Evaluating AI-Powered Predictive Analytics for Public Transport Demand in Dubai

https://doi.org/10.63962/QEHM5235

Sara Ismail Dhalam
Rochester Institute of Technology, Dubai
Dubai, UAE
Sara-dhalam@outlook.com
ORCID: 0009-0006-2039-2269

Ayman Ibrahim
Rochester Institute of Technology Dubai
Dubai, UAE
ayman.ibrahim@rit.edu
ORCID: 0009-0007-5477-5150

***Abstract*—Rapid urbanization has made intelligent transportation systems crucial for facing congestion, improving safety, and enhancing efficiency. This study presents a preliminary evaluation of AI-powered predictive analytics applied to open datasets from Dubai's Roads and Transport Authority (RTA), focusing on multimodal transport modes including metro, tram, marine, and buses. Using SAS Viya's machine learning tools for data analysis was conducted to explore the potential of interpretable AI models for predicting public transport demand and optimizing traffic flow. As this work remains in its early stages, future research will involve full model deployment and testing. This research supports Dubai's vision for a more intelligent, sustainable transport network.**

Keywords—Intelligent transportation systems (ITS), AI-driven predictive analytics, public transport demand, reduce congestion, optimize traffic flow.

I. Introduction

Traffic congestion, driven by rapid population growth, economic development, and rising vehicle ownership, remains a significant challenge in cities worldwide. In Dubai, these elements have exerted considerable pressure on the transportation system, even with substantial investments in public transport. Issues such as traffic signal inefficiencies and reliance on private vehicles lead to increased delays and pollution. To address these challenges, Dubai has implemented Intelligent Transportation Systems (ITS) and developed demand management programs as essential approaches to reduce congestion and encourage public transport usage[1]. Studies estimate that traffic congestion results in yearly losses of around 1.5 billion AED due to wasted fuel and decreased productivity[2]. Additionally, research indicates that potential cumulative losses could reach $1.8 billion if proactive infrastructure investments had not been made [3].

This study aims to identify the benefits and limitations of AI-driven analytics in enhancing the efficiency and reliability of the public transport system. It evaluates AI predictive analytics for RTA open datasets, focusing on public transport demand across various modes. Using the SAS Viya and SAS optimization platform, we explore models to optimize resource allocation[4]. Future research will involve model deployment and evaluations to support Dubai's sustainable transport goals.

II. Literature Review

The extensive literature on intelligent traffic management systems focuses on technologies that improve traffic flow, reduce congestion, and enhance safety. This section will discuss four main topics: real-time traffic management, deep learning and density prediction, machine learning models for autonomous vehicles, and Optimization in Public Transport Systems.

A. Real-Time Smart Traffic Management

Recent studies have introduced IoT-based traffic management systems that use real-time sensors and video data to adjust signals dynamically during congestion. One approach employs VANETs and mobile agents to identify congestion hotspots [5], while another leverages IoT-cloud models for lane-specific signal control [6]. Deep learning methods like YOLO have also been integrated for real-time video analytics to detect abnormal events and enhance urban safety [7]. Additionally, Intelligent Transportation Systems (ITS) have shown effectiveness in improving traffic flow, as demonstrated using Variable Speed Limits (VSL) on major UAE roads, which resulted in reduced delays, travel times, and stops [1].

B. Deep Learning Models and Density Prediction

One notable study introduced DeepCrowd, a deep convolutional LSTM model designed to predict crowd density across fine-grained spatial grids, demonstrating strong performance in spatio-temporal forecasting tasks [6]. Similarly, Spatio-Temporal Convolutional Neural Networks (ST-CNN) were developed to capture complex urban mobility patterns by modeling spatial closeness, temporal periodicity, and long-term trends in mobility data [9]. Furthermore, another study employed Self-Supervised Contrastive Learning techniques—commonly used in fields such as computer vision (CV) and natural language processing (NLP) to extract meaningful traffic features without relying on labeled data, thereby improving the scalability and generalizability of traffic prediction models [10].

C. *Machine Learning Models in Autonomous Vehicle Adoption*

Recent studies have demonstrated the integration of AI, IoT, and autonomous vehicle technologies to enhance urban delivery systems and optimize traffic signal control through real-time data analytics, contributing to reduced carbon emissions and improved urban mobility [11]. In autonomous vehicle adoption, machine learning models such as Naïve Bayes, Random Forest, and Fuzzy Logic were employed to predict user acceptance. Fuzzy Logic outperformed the others, particularly in modeling adoption factors such as safety, trust, security, cost, ethics, and privacy [12]. Random Forest also showed strong performance, highlighting the effectiveness of ensemble methods in both behavioral modeling and traffic forecasting.

A Dubai-based study applied multiple ML models—YOLO, Random Forest, XGBoost, and Graph Neural Networks (GNN)—to improve traffic prediction and congestion classification [7]. Random Forest was particularly effective in forecasting congestion, while Support Vector Machines (SVM) and k-Nearest Neighbors (k-NN) achieved high classification accuracies of 99.9% and 98.6%, respectively. Long Short-Term Memory (LSTM) networks further enhanced the identification of recurring and non-recurring congestion patterns, reaching an accuracy of 98.73% [7].

D. *Optimization in Public Transport Systems*

While traffic prediction has advanced, public transport in urban mobility demands systemic optimization. Here, evolutionary algorithms play an essential role by introducing MOPEA (Multi-objective Evolutionary Algorithm), balancing conflicting goals like minimizing travel time and maximizing operator profits [5]. Additionally, AHP and entropy combined with TOPSIS models have been applied to rank public transport network optimization strategies based on multiple objective functions [6]. Moreover, one study designed an ACO-based approach to adjust existing bus lines without re-planning entire networks, providing a realistic solution for large, congested cities [7].

1. Research Gap:

Despite the growing interest in using artificial intelligence (AI) to improve urban mobility, there remains a significant gap in transparent, interpretable, and deployable AI models for predicting and forecasting public transport demand, particularly for metro and tram systems in complex cities like Dubai. Many existing studies focus mainly on road traffic or bus systems, overlooking rail-based transport and the diverse travel behaviors it involves. Additionally, most AI models used in research rely on complex algorithms, including deep learning and ensemble methods, that are difficult to interpret and require high computational resources. Finally, the lack of transparency and explainability makes them hard to trust for decision-makers, limiting their usefulness in policy, operations, and large-scale implementation.

2. Contribution:

This study proposes a novel predictive framework for estimating demand across multimodal transportation networks in Dubai, leveraging advanced machine learning techniques deployed on the SAS Viya, an advanced analytics platform. It helps cities prioritize investments, such as expanding the metro or tram lines, based on demand predictions [4]. The study aims to identify the benefits and limitations of AI-driven analytics in enhancing the efficiency and reliability of the public transport system. Expected outcomes include optimized traffic flow, reduced congestion, and improved transport efficiency, contributing to a more sustainable and user-friendly system for residents and visitors. Furthermore, utilizing more straightforward machine learning algorithms for data analysis might produce dependable results without requiring costly tools and specialized staff. While numerous studies have focused on bus optimization and road-related problems, this research will examine the metro and tram networks.

III. METHODOLOGY

This section presents the theoretical framework underpinning the methods employed in this project, including the process flowchart, model development stages, research design tools, and data sources used to evaluate the effectiveness of AI-powered predictive analytics for forecasting multimodal transport demand in Dubai.

The research follows a four-phase methodology, as shown in Fig. 1:

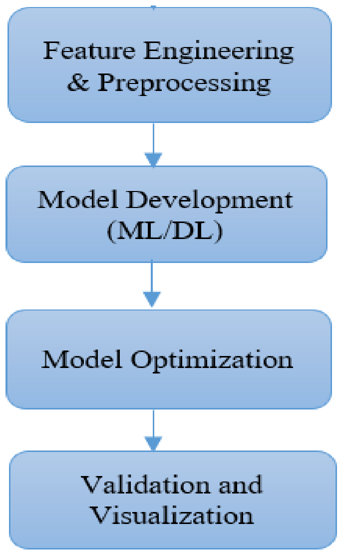

Fig. 1. Methodology Phases

A. *Data preparation and feature engineering*

Data preparation represents a foundational phase in this study, ensuring high-quality, model-ready inputs using SAS Data Preparation within the SAS Viya platform [4]. This process involved acquiring publicly available datasets from the Dubai Roads and Transport Authority (RTA) via the Dubai Pulse Open Data Platform [13]. The datasets encompass multimodal transport records, including metro, tram, marine, and bus service usage data. These data are regularly updated and include critical features such as route names, transaction types, timestamps, and passenger counts. Retrieval is performed by navigating to the "Transport" category on the platform, selecting the desired transport mode (e.g., Metro Ridership, Tram Usage), and downloading the data in CSV or API format.

B. *Demand Prediction Modeling*

Demand prediction is the second phase of the process, focusing on forecasting the most popular transportation mode using machine learning algorithms, as shown in Table 1.

TABLE I: Machine learning algorithms

Algorithm	**Applications and justifications**
Gradient Boosting	Machine learning models, including boosting, were used for predictive analytics and route optimization. Minimizing loss functions can effectively forecast baseline demand [8].
Random Forest	The Random Forest classifier achieved the highest accuracy of 84% in traffic congestion classification due to their robustness against overfitting, ability to handle high-dimensional data, and clear interpretability [9].
SVM	SVM showed moderate accuracy (64–79%) in classifying traffic congestion, making it useful for traffic state modeling, particularly in high-dimensional spaces and with small to medium datasets [9].
clustering (k-means)	Clustering methods, including K-means, were discussed for identifying High-demand hotspots and optimizing traffic flows [10]
Forecasting	Addressed overfitting using contrastive learning, as it outperformed ARIMA, Ridge Regression, SVR, and even Gradient Boosting (GBRT) [11]. Captures spatiotemporal dependencies, attention, and outperformed multiple state-of-the-art methods on large citywide datasets [12].

C. *Optimization*

The third phase is dedicated to optimizing and enhancing the results using SAS optimization techniques, which can be used for scenario-based decision making, what-if analysis, and resource allocation under uncertainty [13].

D. *Validation*

The final phase involves validating the model's results through model comparison and assessment using SAS Visual Analytics [4]. This phase includes evaluating model performance using key validation metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and F1-score. Additionally, classification performance is assessed through the confusion matrix, misclassification rate, and user experience (UX) metrics. These analyses ensure that the selected model meets the desired accuracy and reliability standards for deployment.

IV. PRELIMINARY RESULTS

A preliminary analysis was conducted to visualize the trends in multimodal public transport demand in Dubai over the past five years. Fig.2 presents a line chart illustrating these trends, revealing a general increase in demand across most transport modes, excluding a dip in 2022, primarily attributed to the impact of the COVID-19 pandemic.

The Dubai Metro exhibits the highest passenger demand, increasing ridership from approximately 203 million in 2019 to over 260 million in 2023. This upward trend underscores the city's firm reliance on metro infrastructure. Similarly, public transport buses have shown a robust recovery post-pandemic, with ridership growing from 95 million in 2020 to 173 million in 2023, reflecting their vital role in serving broader geographic areas.

Fig. 2. Line chart - Public transport by type

In contrast, the Metro Green Line shows significantly lower passenger volumes than anticipated. Moreover, despite having comparable infrastructure and connectivity potential, the Dubai Tram recorded the lowest usage levels, as depicted in Fig.3.

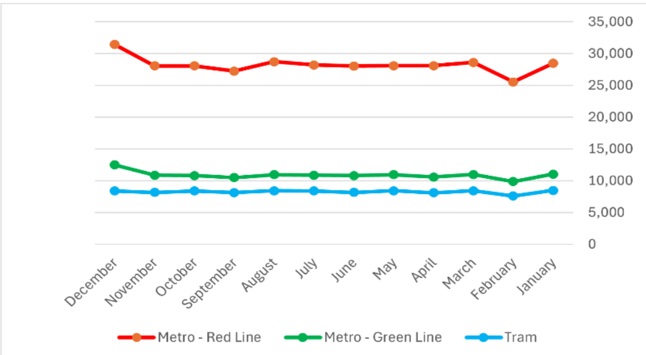

Fig. 3. Line chart – metro vs tram

V. CONCLUSION

This study showcases the potential of AI-driven predictive analytics to enhance traffic management and improve various transit modes in Dubai, such as metro, tram, marine, and bus services. The research demonstrates the benefits of interpretable AI models in urban transport planning by utilizing publicly available data from the Dubai Roads and Transport Authority (RTA) and SAS Viya's advanced machine learning capabilities.

Key findings reveal that AI-driven demand forecasting can significantly reduce traffic congestion by offering actionable insights for infrastructure development. The study emphasizes optimizing traffic flow and improving transport efficiency to create a more sustainable, user-friendly system for residents and visitors.

The next research phase will focus on applying these models to real-world datasets, including comparative testing and model enhancement, to adapt to Dubai's evolving urban mobility landscape. The goal is to support decision-makers in prioritizing funding and creating a more intelligent and sustainable transportation network.

REFERENCES

[1] Y. El-Hansali *et al.*, "Smart Dynamic Traffic Monitoring and Enforcement System," *Computers, Materials & Continua*, vol. 67, no. 3, pp. 2797–2806, 2021, doi: 10.32604/cmc.2021.014812.

[2] S. Shahbandari, "Traffic congestion costs more than Dh700,000 per kilometre in Dubai." Accessed: Apr. 25, 2025. [Online]. Available: https://gulfnews.com/uae/transport/traffic-congestion-costs-more-than-dh700000-per-kilometre-in-dubai-1.1452783

[3] Gulfnews, "RTA projects make Dubai traffic freer than leading cities with similar population." Accessed: Apr. 25, 2025. [Online]. Available: https://gulfnews.com/uae/transport/rta-projects-make-dubai-traffic-freer-than-leading-cities-with-similar-population-1.75275274

[4] "SAS Help Center: SAS Help Center: Welcome." Accessed: Apr. 29, 2025. [Online]. Available: https://documentation.sas.com/doc/ar/helpcenterwlcm/1.0/home.htm

[5] Hou Lin, Li Wen-yong, Ma Li, and Xu Jian-min, "Public transport network optimization based on a Multi-objective Optimization Problems Evolutionary Algorithm," in *2009 Chinese Control and Decision Conference*, Guilin, China: IEEE, Jun. 2009, pp. 4408–4412. doi: 10.1109/CCDC.2009.5192410.

[6] Li Xiaowei, "TOPSIS model for urban public transport network optimization based on AHP and entropy," in *Proceedings 2011 International Conference on Transportation, Mechanical, and Electrical Engineering (TMEE)*, ChangChun, China: IEEE, Dec. 2011, pp. 75–79. doi: 10.1109/TMEE.2011.6199151.

[7] W. Hu, C. Wang, and X. Zuo, "An Ant Colony Optimization based Approach to Adjust Public Transportation Network," in *2019 IEEE Congress on Evolutionary Computation (CEC)*, Wellington, New Zealand: IEEE, Jun. 2019, pp. 2575–2580. doi: 10.1109/CEC.2019.8790117.

[8] B. M. Mohsen, "AI-Driven Optimization of Urban Logistics in Smart Cities: Integrating Autonomous Vehicles and IoT for Efficient Delivery Systems," *Sustainability*, vol. 16, no. 24, p. 11265, Dec. 2024, doi: 10.3390/su162411265.

[9] D. Impedovo, F. Balducci, V. Dentamaro, and G. Pirlo, "Vehicular Traffic Congestion Classification by Visual Features and Deep Learning Approaches: A Comparison," *Sensors*, vol. 19, no. 23, p. 5213, Nov. 2019, doi: 10.3390/s19235213.

[10] R. E. Al Mamlook, M. Zahrawi, H. Gharaibeh, A. Nasayreh, and S. Shresth, "Smart Traffic Control System for Dubai: A Simulation Study Using YOLO Algorithms," in *2023 IEEE International Conference on Electro Information Technology (eIT)*, Romeoville, IL, USA: IEEE, May 2023, pp. 254–264. doi: 10.1109/eIT57321.2023.10187271.

[11] Y. Song, "Effective Traffic Prediction with Self-Supervised Contrastive Learning," in *2022 IEEE 8th International Conference on Computer and Communications (ICCC)*, Chengdu, China: IEEE, Dec. 2022, pp. 2204–2209. doi: 10.1109/ICCC56324.2022.10066048.

[12] R. Jiang *et al.*, "DeepCrowd: A Deep Model for Large-Scale Citywide Crowd Density and Flow Prediction," *IEEE Trans. Knowl. Data Eng.*, pp. 1–1, 2021, doi: 10.1109/TKDE.2021.3077056.

[13] "SAS Optimization." Accessed: May 05, 2025. [Online]. Available: https://www.sas.com/en_ae/software/optimization.html