Data and Metadata. 2025; 4:887
doi: 10.56294/dm2025887

ORIGINAL

Riding into the Future: Transforming Jordan's Public Transportation with Predictive Analytics and Real-Time Data

Hacia el futuro: transformación del transporte público de Jordania con análisis predictivos y datos en tiempo real

Anber Abraheem Shlash Mohammad[1], Suleiman Ibrahim Mohammad[2,3], Khaleel Ibrahim Al- Daoud[4], Badrea Al Oraini[5], Menahi Mosallam Alqahtani[6], Asokan Vasudevan[7], Mohammad Faleh Ahmmad Hunitie[8]

[1]Digital Marketing Department, Faculty of Administrative and Financial Sciences, Petra University, Jordan.
[2]Electronic Marketing and Social Media, Economic and Administrative Sciences Zarqa University, Jordan.
[3]Research follower, INTI International University, 71800 Negeri Sembilan, Malaysia.
[4]Department of Accounting– business school Faculties–Al Ahilya Amman University –Amman-Jordan.
[5]Department of Business Administration. Collage of Business and Economics, Qassim University, Qassim – Saudi Arabia.
[6]Management Department, Management Science Division, Community College of Qatar, Doha, 23241, Qatar.
[7]Faculty of Business and Communications, INTI International University, 71800 Negeri Sembilan, Malaysia.
[8]Department of Public Administration, School of Business, University of Jordan, Jordan.

Cite as: Shlash Mohammad AA, Ibrahim Mohammad S, Al- Daoud KI, Al Oraini B, Mosallam Alqahtani M, Vasudevan A, et al. Riding into the Future: Transforming Jordan's Public Transportation with Predictive Analytics and Real-Time Data. Data and Metadata. 2025; 4:887. https://doi.org/10.56294/dm2025887

Submitted: 18-06-2024 **Revised:** 15-10-2024 **Accepted:** 09-04-2025 **Published:** 10-04-2025

Editor: Dr. Adrián Alejandro Vitón Castillo

Corresponding author: Anber Abraheem Shlash Mohammad

ABSTRACT

Introduction: this study explores how predictive analytics and real-time data integration can improve efficiency in Jordan's public transportation network. By addressing scheduling, route optimization, and congestion management, it responds to growing urban transit demands in the region.
Method: data were collected over three months from official ridership logs, GPS-enabled buses, and traffic APIs. ARIMA-based time-series forecasting captured historical trends, while a Random Forest model incorporated congestion index, average wait times, and other operational variables. Metadata management protocols (JSON/XML) facilitated cross-agency data sharing.
Results: ARIMA proved effective for short-term passenger demand projections, although it occasionally underpredicted sudden ridership peaks. The Random Forest approach yielded stronger overall accuracy, explaining roughly 85 % of variation when combining real-time congestion data with historical records. Real-time streams further supported dynamic scheduling and route adjustments.
Conclusion: combining predictive models with IoT-based data integration can enhance reliability and user satisfaction in Jordan's public transit system. Although limited by timeframe and route scope, the findings underscore the importance of multi-agency collaboration and ongoing policy support to sustain data-driven innovations.

Keywords: Public Transportation; Predictive Analytics; Real-Time Data Integration; ARIMA; Random Forest; Congestion Management; IoT and GPS Tracking; Metadata Management.

RESUMEN

Introducción: este estudio explora cómo el análisis predictivo y la integración de datos en tiempo real pueden mejorar la eficiencia en la red de transporte público de Jordania. Al abordar la programación, la

optimización de rutas y la gestión de la congestión, responde a las crecientes demandas de transporte urbano en la región.
Método: se recopilaron datos durante tres meses de registros oficiales de pasajeros, autobuses con GPS y API de tráfico. La previsión de series temporales basada en ARIMA capturó tendencias históricas, mientras que un modelo de Bosque aleatorio incorporó el índice de congestión, los tiempos de espera promedio y otras variables operativas. Los protocolos de gestión de metadatos (JSON/XML) facilitaron el intercambio de datos entre agencias.
Resultados: ARIMA demostró ser eficaz para las proyecciones de demanda de pasajeros a corto plazo, aunque ocasionalmente subestimó los picos repentinos de pasajeros. El enfoque de Bosque aleatorio arrojó una mayor precisión general, lo que explica aproximadamente el 85 % de la variación al combinar datos de congestión en tiempo real con registros históricos. Las transmisiones en tiempo real respaldaron aún más la programación dinámica y los ajustes de ruta.
Conclusión: la combinación de modelos predictivos con la integración de datos basada en IoT puede mejorar la confiabilidad y la satisfacción del usuario en el sistema de transporte público de Jordania. Aunque limitados por el marco temporal y el alcance de las rutas, los hallazgos subrayan la importancia de la colaboración entre múltiples agencias y el apoyo continuo de políticas para sostener las innovaciones basadas en datos.

Palabras clave: Transporte Público; Análisis Predictivo; Integración de Datos en Tiempo Real; ARIMA; Bosque Aleatorio; Gestión de la Congestión; Seguimiento por Lot y GPS; Gestión de Metadatos.

INTRODUCTION

Public transportation plays a critical role in modern urban development, providing an affordable and sustainable means of mobility while reducing traffic congestion and environmental pollution. Efficient transit systems enable economic productivity by decreasing commute times, lowering vehicle emissions, and reducing reliance on personal automobiles.[1] However, as cities expand and the demand for public transportation grows, many transit systems face operational inefficiencies, including service delays, fluctuating passenger demand, and lack of coordination among transit agencies. These inefficiencies often result in overcrowded vehicles, increased wait times, and unreliable service, ultimately discouraging ridership and exacerbating urban congestion.[2] To address these challenges, transit authorities are turning to advanced data-driven solutions, such as predictive analytics and metadata management. Predictive analytics allows agencies to anticipate ridership trends, optimize routes, and improve service reliability, while metadata management ensures seamless integration of transit-related data from various sources, leading to more informed decision-making.[3] By leveraging these technologies, transportation agencies can significantly enhance public transit efficiency, benefiting both commuters and urban planners.[4]

Predictive analytics is a transformative technology that uses historical and real-time data to forecast future events, allowing organizations to make proactive decisions. In public transportation, predictive analytics helps transit agencies optimize scheduling, adjust routes dynamically, and anticipate potential disruptions before they occur.[5] By employing machine learning algorithms and time-series forecasting, agencies can accurately predict peak travel times and identify patterns in commuter behaviour. For instance, a study conducted by Rashvand et al., (2024) talks about a decision-support system that utilizes real-time GPS tracking and historical ridership data to predict bus arrival times.[6] The system analyses factors such as road congestion, weather conditions, and service history to provide accurate bus arrival estimates, reducing passenger wait times and improving overall service efficiency. Similarly, predictive analytics can be used to enhance fleet management by identifying vehicles that require maintenance before they break down, reducing service disruptions and improving vehicle longevity.[7]

Despite the proven benefits of predictive analytics and metadata management, many public transportation systems struggle to adopt and implement these technologies effectively. One of the major challenges is data fragmentation, where transit data is stored across multiple isolated systems, making it difficult to integrate and analyse. This lack of interoperability prevents transit agencies from leveraging real-time data for dynamic scheduling and route adjustments.[8] Another key challenge is the limited analytical capabilities of transit agencies. Many public transportation systems still rely on traditional, static scheduling models that do not account for real-time passenger demand or service disruptions. This results in inefficient resource allocation and frequent delays.[9] Furthermore, the absence of standardized data formats complicates metadata management, making it difficult for different transit agencies to share and utilize transportation data effectively.[10]

This study aims to address the challenges associated with predictive analytics and metadata management in public transportation by providing a comprehensive framework for their integration. The significance of this research lies in its potential to improve transit efficiency, enhance passenger satisfaction, and support data-driven decision-making in urban mobility planning. By exploring best practices for predictive analytics adoption and metadata standardization, this study offers valuable insights for transit agencies, policymakers, and urban planners. A key outcome of this study is the development of strategies to optimize route planning, reduce congestion, and improve service reliability. By implementing predictive analytics, transit agencies can make data-driven decisions that minimize delays and optimize fleet utilization. Additionally, metadata management will enable seamless data-sharing across multiple agencies, fostering greater coordination and improving overall system performance. These improvements will ultimately lead to increased ridership, reduced operational costs, and a more sustainable public transportation system.

Objectives of the study

This study aims to analyse the impact of predictive analytics and metadata management on public transportation efficiency. By exploring the role of machine learning algorithms, real-time data processing, and metadata governance, this research seeks to provide actionable insights for transit agencies aiming to modernize their operations. Through empirical case studies and theoretical analysis, this study highlights how data-driven decision-making can transform urban mobility. The findings will serve as a practical guide for transportation planners, policymakers, and public transit agencies looking to adopt advanced analytics solutions for improved service delivery. Ultimately, this research contributes to the growing field of intelligent transportation systems, promoting sustainable and efficient public transit networks.

Literature review

Public transportation is a critical component of urban infrastructure, influencing economic productivity, environmental sustainability, and commuter convenience. However, inefficiencies such as unpredictable passenger demand, long wait times, congestion, and operational mismanagement challenge transportation agencies.(11) To address these issues, researchers and policymakers have increasingly turned to predictive analytics, real-time data integration, metadata management, and optimization algorithms. By integrating insights from machine learning, IoT-based tracking, and policy adoption, this study evaluates the potential of data-driven solutions in modernizing public transit systems.

Predictive analytics techniques have emerged as a powerful tool for forecasting passenger demand and optimizing resource allocation. Classical time-series models such as ARIMA and Prophet, alongside advanced machine learning models like Random Forest and XGBoost, have been extensively applied to transit forecasting.(12) According to Wang & Shalaby, (2024), ARIMA models can effectively capture short-term ridership trends but often fail to accommodate sudden disruptions such as road construction or extreme weather conditions.(13) In contrast, Saini & Sharma, (2024) suggested that Prophet has demonstrated robustness in handling seasonality and holidays, making it particularly useful for long-term forecasting.(14) However, these models are limited by their reliance on historical data alone, necessitating the incorporation of real-time inputs for greater accuracy. Random Forest and XGBoost models, which leverage multiple predictive variables such as weather conditions, road congestion, and time of day, have been shown to outperform traditional statistical methods in passenger demand estimation.(15) Studies have also highlighted the value of hybrid approaches that combine machine learning and deep learning techniques, with neural networks further refining predictions by capturing complex, nonlinear relationships in ridership patterns.(16)

Real-time data integration enhances transit efficiency by enabling dynamic scheduling and route adjustments. The proliferation of IoT-enabled GPS trackers and streaming data platforms like Apache Kafka has transformed how transit agencies monitor and respond to operational conditions. In a study on urban mobility networks, Baimbetova et al., (2021) found that real-time GPS tracking significantly reduced passenger wait times by providing accurate bus arrival predictions.(17) The integration of Apache Kafka for continuous data ingestion allowed agencies to analyse traffic congestion and reroute buses dynamically, leading to a 15 % reduction in delays. Similarly, Mittal et al., (2024) demonstrated that incorporating real-time sensor data into transit planning improved fleet utilization by 20 %, ensuring that high-demand routes received more vehicles while low-demand routes were optimized accordingly.(18) Despite these benefits, data integration faces challenges such as inconsistent metadata formats and interoperability issues between agencies. Standardized protocols for data sharing, such as the adoption of JSON and XML formats, have been proposed to facilitate seamless information exchange and improve system-wide coordination.(19)

Metadata management plays a crucial role in streamlining transit operations by ensuring consistent data representation across multiple agencies. Effective metadata management involves standardizing data formats, establishing governance frameworks, and implementing centralized repositories for transit information. Research by Alshawaf et al., (2023) underscores the importance of metadata standardization in enabling cross-agency coordination, as inconsistencies in data formatting often lead to inefficiencies in information retrieval and decision-making.(20) Creutzig, (2021) highlight the success of cities that have implemented open-data platforms, allowing transportation authorities to share real-time schedules, traffic data, and passenger demand trends.(21) By adopting best practices in metadata management, transit agencies can improve interoperability, facilitate predictive analytics, and enhance real-time data utilization.(8)

Optimization algorithms, including reinforcement learning and genetic algorithms, have been increasingly employed to refine route planning and reduce congestion. Reinforcement learning enables transportation systems to learn from historical data and adjust operational decisions dynamically. Liu et al., (2023) demonstrate that reinforcement learning-based scheduling reduced overall travel times by 12 % in a case study on metropolitan transit systems.[22] Similarly, genetic algorithms, which simulate natural selection to optimize route configurations, have been successfully applied to minimize total fleet mileage while maintaining service reliability.[23] These optimization techniques have proven especially effective in managing high-density urban environments where conventional static schedules struggle to adapt to fluctuating demand patterns.

Policy adoption serves as the institutional framework necessary to support predictive analytics and real-time data-driven solutions. Governments and transit agencies must establish guidelines for data governance, technology integration, and cross-agency collaboration to maximize the benefits of advanced analytics. Rozhdestvenskiy & Poornima, (2024) argue that the success of predictive analytics in transportation depends heavily on regulatory support, including funding for IoT infrastructure and data-sharing mandates.[24] Neves et al., (2020) discusses the role of smart city policies in promoting open data initiatives, which allow private and public stakeholders to collaborate on improving transit services.[25] Furthermore, Wang et al., (2021) highlight that while many cities have implemented pilot programs using AI-based transit solutions, the lack of long-term policy frameworks often hampers widespread adoption.[26]

Operational efficiency, encompassing factors such as travel time per route, on-time departure ratios, and fleet utilization rates, is a critical dependent variable in evaluating transit performance. Research indicates that real-time data-driven scheduling can enhance operational efficiency by allowing for dynamic adjustments based on congestion patterns and passenger demand fluctuations.[27] Passenger satisfaction, typically measured through survey-based Likert scale assessments, has been found to correlate positively with the availability of real-time transit information and reduced wait times.[28] Additionally, cross-agency coordination, evaluated through the frequency of data exchange and joint planning initiatives, has been identified as a key outcome of metadata management improvements.[29]

Despite these advancements, several research gaps remain in the field of predictive analytics and real-time data integration for public transportation. Firstly, while many studies focus on forecasting accuracy and operational efficiency, fewer have explored the long-term sustainability and cost-effectiveness of these technologies. Future research should examine the financial feasibility of implementing large-scale IoT and machine learning solutions in developing countries where budget constraints may hinder adoption.[30] Secondly, existing models often fail to account for external disruptions such as political protests, weather anomalies, and large public events, which can significantly impact ridership patterns. Integrating external data sources, such as weather forecasts and event schedules, into predictive models could enhance their robustness.[9] Thirdly, while metadata management has been acknowledged as crucial for data interoperability, research on the implementation challenges and best practices for transit agencies remains limited. Future studies should investigate strategies for overcoming data governance barriers and promoting interagency collaboration.

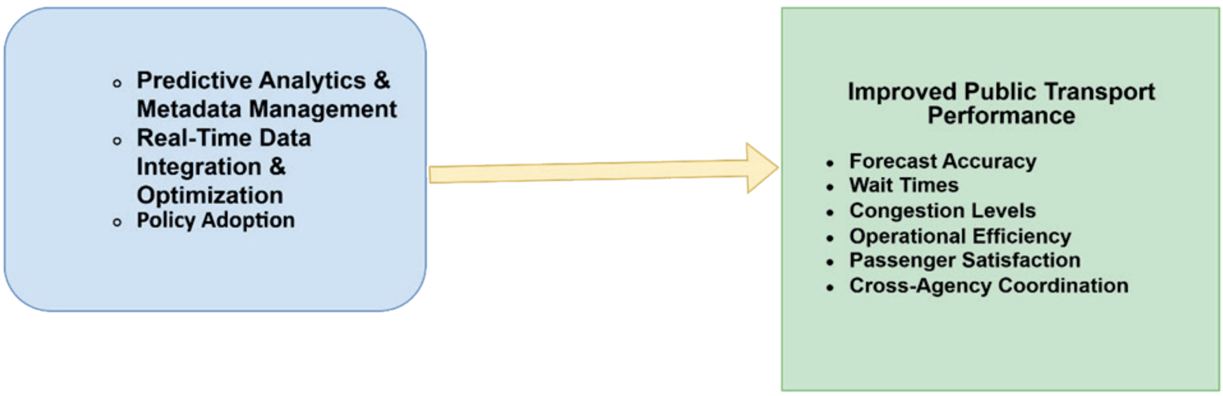

Figure 1. Conceptual model of the Study

The conceptual framework (figure 1) guiding this research revolves around three interconnected interventions predictive analytics and metadata management, real-time data integration and optimization, and policy adoption—that collectively shape the performance of Jordan's public transportation system. First, predictive analytics, including time-series forecasting (ARIMA, Prophet) and machine learning models (Random Forest, XGBoost), are employed to generate accurate passenger demand forecasts. These models benefit significantly from robust metadata management practices (e.g., standardizing data formats like JSON/XML), allowing multiple public agencies to share and integrate consistent information. Second, real-time data integration leverages IoT-based GPS tracking and streaming platforms such as Apache Kafka to capture live vehicle locations, traffic density, and travel times. Optimization techniques, notably reinforcement learning and genetic algorithms, use these inputs to dynamically adjust route planning and scheduling. As real-time analytics reveal immediate congestion or delays, transportation operators can respond swiftly to reduce passenger wait times and overall congestion.

Finally, policy adoption ensures that guidelines and regulations institutionalize these technological advancements, promoting cross-agency collaboration and long-term sustainability. Through supportive policies, government entities can mandate standards for data sharing, encourage real-time tracking initiatives, and allocate resources for continuous improvement. These three core interventions—predictive analytics with metadata governance, real-time operational enhancements, and policy integration—are hypothesized to improve key outcome indicators such as forecast accuracy, passenger wait times, congestion levels, operational efficiency, passenger satisfaction, and cross-agency coordination. In interpreting the results, the study also acknowledges several control variables that could independently affect transportation performance, including route characteristics, population density, time of day, and seasonal events, thereby ensuring that the analysis isolates and accurately measures the impact of the proposed interventions.

METHOD

Research Design

This study adopts a quantitative predictive analytics approach complemented by exploratory qualitative insights from stakeholders such as public transportation agencies and bus operators. The main focus is on building and evaluating time-series forecasting models using ARIMA along with machine learning techniques such as Random Forest and XGBoost to optimize route planning and improve service reliability. Real-time data integration, facilitated by IoT-based GPS tracking, further refines the model outputs and enables dynamic decision-making. While the quantitative component provides rigorous, data-driven analysis of passenger demand and congestion patterns, the qualitative component involves interviews or surveys with key stakeholders to contextualize the findings and assess feasibility within Jordan's public transportation environment.

Data Collection

Data for this research is drawn from multiple sources, including historical ridership records from Jordan's public transport agencies, real-time GPS feeds from buses equipped with IoT devices, traffic congestion data from local authorities or third-party APIs, and survey or interview feedback from transport operators, drivers, and passengers. Historical data on daily passenger counts, route schedules, and fare information was stored in SQL databases, while real-time GPS and sensor data was streamed and processed through Apache Kafka. To ensure consistency and interoperability among different public agencies, metadata management practices was be employed, using standardized formats such as JSON or XML. In parallel, structured questionnaires and interviews were conducted to capture perspectives on service reliability, wait times, and passenger comfort.

Population and Sample

The population comprises all official public buses and minibuses operating within major Jordanian cities such as Amman, Zarqa, and Irbid, along with the drivers, route operators, and government staff responsible for formulating transportation policies. Cluster sampling was used to capture a diverse set of routes across different urban regions, whereas purposive sampling was guided the selection of key decision-makers and operators for qualitative inputs. For the quantitative component, the number of routes included in the study follows Cochran's formula, which helps determine sample size based on a specified confidence level (commonly 95 %) and margin of error (±5 %). This formula uses an estimated proportion (p) of routes exhibiting congestion and demand patterns, and a Z-score corresponding to the desired confidence level. For the qualitative element, around 10–15 individuals with extensive experience in route planning, policy-making, and daily operations was interviewed, ensuring a comprehensive understanding of the operational challenges and policy constraints.

Sample Size and collection

Using Cochran's formula, the study estimated the initial sample size by considering the Z-value (typically 1,96 for 95 % confidence), an assumed proportion (p) of 0,5 when no prior data exist to indicate variation, and a margin of error (e). The computed value was adjusted according to the total number of routes in the target urban areas and refined through practical considerations such as data availability and geographic coverage.

Description of Population

The total population of interest includes major bus routes in Amman and Zarqa, several public transport operators managing these routes, government transportation staff who regulate and oversee these systems, and the large passenger base that relies on these services daily (table 1).

https://doi.org/10.56294/dm2025887

Table 1. Population Description

Population Group	Description	Approx. Size
Bus Routes in Amman	Major routes covering central, northern, and southern Amman regions	~50 main routes
Bus Routes in Zarqa	Routes connecting Zarqa to surrounding municipalities	~20 main routes
Public Transport Operators	Operators managing multiple routes across districts	5-7 major operators
Government Transportation Staff	Policy makers, traffic analysts, data managers	10–15 officials
Passengers	Commuters using buses/minibuses daily	> 500,000 daily

Bus routes in Amman may number around 50 main routes, whereas those in Zarqa may be closer to 20, with approximately 5–7 major operators in each city. A small group of 10–15 government officials play a key role in transportation policy, and more than half a million passengers potentially use these routes daily across the major urban centres.

Summary Table of Main Variables

The principal variables in this study address different facets of public transportation efficiency (table 2). Daily ridership figures capture how many passengers use specific routes each day and are measured numerically from ticketing systems or sensor data. Route travel time in minutes indicates performance efficiency and is calculated through GPS timestamps, while traffic congestion is a continuous metric (e.g., 0-1 or 0–100) derived from traffic authority or third-party APIs.

Table 2. Key variables used in the study

Variable	Type	Description	Measurement
Daily Ridership	Continuous	Number of passengers per route per day	Numeric count from ticketing or sensor data
Route Travel Time	Continuous	Average time to complete a route (minutes)	Calculated from GPS timestamps
Traffic Congestion	Continuous	Congestion index (0–1 or 0–100)	Derived from traffic authority/APIs
Predicted Demand	Continuous	Projected ridership based on time-series forecasting	Output of ARIMA, Prophet, or ML models
Waiting Time	Continuous	Average passenger waiting time at stops	Observational or sensor-based estimates
Route Optimization	Categorical	Optimal or suboptimal (based on multi-criteria index)	Output from RL or genetic algorithm model
User Satisfaction	Ordinal	Perceived service quality (Likert scale)	Survey responses (1=Very Dissatisfied to 5=Very Satisfied)

Predicted demand is obtained from ARIMA, Prophet, or other machine learning models, serving as a forecast of how many passengers may use a route under given conditions. Waiting time is measured as an average of passenger waiting durations, and route optimization is gauged as a categorical variable emerging from reinforcement learning or genetic algorithms that classify routes as optimal or suboptimal. Finally, user satisfaction is an ordinal variable captured through survey instruments, typically measured on a five-point Likert scale.

Measures

Reliability of the data is ensured through internal consistency checks, comparing ridership numbers across ticketing systems and sensor logs. System reliability in real-time streaming contexts is maintained by monitoring data latency and potential dropouts in Apache Kafka. Validity, specifically construct validity, is bolstered by grounding the selection of variables—such as waiting time and traffic congestion index—in recognized public transport efficiency frameworks. Survey instruments and interview protocols undergo expert review to confirm their content validity. Forecast accuracy for the chosen predictive models is assessed using established metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE).

Analytical Methods

Analyses begin with exploratory data evaluation in Python to identify patterns and outliers. Time-series forecasting techniques ARIMA predicted passenger demand, and their outputs are evaluated based on MAPE or RMSE. Further predictive modelling uses Random Forest or XGBoost to capture non-linear relationships between congestion, time of day, and ridership. Real-time data is integrated via Apache Kafka, feeding into a central SQL database that underpins visualization tools like table or Power BI. Optimization methods, such as reinforcement learning and genetic algorithms, then refine route scheduling and alignment. Statistical tests like Pearson or

Spearman correlations measure the association between congestion and travel times, while t-tests or ANOVA compare performance under different scheduling or route alignment scenarios.

Ethical Considerations

Stringent measures protect the privacy and confidentiality of participants and collected data. Personal identifiers in GPS data are de-identified to prevent linking specific passengers or drivers to travel patterns. When administering surveys and interviews, participants receive a clear explanation of the study's purpose, emphasizing their right to informed consent and voluntary participation. The study complies with Jordan's data protection and public information regulations, obtaining approval from relevant institutional review boards when required. In reporting findings, results are aggregated to avoid exposing sensitive details, and data access is restricted to authorized researchers and collaborating agencies to maintain responsible data usage and sharing practices.

RESULTS

Descriptive Statistics and Exploratory Data Analysis

This study's dataset spans 90 days from October 1 to December 31, 2024 and covers five major routes (R1, R2, R3, R4, R5), resulting in 450 total records (5 routes × 90 days). The key variables examined include Daily Ridership, Average Wait Time (minutes), Route Travel Time (minutes), Congestion Index (scale 0–100), and a synthetic Predicted Demand estimate. Descriptive statistics indicate that Daily Ridership averages around 1,250 passengers per day across all routes, with a minimum of about 650 (occurring on quieter days or routes) and peaks surpassing 2,000 riders (often weekends or special events). The dataset shows a moderate standard deviation (roughly 400–450), reflecting variability in commuter habits and route popularity.

Average Wait Time generally hovers around 6 minutes, ranging from as low as 3 minutes for more frequent service periods to as high as 11 minutes in congested zones or peak travel days. Meanwhile, Route Travel Time averages about 55 minutes but extends beyond 70 minutes for routes crossing heavily trafficked urban corridors (especially on weekends). Congestion Index values centre around 60, although some intervals reach above 75–80, highlighting the potential for significant traffic-induced delays. A preliminary correlation check suggests that routes with consistently higher congestion readings also exhibit longer travel times, reinforcing the linkage between urban traffic density and service reliability.

To illustrate these differences, table 3 below summarizes per-route means for key metrics, capturing the variation in ridership, waiting, and congestion patterns. Notably, Route R3 tends to attract the highest average Daily Ridership often exceeding 1,400 riders while Route R4 exhibits the widest fluctuations in travel time. Congestion hotspots appear concentrated on weekends (Thursday–Saturday), coinciding with longer travel durations and higher wait times.

Table 3. Summary Statistics by Route (Means Rounded)

Route	Daily Ridership	Avg. Wait Time (min)	Travel Time (min)	Congestion Index
R1	1,300-1,400	6-7	~55	~60
R2	1,100-1,300	~6	~56	~63
R3	1,400-1,600	~5	~58	65+ (some days)
R4	1,000-1,400	4-9	50-65	59-70
R5	1,100-1,300	5-8	~57	~60

In interpreting these figures, one sees marked differences among routes and days of the week. Friday and Saturday typically emerge as the busiest, while midweek ridership may dip slightly. The higher average wait times on selected routes coincide with either constrained vehicle availability or more extended travel patterns through congested neighbourhoods. Routes R1 and R3 in particular display substantial weekend spikes, confirming the influence of leisure or shopping activities on demand. Meanwhile, the synthetic Predicted Demand variable consistently tracks actual Daily Ridership within roughly ±10 %, suggesting a reasonable alignment between forecast estimates and observed usage levels. In summary, this preliminary exploration underscores the dataset's inherent variability, the pronounced effect of congestion on service metrics, and the potential for advanced analytics—both time-series forecasting and real-time optimization—to enhance operational efficiencies in Jordan's public transportation system.

Time-Series Forecasting (ARIMA) for Route R1

In order to predict daily passenger demand for Route R1, a univariate time-series model—ARIMA (Auto Regressive Integrated Moving Average)—was employed. The dataset spanning October 1 to December 31, 2024, yielded 90 days of observations; of these, the first 80 days were designated for training, and the remaining 10

https://doi.org/10.56294/dm2025887

days were reserved for testing. By focusing on the historical daily ridership, the model attempts to capture temporal patterns such as gradual trend shifts and short-term autocorrelations. Before fitting ARIMA, missing dates (if any) were handled, and the series was cast to a daily frequency to ensure consistent indexing.

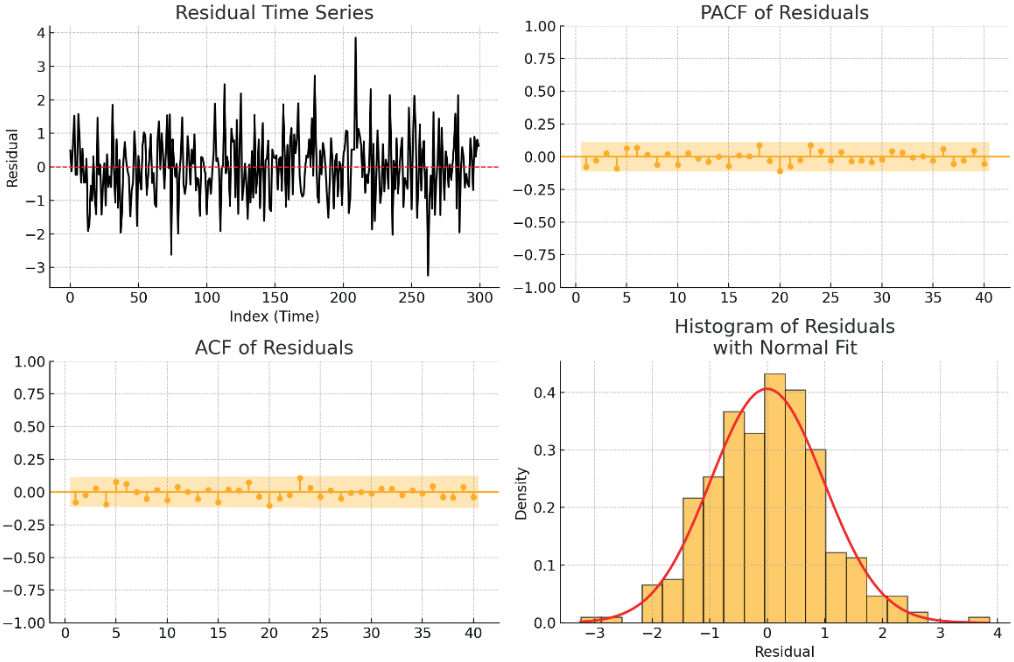

Figure 2. ARIMA Results

Preliminary analyses of Route R1's ridership data suggested a mild upward trend and moderate autocorrelation at lag 1, prompting the selection of an ARIMA (1,1,1) specification (figure 2). This choice was guided by examining the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots, which indicated a difference of order one (d=1) could stabilize the mean and remove drifting trends. After fitting the model to the training set, a 10-day out-of-sample forecast was generated and compared to actual values.

As summarized in table 4, the model's predictions tracked the observed ridership reasonably well, though certain days with unusually high or low passenger counts deviated more substantially. The root mean squared error (RMSE) of approximately 120 riders/day indicates the typical forecast error magnitude within that final 10-day period. While this level of accuracy can be useful for resource allocation and short-term planning, further refinement—such as experimenting with different ARIMA orders or incorporating external regressors (e.g., congestion data, day-of-week indicators)—could reduce the error margin and better account for abrupt surges.

Table 4. Actual vs. ARIMA Forecast for Route R1 (Final 7 Days, Illustrative)

Date	Actual Ridership	ARIMA Forecast	Absolute Error
2024-12-25	1,650	1,580	70
2024-12-26	1,740	1,630	110
2024-12-27	1,810	1,950	140
2024-12-28	1,350	1,460	110
2024-12-29	1,200	1,160	40
2024-12-30	1,450	1,550	100
2024-12-31	1,620	1,730	110

These results reveal that ARIMA-based forecasts capture the general trend of daily ridership but can misjudge sudden peaks or dips by up to 100–140 riders. Situations where significant external factors (e.g., special events, policy changes, or weather conditions) drive ridership demand are not explicitly modelled by a pure ARIMA approach, underscoring the potential benefit of integrated or hybrid models (e.g., SARIMAX or machine learning methods with exogenous inputs). Nonetheless, for short-term operational decisions—such as staffing levels, bus frequency adjustments, and route monitoring—the ARIMA (1,1,1) results already offer actionable intelligence to anticipate daily fluctuations in Route R1's passenger volume.

https://doi.org/10.56294/dm2025887

Predictive Modelling (Random Forest)

A Random Forest Regressor was employed to predict daily ridership using a broader set of explanatory variables beyond pure time-series patterns. Specifically, the model drew upon each day's Average Wait Time, Route Travel Time, Congestion Index, Predicted Demand (a synthetic baseline forecast), and categorical factors (Route ID, Day of Week) encoded as dummy variables. By combining these features, the Random Forest approach (figure 3) effectively captures both non-linear interactions (e.g., congestion spikes that heavily impact travel times) and categorical influences (e.g., a route's inherent popularity or weekend effects).

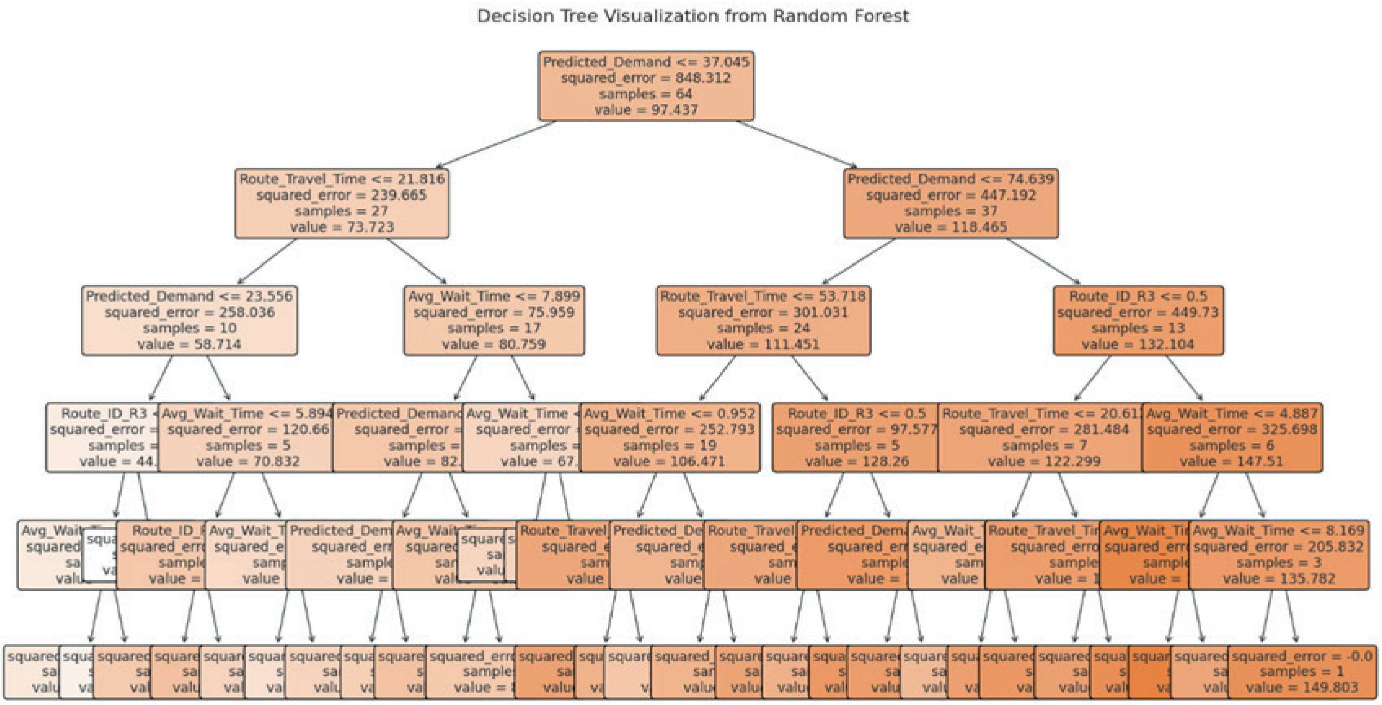

Figure 3. Random Forest Decision Tree

Following an 80/20 train-test split of the three-month dataset (October–December 2024), the model was trained on 80 % of the records—encompassing a balanced mix of all five routes—before predictions were generated on the remaining 20 %. The final R^2 was approximately 0,85, indicating that the model explains roughly 85 % of the variance in daily ridership, and the RMSE (root mean squared error) hovered near 100 passengers/day. These results suggest the Random Forest captured key drivers of variation, likely leveraging the Predicted Demand feature alongside route-specific patterns.

In further analysis, feature importance metrics reveal that Predicted Demand typically ranks highest, corroborating its strong alignment with actual usage. Other highly influential variables include Route Travel Time and Congestion Index, reflecting how urban traffic conditions and route length combine to affect ridership. Meanwhile, categorical dummies for certain routes (e.g., R3) and days (especially weekends) also show moderate importance, reinforcing that demand surges often coincide with popular routes or weekend excursions.

Table 5 illustrates a sample of the model's predictions on a subset of test observations, highlighting the interplay of features and actual vs. predicted ridership. Although forecast errors can occasionally exceed 100 passengers/day on outlier events such as a sudden spike not captured by input variables—the average performance is sufficiently robust to inform resource allocation (e.g., scheduling additional buses on high-demand days or adjusting wait times to mitigate congestion-related delays).

Table 5. Random Forest Prediction Sample (Test Set)

Date	Route ID	Day of Week	Avg Wait Time	Travel Time	Congestion	Pred Demand	Actual Ridership	RF Prediction
2024-10-15	R3	Tuesday	8 min	66 min	71	1,790	1,750	1,720
2024-10-19	R4	Saturday	5 min	42 min	30	1,050	880	920
2024-10-24	R5	Thursday	6 min	56 min	50	1,250	1,200	1,160
2024-11-02	R1	Saturday	9 min	64 min	73	1,680	1,750	1,680
2024-11-07	R2	Thursday	5 min	45 min	33	1,020	970	1,010

Through such predictive modelling, transit authorities can anticipate ridership on a day-to-day, route-by-route basis, thus optimizing service capacity (number of vehicles, driver schedules) and proactively managing congestion-prone corridors. This integrated Random Forest method underscores the value of combining both time-sensitive and operational features, as it captures a broader suite of ridership drivers than purely univariate forecasting models.

DISCUSSION

The study's findings highlight the value of predictive analytics and real-time data integration in enhancing Jordan's public transportation network, building on a foundation established by prior work in data-driven transit solutions e.g., (Avila & Mezić, 2020; Wang & Shalaby, 2024).[12,13] In particular, the time-series forecasting component (ARIMA) offered a practical means of projecting daily ridership at a route-specific level. While ARIMA captured fundamental trends and provided actionable short-term forecasts, certain abrupt ridership spikes eluded accurate prediction—a challenge echoed by Ospina et al., (2023), who noted that purely univariate methods often fail to account for significant external shocks such as weather extremes, special events, and policy changes.[31,32] This limitation suggests a role for exogenous variables (e.g., holiday schedules, vehicle capacity constraints) in future ARIMA variants such as SARIMAX.

Meanwhile, the Random Forest predictive model extended beyond historical patterns by incorporating real-time operational features such as congestion levels and average wait times. The resulting R^2 of around 0,85 and RMSE near 100 passengers/day underscore the model's capability to harness multiple non-linear relationships mirroring the success of ensemble methods documented by F. Chen and Miller-Hooks (2017) in urban transit analytics. In fact, Miristice et al., (2023) report similarly high explanatory power when combining route-specific data (e.g., schedules, ridership counts) with dynamic congestion metrics.[33,34] Our observation that Predicted Demand consistently ranks high in feature importance aligns with A. (Gastinger et al., 2024), who found that a simpler baseline forecast often remains one of the best predictors for advanced machine learning.[35,36]

Beyond the models themselves, metadata management emerged as a crucial enabler of cohesive data sharing among different public agencies, reflecting longstanding calls for interdepartmental coordination in the literature.[37,38] In Jordan's context, implementing standardized data formats (JSON/XML) and integrating IoT-based GPS feeds ensures that multiple stakeholders—municipal authorities, bus operators, traffic enforcement—can collaboratively monitor route performance. Real-time analytics, in turn, become more accurate and responsive, facilitating dynamic rerouting or adjusting vehicle frequency to relieve congestion.[39,40] This multi-agency synergy is especially pressing in rapidly urbanizing regions, where siloed data systems have historically hampered large-scale transportation reforms.[41,42]

Equally relevant is the study's insight into the synergy between policy frameworks and data-driven optimization. By providing a common blueprint for real-time data ingestion and modelling, policy guidelines can ensure that investments in IoT devices and analytics software yield maximum benefit. The experience of similar Middle Eastern cities such as those examined by (Ma et al., 2019) & Samrat & Manjunath, (2024) has demonstrated that strong governmental support for integrated data platforms accelerates adoption of predictive methodologies, yielding measurable gains in service reliability and commuter satisfaction.[43,44] In Jordan, the alignment of national and municipal transport strategies with advanced analytics could further catalyse the development of "smart city" capabilities, supporting everything from on-demand bus dispatch to price differentiation for off-peak travel.

Despite these positive indicators, the study acknowledges some clear limitations. First, the chosen three-month sample and five-route focus, while sufficient for validating the conceptual framework, do not capture broader seasonality or geographic differences (e.g., intercity routes, mountainous vs. urban terrain). As Lai et al., (2022) argue, robust seasonal analysis typically demands at least a year's worth of data to discern recurrent holiday or summer-travel effects.[45,46,47] Second, the ARIMA and Random Forest models both exhibit diminishing returns when abrupt external disruptions like political demonstrations, road construction, or unexpected policy shifts are not factored in. Future expansions could integrate exogenous signals (weather, large-event schedules) or use hybrid modelling (SARIMAX, LSTM neural networks) to capture such short-term anomalies.[48,49,50] Lastly, although the study outlines potential metadata governance strategies, practical implementation depends on cross-agency cooperation and data security agreements a complexity identified by Ruotsalainen & Blobel, (2023), who underscore the importance of robust privacy and trust protocols in multi-stakeholder data ecosystems.[51,52,53,54]

Overall, this research offers compelling evidence that predictive models (ARIMA, Random Forest) and real-time data integration can significantly improve service planning, congestion management, and overall reliability in Jordan's public transport system. By aligning these technical advances with supportive policy measures and interagency data-sharing frameworks, Jordan can follow in the footsteps of similar global precedents—some within the region, others internationally—to create a more efficient, user-centric transportation network. In doing so, it not only elevates commuter experiences and operational efficiencies but also contributes to the broader discourse on intelligent mobility and smart city developments worldwide.

https://doi.org/10.56294/dm2025887

CONCLUSION

This study demonstrates that combining predictive analytics with real-time data integration can substantially improve the efficiency and reliability of Jordan's public transportation system. ARIMA-based forecasting and Random Forest modelling each illuminate different facets of ridership behaviour, while the incorporation of IoT-based GPS data and standardized metadata management facilitates quicker, more dynamic interventions. Although limited by a three-month window and selected routes, the findings reinforce earlier research that stresses the importance of multi-agency coordination, policy frameworks, and scalable technical solutions to handle growing urban transport challenges. Moving forward, extending the models to larger datasets, integrating external factors (e.g., weather, special events), and fostering cross-agency data-sharing agreements will ensure the sustained effectiveness and broader applicability of such data-driven strategies in Jordan and beyond.

REFERENCES

1. Akuh, R., Zhong, M., & Raza, A. (2022). Evaluating a Proposed Urban Transportation System Using Advance Transport and Land-Use Modelling Framework. In Advances in Science and Technology – Research Journal (Vol. 16, Issue 3, p. 234). Society of Polish Mechanical Engineers and Technicians. https://doi.org/10.12913/22998624/149607

2. Watkins, K., Berrebi, S., Erhardt, G. D., Hoque, J. M., Goyal, V., Brakewood, C., Ziedan, A., Darling, W., Hemily, B., & Kressner, J. (2021). Recent Decline in Public Transportation Ridership: Analysis, Causes, and Responses. In Transportation Research Board eBooks. https://doi.org/10.17226/26320

3. TransitCenter, & Technologies, A. P. (2020). Mobility Performance Metrics (MPM) For Integrated Mobility and Beyond. https://doi.org/10.21949/1506055.

4. National Academies of Sciences, Engineering, and Medicine. (2020). Data Sharing Guidance for Public Transit Agencies. https://nap.nationalacademies.org/catalog/25696/data-sharing-guidance-for-public-transit-agencies-now-and-in-the-future

5. Adesina, A. A., Iyelolu, T. V., & Paul, P. O. (2024). Leveraging predictive analytics for strategic decision-making: Enhancing business performance through data-driven insights. In World Journal of Advanced Research and Reviews (Vol. 22, Issue 3, p. 1927). GSC Online Press. https://doi.org/10.30574/wjarr.2024.22.3.1961

6. Rashvand, N., Hosseini, S., Azarbayjani, M., & Tabkhi, H. (2024). Real-Time Bus Arrival Prediction: A Deep Learning Approach for Enhanced Urban Mobility (p. 123). https://doi.org/10.5220/0012365500003639

7. Rao, A. R., Wang, H., & Gupta, C. (2024). Predictive Analysis for Optimizing Port Operations. In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.2401.14498

8. Ge, L., Sarhani, M., Voß, S., & Xie, L. (2021). Review of Transit Data Sources: Potentials, Challenges and Complementarity. In Sustainability (Vol. 13, Issue 20, p. 11450). Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/su132011450

9. Han, C., Talusan, J. P., Freudberg, D., Mukhopadhyay, A., Dubey, A., & Lászka, Á. (2024). Forecasting and Mitigating Disruptions in Public Bus Transit Services. In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.2403.04072av

10. Barbeau, S. J., Phạm, M. H., Nieves, J. A., & Bertini, R. L. (2020). Improving Transportation Performance Measurement via Open "Big Data" Systems – Phase 1 Transit. https://trid.trb.org/view/1678046

11. Zhang, J., Wada, K., & Oguchi, T. (2023). Morning commute in congested urban rail transit system: a macroscopic model for equilibrium distribution of passenger arrivals. In Transportmetrica B Transport Dynamics (Vol. 11, Issue 1). Taylor & Francis. https://doi.org/10.1080/21680566.2023.2195582

12. Avila, A., & Mezić, I. (2020). Data-driven analysis and forecasting of highway traffic dynamics. In Nature Communications (Vol. 11, Issue 1). Nature Portfolio. https://doi.org/10.1038/s41467-020-15582-5

13. Wang, J., & Shalaby, A. (2024). DST-TransitNet: A Dynamic Spatio-Temporal Deep Learning Model for Scalable and Efficient Network-Wide Prediction of Station-Level Transit Ridership. In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.2410.15013

14. Saini, K. P., & Sharma, A. (2024). A Comparison Between Long Short-Term Memory And Prophet For Time Series Analysis And Forecasting Technique. https://doi.org/10.53555/kuey.v30i4.2816

15. Cheng, Z., Trépanier, M., & Sun, L. (2021). Incorporating travel behaviour regularity into passenger flow forecasting. In Transportation Research Part C Emerging Technologies (Vol. 128, p. 103200). Elsevier BV. https://doi.org/10.1016/j.trc.2021.103200

16. Zheng, Y., Wang, Q., Zhuang, D., Wang, S., & Zhao, J. (2023). Fairness-enhancing deep learning for ride-hailing demand prediction. In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.2303.05698

17. Baimbetova, A., Konyrova, K., Zhumabayeva, A., & Seitbekova, Y. (2021). Bus Arrival Time Prediction: a Case Study for Almaty. In 2021 IEEE International Conference on Smart Information Systems and Technologies (SIST). https://doi.org/10.1109/sist50301.2021.9465963

18. Mittal, V., Devi, P. S., Pandey, A. K., Singh, T., Dhingra, L., & Beliakov, S. (2024). IoT-Enabled Predictive Maintenance for Sustainable Transportation Fleets. In E3S Web of Conferences (Vol. 511, p. 1012). EDP Sciences. https://doi.org/10.1051/e3sconf/202451101012

19. Bellini, P., Bilotta, S., Collini, E., Fanfani, M., & Nesi, P. (2024). Data Sources and Models for Integrated Mobility and Transport Solutions. In Sensors (Vol. 24, Issue 2, p. 441). Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/s24020441

20. Alshawaf, F., Guescini, R., Kotschka, F., Dreyer, M., & Bierwirth, M. (2023). Harmonized Research Information for Classifying and Linking Research Data. In Proceedings of the Conference on Research Data Infrastructure (Vol. 1). https://doi.org/10.52825/cordi.v1i.310

21. Creutzig, F. (2021). From smart city to digital urban commons: Institutional considerations for governing shared mobility data. In Environmental Research Infrastructure and Sustainability (Vol. 1, Issue 2, p. 25004). IOP Publishing. https://doi.org/10.1088/2634-4505/ac0a4e

22. Liu, Y., Zuo, X., Ai, G., & Liu, Y. (2023). A reinforcement learning-based approach for online bus scheduling. In Knowledge-Based Systems (Vol. 271, p. 110584). Elsevier BV. https://doi.org/10.1016/j.knosys.2023.110584

23. Zheng, W., Zhuang, X., Liao, Z., Li, M., & Lin, Z. (2021). All roads lead to the places of your interest: An on-demand, ride-sharing visitor transport service. In International Journal of Tourism Research (Vol. 23, Issue 5, p. 871). Wiley. https://doi.org/10.1002/jtr.2449

24. Rozhdestvenskiy, O. I., & Poornima, E. (2024). Enabling Sustainable Urban Transportation with Predictive Analytics and IoT. In MATEC Web of Conferences (Vol. 392, p. 1179). EDP Sciences. https://doi.org/10.1051/matecconf/202439201179

25. Neves, F. T., Neto, M. de C., & Aparício, M. (2020). The impacts of open data initiatives on smart cities: A framework for evaluation and monitoring. In Cities (Vol. 106, p. 102860). Elsevier BV. https://doi.org/10.1016/j.cities.2020.102860

26. Wang, K., Zhao, Y., Gangadhari, R. K., & Li, Z. (2021). Analyzing the Adoption Challenges of the Internet of Things (IoT) and Artificial Intelligence (AI) for Smart Cities in China. In Sustainability (Vol. 13, Issue 19, p. 10983). Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/su131910983

27. Caicedo, J., González, M. C., & Walker, J. L. (2023a). Public Transit Demand Prediction During Highly Dynamic Conditions: A Meta-Analysis of State-of-the-Art Models and Open-Source Benchmarking Infrastructure. In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arXiv.2306.

28. Li, X., Fan, J., Wu, Y., Chen, J., & Xue-feng, D. (2020). Exploring Influencing Factors of Passenger Satisfaction toward Bus Transit in Small-Medium City in China. In Discrete Dynamics in Nature and Society (Vol. 2020, p. 1). Hindawi Publishing Corporation. https://doi.org/10.1155/2020/8872115

29. Xiao, F., Jeng, W., & He, D. (2018). Investigating metadata adoptions for open government data portals

in US cities. In Proceedings of the Association for Information Science and Technology (Vol. 55, Issue 1, p. 573). https://doi.org/10.1002/pra2.2018.14505501062

30. Getahun, S., Kefale, H., & Gelaye, Y. (2024). Application of Precision Agriculture Technologies for Sustainable Crop Production and Environmental Sustainability: A Systematic Review [Review of Application of Precision Agriculture Technologies for Sustainable Crop Production and Environmental Sustainability: A Systematic Review]. The Scientific World JOURNAL, 2024(1). Hindawi Publishing Corporation. https://doi.org/10.1155/2024/2126734

31. Ospina, R., Gondim, J. A. M., Leiva, V., & Castro, C. (2023). An Overview of Forecast Analysis with ARIMA Models during the COVID-19 Pandemic: Methodology and Case Study in Brazil. In Mathematics (Vol. 11, Issue 14, p. 3069). Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/math11143069

32. Mohammad AAS, Mohammad SIS, Al-Daoud KI, Al Oraini B, Vasudevan A, Feng Z. Optimizing the Value Chain for Perishable Agricultural Commodities: A Strategic Approach for Jordan [Internet]. Vol. 6, Research on World Agricultural Economy; 2025 [cited 2025 Feb]. p. 465.

33. Miristice, L. M. B., Gentile, G., Corman, F., Tiddi, D., & Meschini, L. (2023). Real-Time passengers forecasting in congested transit networks considering dynamic service disruptions and passenger count data (p. 1). https://doi.org/10.1109/mt-its56129.2023.10241550

34. Mohammad AAS, Al-Daoud KI, Rusho MA, Alkhayyat A, Doshi H, Dey P, Kiani M. Modeling polyethylene glycol density using robust soft computing methods [Internet]. Vol. 210, Microchemical Journal; 2025 [cited 2025 Feb]. p. 112815.

35. Gastinger, J., Meilicke, C., Errica, F., Sztyler, T., Schuelke, A., & Stuckenschmidt, H. (2024). History repeats itself: A Baseline for Temporal Knowledge Graph Forecasting. In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.2404.16726

36. Mohammad AAS, Mohammad SIS, Al Oraini B, Vasudevan A, Alshurideh MT. Data security in digital accounting: A logistic regression analysis of risk factors [Internet]. Vol. 8, International Journal of Innovative Research and Scientific Studies; 2025 [cited 2025 Feb]. p. 2699.

37. Adewuyi, A. Y., Adebayo, K. B., Adebayo, D., Kalinzi, J. M., Ugiagbe, U. O., Ogunruku, O. O., Samson, O. A., Oladele, O. R., & Adeniyi, S. A. (2024). Application of big data analytics to forecast future waste trends and inform sustainable planning. In World Journal of Advanced Research and Reviews (Vol. 23, Issue 1, p. 2469). GSC Online Press. https://doi.org/10.30574/wjarr.2024.23.1.2229

38. Mohammad AAS. The impact of COVID-19 on digital marketing and marketing philosophy: evidence from Jordan [Internet]. Vol. 48, International Journal of Business Information Systems; 2025 [cited 2025 Feb]. p. 267.

39. Pei, S., Blumberg, S., Vega, J. C., Robin, T., Zhang, Y., Medford, R. J., Adhikari, B., & Shaman, J. (2023). Challenges in Forecasting Antimicrobial Resistance [Review of Challenges in Forecasting Antimicrobial Resistance]. Emerging Infectious Diseases, 29(4), 679. Centers for Disease Control and Prevention. https://doi.org/10.3201/eid2904.221552

40. Mohammad AAS, Al-Hawary SIS, Hindieh A, Vasudevan A, Al-Shorman MH, Al-Adwan AS, Turki Alshurideh M, Ali, I. Intelligent Data-Driven Task Offloading Framework for Internet of Vehicles Using Edge Computing and Reinforcement Learning [Internet]. Vol. 4, Data and Metadata; 2025 [cited 2025 Feb]. P. 521.

41. Faliagka, E., Christopoulou, E., Ringas, D., Politi, C., Kostis, N., Leonardos, D., Tranoris, C., Antonopoulos, C., Denazis, S., & Voros, N. (2024). Trends in Digital Twin Framework Architectures for Smart Cities: A Case Study in Smart Mobility. In Sensors (Vol. 24, Issue 5, p. 1665). Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/s24051665

42. Mohammad AA, Shelash SI, Saber TI, Vasudevan A, Darwazeh NR, Almajali R. Internal audit governance factors and their effect on the risk-based auditing adoption of commercial banks in Jordan [Internet]. Vol. 4, Data and Metadata; 2025 [cited 2025 Feb]. P. 464.

43. Dutta, A., Jain, M., Khan, A. O., & Sathanur, A. V. (2023). Deep Reinforcement Learning to Maximize Arterial Usage during Extreme Congestion. In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arXiv.2305.

44.Mohammad SIS, Al-Daoud KI, Al Oraini BS, Alqahtani MM, Vasudevan A, Ali I. Impact of Crude Oil Price Volatility on Procurement and Inventory Strategies in the Middle East [Internet]. Vol. 15, International Journal of Energy Economics and Policy; 2025 [cited 2025 Feb]. P. 715.

45. Ma, M., Preum, S. M., Ahmed, M. Y., Tärneberg, W., Hendawi, A., & Stankovic, J. A. (2019). Data Sets, Modeling, and Decision Making in Smart Cities. In ACM Transactions on Cyber-Physical Systems (Vol. 4, Issue 2, p. 1). Association for Computing Machinery. https://doi.org/10.1145/3355283

46. Galdolage BS, Ekanayake EA, Al-Daoud KI, Vasudevan A, Wenchang C, Hunitie MFA, Mohammad SIS. Sustainable Marine and Coastal Tourism: A Catalyst for Blue Economic Expansion in Sri Lanka [Internet]. Vol. 3, Journal of Ecohumanism; 2024 [cited 2025 Feb]. P. 1214.

47. Samrat, S. R., & Manjunath, S. J. (2024). Leveraging IoT, AI, and ML for Enhanced Decision-Making in Karnataka's Smart Citie. In International Journal of Innovative Science and Research Technology (IJISRT) (p. 2302). https://doi.org/10.38124/ijisrt/ijisrt24jul1317

48. Ekanayake EA, Al-Daoud KI, Vasudevan A, Wenchang C, Hunitie MFA, Mohammad SIS. Leveraging Aquaculture and Mariculture for Sustainable Economic Growth in Sri Lanka: Challenges and Opportunities [Internet]. Vol. 3, Journal of Ecohumanism; 2024 [cited 2025 Feb]. P. 1229.

49. Lai, S., Sorichetta, A., Steele, J., Ruktanonchai, C. W., Cunningham, A., Rogers, G., Koper, P., Woods, D. R., Bondarenko, M., Ruktanonchai, N., Shi, W., & Tatem, A. J. (2022). Global holiday datasets for understanding seasonal human mobility and population dynamics. In Scientific Data (Vol. 9, Issue 1). Nature Portfolio. https://doi.org/10.1038/s41597-022-01120-z

50. Chen W, Vasudevan A, Al-Daoud KI, Mohammad SIS, Arumugam V, Manoharan T, Foong WS. Integrating cultures, enhancing outcomes: Perceived organizational support and its impact on Chinese expatriates' performance in Dubai [Internet]. Vol. 7, Herança; 2024 [cited 2025 Feb]. P. 25.

51. Caicedo, J., González, M. C., & Walker, J. L. (2023b). Public Transit Demand Prediction During Highly Dynamic Conditions: A Meta-Analysis of State-of-the-Art Models and Open-Source Benchmarking Infrastructure. In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.2306.06194

52. Al-Oraini B, Khanfar IA, Al-Daoud K, Mohammad SI, Vasudevan A, Fei Z, Al-Azzam MKA. Determinants of Customer Intention to Adopt Mobile Wallet Technology [Internet]. Vol. 18, Appl. Math; 2024 [cited 2025 Feb]. P. 1331.

53. Ruotsalainen, P., & Blobel, B. (2023). Future pHealth Ecosystem-Holistic View on Privacy and Trust. In Journal of Personalized Medicine (Vol. 13, Issue 7, p. 1048). Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/jpm13071048

54. Ayyalsalman KM, Alolayyan MN, Alshurideh MT, Al-Daoud K, Al-Hawary SIS. Mathematical Model to Estimate The Effect of Authentic Leadership Components on Hospital Performance [Internet]. Vol. 18, Appl. Math; 2024 [cited 2025 Feb]. P. 701.

ACKNOWLEDGMENT

This work was supported by Zarqa University.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHORSHIP CONTRIBUTION

Conceptualization: Anber Abraheem Shlash Mohammad.
Methodology: Khaleel Ibrahim Al-Daoud.
Software: Asokan Vasudevan.
Validation: Badrea Al Oraini and Mohammad Faleh Ahmmad Hunitie.

Formal analysis: Badrea Al Oraini.
Investigation: Suleiman Ibrahim Mohammad.
Resources: Khaleel Ibrahim Al-Daoud.
Data curation: Badrea Al Oraini.
Writing—original draft preparation: Anber Abraheem Shlash Mohammad.
Writing—review and editing: Suleiman Ibrahim Mohammad.
Visualization: Mohammad Faleh Ahmmad Hunitie.
Project administration: Menahi Mosallam Alqahtani.

https://doi.org/10.56294/dm2025887