

Comparación en la aplicación de métodos de clasificación para determinar el modo de transporte de los estudiantes para acceder a la sede Rodrigo Facio de la Universidad de Costa Rica en Montes de Oca, San José, Costa Rica

Comparison in the application of classification methods to determine the mode of transportation of students to access the Rodrigo Facio campus of the Universidad Costa Rica in Montes de Oca, San Jose, Costa Rica

Diana Sanabria Barboza

Escuela de Ingeniería Civil
Universidad de Costa Rica
San José, Costa Rica
diana.sanabriabarboza@ucr.ac.cr

Henry Hernández

Escuela de Ingeniería Civil
Universidad de Costa Rica
San José, Costa Rica
henry.hernandezvega@ucr.ac.cr

Fecha de recepción: 31/08/2021 - Fecha de aprobación: 03/01/2022

RESUMEN

Este trabajo presenta los resultados de un proceso exploratorio donde se aplicaron diferentes métodos de clasificación para determinar el modo de transporte de los estudiantes para acceder a la sede Rodrigo Facio de la Universidad de Costa Rica. Dentro de los modelos analizados se encuentran la regresión logística binomial, análisis discriminante lineal, árboles de decisión, K-vecinos más cercanos, máquinas de soporte vectorial y redes neuronales. Se realizó una validación con el método de K-pliegues y se obtuvo una precisión superior al 83% para todos los modelos analizados. De manera similar, se aplicó el modelo de ensamble apilamiento para las técnicas de árboles de decisión, K-vecinos más cercanos, máquinas de soporte vectorial, bosques aleatorios, agregación de Bootstrap, regresión logística binomial y método de potenciación obteniendo valores de precisión superiores al 86% en todos los casos, siendo el modelo de bosques aleatorios el que presentó una mayor precisión.

Palabras clave: modo de transporte, análisis multivariado, métodos de ensamble, métodos de clasificación.

ABSTRACT

This work presents the results of an exploratory process where different classification methods were applied to determine the mode of transportation for students to access the Rodrigo Facio campus of the University of Costa Rica. Among the analyzed models are binomial logistic regression, linear discriminant analysis, decision trees, K-closest neighbors, vector support machines and neural networks. A validation was carried out with the K-folds method and a precision higher than 83% was obtained for all the models analyzed. Similarly, the stacking assembly model was applied for the decision tree techniques, K-nearest neighbors, vector support machines, random forests, Bootstrap aggregation, binomial logistic regression, and the potentiation method, obtaining precision values higher than 86% in all cases. The random forest method gives the highest precision.

Keywords: mode of transport, multivariate analysis, ensemble methods, classification methods.

Introducción

El proceso de gestión de la infraestructura de transporte pasa por una etapa de planificación, en donde es necesario evaluar diferentes escenarios futuros. Las etapas de planificación también incluyen la evaluación de proyectos y políticas públicas con el fin de alcanzar metas y objetivos. Por ejemplo, la priorización del transporte público tiene un efecto significativo en la reducción de la congestión. En Costa Rica, de acuerdo con el Programa Estado de la Nación (2019), para el año 2018 la congestión, como una externalidad del transporte, significó en promedio una pérdida de 2869 dólares anuales por persona (p.161). Adicionalmente, el fomentar el transporte público puede mejorar la calidad del aire y, por ende, el bienestar social (Alpizar, Piaggio y Pacay, 2017) o contribuir a las metas país en temas del uso de energía para el transporte (Ministerio de Ambiente y Energía, 2015). Adicionalmente, Sagaris, *et al.* (2021) mencionan los beneficios a la población relacionados con la reducción del uso del vehículo privado. De acuerdo con Jiménez-Serpa, Rojas-Sánchez, y Salas-Rondón (2015) la implementación de medidas como la integración tarifaria son fundamentales para proveer un transporte público eficiente.

Para evaluar el impacto de los diferentes proyectos y políticas públicas en la selección de modo de transportes, es necesaria su modelación. De acuerdo con Garber y Hoel (2005) una de las etapas del análisis de demanda corresponde a la selección de modo, el cual permite determinar el porcentaje de viajes que se realizan en vehículo particular y en transporte público. La elección del modo es una de las etapas más importantes en el proceso de planificación del transporte y tiene un impacto directo en las decisiones de formulación de políticas de transporte (Sekhar, 2014).

Por otro lado, el identificar el modo de transporte puede ofrecer varias ventajas en diferentes campos de la ingeniería del transporte como la planificación del transporte y los sistemas de transporte inteligentes que conducen a una amplia gama de aplicaciones con impactos medioambientales y de seguridad vial (Jahangiri y Rakha, 2014).

Conceptualización

Las modelos de elección de modo de transporte más comúnmente utilizados son los modelos Logit multinomial (creado en 1970) y Logit anidado. El primero asume que las alternativas son independientes entre sí y el segundo se utiliza generalmente en casos que involucran modos combinados, en los que las alternativas se agrupan en nido, según Bekhor y Shiftan (2009). Por ejemplo, en Costa Rica, se utilizó en los años noventa el modelo Logit para estimar la distribución modal entre el transporte público y el vehículo privado para el Estudio del Transporte Urbano del Gran Área Metropolitana de 1991. Similarmente, se aplicaron modelos Logit para el proyecto de Planificación Regional y Urbana de la GAM (PRUGAM) en el año 2007 y para el Modelo de Demanda de Transporte Urbano en la GAM del 2016 para estimar la distribución modal entre el transporte público y el vehículo privado (Castro-Rodríguez, Picado-Aguilar y Rodríguez-Shum, 2018). Otro ejemplo a nivel latinoamericano es el de París-Bravo (2019), quien aplicó modelos Logit multinomiales para determinar los factores que intervienen en la selección de modo acceder al Transmilenio en Bogotá, Colombia o los aplicados por Jiménez-Serpa y Salas-Rondón (2016) para estudiar los factores que influyen en la selección de taxi compartido como opción para viajes relacionados con el aeropuerto en el departamento Santander en Colombia. Similarmente, Jiménez-Serpa y Salas-Rondón (2017) aplicaron modelos logit para estudiar la probabilidad de uso del vehículo particular en caso de implementarse un sistema de tarificación por congestión y Salas-Rondón, Jiménez-Serpa, y Martínez-Estupiñán (2021), quienes aplicaron este tipo de modelos para evaluar las acciones para mejorar la futura operación de los sistemas estratégicos de transporte público.

Para estimar la utilidad de un modo de transporte, lo usual es usar una combinación lineal de atributos para evaluar el costo que representa para el usuario el utilizar cierto modo de transporte. En el campo de la ingeniería del transporte, el modelo más popular es el Logit múltiple (multinomial) (Ortúzar, 2012).

Sin embargo, después del año 2007 se ha visto un incremento significativo en las publicaciones científicas relacionadas con la aplicación de técnicas de clasificación de aprendizaje de máquinas (Hillel, Bierlaire, Elshafie y Jin, 2020). Por ejemplo, Zenina y Borisov (2011) utilizaron análisis discriminantes escalonados y progresivos para la elección del modo del modelo y compararon los resultados con el modelo Multinomial Logit (MNL, por sus siglas en inglés) y los modelos de árbol de decisión (DT, por sus siglas en inglés).

Adicionalmente, Sekhar (2014) indica que modelos como el logit, redes neuronales artificiales, sets de teoría difusa y neuro-difusos tienen alta precisión. Por otro lado, Souza Pitombo, Schindler Gomes Da Costa y Salgueiro (2015) aplicaron la metodología de árboles de decisión donde la mayoría de los encuestados usaban el modo motorizado privado (60,6%), seguido del modo no motorizado (21,2%) y el transporte público (18,2%). La variable más importante (que explica mejor la variabilidad de los datos teniendo en cuenta la elección del modo) fue "Licencia de conducir", que divide los datos en dos ramas principales.

Vassilev (2018) combinó tres herramientas de minería de datos: análisis de componentes principales, distancia de Mahalanobis y análisis discriminante lineal para explorar el modo de transporte utilizado, a partir de datos recolectados con teléfonos celulares. Similarmente, Hillel, Elshafie y Jin (2018) aplicaron la metodología de árboles de decisión con aumento de gradiente estocástico y realizaron una validación de K-pliegues. Adicionalmente, Sekhar y Madhu (2016) aplicaron el modelo de bosque aleatorio para modelar la selección de modo en Delhi, India.

Existen estudios donde se determina el modo de transporte a partir de datos recolectados de manera automática. Por ejemplo, Dabiri y Heaslip (2018) aplicaron redes neuronales convolucionales para predecir el modo de transporte a partir de datos obtenidos de Sistemas de Posicionamiento Global. Similarmente, Jahangiri y Rakha (2014) aplicaron máquinas de soporte vectorial utilizando datos de diferentes sensores para predecir el modo de transporte utilizados por los diferentes usuarios.

Metodología

El objetivo general de esta investigación es comparar la precisión de técnicas de clasificación para la determinación del modo de transporte que utilizan los estudiantes activos de la Universidad de Costa Rica en el 2018 con base en su frecuencia de visita, las características del hogar y la ubicación de su residencia.

Interesa explorar el rendimiento de modelos paramétricos de clasificación posteriores al modelo Logit multinomial, basadas en la conformación de grupos de personas de un entorno común con base en coincidencia de características y relaciones de independencia entre sí: tanto técnicas basadas en cálculos estadísticos, como los modelos previamente mencionados y el análisis de discriminante lineal, así como otras basadas en inteligencia artificial como árboles de decisión y clasificación, K-vecinos más cercanos, máquinas de soporte vectorial y redes neuronales.

El registro utilizado para esta investigación es la Encuesta de Transporte UCR llevada a cabo por el LanammeUCR (Hernández-Vega y Umaña-Marín, 2018), la cual tuvo el objetivo de caracterizar a los viajes de las personas usuarias de las instalaciones de la Universidad de Costa Rica (UCR), identificando zonas de generación de viajes, perfil personal y de viajes, medios de transporte utilizados, rutas empleadas, entre otros. El cuestionario se aplicó en formato electrónico mediante la plataforma LimeSurvey, tomando en cuenta aspectos prácticos, especialmente considerando su bajo costo. Se enviaron correos electrónicos al estudiantado y al personal docente y administrativo de la universidad. Las respuestas recibidas al cuestionario fueron totalmente voluntarias, por lo que es de esperar de que exista un sesgo de participación.

El estudiantado y el personal docente y administrativo tuvieron acceso a la encuesta durante el mes de junio del 2018, a través del enlace <http://www.encuestas.ucr.ac.cr>. La duración promedio del llenado de las preguntas fue de 13 minutos para el caso de los estudiantes y de 12 minutos para el caso de las personas funcionarias de la universidad (Hernández-Vega y Umaña-Marín, 2018).

La base de datos disponible cuenta con 1733 registros. En cuanto a tipo de usuario cabía la posibilidad de seleccionar múltiples roles por respuesta, se resalta que el 82,23% declaró ser estudiante regular de la Universidad (diplomado, bachillerato o licenciatura), paralelamente 7,16% se identificó como estudiante de posgrado, 6,41% como docente y 6,06% como funcionario administrativo, las personas que también se denominaron deportistas, visitantes, estudiantes de curso libre, egresados, proveedores no alcanzan el 1% en cada categoría.

Así, el presente trabajo buscó enfocarse en personas usuarias que son estudiantes regulares, cuyo nivel más alto de educación declarado es universitaria incompleta y son estudiantes activos. Además, el análisis se limitó a estudiantes que visitan el campus universitario Rodrigo Facio, en San Pedro Montes de Oca, y excluyendo a quienes visitan exclusivamente alguna otra sede. Finalmente, interesó considerar datos de ingreso monetario, por lo cual se excluyeron los registros que no contaban con respuestas asociadas. La selección final se compone de 626 registros válidos.

Seguidamente, se detalla la totalidad variables de la encuesta consideradas para el presente trabajo, además de su descripción y tratamiento preliminar. En la mayoría se agruparon ciertas categorías originales para generar rangos representativos y evitar rangos con poca representación de respuestas:

- **Edad.** Variable discreta, edad de los estudiantes, cuyo rango va desde los 17 a los 47 años, con una media de 21,8 años. El 75% de las personas tenía 23 años o menos.
- **Sexo.** La muestra está compuesta exactamente por 50% estudiantes mujeres y 50% hombres.
- **Ingreso bruto mensual del hogar.** Esta variable es categórica, tiene rangos que van desde menos de 250 mil colones (443 dólares de los Estados Unidos de América) mensuales hasta 1 millón de colones (1773 dólares de los Estados Unidos de América), los rangos superiores a este valor se agruparon en la categoría "Más de 1 millón".
- **Cantidad de personas en el hogar.** Variable categórica, número de personas que conforman el hogar de la persona encuestada en rangos: uno o dos miembros, tres, cuatro, más de cinco miembros.

- **Cantidad de automóviles en el hogar.** Variable categórica que ofrece tres opciones: 0, 1 o más de 2.

Las siguientes variables que se refieren a la provincia, consideran la dinámica de movilidad de los estudiantes que viven fuera de la Gran Área Metropolitana, principalmente, y su movilización a residencias alquiladas o temporales en el campus durante el periodo lectivo, y su retorno a su lugar de origen en los periodos de receso de lecciones: fines de semana y meses de vacaciones.

- **Provincia de residencia (entre semana, de marzo a junio y de julio a diciembre).**
- **Provincia de residencia durante los fines de semana.**
- **Provincia de residencia durante los meses de enero y febrero.**
- **Fincas de la sede Rodrigo Facio que visita.** Cantidad de fincas del campus que suelen visitar: campus central, Ciudad de la Investigación e Instalaciones Deportivas, por lo cual las respuestas se categorizaron, respectivamente, en 1, 2 o 3.
- **Días de la semana que visita la Universidad.** Cantidad de días que visita la Universidad en una semana, los valores parten desde 1 día hasta la categoría 5 o más.
- **Tiempo promedio de llegada a la Universidad.** Tiempo en rangos que parten desde los 15 minutos hasta más de 1 hora.
- **Tiempo promedio de retorno de la Universidad.** Tiempo en rangos que parten desde los 15 minutos hasta más de 1 hora.

Entre todas las posibilidades que ofrece la información de la encuesta, la variable categórica que se planteó como respuesta de los modelos por construir es el **modo de transporte** que cada persona utiliza más frecuentemente (es decir, que se está ante una encuesta de preferencias reveladas). Considerando el objetivo de la investigación y la versatilidad que una variable dependiente dicotómica ofrece para la simplificación y aplicación de diversos métodos, los modos de transporte encuestados se categorizaron previamente en dos grupos:

- **Movilidad activa - Servicio público colectivo (MASP):** camina, bicicleta, servicio de autobús externo de la universidad (conocido como ruta bus UCR) para el transporte de estudiantes, bus regular, tren, otro (87,06% de la selección final).
- **Vehículo particular o servicio público individual (motorizado) (VP):** Carro (conductor o pasajero), Moto (conductor o pasajero), Uber, Taxi (12,94%).

Para alcanzar el objetivo planteado primeramente se aplicaron seis técnicas diferentes al 80% del conjunto de datos partiendo de la regresión logística binomial, en combinación con métodos de validación para evitar el sesgo de estimación de la precisión en las predicciones al emplear el conjunto total de registros para su construcción. El 20% restante se reservó para corroboraciones finales de validación de los resultados.

La validación empleada parte del principio de retención, utilizando una parte de los datos para construir el modelo y otra parte para ensayarlo, proceso que se optimiza al aprovechar la técnica de validación cruzada K-pliegues que genera grupos aleatorios del total de datos. De esta manera, es posible iterar la construcción de cada modelo excluyendo determinado "K-pliegue" a la vez,

calculando los indicadores de precisión y, finalmente, promediando los resultados de todas las iteraciones.

En una segunda etapa, se exploraron los resultados de la combinación de modelos predictivos, a través de los modelos de ensambles que incluyen la generación de pliegues de construcción e iteraciones: Agregación de Bootstrap, Bosques Aleatorios, Método de potenciación y Apilamiento.

Resultados y discusión

Para la evaluación de los modelos se aplicó un método de validación cruzada de K-pliegues 10 pliegues y 5 repeticiones, y se obtuvo una precisión superior al 87% para todos los modelos analizados, con excepción de la aplicación de redes neuronales (83%). La técnica de K-vecinos más cercanos presentó la precisión más alta, tal como se aprecia en el Cuadro 1.

Respecto a la regresión logística binomial aplicada, solamente las variables, ningún vehículo en el hogar, un vehículo en el hogar, si el estudiante reside fuera de la Gran Área Metropolitana (GAM) los fines de semana, o si el tiempo de llegada es entre 45 y 60 minutos, resultaron ser significativos en el modelo, como se evidencia en el Cuadro 2.

Cuadro 1. Medidas promedio de evaluación de predicciones de modo de transporte estudiantil, según modelos de clasificación base construidos y aplicación de validación cruzada K-pliegues

Técnica	Error	Precisión	Sensibilidad	Especificidad
Regresión Logística Binomial	0,10609	0,89391	0,96798	0,38500
Análisis Discriminante Lineal	0,12202	0,87798	0,95173	0,38905
Árboles de decisión	0,12386	0,87614	0,96353	0,27881
K-vecinos más cercanos	0,08000	0,92000	0,97268	0,33405
Máquinas de soporte vectorial	0,11798	0,88202	0,97475	0,24048
Redes neuronales	0,16115	0,83885	0,90981	0,33833

Cuadro 2. Coeficientes significativos del modelo de Regresión Logística Binomial de predicciones de modo de transporte estudiantil

Variable	Coeficiente	Estadístico z	Probabilidad (> z)	Significancia
Ningún vehículo en el hogar	0,057	-3,0217	0,0025	**
Un vehículo en el hogar	0,121	-3,4757	0,0005	***
Fuera de la GAM los fines de semana	0,16	-1,9213	0,0547	.
Tiempo de llegada entre 45-60 min.	0,081	-2,1636	0,0305	*

Similarmente, el método de árbol de decisión (ver Figura 1) utiliza la tenencia de vehículo en el hogar como primer criterio de selección, seguido del tiempo de retorno, del ingreso en el hogar, si el estudiante

vive fuera de la Gran Área Metropolitana (GAM) durante los fines de semana, seguido del tiempo de retorno (en este caso entre 45 y 60 minutos).

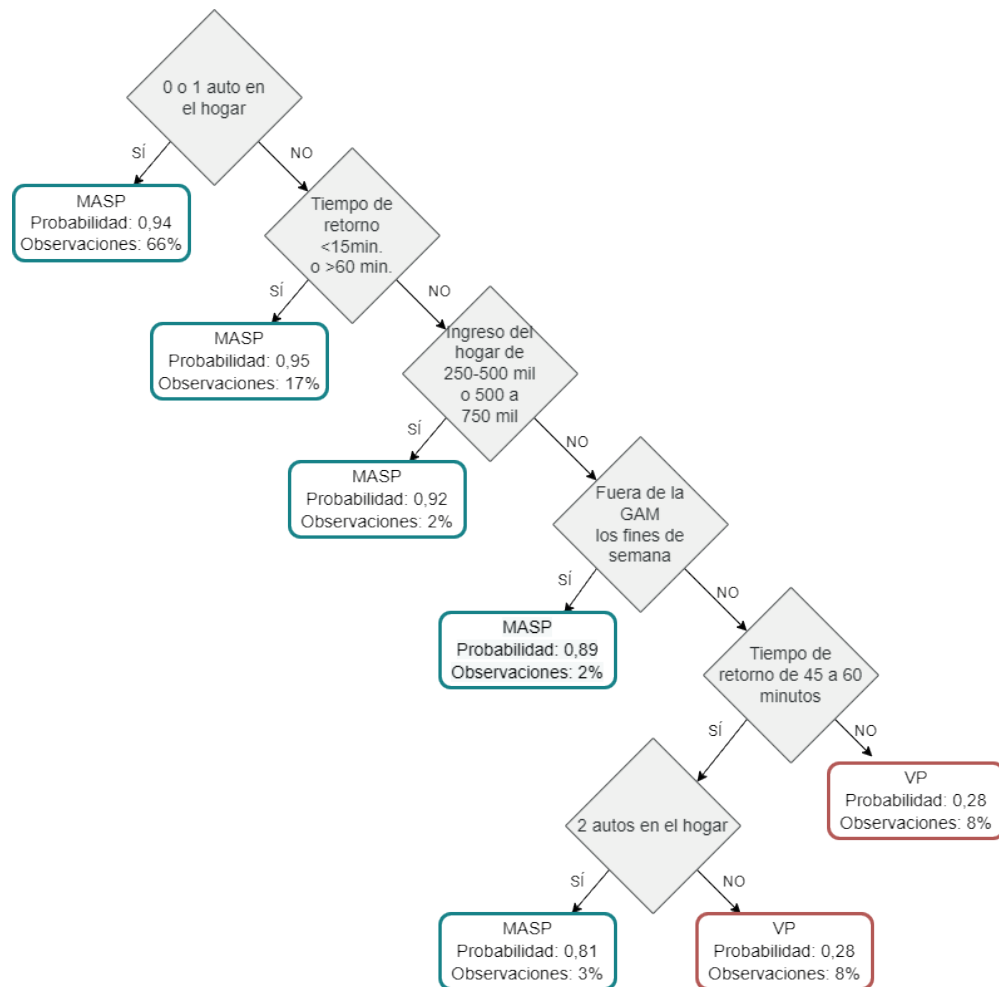


Figura 1. Árbol de decisión y clasificación del modo de transporte de estudiantes de la Universidad de Costa Rica

Específicamente, para el modelo de K-vecinos, y con base en un 20% de los datos originales reservados para validación, se obtuvo una precisión superior al 85%, una sensibilidad del 98% y una especificidad del 11%, como se detalla en el Cuadro 3.

Respecto a la precisión de predicciones de modo de transporte, se aplicó el modelo de ensamble de apilamiento y para las diferentes técnicas analizadas se obtuvieron precisiones con medias superiores al 86% (ver Cuadro 4), resultados muy similares se obtuvieron con los modelos individuales.

Cuadro 3. Medidas promedio de evaluación de predicciones de modo de transporte estudiantil, según el modelo K-vecinos más cercanos y con base en 20% de los datos originales reservados para validación

Técnica	Error	Precisión	Sensibilidad	Especificidad
K-vecinos más cercanos	0,14286	0,85714	0,98148	0,11111

Cuadro 4. Distribución estadística de la precisión de predicciones de modo de transporte estudiantil según el modelo de ensamble Apilamiento, para diferentes técnicas

Técnica	Mínimo	Primer cuartil	Mediana	Media	Tercer cuartil	Máximo
Árboles de decisión	0,8000	0,8600	0,8776	0,8676	0,8800	0,9000
K-vecinos más cercanos	0,8235	0,8627	0,8800	0,8749	0,8800	0,9184
Máquinas de soporte vectorial	0,8235	0,8782	0,8824	0,8885	0,9015	0,9388
Bosques aleatorios	0,8400	0,8800	0,8980	0,8937	0,9184	0,9600
Agregación de Bootstrap	0,8000	0,8600	0,8788	0,8789	0,9000	0,9600
Regresión Logística Binomial	0,8235	0,8600	0,8800	0,8765	0,9000	0,9600
Método de potenciación	0,8235	0,8782	0,8824	0,8901	0,9000	0,9600

Cuadro 5. Medidas promedio de evaluación de predicciones de modo de transporte estudiantil, según modelos de clasificación de ensamble

Modelo	Error	Precisión
Agregación de Bootstrap	0,12090	0,87910
Bosques Aleatorios	0,14180	0,85820
Método de potenciación	0,11800	0,88200
Apilamiento	0,10700	0,89300

Las medidas promedio de evaluación de predicciones para el modelo de ensamble Apilamiento con base en 20% de los datos originales reservados para validación tienen una precisión del 89% siendo este modelo el que presenta mejores resultados, según lo observado en el Cuadro 5.

Los resultados obtenidos se encuentran dentro de los rangos obtenidos por Hagenauer y Helbich (2017) y superiores a los reportados por Omrani (2015). Al comparar los resultados obtenidos en el presente análisis con otros estudios de tenencia de vehículo, por ejemplo, se puede decir presentan valores de precisión ligeramente inferiores y similares a los encontrados por Muhsin Zambang, Jiang, y Wahab (2021) en un estudio realizado en Ghana, pero superiores a los determinados por Kaewwichian, Tanwanichkul y Pitaksringkarn (2019) en otro estudio similar en Tailandia. En estos estudios se aplicaron algunas de las técnicas incluidas en el presente análisis.

Conclusiones y recomendaciones

En el presente estudio se logró comparar la precisión de técnicas de clasificación para la determinación del modo de transporte que utilizan los estudiantes activos de la Universidad de Costa Rica en el 2018 con base en su frecuencia de visita, las características del hogar y la ubicación de su residencia. Los métodos estudiados brindan en términos generales buenos resultados, por lo que su aplicación podría tener un potencial similar a los métodos logit tradicionales.

Respecto a los modelos de clasificación base construidos y aplicación de validación cruzada K-pliegues, la técnica de K-vecinos más cercanos presentó la precisión más alta. Cuando se aplicó el modelo de ensamble de apilamiento y para las diferentes técnicas analizadas se obtuvieron precisiones con medias superiores al 86%, siendo la técnica de bosques aleatorios la que presentó resultados ligeramente superiores a las otras técnicas. Los resultados fueron muy similares a los que se obtuvieron con los modelos individuales.

Se recomienda explorar, en estudios posteriores, otros métodos de clasificación de aprendizaje bayesiano como el explorador bayesiano ingenuo aplicado por Hagenauer y Helbich (2017), o realizar pruebas de hipótesis, como la Kruskal-Wallis sobre las diferencias de rendimiento entre los clasificadores que también fue aplicada por Hagenauer y Helbich (2017).

Sería interesante utilizar en futuros proyectos similares otros modelos para la clasificación, como el enfoque de aprendizaje automático supervisado utilizado por Bjerre-Nielsen, Minor, Sapieżyński, Lehmann y Lassen (2020) para determinar el modo de transporte utilizado a partir de datos de Wi-fi y Bluetooth.

Finalmente, sería recomendable aplicar estos métodos con datos más recientes con el fin de valorar posibles variaciones recientes en los patrones de movilidad.

Agradecimientos

Se agradece a Deiby Solano, profesor del posgrado en Estadística en la Universidad de Costa Rica, los comentarios y sugerencias para mejorar el presente estudio. Además, se agradece al Laboratorio Nacional de Materiales y Modelos Estructurales de la Universidad de Costa Rica por proveer la información requerida para el presente estudio.

Referencias

- Alpizar, F., Piaggio, M., y Pacay, E. (2017). *Valoración económica de los beneficios en la salud asociados a la reducción de la contaminación del aire*. Santiago, Chile: CEPAL.
- Bekhor, S., y Shiftan, Y. (2009). Specification and Estimation of Mode Choice Model Capturing Similarity between Mixed Auto and Transit Alternatives. *Journal of Choice Modelling*, 3(2), 29-49. DOI: 10.1016/S1755-5345(13)70034-4
- Bjerre-Nielsen, A., Minor, K., Sapieżyński, P., Lehmann, S., y Lassen, D. D. (2020). Inferring transportation mode from smartphone sensors: Evaluating the potential of Wi-Fi and Bluetooth. *PLOS ONE*, 15(7), e0234003. DOI: 10.1371/journal.pone.0234003

- Castro-Rodríguez, L, Picado-Aguilar, G., y Rodríguez-Shum, S. (2018). Evolución histórica de la modelación de demanda de transporte urbano en Costa Rica. *Infraestructura Vial*, 20, 4-47. DOI: 10.15517/iv.v20i1.33541
- Dabiri, S., y Heaslip, K. (2018). Inferring transportation modes from GPS trajectories using a convolutional neural network. *Transportation Research Part C: Emerging Technologies*, 86, 360-371. DOI: 10.1016/j.trc.2017.11.021
- Garber, N. y Hoel, L. (2005). *Ingeniería de tránsito y carreteras*. México, México D. F.: Thomson.
- Hagenauer, J., y Helbich, M. (2017). A comparative study of machine learning classifiers for modeling travel mode choice. *Expert Systems with Applications*, 78, 273-282. DOI: 10.1016/j.eswa.2017.01.057
- Hernández-Vega, H. y Umaña-Marín, G. (2018). Encuesta de Transporte 2018. Sede Rodrigo Facio Universidad de Costa Rica. Recuperado de: <https://www.lanamme.ucr.ac.cr/repositorio/handle/50625112500/1582>
- Hillel, T., Bierlaire, M., Elshafie, M., y Jin, Y. (2020). A systematic review of machine learning classification methodologies for modelling passenger mode choice. *Journal of Choice Modelling*, 38, 100221. DOI: 10.1016/j.jocm.2020.100221
- Hillel, T., Elshafie, M. Z., y Jin, Y. (2018). Recreating passenger mode choice-sets for transport simulation: A case study of London, UK. *Proceedings of the Institution of Civil Engineers-Smart Infrastructure and Construction*, 171(1), 29-42. DOI: 10.1680/jsmic.17.00018
- Jahangiri, A., y Rakha, H. (2014). Developing a support vector machine (SVM) classifier for transportation mode identification by using mobile phone sensor data. En *Transportation Research Board 93rd Annual Meeting*. Conferencia de Transportation Research Board, Washington D.C., Estados Unidos.
- Jiménez-Serpa, J. C., Rojas-Sánchez, A. E., y Salas-Rondón, M. H. (2015). Tariff integration for public transportation in the metropolitan area of Bucaramanga. *INGE CUC*, 11(1), 25-33. DOI: 10.17981/ingecuc.11.1.2015.02
- Jiménez-Serpa, J. C., y Salas-Rondón, M. H. (2016). Un caso de estudio sobre los factores que influyen para viajar en taxi compartido desde y hacia el aeropuerto. *Ingeniería de Transporte*, 20(01), 33-46.
- Jiménez-Serpa, J. C., y Salas-Rondón, M. H. (2017). Aplicación de modelos econométricos para estimar la aceptabilidad de una tasa por congestión vehicular. *INGE CUC*, 13(2), 60-78. DOI: 10.17981/ingecuc.13.2.2017.08
- Kaewwichian, P., Tanwanichkul, L., y Pitaksringkarn, J. (2019). Car ownership demand modeling using machine learning: Decision trees and neural networks. *International Journal of GEOMATE*, 17(62), 219-230. DOI: 10.21660/2019.62.94618
- Ministerio de Ambiente y Energía (2015). VII Plan Nacional de Energía 2015-2030. Programa de las Naciones Unidas para el Desarrollo PNUD. Recuperado de: <https://minae.go.cr/recursos/2015/pdf/VII-PNE.pdf>
- Muhsin Zambang, M. A., Jiang, H., y Wahab, L. (2021). Modeling vehicle ownership with machine learning techniques in the Greater Tamale Area, Ghana. *PLOS ONE*, 16(2), e0246044. DOI: 10.1371/journal.pone.0246044
- Omrani, H. (2015). Predicting travel mode of individuals by machine learning. *Transportation Research Procedia*, 10, 840-849. DOI: 10.1016/j.trpro.2015.09.037
- Ortúzar, J.D. (2012). *Modelos de demanda de transporte*. Santiago, Chile: Ediciones UC.
- París-Bravo, D. (2019). *Factores determinantes para la selección de modo auxiliar para acceder a Transmilenio* (Tesis de grado). Universidad de los Andes, Colombia.

- Programa Estado de la Nación (2019). Informe Estado de la Nación 2019. Recuperado de: https://estadonacion.or.cr/wpcontent/uploads/2019/11/informe_estado_nacion_2019.pdf
- Sagaris, L., Mindell, J., Rojas-Rueda, D., Cortínez-O’Ryan, A., Sadrangani, K., Casanave-Macías, J., González-Sánchez, Y., y Hernández-Vega, H. (2021). *Transporte, Salud, Equidad Acercamientos urgentes en un mundo con y post-Covid 19*. Recuperado de: <https://www.cambiarnos.cl/transporte-salud-y-equidad/>
- Salas-Rondón, M. H., Jiménez-Serpa, J. C., y Martínez-Estupiñán, Y. F. (2021). Subsidio a la tarifa para fortalecer la operación de los sistemas estratégicos de transporte público en Colombia. *Revista UIS Ingenierías*, 20(3), 77-90. DOI: 10.18273/revuin.v20n3-2021005
- Sekhar, C. (2014). Mode Choice Analysis: The Data, the Models and Future Ahead. *International Journal for Traffic & Transport Engineering*, 4(3), 269 - 285. DOI: 10.7708/ijtte.2014.4(3).03
- Sekhar, C. R., y Madhu, E. (2016). Mode choice analysis using random forrest decision trees. *Transportation Research Procedia*, 17, 644-652. DOI: 10.1016/j.trpro.2016.11.119
- Souza Pitombo, C., Schindler Gomes Da Costa, A., y Salgueiro, A. R. (2015). Proposal of a sequential method for spatial interpolation of mode choice. *Boletim de Ciências Geodésicas*, 21(2), 274-289. DOI: 10.1590/S1982-21702015000200016
- Vassilev, A. (2018). Data Mining Applied to Transportation Mode Classification Problem. *Proceedings of the 4th International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS)*, 36-46. DOI: 10.5220/0006633300360046
- Zenina, N., y Borisov, A. (2011). Transportation Mode Choice Analysis Based on Classification Methods. *Scientific Journal of Riga Technical University Computer Sciences*, 45(1), 49-53. DOI: 10.2478/v10143-011-0041-2

Los contenidos de este documento, representan insumos dentro de un proceso de gestión del conocimiento que, por sí mismos, a priori, no constituyen una declaración de una normativa, procedimiento, criterio o herramienta oficial de acatamiento obligatorio en la gestión de proyectos de obra vial pública de Costa Rica, por parte del LanammeUCR. Cualquier posición oficial para Costa Rica sobre aspectos puntuales contemplados en este documento, se realizarán por los medios que corresponden, según los lineamientos de la Universidad de Costa Rica, de la Ley 8114 y su Reglamento al Art 6 (Decreto 37016 – MOPT).