

Minería de Datos en el estudio de comorbilidades relacionadas a las defunciones ocurridas por COVID-19 en la CDMX durante el 2020.

Litzy Sarai-López Pineda^a, Dra. Doricela-Gutiérrez Cruz^b, Dra. Carmen Liliana-Rodríguez Páez^c.

^a Centro Universitario UAEM Nezahualcóyotl, sarailopez.lslp@gmail.com, Nezahualcóyotl, Edo. Méx., México.

^b Centro Universitario UAEM Nezahualcóyotl, dgutierrezcr@uaemex.mx, Nezahualcóyotl, Edo. Méx., México.

^c Centro Universitario UAEM Nezahualcóyotl, clrodriguezp@uaemex.mx, Nezahualcóyotl, Edo. Méx., México.

Resumen

En el presente trabajo, se identificaron las comorbilidades relacionadas con las defunciones ocurridas en el 2020 en la Ciudad de México (CDMX), durante la pandemia por Covid-19. Se analizaron 68 750 datos respecto a las variables: Alcaldía, Causa de Defunción, Edad, Escolaridad y Sexo, mediante la aplicación de los algoritmos J48 (tree), Decision Table (rules) y JRip (rules), en el software de código abierto WEKA y con el empleo de la metodología de Minería de Datos, KDD “Knowledge Discovery in Databases”. Únicamente se consideraron las enfermedades renales, cáncer, hipertensión y diabetes. Obteniendo como resultados, que las comorbilidades a las que se les atribuye la cantidad mayoritaria de defunciones suscitadas por Covid-19, corresponden a diabetes y cáncer, específicamente en hombres mayores de 64 años.

Palabras clave—Algoritmo J48, Análisis de Datos, Decision Table, JRip, Proceso KDD.

Abstract

In the present work, the comorbidities related to the deaths that occurred in 2020 in Mexico City (CDMX), during the Covid-19 pandemic, were identified. 68,750 data were analyzed regarding the variables: Mayor's Office, Cause of Death, Age, Education and Sex, by applying the algorithms J48 (tree), Decision Table (rules) and JRip (rules), in open source software. WEKA and with the use of the Data Mining methodology, KDD "Knowledge Discovery in Databases". Only kidney diseases, cancer, hypertension and diabetes were considered. Obtaining as results, that the comorbidities to which the majority of deaths caused by Covid-19 are attributed, correspond to diabetes and cancer, specifically in men over 64 years of age.

Keywords—Data Analysis, Decision Table, J48 Algorithm, JRip, KDD Process.

1. INTRODUCCIÓN

La población mexicana asciende a casi 130 millones de habitantes, de los cuales el 52% son mujeres y el 48% hombres [1], la cual se caracteriza por una alta prevalencia de hipertensión, obesidad y diabetes, [2] enfermedades que se mencionan como comorbilidades asociadas a la de mortalidad

por COVID-19 en México. El estudio epidemiológico de esta enfermedad es fundamental para entender el riesgo en poblaciones específicas [3].

Con el paso de la edad el sistema inmunológico de las personas se va deteriorando y por consiguiente la exposición a padecer alguna enfermedad va en aumento. Es aquí, donde el término comorbilidad se hace presente, puesto que refiere a la presencia de más de una enfermedad en un mismo individuo [4, 5].

La Comisión Nacional de Población (CONAPO), señala que México es un país que involucra una situación alarmante, ya que la población infantil es menor a la adulta actualmente, y para el año 2025 se seguirá manteniendo esta diferencia [6].

Así mismo, estadísticas de la Organización Mundial de la Salud (OMS), indican que las enfermedades que forman parte de las principales causas de muerte en el mundo y, por consiguiente, provocarán crecientes afectaciones, son: las respiratorias, pulmonares, tuberculosis y cáncer [6].

Lo anterior, en conjunto con la carencia de medidas preventivas funcionales dificultan la condición de salud [6]. El COVID-19 requiere una atención especializada y oportuna para evitar complicaciones, sin embargo, el hecho de contar con alguna comorbilidad potencia el índice de hospitalización y en dado caso de defunción de la persona infectada, en contraste con el de una persona completamente sana [7].

Los decesos provocados en México por esta enfermedad (COVID-19), en el 2020 fueron 683 823 en comparación con el 2019, en donde, se reportaron 499 784, y de los cuales la Ciudad de México (CDMX) asentó 82 449, según fuentes del Registro Civil, Servicios Médicos Forenses y Agencias del Ministerio Público, siendo esta la segunda causa de muerte a nivel mundial, teniendo un impacto superior en el mes de julio con un 17.1% [8].

La importancia de reportar estas cifras por parte de las instituciones, sin duda genera conocimiento de la problemática, pero esto no sería posible sin el almacenamiento de la misma, en este sentido radica la relevancia de analizar dicha información a través de Tecnologías de la Información y Comunicación (TIC'S), entre las que se encuentra la inteligencia artificial, Big Data y Minería de Datos [9, 10], esta última como instrumento de análisis y en disparidad con otras técnicas convencionales, es útil para resolver problemas concretos de manera ágil, y al ponerse en práctica en el sector salud apoya estratégicamente evitando nuevos contagios y reduciendo consecuencias ante padecimientos [11, 12].

La Minería de Datos (MD) o Data Mining es el proceso de descubrir patrones, modelos y otros tipos de conocimiento interesantes [13], a través de la información extraída de una Base de Datos (BD). Con tal propósito, es como la MD aplica algoritmos, los cuales son una secuencia de pasos para

construir por medio de los datos un modelo con criterios adecuados [14].

En la MD se emplean metodologías con la finalidad de alcanzar resultados exitosamente, entre las más dominantes se encuentra el KDD “Knowledge Discovery in Databases” [15], un proceso que distingue patrones en grandes bases de datos, para hallar conocimiento [16].

Aunado a esto, el principal objetivo de este trabajo es identificar mediante la aplicación de los algoritmos J48, Decision Table y JRip, las comorbilidades relacionadas con las defunciones ocurridas en el 2020 en la Ciudad de México, durante la pandemia por Covid-19.

2. MATERIALES Y MÉTODOS

Los datos recabados fueron emitidos por la Secretaría de Salud de la Ciudad de México [17], institución que proporciona información precisa y confiable sobre el tema abordado, y se extrajeron del Portal de Datos Abiertos de la Ciudad de México [18].

Se analizaron 68 750 datos correspondientes a los fallecimientos ocurridos en la Ciudad en el transcurso del 01 de enero al 31 de diciembre de 2020, desde una perspectiva epidemiológica. Mismos que fueron actualizados hasta el día 20 de septiembre de 2022.

Donde se empleó la metodología de descubrimiento de conocimiento KDD, por sus siglas en inglés, que consta primeramente de un entendimiento del problema, seguido de una selección de fuentes y preprocesamiento de datos, y por último de un análisis de datos [19].

2.1 Entendimiento del Problema

En esta etapa del proceso se prioriza la asimilación de cada aspecto del problema en cuestión para brindar una solución coherente a través de la minería de datos, obteniendo con esto resultados que efectivamente permitan cumplir el objetivo previamente planteado.

2.2 Selección y Preprocesamiento de Datos

Se efectuó la identificación de atributos o variables significativas, es decir, aquellas en las que la información contenida se puede aprovechar para adquirir mayor conocimiento, y son las siguientes: (Alcaldía, Causa de Defunción, Edad, Escolaridad y Sexo), las cuales se muestran en la Tabla 1.

Tabla 1. Descripción de los atributos con base al certificado de defunción de cada persona.

ATRIBUTO	DESCRIPCIÓN
ALC	Denota la alcaldía de registro de la defunción de la persona.

CD	Denota la enfermedad por la que fallece la persona.
EDAD	Denota la edad de la persona tras su defunción.
ESC	Denota el nivel educativo alcanzado por la persona fallecida.
SEXO	Denota el género de la persona fallecida (Hombre o Mujer).

Fuente: Elaboración propia a partir del “Diccionario de Datos sobre los Certificados de defunción de la Secretaría de Salud de la Ciudad de México a partir del 01 de enero de 2020, del Portal de Datos Abiertos de la Ciudad de México”.

La limpieza realizada a los datos extraídos abarcó la corrección de caracteres especiales existentes y la eliminación de registros faltantes que generan una disminución de la calidad del minado de datos, además de ruido e inconsistencias.

En las Tablas 2 y 3 se describen los acrónimos correspondientes al atributo ALC y ESC.

Tabla 2. Acrónimos existentes en la Base de Datos a analizar, respecto al atributo ALC.

ACRÓNIMO	DESCRIPCIÓN
AC	Acalipa.
ACF	Acopiaco [Campamento Forestal].
ALA	Albergue Alpino Ajusco.
AMA	AA Milpa Alta.
AO	Álvaro Obregón.
ASM	Ampliación San Miguel.
AZC	Azcapotzalco.
BJ	Benito Juárez.
CA	Cansacaballos.
CB	Cruz Blanca.
CM	Cuajimalpa de Morelos.
CP	Cola de Pato.
CU	Cuauhtémoc.
CY	Coyoacán.
DES	Desconocido.
ESA	Ejidos de San Andrés Totoltepec.
GA	Gustavo A. Madero.
IZP	Iztapalapa.
IZT	Iztacalco.
MC	La Magdalena Contreras.
MH	Miguel Hidalgo.
OA	Ojo de Agua.
SAM	San Andrés Mixquic.
SAN	Santa Ana Tlacotenco.
SAT	San Antonio Tecomitl.
SCY	Santa Catarina Yecahuitzotl.
SFT	San Francisco Tecoxpa.
SJI	San Juan Ixtayopan.
SLA	San Lorenzo Acopilco.
SLT	San Lorenzo Tlacoyucan.
SM	San Miguel.
SMT	San Miguel Topilejo.
SNT	San Nicolás Tetelco.

SPA	San Pedro Atocpan.
SPO	San Pablo Oztotepec.
SSC	San Salvador Cuauhtenco.
TEM	Temaxcatitla (Kilómetro 32.2).
TH	Tláhuac.
TM	Tlamenolpa (Los Encinos).
TP	Tlalpan.
VC	Venustiano Carranza.
VMA	Villa Milpa Alta.
XOC	Xochimilco.

Fuente: Elaboración propia.

Tabla 3. Acrónimos existentes en la Base de Datos a analizar, respecto al atributo ESC.

ACRÓNIMO	DESCRIPCIÓN
BPC	Bachillerato o Preparatoria Completa.
BPI	Bachillerato o Preparatoria Incompleta.
LPC	Licenciatura o Profesional Completo.
LPI	Licenciatura o Profesional Incompleto.
NIN	Ninguna.
PC	Preescolar Completo.
PI	Preescolar Incompleto.
POC	Posgrado Completo.
POI	Posgrado Incompleto.
PRC	Primaria Completa.
PRI	Primaria Incompleta.
SC	Secundaria Completa.
SI	Secundaria Incompleta.

Fuente: Elaboración propia.

2.3 Análisis de Datos

En esta actividad se encuentran diferentes tipos de tareas, las cuales pueden considerarse un problema para ser resuelto a través de un algoritmo de minería de datos [20]. Entre las más importantes se encuentran la clasificación, segmentación o clustering, asociación y patrones secuenciales. En este caso particular fue aplicada la clasificación, la cual permite lograr resultados a partir de un proceso de aprendizaje supervisado, además, en esta se conduce a encontrar propiedades comunes entre un conjunto de objetos de una base de datos y son catalogados en diferentes clases, de acuerdo con el modelo de clasificación [21], como fue con el J48 (tree). Dicho algoritmo, J48 incluye cálculos estadísticos, y pertenece al algoritmo C4.5 expuesto por Quinlan, el cual se usa para generar árboles de decisión, determinando el valor de la variable dependiente para nuevos casos [22, 23, 24].

De igual forma, se emplearon los algoritmos de asociación, Decision Table (rules) y JRip (rules), cuya finalidad fue descubrir patrones en forma de reglas, que muestren los hechos que ocurren frecuentemente en un conjunto de datos determinado [25]. Decision Table o bien DTM (Decision Table Majority), durante la clasificación busca las correspondencias de cada dato de entrada, y al no encontrarlas simplemente asigna el dato a la clase mayoritaria [26, 27, 28]. Mientras que el algoritmo Ripper (JRip), proporciona un enfoque basado en reglas, y hace uso del error incremental

reducido JRip, con el propósito de ir hallando un conjunto de reglas para cada clase [29, 30].

Para las cuestiones propiamente de análisis y construcción del modelo de minería de datos, se utilizó el software de código abierto, desarrollado en la Universidad de Waikato en Nueva Zelanda, conocido como WEKA (Waikato Environment for Knowledge Analysis), que precisamente contiene herramientas para la preparación, clasificación, regresión, agrupación, minería de reglas de asociación y visualización de datos [31]. Donde la ejecución de los algoritmos mencionados con anterioridad, se llevó a cabo, manteniendo los parámetros que establece por defecto WEKA.

3. RESULTADOS

A continuación, se describen los resultados obtenidos conforme a cada tarea de clasificación realizada, principalmente el algoritmo que alcanzó una mayor efectividad fue Decision Table con un 63.35% de instancias correctamente clasificadas. En cuanto al tema abordado, se tiene que el género con más repercusiones por decesos ocurridos durante la pandemia es el masculino con un 83.7%, frente a un 33.6% del femenino, tal y como se muestra en la Figura 1.

Fig. 1. Ejecución de los datos con base a los decesos ocurridos en el 2020, con el algoritmo Decision Table.

```

Time taken to build model: 4.06 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      43555      63.3527 %
Incorrectly Classified Instances    25195      36.6473 %
Kappa statistic                    0.186
Mean absolute error                 0.4479
Root mean squared error             0.4741
Relative absolute error             92.8153 %
Root relative squared error         96.5121 %
Total Number of Instances          68750

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
Weighted Avg.   0.634   0.460   0.623     0.634   0.607     0.201   0.652   0.547   Hombre
                0.336   0.163   0.586     0.336   0.427     0.201   0.652   0.547   Mujer

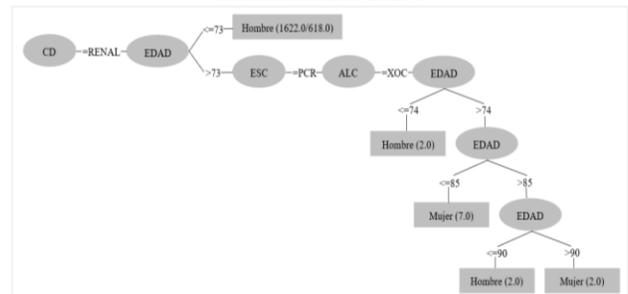
=== Confusion Matrix ===
      a    b  <-- classified as
34150  6647 | a = Hombre
18548  9405 | b = Mujer

```

Fuente: Elaboración propia.

Con el algoritmo J48 se obtuvo un 63.26% de instancias correctamente clasificadas. En la Figura 2, se visualiza que los hombres menores de 74 años son los más afectados por enfermedades renales.

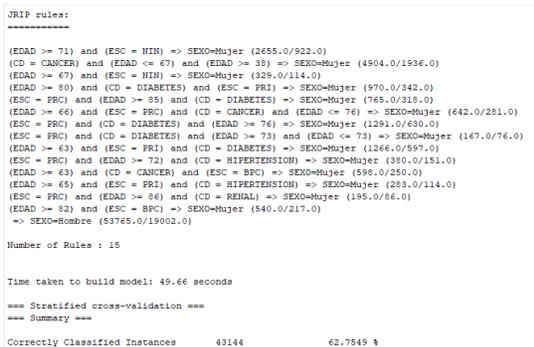
Fig. 2. Árbol de decisión del algoritmo J48, con base a los decesos ocurridos en el 2020.



Fuente: Elaboración propia.

En la Figura 3 se observa que el sexo predominante al ejecutar el último algoritmo, JRip, es el masculino con un 85.3%, mientras que el femenino con un 29.9% y un 62.75% de instancias correctamente clasificadas.

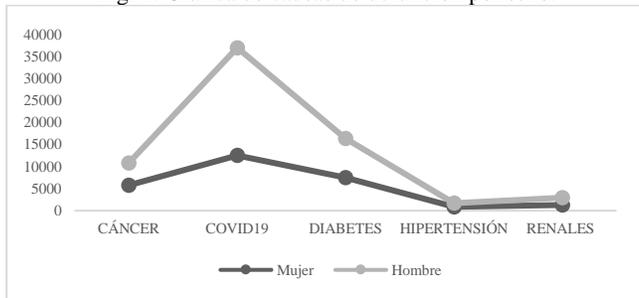
Fig. 3. Ejecución de los datos con base a los decesos ocurridos en el 2020, con el algoritmo JRip.



Fuente: Elaboración propia.

De manera general, se puede apreciar en la Figura 4 la cantidad de defunciones ocurridas según el sexo de la persona y la enfermedad. Hombres y mujeres fallecieron en menor medida, por padecimientos de hipertensión y renales.

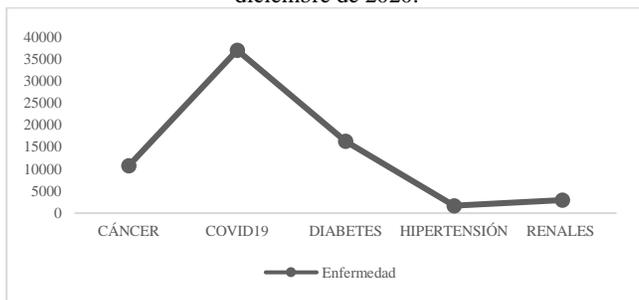
Fig. 4. Gráfica de causas de defunción por sexo.



Fuente: Elaboración propia.

La Figura 5 muestra las causas primordiales en el aumento de muertes durante el año 2020, las cuales son: Covid-19 con 36 973, Diabetes con 16 349 y Cáncer con 10 783, así como en los estudios reportados por [32, 33].

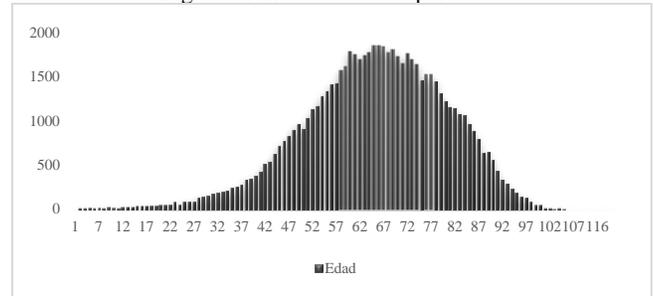
Fig. 5. Gráfica de defunciones ocurridas del 01 de enero al 31 de diciembre de 2020.



Fuente: Elaboración propia.

De los 68 750 decesos totales, 36 965 equivalentes al 53.76%, se asocian al grupo de adultos mayores de 64 años y tan solo un 0.46% (323), a la población menor de edad. Lo dicho con anterioridad, se expone en la Figura 6 y coincide con el trabajo realizado por [34].

Fig. 6. Gráfica de decesos por edad.

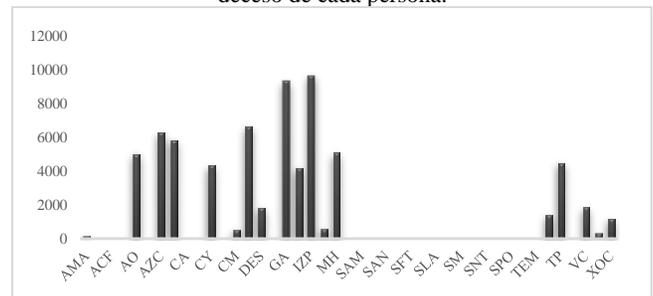


Fuente: Elaboración propia.

La estadística indica que un 27.02% de la población fallecida contaba con PRC (Primaria Completa), un 18.31% con SC (Secundaria Completa), y 13.65% con BPC (Bachillerato o Preparatoria Completa). Solo, 0.06% tenían estudios PI de (Preescolar Incompleto).

La Figura 7 revela el impacto en los decesos registrados en la CDMX, exactamente en las alcaldías (IZP (Iztapalapa) y GA (Gustavo A. Madero)), la primera con 9 648 y la segunda con 9 325 respectivamente.

Fig. 7. Gráfica de alcaldías de la CDMX en las que se registró el deceso de cada persona.



Fuente: Elaboración propia.

4. CONCLUSIONES Y RECOMENDACIONES

De acuerdo con la información analizada en esta investigación, mediante los algoritmos J48, Decision Table y JRip, se concluye que las comorbilidades a las que se les atribuye la cantidad mayoritaria de defunciones suscitadas por COVID-19 en la CDMX en el 2020, corresponden a diabetes y cáncer, presentes específicamente en hombres con un rango de edad de 65 años y más, cuya escolaridad era Primaria Completa, y con un registro de fallecimiento proveniente de la alcaldía Iztapalapa.

Para este caso en particular, se resalta la importancia de la aplicación de la Minería de Datos, puesto que las relaciones y patrones descubiertos son de utilidad y en beneficio de los

habitantes, a fin de proteger y contrarrestar el ascenso en la mortalidad en el país. Ayudando a indicar las dosis de medicamentos suministrados a cada paciente, contabilizando cuantas personas son atendidas en cada hospital por una infección en cierto periodo, o bien cuales enfermedades son frecuentes con base a las estaciones del año.

En este contexto, crear un plan de salud es fundamental para que la población sin importar su origen de residencia o edad, mejore sus hábitos para que su sistema inmune produzca una respuesta deseable, y si ya se padece alguna enfermedad, se pueda disminuir la probabilidad de muerte con la debida medicación y medidas brindadas.

5. REFERENCIAS

- [1]. World Health Organization (WHO). (2020). Coronavirus disease (COVID-19): Situation Report - 142. Disponible en: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200610-covid-19-sitrep-142.pdf?sfvrsn=1_80898cd_6
- [2]. The World Bank. (2022). Mexico Overview: Development news, research, data | World Bank. Disponible en: <https://www.worldbank.org/en/country/mexico/overview>
- [3]. Parra-Bracamonte, G. M., Lopez-Villalobos, N., & Parra-Bracamonte, F. E. (2020). Clinical characteristics and risk factors for mortality of patients with COVID-19 in a large data set from Mexico. *Annals of epidemiology*, 52, 93-98. Disponible en: <https://www.sciencedirect.com/science/article/pii/S1047279720302866>
- [4]. Lifshitz, A. (2016). Sobre la "comorbilidad". *Acta médica Grupo Ángeles*, 14(2), 61-62. Disponible en: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1870-72032016000200061&lng=es&tlng=es
- [5]. Marín-Navarrete, R., & Medina-Mora, M. E. (2015). Comorbilidades en los Trastornos por Consumo de Sustancias: Un desafío para los servicios de salud en México. *La depresión y otros trastornos psiquiátricos. México: Academia Nacional de Medicina de México AC*, 39-58.
- [6]. Instituto Nacional de Salud Pública (INSP). (2015). Enfermedad respiratoria en adultos; Reto para la salud pública de México. Disponible en: <https://www.insp.mx/avisos/3827-enfermedad-respiratoria-adultos.html>
- [7]. Instituto Mexicano del Seguro Social (IMSS). (2020). COM. 521 Redoblar medidas de autocuidado para personas con comorbilidades es clave para evitar riesgos por COVID-19. Disponible en: <https://www.gob.mx/imss/prensa/com-521-redoblar-medidas-de-autocuidado-para-personas-con-comorbilidades-es-clave-para-evitar-riesgos-por-covid-19>
- [8]. Instituto Nacional de Estadística y Geografía (INEGI). (2021). Características de las defunciones registradas en México durante enero a agosto de 2020. Disponible en: https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2021/EstSociodemo/DefuncionesRegistradas2020_Pnles.pdf
- [9]. Torres Fernández, J. P., et. al. (2017). Gestión de la información como herramienta para la toma de decisiones en salud: escenarios más probables. *Revista Cubana de Investigaciones Biomédicas*, 36(3). Disponible en: http://scielo.sld.cu/scielo.php?pid=S086403002017000300010&script=sci_arttext&tlng=en
- [10]. Rodríguez Casiano, J., et al. (2021). Análisis de comorbilidad asociados a la mortalidad por COVID 19 en el municipio de Nezahualcóyotl mediante Algoritmo K-means y EM. *Número Especial de la Revista Aristas: Investigación Básica y Aplicada*, 8(16). Disponible en: http://revistaaristas.tij.uabc.mx/index.php/revista_aristas/article/view/112
- [11]. Buczak, A. L., et. al. (2014). Prediction of high incidence of dengue in the Philippines. *PLoS neglected tropical diseases*, 8(4), e2771. Disponible en: <https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0002771>
- [12]. Sanchez-Moreno, E. A., Torres-Peralta, R., Cirett-Galán, F., & Barceló-Valenzuela, M. (2017). Propuesta de una Metodología Utilizando Minería de Datos para Estrategias de Medicina Preventiva Más Eficaces. *Avances de Investigación en Ingeniería en el Estado de Sonora*, 3(1), 211-217. Disponible en: <http://www.aviies.ith.mx/revistas/AvIIES-2017.pdf>
- [13]. Han, J., Pei, J., & Tong, H. (2022). *Data mining: concepts and techniques*. Morgan Kaufmann. Disponible en: https://books.google.es/books?hl=es&lr=&id=NR1oEAAAQBAJ&oi=fnd&pg=PP1&dq=Data+Mining:+Concepts+and+Techniques&ots=_M8GMNvhs0&sig=LIW9CSrrokjSV2UfOIB_eDC9QWs#v=onepage&q=Data%20Mining%3A%20Concepts%20and%20Techniques&f=false
- [14]. Microsoft. (2022). Algoritmos de minería de datos (Analysis Services: Minería de datos). Disponible en: <https://learn.microsoft.com/es-es/analysis-services/data-mining/data-mining-algorithms-analysis-services-data-mining?view=asallproducts-allversions>
- [15]. Guzmán León, E. Metodologías aplicadas al proceso de Minería de Datos. Disponible en: https://disi.unal.edu.co/~eleonguz/cursos/md/presentaciones/Sesion5_Metodologias.pdf
- [16]. Mancilla-Vela, G., Leal-Gatica, P., Sánchez-Ortiz, A., & Vidal-Silva, C. (2020). Factores asociados al éxito de los estudiantes en modalidad de aprendizaje en línea: un análisis en minería de datos. *Formación universitaria*, 13(6), 23-36. Disponible en: <https://dx.doi.org/10.4067/S0718-50062020006000023>

- [17]. Gutiérrez, Pérez, L. A., et. al. (2021). Descubrimiento de conocimiento en incidencia de tipo de cáncer para pacientes terminales mediante minería de datos. *CIENCIA ergo-sum, Revista Científica Multidisciplinaria de Prospectiva*, 28(1). Disponible en: <https://www.redalyc.org/journal/104/10464915006/10464915006.pdf>
- [18]. Portal de Datos Abiertos de la CDMX. (Consultado el 12 de diciembre del 2022). Disponible en: <https://datos.cdmx.gob.mx/>
- [19]. García-González, J. R., Sánchez-Sánchez, P. A., Orozco, M., & Obredor, S. (2019). Extracción de Conocimiento para la Predicción y Análisis de los Resultados de la Prueba de Calidad de la Educación Superior en Colombia. *Formación universitaria*, 12(4), 55-62. Disponible en: <https://dx.doi.org/10.4067/S0718-50062019000400055>
- [20]. Hernández, J., Ramírez, M. J., & Ferri, C. (2005). *Introducción a la minería de datos*. Editorial Pearson Prentice Hall. Disponible en: <https://es.scribd.com/document/496554106/Jose-Hernandez-Orallo-Cesar-Ferri-Ramirez-Maria-Jose-Ramirez-Quintana-Introduccion-a-La-Mineria-de-Datos-2004-Pearson-Educacion-Libgen-li#>
- [21]. Agrawal, R., Ghosh, S., Imielinski, T., Iyer, B., & Swami, A. N. (1992). An interval classifier for database mining applications. *VLDB*, 92, 560-573. Disponible en: <https://www.vldb.org/conf/1992/P560.PDF>
- [22]. Tabora, C. H. C., García, N. G., & Roza, J. J. P. (2016). Análisis de datos mediante el algoritmo de clasificación J48, sobre un cluster en la nube de AWS. *Redes de Ingeniería*, 3-15. Disponible en: <https://revistas.udistrital.edu.co/index.php>
- [23]. Rivera, M. S. I. W., Suárez, A. R., & Díaz, A. R. (2010). Aplicación de la Minería de Datos para el análisis de información clínica. Estudio Experimental en cardiopatías isquémicas. *Conjunto*, 1223, 100. Disponible en: http://www.rcim.sld.cu/revista_18/articulos_pdf/mineriadatos.pdf
- [24]. Charris, L., Henríquez, C., Hernández, S., Jimeno, L., Guillen, O., & Moreno, S. (2018). Análisis comparativo de algoritmos de árboles de decisión en el procesamiento de datos biológicos. *Investigación y desarrollo en TIC*, 9(1), 26-34. Disponible en: <https://revistas.unisimon.edu.co/index.php/identific/article/download/3158/3905>
- [25]. Timarán-Pereira, S. R., Hernández-Arteaga, I., Caicedo-Zambrano, S. J., Hidalgo-Troya, A. y Alvarado-Pérez, J. C. (2016). El proceso de descubrimiento de conocimiento en bases de datos. En *Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional* (pp. 63-86). Bogotá: Ediciones Universidad Cooperativa de Colombia. Disponible en: <http://dx.doi.org/10.16925/9789587600490>
- [26]. Alcántara-Ramírez, A., López-García, L., & Paz, J. R. (2017). Generación de un vector característico para la detección de intrusos en redes computacionales. *Research in Computing Science*, 139, 19-31. Disponible en: https://www.rcs.cic.ipn.mx/rcs/2017_139/Generacion%20de%20un%20vector%20caracteristico%20para%20la%20deteccion%20de%20intrusos%20en%20redes%20computacionales.pdf
- [27]. Berdun, F., Armentano, M. G., & Amandi, A. (2016). Inferencia de roles de equipo a partir de conductas colaborativas detectadas en interacciones textuales. *Simposio Argentino de Inteligencia Artificial (ASAI 2016)-JAIIO*, 45, 78-85. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/56972>
- [28]. Univaso, P., Ale, J. M., & Gurlekian, J. A. (2015). Data mining applied to forensic speaker identification. *IEEE Latin America Transactions*, 13(4), 1098-1111. Disponible en: <https://ieeexplore.ieee.org/abstract/document/7106363>
- [29]. Torres-Vásquez, M., Hernández-Torruco, J., Hernández-Ocaña, B., & Chávez-Bosquez, O. (2021). Impacto de los algoritmos de sobremuestreo en la clasificación de subtipos principales del síndrome de Guillain-Barré. *Ingenius. Revista de Ciencia y Tecnología*, (25), 20-31. Disponible en: <https://doi.org/10.17163/ings.n25.2021.02>
- [30]. Rajput, A., Aharwal, R. P., Dubey, M., Saxena, S., & Raghuvanshi, M. (2011). J48 and JRIP rules for e-governance data. *International Journal of Computer Science and Security (IJCSS)*, 5(2), 201-207. Disponible en: <https://www.cscjournals.org/library/manuscriptinfo.php?mc=IJCSS-448>
- [31]. Weka 3: Machine Learning Software in Java. (Consultado el 7 de enero del 2023). Disponible en: <http://www.cs.waikato.ac.nz/ml/weka/>
- [32]. Plasencia-Urizarri, T. M., Aguilera-Rodríguez, R., & Almaguer-Mederos, L. E. (2020). Comorbilidades y gravedad clínica de la COVID-19: revisión sistemática y meta-análisis. *Revista Habanera de Ciencias Médicas*, 19. Disponible en: <https://www.redalyc.org/articulo.oa?id=180465399002>
- [33]. Calixto-Calderón, B., et. al. (2021). Comorbilidad preexistente, el mayor factor de riesgo de mal pronóstico de COVID-19 en la población Mexicana. *Nova Scientia*, 13. Disponible en: <https://doi.org/10.21640/ns.v13ie.2823>.
- [34]. Centros para el Control y la Prevención de Enfermedades (CDC). (2022). El COVID-19 y las muertes por cancer. Disponible en: <https://www.cdc.gov/spanish/cancer/dcpc/research/articles/covid-19-cancer-deaths.html#:~:text=Entre%20el%202018%20y%20el%20vari%C3%B3l%20de%202028%20a%201055>