

Uso de la minería de datos para evaluar la calidad y movilidad del transporte de estudiantes universitarios.

José Antonio Morales Salmoran^a, Javier Romero-Torres^b, Doricela Gutiérrez Cruz^c

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

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

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

Resumen

Se analiza el nivel de calidad y movilidad de un transporte estudiantil universitario, utilizando minería de datos. Los 6779 datos analizados se generaron de una encuesta origen destino y de satisfacción en relación con la calidad del transporte y el uso de las rutas de transporte universitario, la aplicación de la encuesta fue en 2024 en la Universidad Autónoma del Estado de México en la ciudad de Toluca. Las variables evaluadas fueron 12CT, 12SPBT, 9IS, 12TT, 12CTR, 7TVPB, 13IEPB, 13IVPB, 4TDP, 6MD. En el proceso de clasificación se utilizaron los algoritmos J48, en que el primero se identificó una relación entre en tiempo de viaje y la zona en la que el usuario aborda el autobús. Mientras, con el algoritmo RandomForest se observó que se tiene una tendencia a que se inicia el traslado en las zonas de la ciudad de Toluca. Los resultados pueden contribuir a una mejor planificación de rutas, a la eficiencia en la distribución de los recursos de transporte universitario, y a la mejora en la percepción de los niveles de servicio.

Palabras clave— Árboles de clasificación, Descubrimiento de Conocimiento en Bases de Datos, Minería de Datos, Movilidad, Transporte público

Abstract

The quality and mobility level of a university student transport is analyzed, using data mining. The 6779 analyzed data were generated from an origin-destination and satisfaction survey in relation to the quality of transport and the use of university transport routes. The application of the survey was in 2024 at the Autonomous University of the State of Mexico in the city of Toluca. The variables evaluated were 12CT, 12SPBT, 9IS, 12TT, 12CTR, 7TVPB, 13IEPB, 13IVPB, 4TDP, 6MD. In the classification process, the J48 algorithms were used, in which the first one identified a relationship between travel time and the area in which the user boards the bus. Meanwhile, with the RandomForest algorithm it was observed that there is a tendency for the transfer to begin in the areas of the city of Toluca. The results can contribute to better route planning, efficient distribution of university transportation resources, and improved perceptions of service levels.

Keywords: *Decision trees, Knowledge Discovery in Databases, Data Mining, Mobility, Public transportation*

1. INTRODUCCIÓN

La movilidad y logística del transporte universitario son factores determinantes en el desempeño académico y la calidad de vida de los estudiantes. Garantizar un sistema de transporte eficiente, seguro y accesible representa un desafío para las instituciones educativas y los responsables de la planificación urbana. La movilidad urbana no solo impacta la calidad de vida, sino que también influye en la gobernabilidad y la sostenibilidad ambiental. La expansión metropolitana ha incrementado la dispersión de los destinos, lo que a su vez ha diversificado las necesidades de transporte. Factores como la accesibilidad, el tiempo de traslado, los costos y la calidad del servicio son determinantes en la experiencia del usuario. En este contexto, la movilidad se posiciona como un eje central en los estudios urbanos y de transporte, reflejando la separación entre vivienda, trabajo, abasto y recreación. Esta dinámica, influenciada por la estructura urbana y las desigualdades socioeconómicas, caracteriza a las aglomeraciones urbanas [2] y genera nuevas demandas de transporte y servicios [3].

La información sobre las pautas de movilidad es esencial para la formulación de diagnósticos precisos, el desarrollo de modelos de planificación y la estimación del impacto de las medidas implementadas. [4,5] En este sentido, la digitalización emerge como una herramienta clave para mejorar la recopilación y análisis de datos sobre la demanda de transporte [6].

Las tecnologías de la información y comunicación (TIC) desempeñan un papel fundamental en la gestión eficiente de los recursos urbanos, permitiendo abordar los desafíos derivados del crecimiento poblacional. En este marco, la minería de datos se presenta como una herramienta estratégica para el análisis de la movilidad universitaria, facilitando la identificación de patrones, la optimización de rutas y la evaluación del desempeño de los servicios de transporte. La minería de datos permite procesar grandes volúmenes de información para identificar tendencias y optimizar la asignación de recursos.

A través de la metodología de Descubrimiento de Conocimiento en Bases de Datos (KDD), es posible estructurar un proceso sistemático que transforma datos complejos en conocimiento útil para la toma de decisiones estratégicas. Entre las tareas más comunes del KDD se encuentran la inducción de reglas, la clasificación, el clustering y el reconocimiento de patrones.

Para Molina, la minería de datos es un proceso de extracción de conocimiento a partir de bases de datos, cuyo objetivo es identificar patrones, anomalías, tendencias y secuencias relevantes para la toma de decisiones [7]. En este contexto, el presente estudio tiene como objetivo analizar la calidad y movilidad del transporte universitario mediante técnicas de minería de datos, con el fin de mejorar la eficiencia del servicio y la logística de transporte. Diversos estudios han demostrado la efectividad de la minería de datos y los métodos heurísticos en la optimización del transporte. Calderón y Bermeo [8] diseñaron un modelo de optimización de rutas de transporte con el objetivo de reducir tiempos y costos en los procesos de entrega. Para ello, emplearon métodos heurísticos y metaheurísticos, incluyendo algoritmos

genéticos para optimizar rutas en términos de distancia, capacidad y tiempo de traslado. Mediante el uso del software VRP Solver 1.3, lograron reducir significativamente los costos operativos y mejorar la eficiencia del transporte. De manera similar, Jaramillo-Álvarez, González-Calderón y Guillermo-González [9] propusieron un modelo de optimización de redes de autobuses mediante programación lineal entera mixta. Utilizando el software GAMS 22.7 y un método de ponderación para la combinación de múltiples funciones, diseñaron una red de transporte con 15 nodos estratégicos. Los resultados incluyeron la creación de cuatro rutas con una extensión total de 98.3 km, logrando cubrir el 77% de la demanda proyectada, reducir los costos operativos y mejorar la experiencia del usuario mediante la disminución de transbordos.

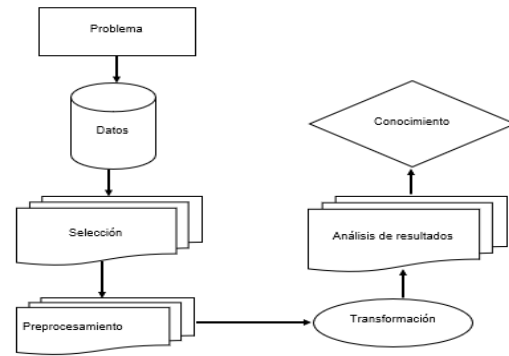
Por otro lado, Quintana Cruz, Ojeda Mesa y Almeida Franco [10] diseñaron un sistema de rutas variables basado en Sistemas de Información Geográfica (SIG) para optimizar la distribución de transporte. Utilizando ArcGIS y Network Analysis, lograron diseñar rutas óptimas considerando la capacidad de carga, tiempos de servicio y restricciones horarias. Mediante el análisis de 410 viajes y la simulación de 30 escenarios aleatorios, validaron el sistema con datos reales, demostrando mejoras en la eficiencia operativa, reducción de tiempos de viaje y ahorro en el consumo de combustible.

El presente trabajo busca analizar la movilidad y satisfacción en el transporte universitario mediante minería de datos, identificando patrones de demanda y optimizando rutas. Se evalúan factores como tiempos de traslado, accesibilidad, costos y calidad del servicio, con el fin de mejorar la movilidad urbana y la eficiencia del transporte, promoviendo soluciones sostenibles que optimicen la experiencia del usuario.

2. CONTENIDO (Materiales y Métodos)

Se analizaron un total de 6,779 registros obtenidos a partir de las respuestas de 171 participantes en una encuesta sobre movilidad y satisfacción en relación con la calidad del transporte y la movilidad urbana. Esta encuesta fue realizada en 2024 por la Universidad Autónoma del Estado de México en la ciudad de Toluca, con el propósito de evaluar las percepciones y experiencias de los usuarios en estos aspectos. La investigación adoptó un enfoque cuantitativo de tipo descriptivo. Para el tratamiento de los datos, se empleó la metodología **Knowledge Discovery in Databases (KDD)**, que comprende las fases de comprensión del problema, selección de datos, preprocesamiento, limpieza y transformación de la información, garantizando un análisis estructurado y preciso.

Fig1. Diagrama del proceso KnowledgeDiscovery in Databases



Fuente: Elaboración propia.

2.1 Comprensión del Problema

Para garantizar un análisis efectivo mediante minería de datos, es fundamental comprender con precisión el problema a resolver. Esto permite recolectar la información necesaria y asegurar una interpretación adecuada de los resultados obtenidos [11].

2.2 Selección de Datos

En esta fase, se define el conjunto de datos relevantes para el estudio, identificando los atributos que lo conforman (12CT, 12SPBT, 9IS, 12TT, 12CTR, 7TVPB, 13IEPB, 13IVPB, 4TDP, 6MD), los cuales se ajustan a los requerimientos específicos de la investigación.

2.3 Preprocesamiento y Limpieza de Datos

Este proceso implica el análisis de la cantidad y calidad de los datos, eliminando registros ruidosos, desconocidos, nulos o vacíos. Para ello, se emplean técnicas estadísticas, como la imputación por media o moda, garantizando la coherencia y fiabilidad de la información [12]. La Tabla 1 presenta la descripción de cada uno de los atributos seleccionados.

Tabla 1. Descripción de atributos basados a la encuesta sobre la calidad y movilidad del transporte universitario.

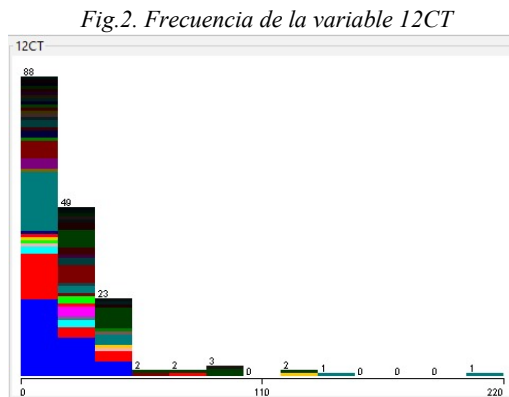
Atributo	Descripción
12CT	Costo por cada transporte utilizado.
12SPBT	Alternativa de transporte del domicilio a Centro Universitario.
9IS	Ingreso semanal ya sea por mesada, trabajo o beca.
12TT	Tiempos por cada transporte utilizado.
12CTR	Costo por cada transporte utilizado.
7TVPB	Tiempo del viaje a bordo del potrobús.
13IEPB	Importancia del tiempo de espera para abordar potrobús.
13IVPB	Importancia del tiempo de viaje a bordo del potrobús.
4TDP	Tiempo de traslado del domicilio a la parada del potrobús donde lo aborda el alumno.
6MD	Municipio del domicilio donde inició su viaje el alumno.

Fuente: Elaboración propia.

2.4 Análisis de datos

El análisis de datos de los atributos que fueron evaluados (12CT,12SPBT,9IS,12TT,12CTR,7TVPB,13IEPB,13IVPB,4TDP,6MD), este proceso se realizó mediante el uso de la herramienta WEKA, el cual es una plataforma de uso libre para el aprendizaje automático y la minería de datos, desarrollado con el lenguaje de programación Java. [13]

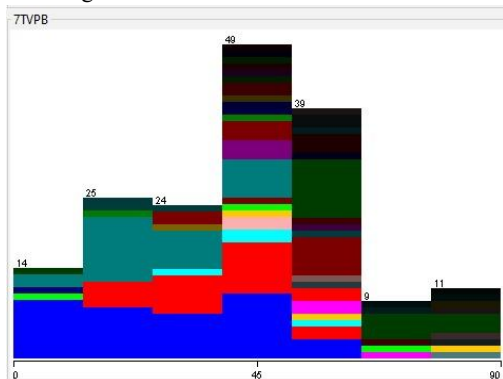
En la Figura 2 se presenta la gráfica de frecuencia de la variable 12CT, que muestra el costo del transporte siendo un factor determinante en la movilidad de los alumnos. Se observa que el valor más frecuente es de \$12, con 47 alumnos, la cual sugiere que la mayoría usa transportes económicos, siendo algunos casos que se enfrentan a costos elevados. Teniendo un incremento en sus tarifas los transportes afectarían directamente la asistencia de los alumnos con ingresos menores, por ello la importancia del potrobús como una alternativa accesible. La distribución de costos se desglosa de la siguiente manera: entre \$10 y \$20, 92 alumnos; entre \$21 y \$40, 49 alumnos; entre \$42 y \$70, 17 alumnos; entre \$75 y \$220, 8 alumnos; y con un costo de \$0, 5 alumnos.



Fuente: Elaboración propia.

La Figura 3 muestra la frecuencia de la variable 7TVPB, mostrando que la mayoría de los estudiantes viajan entre 31-60 minutos, siendo así de 60 minutos para 36 alumnos. En un rango de 0-30 minutos, hay 55 alumnos; de 31-60 minutos, 96 alumnos; y de 61-90 minutos, 20 alumnos. Demostrando así la poca eficiencia temporal, enfrentando así problemas de planificación en rutas o largas esperas para abordar el transporte.

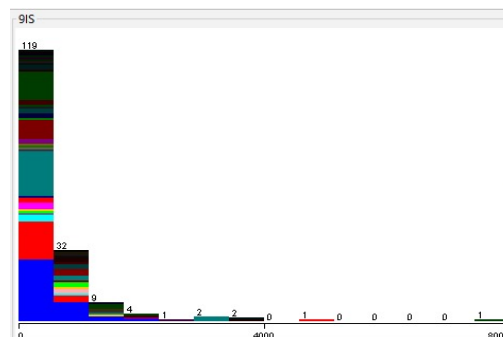
Fig. 3. Frecuencia de la variable 7TVPB



Fuente: Elaboración propia

La Figura 4 muestra la frecuencia de la variable 9IS, donde el ingreso semanal más común es de \$500 con 36 alumnos. En un rango de \$0-\$500 hay 119 alumnos; de \$501-\$1000, 32 alumnos; de \$1001-\$2000, 13 alumnos; y de \$2001-\$8000, 7 alumnos. Donde se observa que la mayoría de los estudiantes tienen ingresos menores a \$500 semanales, indicando que muchos dependen de apoyos externos o cuentan con recursos limitados, lo cual confirma que el costo del transporte es un factor importante en su movilidad.

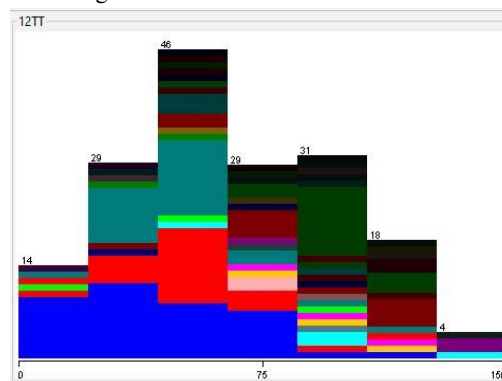
Fig. 4. Frecuencia de la variable 9IS



Fuente: Elaboración propia.

La Figura 5 muestra la frecuencia de la variable 12TT, indicando que, sin el servicio del Potrobús, el tiempo de traslado más común sería de 60 minutos (27 alumnos). En un rango de 0-30 minutos hay 27 alumnos; de 31-60 minutos, 62; de 61-90 minutos, 54; de 91-120 minutos, 24; y de 121-150 minutos, 4. Mostrando que el servicio que se brinda no solo reduce los costos, sino que de igual manera mejora los tiempos de traslado.

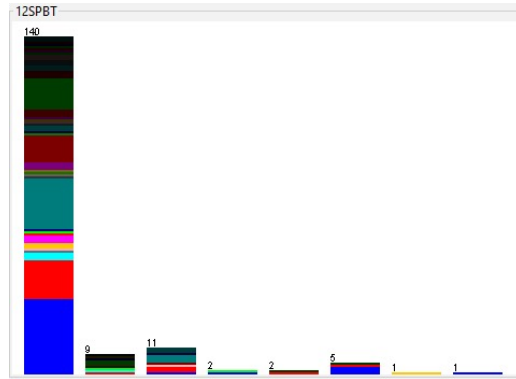
Fig .5. Frecuencia de la variable 12TT



Fuente: Elaboración propia.

La Figura 6 muestra la frecuencia de la variable 12SPBT, indicando que, sin el servicio del Potrobús, la mayoría de los alumnos (140) usarían transporte público. Otros optarían por taxi (11), taxi y transporte público (9), automóvil (5), caminata (2), transporte público y automóvil (1) o ningún medio (2), por lo anterior se puede observar que es un servicio esencial para la comunidad universitaria.

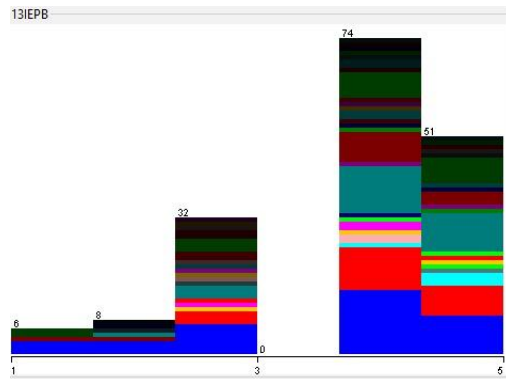
Fig.6. Frecuencia de la variable 12SPBT



Fuente: Elaboración propia.

En la figura 7 se puede observar la gráfica de frecuencia de la variable 13IEPB donde se observa que la importancia del tiempo de espera para abordar el potrobús es importante con un total de 74 alumnos, muy importante con 51 alumnos, indiferente con 32 alumnos, poco importante con 8 alumnos y sin importancia con 6 alumnos.

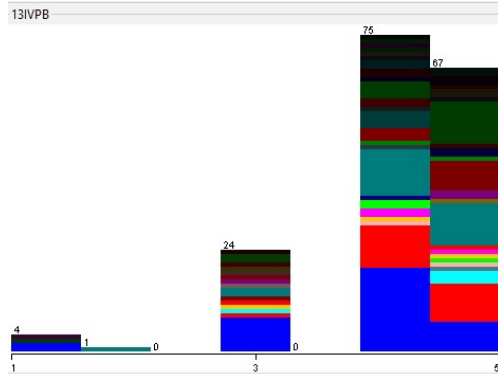
Fig. 7. Frecuencia de la variable 13IEPB



Fuente: Elaboración propia.

En la figura 8 se puede observar la gráfica de frecuencia de la variable 13IVPB donde se observa que la importancia del tiempo de viaje a bordo del potrobús es importante con 75 alumnos, muy importante con 67 alumnos, indiferente con 24 alumnos, poco importante con 1 alumno y sin importancia con 4 alumnos.

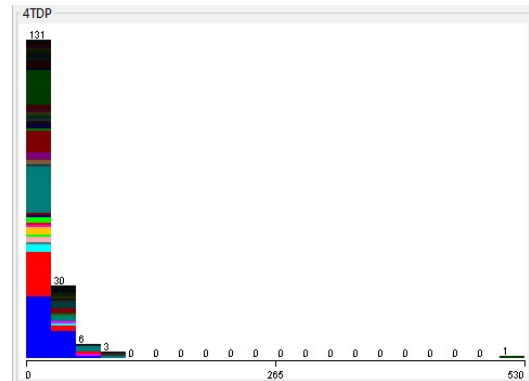
Fig.8. Frecuencia de la variable 13IVPB



Fuente: Elaboración propia.

La Figura 9 muestra la frecuencia de la variable 4TDP, donde el tiempo de traslado más común a la parada del Potrobús es de 10 y 15 minutos (36 alumnos cada uno). En un rango de 0-30 minutos hay 155 alumnos; de 31-60 minutos, 11; de 61-90 minutos, 3; y de 91-530 minutos, 2., lo anterior refiere a que es necesario realizar una adecuada evaluación de rutas.

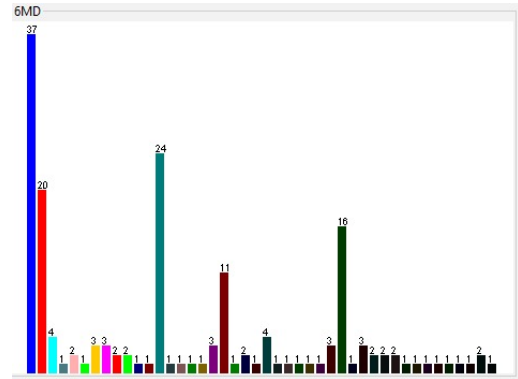
Fig.9. Frecuencia de la variable 4TDP



Fuente: Elaboración propia.

La Figura 10 muestra la frecuencia de la variable 6MD, donde la mayoría de los alumnos inician su viaje en Toluca (37), seguido de Metepec (30) y Zinacantepec (26). Estos municipios concentran la mayor cantidad de estudiantes, lo que indica que las rutas deben ser priorizadas, en estas zonas, ya que mejoraría la conexión con los municipios más alejados y así reducir los tiempos de traslado y movilidad de traslado en los estudiantes.

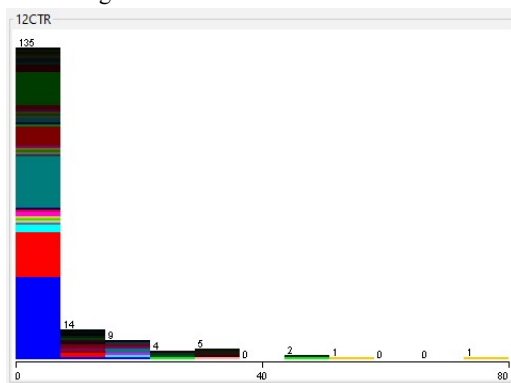
Fig.10. Frecuencia de la variable 6MD



Fuente: Elaboración propia

La Figura 11 muestra la frecuencia de la variable 12CTR, donde 135 alumnos no pagan por un segundo transporte. En un rango de \$0-\$15 hay 153 alumnos; de \$16-\$30, 13; y de \$35-\$80, 5.

Fig.11. Frecuencia de la variable 12 CTR



Fuente: Elaboración propia.

2.5 Clasificación

Los datos seleccionados son nominales, es decir, categóricos y numéricos. Se optó para utilizar de los algoritmos disponibles el J48, el cual permite obtener resultados desde un aprendizaje supervisado. Este es un método de aprendizaje supervisado, el cual construye árboles de decisión a partir de un conjunto de datos o ejemplos los cuales se denominan conjunto de entrenamiento extraídos de la base de datos. De igual manera es seleccionado un conjunto de prueba, en los cuales sus características ya son conocidas, para que pueda ser evaluado el árbol, siendo este útil por su capacidad para generar reglas de clasificación comprensibles, facilitar la interpretación de los resultados y su ayuda en la toma de decisiones, asimismo por ser eficiente al trabajar con conjuntos de datos categóricos y numéricos.

Se seleccionó el algoritmo RandomForest, un método de aprendizaje supervisado que resuelve problemas de regresión y clasificación. Este algoritmo utiliza múltiples árboles de decisión, cada uno entrenado con una muestra aleatoria con reemplazo del conjunto de datos. La predicción final varía según el problema y depende de parámetros como el tamaño del nodo, la cantidad de árboles y las características seleccionadas. Para la ejecución de los algoritmos, se utilizó la configuración por defecto del software WEKA. Su elección es por la capacidad para mejorar la precisión y evitar un sobreajuste en el uso de múltiples árboles de decisión, este método proporciona mayor estabilidad al combinar varias predicciones y en generalización, en comparación con un solo árbol, permitiendo observar que existe una tendencia en la ubicación donde los usuarios inician su traslado, lo cual resulta viable para la planificación y optimización del servicio del transporte.

2.6 Resultados

Se utilizó un árbol de decisión J48, mostrados en la Tabla 2, siendo 10 atributos analizados con 171 instancias de las cuales solo el 30.41% fueron correctamente clasificadas, teniendo un error del 69.59% de instancias mal clasificadas, un error absoluto relativo de 82.95%, un error cuadrático relativo de 111.52% y un Kappa Statistic de 0.2118 (baja concordancia con los datos reales). Creando un árbol de 83

nodos, con un mínimo de 2 instancias por nodo, un total de 45 hojas y un criterio de confianza de 0.25.

Por lo anterior se puede identificar que existe una relación entre el tiempo de traslado y la zona de residencia, donde se mostró que los estudiantes de los municipios más alejados presentan mayor tiempo de viaje y se ve reflejado en los costos de transporte. Así como el número de viajes semanales en transporte público y su municipio de origen, mostrando que quienes realizan más viajes suelen vivir en zonas más alejadas del centro de Toluca. De igual manera se observa que la frecuencia de uso de transporte público es mayor en zonas periféricas, donde los estudiantes tienden a utilizar diversos medios de transportes. En cuanto a los medios de transporte y destinos principales, el transporte público es la opción que más predomina en zonas urbanas, mientras que en municipios lejanos es común la combinación de taxi y autobús. Estos resultados se deben a la posible ausencia de atributos clave lo cual hizo que redujera su desempeño, es decir, que existieran campos con valores nulos, su bajo desempeño se debe a que se generó un árbol grande y sobre ajustado con variables categóricas de múltiples valores.

Tabla2.Resultados obtenidos algoritmo J48

```

7TVFB <= 45
| 12TT <= 75
| | 12CT <= 12
| | | 12TT <= 45: Toluca (17.0/4.0)
| | | 12TT > 45
| | | | 12TT <= 60
| | | | | 13IEPB <= 3
| | | | | | 13IVPB <= 4: Toluca (2.0)
| | | | | | 13IVPB > 4: Metepec (2.0/1.0)
| | | | | | 13IEPB > 3: Metepec (10.0/2.0)
| | | | | 12TT > 60: Toluca (3.0/2.0)
| | | 12CT > 12
| | | | 12CT <= 18: Zinacantepec (24.0/9.0)
| | | | 12CT > 18
| | | | | 12CT <= 27: Toluca (13.0/3.0)
| | | | | 12CT > 27
| | | | | | 12CT <= 46: Zinacantepec (6.0/1.0)
| | | | | | 12CT > 46
| | | | | | | 12CT <= 60
| | | | | | | 7TVFB <= 20: Toluca (3.0)
| | | | | | | 7TVFB > 20: Metepec (2.0)
| | | | | 12CT > 60: Metepec (2.0/1.0)
| 12TT > 75
| | 7TVFB <= 25: Zinacantepec (3.0/1.0)
| | 7TVFB > 25
| | | 13IVPB <= 4
| | | | 4TDP <= 12: Metepec (2.0)
| | | | 4TDP > 12
| | | | | 4TDP <= 20: Lerma (3.0)
| | | | | 4TDP > 20: Toluca (2.0/1.0)
| | | | 13IVPB > 4
| | | | | 12TT <= 85: Metepec (2.0/1.0)
| | | | | 12TT > 85
| | | | | | 12TT <= 135: Calimaya (3.0)
| | | | | | 12TT > 135: Ocoyoacac (2.0)
7TVFB > 45
| 12SPBT = Transporte publico
| | 9IS <= 300
| | | 13IEPB <= 4
| | | | 13IVPB <= 3: Calimaya (3.0/2.0)
| | | | 13IVPB > 3
| | | | | 12TT <= 65: Toluca (3.0/2.0)
| | | | | 12TT > 65: Xonacatlan (5.0/2.0)
| | | | 13IEPB > 4
| | | | | 12CT <= 20: Zinacantepec (2.0/1.0)
| | | | | 12CT > 20: Metepec (2.0/1.0)
| | | 9IS > 300
| | | | 12CT <= 21
| | | | | 13IEPB <= 4: Ixtlahuaca y Metepec (3.0/2.0)
| | | | | 13IEPB > 4
| | | | | | 7TVFB <= 55: Lerma (4.0/1.0)
| | | | | | 7TVFB > 55
| | | | | | 13IVPB <= 4: Toluca (3.0/1.0)

```

Fuente: Elaboración propia.

En la tabla 3 se muestran los resultados del algoritmo RandomForest, siendo 10 atributos analizados, 171 instancias, de las cuales sólo el 33.34% fueron correctamente clasificadas. Teniendo un error de 66.66% de instancias mal

clasificadas, un Kappa Statistic de 0.2433 (baja concordancia con los datos reales), un error absoluto relativo: de 84.38% y un error cuadrático relativo de 95.10%. Por lo que, la matriz de confusión obtenida permite analizar el desempeño por cada localidad, la diagonal principal muestra las predicciones correctas, mientras que los valores fuera de la diagonal muestran los errores de la clasificación. El modelo mostró un alto desempeño, teniendo mejor manejo en datos complejos, en la clasificación de la categoría 'Toluca', con 19 predicciones correctas, aunque hubo confusión con 'Metepec' en 7 casos, lo que sugiere una similitud en las características de ambas clases, esto porque se encontraron patrones similares en costo y tiempo de viaje, por lo cual el algoritmo confundió estas clases. Esta matriz muestra la distribución de las predicciones de un modelo sobre diferentes categorías, siendo 35 diferentes ubicaciones o entidades. Se observa una tendencia a clasificar correctamente las instancias, aunque presentan algunos errores entre clases cercanas, tales como Toluca y Metepec, siendo estos municipios cercanos. Los valores principales en cada fila se concentran a lo largo de la diagonal, lo que indica que el modelo clasifica correctamente muchas instancias en sus respectivas clases. Apesar de que la mayoría de las predicciones son correctas, hay errores de clasificación, donde las instancias de las clases llegan a ser clasificadas incorrectamente como otra clase, debido a su poca representación o con patrones similares. En la fila de Toluca, se encuentran algunas predicciones incorrectas, como la clasificación Metepec o Almoloya del Río. Por otra parte, en la fila de Metepec, existe una cantidad considerable de errores de clasificación hacia Toluca y otras localidades cercanas. En las filas correspondientes a localidades menos representadas, como Xonacatlán, San Mateo Atenco, Calimaya, San Antonio la Isla, los errores de clasificación son más frecuentes, ya que este algoritmo requiere de suficientes ejemplos por clase para poder clasificar correctamente.

Tabla3.Resultados obtenidos del algoritmo RandomForest.

[illegible]

Fuente: Elaboración propia.

3. CONCLUSIONES Y RECOMENDACIONES

A partir del análisis realizado, se identificaron patrones significativos en la relación entre el tiempo de traslado, el número de viajes, el costo del transporte y la ubicación geográfica de los usuarios, similar a lo encontrado en la

movilidad de estudiantes del estado de Jalisco [14]. Se encontró que los tiempos de traslado más cortos y menores costos de transporte están asociados con municipios cercanos a Toluca, mientras que los traslados más largos y costosos corresponden a zonas más alejadas. Asimismo, la frecuencia de uso del transporte público varía según la ubicación, siendo mayor en municipios periféricos donde se observa una mayor combinación de medios de transporte, lo que es acorde a lo encontrado en viajeros que utilizan el transporte público para trasladarse hacia y fuera de la Ciudad de México [15].

Estos hallazgos sugieren que el acceso y el costo del transporte son factores determinantes en la movilidad urbana y pueden influir en la planificación de rutas y la distribución del servicio de transporte. A pesar del bajo desempeño del modelo, la información obtenida podría contribuir a la optimización de la movilidad y a la toma de decisiones en infraestructura y transporte público, favoreciendo una mejor distribución de recursos y una mayor eficiencia en los desplazamientos urbanos y suburbanos.

Dado esto, la planificación de rutas debe priorizar las zonas con mayor demanda como lo son Toluca, Metepec y Zinacantepec, esto con el fin de reducir los tiempos de espera y tener una mejor conectividad, de igual manera se recomienda evaluar una expansión en las zonas de baja cobertura, para que con esto mejore la conectividad con otros modos de transporte público y así poder reducir tiempos de espera y trasbordo en el caso de los alumnos que utilizan más de un modo de transporte para poder llegar a su espacio educativo, esto debido a que viven en zonas periféricas y lejanas a Toluca.

Para mejorar la eficiencia de los algoritmos seleccionados, se recomienda mejorar la recopilación de datos, aumentando la cantidad de datos y su calidad, al igual que explorar modelos más avanzados, como lo son las redes neuronales artificiales, que permiten identificar patrones complejos en los datos y mejorar de la optimización de los tiempos de traslado de los usuarios. La minería de datos mostró que es una herramienta útil para analizar los patrones de movilidad en el conjunto de datos estudiado, aun así, su efectividad puede incrementar haciendo uso de técnicas avanzadas como podrían ser redes neuronales artificiales y redes neuronales recurrentes, asegurando así una mejor predicción y una mejor optimización en las rutas del servicio universitario, asegurando una movilidad más eficiente para los estudiantes.

3.1 Observaciones generales

El análisis mediante técnicas de minería de datos permitió identificar relaciones clave entre el tiempo de traslado, la frecuencia de uso del transporte y la ubicación geográfica de los usuarios. Se observó que los tiempos y costos de transporte más bajos corresponden a municipios cercanos a Toluca, mientras que los traslados más largos y costosos están asociados con zonas periféricas, donde además se combinan más medios de transporte.

Estos hallazgos destacan la utilidad de la minería de datos para optimizar la movilidad urbana, facilitando la toma de decisiones en infraestructura y transporte público, para entender los patrones de viaje y los requerimientos de los viajeros [16]. A pesar de las limitaciones del modelo, la

información obtenida puede contribuir a una mejor planificación de rutas y a la eficiencia en la distribución de los recursos de transporte.

4. REFERENCIAS

- [1] Gutiérrez, J. y García, J. (2005), Movilidad por motivo de trabajo en la comunidad de Madrid. Madrid: Universidad Complutense de Madrid.
- [2] Arango, A. (2010). La periferia conurbada de la Ciudad de México: Movilidad cotidiana y manejo de tiempo de la población en unidades habitacionales de Ixtapaluca. *Verteidigung*.
- [3] Jirón Martínez, P. (2015). La movilidad como oportunidad para el desarrollo urbano y territorial. En E. U. Valparaíso (Ed.), *La ciudad que queremos*. Santiago de Chile: Biblioteca del Congreso Nacional de Chile.
- [4] Miralles-Guasch, C. (2012). “Las encuestas de movilidad y los referentes ambientales de los transportes”. *Eure*, 38(115), 33–45.
- [5] Miralles-Guasch, C., y Martínez, M. (2013). “Las fuentes de información sobre movilidad: la visión de los profesionales. Ejemplo de aplicación de metodología DELPHI”. *Revista Transporte Y Territorio*, (8), 99–116.
- [6] Galán, J. B. (2022). Estudio de la movilidad con la tecnología big data: Posibilidades por explorar. *Dialnet*.
<https://dialnet.unirioja.es/servlet/articulo?codigo=8435917>.
- [7] Molina, L. C. (2000). Data mining: torturando los datos hasta que confiesen. *FUOC*.
- [8] Bermeo Muñoz, E., & Calderón Sotero, J. (2009). Diseño de un modelo. *Redalyc*, 52-66.
- [9] González Calderón, G., Jaramillo Álvarez, P., & González Calderón, C. (2013). Optimización De Rutas De Transporte Público Urbano. *Redalyc*, 42-48.
- [10] Quintana Cruz, E., Ojeda Meza, L., & Almeida Franco, O. (2019). Diseño de un sistema de rutas variable de transportación basado en sistemas de información geográfica. *Avances*, vol. 21, núm. 4, 1-14.
- [11] Noriega Díaz, E., Romero-Torres, J., Gutiérrez Cruz, D., & Rico Molina, R. (2023). 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. *Revista Aristas: Investigación Básica y Aplicada*, 164.
- [12] Timarán Pereira, S., Hernández Arteaga, I., Caicedo Sambrano, S., Hidalgo Troya, Á., & Alvarado Pérez, J. (2016). El proceso del descubrimiento en bases de datos. *Universidad Cooperativa de Colombia*.
- [13] Gutiérrez Cruz, D., Albarrán Fernández, Y. A., & Rico Molina, R. (2017). *RI UAEMex*. Obtenido de http://ri.uaemex.mx/bitstream/handle/20.500.11799/69982/secme-30553_1.pdf?sequence=1.
- [14] IIEG. (2019). Obtenido de Encuesta del uso del tiempo de estudiantes universitarios. Movilidad, costo de traslado y uso del tiempo de estudiantes universitarios de Jalisco.
https://iieg.gob.mx/ns/?page_id=6971.
- [15] INEGI. (2019). Obtenido de Encuesta de Origen Destino en Hogares de la Zona Metropolitana del Valle de México: <https://www.inegi.org.mx/>.
- [16] Secretaría de Movilidad. (2020). Programa Integral de Movilidad de la Ciudad de México 2020-2024, Diagnóstico Técnico.