

Percepción de seguridad en el transporte público del Estado de México en relación con la contingencia COVID 19 mediante el uso de Minería de Datos.

Erick Noriega Díaz, Javier Romero Torres, Doricela Gutiérrez Cruz, Ricardo Rico Molina.

Universidad Autónoma del Estado de México, Centro Universitario UAEM Nezahualcóyotl, Av. Bordo de Xochiaca S/N, Benito Juárez, 57000 Nezahualcóyotl, Méx., ericknd29@hotmail.com.

Universidad Autónoma del Estado de México, Centro Universitario UAEM Nezahualcóyotl, Av. Bordo de Xochiaca S/N Benito Juárez, 57000 Nezahualcóyotl, Méx., jromero@uaemex.mx.

Universidad Autónoma del Estado de México, Centro Universitario UAEM Nezahualcóyotl, Av. Bordo de Xochiaca S/N Benito Juárez, 57000 Nezahualcóyotl, Méx., dgutierrezcr@uaemex.mx.

Universidad Autónoma del Estado de México, Centro Universitario UAEM Nezahualcóyotl, Av. Bordo de Xochiaca S/N Benito Juárez, 57000 Nezahualcóyotl, Méx., ricom@uaemex.mx.

Resumen

Se analiza el nivel de percepción de seguridad en el transporte público del Estado de México en relación con la contingencia COVID-19 mediante el uso de minería de datos. Los 1788 datos fueron obtenidos de la encuesta movilidad y medidas sanitarias-contingencia COVID-19 realizada por la UAEM en el año 2020. Las variables evaluadas fueron seguridad accediendo, seguridad esperando, seguridad a bordo, seguridad general. Para la clasificación se utilizó el algoritmo J48. La percepción de seguridad en el uso del transporte público en el Estado de México se mantiene constante. Con el algoritmo apriori se identifica que el 88 % de los usuarios percibieron inseguridad en el transporte durante su trayecto.

Palabras clave—Arboles de clasificación, COVID-19, Knowledge Discovery in Databases, Minería de Datos, Transporte público.

Abstract

The objective of this work is to analyze the level of perception of security in public transport in the State of Mexico in relation to the COVID-19 contingency using data mining. The 1,788 data were obtained from the mobility and sanitary measures-COVID-19 contingency survey carried out by the UAEM in 2020. The variables evaluated were security accessing, security waiting, security on board, and general security. The J48 algorithm was used for classification. The perception of safety in the use of public transport in the state of Mexico remains constant. With the a priori algorithm, it is identified that 88% of users perceived insecurity in transport during their journey.

Keywords— COVID-19, classification tree, Data Mining, Knowledge Discovery in Databases, public transport

1. INTRODUCCIÓN

En México se detectó el primer caso de COVID-19 el 27 de febrero de 2020. Se trataba de un mexicano que había viajado

a Italia y tenía síntomas leves. El 20 de marzo del 2020 se declara la suspensión de clases presenciales debido a la pandemia y tres días más tarde da inicio la jornada de sana distancia, ya para el 30 de marzo el consejo de salud general decretaría una emergencia sanitaria lo cual significó el cierre de una gran parte de los servicios de los gobiernos en todos los niveles [1].

La pandemia trajo consigo una nueva normalidad con medidas de contención contra la COVID-19 como lo fue el uso del cubrebocas obligatorio, la sana distancia, accesos restringidos, reducción de la movilidad y en un punto una total restricción a la movilidad.

El aislamiento social fue la principal medida de contingencia por lo que los medios de comunicación fomentaron que las personas realizaran sus actividades desde casa y solo salieran en caso de ser necesario, dando como resultado un cambio en la cantidad de gente que usaba el transporte público, todos los países mostraron un sensible descenso de uso del transporte público, variando del 53.5% hasta el 85.6% de acuerdo con el Instituto Mexicano del transporte en 2020[2].

La pandemia causó desconfianza al transporte público por el miedo al alto riesgo de contagio debido a que las personas se trasladaban en espacios limitados por el tamaño del propio vehículo, sin embargo, hay una variedad de investigaciones que demuestran lo contrario.

Un estudio publicado en septiembre de 2020 por la NTU (Asociación Nacional de Empresas de Transporte Urbano), entidad nacional brasileña que representa las empresas de transporte por autobús, evaluó 15 sistemas de transporte público por buses, responsables de 177 municipios, y concluyó que no existe relación entre el número de pasajeros transportados y la variación en el número de casos del nuevo coronavirus, así mismo se informa que no hay una relación entre los vínculos de los brotes de COVID-19 y las redes de transporte público [3]. Lo anterior es posible gracias a la recopilación de los datos y su importancia en el tratamiento, por lo que los sistemas de información forman parte de las herramientas y métodos utilizados, ya que permiten una mejor forma de adquisición, tratamiento y análisis de los datos. Siendo estos fundamentales debido a que almacenan información en donde es necesaria la medición de la calidad de los datos [4].

Actualmente en temas de movilidad ya se cuenta con una gran variedad de estudios donde es utilizada la minería de datos en diversas problemáticas, tal es el caso del análisis respecto a la demanda del transporte aéreo [5], identificar factores más influyentes en la ocurrencia de siniestros de tránsito [6], hallar patrones que clasifiquen y predigan comportamientos de usuarios que utilizan Smart Cards [7].

Dada la relevancia de la minería de datos en esta área, resulta importante su aplicación, entendiendo esta como el proceso que consiste en extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos

almacenados en distintos formatos. Entonces el objetivo fundamental de la minería de datos es encontrar modelos inteligibles a partir de los datos recogidos [8].

Los tipos de modelo estándar en la minería de datos incluyen la regresión (regresión normal, para la predicción, la regresión logística para la clasificación), redes neuronales, y los árboles de decisión, siendo estas técnicas las más conocidas [9].

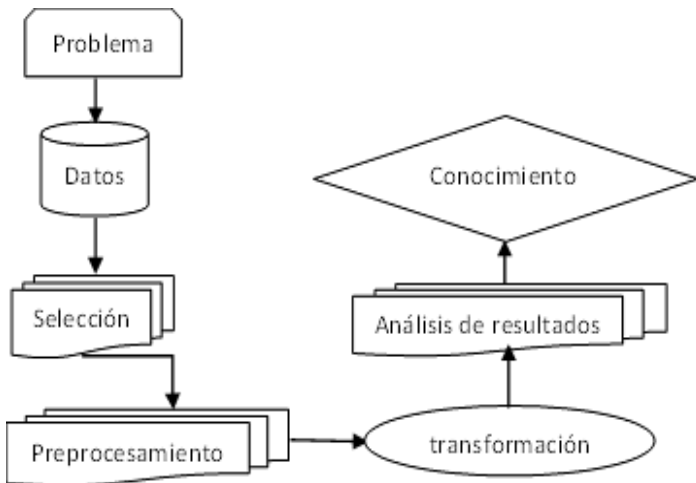
El presente trabajo de estudio tiene como objetivo hacer uso de la minería de datos para descubrir el nivel de seguridad que percibían las personas al usar el transporte público con las medidas de contingencia durante la pandemia COVID-19 en el año 2020.

2. Materiales y Métodos.

Se analizaron 1788 datos correspondientes a las 447 personas que respondieron la encuesta sobre movilidad y medidas sanitarias-contingencia COVID-19 realizada por la Universidad Autónoma del Estado de México en el Centro Universitario Nezahualcóyotl en el año 2020.

Se utilizó el enfoque cuantitativo de tipo descriptivo para el desarrollo de la investigación. Haciendo uso de la metodología del KDD (Knowledge Discovery in Databases) para los datos recabados, donde se contempla: entendimiento del problema, selección de datos, preprocesamiento y limpieza y transformación de datos como se puede observar en la ilustración 1.

Ilustración 1. Esquema del proceso Knowledge Discovery in Databases.



Fuente: Elaboración propia.

2.1 Entendimiento del problema.

En esta fase se requiere comprender con exactitud el problema al que se le va a dar solución por medio de la minería de datos. Esto permitirá recolectar la información necesaria para interpretar con asertividad los resultados encontrados [10].

2.2 Selección y preprocesamiento de datos.

En esta etapa se lleva a cabo el proceso de selección y preprocesamiento de los datos identificando los atributos (seguridad accediendo, seguridad esperando, seguridad a bordo, seguridad general) y la limpieza de datos consistió en seleccionar de manera coordinada los datos más relevantes; por el tipo de rigor, al recabar la información y procesarla, no se tuvieron inconsistencia de datos o valores atípicos, por lo que no fue necesario subsanar datos faltantes. En la tabla 1 se muestra la descripción de cada uno de los atributos seleccionados.

Tabla 1. Descripción de atributos basados en la encuesta sobre movilidad y medidas sanitarias contingencia COVID-19.

Atributo	Descripción
Seguridad accediendo	Es la percepción de seguridad que el encuestado percibió al acceder al transporte público bajo las medidas preventivas de COVID-19.
Seguridad esperando	Es la percepción de seguridad que el encuestado percibió al esperar al transporte público bajo las medidas preventivas de COVID-19.
Seguridad a bordo	Es la percepción de seguridad que el encuestado percibió al abordar al transporte público bajo las medidas preventivas de COVID-19.
Seguridad general	Es la percepción de seguridad del encuestado al viajar en transporte público bajo todas las medidas preventivas del COVID-19.

Fuente: Elaboración propia.

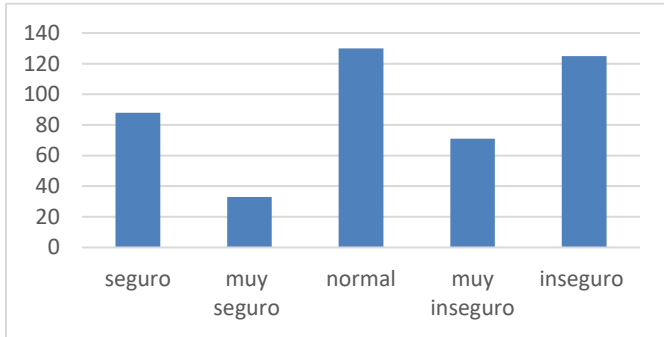
2.3 Análisis de datos

El análisis de datos de los atributos evaluados (seguridad accediendo, seguridad esperando, seguridad a bordo, seguridad general) se realizó haciendo uso del software WEKA, el cual es un software de uso libre para el aprendizaje automático y la minería de datos, desarrollado con el lenguaje de programación java en la Universidad de Waikato en Nueva Zelanda.

En la figura 1 se muestra la gráfica de frecuencia de la variable seguridad ascendiendo donde se observa que haciendo uso de las medidas de contingencia 88 personas se sentían seguras, 33 personas muy seguras, 130 personas se

sentían normal, 71 personas muy inseguras y 125 personas se sentían inseguras.

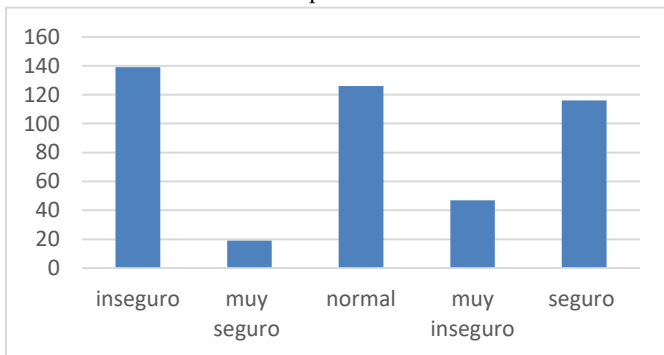
Figura 1. Frecuencia de la variable nivel de seguridad ascendiendo.



Fuente: Elaboración propia.

En la figura 2 se muestra la gráfica de frecuencia de la variable seguridad esperando donde se observa que haciendo uso de las medidas de contingencia 139 personas se sentían inseguras, 19 personas muy seguras, 126 personas se sentían normal, 47 personas muy inseguras y 116 personas se sentían seguras.

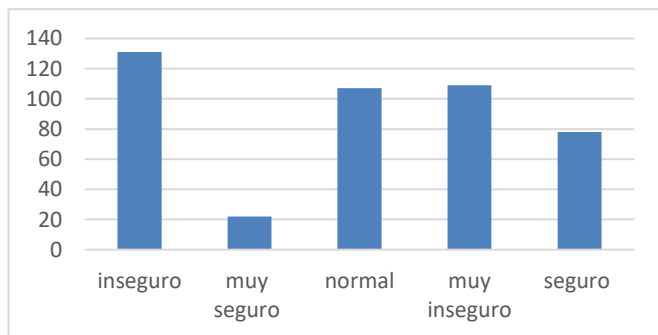
Figura 2. Frecuencia de la variable nivel de seguridad esperando.



Fuente: Elaboración propia.

En la figura 3 se muestra la gráfica de frecuencia de la variable seguridad abordando donde se observa que haciendo uso de las medidas de contingencia 131 personas se sentían inseguras, 22 personas muy seguras, 107 personas se sentían normal, 109 personas muy inseguras y 78 personas se sentían seguras.

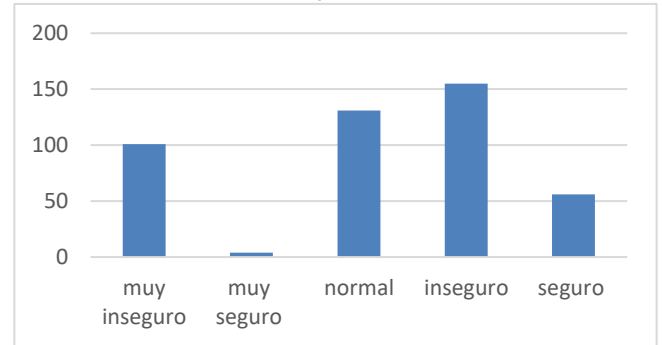
Figura 3. Frecuencia de la variable nivel de seguridad abordando.



Fuente: Elaboración propia.

En la figura 4 se muestra la gráfica de frecuencia de la variable seguridad general donde se observa que haciendo uso de las medidas de contingencia 101 personas se sentían muy inseguras, 4 personas muy seguras, 131 personas se sentían normal, 155 personas inseguras y 56 personas se sentían seguras.

Figura 4. Frecuencia de la variable nivel de seguridad general.



Fuente: Elaboración propia.

2.4 Clasificación

Los datos seleccionados eran nominales, por lo que se optó para utilizar de los algoritmos disponibles el J48, este algoritmo nos permite obtener resultados a partir de un proceso de aprendizaje supervisado, además, el proceso por medio del cual se encuentran propiedades comunes entre un conjunto de objetos de una base de datos y los cataloga en diferentes clases, de acuerdo con el modelo de clasificación [11].

Es un método de aprendizaje supervisado que construye árboles de decisión a partir de un conjunto de casos o ejemplos denominados conjunto de entrenamiento extraídos de la base de datos. También se escoge un conjunto de prueba, cuyas características son conocidas, con el fin de evaluar el árbol.

Para la ejecución de los algoritmos se utilizó la configuración por defecto que tiene WEKA, así como también se utilizó el algoritmo apriori el cual nos permite encontrar eficientemente conjuntos de ítems frecuentes los cuales sirven para obtener reglas de asociación.

2.5 Resultados

En las pruebas realizadas en la etapa de minería de datos con la tarea de clasificación y con las configuraciones por defecto que arroja el software WEKA el árbol de decisión obtenido tiene como nodo raíz la variable seguridad a bordo la cual da 5 posibles rutas, si la percepción de seguridad a bordo es igual a inseguro, la seguridad ascendiendo es igual a seguro y la seguridad esperando es inseguro entonces la percepción de seguridad en el transporte público es inseguro, de 3 casos en 1 se cumple, pero si la seguridad esperando es igual a normal la percepción de seguridad en el transporte público es normal, de 5 casos en 1 se cumple, si la seguridad esperando es igual

a seguro entonces la percepción de seguridad en el transporte público es normal, de 6 casos en 3 se cumple. Si la variable seguridad ascendiendo es igual a muy inseguro entonces la percepción de seguridad en el transporte público es muy inseguro, de 2 casos en 1 se cumple, pero si la seguridad ascendiendo es igual a normal entonces la percepción de seguridad en el transporte público es normal, de 35 casos en 17 se cumple, pero si la seguridad ascendiendo es igual a muy inseguro entonces la percepción de seguridad en el transporte público es muy inseguro, de 11 casos en 5 se cumple, si la seguridad ascendiendo es igual a inseguro la percepción de seguridad en el transporte público es inseguro, de 69 casos en 14 se cumple. Si la seguridad a bordo es igual a muy seguro y la seguridad esperando es igual a muy seguro la percepción de seguridad en el transporte público es seguro, de 15 casos en 10 se cumple. Si la seguridad esperando es igual a normal la percepción de seguridad en el transporte público es normal, en 2 casos se cumple. Si la seguridad esperando es igual a seguro entonces la percepción de seguridad en el transporte público es seguro, de 4 casos 1 se cumplió. Si la seguridad abordó es igual a normal entonces la percepción de seguridad en el transporte público es igual a normal, de 107 casos en 37 se cumplió, si la seguridad a bordo es muy inseguro entonces la percepción de seguridad en el transporte público es muy inseguro, de 109 casos en 36 se cumplió. Si la seguridad a bordo es igual a seguro y la seguridad esperando es igual a inseguro entonces la percepción de seguridad en el transporte público es inseguro, de 7 casos en 4 se cumplió, pero si la seguridad esperando es igual a muy seguro entonces la percepción de seguridad en el transporte público es normal, de 3 casos en 1 se cumplió, si la seguridad esperando es igual a normal entonces la percepción de seguridad en el transporte público es igual a normal, de 9 casos en 3 se cumplió, si la seguridad esperando es igual a seguro entonces la percepción de seguridad en el transporte público es seguro, de 59 casos en 25 se cumplió (tabla 2).

Tabla 2. Resultados obtenidos con el algoritmo J48.

Seguridad_abordo = inseguro
seguridad_ascendiendo = seguro
seguridad_esperando = inseguro: inseguro (3.0/1.0)
seguridad_esperando = Mseguro: normal (0.0)
seguridad_esperando = normal: normal (5.0/1.0)
seguridad_esperando = Minseguro: normal (0.0)
seguridad_esperando = seguro: normal (6.0/3.0)
seguridad_ascendiendo = Mseguro: Minseguro (2.0/1.0)
seguridad_ascendiendo = normal: normal (35.0/17.0)
seguridad_ascendiendo = Minseguro: Minseguro (11.0/5.0)
seguridad_ascendiendo = inseguro: inseguro (69.0/14.0)
seguridad_abordo = Mseguro
seguridad_esperando = inseguro: inseguro (1.0)
seguridad_esperando = Mseguro: seguro (15.0/10.0)
seguridad_esperando = normal: normal (2.0)
seguridad_esperando = Minseguro: seguro (0.0)
seguridad_esperando = seguro: seguro (4.0/1.0)
seguridad_abordo = normal: normal (107.0/37.0)
seguridad_abordo = Minseguro: Minseguro (109.0/36.0)
seguridad_abordo = seguro
seguridad_esperando = inseguro: inseguro (7.0/4.0)
seguridad_esperando = Mseguro: normal (3.0/1.0)
seguridad_esperando = normal: normal (9.0/3.0)
seguridad_esperando = Minseguro: seguro (0.0)
seguridad_esperando = seguro: seguro (59.0/25.0)

Fuente: Elaboración propia.

En la tabla 3 se muestran las reglas obtenidas con el algoritmo apriori aplicado para identificar el nivel de percepción de seguridad en el transporte público por parte de los usuarios. Como primera regla tenemos que, en 51 casos la seguridad ascendiendo es igual a muy inseguro y la seguridad en general entonces la percepción de seguridad a bordo del transporte público es muy inseguro para 45 de los casos, con una confianza de 0.88. La siguiente regla nos dice que en 68 casos la seguridad ascendiendo es igual a normal y la seguridad a bordo es igual a normal entonces la percepción de seguridad general en el transporte público es normal en 55 de los casos, con una confianza de 0.81. Como tercera regla tenemos que en 62 casos la seguridad esperando es igual a normal y la seguridad a bordo es igual a normal entonces la percepción de seguridad ascendiendo al transporte público es normal en 50 de los casos, con una confianza de 0.81. La cuarta regla nos dice que en 71 casos donde la seguridad ascendiendo es igual a muy inseguro entonces la percepción de seguridad a bordo del transporte público es muy inseguro en 57 de los casos, con una confianza de 0.8. En la quinta regla encontramos que en 69 casos donde la seguridad ascendiendo es igual a inseguro y la seguridad a bordo es igual a inseguro entonces la percepción de seguridad general en el transporte público es inseguro en 55 de los casos, con una confianza de 0.8. Para la sexta regla encontramos que en 76 casos donde la seguridad esperando es igual a normal y la seguridad general es igual a normal entonces la percepción de seguridad ascendiendo al transporte público es normal en 60 de los casos, con una confianza de 0.79. La séptima regla nos dice que en 57 casos donde la seguridad ascendiendo es igual a muy inseguro y la seguridad a bordo es muy inseguro entonces la percepción de seguridad general en el transporte público es muy inseguro en 45 de los casos, con una confianza de 0.79. En la octava regla tenemos que en 70 casos donde la seguridad a bordo es igual a normal y la seguridad general es igual a normal entonces la percepción de seguridad ascendiendo es normal en 55 de los casos, con una confianza de 0.79. En la penúltima regla tenemos que en 78 casos donde la seguridad esperando es igual a normal y la seguridad ascendiendo es igual a normal entonces la percepción de seguridad general es normal en 60 de los casos, con una confianza de 0.77. Como última regla tenemos que en 62 casos donde la seguridad esperando es igual a normal y la seguridad a bordo es igual a normal entonces la percepción de seguridad general es normal en 47 de los casos, con una confianza de 0.76.

Tabla 3. Reglas obtenidas con el algoritmo apriori.

1. seguridad_ascendiendo=Minseguro seguridad_general=Minseguro 51 ==> seguridad_abordo=Minseguro 45 <conf:(0.88)>
2. seguridad_ascendiendo=normal seguridad_abordo=normal 68 ==> seguridad_general=normal 55 <conf:(0.81)>
3. seguridad_esperando=normal seguridad_abordo=normal 62 ==> seguridad_ascendiendo=normal 50 <conf:(0.81)>
4. seguridad_ascendiendo=Minseguro 71 ==> seguridad_abordo=Minseguro 57 <conf:(0.8)>

5. seguridad_ascendiendo=inseguro seguridad_abordo=inseguro 69 ==> seguridad_general=inseguro 55 <conf:(0.8)>
6. seguridad_esperando=normal seguridad_general=normal 76 ==> seguridad_ascendiendo=normal 60 <conf:(0.79)>
7. seguridad_ascendiendo=Minseguro seguridad_abordo=Minseguro 57 ==> seguridad_general=Minseguro 45 <conf:(0.79)>
8. seguridad_abordo=normal seguridad_general=normal 70 ==> seguridad_ascendiendo=normal 55 <conf:(0.79)>
9. seguridad_esperando=normal seguridad_ascendiendo=normal 78 ==> seguridad_general=normal 60 <conf:(0.77)>
10. seguridad_esperando=normal seguridad_abordo=normal 62 ==> seguridad_general=normal 47 <conf:(0.76)>

Fuente: Elaboración propia.

3. CONCLUSIONES

Con base en los resultados obtenidos por el algoritmo J48 y el algoritmo apriori los cuales coincidieron, se concluye que la percepción de seguridad en el uso del transporte público con las medidas de contingencia relacionadas a la COVID-19 en el Estado de México no solía cambiar desde la primera sensación del usuario, si se sentía muy inseguro al estar esperando el transporte o ascendiendo al mismo su nivel de seguridad no cambiaba durante todo el trayecto, siendo esto una de las causas que influyo en el descenso que sufrió el transporte público en México durante la contingencia COVID-19, lo cual coincide con los resultados obtenidos en el artículo observando el transporte público desde los ojos de sus usuarios: Ciudad de México [12].

Los algoritmos J48 y apriori resultaron ser eficientes de acuerdo con la investigación, ya que clasificaron correctamente los datos. Cabe resaltar la importancia de la minería de datos en este tipo de estudios, ya que ayuda a analizar y encontrar comportamiento o patrones en los datos los cuales en este caso nos permitieron saber cuál era el nivel de percepción de seguridad de las personas al usar el transporte público durante la contingencia COVID-19.

4. REFERENCIAS

[1] Guzmán, C. S. (2021). 'Línea de tiempo COVID-19'; a un año del primer caso en México. Noticias| Capital 21. <https://www.capital21.cdmx.gob.mx/noticias/?p=12574>

[2] Instituto Mexicano del Transporte, I. M. T. (s/f). Impacto del COVID-19 en el transporte público. gob.mx. Recuperado el 5 de enero de 2023, de <https://www.gob.mx/imt/articulos/impacto-del-covid-19-en-el-transporte-publico>

[3] Asquini, A., & Asquini, A., Barcelona, España, acogerá la próxima 'UITP Global Public Transport Summit', organizada por la Unión Internacional de Transporte Público (UITP). Descarga el folleto del evento. Mobilitas. <https://mobilitas.lat/es/2020/11/22/transporte-colectivo-reconquista-de-la-confianza-y-recuperacion-da-la-demanda>

gpor-roberto-sganzerla-especialista-internacional-en-marketing-de-transporte-y-movilidad-urbana/.

[4] Laudon Kennet, Sistema de Información Gerencial.Administración de la empresa digital, Pearson Educación, México, Octava Edición, 2004.

[5]. A Rosado Gómez, A. A., y Verjel Ibáñez, A. (2015). Minería de datos aplicada a la demanda del transporte aéreo en Ocaña, Norte de Santander. Revista Tecnura, 19(45), 101-113. doi: 10.14483/udistrital.jour.tecnura.2015.3. a08.

[6] Torres-Quezada (2021). Minería de datos para determinar los factores más influyentes en la ocurrencia de siniestros de tránsito en Ecuador en el año 2020. CEDAMAZ, 11(2), 124–132. <https://doi.org/10.54753/cedamaz.v11i2.1181>.

[7] Carlos, J. y Cristóbal, R. (2014). Técnicas de Data Mining Aplicadas a Datos de Transporte Público. Edu.ar,http://sedici.unlp.edu.ar/bitstream/handle/10915/41184/Documento_completo.pdf?sequence=1&isAllowed=y.

[8] Hernández, J., Ramírez, M. J. y Ferri, C. (2004). Introducción a la minería de datos. Madrid: Pearson-Prentice Hall.

[9] Olson, D., & Delen, D. (2008). Advanced Data Mining Techniques. doi:<http://doi.org/10.1007/978-3-540-76917-0>

Villalobos J. Ángel, Expediente-e. Telemedicina, "Minería de Datos" y apoyo por Teléfono", México, 2009.

[10] Gallardo, J. (2009). Metodología para el desarrollo de proyectos en minería de datos CRISP-DM. Disponible enhttp://www.oldemarrodriguez.com/yahoo_site_admin/assets/docs/Documento_CRISP-DM.2385037.

[11] Agrawal, R., Ghosh S., Imielinski, T., Iyer, B. y Swami, A. (1992). An Interval Classifier for Database Mining Applications. Proceedings VDLB Conference, Vancouver.

[12] Observando el transporte público desde los ojos de sus usuarios: Ciudad de México. (2021). Caf.com. Recuperado el 12 de enero de 2023, de <https://www.caf.com/es/conocimiento/visiones/2021/09/observando-el-transporte-publico-desde-los-ojos-de-sus-usuarios-ciudad-de-mexico/>.