Linguistic and Philosophical Investigations
ISSN: 1841-2394, e-ISSN: 2471-0881
Vol 23 (1), 2024
pp. 1198 - 1211

Advances in Solutions on Public Transportation Problems with Artificial Intelligence

[1]Carlos D. Pimentel-Díaz, [2]Flor E. Romero-Rueda, [3]Eduardo J. Duarte-Peña

[1]Facultad Ingeniería Mecatrónica, Universidad Santo Tomás Seccional Bucaramanga

Carrera 18 No. 9-27, 680002

Bucaramanga Colombia

[2]Facultad Ingeniería Industrial, Universidad Santo Tomás Seccional Bucaramanga

Carrera 18 No. 9-27, 680002

Bucaramanga Colombia

flor.romero01@ustabuca.edu.co

[3]Facultad Ingeniería Mecatrónica, Universidad Santo Tomás Seccional Bucaramanga

Carrera 18 No. 9-27, 680002

Bucaramanga Colombia

Abstract: The public transportation service offered in medium and large cities has been under scrutiny in recent years due to declining user numbers. This trend is commonly attributed to internal factors such as poor management or suboptimal routes leading to extended waiting times, as well as external factors like competition from informal transportation services, which have gained popularity through mobility applications, among others. This article compiles a series of studies focused on the application of artificial intelligence to tackle various issues within public transportation services, published in recent years. The goal is to determine the most suitable technique among those implemented to address these challenges. Analyzing the research presented herein allowed for the identification of public transportation-related problems encountered by researchers and the techniques that yielded superior outcomes. It is concluded that techniques like Random Forest (RF) and Long Short-Term Memory (LSTM) neural networks are better suited for resolving problems requiring geographical coordinate data, latitude, longitude, and time series.

Keywords - Artificial intelligence, public transport, random forest, neural networks.

Inteligencia artificial, transporte público, bosques aleatorios, redes neuronales.

Received: 12 May 2024 **Revised**: 06 June 2024 **Accepted**: 22 June 2024

1. Introduction

The mobility of citizens in any country is essential to support economic growth, create jobs and connect people with essential services such as health care or education. In the field of public transportation, developing nations face a double challenge. On the one hand, they must ensure that mobility is efficient, safe, and economical for all their citizens, while, on the other hand, they must achieve this with a reduced climate footprint [1].

The importance of public transportation not only lies in the accessibility it gives people a way to move around the city, but it is also an important factor if we want to reduce traffic congestion [2]. However, getting around daily can be complicated due to traffic, delays on routes, the lack of buses to meet demand or the constant congestion that usually occurs in passenger transportation services. These problems increasingly motivate people to use different mobility applications such as Uber, InDriver or DiDi that provide safety, convenience and save time [3]. Due to this, we seek to implement Artificial Intelligence (AI) tools in public transportation, to provide a more effective and sustainable service.

Lately, due to the rise of artificial intelligence and sciences derived from it such as Data Science, Big data, Machine learning and Deep learning, these have been entering public management and administration [4]. New advances within AI have created new opportunities for continuous innovation [5]. What is leading us to a new era of digital disruption and institutions, companies and organizations must be prepared for it [6].

Within the search for solutions to public transport problems, the term route optimization is normally used. AI offers advanced analytical capabilities to precisely optimize routes. An optimized route saves time and fuel, reducing the expense of the provider company and waiting times for users, improving the quality of service [7]. Optimized route management is a crucial issue to improve efficiency in transportation, delivery, and distribution [8].

The document brings together research carried out on topics related to public transportation. The different authors worked with predictive models, which were designed using various AI techniques, with the purpose of addressing relevant issues in public transportation, such as time optimization, route management, decision making, among others.

2. AI applications in public transportation

In 2013, a desktop application was developed capable of generating a passenger transportation route using Genetic Algorithms (GA) supported by the Google Maps API. The algorithm takes the addresses of each passenger and groups them, as much as possible, based on certain parameters into stops using clustering techniques. It determines the optimal way to pick them up and transport them, reflecting the route on real maps [9].

To work on traffic issues in cities, a prototype was developed for a predictive model of bus arrival times intended for implementation in an application providing information services for Chile's public transportation. Two types of neural networks are employed: the first model utilizes a Long Short-Term Memory (LSTM) artificial neural network, while the second model is a hybrid, incorporating a Particle Filter (PF), which implements a Bayesian Filter through Monte Carlo simulations, complementing its operation with data obtained from a Bayesian neural network. The research concludes with positive results, measuring the performance of the second model using Root Mean Square Error (RMSE), alongside a recommendation to implement this model within the application [10]. Table 1 demonstrates that the PF technique yielded superior results for various routes across different times of the day, in contrast to the LSTM technique.

Table 1 Comparison between results of LSTM and Particle Filter models.

Route	Schedule	RMSE LSTM	RMSE FP
301	Off-peak hours	2.66	0.32
	Rush hour	4.99	1.56
315e	Off-peak hours	6.44	0.26
	Rush hour	14.35	1.30
506	Off-peak hours	3.11	0.91
	Rush hour	6.67	0.26

Taken and adapted from [10].

In another investigation, route data was analyzed within the Integrated Public Transportation System of Bogotá, Colombia, aiming to propose the implementation of a machine learning-based prediction model. Although the technique used for model training is not explicitly mentioned, it focuses on scrutinizing user behavior across days to highlight periods when the analyzed route experiences higher usage. This insight aids in making administrative decisions regarding the service. Figure 1 showcases the model results. The X-axis comprises dates at 15-day intervals, while the Y-axis illustrates the frequency of route usage by users. This investigation concludes with the development of a predictive tool for resource and capacity allocation for the concessionaire Consorcio Express S.A., additionally revealing bottlenecks in the city due to inadequate resource planning [11].

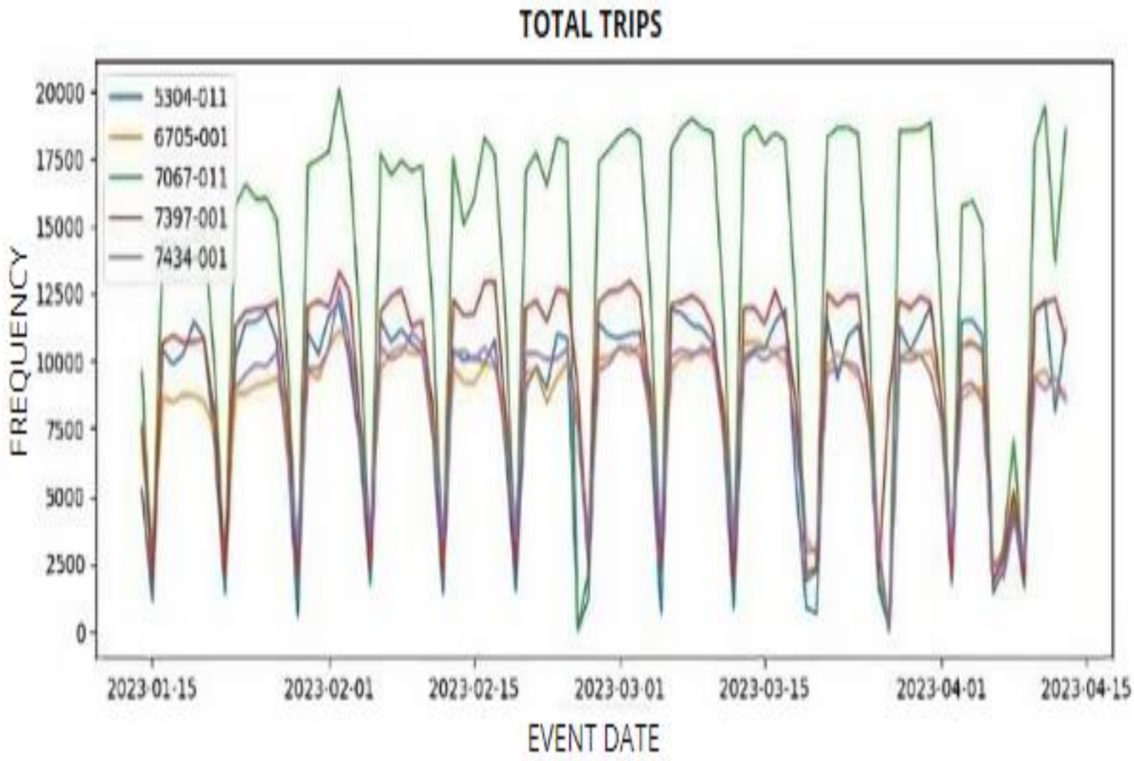

Figure 1 Total trips from February to April. Taken and edited from [11].

On the other hand, research was conducted that compares different machine learning models to address "choice of modeling travel mode" problem, which examines how individuals choose their mode of travel. This problem relies on numerous variables, including individual, household, and exogenous factors such as safety and comfort during travel, weather conditions, and the built environment. Algorithms such as Artificial Neural Networks (ANN), Decision Trees (DT), Support Vector Machines (SVM), and Cluster Analysis (CA) were employed. The researcher concludes that the optimal algorithm for this problem is a variant of DT known as Random Forest (RF) [12].

Other researchers propose the use of a Reinforcement Learning (RL) algorithm for controlling a public transportation system. To substantiate this proposal, the researcher devises a simulation environment enabling the testing and evaluation of reinforcement learning algorithms in administrative decision-making concerning routes, buses, and stations within a city. This investigation culminates in the creation of *TranCity*, a simulator representing a complex system such as a city's public transportation system [13].

Research was also conducted that addresses the problem of route optimization with the help of genetic algorithms (GA). Using Python, they designed two types of agents that communicate until finding the best solution: the optimizing agents, called OptiAgent (OA), and the SolutionManager agent (SMA), responsible for exchanging information between the initial population and the optimization agents, implementing changes upon finding improvements [14], as depicted in Figure 2.

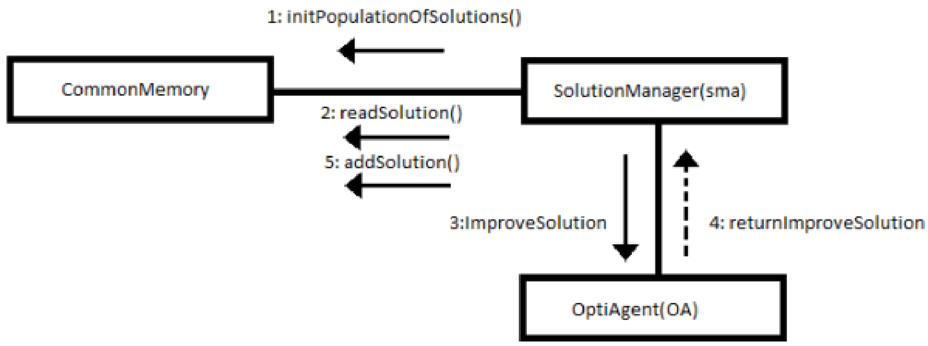

Figure 2 Communication diagram of the JABAT resolution process. Taken and adapted from [14].

Based on intelligent agents, the use of Java Agent Development (JADE)-based A-Team, named JABAT. JABAT employs a group of agents, each with its own algorithm to solve the problem, aiming to avoid a local optimum. Every solution is taken and enhanced by external and independent agents to increase the chances of reaching the global optimum [14,15]. This research concludes with a design of a multi-agent system addressing the vehicle routing problem with time windows, featuring a single depot (vehicle's starting point) and any number of customers, as depicted in Figure 3. The red circle denotes the depot, while the yellow circles represent the customers. The initial customers for each route are highlighted in blue, and the blue lines simulate the vehicle's route [15].

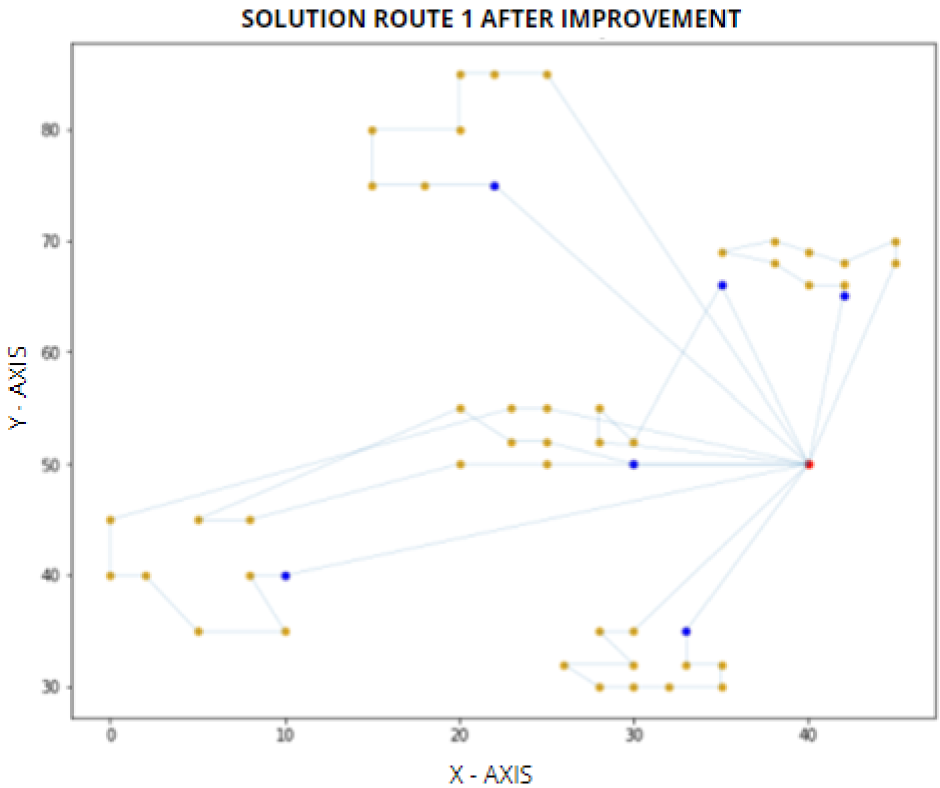

Figure 3 Graphical representation of the final solution for one of the routes. Taken from [15].

Similarly, the use of the Particle Swarm Optimization (PSO) algorithm is proposed, belonging to the GA family, to solve the Vehicle Routing Problem with Simultaneous Pickup and Delivery (VRPSPD). Initially, they decode the data used in their algorithm, creating a customer list with different priorities for each. Subsequently, they construct a matrix of vehicle priorities to then build their routes. The study concludes with a solution to the VRPSPD problem using PSO implemented in Matlab, which, compared to the Tabu

Search (TS) method—utilized for solving salesman travelling problems—shows promising results with a maximum deviation of only 13.043% with 200 clients. With 400 or 100 clients, the deviation remains under 9% [16].

In 2015, research was conducted to verify whether a linear regression model could estimate travel time in a mass transit system, utilizing real data from MEGABUS in the central-western metropolitan area of Pereira city, Colombia. In this investigation, they constructed two linear regression models. The first model utilized travel time variables against distance, while the second model utilized vehicle speed (obtained through a sensor) against travel times. Their conclusion was that employing classical linear regression models to estimate travel times is not feasible [17].

An analysis of public transport data in Madrid was carried out and an artificial intelligence model was proposed for predicting drop-off stops along each route. The researchers used Python and libraries such as scikit-learn, Keras, and TensorFlow to build two models, which they compare in their research. The first model employs the Random Forest machine learning technique, while the second model is a neural network with a single hidden layer of 48 neurons using ReLu activation, alongside an output layer with softmax activation. The research demonstrates and concludes that the Random Forest method is the superior choice for this problem, achieving an 87% accuracy compared to the neural network's 75% [18].

Other research demonstrates that artificial neural networks yield better mobility demand estimations compared to classical methods such as the direct method and gravitational methods. The classical methods were modeled using multiple linear regression in Matlab software. These models were evaluated based on the coefficient of correlation (R). The direct method, considered the best model, yielded modest results since, on average, in the Campo de Gibraltar and Área de Málaga (Spain), the study areas in this research, the average R value was 0.39. The neural network model was also developed in Matlab using the Neural Networks toolbox. The neural network provided mobility estimations very close to reality, obtaining an average R value of 0.81691 for the Campo de Gibraltar and 0.89669 for the Área of Málaga [19].

Similarly, the SmartBus system was developed, which applies Big Data and Data Science techniques to Madrid's urban transportation system. To achieve this, with the aid of data from the Municipal Transport Company of Madrid (EMT), they employed various machine learning techniques to segment bus lines. Models like SAX-K-Means did not yield satisfactory results as they grouped routes with highly diverse behaviors. Consequently, the researcher opted to use K-Means and Hierarchical Clustering techniques. They utilized K-Means to cluster all lines into 6 distinct groups, while employing Hierarchical Clustering, designed using the hierarchical clustering technique from the sciPy library, to validate the groupings made with K-Means. The research designs a Deep Learning model whose results are compared with outcomes derived from a statistical method called ARIMA (Auto Regressive Integrated Moving Average). These models underwent testing to predict route demand. For the Deep Learning model, they utilize an LSTM layer employing the Keras library. The model demonstrates satisfactory performance, achieving an RMSE of 271.13 during training and 280.2 in its testing phase. As depicted in Table 2, the model proposed by the researcher yielded superior results [20].

Table 2 Comparison of results of LSTM and ARIMA models

MODEL	Medium RMSE on Training	Medium RMSE on Test
ARIMA (No external data)	299.41	1216.20
ARIMA (With external data)	350.98	713.97
LSTM (No external data)	411.82	515.25

LSTM (With external data)	271.13	280.20

Taken and adapted from [20].

Another research employs machine learning methods in public transportation. The researchers compare five machine learning techniques to solve the problem of predicting bus arrival times. The technique proposed by the researchers is named Random Forest Based on Near Neighbors (RFNN), which involves a preselection of training data using the Near Neighbors technique to enhance the model's performance obtained with Random Forest by training it with the preselected data. They compare the performance of the RFNN model with models such as Linear Regression, K-nearest Neighbors, SVM, and the classic RF. The results demonstrated that the model proposed in the research is the most effective for solving the problem, achieving scores of 13.65 in Mean Absolute Error (MAE) and 26.37 for RMSE.

A model was designed to predict and estimate travel times between two points using latitude and longitude data. provided by GPS installed in vehicles. Using the Random Forest technique to train the model, they calculated the average speed of the vehicles to subsequently derive an algorithm for calculating historical travel times. The model predictions showcase a 96.88% accuracy validated using the Mean Absolute Percentage Error (MAPE). The researcher concludes by comparing their model with the Google Maps API, noting a difference of ±5 minutes, demonstrating the satisfactory performance of their model [22].

An illustrative example is the research conducted in 2017. The aim was to enhance the quality and safety of public transportation services in Greece. The study combined spatial clustering methods and neural network models to forecast high-risk crime areas in transportation. The developed models were evaluated using Mean Squared Error (MSE) and implemented with the Scaled Conjugate Gradient Algorithm (SCG) due to its speed. The research concludes with a model exhibiting a mean squared error of 0.00013833 at epoch 153 of training, capable of predicting transportation stations with a higher risk of crime occurrence [23].

Another notable investigation was carried out in 2019, addressing the application of AI in the transportation domain to address various challenges like the increasing travel demand, CO_2 emissions, environmental concerns, and safety issues. The findings from their analysis in the study demonstrate that AI, such as neural networks and genetic algorithms, can be effectively utilized to resolve challenges related to increasing travel demand, security, carbon dioxide emissions, and fuel wastage [24].

In 2020, the comparison was made through several neural networks trained for computer vision. Utilizing Yolo3, Yolo3Tiny, HOGGCV, AWS Rekognition, and Azure Cognitive Services, they predicted the number of people in an image and assessed their performance using the F1 Score of the predictions. The aim was to determine which model to implement for calculating demand in public transportation. The researchers concluded by noting that the F1 score could occasionally tie across different algorithms, rendering it an ineffective measure to determine the best model. They also found that AWS Rekognition achieved a high percentage in detections; however, it wasn't defined as the best among the five due to the tie observed in the comparisons [25].

A similar investigation, proposed the development and integration of an artificial vision system, incorporating a person recognition algorithm capable of detecting the number of passengers boarding public transportation vehicles, thus determining the number of actual trips. Through the implementation of this system, they aim to enhance the accuracy of passenger counts, currently conducted using bars equipped with infrared sensors, by employing the YOLO model and a Deep Simple Online Real-time Tracker. They conclude with a YOLOv4 model achieving an average accuracy of 46.1% [26].

In 2021, research was carried out with aiming to enhance the Public Transportation System operating in Metropolitan Lima, Peru, through route optimization. The researcher utilized Big Data and Machine Learning sciences in developing an algorithm for generating optimized routes. Preprocessing large volumes of data from electronic card reader devices used in service stations and user origin-destination data with Big Data techniques. The research employed the Machine Learning capabilities within the Weka software, used in the algorithm's creation. This involved attribute selection and model testing for predicting the hours a card is used at a specific stop. Through comparison of techniques, it was discovered that the model with the highest performance, measured by the Accuracy metric, was the BayesNet, achieving 51.76% accuracy over intervals of 10 hours and 22 days of data. The research concludes with an algorithm reducing 51 trips in a simulation using the mentioned system's database, reducing waiting time by 66.6%, transfer time by 83.44%, and the number of transfers by 69.06% [27].

The implementation of Big Data in the Integrated Public Transportation System of Fortaleza, Brazil was addressed (SIT-FOR) to identify common usage patterns. Data was processed using K-means and supervised models were applied, with Random Forest demonstrating notable performance. The outcomes support the research's hypothesis, affirming that "understanding different usage patterns can enhance the identification of travel attributes in ticket emission data" [28].

In 2019, research was carried out in which the objective is to improve the quality of public transport service for citizens by creating a predictive model that classifies citizen queries about public transport. This work proposes the development of a predictive mathematical model serving as a foundation for a chatbot program, offering uninterrupted service via instant messaging and optimizing query times, thereby reducing costs for the company. The model was developed in Python using the SpaCy library and utilized data from inquiries sent to the Municipal Transportation Company in Valencia, Spain. The resulting model achieved an accuracy of 95.22%, an excellent outcome considering the seasonality and sample size collected [29].

In 2020, the Supqua mobile application is developed. The application, designed for visually impaired individuals in Bogotá, utilizes Machine Learning and open data from the SITP to provide route information at zonal stops without requiring an internet connection. It implements Transfer Learning in TensorFlow using the MobilNet model for object detection in signs and employs Tesseract OCR for optical character recognition. The application queries Bogotá's open data, storing it locally, and presents the results through an interface with audible feedback for users [30].

Another investigation addresses the challenges faced by visually impaired individuals while taking buses, lacking tools that provide them with autonomy in public transportation. In their project, they developed an "Assistive Device for Public Transportation for Individuals with Severe Visual Disabilities (SVD)" during the year 2022. The focus is on providing autonomy to these individuals when using public transportation in the city of Bogotá, Colombia. The solution incorporates the Amazon Rekognition model and utilizes the App Inventor application to create a mobile app serving as a user-machine interface [31].

Another approach addresses the issue of low accessibility in the public transportation system of the city of Fortaleza, Brazil, using Big Data. Their research focuses on establishing and applying methods to characterize this problem by consolidating data, calculating indicators, and describing the current situation. Using R programming, the researchers consolidate validation and vehicle positioning data, compute accessibility, mobility, and public transportation operation indicators, and apply spatial analysis methods alongside the K-Nearest Neighbors (KNN) machine learning technique. The aim is to establish methodologies for analyzing the vast amount of data and contribute to directing interventions in Fortaleza's transportation system [32].

Another research work delving into data handling, focused on optimizing user flow in public transportation. In 2019, this study analyzed data acquired from a device developed in a previous investigation. A device installed in the center of buses capable of collecting information about the location of people's cell phones and correlating this data with the location of buses and nearby stops [33]. The research aimed to obtain the occupancy of a bus at each point of its route, verify its reliability and efficiency, and explore the possibility of predicting this occupancy behavior from the collected data. Their work concludes with the launch of a webpage capable of preprocessing the collected data. Additionally, they recommend applying the K-Means technique when there is enough data for constructing a machine learning model [34].

In 2023, research focused on understanding the spatio-temporal patterns of user behavior in a public transportation system was presented. The primary objective was to diagnose the temporal and spatial displacement patterns of users in the public transport network of Fortaleza (Brazil), utilizing pre-pandemic data, specifically from November 2018. To diagnose these patterns, they employed Machine Learning models with clustering techniques. The results revealed that the initial daily validations by users exhibit a higher percentage of spatial and spatiotemporal patterns compared to the final daily validations. Additionally, it was observed that it is rare for users to exhibit spatial and temporal patterns simultaneously, emphasizing the need for individual attention to both aspects [35].

In 2021, research was developed with the objective of creating a predictive model of failures through machine learning to improve availability on the buses of the company Turismo Nacional Imperial Cusco S.A.C. The consistent rise in costs to maintain bus availability compels companies to replace parts only when damage becomes evident, which can halt services in the event of failures. The applied experimental design analyzed 7 employee transport buses using the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology to develop the predictive model. Processing the machine learning model in Matlab revealed an accuracy of 82.8% with the decision tree algorithm. Implementing the predictive model resulted in a significant reduction in the Mean Time Between Failures (MTBF) indicator from 53.65 to 10.26 minutes and the Mean Time to Repair (MTTR) indicator from 786.81 to 11.79 minutes. The research concludes that the decision tree predictive model contributes to increasing the availability of buses within the Turismo Nacional Imperial Cusco S.A.C. company [36].

In 2023, work was presented focused on addressing one of the sources of delays in public transportation systems: the stations. Traditionally, data about people boarding and/or alighting from transportation are collected manually or through visually processed videos. This work employs computer vision with models trained from scratch, such as YOLO, EspiNetV2, and Faster R-CNN, and conducts a comparative analysis among these models. The study concludes with a YOLOv3 model capable of accurately counting individuals with 92% accuracy and the creation of a dataset containing 348 video sequences [37].

In 2022, a work resulting from the collaboration between the Ontological Engineering Group of the Universidad Politécnica de Madrid and the Regional Transport Consortium of Madrid. This study focuses on analyzing public transportation usage, particularly among the elderly population in the Madrid region. The work addresses two cases: firstly, it examines the effect of weather conditions on public transportation usage by elderly users, incorporating these meteorological variables into forecasting models and evaluating their impact on previous models. Additionally, new forecasting models are introduced to enhance existing ones. Secondly, the feasibility of clustering time series data of public transportation usage is evaluated to identify users with similar patterns and improve predictions of public transportation behavior for the entire group [38].

In 2021, a work is presented that highlights the advancements in the implementation of an automated vehicular traffic counting system based on videos in Bahía Blanca, Argentina. They employed YOLOv4 and DeepSORT algorithms across 85 surveillance cameras distributed throughout the urban area. Despite encountering certain technical challenges in the records, these were categorized for correction in both algorithms [39].

In a study, the use of the Internet of Things (IoT) tools and Computer Vision for real-time monitoring of mass transportation was proposed. Utilizing the OpenCV library, they employed a pre-trained Convolutional Neural Network (CNN) to count the number of people waiting for buses at stations. Through a prototype built with a Raspberry Pi 3 connected to a camera, it captured an image every five seconds for analysis by the CNN and sent the results via KafkaStreams to the database. It was concluded that this is a technologically viable way to obtain real-time public transportation information, although, according to the researchers, further experimentation in real environments and finer adjustments of neural network parameters were needed [40].

A 2021 study, address pedestrian fatalities, one of the leading causes of death worldwide, categorized as a pandemic by the World Health Organization (WHO) [41]. The research employs Machine Learning methodology aiming to predict whether a pedestrian involved in a traffic accident would end up dead or injured. The results show that the model, utilizing Oversampling for training data, adequately predicts injuries with 90% precision, accuracy, and reliability. However, it only manages to predict half of the death cases. Conclusions highlight that the model closest to achieving the proposed objective was a penalized model that generated synthetic samples of the minority class "deceased" using the RandomOverSampler method. The study remains ongoing in search of better performance in predicting death cases [42].

In an evaluation of transportation models carried out in 2020, information is collected from various sources related to mobility, using publicly accessible data such as the 2011 Population and Housing Census, the Atlas of Household Income Distribution, Cadastre, and Spatial Distribution of the Population in Andalusia, Spain. The study focuses on the city of Seville, dividing it into Transport Analysis Zones according to the Sustainable Urban Mobility Plan (PMUS) of January 2017. Main objectives include gathering information into an extensive database, creating modal split and travel distribution models, seeking additional utilities like defining transport zone centroids, and applying modal split models and Origin-Destination matrices calculation at the transport zone level using Machine Learning techniques. The investigation concludes by demonstrating that the Decision Trees model performed best in predicting the type of vehicle a person would use, achieving 42.6% Accuracy [43].

In 2017, research was conducted focused on machine learning for visual attention models to support video surveillance in public transportation. They utilized technologies that acquired task-based features, developing models known as top-down. The project aimed to analyze areas in images that draw visual attention in public transportation video surveillance. Classification trees were employed to achieve performance close to real-time implementations. In the conclusions, the researcher highlights the creation of the database and the top-down visual attention model implemented using the Random Forest technique. This model exhibits high performance in supporting video surveillance in public transportation [44].

An investigation carried out in 2019, delves into modeling the analysis of demand and service quality in regular road passenger public transportation through data mining. The thesis aimed to realign mobility towards a sustainable approach and promote the use of public transportation, resulting in an accumulation of three publications addressing the same topic [45]. In the first one, they utilized time-dependent attributes to represent public transportation demand, achieving a *silhouette_score* of less than 0.1 with eight groups [46]. The second publication describes a system for continuous and comprehensive control of punctuality and regularity for a mass transit system using GPS-provided data, implemented and tested at the Gran Canaria Transport Authority in Spain. This article employed the K-Means Classifier (KMC) technique with three groups to analyze bus stops [47]. In their third publication, they again utilized the KMC technique, and the data obtained from their previous publication to analyze travel times within the public transportation service, using a model with three clusters that scored 0.35 on the silhouette function. This publication concludes by identifying patterns present in different travel times and where these times are longer [48].

In 2020, the perception of safety in public transportation in the State of Mexico concerning the COVID-19 contingency using data mining. They analyzed 1788 data points from the mobility survey and health measures during the COVID-19 contingency conducted by the Autonomous University of the State of Morelos (UAEM) in 2020. Evaluated variables included access safety, waiting safety, on-board safety, and overall safety. They employed Weka's J48 and a priori algorithms for classification. The results indicate that the perception of safety in using public transportation remains constant, with the apriori algorithm revealing that 88% of users perceived insecurity during their journey. It's important to note that the algorithms used operate based on the Decision Trees technique [49].

In a recent investigation, a computer vision method was proposed to estimate the number of people at TransMilenio stations in Bogotá. They employ deep learning on the TransMilenio-Javeriana database, which provides nearly 900,000 head labels on buses and stations. They achieve an 11% relative error accuracy, surpassing manual methods, and offer a scalable, low-cost solution to enhance public transportation planning and operations [50].

Another research presents a project that utilizes machine learning techniques and onboard monitoring in vehicles equipped with On Board Diagnostics II (OBD II) to classify driver behavior under real road conditions. They conducted a monitoring campaign involving 5 vehicles, capturing data through an OBD II device and a smartphone application named Torque Pro. Using the K-Means technique, they categorized drivers into 3 groups: normal, aggressive, and hazardous. The results indicate that hazardous drivers, when driving at high speeds, exhibit increased fuel consumption and pose a risk of accidents on the road [51].

In research carried out in 2023, the advantages of using Big Data to improve mobility in the transportation sector are described. This work showcases mobility improvements through this branch of artificial intelligence, using a case study involving the implementation of an Intelligent Transportation System (ITS) in the city of Bogotá, Colombia [52]. The Bogotá Mobility Secretary, in collaboration with the technology company SKG, implemented the ITS, resulting in a 12% increase in the city's average speed and a 42% reduction in response time to traffic incidents. Furthermore, it allowed the identification of key issues impacting mobility within the city, facilitating informed decision-making [53].

Likewise, a prototype of an advanced driver assistance system was developed using computer vision in public transportation vehicles to prevent traffic accidents. This prototype is positioned at the front center of the vehicle and employs a Raspberry Pi camera to detect obstacles such as pedestrians and lane markings using the Haar Cascades computer vision technique. The system issues visual and auditory alerts in case of infractions, sending detailed messages to the database. They conclude with a functional prototype that successfully detects 93% of the cases in the conducted tests [54].

To conclude, research addressed the prediction of psychosocial occupational risk in urban transportation, specifically among urban bus drivers. The study highlighted the exposure to psychosocial occupational risks in this stressful occupation. Machine learning techniques such as K-Nearest Neighbors, Decision Tree algorithms, and Support Vector Machines were employed. Classification was performed using psychosocial risk data collected through the Fpsico 4.0 questionnaire from urban transport drivers in the city of Ambato, Ecuador. Experimental results indicated that the Support Vector Machine model achieved an 88.53% performance in predicting psychosocial occupational risks [55].

3. Conclusions

This article compiles various research endeavors that apply artificial intelligence techniques in the field of public transportation to address characteristic problems such as route optimization, service demand estimation, arrival time predictions, among others. It can be inferred that the techniques yielding better results in route optimization tasks were Random Forest and Neural Networks as they better adapted to the data.

Artificial intelligence can become a significant ally in enhancing the quality of public transportation services because, with the aid of computer vision techniques, we can create a service demand control system to identify the most used vehicles or routes, enabling better administrative decision-making.

The Random Forest technique proved to be the most used in problems such as arrival time prediction, stop prediction, and predicting people's travel modes, demonstrating its high capability to accommodate vehicle trajectory and time data, in addition to analyzing the behavior of service users.

Models obtained using linear regression-based techniques are unsuitable for issues encountered in public transportation services because they fail to adapt to the diverse factors influencing routes and travel times, such as city traffic congestion, unforeseen stops, delays in user pickups and drop-offs.

The use of neural networks with LSTM layers is common in solving arrival time estimation and vehicle demand problems, showing good results compared to classical techniques used for such estimations. These neural networks are known for their long-term memory, allowing them to have more information when predicting their target variable, hence, they are widely used in environments facing Time Series Forecasting (TSF) problems, where predicting a future event requires analyzing the past.

The quality and relevance of artificial intelligence models applied in public transportation fundamentally relies on the quality of the datasets used. The pertinence and effectiveness of these models largely hinge on the breadth, precision, and timeliness of the gathered information. The fusion of valuable data with sophisticated artificial intelligence algorithms represents a promising advancement in the evolution of more efficient and adaptive public transportation systems tailored to the dynamic needs of urban communities.

The transformative potential of IoT, artificial intelligence, artificial vision systems, and Information and Communication Technologies (ICT) in the domain of public transportation is undeniable. The synergy among these technologies facilitates continuous monitoring of vehicles and passengers, captures valuable data for real-time decision-making, and transmits information instantly among system components, crucial for enhancing operational efficiency and user experience. Leveraging the benefits offered by technological advancements paves the way to revolutionize the public transportation industry, offering smarter, safer, and more adaptive services, thus paving the way toward a more efficient and sustainable urban mobility.

Based on the research analyzed in this article, we propose as a future line of work the comparison of performance between models obtained using neural network-based techniques like LSTM, contrasted with classical techniques such as Random Forest, in a common task of generating optimal routes for public transportation. Both techniques demonstrated great adaptability in problems requiring geographic coordinate data, latitude and longitude, and time series.

4. Declaration of competing interest

We declare that we have no significant competing interests including financial or non-financial, professional, or personal interests interfering with the full and objective presentation of the work described in this manuscript.

5. Funding

This article was sponsored by the Research Fund of the University Santo Tomás Sectional Bucaramanga.

6. Author contributions

Carlos D. Pimentel-Díaz, professor at the Mechatronics Engineering Faculty, directed the research, wrote and corrected the article. Flor E. Romero-Rueda, professor at the Industrial Engineering Faculty, conducted the state of the art and co-wrote the paper, Eduardo J. Duarte-Peña student at the Mechatronics Engineering Faculty, co-authored.

References

[1] Banco Mundial. “Transporte: Panorama general”. Accessed September 28, 2023. [Online]. Available in: https://www.bancomundial.org/es/topic/transport/overview

[2] A. Bull. “Congestión de Tránsito, el problema y cómo enfrentarlo”. Santiago de Chile: CEPAL. 2003, pp. 21. Available in: https://repositorio.cepal.org/server/api/core/bitstreams/baac8944-5176-4792-9b19-b49ea4062290/content

[3] X. González. “Cuatro aplicaciones que le permiten transportarse según su necesidad”. La República. Accessed September 28, 2023. [Online]. Available in: https://www.larepublica.co/especiales/especial-transporte-moderno-noviembre-2019/cuatro-aplicaciones-que-le-permiten-transportarse-segun-su-necesidad-2938650

[4] Haenlein, M., & Kaplan, A. “A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence”. California Management 2019. Available in: https://doi.org/10.1177/0008125619864925

[5] Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., ... & Galanos, V. “Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda

for research, practice and policy". International Journal of Information Management, 2019. Available in: https://doi.org/10.1016/j.ijinfomgt.2019.08.002
[6] Chui, M. "Artificial intelligence the next digital frontier?". McKinsey and Company Global Institute, Shanghai: China. 2017. Available in: https://bit.ly/33TVo6N.
[7] S. Fuentes. "Aplicación de multiagentes en Python para optimización de rutas". 2021. Available in: https://biblus.us.es/bibing/proyectos/abreproy/93448/fichero/TFG-3448+FUENTES+MORENO%2C+SERGIO.pdf
[8] L. Abarca. "Cómo ahorrar en gasolina con la optimización de rutas: la planificación al servicio de la eficiencia". 2022. Available in: https://www.praxedo.es/blog/como-ahorrar-en-gasolina-con-la-optimizacion-de-rutas-la-planificacion-al-servicio-de-la-eficiencia/
[9] A. Aguado & J. Jiménez, "Optimización de rutas de transporte", 2013. Available in: https://hdl.handle.net/20.500.14352/36218
[10] D. Jorquera, "Desarrollo de un prototipo de modelo predictor de tiempos de llegada de buses para una aplicación que brinda servicios de información sobre el transporte público", 2020. Available in: https://repositorio.uchile.cl/handle/2250/178520
[11] E. Benito & C. Garzón, "Propuesta de diseño de una herramienta tecnológica que ayude en la programación de recursos en el sitp de Bogotá para el concesionario consorcio express s.a. Sistematización. Corporación Universitaria Minuto de Dios-Soacha Colombia", 2023. Available in: https://repository.uniminuto.edu/handle/10656/17413
[12] J. Pineda, "A review of Machine Learning (ML) algorithms used for modeling travel mode choice". 2019. Available in: http://www.scielo.org.co/scielo.php?pid=S0012-73532019000400032
[13] M. Salcedo, "Algoritmo de aprendizaje por refuerzo para el control de un sistema de transporte público", 2020. Available in: http://hdl.handle.net/1992/51469
[14] S. Fuentes. "Aplicación de multiagentes en Python para optimización de rutas". 2021. Available in: https://biblus.us.es/bibing/proyectos/abreproy/93448/fichero/TFG-3448+FUENTES+MORENO%2C+SERGIO.pdf
[15] D. Barbucha, I. Czarnowski, P. Jedrzejowicz, E. Ratajczak y I. Wierzbowska, "JADE-based A-Team as a tool for implementig population-based algorithms", vol. 3, 2006. Available in: https://ieeexplore.ieee.org/document/4021874
[16] H. Lamos Diaz, S. Galvan Nuñez, L. Gonzalez Villamizar, C. Cruz Jimenez. "Algoritmo PSO-Híbrido para solucionar el problema de ruteo de vehículos con entrega y recolección simultáneas". 2013. Recovered from: http://www.scielo.org.co/pdf/rfing/v22n35/v22n35a07.pdf
[17] O.A. Sabogal Cardona, J.D. Hincapié Zea, J.J. Santa Chávez, J.W. Escobar. "Modelos de regresión lineal para estimación de tiempos de viaje en sistemas de transporte masivo". 2015. Available in: http://www.scielo.org.co/pdf/cein/v25n1/v25n1a06.pdf
[18] G. Díaz Ferreiro. "Análisis de los datos del transporte público de Madrid y deducción de la parada de bajada para cada trayecto". 2022. Available in: https://ruc.udc.es/dspace/bitstream/handle/2183/31977/DiazFerreiro_Gema_TFG_2022.pdf?sequence=3
[19] P. Rodríguez, I. Turias. "Una comparativa entre redes neuronales artificiales y métodos clásicos para la predicción de la movilidad entre zonas de transporte. Aplicación práctica en el Campo de Gibraltar, España". 2016. Available in: http://www.scielo.org.co/pdf/dyna/v84n200/0012-7353-dyna-84-200-00209.pdf
[20] C. Rosado. "SmartBus: Big Data y Data Science en Transporte Urbano". 2017. Available in: https://repositorio.uam.es/bitstream/handle/10486/681129/Rosado_Moral_Carlos_tfm.pdf
[21] B. Yu, H. Wang. W. Shan, B. Yao. "Prediction of Bus Travel Time Using Random Forests Based on Near Neighbors". 2017. Available in: https://doi.org/10.1111/mice.12315
[22] J. Morales. "Modelo de predicción y estimación de tiempos de traslado entre dos puntos utilizando datos de GPS". 2020. Available in: https://repositorio.uchile.cl/bitstream/handle/2250/177134/Modelo-de-predicción-y-estimación-de-tiempos-de-traslado-entre-dos-puntos-utilizando-datos-de-GPS.pdf

[23] G. Kouziokas, “The application of artificial intelligence in public administration for forecasting high crime risk transportation areas in urban environment”. 2017. Available in: https://doi.org/10.1016/j.trpro.2017.05.083
[24] R. Abduljabbar, H. Dia, S. Liyanage & S. Bagloee, “Applications of artificial intelligence in transport: An overview”. Sustainability, 2019. Available in: https://doi.org/10.3390/su11010189.
[25] R. Contreras, A. Ochoa & V. Torres, “Avances en la medición de la demanda en el transporte público con inteligencia artificial”. 2020. Available in: http://econferencias.uacj.mx/ocs/public/conferences/26/2020/Ciudadessustentables/RobertoContreras.pdf
[26] E. Santiago & F. Gonzalez, “Propuesta de herramienta de reconocimiento de personas mediante técnicas de inteligencia artificial para conteo de abordaje al transporte público”. 2023. Available in: http://ciateq.repositorioinstitucional.mx/jspui/handle/1020/630
[27] B. Tapia, “Generación de rutas óptimas en sistemas de transporte público masivo usando ciencia de datos con Big Data y Machine Learning” 2021. Available in: https://repositorio.uni.edu.pe/handle/20.500.14076/22099
[28] K. Mesquita, “Método de identificação dos padrões de uso e locais de embarque a partir do Big Data de transporte público: uma abordagem baseada em Machine Learning”. 2023. Available in: http://repositorio.ufc.br/handle/riufc/74404
[29] E. Gallardo, “Creación de un modelo predictivo para clasificar las consultas ciudadanas sobre el transporte público”. 2019. Available in: http://hdl.handle.net/10251/128834
[30] A. Velosa, “Supqua: aplicación móvil para interpretar datos de las señales del SITP zonal para la población en condición de discapacidad visual usando machine learning”. 2020. Available in: http://hdl.handle.net/11349/24820
[31] P. Galindo, “Dispositivo de asistencia en el transporte público para personas con DVS”. 2022. Available in: http://hdl.handle.net/1992/64450
[32] C. Vieira & C. Grangeiro, “Caracterizando a problemática da baixa acessibilidade do transporte público de fortaleza com o uso de big data”. Available in: http://www.anpet.org.br/anais/documentos/2018/Planejamento%20Territorial%20do%20Transporte/Transporte%20e%20Uso%20do%20Solo/5_241_RT.pdf
[33] G. Pérez, “Optimización del tráfico de usuario en el transporte público”. 2019. Available in: http://hdl.handle.net/2117/167309
[34] T. Torres, “Elaboración de un sistema alternativo no invasivo que permita optimizar las líneas de transporte público”. 2019. Available in: http://hdl.handle.net/2117/171876
[35] K. Mesquita & L. Barros, “Análise dos Padrões Espaço-Temporais de Validação em Sistemas de Transporte Público: Uma Abordagem Baseada em Mineração de Dados”. 2023. Available in: https://doi.org/10.12957/cadinf.2023.73912
[36] R. Ruiz, “Machine learning en matlab para mejorar la disponibilidad en los buses de la empresa Turismo Nacional Imperial Cusco SAC” 2021. Available in: https://hdl.handle.net/20.500.12692/86535
[37] S. Velastin, & R. Fernández, “Conteo de personas que suben y bajan del transporte público mediante visión artificial” 2023. Available in: https://estudiosdetransporte.org/sochitran/article/view/289
[38] G. González, “Análisis y predicción individual del comportamiento de los usuarios con abono tercera edad en el transporte público de la Comunidad de Madrid”. 2022. Available in: https://oa.upm.es/71393/
[39] M. De Meio & V. Viego, “Conteo automatizado de tráfico en Bahía Blanca basado en videos”. 2021. Available in: http://sedici.unlp.edu.ar/handle/10915/140138
[40] L. Forero, C. Martinez, D. Ramirez & J. Zambrano, “IoTransit: Usando IoT y Visión Artificial para monitoreo en tiempo real de medios de transporte masivo”. 2019. Available in: http://dx.doi.org/10.13140/RG.2.2.33510.24648

[41] Naciones Unidas. (2023, Nov 20). “Accidentes viales: “Una epidemia silenciosa y ambulante” que mata a 1,3 millones de personas por año”. Accessed November 20, 2023. [Online]. Available in: https://news.un.org/es/story/2022/06/1511112
[42] M. Martínez, L. Vásquez & J. Gonzalez, “Análisis de Siniestralidad Peatonal A partir de la metodología de Machine Learning”. 2021. Available in: http://hdl.handle.net/11371/4978
[43] J. Herrera, “Evaluación de Modelos de Transporte mediante datos abiertos y técnicas de Aprendizaje Automático”. 2020. Available in: https://idus.us.es/handle/11441/107094
[44] S. Parejo, “Aprendizaje automático de modelos de atención visual como apoyo a la videovigilancia en el transporte público”. 2017. Available in: http://hdl.handle.net/10016/27649
[45] T. Cristóbal, “Modelización del análisis de la demanda y de la calidad de servicio en el transporte público regular de viajeros por carretera mediante minería de datos”. 2019. Available in: http://hdl.handle.net/10553/77295
[46] T. Cristóbal, G. Padrón, J. Lorenzo, A. Quesada & C. Gracía, “Applying Time-Dependent Attributes to Represent Demand in Road Mass Transit Systems”. 2018. Available in: https://www.mdpi.com/1099-4300/20/2/133
[47] G. Padrón, T. Cristobal, F. Alayon, A. Quesada & C. Garcia, “System Proposal for Mass Transit Service Quality Control Based on GPS Data” 2017. Available in: https://www.mdpi.com/1424-8220/17/6/1412
[48] T. Cristobal, G. Padrón, A. Quesada, F. Alayon & C. Garcia, “Systematic Approach to Analyze Travel Time in Road-Based Mass Transit Systems Based on Data Mining”. 2019. Available in: https://doi.org/10.1109/ACCESS.2018.2837498
[49] E. Noriega, J. Romero, D. Gutiérrez, & R. Rico, “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”. 2023. Available in: http://revistaaristas.tij.uabc.mx/index.php/revista_aristas/article/view/291
[50] W. Moreno, C. Burgos, D. Jaramillo & H. Carrillo, “Passenger Counting in Mass Public Transport Systems using Computer Vision and Deep Learning”. 2023. Available in: https://latamt.ieeer9.org/index.php/transactions/article/view/7331
[51] J. Aguilar, “Clasificación del comportamiento del conductor mediante técnicas de aprendizaje automático y monitoreo a bordo con OBD ll en condiciones reales de carretera”. 2023. Available in: https://repository.unab.edu.co/handle/20.500.12749/22585
[52] A. Rojas, “Uso de big data para mejorar la movilidad”. 2023. Available in: https://repository.unimilitar.edu.co/handle/10654/44941
[53] SKG Tecnología. “Secretaría Distrital de Movilidad Bogotá, Sistema Inteligente de Transporte (SIT) de Bogotá”. Accessed November 19, 2023. [Online]. Available in: https://skgtecnologia.com/secretaria-distrital-de-movilidad-bogota-sit/
[54] F. Salazar & G. Pico, “Sistema avanzado de asistencia al conductor empleando visión artificial en vehículos de transporte público”. 2019. Available in: https://repositorio.uta.edu.ec/jspui/handle/123456789/29951
[55] A. Lara, “Predicción del riesgo psicosocial laboral en el transporte urbano aplicando técnicas de aprendizaje automático”. 2020. Available in: https://www.proquest.com/openview/bbad65ac2af742e1958032097225379d/1?pq-origsite=gscholar&cbl=1006393