Keywords: rail freight; forecasting models; Uzbekistan rail transportation; freight prediction; multi-model comparison; machine learning; traditional regression equation

Komil TURAEV[1]*, Artur BUDZYNSKI[2], Aleksander SŁADKOWSKI[3]

A COMPARATIVE STUDY OF TRADITIONAL EXCEL FORECASTING METHODS AND MACHINE LEARNING TECHNIQUES FOR FREIGHT VOLUME PREDICTION IN UZBEKISTAN

Summary. Accurate forecasting of freight volumes is essential for effective transportation planning and infrastructure development. Previous research on Uzbekistan's railway sector primarily relied on single-method approaches, either using traditional statistical tools or machine learning techniques. This study adopts an innovative dual-method framework, combining Excel-based models—such as regression equation, exponential smoothing, and moving average—with advanced machine learning techniques, including decision tree, random forest, gradient boosting, and extreme gradient boosting. Freight shipment data and socio-economic variables, such as gros domestic product and operational railway length. Model performance was evaluated using root mean square error and mean absolute percentage error. The regression equation model demonstrated exceptional precision with a mean absolute percentage error of 0.001%, though its simplicity raised concerns about overfitting and limited scalability. Meanwhile, machine learning models showcased superior robustness and generalization capabilities, achieving low and balanced error rates, making them more suitable for capturing complex, non-linear relationships in freight dynamics. According to the compound annual growth rate projection, freight volumes are expected to increase significantly, reaching 106 million tons by 2030. This underscores the growing importance of strategic infrastructure investment, modernization, and policy interventions to accommodate future demand. The findings provide valuable insights for policymakers and transportation planners, offering a practical and comprehensive framework for sustainable development in Uzbekistan's railway sector. This study aims to lay a foundation for informed decision-making and long-term growth planning by leveraging a mix of traditional and modern forecasting approaches.

1. INTRODUCTION

Freight transportation is pivotal in economic development, trade facilitation, industrial growth, and infrastructure planning. Accurate forecasting of freight volumes is essential for decision-makers to optimize resources, plan for capacity, and meet future demands. Traditional statistical methods such as linear regression, exponential smoothing, and moving average have been widely used in forecasting due to their simplicity and accessibility [1]. However, the rise of machine learning (ML) has introduced more robust and flexible techniques capable of capturing complex, non-linear relationships in data, such as random forest, decision tree, extreme gradient boosting (XGBoost), and gradient boosting [2].

[1] Silesian University of Technology; Krasińskiego 8, 40-019 Katowice, Poland; e-mail: komil.turaev@polsl.pl; orcid.org/0009-0008-5316-0346
[2] Cracow University of Economics; Rakowiecka 27, 31-510 Cracow, Poland; e-mail: abudzyns@uek.krakow.pl; orcid.org/0000-0002-5803-6749
[3] Silesian University of Technology; Krasińskiego 8, 40-019 Katowice, Poland; e-mail: aleksander.sladkowski@polsl.pl; orcid.org/0000-0002-1041-4309
* Corresponding author. E-mail: komil.turaev@polsl.pl

Uzbekistan's rail freight sector has been a cornerstone of national transportation, witnessing growth amid socio-economic changes. However, accurately predicting freight volumes remains challenging due to dynamic factors like GDP, population, employment rates, and infrastructure expansion [3, 4]. To address this matter, the present study evaluates and compares the accuracy of traditional Excel-based forecasting models and advanced machine-learning techniques using freight shipment data from Uzbekistan.

Previous research on forecasting freight transport in Uzbekistan's railway sector has primarily focused on traditional statistical and mathematical models. For instance, [5] employed a mathematical model to predict cargo flow during the construction of the Uzbekistan–Kyrgyzstan–China railway, while [6] developed a traditional model to forecast the dynamics and growth of throughput in Central Asian transport corridors. Similarly, [7] utilized the autoregressive integrated moving average (ARIMA) model, a time series technique, to forecast the share of railways in the industry. While these studies provide valuable insights, they are limited to single-method approaches, either traditional or statistical, highlighting the gap and the novelty of integrating both traditional and machine learning methods in forecasting freight transport in Uzbekistan.

This study employs Excel-based forecasting techniques, including linear regression, exponential smoothing, and moving average, alongside advanced machine learning models such as random forest regression, decision tree regression, XGBoost, gradient boosting, and polynomial regression. The performance of these models is evaluated using metrics such as mean absolute percentage error (MAPE) and root mean square error (RMSE) to determine their accuracy and suitability for freight forecasting. By providing a detailed comparison of traditional and modern forecasting approaches, this research offers valuable insights into their respective strengths and limitations, ultimately supporting policymakers in making more informed decisions about managing and predicting freight volumes.

Additionally, the compound annual growth rate (CAGR) method is utilized to forecast future values of critical socio-economic and infrastructure-related variables, including GDP, population, and railway operational length. CAGR's ability to provide a consistent and smooth growth rate over a defined period makes it a reliable tool for projecting these factors. When combined with regression analysis, this approach ensures that the forecasts are well-aligned with historical trends, enhancing the reliability and validity of the predictions for freight volumes in 2030.

2. LITERATURE REVIEW

Traditional methods like linear regression, exponential smoothing, and moving average have been extensively applied in transportation forecasting due to their ease of use and interpretability. Linear regression assumes a linear relationship between predictors and the target variable, making it suitable for trend-based data. However, it fails to capture complex, non-linear patterns [1].

Exponential smoothing is a popular method for time series data, as it assigns exponentially decreasing weights to past observations. Moving average methods, on the other hand, smooth fluctuations over time but may not adapt well to sudden changes in trends [1]. Despite their simplicity, these models often yield higher errors when applied to large and complex datasets.

Advancing technologies have significantly influenced forecasting methods, with researchers increasingly combining traditional techniques with machine learning approaches. A relevant example is the study on Kazakhstan's National Railway Company (KTZ), which transitioned from traditional expert-based forecasting methods to quantitative techniques, specifically the ARIMA model, to predict railway freight demand. By utilizing historical data and validating results through mean absolute error (MAE) and MAPE, the study highlighted the ARIMA model's superiority over qualitative approaches in enhancing forecast accuracy. The findings underscore the importance of modernized planning and resource allocation practices, aligning with contemporary industry standards [8].

ML techniques offer a powerful alternative to traditional methods by accommodating complex, non-linear relationships and leveraging large datasets. Models such as random forest regression and decision tree regression have shown significant accuracy in time series forecasting due to their ability to handle both categorical and numerical predictors. The article [9] examined the application of machine learning techniques in rail transit signaling systems, analyzed their impact on safety and efficiency, and proposed an evaluation method to ensure system reliability and safety from the perspective of independent safety assessment entities. The article [10] presented a novel machine learning-based voting regression model for accurately estimating wheel-rail adhesion in railway vehicles, leveraging weighted combinations of models such as histogram-based gradient boosted trees, random forest, and linear regression, and demonstrating superior performance compared to traditional methods.

XGBoost [11] is a scalable, efficient tree-boosting system that leverages novel algorithms for sparse data, optimized cache access, data compression, and sharding to achieve state-of-the-art performance on massive datasets with minimal resources. XGBoost and gradient boosting are ensemble learning methods that combine weak learners to build a strong predictive model. These models are robust against overfitting and often outperform traditional regression approaches. Studies by Zhang et al. [2] demonstrated that XGBoost yields lower forecasting errors in transportation demand prediction compared to classical models. Polynomial regression, while often categorized as a traditional model, can also be implemented with machine learning tools to capture curvilinear relationships. However, its overfitting tendencies limit its applicability to smaller datasets.

The article [12] introduced a machine learning-based analytical framework to predict broken rail occurrences on commuter railroads, utilizing algorithms such as light gradient boosting machine (LightGBM), XGBoost, and random forests, and addressing data imbalance with oversampling techniques, highlighting gradient, operational speed, and prior rail defects as key predictive factors.

The article [13] presented a machine learning-based methodology to enhance freight capacity utilization by predicting freight weight and traffic counts along transport routes, utilizing models such as random forest, artificial neural networks, and support vector machines, and demonstrates the potential of this approach for optimizing freight operations in urban environments despite challenges related to feature data limitations.

Error metrics play a crucial role in evaluating forecasting models. Commonly used metrics include MAPE and absolute forecasting error. MAPE is widely preferred due to its interpretability and ability to provide a percentage-based error, making it suitable for comparing models across scales.

The CAGR method is a widely used metric for assessing the consistent growth of variables over a specified period. Its primary advantage lies in its ability to smooth out fluctuations, providing a clear picture of the average annual growth rate, even in the presence of volatile changes in the data. As noted by Investopedia, [17] CAGR is especially useful in financial and economic forecasting, enabling analysts to extrapolate trends for long-term planning. Its application is highlighted [18] in a research article where compound growth rate analysis was used for agricultural outputs, emphasizing its relevance in forecasting and policy decision-making. Furthermore, Wall Street Prep [19] underscored its significance in evaluating performance across sectors, including transportation and logistics, where consistent growth indicators like GDP and infrastructure expansion are crucial. By integrating CAGR with other forecasting models, researchers ensure alignment with historical data trends, enhancing prediction accuracy and reliability for strategic planning.

3. METHODOLOGY

The dataset used for this study includes historical data on freight transport volume (the dependent variable) and eight socio-economic indicators (the independent variables) from 2013 to 2022. These indicators include the number of rolling stocks, population, average annual employment, GDP, operational length of public railways, registered enterprises, pipeline cargo turnover, and freight turnover by automotive transport. The data were sourced from the official Statistics Agency under the President of the Republic of Uzbekistan [20].

Prior to analysis, the dataset was preprocessed to ensure completeness and consistency. Any missing values were addressed using appropriate imputation methods, and all variables were normalized to enable comparability. The selected variables, which exhibited relatively high correlations with freight shipment volume, are presented in Table 1.

3.1. Excel-Based Forecasting Models

Three traditional forecasting methods were implemented in Excel:

A linear regression model was employed to establish a relationship between the dependent variable, freight volume, and various independent variables. The model is represented by the following equation:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon. \quad (1)$$

This formula predicts the target variable Y (freight volume) based on a set of independent variables, $X_1 + X_2 + \cdots X_n$ (e.g., GDP, employment rate, or population).

- β_0: The intercept, which represents the predicted value of Y when all X variables are zero.
- $\beta_1 + \beta_2 + \cdots + \beta_n$: Coefficients that quantify the impact of each independent variable on Y.
- ϵ: The error term capturing the deviation between the observed and predicted values.

This method finds the best-fit line through the data points by minimizing the sum of squared residuals.

The exponential smoothing model is a time-series forecasting technique for predicting freight volumes. This method is particularly useful for handling data with trends or seasonality by assigning exponentially decreasing weights to past observations, thereby emphasizing more recent data. The model is expressed mathematically as:

$$F_t = \alpha Y_{t-1} + (1 - \alpha) F_{t-1}. \quad (2)$$

This formula forecasts F_t, the value of freight volume for time t, by weighting the most recent actual observation (Y_{t-1}) and the previous forecast (F_{t-1}).

- α: The smoothing constant, where $0 < \alpha < 1$, determines the emphasis on recent versus past data.
- Higher α gives more weight to recent observations, making the forecast more reactive to recent changes.
- This iterative model updates the forecast as new data become available, making it suitable for trends or seasonality.

Table 1

Selected freight transport volume as a dependent variable Y and eight socio-economic indicators as independent variables X1…X8 [20]

Year	Y Freight shipment by railway transport (million tons)	X1 (Number of rolling stock, pcs)	X2 (Population, thousand people)	X3 (Annual average number of employed people, thousand)	X4 (Gross domestic product, billion soums)	X5 (Operational length of public railways, km)	X6 (Registered enterprises and organizations, total)	X7 Pipeline cargo turnover (million ton-km)	X8 (Freight turnover in automotive transport, million ton-km)
2013	63.7	7000	30,500	12,500	153,000	4190	266,000	31.5	11.2
2014	65.7	6870	31,000	12,800	187,000	4200	274,000	31.2	11.9
2015	67.2	6770	31,600	13,100	221,000	4240	279,000	30.0	12.8
2016	67.6	6810	32,100	13,300	256,000	4300	285,000	28.9	13.3
2017	67.9	6780	32,700	13,500	318,000	4640	300,000	30.2	13.6
2018	68.4	6870	33,300	13,300	427,000	4720	339,000	33.6	14.6
2019	70.1	6620	33,900	13,500	533,000	4740	420,000	33.2	15.9
2020	70.6	6570	34,600	13,200	606,000	4730	504,000	26.8	16.2
2021	72.0	6720	35,300	13,500	738,000	4730	558,000	30.8	19.1
2022	73.4	6730	36,000	13,700	888,000	4730	628,000	29.7	20.5

The moving average model is applied as a simple forecasting technique to predict freight volumes. This method calculates the forecast by averaging a fixed number of the most recent observations, effectively smoothing out short-term fluctuations. The equation for the moving average used is:

$$F_t = \frac{1}{n}\sum_{i=t-n}^{t-1} Y_i \ . \tag{3}$$

This formula predicts the freight volume F_t by averaging the actual values (Y_i) from the most recent n periods.

- n: the number of past periods included in the average, chosen to smooth out short-term fluctuations in the data.
- This method provides a straightforward approach to forecasting by assuming that future values are influenced equally by the most recent n periods.
- Different configurations (e.g., 2-year, 3-year, or 4-year moving averages) help assess the sensitivity of the forecast to the number of periods considered.

The error analysis showed that the two-year moving average achieved the lowest MAPE, followed by the three-year and four-year averages with MAPE. These results indicate that the two-year moving average provided the most accurate short-term predictions by closely following recent trends, while the four-year moving average effectively smoothed long-term trends but at the cost of reduced precision in short-term forecasting. This trade-off highlights the importance of selecting an appropriate moving average period based on the specific forecasting objective.

3.2. Machine Learning Forecasting Models

ML models were implemented in Python using libraries such as Scikit-Learn and XGBoost and are presented below. Scikit-learn [14] is a Python library offering a variety of machine learning algorithms designed for ease of use, efficiency, and versatility, catering to both supervised and unsupervised tasks in academic and commercial contexts.

Random forest regression: This ensemble learning method constructs multiple decision trees during training and combines their outputs to generate a more robust and accurate prediction. By averaging predictions, random forest reduces overfitting and improves generalization, making it ideal for capturing complex, non-linear relationships and handling large datasets with varied features.

Decision tree regression: A straightforward and interpretable regression technique that segments the dataset into distinct subsets by repeatedly splitting the data based on feature thresholds. This process minimizes prediction errors within each subset, effectively capturing local patterns in the data.

XGBoost: A cutting-edge gradient boosting algorithm designed for exceptional efficiency and scalability. XGBoost incorporates advanced regularization techniques to mitigate overfitting and leverages optimized data processing to handle large-scale datasets, delivering state-of-the-art predictive performance.

Gradient boosting: An iterative machine learning approach that sequentially builds a strong predictive model by combining the outputs of several weak models, typically decision trees. Each subsequent model focuses on correcting the errors of the previous ones, improving accuracy over time.

Polynomial regression: A type of regression analysis that extends linear models to account for non-linear relationships by fitting a polynomial equation to the data. This method is particularly effective for capturing curved trends in datasets, enabling more precise predictions in such scenarios.

The paper [15] applied machine learning methods to predict truck fuel consumption, demand, and prices in road freight management, detailing data collection and preprocessing, model training and validation, theoretical underpinnings, and practical implementations. All materials were made publicly available to foster transparency and collaboration.

3.3. Evaluation Metrics

The feasibility of the models was assessed using two error metrics: RMSE and MAPE. These were calculated using the following formulas [16]:

RMSE:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2}, \tag{4}$$

The root mean squared error, which represents the average magnitude of errors between the actual and predicted values. It penalizes larger errors more heavily due to squaring the differences.
Y_i: The observed (actual) value for the i-th data point. This is the real measurement from the dataset.
$\widehat{Y}_i$: The predicted value for the i-th data point, as generated by the model.
$\sum_{i=1}^{n}$:The summation symbol, which indicates that the squared residuals are added together across all n data points.

MAPE:

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \widehat{Y}_i}{Y_i}\right| \times 100, \tag{5}$$

The mean absolute percentage error expresses the average prediction error as a percentage of the actual values.
Y_i: The observed (actual) value for the i-th data point.
$\widehat{Y}_i$: The predicted value for the i-th data point, as generated by the model.
$Y_i - \widehat{Y}_i$: The residual or error for the i-th data point, showing the difference between observed and predicted values.
$\left|\frac{Y_i - \widehat{Y}_i}{Y_i}\right|$: The absolute deviation between the actual and predicted values. Dividing by the observed value normalizes the error as a proportion of the actual value to ensure that all errors are positive.
$\times \mathbf{100}$: Converts the average error into a percentage, making it easier to interpret and compare.

3.4. CAGR

CAGR provides a smooth annual growth rate by assuming that the variable grows constantly over the entire period. It is useful when comparing different time periods or datasets with fluctuating growth, as it removes the effects of short-term volatility and focuses on the long-term trend. For example, if analyzing the growth of rail infrastructure, GDP, or population, CAGR offers a clear, compounded growth rate, making it an ideal metric for projecting future trends based on past data. It simplifies the understanding of complex growth patterns, enabling more accurate forecasting and long-term planning. The equation is calculated by using the following formula [17, 19]:

$$\mathrm{CAGR} = \left(\frac{V_f}{V_i}\right)^{\frac{1}{n}} - 1, \tag{6}$$

V_f: Final value of the variable at the end of the period (e.g., population at the end of 2030).
V_i: Initial value of the variable at the start of the period (e.g., population at the start of 2020).
n: Time duration (e.g., the 10 years from 2020 to 2030).
$\left(\frac{V_f}{V_i}\right)$: Calculation of the growth ratio.

Annualizing growth: Take the n-th root of the growth ratio to determine the average annual growth rate, accounting for compounding. Subtract 1 from the result to find the annual growth rate as a percentage (e.g., a CAGR of 1.05 indicates 5% growth per year).

After applying the CAGR to project the values of the independent variables for 2030, these calculated values were used as inputs in the regression predictive model. This approach enables the estimation of the freight transport volume by railway in Uzbekistan for the target year, providing insights into potential future trends based on historical data and socio-economic growth.

4. RESULTS

4.1. Excel-Based Results

As shown in Fig. 1, the regression equation forecast predicts freight shipment trends with a very close fit to the actual data. The model forecasts the freight shipment for 2022 at 73.37 million tons, compared to the actual observed value of 73.4 million tons, demonstrating negligible error. While the regression equation provides the best fit among the tested models, the near-perfect alignment raises concerns about potential overfitting. Overfitting occurs when a model captures not only the trend but also the noise or irregularities in the historical data, which can limit its ability to generalize to future unseen data.

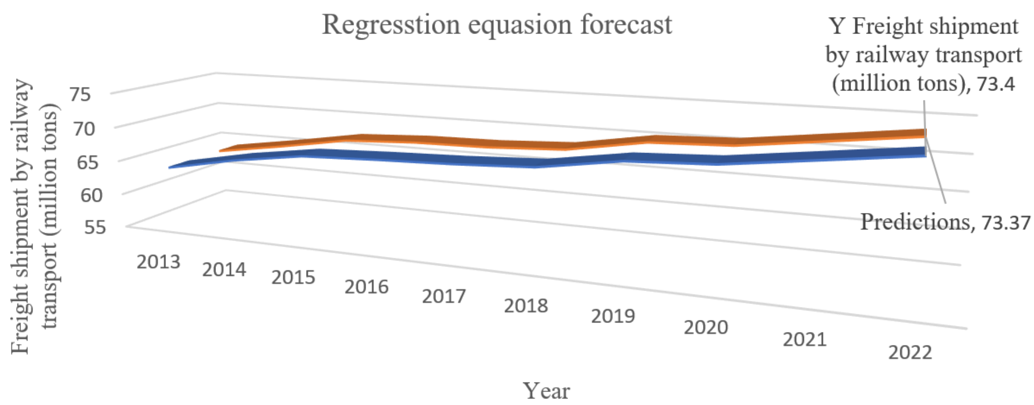

Fig. 1. Regression equation predictive model chart comparing actual vs. predicted data for 2022

This concern should be kept in mind when interpreting the performance of the regression model, as such a close fit may not always translate into reliable predictions for significantly different datasets.

The results of the exponential smoothing model, with a smoothing factor of α = 0.8, are illustrated in Fig. 2. This method smooths historical data by assigning higher weights to recent values, effectively capturing trends while reducing short-term fluctuations. The model predicts a gradual upward trend, with the freight shipment for 2022 forecasted at 71.68 million tons. While exponential smoothing provides reasonable forecasts, deviations are observed during periods of sharper increases, such as 2021–2022, due to its reliance on weighted averages. Compared to the regression model, exponential smoothing appears less prone to overfitting but sacrifices some accuracy during rapid changes in the data trend.

The moving average model generates forecasts by averaging recent observations, effectively smoothing out short-term fluctuations. Forecasted freight shipment values for 2021 and 2022 were 71.3 and 72.7 million tons, respectively, as illustrated in Fig. 3 below.

While the moving average model captures overall trends, it lags during periods of significant change due to its dependence on historical averages. This lag effect makes the moving average less suitable for forecasting data with clear upward or downward trends over time.

4.2. Machine Learning Approach-Based Results

The line chart in Fig. 4 illustrates the actual freight shipment data from 2013 to 2021 alongside the 2022 forecasts from various regression models, with selective offsets applied for better differentiation of points. The chart provides a comparative understanding of the models' forecasting capabilities relative to the actual value for 2022.

The polynomial regression forecast aligns very closely with the actual 2022 value, indicating its capacity to fit the test data precisely. However, this performance may be attributed to overfitting, as suggested by its high MAPE in cross-validation. The decision tree, gradient boosting, and XGBoost models forecast similar values slightly below the actual data point, demonstrating consistent but slightly conservative predictions. Among these, gradient boosting achieved the best generalization performance during cross-validation (lowest MAPE), making it a reliable model for future predictions.

The random forest forecast, while reasonably aligned with the historical trend, deviated the most from the actual value in 2022, suggesting that it may struggle with unseen data despite performing well in training. The visualization emphasizes the importance of balancing trend-capturing ability and generalization to avoid underperforming in dynamic test scenarios.

The presented feature importance results for the gradient boosting model (Fig. 5) reveal several key insights into the drivers behind railway freight transport. The operational length of public railways (X5) emerged as the most influential factor, underscoring the critical role of infrastructure in enabling and expanding freight transport capabilities. This result is logical, as a larger and more developed railway network directly translates to greater capacity for cargo movement. Additionally, gross domestic product (X4) and the number of registered enterprises (X6) ranked as the second and third most important features, emphasizing the strong link between economic activity and demand for freight transport. A higher GDP and a greater number of businesses signal increased industrial activity and trade, driving logistics needs. Other notable factors, such as the number of rolling stock (X1) and population size (X2), further highlight the importance of available transport resources and demographic pressures. These findings indicate that investments in railway infrastructure, coupled with economic growth, are fundamental to enhancing freight transport efficiency.

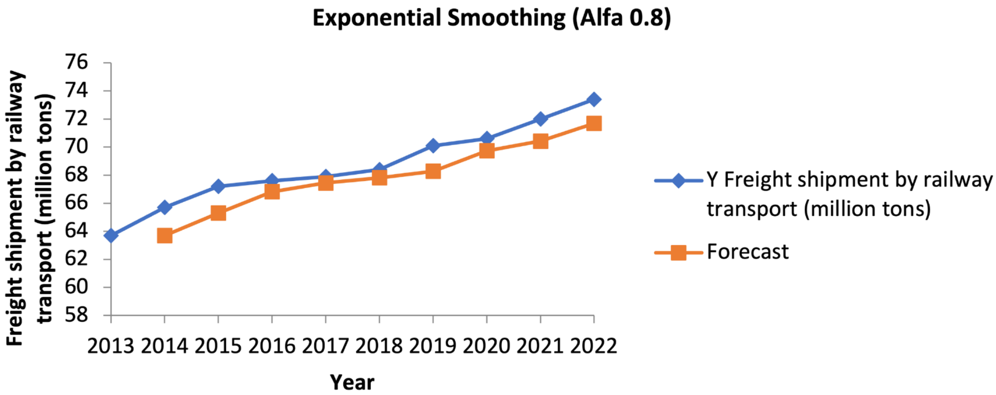

Fig. 2. Exponential smoothing forecast model

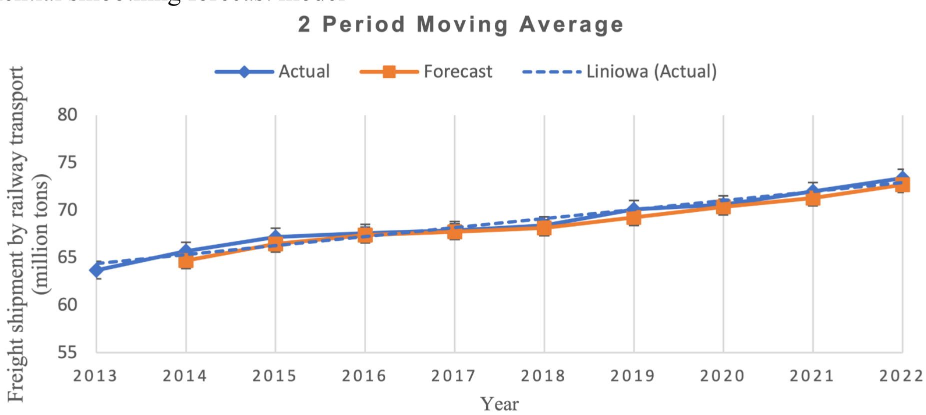

Fig. 3. Moving average forecast model

4.3. Error Metrics Comparison of Excel and ML-Based Predictive Models

Error metrics are critical in evaluating and comparing the accuracy of predictive models. Two widely used metrics, RMSE and MAPE, offer complementary insights. Fig. 6 below comprehensively compares RMSE and MAPE values for both Excel-based and ML-based models, highlighting their respective strengths and limitations.

- RMSE, expressed in absolute terms, measures the average magnitude of prediction errors in the same unit as the dependent variable (e.g., millions of tons). It highlights the overall error magnitude, making it useful for comparing models with the same scale.
- MAPE, presented as a percentage, evaluates prediction errors relative to actual values. By normalizing errors, it enables fair comparison across datasets with different scales.

The performance of Excel-based models shown in Fig. 6 reveals that the regression equation achieves exceptional accuracy, with an RMSE of 0.107 and a remarkably low MAPE of 0.001%. This precision indicates a strong alignment with historical data. However, it raises concerns about potential overfitting and limited generalization to new data. Other Excel-based models, such as exponential smoothing and moving average, demonstrate higher RMSE values (1.238 and 1.620, respectively) and MAPE percentages (1.55% and 2.11%). These models exhibit moderate predictive accuracy with better generalization than the regression equation, making them suitable for short-term forecasting.

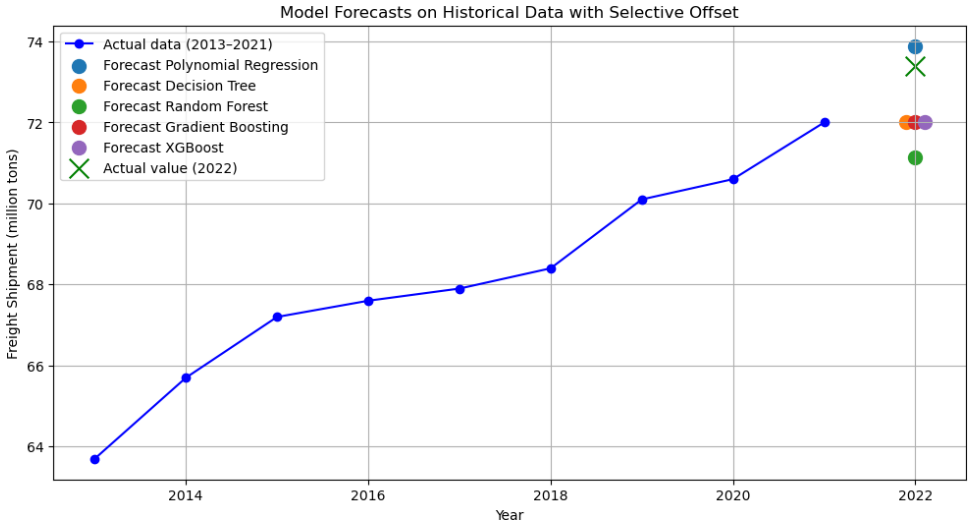

Fig. 4. Model forecasts on historical data with selective offset

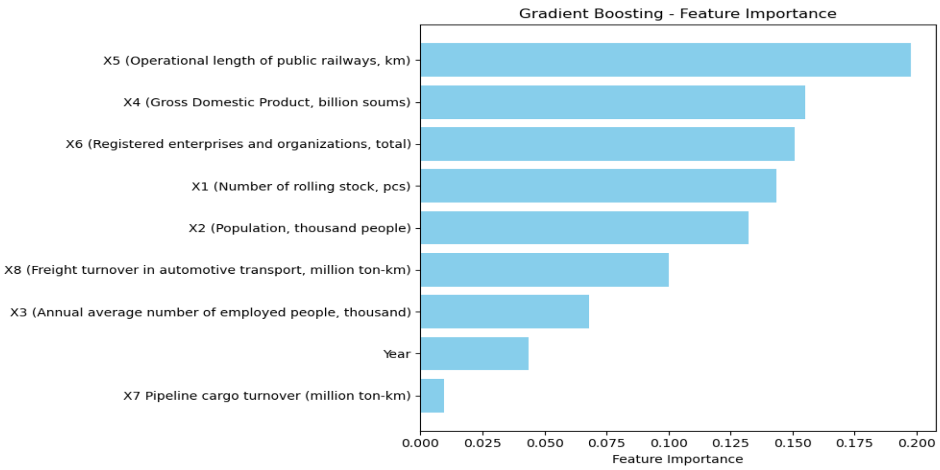

Fig. 5. Feature importance for the gradient boosting regressor

The performance of ML-based models (Fig. 6.) showcases their ability to handle complex, non-linear relationships. Gradient boosting emerged as the most balanced model, with the lowest MAPE (1.0%), reflecting both high accuracy and robust generalization. Random forest follows closely, achieving a MAPE of 1.1%, demonstrating reliable performance across diverse datasets. XGBoost, with a MAPE of 1.8%, provides robust results due to its strong regularization capabilities. Other models, such as decision trees, exhibit moderate accuracy (MAPE of 2%), but their performance is less consistent than that of ensemble methods like random forest. On the other hand, polynomial regression, while capable of capturing non-linear trends, has the highest RMSE (5.74) and MAPE (4.8%), indicating significant overfitting and limited applicability for unseen data.

The analysis highlights the trade-offs between accuracy and generalization in predictive modeling. While traditional Excel-based models like the regression equation provide precise results, their simplicity can lead to overfitting. Machine learning models, particularly gradient boosting and random forest, offer robust performance, balancing predictive accuracy and generalization.

This study emphasizes the importance of a mixed forecasting strategy, combining the interpretability of traditional methods with the adaptability of machine learning techniques. Such a comprehensive approach equips decision-makers in freight transportation planning with reliable tools to address diverse forecasting needs.

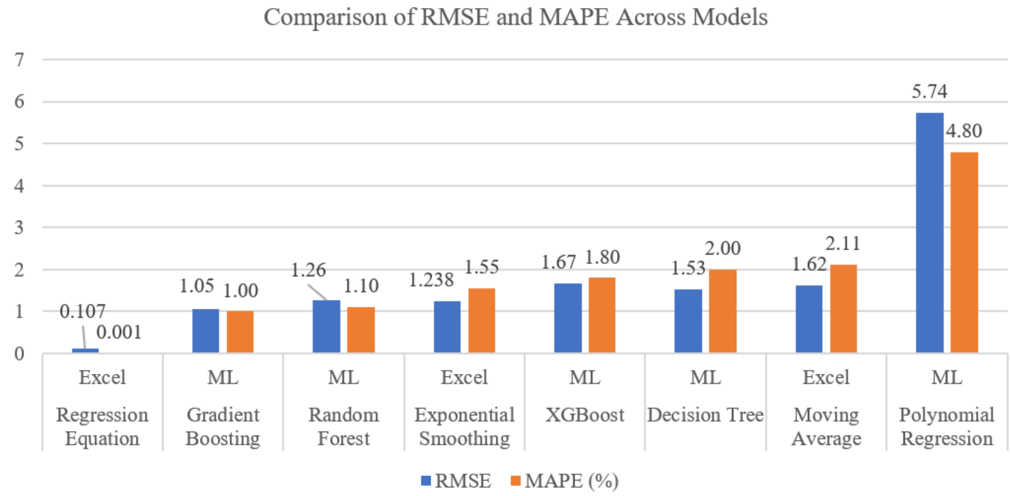

Fig. 6. Comparison of RMSE and MAPE across models

5. CAGR APPLICATION FOR FUTURE PREDICTION

Independent variables were extrapolated for the year 2030 using the CAGR equation outlined in the methodology. The historical data for this analysis, covering the period from 2013–2022, was sourced from the Statistics Agency under the President of the Republic of Uzbekistan [20]. This dataset provided an accurate and reliable foundation for analyzing trends and projecting future values.

- **The extrapolated values of the key independent variables for 2030 are summarized in Table 2.**
- **Substitution into the Regression Equation as a chosen model:**

The extrapolated values for 2030 were substituted into the regression equation given in the methodology section, with the coefficients determined from the historical data analysis (2013–2022). The computation was carried out step-by-step using Excel to ensure accuracy.

- **Substitution into the Regression Model:**

The freight shipment Y was calculated using the extrapolated 2030 values and the regression coefficients:

$$\mathrm{Y} = 98.4 + (-0.00553 \cdot 6500) + (0.000593 \cdot 41800) + (-0.000585 \cdot 14900) +$$
$$+(0.000010 \cdot 4230000) + (-0.00188 \cdot 5260) + (-0.000024 \cdot 1340000) +$$
$$+(-0.0901 \cdot 28) + (0.798 \cdot 35)$$
$$\mathbf{Y = 106\ million\ tons}$$

The predicted railway freight volume for 2030 was 106 million tons, representing a 43.5% increase from the 2022 value of 73.4 million tons.

Interpretation of Results and Suggestions

The analysis underscores the need for strategic investments to modernize and expand the capacity of the railway system. After various predictive models were evaluated, regression analysis, which yielded the least error, was identified as the most reliable approach. Using this model, freight transport predictions for 2030 were calculated through the CAGR method. The results indicate a significant increase in railway freight volume by 2030. Specifically, total freight shipments are projected to reach 106 million tons, a 43.5% increase from 73.4 million tons in 2022. This growth is attributed to several factors, including population growth, economic expansion, and improvements in railway infrastructure, such as rolling stock and expanded operational rail lengths. These projections signal a positive outlook for the railway sector, highlighting its increasingly critical role in meeting national transportation needs.

Table 2

The extrapolated values of variables for the year 2030 and coefficients, intercept subtracted from the regression equation

Independent Variables	CAGR	Extrapolated Values (2030)	Coefficients	Intercept
X1 (Number of rolling stock, pcs)	-0.4%	6500	-0.005530	98.4
X2 (Population, thousand people)	1.9%	41,800	0.000593	
X3 (Annual average number of employed people, thousand)	1.0%	14,900	-0.000585	
X4 (Gross Domestic Product, billion soums)	22.0%	4,234,500	0.000010	
X5 (Operational length of public railways, km)	1.4%	5260	-0.001880	
X6 (Registered enterprises and organizations, total)	10.0%	1,343,850	-0.000024	
X7 Pipeline cargo turnover (million ton-km)	-0.6%	28	-0.090088	
X8 (Freight turnover in automotive transport, million ton-km)	7.0%	35	0.798273	

The model's projections suggest that freight volume in the railway sector will grow by 43.5% by 2030, with the total freight volume increasing from 73.4 million tons in 2022 to 106 million tons. This rise reflects the sector's growing importance in fulfilling Uzbekistan's transportation demands, driven by socio-economic and infrastructural factors. The expected 21.56% increase in GDP and continued infrastructure expansion, including improvements in rolling stock and railway length, are crucial drivers of this growth.

Socio-economic factors, including population growth and rising employment levels, are expected to drive greater demand for freight transport. As the economy expands, industries will require more efficient and reliable transportation networks to move goods, further emphasizing the need to develop the railway sector. Additionally, infrastructure enhancements—such as expanding rail networks and rolling stock—will improve the sector's capacity and efficiency, enabling it to manage the expected increase in freight volumes. These factors demonstrate that the railway sector will play a pivotal role in supporting the nation's economic growth and facilitating the movement of goods.

While these projections are optimistic, several challenges must be addressed to ensure that the railway sector can meet the forecasted demand. Continuous investment in infrastructure is essential to keep pace with the increasing freight volumes. Prioritizing the expansion of rail networks and increasing the number of rolling stock will be key to improving capacity and alleviating congestion on existing lines. Additionally, integrating rail transport with other modes of transport, such as road and pipeline systems, will be critical in creating a more efficient and seamless logistics network.

To support the anticipated growth in railway freight transport, several strategies should be considered:

- Infrastructure investment: The government must prioritize substantial investments in railway infrastructure, focusing on extending rail networks and modernizing rolling stock. This will enable the sector to meet future freight transport demands.
- Public-private partnerships (PPPs): Encouraging collaboration between the public sector and private investors can help finance and develop the necessary railway infrastructure, leveraging both funding and expertise for system modernization.
- Digitalization and automation: Integrating digital technologies and automation into railway operations will enhance efficiency and reduce operational costs. Advanced logistics systems,

real-time tracking, and route optimization can help manage the projected increase in freight traffic.
- Intermodal transportation development: Strengthening the integration of rail systems with road and pipeline transport systems will improve the efficiency of the overall logistics network. The development of intermodal terminals and better connectivity between transport modes will streamline operations and reduce costs.
- Environmental sustainability: As the railway sector expands, it is important to focus on sustainability. Implementing environmentally friendly technologies, such as electric trains, and strategies to reduce carbon emissions will ensure that the railway system's growth aligns with global environmental goals.

By addressing these key areas, Uzbekistan's railway sector can maximize its growth potential, effectively support economic development, and meet the increasing transportation needs of the country sustainably and efficiently.

6. CONCLUSIONS

This study evaluated traditional Excel-based models and ML techniques for forecasting freight volumes in Uzbekistan. While Excel-based models like the regression equation were highly accurate, with a low MAPE (0.001%), their simplicity raises concerns about overfitting. On the other hand, ML models such as gradient boosting and random forest demonstrated stronger robustness and generalizability, making them more suitable for complex, non-linear forecasting. A combined approach, leveraging the interpretability of traditional methods and the adaptability of ML techniques, is recommended for more reliable forecasting.

The analysis also identifies key predictors of freight volumes, such as GDP, rail network length, and the number of registered enterprises, offering valuable insights for transportation planning. The study recommends further exploration of hybrid models and additional variables to refine forecasting accuracy.

In conclusion, the forecast for railway freight transport in Uzbekistan indicates substantial growth by 2030, with a projected 43.5% increase from 2022 to 2030. This growth is driven by socio-economic factors and infrastructure improvements. Strategic investments in infrastructure, fleet modernization, and the integration of digital technologies and sustainable practices are essential to realizing this potential. A comprehensive approach focusing on infrastructure and intermodal connectivity, will ensure a robust and efficient railway system to meet future transportation demands.

References

1. Hyndman, R. & Athanasopoulos, I. *Forecasting: principles and practice. 3rd edition*. OTexts, 2021. Available at: https://Otexts.com/Fpp3/.
2. Zhang, J. & Chen, F. & Cui, Z. et al. Deep learning architecture for short-term passenger flow forecasting in urban rail transit. *IEEE Transactions on Intelligent Transportation Systems*. 2021. Vol. 22. No. 11. P. 1-12.
3. Гуламов, А.А. „Прогнозирование объёмов перевозок грузов на Узбекской железной дороге". *Известия ПГУПС*. 2010. No. 1. P. 12-23. ISSN 1815-588X. [In Russian: Gulamov, A.A. Forecasting the volume of freight transportation on the Uzbek railway. *News of PGUPS*].
4. Сабуров, М. & Бутунов, Д. Прогнозирование погрузки грузов на железных дорогах Узбекистана. *Universum: Технические Науки*. 2021. No. 1(82). P. 71-74. [In Russian: Saburov, M. & Butunov, D. Forecasting of cargo loading on the railways of Uzbekistan. *Universum: Technical Sciences*].
5. Umarov, K. & Tursinaliyeva, Y. & Khurramov, I. & Shayakhmetov, S. Mathematical model for prediction of cargo flow during the construction of the railway line Uzbekistan-Kyrgyzstan-China. *E3S Web of Conferences*. 2023. No. 403. P. 1-9. DOI: 10.1051/e3sconf/202340103018.

6. Kotenko, A. & Sattorov, B. & Nehoroshkov, P. & Timuhin, M. Model for forecasting the dynamics and growth of the throughput of the Central Asian transport corridor lines. *Journal of Physics: Conference Series. Intelligent Information Technology and Mathematical Modeling (IITMM 2021)*. Montreal, Canada. 2021. P. 1-8. DOI: 10.1088/1742-6596/2131/3/032102.
7. Toshaliyeva, S. Using the ARIMA model to forecast the share of railways in the industry. *E3S Web of Conferences*. 2024. No. 531. P. 1-11. DOI: 10.1051/e3sconf/202453102024.
8. Sultanbek, M. & Adilova, N. & Sładkowski, A. & Karibayev, A. Forecasting the demand for railway freight transportation in Kazakhstan: a case study. *Transportation Research Interdisciplinary Perspectives*. 2024. Vol. 23. No. 101028. P. 1-9. ISSN: DOI: 10.1016/j.trip.2024.101028.
9. Hu, G. & Han, J. Applications of machine learning in rail transit signaling system and its independent safety assessment method. *IEEE 8th International Conference on Intelligent Transportation Engineering (ICITE)*. Beijing, China: IEEE. October 2023. P. 1-4. DOI: 10.1109/ICITE59717.2023.10733904.
10. Zirek, A. & Uysal, C. A machine learning-based voting regression method for adhesion estimation in wheel-rail contact. *Vehicle System Dynamics.* 2024. P. 1-18. DOI: 10.1080/00423114.2024.2390578.
11. Chen, T. & Guestrin, C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, California, USA. 2016. P. 785-794. DOI: 10.1145/2939672.293.
12. Kang, D. & Dai, J. & Liu, X. et al. Estimating the occurrence of broken rails in commuter railroads with machine learning algorithms. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit.* 2024. Vol. 238. No. 10. P. 1338-1350. DOI: 10.1177/09544097241280848.
13. Tabib, M.V. & Stene, J.K. & Rasheed, A. et al. Machine learning for capacity utilization along the routes of an urban freight service. Intelligent technologies and applications. *Communications in Computer and Information Science, Cham: Springer International Publishing*. 2022. Vol. 1616. P. 419-432. DOI: 10.1007/978-3-031-10525-8_33.
14. *Scikit-learn development team. Scikit-learn: machine learning in Python*. Available at: http://arxiv.org/abs/1201.0490.
15. Budzyński, A. & Sładkowski, A. Machine learning in road freight transport management. In: *Using Artificial Intelligence to Solve Transportation Problems*. *Studies in Systems, Decision and Control*. Cham: Springer Nature Switzerland. 2024. P. 485-565. DOI: 10.1007/978-3-031-69487-5_9.
16. Chai, T. & Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*. 2014. Vol. 7. No. 3. P. 1247-1250. DOI: 10.5194/gmd-7-1247-2014.
17. *Investopedia*. *Compound Annual Growth Rate (CAGR) Formula and Calculation*. Available at: https://www.investopedia.com/terms/c/cagr.asp.
18. Prajneshu, & Chandran, K.P. Computation of compound growth rates in agriculture: Revisited. *Agricultural Economics Research Review.* 2005. Vol. 18. P. 317-324.
19. *Wall Street Prep. Compound Annual Growth Rate (CAGR) Formula + Calculator.* Available at: https://www.wallstreetprep.com/knowledge/cagr-compound-annual-growth-rate/.
20. *Annual Statistical Collection of the Republic of Uzbekistan 2013-2022*. Tashkent. Statistics Agency under the President of the Republic of Uzbekistan. 2023. Available at: https://stat.uz/uz/nashrlar.

Received 21.11.2023; accepted in revised form 04.06.2025