UNIVERSIDAD DEL AZUAY

DEPARTAMENTO DE POSGRADOS

Identificación de patrones de movilidad urbana y su relación con variables meteorológicas registradas en el cantón Cuenca

Trabajo de graduación previo a la obtención del título de:

Magister en Matemática Aplicada

Autor:

Darwin Darío Espinoza Saquicela

Director:

Iván Andrés Mendoza Vázquez, PhD.

Cuenca – Ecuador

2023

AGRADECIMIENTOS

A mi familia, amigos y sobre todo a mi prometida, por sus palabras de aliento y momentos compartidos antes y durante se materializaba este trabajo, sepan que han servido de inspiración para superar los obstáculos que se presentaron.

A los docentes que formaron parte de este programa de maestría por haber compartido su experiencia, consejos, ayuda y conocimiento en cada módulo. A los profesores Iván Mendoza, Julio Mosquera y Jonnatan Avilés, por el acertado y oportuno acompañamiento que ha servido para que este trabajo se realice.

Darwin Darío Espinoza Saquicela.

DEDICATORIA

De manera especial a mi prometida, María José, que ha estado presente desde el primer día en que este desafío académico inició, brindándome su apoyo, paciencia y amor incondicional. Gracias por creer en mi y acompañarme en cada etapa de esta maestría, este logro es nuestro.

También a mi madre, padre y hermanos, por la motivación y apoyo que me han brindado una vez más para afrontar un nuevo desafío académico. El saber que estaban presentes, con palabras de apoyo y gestos de cariño, fue un respaldo necesario para soportar cada reto que se presentó durante el curso de esta maestría.

Darwin Darío Espinoza Saquicela.

ÍNDICE

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

Identificación de patrones de movilidad urbana y su relación con variables meteorológicas registradas en el cantón Cuenca

Darío Espinoza S. – despinoza_maestria@es.uazuay.edu.ec
Iván Mendoza V. – imendoza@uazuay.edu.ec

RESUMEN

Resumen

Los estudios referentes a la movilidad urbana proporcionan datos y análisis fundamentados que respaldan la toma de decisiones y permiten a los gobernantes desarrollar estrategias efec tivas y eficientes para enfrentar los desafíos que presenta la movilidad urbana, mejorando la calidad de vida de los habitantes. En este estudio se realizaron análisis de datos de movilidad registrados mediante dispositivos móviles de personas que transitaron en Cuenca, y de las variables meteorológicas, temperatura del aire y precipitación, registradas durante sus movi lizaciones. Se encontraron relaciones entre variables de cada tipo de datos, y posteriormente estos se combinaron encontrando relaciones con menor fuerza. Además, se notó superioridad en uso de vehículos por sobre otros modos de transporte. Y finalmente se produjeron dos clasificadores usando la técnica *KKN* para determinar el modo de movilización en horarios pico y horarios valle, y para destinos de movilización a pie, con un 70 % y 30 % de aciertos, respectivamente.

Palabras clave: Patrones de movilidad. Variables meteorológicas. Combinación de datos. Agrupación. Modelos de clasificación.

ABSTRACT

Abstract

Studies on urban mobility provide data and informed analysis that support decision making and decision makers to develop effective and efficient strategies to face the challenges of urban mobility, by improving the quality of life of the inhabitants. In this study, mobility data recorded by mobile devices of people who traveled in Cuenca, and the meteorological variables, air temperature and precipitation, recorded during their travels were analyzed. Relations were found between variables of each type of data, and subsequently these were combined. Later, these findings led to finding relations of lesser strength. In addition, superiority was noted in the use of vehicles over other means of transportation. Finally, two classifiers were produced by using the *KKN* technique to determine the mode of transportation during peak and off-peak hours, and for walking destinations, with a 70 % and 30 % success rate, respectively.

Keywords: Mobility patterns. Meteorological variables. Data merging. Clustering. Classifi cation models.


UNIVERSIDAD DEL
AZUAY
Dpto. Idiomas

INTRODUCCIÓN

Las ciudades crecen debido a diversos factores como la urbanización, migración o el crecimiento mismo de la población. Según la *ONU*, se prevé que para el año 2050 un 68 % de la población viva en áreas urbanas (United-Nations, 2018). Es de esperarse, entonces, que el crecimiento de las ciudades demande incremento y adaptación de aquellos servicios que satisfacen las necesidades de los habitantes. Un área de especial importancia para las ciudades es la diversa infraestructura que ha de ser dedicada a las distintas necesidades de las personas; para el caso de la movilidad son las calles, autopistas y sistemas de transporte. En Cuenca, la movilidad tiene que ver principalmente con el transporte terrestre siendo usados para movilizarse buses, taxis y vehículos privados; sin embargo, a pesar de los beneficios que estos medios representan, un problema generado de forma recurrente es la congestión vehicular producida, especialmente durante horas pico.

La congestión vehicular ocasiona impactos sobre la calidad de vida de sus habitantes. Los retrasos e incrementos en el tiempo de viaje son preocupaciones menores frente a problemas de gran escala que impactan negativamente sobre la salud de las personas (Louiza y cols., 2015), como la contaminación del aire e incluso contaminación sonora. En años recientes se han realizado esfuerzos dirigidos a mejorar la movilidad de las personas y reducir la congestión vehicular, con la implementación de sistemas de bicicleta pública, incremento de ciclo vías, semaforización adaptable e implementación de nuevos medios como tranvía.

Este estudio tiene como objetivo evidenciar la influencia de un factor externo en la movilidad urbana, como son las condiciones meteorológicas. Para ello, se cuenta con datos de movilidad recolectados de forma pasiva mediante *GPS* de dispositivos móviles provenientes de personas que se movilizaron por el cantón Cuenca durante tres meses del año 2019. Para el caso de las condiciones meteorológicas, se obtuvieron datos de variables meteorológicas de fuentes abiertas que se contrastan con los periodos observados en los datos de movilidad.

Revisión bibliográfica

Los patrones de movilidad son una parte importante en el entendimiento de los motivos que llevan a las personas a movilizarse. En la ciudad de Shenzhen-China (Liu y cols., 2009) se usan datos en tiempo real provenientes de dispositivos *GPS* en taxis y tarjetas inteligentes para bus o subterráneo con el objetivo evaluar dinámicas de movilidad; usando técnicas de *clustering* para analizar cualitativamente la relación de viajes diarios con locaciones y uso del suelo. Como resultado se obtuvo una plataforma denominada *Urban Mobility Landscape System* (*UMSL*) la cual es descrita a detalle con el propósito de que sea replicada, y se demuestra que con los datos de entrada se pueden inferir aspectos culturales y geográficos de una ciudad, por ejemplo, que las personas eligen realizar actividades nocturnas los fines de semana y no entre semana.

En el trabajo de Yuan y Rauball (Yuan y Martin, 2012) se identifican patrones de movilidad agregados con datos provenientes de telefonía móvil. Partiendo de series de tiempo

se extraen y representan patrones de movilidad dinámicos en diferentes zonas urbanas, aplicando un algoritmo de *Dynamic Time Warping (DTW*) para medir la similaridad entre series de tiempo; pudiendo clasificar distintas áreas urbanas por sus patrones y notando diferencias en las series temporales de distritos con negocios y zonas suburbanas, por ejemplo.

En la ciudad de Viena-Austria (Rudloff y cols., 2015) se obtiene evidencia de que la demanda de viajes es influenciada por las condiciones climáticas, mediante análisis descriptivo de encuestas a largo plazo y modelado estadístico con datos de encuestas cortas; concluyendo que la mayor reducción de movilidad se da por la precipitación y temperaturas extremas, siendo la caminata, bicicleta y motocicleta los tipos de movilidad más afectados por el clima, mientras que el auto privado no se ve fuertemente afectado.

Se puede observar que en estos primeros trabajos se usan distintos orígenes de datos para el estudio de patrones. Usualmente los medios de recolección pasiva como *GPS*, redes móviles y su combinación entregan niveles de precisión considerables (Rojas y cols., 2016), sin embargo estudios basados en orígenes de datos como encuestas o redes sociales brindan un acercamiento interesante a los patrones de movilidad. La gran cantidad de dispositivos móviles en tiempos recientes y su capacidad de proporcionar información, por otra parte, incrementan la complejidad y volumen de datos. En el estudio de Sun y Axhausen (Sun y Axhausen, 2016) se propone un marco analítico para tratar grandes cantidades de datos, formulándolos con ajuste probabilístico y considerando cada registro como una observación multivariante de una distribución estadística subyacente. En una aplicación numérica, en Singapur, el modelo es aplicado a un data-set con cuatro orígenes de datos y catorce millones de registros de viajes en transporte público.

Estudios en tiempo real proporcionan una visión muy cercana a la realidad de la movilidad de una ciudad, dada naturaleza de su origen de datos, sin embargo un problema a considerar es que las aplicaciones pueden verse afectadas negativamente por altas demandas derivadas de numerosas solicitudes de usuarios o grandes latencias que dan lugar a respuestas correctas pero tardías. Para mitigar esto, el proyecto denominado *APOLO* (*context-Aware and PeOple-centric vehicuLar traffic rerOuting*) (Akabane y cols., 2017) propone un enfoque en dos etapas, fuera de línea y en línea; el primero dependiendo del procesamiento de patrones provenientes de datos históricos de la red de carreteras para generar rutas a seguir sin necesidad de información en tiempo real, y el segundo en el que vehículos que estén próximos a rutas congestionadas son redirigidos; obteniendo una reducción en tiempos de viaje de 17 %, tiempo de inactividad para conductores de 50 % e incremento de velocidad promedio de 6 %.

En el trabajo sobre identificación de patrones de movilidad urbana de Lucas M. Rodríguez (Rodríguez, 2018) se generan modelos predictores tomando *tweets* como fuente de información georreferencial, comparando las rutas generadas por el modelo predictor con rutas reales y encontrando diferencias por género en los caminos analizados. Puesto que el origen de datos es una red social la precisión de los caminos es baja, requiriendo un agregado de georreferenciación de rutas por *GPS* para mejorar las predicciones.

En cuanto al estudio de patrones de movilidad, con ámbito local, Orellana (Orellana, 2016) documenta elementos para un marco analítico multidisciplinar sobre la movilidad no motorizada, compuesto por tres ejes, metodológico, comportamental y perceptual. En un caso de uso del esquema que propone, asigna locaciones que voluntarios deben visitar durante un periodo de tiempo, registrando sus movimientos con dispositivos GPS; para luego comparar sus rutas elegidas con las consideradas óptimas estudiando las diferencias y validando resultados con entrevistas semi-estructuradas. También en Cuenca, Gallegos y García (Gallegos Barros y García Torres, 2022) analizan rutas de empresas privadas enfocadas al turismo y las opiniones de sus actores involucrados; concluyendo que los usuarios del servicio mencionan no sentirse afectados por la congestión vehicular, a pesar que el mismo contribuya debido a acciones como transitar a muy bajas velocidades por calles de alto tráfico vehicular.

Como se ha expuesto, el estudio de patrones de movilidad puede involucrar distintos planteamientos y acercamientos, esta breve investigación ha permitido conocer estudios y análisis con una motivación común. El crecimiento de las ciudades origina nuevos escenarios y demanda decisiones oportunas para satisfacer problemas de distinta índole como lo es en el ámbito de la movilidad. Este trabajo propone la identificación de patrones de movilidad partiendo de un *dataset* de movilidad con variables referentes a los orígenes, destinos, distancias, fechas y tiempos de viaje, de personas que transitan por las calles de Cuenca. A diferencia de los estudios citados, este pretende evidenciar la influencia del factor externo clima, combinando *datasets* de movilidad y variables meteorológicas registradas durante el tránsito, para finalmente generar y evaluar modelos de predicción de medios de transporte y locaciones de destino. En la siguiente sección se revisará la metodología para preparar los datos, combinar los *datasets* y producir modelos, luego se discutirá las ventajas y desventajas de la propuesta, finalmente se tendrán las conclusiones y mención de posibles trabajos futuros.

METODOLOGÍA

La escasez de datos meteorológicos y presencia de errores son riesgos que se tuvieron presentes durante la formulación de este proyecto, por lo que para su mitigación se adquirieron datos de distintas fuentes de registros de variables meteorológicas, con especial interés en temperatura del aire y precipitación.

La primera fuente de datos consultada fue una red de sensores referenciales que monitorea distintos puntos de la ciudad. Tras obtener acceso a los datos se notó que las variables disponibles son temperatura del aire, presión barométrica y humedad relativa, con registros desde el 30 de septiembre al 15 de octubre de 2019. Poca continuidad y falta de observaciones durante el periodo definido por los datos de movilidad llevó a que esta fuente se descarte.

Otra fuente consultada fue la perteneciente al Aeropuerto Mariscal Lamar de la ciudad de Cuenca. Entre las variables disponibles se encontró temperatura promedio del aire, precipitación, humedad, presión, dirección y velocidad del viento, con frecuencia mensual y

horaria. Sin embargo, no contar con las especificaciones de esta estación y ausencia de registros de precipitación para todo el periodo de las movilizaciones hizo que esta fuente también se descarte.

Finalmente se obtuvieron registros por hora, desde el 26 de octubre de 2019 en adelante, de temperatura del aire y precipitación provenientes de la estación automática de monitoreo en tiempo real de contaminantes atmosféricos. La estación se encuentra ubicada en los altos de la empresa pública EMOV con coordenadas -2.89 y -79.00, latitud y longitud respectivamente (Sellers Walden, 2017), y tiene un rango efectivo de cobertura de 4 Km de radio, lo que cubriría gran parte de la ciudad, como se observa en la Figura 1.

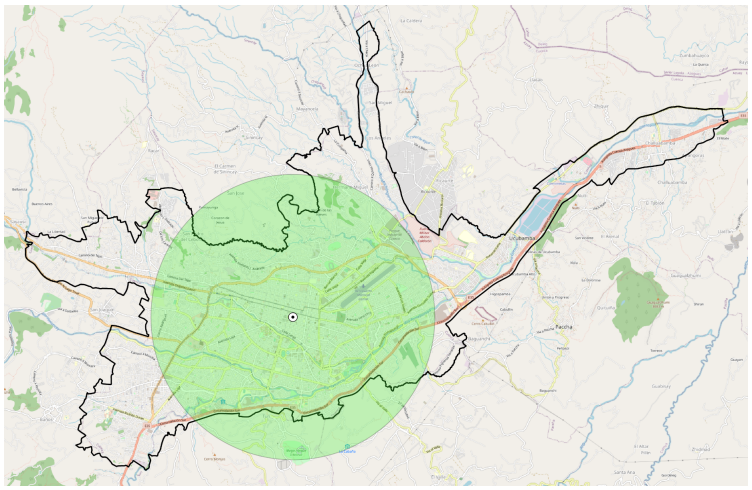

Figura 1: Área efectiva cubierta por la estación automática de monitoreo en tiempo real.

Preparación de *datasets*

Dataset de movilidad

En el caso del primer conjunto de datos se omitieron las variables que corresponden a coordenadas *UTM* ya que para el análisis se usaron coordenadas geográficas. Para el segundo conjunto de datos se omitió el identificador de modelo de dispositivo puesto que es una variable complementaria al identificador de usuario, mismo que relaciona ambos conjuntos de datos. Conocido el alcance efectivo de la estación automática de monitoreo en tiempo real de contaminantes atmosféricos, de 4 Km, se realizó un primer filtrado de los datos de movilidad que estén dentro del área urbana de Cuenca. Para ello se usó el algoritmo presentado en la Figura 2, en el que mediante un bucle se analizan cada uno de los viajes, verificando que los puntos de viaje se encuentren dentro del límite urbano.

Tras completar el algoritmo es posible visualizar los puntos que se encuentran dentro del área delimitada usando sus coordenadas, como se observa en la Figura 3, y se obtienen en total 2724 viajes realizados.

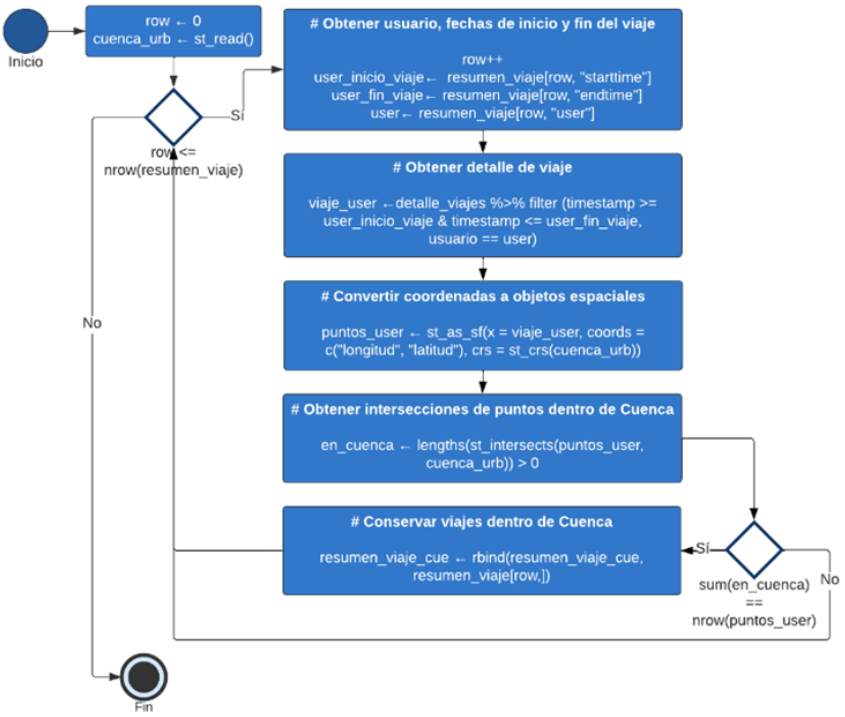

Figura 2: Algoritmo para seleccionar puntos de movilización dentro del área urbana de Cuenca.

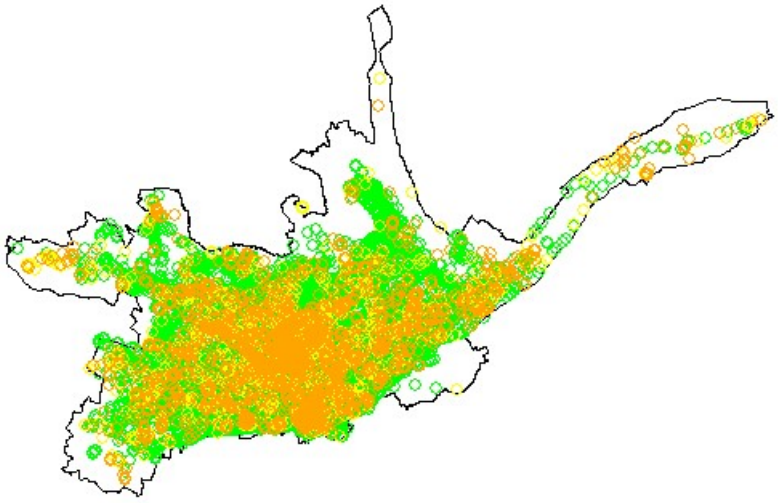

Figura 3: Puntos de movilización dentro del área urbana de Cuenca. En color amarillo y naranja, orígenes y destinos, respectivamente, y en verde, los punto registrados por cada viaje.

Dataset de meteorología

Los datos meteorológicos constan de las variables hora, fecha, temperatura del aire máxima, mínima y promedio, y suma de precipitación. Para este *dataset* fue necesario generar una nueva variable de tipo marca de tiempo o *timestamp*, obtenida al combinar las variables fecha y hora, y se seleccionó las variables temperatura promedio del aire y suma de precipitación. Consta de datos continuos por hora, desde el 26 de octubre al 6 de noviembre de 2019.

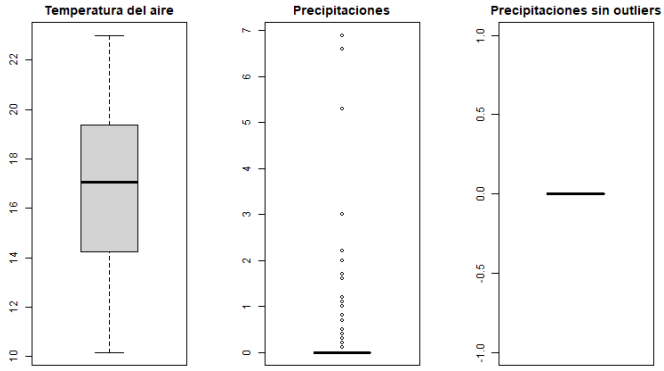

Figura 4: Diagramas de caja de temperatura promedio del aire y precipitación, registrados del 26 de octubre al 6 de noviembre de 2019.

Una observación rápida permitió notar que la temperatura del aire para este periodo, oscila entre 10 y 22 grados centígrados, con una media cercana a los 16; y en el caso de precipitaciones se observó valores de hasta 6.9 milímetros advirtiendo una gran presencia de valores atípicos, como se puede observar en la Figura 4.

Combinación de *datasets*

¿Cuáles fueron los valores de temperatura y precipitación registrados durante cada viaje? Para resolver esta incógnita y conocer las condiciones climáticas presentadas durante los viajes fue necesario encontrar las variables comunes, presentes en los *datasets* de movilidad y meteorología, que permitan combinarlos.

Al delimitar el área de estudio por el rango efectivo de la estación automática de monitoreo, de forma indirecta, se realizó un acercamiento a la combinación o *merge* de los *datasets* de movilidad y meteorología usando sus coordenadas. Debido a la alta precisión y cobertura de la fuente de datos meteorológicos se hizo posible trabajar con un área que cubre casi la totalidad de la ciudad; omitiendo escenarios en los que se requieren interpolaciones debido a que los puntos de monitoreo suelen estar dispersos. La segunda variable que ha sido usada para combinar los *datasets* tiene que ver con las marcas de tiempo o *timestamps*. En el caso del *dataset* de movilidad se tienen las marcas de tiempo para inicio y final de cada viaje, y en del *dataset* de variables meteorológicas se tienen *timestamps* que corresponden al promedio de las mediciones registradas en la hora pasada. Para realizar el *merge* se usó unión izquierda *(left join)*, generando un nuevo *dataset* en el que están

presentes las variables de resumen de viaje, con dos adicionales en las que se almacenan el promedio de temperatura y suma de precipitación.

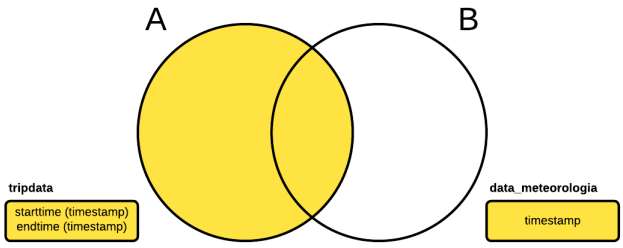

Figura 5: *Left-join* entre datasets de movilidad y variables meteorológicas.

Se implementó un algoritmo que, a breves rasgos, itera con cada uno de los viajes que se encuentran dentro del área de estudio, obteniendo los valores de las variables meteorológicas registradas entre las marcas de tiempo de inicio y final de viaje; es decir, son considerados todos aquellos valores de temperatura y precipitación presentes durante cada viaje. Tras obtener todos los registros de las variables meteorológicas el algoritmo asocia el promedio de las temperaturas y suma de precipitaciones, a cada uno de los viajes. Finalmente, completada la combinación de los *datasets* se descartan aquellos viajes en los que no se obtuvieron promedios y suma de precipitaciones, obteniendo un nuevo *dataset*. Se ilustra en la Figura 5 la operación de combinación siendo "A" el *dataset* de movilidad y "B" el de datos meteorológicos, y en la Figura 6 se muestra el algoritmo usado.

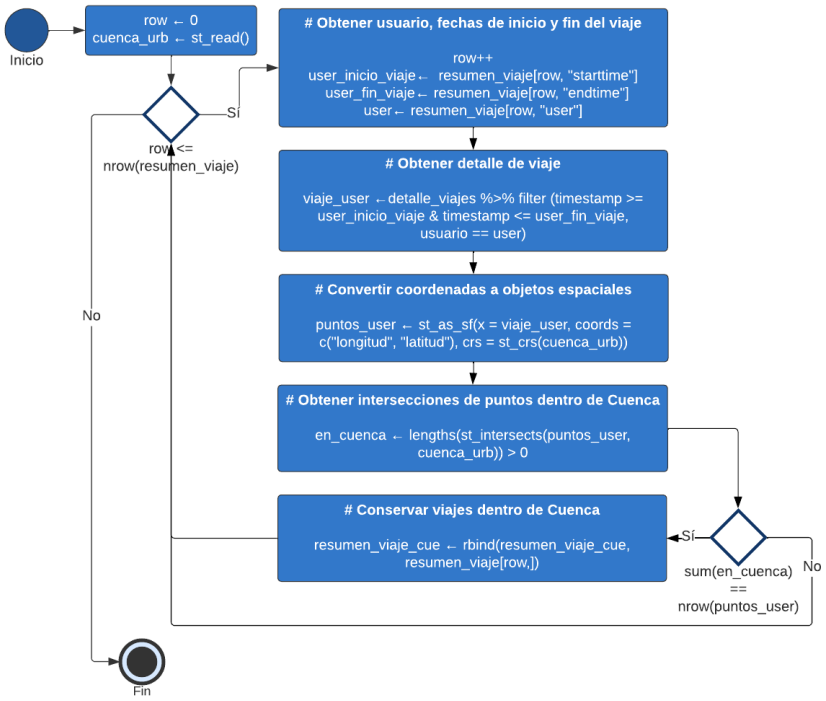

Figura 6: Algoritmo para obtener valores de temperatura promedio y precipitación presentados durante viajes.

Relación entre movilidad y meteorología

Con el *dataset* combinado, de datos meteorológicos y de movilidad, se realizó un análisis exploratorio en busca de patrones. Es evidente que la movilidad puede verse afectada por las condiciones climáticas, la elección de un medio de transporte u otro puede depender de la temperatura. Por ejemplo, en un clima cálido, las personas pueden preferir caminar o andar en bicicleta, mientras que en un clima frío o lluvioso, pueden preferir usar un automóvil o el transporte público. Sin embargo, la movilidad se ve afectada por factores adicionales como cultura, situación económica, geografía, etc., por lo que definir un patrón que relacione movilidad y meteorología puede presentar un desafío.

Se analizó el *dataset* combinado con el objetivo de encontrar relaciones, entre los datos de movilidad y meteorología, que guíen a la identificación de los patrones:

1. Modos de transporte usados en horas pico y horas valle, en días hábiles. Trata de estudiar si los modos de transporte se ven influenciados por las condiciones meteorológicas, con especial interés en movilización pedestre y vehículos de motor.
2. Destinos de personas que se movilizan a pie, en fines de semana. Trata de estudiar si los destinos elegidos por este grupo de personas, está afectado por condiciones meteorológicas.

Análisis inicial

Usando la totalidad del *dataset* combinado y tomando las variables numéricas distancia de viaje en metros, tiempo de viaje en segundos, temperatura promedio del aire en grados centígrados y precipitación en milímetros; se encontró correlaciones, como se observa en la Figura 7, positiva baja entre tiempo de viaje y distancia de viaje, e inversamente proporcional entre precipitación y temperatura promedio, algo común entre estas dos variables meteorológicas.

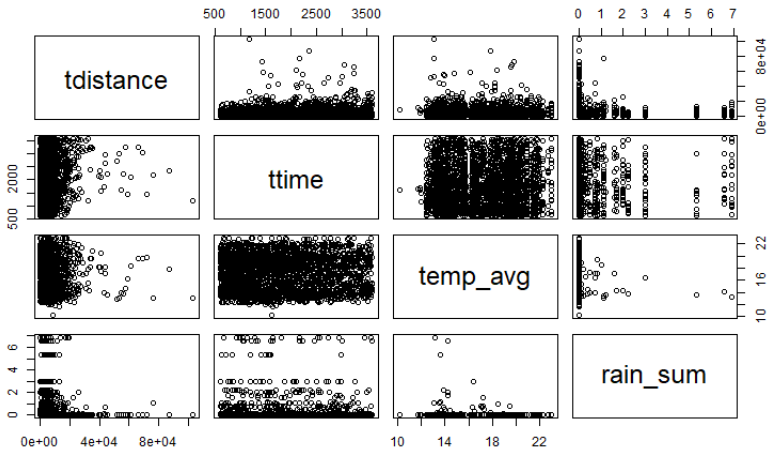

Figura 7: Correlación entre distancia de viaje, tiempo de viaje, temperatura promedio del aire y precipitación.

Por otro lado, adjuntando la variable tipo factor "modo", en busca de un patrón que relacione los modos de movilización y las variables numéricas, temperatura promedio, precipitación, tiempo de viaje y distancia de viaje, se encontró que las variables meteorológicas no son tan buenos predictores como las variables de movilidad. Los resultados de este análisis se muestran en la sección siguiente.

Clustering

Tras no encontrarse patrones fuertes que relacionen las variables meteorológicas y las variables de movilidad, se usó *clustering* para encontrar una forma de agrupar los datos de manera que se relacionen movilidad y meteorología. Las técnicas seleccionadas son *clustering* jerárquico complete y *k-means*¸ experimentando con distintos números de clúster y analizando métricas internas. Los resultados de aplicar *clustering* también se muestran en la sección siguiente.

RESULTADOS

Análisis inicial

Se generó modelos de regresión lineal para predecir una variable numérica en función de otra, se usó las variables tiempo de viaje y distancia de viaje, temperatura promedio del aire y distancia de viaje. Se preparó conjuntos de entrenamiento y validación correspondientes al 60 % y 40 % del *dataset* combinado, respectivamente; los resultados son presentados en la Tabla 1. Las métricas indican que la temperatura promedio del aire no sería un buen predictor de la distancia de viaje, mientras que la variable "tiempo de viaje" podría ser un predictor aceptable de la distancia de viaje. Durante este análisis no se logró remover *outliers* haciendo uso de la región intercuartil.

	Tiempo de viaje y Distancia de viaje	**Temperatura promedio del aire y Distancia de viaje**
RMSE	7849.125	6897.402
p-value	1.17E-06	0.6481
Adjusted R-squared	0.02685	-0.0005978

Tabla 1: Métricas obtenidas para modelos de regresión lineal generados con variables numéricas

Analizando variables tipo factor, en busca de un patrón entre los modos de movilización y, las variables temperatura promedio, precipitación, tiempo de viaje y distancia de viaje, se preparó un subconjunto de datos. Los niveles que la variable modo puede tener, son: caminando, en vehículo, en bicicleta, a pie y corriendo. Como se puede observar en los diagramas de caja de la Figura 8, para el caso de temperatura promedio, no se observan grandes diferencias entre los grupos, de igual manera, en la Figura 9 donde se observa gran presencia de *outliers* para la variable precipitación, y en la Figura 10 los valores atípicos fueron removidos.

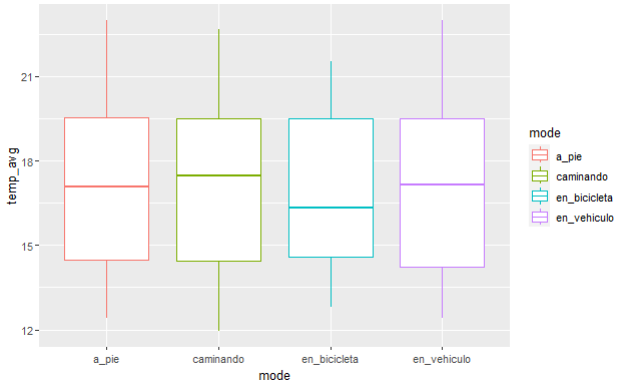

Figura 8: Diagramas de caja por grupos de modos de transporte para temperatura promedio del aire

Esto se confirmó realizando un análisis de varianza *ANOVA*, obteniendo un *p-value* de 0.819 para el caso de temperatura promedio. Al usar la variable precipitación, se obtiene un valor de 0.079 convirtiéndolo en un mejor predictor que la variable temperatura promedio. Para fines didácticos, al usar la variable distancia de viaje, proveniente del *dataset* de movilidad, se obtiene un *p-value* muy bajo de 2e-16, mostrando que sería un buen predictor del tiempo de viaje.

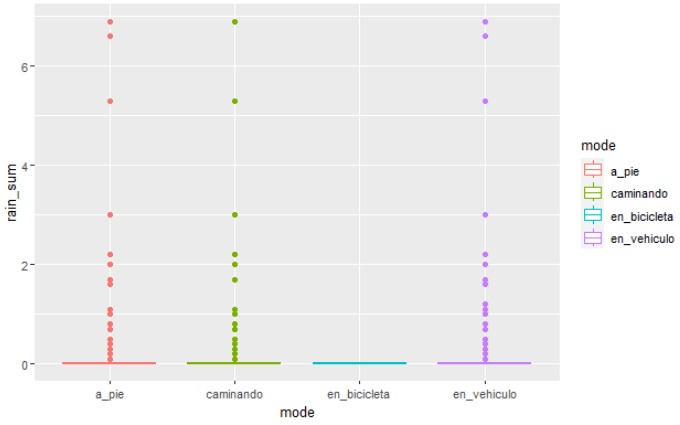

Figura 9: Diagramas de caja por grupos de modos de transporte para precipitación

Los resultados son sustentados también por pruebas de *Tuckey Honest Significant Differences*, que se pueden ver en la Figura 11, notándose que existen pocas diferencias en las medias al usar las variables temperatura promedio o precipitación, mientras que la distancia de viaje muestra diferencias en sus medias lo que podría convertirla en un buen predictor del modo de movilización.

Finalmente se elaboró un modelo de árbol de decisión, visible en la Figura 12, en el que se puede evidenciar que para su generación las variables meteorológicas no son elegidas, y en su defecto se tomó las variables de movilidad tiempo de viaje y distancia de viaje para la generación del modelo, lo que es sustentado por los análisis mostrados.

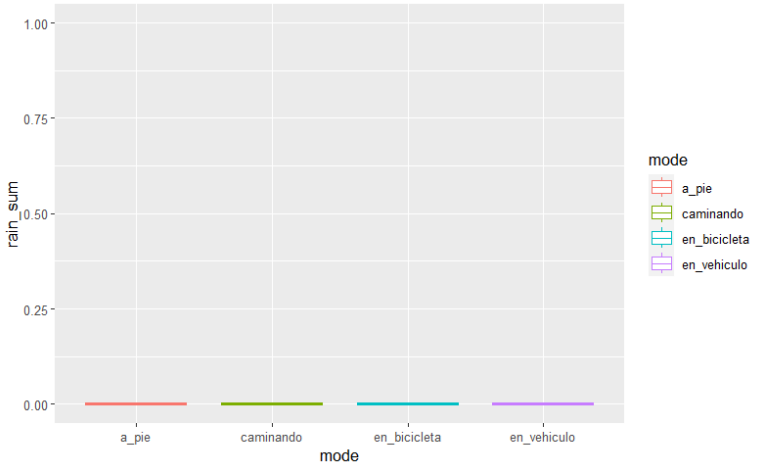

Figura 10: Diagramas de caja por grupos de modos de transporte para precipitación, sin valores atípicos

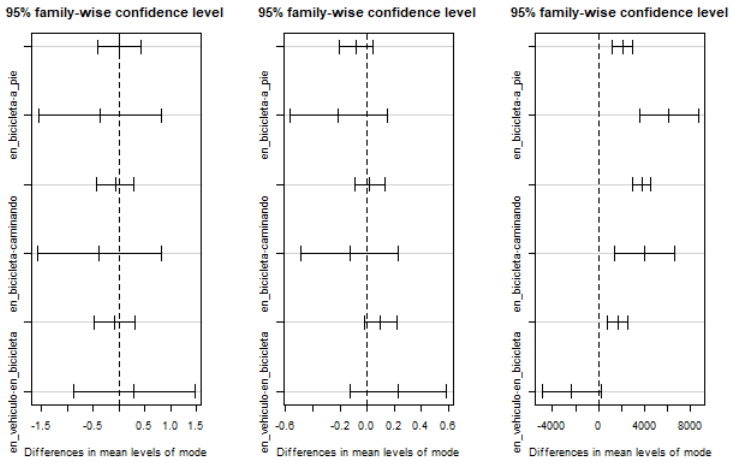

Figura 11: *Tuckey Honest Significant Differences* para modos de movilización en función de temperatura, precipitación y distancia de viaje, de izquierda a derecha

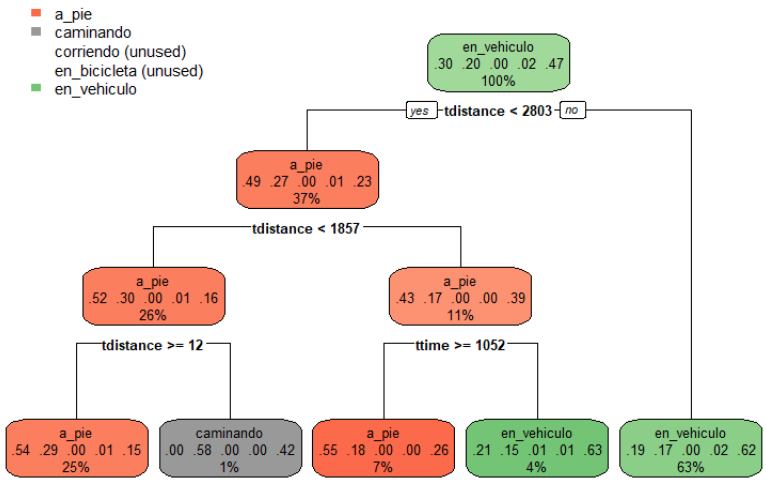

Figura 12: Árbol de decisión para predecir modo de movilización

Clustering

Debido a no hallar una fuerte relación de las variables meteorológicas y las de movilidad, se usó *clustering* para tratar de encontrar una agrupación no detectable a simple vista. Las técnicas elegidas son *clustering* jerárquico *complete* y *k-means*, obteniendo las métricas internas, conectividad o *connectivity*, ancho de silueta o *silhoutte width* e índice de *Dunn* o *Dunn index*, para un número de dos, tres y cuatro clúster. Tras normalizar la data y aplicar *clustering* con números para n entre 2 y 5 , se obtienen los resultados presentados en la Tabla 2. Analizando las métricas obtenidas, se concluye que el método de *clustering* con mejores resultados es el jerárquico con 2 clúster.

	Métrica	**n = 2**	**n = 3**	**n = 4**	**n = 5**
Jerárquico	*Connectivity*	3.1790	3.1790	9.7734	17.1913
	Dunn Index	0.2143	0.2145	0.1069	0.1231
	Silhouette width	0.8315	0.7070	0.6847	0.6397
K-means	*Connectivity*	266.8151	256.4425	273.9222	273.922
	Dunn Index	0.0041	0.0045	0.0059	0.0051
	Silhouette width	0.2848	0.3059	0.3016	0.2938

Tabla 2: Métricas internas de *clustering* obtenidas para *clustering* jerárquico y *k-means*

En la Figura 13 se observa el dendograma generado tras aplicar *clustering* jerárquico *complete* para n igual a 2, obteniendo dos grupos, uno en color morado y otro en cian. Sin embargo, la abundancia de individuos para el grupo en cian hace que la clasificación generada no sea de gran aporte, ya que en el grupo en morado cuenta apenas un 0.7 % del total de individuos.

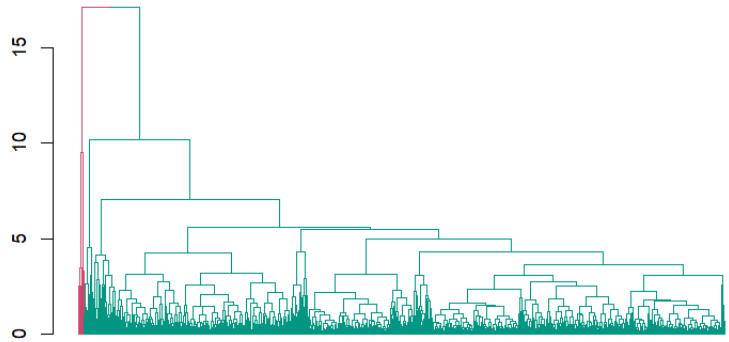

Figura 13: Dendograma obtenido tras realizar *clustering* jerárquico *complete*

Patrones

El análisis inicial anterior, fue realizado al considerar un único conjunto de datos, lo cual no es del todo correcto, puesto que la movilidad es un fenómeno humano cambiante que se ve influenciado por múltiples factores. En esta sección se trabajó con distintos subconjuntos de datos, para analizar datos de movilidad y meteorología en horarios con distintos

niveles tráfico vehicular, para días hábiles y fines de semana. Cabe mencionar la superioridad que tiene el uso de vehículos frente a otros modos de transporte, como se ve en la Figura 14, algo mencionado en estudios de movilidad realizados en la ciudad de Cuenca por Orellana (Orellana, 2016) , Gallegos y García (Gallegos Barros y García Torres, 2022).

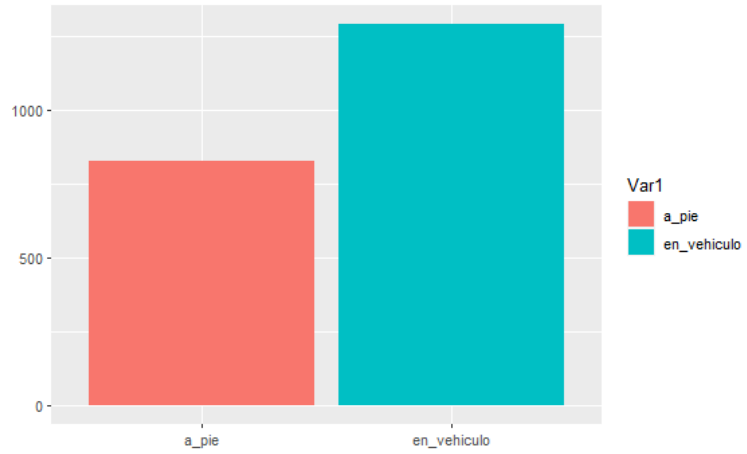

Figura 14: Comparativa de uso de vehículos y movilización pedestre

En la Figura 15 se pueden observar las gráficas de rutas aproximadas para un subconjunto de individuos que transitaron por el área urbana de Cuenca a las 07H00 siendo notable el incremento de movilizaciones en días hábiles, de lunes a viernes, frente a los días sábados y domingos para un mismo horario.

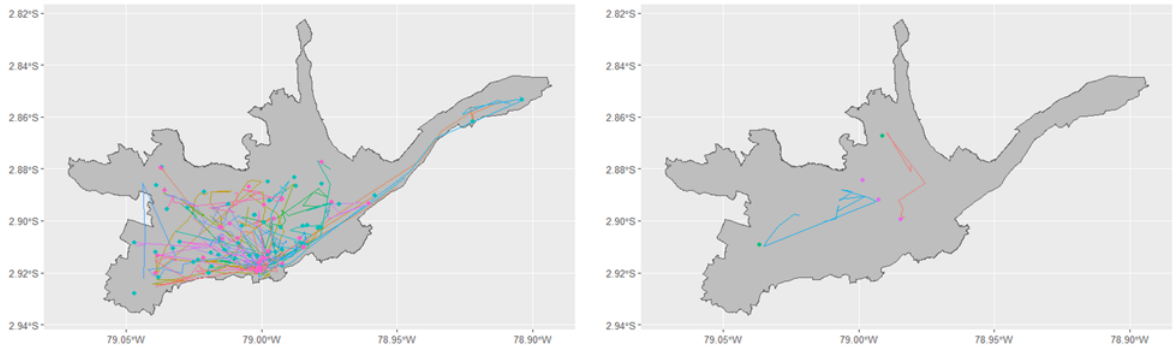

Figura 15: Movilizaciones a las 07H00 en días hábiles (izquierda) y fines de semana (derecha)

Modos de transporte usados en horas pico y horas valle

Este patrón es buscado en días hábiles, con el objetivo de estudiar si los modos de transporte se ven influenciados por las condiciones meteorológicas en horas pico y horas valle, prestando especial interés a movilización pedestre y vehículos. Para este estudio se obtuvo un subconjunto de datos correspondiente a movilizaciones realizadas de lunes a viernes,

a pie y en vehículo, en los horarios 07H00, 13h00 y 18:00 que corresponden a horarios con mayor flujo de movilizaciones; y también, a las 10H00 y 15H00 que corresponden a horarios de disminución (Delgado Inga y cols., 2018). En la Figura 16 se puede observar un árbol de decisión generado con datos de movilización pertenecientes a las 18H00, en su generación se consideró las variables modo de transporte, temperatura promedio del aire y suma de precipitaciones al momento de la movilización; y distancia de viaje, variable que mostró tener relación con la variable tiempo de viaje en la sección Análisis inicial de este documento. La aparición de la temperatura del aire en uno de los nodos es remarcable, ya que en el análisis presentado en la sección anterior las variables meteorológicas no fueron consideradas.

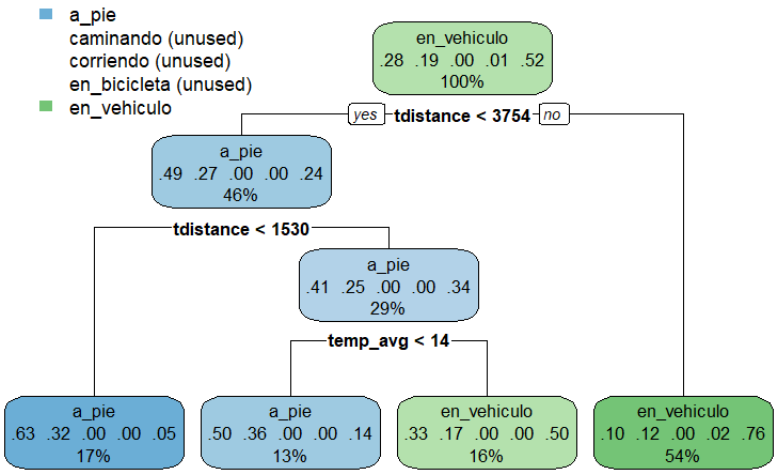

Figura 16: Árbol de decisión para movilizaciones dadas a las 18H00

Con este precedente, se usó el subconjunto de datos para la generación de un modelo de clasificación *KNN* en el que se consideraron las variables modo de movilización, distancia de viaje, día de la semana, hora de movilización, temperatura promedio y sumatoria de las precipitaciones. El software usado para la generación del modelo fue *Weka* (Frank y cols., 2016), usando validación cruzada o *Cross-validation* con 10 *folds*.

$$TPR = \frac{TP}{TP + FN} \tag{1}$$

$$FPR = \frac{FP}{FP + TN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$F_1 = \frac{2TP}{2TP + FP + FN} \tag{4}$$

$$1 - \frac{FP}{FP + TN} = \frac{TP}{TP + FN} \tag{5}$$

TP: True Positives o Verdaderos Positivos
TN: True Negatves o Verdaderos Negativos
FP: False Positives o Falsos Positivos
FN: False Negatives o Falsos Negativos

En la Tabla 3 se observan las métricas, tasa de falsos positivos (*False Positive Rate*) (1), tasa de verdaderos positivos (*True Positive Rate*) o recall (2), precisión (3), *F-measure* (4) y área *ROC* (5), obtenidas al variar k entre 1 y 5, obteniendo para un k igual a 2 un 70 % de clasificaciones correctas.

Métrica	**k = 1**	**k = 2**	**k = 3**	**k = 4**	**k = 5**
TPR	0.690	0.709	0.680	0.676	0.644
FPR	0.333	0.408	0.371	0.448	0.408
Precision	0.695	0.709	0.676	0.668	0.640
F-Measure	0.692	0.685	0.678	0.648	0.642
ROC Area	0.670	0.705	0.701	0.680	0.666

Tabla 3: Métricas obtenidas para el modelo de clasificación *KNN*, con k igual a 2

En la Figura 17 se puede observar la curva *ROC* relacionada al clasificador *KNN* generado, con k igual a 2, para modo de movilización en vehículo y a pie, de izquierda a derecha respectivamente.

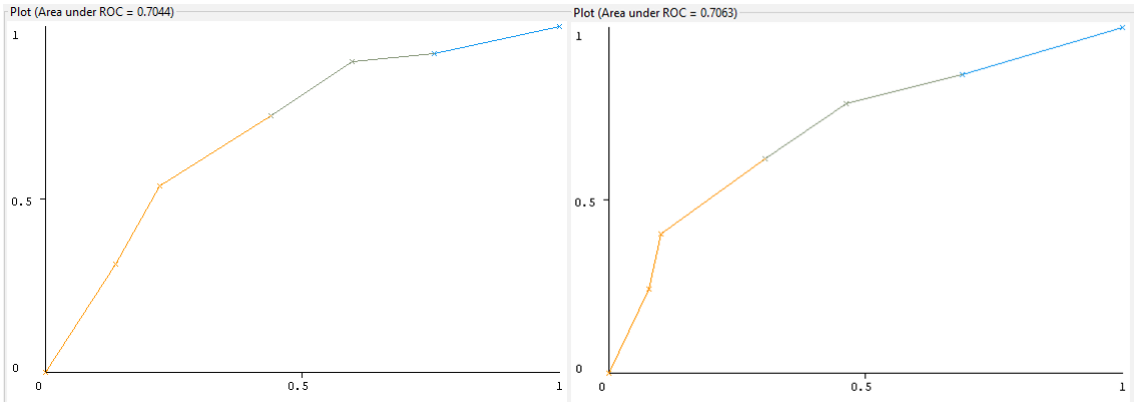

Figura 17: Curva *ROC* para el modelo de clasificación *KNN*, k igual a 2

Destinos de personas que se movilizan a pie

Este patrón requirió agrupar los destinos de personas que se movilizan a pie los sábados y domingos, considerando que estos días corresponden a un periodo de descanso y por ello las personas podrían modificar sus hábitos de movilidad, eligiendo lugares recreativos por ejemplo. El interés sobre este patrón fue analizar si los destinos elegidos por este grupo de personas son afectados por condiciones meteorológicas, como la temperatura. Es entonces necesaria la agrupación de destinos, por lo que se usó *clustering k-means*, partiendo de un

subconjunto conformado por las movilizaciones ocurridas desde las 07H00 a las 19H00, las coordenadas de destino y la temperatura promedio presente en el viaje. En la Tabla 4 se muestran las métricas internas al variar el número de clústers de dos a seis, siendo elegido un número de 5 como el que produce mejor agrupación.

Métrica	**n = 2**	**n = 3**	**n = 4**	**n = 5**	**n = 6**
Connectivity	3.6679	2.7000	3.6679	4.0171	7.9460
Dunn Index	0.1069	0.1001	0.2011	0.3313	0.3313
Silhouette width	0.6123	0.6423	0.6866	0.7532	0.7293

Tabla 4: Métricas internas de *clustering* obtenidas para agrupación destinos de movilización pedestre

En la Figura 18 se puede apreciar los destinos y su agrupación correspondiente, donde se hace notable la superioridad del grupo 5. Dado a que se pretende generar un modelo de clasificación de los grupos o clústers de destinos, se aplicó, durante el pre-procesamiento, un balanceo de clases o grupos, en el que se asigna un peso igual a cada uno de los grupos o clústers, con el objetivo de que el modelo a generar evite especializarse en la clasificación de un grupo en particular debido a su mayoritaria presencia.

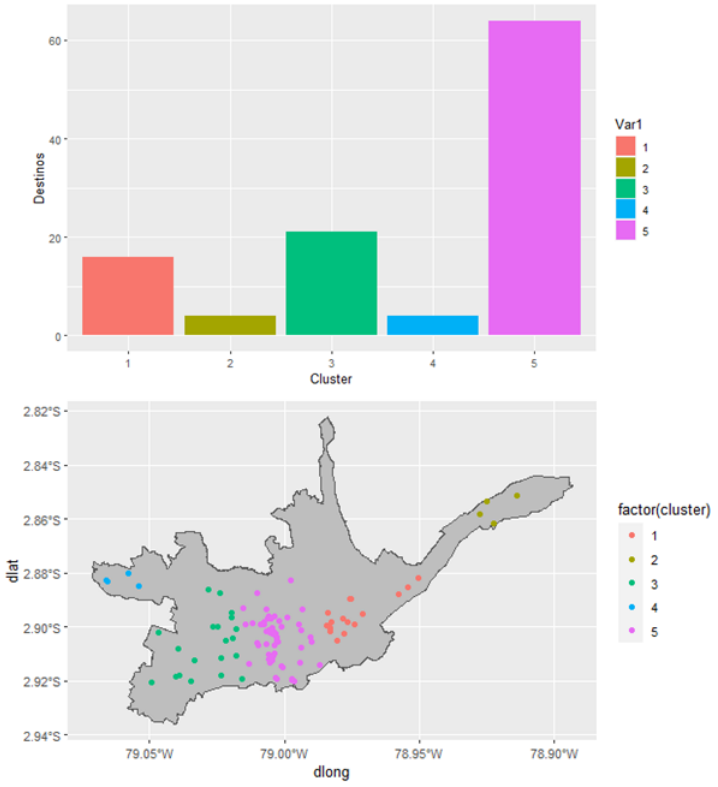

Figura 18: Clústers de destinos para personas de movilización terrestre

En la Tabla 5 se presentan los valores promedio y sumatoria de las variables meteorológicas temperatura y precipitación, respectivamente, para cada uno de los grupos o clústers

de destinos generados. Tras obtener las agrupaciones de los destinos de movilización, mediante *clustering*, se usó el subconjunto de datos resultante para la generación de un modelo de clasificación de destinos de movilización. Para esto se consideraron las variables distancia de viaje, hora del viaje, temperatura promedio, suma de precipitaciones y clústers de destinos. Se usó la técnica *KNN* con validación *Cross-validation* y 10 *folds*.

Clúster	Temperatura promedio	Suma precipitaciones
1	17.13	0.08
2	18.37	0.25
3	17.19	0.99
4	18.01	0.25
5	17.62	0.46

Tabla 5: Promedio de temperatura y suma de precipitaciones por clúster

De igual forma que para el modelo generado anteriormente, se usó el software *Weka*. Las métricas consideras para evaluar el modelo generado son tasa de falsos positivos (*FPR*), tasa de verdaderos positivos (*TPR*) o *recall*, precisión, *F-measure* y área *ROC*, presentadas en la Tabla 6. Para evidenciar la ventaja del balanceo de clases o grupos se consideraron dos escenarios, generando clasificadores con y sin data balanceada por clases o grupos. Es de notar que cuando la data no ha pasado por balanceo de clases se omiten métricas, como lo es en el caso de k igual a 4 y 5 para *precision y F-measure*; esto debido a que la clasificación obtuvo resultados nulos para grupos con presencia en minoría. El clasificador con mejores resultados es el que usa balanceo de clases y un k igual a 5.

Métrica	Sin balanceo de clases				Con balanceo de clases			
	k = 2	k = 3	k = 4	k = 5	k = 2	k = 3	k = 4	k = 5
TPR	0.330	0.477	0.468	0.523	0.185	0.197	0.301	0.310
FPR	0.301	0.362	0.366	0.373	0.204	0.201	0.175	0.173
Precision	0.423	0.479	0.460		0.131	0.185	0.294	0.300
F-Measure	0.353	0.472	0.464		0.148	0.187	0.294	0.300
ROC Area	0.569	0.583	0.635	0.626	0.541	0.530	0.624	0.610

Tabla 6: Métricas obtenidas para los clasificadores *KNN*, con y sin balanceo de clases

En la Figura 19 se observan las curvas *ROC* producidas por clasificadores *KNN* con un k igual a 4, usando la misma data y para un mismo grupo. A la izquierda sin aplicar balanceo de clases y a la derecha con balanceo de clases. El primer clasificador obtiene un 47 % de clasificaciones correctas y el segundo un 30 %, sin embargo al observar las curvas *ROC* es notable que el segundo clasificador tiene un comportamiento mejor que el primero, evidenciando la utilidad del balanceo de clases.

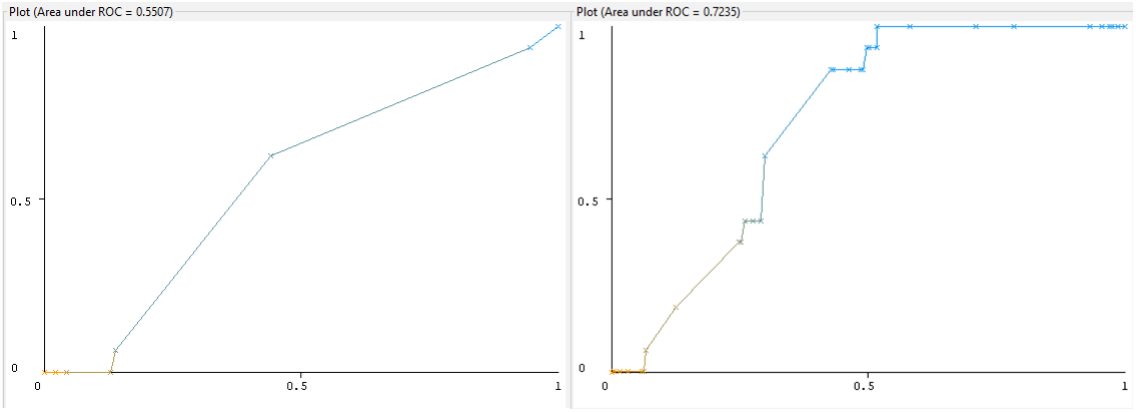

Figura 19: Comparación de curvas *ROC* para clasificadores sin balanceo de clases (izquierda) y con balanceo de clases (derecha)

DISCUSIÓN

En este estudio se analizaron datos combinados de movilidad y meteorología, buscando relaciones entre estos dos tipos de datos. Por una parte, los datos de movilidad son enfocados a estudios sobre el movimiento de las personas, mientras que los de meteorología son dirigidos a estudiar las condiciones atmosféricas y clima. Como fuente de datos meteorológicos se recolectó y usó los pertenecientes a una estación de monitoreo automático, cuyo alcance efectivo cubre casi la totalidad de la ciudad de Cuenca (Sellers Walden, 2017), y como fuente de datos de movilidad se tuvo acceso a registros de individuos que permitieron el seguimiento de su tránsito mediante el uso de dispositivos móviles con *A-GPS*.

Si bien es evidente que condiciones meteorológicas extremas pueden influir cambios en los comportamientos de movilidad de las personas (Rudloff y cols., 2015) fue necesario estudiar los dos tipos de datos de forma separada, inicialmente. De los datos de movilidad se encontró que es posible generar regresiones lineales entre el tiempo y distancia de viaje (Tabla 1); mientras que en los datos meteorológicos, se encontró correlación inversamente proporcional entre las variables, temperatura promedio del aire y suma de precipitación (Figura 7). Como paso siguiente, se aplicó un algoritmo para la combinación de *datasets*, usando las coordenadas geográficas (Figura 2) y marcas de tiempo tanto de los viajes como de las variables meteorológicas (Figura 6).

Al analizar la totalidad de los datos combinados se observó que las variables de movilidad son más influyentes en el modo de movilización por sobre las variables meteorológicas (Figura 12 y Figura 11), lo que llevó a buscar otra forma de relacionar los dos tipos de datos, esta vez usando técnicas dos técnicas de *clustering* (Tabla 2) con distintos escenarios. Sin embargo, la mejor clasificación generada por las métricas derivó en una agrupación de poco aporte al *dataset* debido a que la mayor parte de datos fueron acaparados un grupo en particular (Figura 13).

Se generó subconjuntos del *dataset* combinado, con el objetivo de estudiar los patrones: modos de movilización usados en horas pico y horas valle, y destinos de movilización de

personas que se movilizan a pie. El primer patrón analizado consideró los viajes realizados a pie y en vehículo, de lunes a viernes en los horarios 07H00, 10H00, 13H00, 15H00 y 18H00, produciendo un clasificador con 70 % de aciertos al usar las variables de entrada, modo de transporte, distancia de viaje, temperatura promedio y suma de precipitaciones. Trabajar con esta data presentó la inclusión de la variable meteorológica temperatura del aire (Figura 16), sin embargo la variable precipitación fue ignorada. El modelo generado usó la técnica *KNN* y obtuvo un 30 % de clasificaciones incorrectas, lo cual es alentador considerando que se tuvieron como entrada dos variables meteorológicas y una única variable de movilidad para clasificar el modo de movilización.

El segundo patrón estudiado consideró los viajes realizados a pie en los días sábado y domingo, entre las 07H00 y 19H00, y requirió la agrupación de los destinos de cada viaje, lo cual se realizó aplicando *clustering* basado en *k-means* (Tabla 4). El modelo de clasificación generado tomó como entrada la agrupación de destinos proveniente del *clustering*, distancia y hora de viaje, temperatura y precipitaciones, obteniendo en el mejor escenario un 30 % de clasificaciones correctas. Este patrón presentó la ventaja de aplicar balanceo de clases o grupos (Tabla 6) durante el pre-procesamiento de la data, algo que se consideró necesario debido a la mayoritaria presencia de un grupo en particular (Figura 18). Sin embargo, este modelo tuvo una alta tasa de fallos de clasificación, algo que podría estar relacionado a la poca variación de las variables meteorológicas presentes en los viajes, (Tabla 5) teniendo en cuenta que corresponden al mes de septiembre, estación seca.

Un punto que condicionó el análisis de los datos y la generación de los modelos de clasificación para los patrones estudiados fue la escasez de los mismos, principalmente en data de meteorología, considerando que el conjunto de datos combinado corresponde a 11 días del mes de septiembre de 2019. Contar con una mayor cantidad de datos y variables podría guiar a mejorar el análisis de datos y la generación de modelos, tomando en cuenta que la movilidad es un fenómeno complejo que obedece a factores adicionales que no se consideraron en este estudio, como lo son cultura o situación socio-económica.

Esta breve investigación abordó el estudio de patrones de movilidad y su relación con variables meteorológicas, haciendo uso de distintas herramientas de software libre para análisis estadístico, espacial y de minería de datos; aporta con un análisis de datos de movilidad y meteorología, un algoritmo para combinar dos tipos de *datasets* y dos modelos de clasificación que toman como entrada datos combinados de movilidad y meteorología, usando validación cruzada o *cross validation*.

CONCLUSIONES

Los estudios referentes a la movilidad urbana son de importancia por su aporte en ámbitos como plantificación de transporte, infraestructura de movilidad, e incluso salud de las personas (Louiza y cols., 2015). Proporcionan datos y análisis fundamentados que respaldan la toma de decisiones, permitiendo que los gobernantes desarrollen estrategias efectivas y eficientes para abordar los desafíos que presenta la movilidad urbana, y mejorar la calidad de vida de los ciudadanos.

Para este estudio se realizó análisis de datos de movilidad y variables meteorológicas, encontrando relaciones notables entre cada tipo de datos pero con menor fuerza al combinarlos. Se notó que existe superioridad en el uso de vehículos por sobre otros modos de transporte. Se produjeron dos modelos clasificadores con la técnica *KKN* para determinar el modo de movilización en horarios pico y horarios valle, y para destinos de movilización a pie, con un 70 % y 30 % de aciertos, respectivamente, evaluando distintos escenarios y las métricas producidas por cada uno.

Este trabajo abordó el estudio de patrones de movilidad y buscó relacionarlos con variables meteorológicas, aporta con un análisis de datos de movilidad y meteorología, un algoritmo para combinar los dos tipos de datos mencionados, y dos modelos de clasificación que toman como entrada datos combinados de movilidad y meteorología. Este estudio da paso a trabajo futuro en el que se consideren la incorporación de variables meteorológicas adicionales, como radiación solar y velocidad del viento, y uso de técnicas adicionales para la generación de los clasificadores, como redes neuronales. Si bien este trabajo de investigación uso las variables temperatura y precipitación, existen más variables meteorológicas que podrían tener influencia en los modos de movilización que un mismo individuo puede elegir; y, por otra parte, el uso de técnicas adicionales a la presentada (*KNN*) podría generar clasificadores con métricas distintas a las presentadas.

Otra área de interés al realizar estudios de movilidad tiene que ver con la reconstrucción de rutas o *path reconstruction*. La data de movilidad usada en este estudio corresponde una aproximación a los comportamientos de movilidad de los individuos y no provee de la precisión suficiente para determinar las calles o ejes viales que fueron usados en el viaje. Esto la vuelve insuficiente para realizar análisis de *hotspots* o detección embotellamientos de tráfico, un tema de interés en estudios para infraestructura de movilidad, sin antes realizar una reconstrucción de las rutas usadas.

También, respecto al análisis de data de movilidad, un área de atención es el estudio a detalle de los modos de transporte usados para un mismo viaje. Este trabajo consideró un único modo de transporte por cada viaje, sin embargo, un mismo individuo puede elegir usar mas de un modo en un mismo viaje, intercambiando entre transporte público y caminata, lo que guiaría estudios para determinar la ubicación de estaciones, paradas y rutas de transporte público e.g..

Finalmente, la incorporación de un componente descriptivo que complemente los análisis realizados sobre datos combinados de movilidad y meteorología. Si bien los datos de movilidad provenientes de dispositivos *GPS* representan un acercamiento notable a los comportamientos de movilidad, incorporar un componente descriptivo, a base de encuestas e.g., brindaría un mayor entendimiento de las razones que llevan a las personas a usar un modo de transporte u otro; ya que la movilidad es influenciada también por aspectos culturales, demográficos, económicos, sociales, entre otros.

REFERENCIAS

Akabane, A. T., Gomes, R. L., Pazzi, R. W., Madeira, E. R. M., y Villas, L. A. (2017). Apolo: A mobility pattern analysis approach to improve urban mobility. En *Globecom 2017 - 2017 ieee global communications conference* (p. 1-6). doi: 10.1109/GLOCOM.2017.8253942

Delgado Inga, O., Universidad del Azuay, Martínez Gavilanes, J., Sellers Walden, C., Salgado Castillo, F., y Carranco Zumba, S. (2018, nov). *Ruido en cuenca 2012-2018.* Casa Editora Universidad del Azuay.

Frank, E., Hall, M. A., y Witten, I. H. (2016). The WEKA workbench. online appendix for, data mining: Practical machine learning tools and techniques. Morgan Kaufmann.

Gallegos Barros, D. A., y García Torres, □. E. (2022, nov). Análisis de patrones de movilidad del transporte turístico motorizado privado en el centro histórico de cuenca ecuador.

Liu, L., Biderman, A., y Ratti, C. (2009, 01). Urban mobility landscape: Real time monitoring of urban mobility patterns.

Louiza, H., Zeroual, A., y Haddad, D. (2015, 09). Impact of the transport on the urban heat island. *INTERNATIONAL JOURNAL FOR TRAFFIC AND TRANSPORT ENGINEERING*, *5*, 252-263. doi: 10.7708/ijtte.2015.5(3).03

Orellana, D. (2016). Métodos para el análisis de patrones de movilidad no motorizada. En *Comunidades urbanas energéticamente eficientes* (p. 146–153). Editora da Universidade Federal do Espiritu Santo.

Rodríguez, L. M. (2018). Identificación de patrones de movilidad urbana.

Rojas, M., Sadeghvaziri, E., y Jin, X. (2016, 01). Comprehensive review of travel behavior and mobility pattern studies that used mobile phone data. *Transportation Research Record: Journal of the Transportation Research Board*, *2563*, 71-79. doi: 10.3141/2563-11

Rudloff, C., Leodolter, M., Bauer, D., Auer, R., Brög, W., y Kehnscherper, K. (2015, 01). Influence of weather on transport demand: A case study from the vienna region..

Sellers Walden, C. A. (2017, abril). Publicación de contaminantes atmosféricos de la estación de monitoreo de la ciudad de cuenca, utilizando servicios estándares OGC. *ACI Avances en Ciencias e Ingenierías*, *9*(1). Descargado de `https://doi.org/10.18272/aci.v9i15.300` doi: 10.18272/aci.v9i15.300

Sun, L., y Axhausen, K. W. (2016). Understanding urban mobility patterns with a probabilistic tensor factorization framework. *Transportation Research Part B: Methodological*, *91*, 511-524. Descargado de `https://www.sciencedirect.com/science/article/pii/S0191261516300261` doi: https://doi.org/10.1016/j.trb.2016.06.011

United-Nations. (2018). *Las ciudades seguirán creciendo, sobre todo en los países en desarrollo | noticias onu.* Descargado de `https://news.un.org/es/story/2018/05/1433842`

Yuan, Y., y Martin, R. (2012, 09). Extracting dynamic urban mobility patterns from mobile phone data. En (p. 354-367). doi: 10.1007/978-3-642-33024-7_26