).

Fig. 12. Comparación de los resultados.

Variable	J48		Naive Bayes	
	Hombres	Mujeres	Hombres	Mujeres
Año	2015	2009	2001	2004
Mes	junio	junio	junio	junio
Escolaridad	sc	sc	sc	sc
Edad	23	21	23	22
Defunción	ho	ho	ho	ho

Fuente: Elaboración propia.

4. CONCLUSIÓN

Este estudio se focalizo en el municipio de Nezahualcóyotl, debido al interés de obtener información de uno de los municipios más poblado en la zona metropolitana teniendo un impacto mayor por el número de población y se pueda tener resultados acercados en tema de seguridad acerca de las defunciones violentas que se comenten día a día en nuestro país y focalizándolo en el municipio de Nezahualcóyotl.

El presente estudio da evidencia de las necesidades de seguridad, vigilancia y educación de una población vulnerable en el municipio, de tal forma que se analiza que de los resultados obtenidos mediante los clasificadores, sería importante conocer porque el mes de junio y la edad promedio que es donde mayor incidencia de defunciones hay de este tipo y de esta manera evidenciar de las necesidades de campañas adecuadas para dirigir a la comunidad juvenil a estudiar o continuar con sus estudios. Para trabajos futuros se recomienda hacer una valoración por gestiones gubernamentales ya que esto podría permitir valorar las políticas públicas aplicadas en cada periodo.

5. REFERENCIAS

- [1] OMS. (2018, 10 mayo). Biblioguias: Mortalidad y salud: Mortalidad y Salud @ ONU. Recuperado 28 de febrero de 2022, de <https://biblioguias.cepal.org/c.php?g=398214&p=3602022#:~:text=Los%20datos%20de%20mortalidad%20indican,codificadas%20por%20las%20autoridades%20nacionales>.
- [2] El sol de Toluca, S. E. R. E. |. (2020, 2 noviembre). Casi el 10 por ciento de muertes violetas y accidentales ocurrieron en Edomex en 2019: INEGI. El Sol de Toluca | Noticias Locales, Policiacas, sobre México, Edomex y el Mundo. Recuperado 28 de febrero de 2022, de <https://www.elsoldetoluca.com.mx/local/casi-el-10-por-ciento-de-muertes-violetas-y-accidentales-ocurrieron-en-edomex-en-2019-inegi-5963883.html>
- [3] INEGI. (2021, 27 julio). Datos preliminares revelan que en 2020 se registraron 36 579 homicidios. Recuperado 28 de febrero de 2022, de <https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2021/EstSociodemo/Defuncioneshomicidio2020.pdf>
- [6] INEGI. (2020, 14 abril). Defunciones registradas (mortalidad general). Recuperado 28 de febrero de 2022, de <https://www.inegi.org.mx/sistemas/olap/Proyectos/bd/continuas/mortalidad/MortalidadGeneral.asp>
- [4] El Universal. (2020, 6 noviembre). Sobre los homicidios en Neza. Recuperado 28 de febrero de 2022, de <https://www.eluniversal.com.mx/opinion/alejandrohope/sobre-los-homicidios-en-neza>
- [5] Hot Spot Neza. (2019, 3 diciembre). Neza 10,000 cuadras resguardadas por vecinos. Recuperado 28 de febrero de 2022, de <http://seguridadacontracorriente.mexicoevalua.org/wp-content/uploads/2020/11/HotSpotNeza.pdf>
- [6] Troche Clavijo, Alvaro. (2014). Aplicación de la minería de datos sobre bases de datos transaccionales. *Fides et Ratio - Revista de Difusión cultural y científica de la Universidad La Salle en Bolivia*, 7(7), 58-66. Recuperado en 13 de febrero de 2023, de http://www.scielo.org.bo/scielo.php?script=sci_arttext&pid=S2071-081X2014000100005&lng=es&tlng=es.
- [7] Valenga, F., Fernández, E., Merlino, H., Rodríguez, D., Procopio, C., Britos, P., & García-Martínez, R. (2008). Minería de Datos Aplicada a la Detección de Patrones Delictivos en Argentina. In *JIISIC* (pp. 31-40).
- [8] Velarde Martínez, A., (2003). Minería de Datos. Una Introducción. *Conciencia Tecnológica*, (23).
- [9] H.-A. Ordoñez-Eraso, C.-J. Pardo-Calvache, C.-A. Cobos-Lozada, "Detection of Homicide Trends in Colombia Using Machine Learning," *Revista Facultad de Ingeniería*, vol. 29 (54), e11740, 2020. <https://doi.org/10.19053/01211129.v29.n54.2020.11740>
- [10] Colina Vargas, A. M., & Espinoza Mina, M. A. (2022). Aplicación de minería de datos en datos abiertos de Ecuador: Delitos. *UCV Hacer*, 11(1), 79-93. <https://doi.org/10.18050/RevUCVHACER.v11n1a8>.
- [11] Jaramillo, A., & Arias, H. P. P. (2015). Aplicación de Técnicas de Minería de Datos para Determinar las Interacciones de los Estudiantes en un Entorno Virtual de Aprendizaje. *Revista Tecnológica-ESPOL*, 28(1).
- [12] Gallardo, J.A. (2018). Metodología para el desarrollo de proyectos en Minería de Datos CRIPS-DM. Recuperado de http://www.oldemarrodriguez.com/yahoo_site_admin/assets/docs/Documento_CRISP-DM.2385037, de <https://biblioguias.cepal.org/c.php?g=398214&p=3602022#:~:text=Los%20datos%20de%20mortalidad%20indican,codificadas%20por%20las%20autoridades%20nacionales>.
- [13] INEGI. (2020, 14 abril). Defunciones registradas (mortalidad general). Recuperado 28 de febrero de 2022, de <https://www.inegi.org.mx/sistemas/olap/Proyectos/bd/continuas/mortalidad/MortalidadGeneral.asp>
- [14] ONU. (2019, 08, julio). El homicidio causa muchas más muertes que los conflictos armados, según nuevo estudio de la UNODC. Recuperado 28 de febrero de 2022, de <https://www.onu.org.mx/el-homicidio-causa-muchas-mas-muertes-que-los-conflictos-armados-segun-nuevo-estudio-de-la-unodc/>
- [15] Wu, Xindong, Kumar, Vipin, Quinlan, J. Ross, Ghosh, Joydeep, Yang, Qiang, Motoda, Hiroshi, Mclachlan, Geoffrey J., Ng, Angus, Liu, Bing, Yu, Philip S., Zhou, Zhi-Hua, Steinbach, Michael, Hand, David J. Y Steinberg, Dan, 2008. Top 10 algorithms in data mining. *Knowledge and Information Systems*. 2008. Vol. 14, p. 1-37. DOI 10.1007/s10115-007-0114-2.
- [16] Marbán, M.S. (2018). A Data mining & knowledge discovery process model. Recuperado de http://cdn.intechopen.com/pdfs/5937/InTech-A_data_mining_amp_knowledge_discovery_process_model.pdf
- [17]. Weka 3. Data Mining with Open Source Machine Learning Software in Java. (s. f.). <https://www.cs.waikato.ac.nz/ml/weka/>
- [18] Jaulis, J. J. J. R., & Vilcarromero, J. R. V. G. (2015). Sistema De Predicción De Hechos Delictivos Para La Mejora Del Proceso De Prevención Del Delito En El Distrito De La Molina Utilizando Minería De Datos [Para Optar El Título Profesional De Ingeniero De Computación Y Sistemas]. Facultad De Ingeniería Y Arquitectura Escuela Profesional De Ingeniería De Computación Y Sistemas. https://repositorio.usmp.edu.pe/bitstream/handle/20.500.12727/2022/jaulis_vilcarromero.pdf?sequence=1&isAllowed=y