

Minería de Datos en el Análisis de la Factibilidad de la Capacidad de Servicio: Caso Centro de Verificación

Data Mining in Service Capacity Feasibility Analysis: Case Verification Center

Mariel Juárez-Ortuño¹, María Eva Juárez-Ortuño¹,
Nayeli Montalvo-Romero², Aarón Montiel-Rosales^{1,*}

¹Tecnológico Nacional de México / ITS de Teziutlán

²Tecnológico Nacional de México / ITS de Misantla

* **Corresponding author:** ar.chimedes@hotmail.com

Resumen: El tiempo de servicio en los centros de verificación vehicular que como ente evalúan el cumplimiento de normas oficiales mexicanas, de conformidad con las emisiones contaminantes, es importante. Tener la capacidad adecuada para prestar el servicio se ve directamente reflejado en la satisfacción del cliente; al mismo tiempo, en que se disminuyen inconformidades y costos de oportunidad. Actualmente, la complejidad de los sistemas y del gran volumen de información que se genera hacen necesario el procesamiento de datos masivos para identificar patrones de comportamiento, tendencias y relaciones entre variables. En este estudio, con la finalidad de evaluar la apertura de una nueva línea de servicio de verificación vehicular se presenta la exploración de datos históricos como herramienta del proceso de toma de decisiones. Orange *Data Mining* versión 3.36.2 como herramienta tecnológica de código abierto se empleó para organizar y analizar de forma objetiva la información implicada en el tiempo de servicio. Del análisis estadístico de la información se desprende la necesidad de abrir una nueva línea de atención que pueda ayudar en la distribución de la carga de trabajo, y así reducir tiempos de espera para los clientes.

Palabras clave: Toma de Decisión, Minería de Datos, Proceso KKD, Análisis de Capacidad.

Abstract: The time of service in the vehicle verification centers that as an entity evaluate compliance with official Mexican standards, in accordance with polluting emissions, is important. Having the right capacity to provide the service is directly reflected in customer satisfaction; at the same time, in which nonconformities and opportunity costs are reduced. Currently, the complexity of systems and the large volume of information that is generated make it necessary to process big data to identify patterns of behavior, trends and relationships between variables. In this study, to evaluate the opening of a new vehicle verification service line, the exploration of historical data is presented as a tool for the decision-making process. Orange *Data Mining* version 3.36.2 as an open-source technological tool was used to objectively organize and analyze the information involved in the service time. The statistical analysis of the information shows the need to open a new service line that can help in the distribution of the workload, and thus reduce waiting times for customers.

Keywords: Decision Making, Data Mining, KKD Process, Capacity Analysis.

1 Introducción

La capacidad de servicio al cliente es un parámetro que permite evaluar la eficiencia de un sistema. Hanukov (2022), establece que las capacidades de gestión y organización fortalecen la calidad del servicio y las capacidades de las organizaciones. En (Bin Taliah & Zervopoulos, 2023), se menciona que el sector servicios representa aproximadamente dos tercios de la producción mundial.

El concepto de servicio ha evolucionado, desde un enfoque de valor material a un valor centrado en el cliente (Kim & Lee, 2024). Por lo que, hoy en día la satisfacción del cliente va más allá de cumplir con el servicio que se le presta, sino es un esquema que genera competitividad organizacional a partir del valor agregado del servicio, lo que conlleva al cumplimiento excedente de las expectativas del cliente.

El conocimiento en el comportamiento del tiempo de servicio en los centros de verificación vehicular se considera importante, así como todo tiempo de servicio dentro de un sistema. Debido a que, con esta información es posible realizar una programación eficiente de los recursos y así cumplir en tiempo y forma con el tiempo de servicio.

La verificación vehicular tiene por objeto, mantener el parque vehicular en circulación en la mejor condición ambiental posible, motivando el mantenimiento preventivo de aquellas unidades que se presentan a verificar, y obligar el mantenimiento correctivo de los vehículos que presenten niveles de contaminación mayores a lo permisible (Puebla, 2024); siendo de este modo, el derecho al medio ambiente sano y la calidad del aire de los habitantes.

El análisis del tiempo de servicio en un centro de verificación vehicular es complejo por ser una variable que tiene correlación con múltiples factores. Para poder analizar y evaluar, la capacidad de servicio e identificar la necesidad de la apertura de una línea de atención, este estudio emplea la exploración de datos mediante minería de datos como herramienta de apoyo a la decisión. La minería de datos permite explorar un conjunto de datos, para identificar patrones de comportamiento y con ello asistir al proceso de toma de decisiones (Jinghua & Haiying, 2023; Klimek & Klimek, 2020; Zhang & Song, 2024).

En este estudio se analiza mediante minería de datos, el comportamiento del tiempo de

servicio en un centro de verificación vehicular ubicado en el estado de Puebla-México; con la finalidad de analizar el comportamiento del tiempo de servicio y con ello la capacidad del servicio de verificación vehicular, y así establecer la viabilidad de habilitación de una nueva línea de atención.

Para esto, se analiza una base de datos conformada por 2,606 registros correspondientes a un trimestre; de los cuales, se analizan las variables: hora de recepción, hora de entrega, tiempo real de atención al cliente, multa, entidad, tipo de pago, tipo de verificación entregada, línea, modelo, marca, tipo de combustible, tipo de servicio del vehículo y, tipo de prueba.

2 Metodología

El proceso de minería de datos implica la extracción de información valiosa a partir de grandes volúmenes de datos.

Desde la selección de las fuentes de información hasta la interpretación de los resultados, este proceso comprende un exhaustivo análisis de datos; el cual implica tres pasos cruciales: primero, limpiar y preparar los datos; segundo, aplicar algoritmos o modelos; y tercero, interpretar los resultados obtenidos.

Hay múltiples maneras de abordar el tema que apoyan este proceso, agrupados según los objetivos del análisis. En el caso de la aplicación de minería de datos, los datos abiertos relacionados con la verificación de vehículos en un verificentro, sigue una secuencia metodológica basada en el proceso KDD (*Knowledge Discovery in Databases*). El KDD, implica la extracción de información, la preparación de los datos y la interpretación de los resultados obtenidos (Shu & Ye, 2023).

Para lograrlo es necesario habilidades en: aprendizaje automático, estadísticas, manejo de bases de datos, representación de conocimiento, razonamiento basado en casos, razonamiento aproximado, adquisición de conocimientos, uso de redes neuronales y análisis de datos. Tareas comunes en KDD son la inducción de reglas, los problemas de clasificación y *clustering*, el reconocimiento de patrones, el modelado predictivo, y la detección de dependencias. (Joyanes Aguilar, 2019).

El proceso de aplicación de KDD, en esta investigación sigue una secuencia ordenada, interactiva e iterativa de pasos, que incluye la

detección de patrones o tendencias en los datos, como la hora pico de atención, modelos de vehículo más común, tipo de licencia con mayor comercialidad, entre otros factores (Timarán, et al., 2016).

La Figura 1, describe la interrelación entre las etapas del KDD.

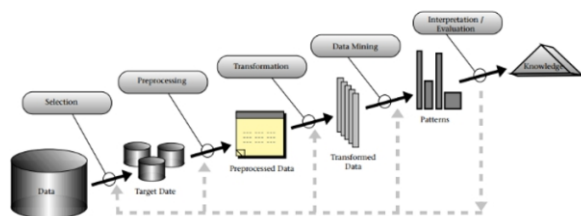


Figura 1: Visión general de los pasos que componen el proceso KDD

Figura 1. Etapas del Proceso de KDD.

Fuente. (Joyanes Aguilar, 2019).

El proceso de extracción/descubrimiento del conocimiento en base de datos, se entrega de las siguientes fases:

- Etapa 1. Selección.
- Etapa 2. Preprocesamiento.
- Etapa 3. Transformación.
- Etapa 4. Minería de Datos.
- Etapa 5. Interpretación/Evaluación.

En el método KDD, es importante la selección de datos que se emplearan para el procesamiento, de parte de analista; debe haber una comprensión clara del analista. Las etapas en el proceso global del KDD no están claramente diferenciadas, al ser un proceso reiterado e interactivo entre el analista y el tomador de decisiones (Joyanes Aguilar, 2019).

Las diferencias entre KDD y minería de datos, radica principalmente en que, en la minería de datos es una etapa particular del proceso KDD, donde la minería de datos aplica algoritmos o técnicas particulares para extraer patrones de datos que se utilizan en la minería de datos, que es la fase más crítica del KDD. Esta fase involucra el uso de métodos estadísticos y de aprendizaje para generar hipótesis de patrones y modelos. De modo informal se asocia minería de Datos con KDD (Joyanes Aguilar, 2019).

El enfoque empleado, permitirá evaluar las diferentes variables y sus interrelaciones; para establecer con ello, la viabilidad de apertura de una nueva línea para seguir brindando un mejor servicio a los clientes del centro de verificación vehicular bajo estudio.

3 Resultados

En esta sección, se describen los resultados obtenidos de la analítica de datos.

3.1 Aplicación del Proceso de Minería de Datos en CVV

La aplicación de minería de datos en las fuentes de datos abiertos extraídas del CVV (Centro de Verificación Vehicular), ubicado en un municipio del estado de Puebla-México, se desarrolló considerando las cinco fases del proceso KDD, las cuales incluyen:

Etapa 1. Selección

En el comienzo del proceso, se identifican las fuentes de información y se selecciona el tipo de datos que se empleará.

Se analizó la base de datos que cuenta con 2,606 registros correspondientes a un trimestre y se obtuvieron las principales variables: hora de recepción, hora de entrega, tiempo real de atención al cliente, multa, entidad, tipo de pago, tipo de verificación entregada, línea, modelo, marca, tipo de combustible, tipo de servicio del vehículo, y tipo de prueba.

Etapa 2. Preprocesamiento

Durante esta fase, se lleva a cabo el proceso de organización y purificación de la información obtenida de diversas fuentes de datos. Esta labor es esencial para poder trabajar con los datos de manera efectiva en las etapas subsiguiente. En esta etapa se utilizan diversas estrategias para manejar datos faltantes o ausentes, en blanco, datos inconsistentes o que están fuera de rango, obteniéndose al final una estructura de datos adecuada para posterior transformación (Joyanes Aguilar, 2019).

En este sentido, se llevó a cabo la limpieza de los archivos de origen utilizando el programa Microsoft Excel, eliminando la información que no resultaba relevante para el análisis posterior.

Etapa 3. Transformación

El proceso implica la preparación inicial de los datos disponibles, la conversión y creación de variables adicionales basadas en las ya existentes, todo dentro de una estructura de datos adecuada. Aquí se realizan operaciones de agregación o normalización, consolidando los datos de una forma necesaria para la fase siguiente (Joyanes Aguilar, 2019).

La implementación de la teoría de colas en la atención al cliente es una estrategia muy relevante para abordar la gestión de tiempos de espera, y la distribución eficiente de recursos en centros vehiculares y otros entornos de servicio. La teoría de colas se centra en el estudio matemático de la formación y el comportamiento de las líneas de espera, lo que permite diseñar sistemas que minimicen los tiempos de espera y maximicen la eficiencia del servicio. Se utiliza para modelar y analizar situaciones en las que llegan clientes o entidades a un servicio, esperando ser atendidos y luego salen del sistema.

Al aplicar la teoría de colas, es posible determinar el número óptimo de servidores (personal atendiendo a los clientes), analizar los patrones y tiempos de llegada de los clientes, así como establecer estrategias para gestionar las colas y los tiempos de espera. Esto puede incluir la asignación dinámica de recursos, la implementación de sistemas de turnos o citas, y el diseño de procesos que minimicen la variabilidad en los tiempos de atención.

Para la transformación de información los datos se someten a observación en el sistema Orange *Data Mining* versión 3.36.2; basado en componentes para la minería de datos y análisis predictivo, dicho software es libre y multiplataforma, los flujos de trabajo son mediante *widgets*, agrupados según su funcionalidad.

Las ventajas de dicho software son: una programación visual y versátil para el análisis de datos, es amigable para el usuario y multiplataforma (Burbano Pantoja, et al., 2018).

Etapa 4. Minería de Datos

Es la fase de modelado propiamente dicho, en donde métodos inteligentes son aplicados con el objetivo de extraer patrones previamente desconocidos, válidos, nuevos, potencialmente útiles y comprensibles, y que están contenidos u “ocultos” en los datos (Joyanes Aguilar, 2019).

Después de transformar y depurar los datos de cada fuente, el siguiente paso consistió en identificar el modelo de minería que se adecuara al tema seleccionado, situándose dentro del proceso supervisado o predictivo. En esta línea, el proceso implica predecir datos mediante la inducción de una relación entre una variable objetivo y una serie de variables adicionales (Rodríguez & Díaz, 2009). En otras palabras, se busca identificar la variable más influyente en el

problema sin comprometer la calidad del modelo. Las variables más influyentes son:

x_1 = **Hora de Recepción**, representa la hora en que son ingresados los datos del vehículo a verificar.

x_2 = **Hora de Entrega**, muestra la hora en que se les entrega a los usuarios el resultado de la prueba de emisiones realizada a sus vehículos.

x_3 = **Tiempo Real de Atención al Cliente**, tiempo que se hace el personal del verificentro en realizar el proceso de verificación vehicular desde que llega el vehículo hasta que se entrega el resultado de la prueba el cual está establecido en 8 minutos por vehículo.

x_4 = **Multa**, importe pagado por no verificar en tiempo y forma su vehículo, de acuerdo con su color de engomado y terminación del último dígito de su placa.

x_5 = **Entidad**, es el estado donde se encuentran registradas las placas del vehículo que se verifica.

x_6 = **Tipo de Pago**, muestra si el pago de la verificación fue al contado o con tarjeta de crédito o débito.

x_7 = **Tipo de Verificación Entregada**, es el holograma (00, 0, 1, 2) y que por año de modelo del auto se le entrega al usuario después de la prueba de emisiones; la cual puede ser aprobada o reprobada en cuyo caso se le entrega y una constancia de no aprobado.

x_8 = **Línea**, área de un CVV que se utiliza para llevar a cabo la exploración físico-mecánicas, pruebas de Sistema de Diagnóstico a Bordo, y medición de emisiones de gases y opacidad de vehículos. Esta zona está compuesta por tres estaciones: una para la inspección físico-mecánica, otra para la prueba de SDB, y una tercera para medir las emisiones contaminantes. En este caso solo se cuenta con dos. La línea uno es exclusivo para combustible gasolina y la línea dos es una línea dual; es decir, para combustible diésel y gasolina.

x_9 = **Modelo**, año calendario que el fabricante designe al vehículo y que es consignado en la tarjeta de circulación vehicular

x_{10} = **Marca**, es el fabricante del vehículo.

x_{11} = **Tipo de Combustible**, es el combustible que requiere para que el vehículo pueda realizar su función.

x_{12} = **Tipo de Servicio del Vehículo**, es el uso que se le da al vehículo; por ejemplo: si es taxi o automóvil particular.

x_{13} = **Tipo de Prueba**, este proceso implica la evaluación de los gases emitidos por vehículos con motores que usan gasolina, gas licuado de

petróleo, gas natural y otros combustibles alternativos. Los gases medidos incluyen HC, CO, CO₂, O₂ y NO_x, y se miden mientras el vehículo está sometido a aceleraciones simuladas mediante la aplicación de una carga externa controlada por el dinamómetro. La Figura 2, presenta la integración de las 13 variables en estudio en el programa Orange.

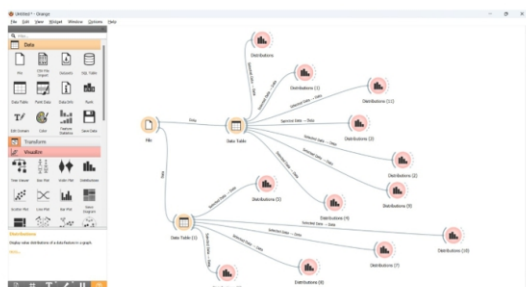


Figura 2. Flujo de Datos del CVV.

Etapla 5. Interpretación/Evaluación

Se identificaron patrones de comportamiento con lo que una evaluación de los resultados obtenidos fue determinada (Joyanes Aguilar, 2019).

Para esto, se emplean técnicas de visualización y representación del conocimiento que permiten presentar los hallazgos a los analistas (Shafique y Kaiser, 2014).

Los resultados de la aplicación de los diversos clasificadores a los datos de entrenamiento de cada fuente de datos son presentados, evaluados y analizados utilizando el método *Test and Score* de Orange. A partir de esta evaluación, se procede a interpretar los resultados obtenidos (ver Figura 3).

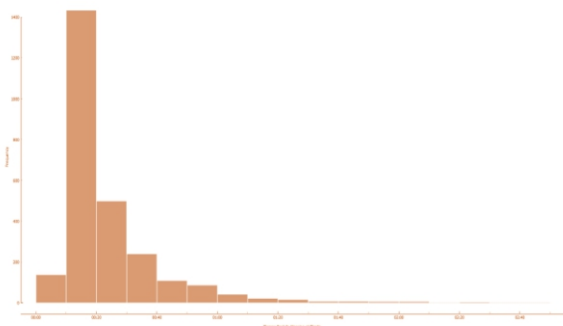


Figura 3. Tiempo Real de Atención al Cliente.

La base de datos con 2,606 registros permitió identificar el no cumpliendo con los 8 minutos que debe tardar todo el procedimiento de la verificación desde el momento en que le reciben los documentos en la recepción hasta la entrega del resultado.

Solo el 5.26% (137) de los vehículos verificaron en un promedio de 10 minutos, el 54.99% (1433) en un promedio de 20 minutos, el 19.11% (498) en 30 minutos, el 9.17% (239) en 40 minutos, el 4.14% (108) en 50 minutos, el 3.30% (86) en una hora y un 4.03 (105) verifico en más de una hora hasta un máximo de 2 horas 40 minutos.

El 62.01% de las verificaciones fueron sin pago de multa y el 37.99% con multa, lo que indica que más de la mitad de la muestra sí cumple en tiempo y forma su calendario de verificación anual (véase Figura 4).

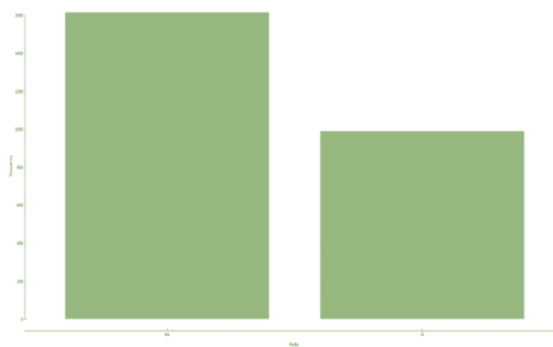


Figura 4. Comportamiento de la Variable Multa.

El 95.51% del parque vehicular es del estado de Puebla y de manera voluntaria vehículos con placas del estado de México, la CDMX, Tlaxcala, Veracruz, entre otras entidades, son solo un 4.49% (ver Figura 5).

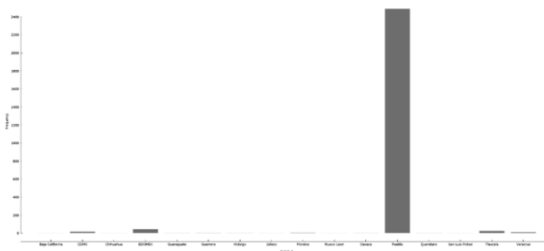


Figura 5. Comportamiento de la Variable Entidad.

En la variable de tipo de pago, se observa que de un total de 2,606 registros se obtuvo porcentaje de un 89.52% de pagos que se realizaron de contado, también se aprecia que el 6.72% fue de los pagos fueron con tarjeta de débito y el 3.76% restante corresponde a pagos con tarjeta de crédito (ver Figura 6).

De acuerdo con la variable Tipo Verificación entregada, se destacan los siguientes parámetros, de las pruebas realizadas durante el mes enero 2024, se tiene que el 78.97% corresponde a verificaciones aprobadas, de las cuales, se desprende que el 66% son de hologramas tipo 0,

el 8.40% son hologramas tipo 1, el 3.95% es de hologramas tipo 00 y el 0.61% de hologramas tipo 2, lo que indica que solo se tiene un porcentaje del 21.03% de verificaciones no aprobadas (véase Figura 7).



Figura 6. Comportamiento de la Variable Tipo de Pago.

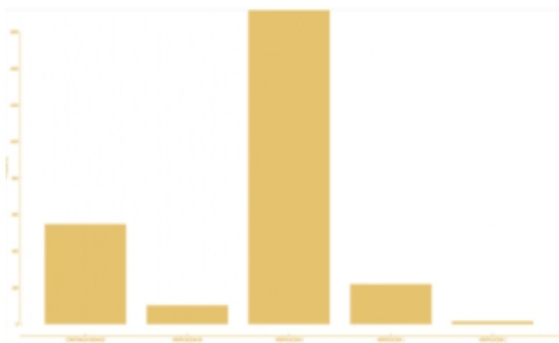


Figura 7. Comportamiento de la Variable Tipo de verificación Entregada.

De esta variable se desprende que en la línea uno se realizaron 1,434 pruebas de verificación, que corresponden a un 55.03% ya que es únicamente para vehículos que utilizan para su función gasolina, en comparación de la línea dos que sólo realizó 1,172 pruebas que representan 44.97% debido a que se trata de una línea dual; es decir, realiza pruebas de diésel y gasolina, lo anterior genera la diferencia, esto debido a que en su mayoría de los vehículos son de combustible gasolina, los diésel únicamente se atienden en la línea dos (ver Figura 8).

La Figura 9, proporciona un panorama del año modelo de los vehículos que han iniciado a verificar el 1er semestre de 2024, y se observa que la tendencia es en los modelos de los años 2015 a 2020 con un 32.46%, seguido de los modelos entre los años 2010 a 2015 con un 19.53%, continuando con los modelos 2005 a 2010 con un 15.31%, en la siguiente posición se ubican los modelos más recientes que van de 2020 a la fecha con un 15% quedando en siguiente termino los modelos del 2000 al 2005

con solo un 11.01% y al final con un 6.69% están los modelos más clásicos que son de 1980 a los 1999.



Figura 8. Comportamiento de la Variable Línea.

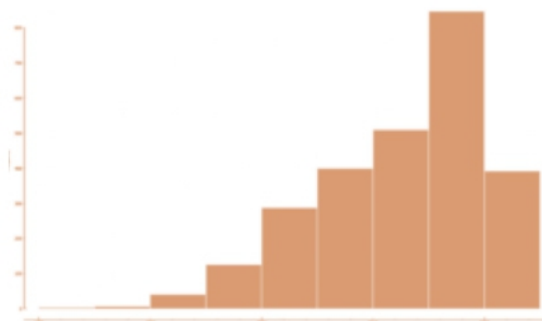


Figura 9. Comportamiento de la Variable Modelo.

La Figura 10, presenta la gráfica de la variable Marca; la cual, describe los vehículos que han se han presentado a verificar en el mes de enero de 2024, describiendo a continuación a cada una de las marcas existentes en el mercado, resaltando que las marcas preferidas por los usuarios son, Chevrolet con el 15.16%, Ford 10.40%, Nissan 30.66% y VW18.99%.

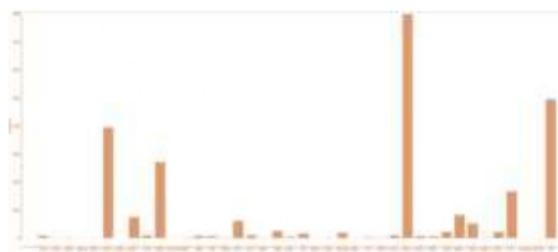


Figura 10. Comportamiento de la Variable Marca.

La Figura 11, presenta el comportamiento de la variable Tipo de Combustible; esta variable confirma el hallazgo de la variable de Línea, el 98.08% de los vehículos son de gasolina, el 1.57% de diésel y el 0.35% de gas natural.

Los datos de la variable Tipo de Servicio del Vehículo, permite identificar los vehículos que se clasifican de acuerdo con el tipo del uso que

se requieren usar, de los cuales se presentan a realizar su verificación, son de tipo servicio particular en un 70.76%, seguido del tipo servicio mercantil con un 20.72% y por último el tipo servicio público en un 8.52% (ver Figura 12).



Figura 11. Comportamiento de la Variable Tipo de Combustible.



Figura 12. Comportamiento de la Variable Tipo de Servicio del Vehículo.

La Figura 13, describe el comportamiento de los diversos tipos de prueba que se realizan en el verificentro, bajo la simulación de aceleraciones y el uso de una carga externa controlada por un dinamómetro, se descubrió que el tipo de prueba Dinámica produce resultados favorables (auto con movimiento) se tienen 2,478 pruebas con un 95.09% y en la prueba de modo Estático (auto encendido en punto muerto) se obtienen 128 pruebas siendo esto solo un 4.91%, de un total de 2,606 datos que corresponde a un 100%.

4 Discusión

En función de los resultados obtenidos empleando teoría de colas, mediante KDD y minería de datos se ha identificado que, el límite de tiempo en la atención al cliente, de 8 minutos con una holgura de 10 minutos establecido por el CVV no cumple, se está brindando el servicio con un tiempo superior al establecido. Los datos

recopilados y analizados son importantes en la estimación del tiempo promedio de espera; en la evaluación del tiempo de servicio actual y de su cumplimiento con el sistema.



Figura 13. Comportamiento de la Variable Tipo de Prueba.

La evaluación del cumplimiento de los tiempos promedio de espera y las variaciones significativas en la distribución de tiempos, evidenciaron la necesidad de incrementar una línea más de atención con la finalidad de reducir tiempos.

El estudio realizado mediante minería de datos permitió calcular el tiempo promedio de espera, el tiempo promedio de atención y la distribución de actividades realizadas. Se observa que nuestra variable principal es el tiempo de servicio, mismo que permite establecer parámetros de observación; con lo que es posible establecer la base para el desarrollo de un manual de operaciones.

5 Conclusiones

Después del análisis de datos realizado al tiempo de espera en el verificentro, se establece que el sistema de atención al cliente; no cumple con los estándares establecidos, el tiempo promedio de espera no se encuentra dentro del rango objetivo, teniendo en cuenta la holgura ya que se realiza en un promedio de 20 minutos. Con esta información es posible establecer mejoras en la infraestructura para brindar mayor atención, y así hacer más rápida y eficaz la atención al cliente.

Generalmente la Teoría de Colas o Línea de Espera, es el enfoque empleado para evaluar el comportamiento de un servidor, y así determinar la eficiencia en el tiempo de servicio; sin embargo, no es el único enfoque que puede ser empleado. En este estudio se adoptó la minería de datos como estrategia para identificar

patrones de comportamiento de variables que influyen en el tiempo de servicio en un verificentro vehicular, y así determinar si la capacidad instalada era suficiente o no. Para este caso se evidencio que la capacidad instalada, no cumplía con la capacidad demanda; por lo que, una nueva línea de servicio es necesaria.

Con el desarrollo de este estudio, es posible establecer que la minería de datos como herramienta de apoyo al proceso de toma de decisiones, permite asistir al decisor al analizar e identificar patrones de comportamiento en grandes volúmenes de información.

6 Referencias

- Bin Taliah, A. R., & Zervopoulos, P. D. (2023). Using a Bayesian meta-frontier approach to evaluate airline performance: The nexus between airline efficiency and customer satisfaction. *Heliyon*, 9(10), e20391. <https://doi.org/10.1016/J.HELIYON.2023.E20391>
- Burbano Pantoja, V., Valdivieso Miranda, M., & Burbano Valdivieso, A. (2018). Aplicaciones de la teoría de colas y líneas de espera en contextos específicos de investigación. *Editorial UPTC*. <https://doi.org/https://elibro.net/es/lc/its/teziutlan/titulos/132820>
- Hanukov, G. (2022). Improving efficiency of service systems by performing a part of the service without the customer's presence. *European Journal of Operational Research*, 302(2), 606–620. <https://doi.org/10.1016/J.EJOR.2022.01.045>
- Jinghua, Z., & Haiying, R. (2023). Multi-attribute decision-making based on data mining under a dynamic hybrid trust network. *Computers & Industrial Engineering*, 185, 109672. <https://doi.org/10.1016/J.CIE.2023.109672>
- Joyanes Aguilar, L. (2019). *Inteligencia de Negocios y Analítica de Datos* (1ra ed.). México, México: Alfaomega.
- Kim, C., & Lee, J. (2024). Discovering patterns and trends in customer service technologies patents using large language model. *Heliyon*, 10(14), e34701. <https://doi.org/10.1016/J.HELIYON.2024.E34701>
- Klimek, J., & Klimek, J. A. (2020). IT and Data Mining in Decision-Making in the Organization. *Education Management in the Culture of Late Modernity. Procedia Computer Science*, 176, 1990–1999. <https://doi.org/10.1016/J.PROCS.2020.09.235>
- Puebla, G. d. (2024). Programa de Verificación Vehicular. *Diario Oficial de Federación*.
- Rodríguez, Y., & Díaz, A. (2009). Herramientas de minería de datos. *Revista Cubana de Ciencias Informáticas*, 3(3-4), 73-80.
- Shafique, U., & Qaiser, H. (2014). Un estudio comparativo del proceso de minería de datos Modelos sin (KDD, CRISP-DM y SE. *Revista Internacional de Innovación y Investigación científica*, 217-222.
- Shu, X., & Ye, Y. (2023). Knowledge Discovery: Methods from data mining and machine learning. *Social Science Research*, 110, 102817. <https://doi.org/10.1016/J.SSRESEARC.2022.102817>
- Timarán Pereira, S. R., Hernández Arteaga, I., Caicedo Zambrano, S. J., Hidalgo Troya, A., & Alvarado Pérez, J. C. (2016). El proceso de descubrimiento de conocimiento en bases de datos. *Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional*.
- Zhang, F., & Song, W. (2024). Product improvement in a big data environment: A novel method based on text mining and large group decision making. *Expert Systems with Applications*, 245, 123015. <https://doi.org/10.1016/J.ESWA.2023.123015>