Time series forecast of inbound call volume in call center using machine learning methods

Ruslan Krasnozhonov[1,†], Aizhan Altaibek[1,2,†], Aizhan Ydydrys[1,†] and Marat Nurtas[1,2,†]

[1] *International Information Technology University, 34/1 Manas St., Almaty, Kazakhstan*
[2]*Institute of Ionosphere, Gardening community IONOSPHERE 117, Almaty, 050020, Kazakhstan*

Abstract

Time series analysis involves examining data collected at various time points to identify patterns, trends, and seasonal changes. Accurate forecasting of future values is crucial for call centers to optimize staff scheduling and manage workloads effectively. While traditional statistical methods and manual forecasting have been widely used, machine learning techniques have shown promising results in enhancing forecast accuracy. This paper explores the application of machine learning for forecasting incoming call volumes, with a focus on comparing Long Short-Term Memory (LSTM) networks and Random Forest Regression models. The models are evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R2) metrics. Experimental results indicate that Random Forest Regression performs well with limited data, achieving competitive MAE and RMSE values. However, LSTM networks outperform Random Forest Regression on an hourly scale, showing superior accuracy and higher R2 scores as the dataset size increases. This study demonstrates that while Random Forest Regression provides stable performance across different data sizes, LSTM models offer significant improvements in forecasting accuracy, particularly with larger datasets and high temporal granularity.

1. Introduction

Time series analysis involves examining sequences of data points collected at successive time intervals to identify patterns, trends, and seasonal variations. This analysis is crucial for forecasting future values, particularly in domains where understanding and predicting temporal patterns can significantly impact decision-making and resource management. One such domain is call center operations, where accurate forecasting of inbound call volumes is essential for optimizing staffing levels and enhancing customer service [1].

Call centers serve as a primary point of contact for customers, making effective workload forecasting critical to prevent operational inefficiencies and improve service quality. Traditional methods for forecasting call volumes often rely on statistical models or manual approaches, which may not fully capture the complex and dynamic nature of call volume patterns. As call centers continue to be a key interaction channel, there is a growing need for advanced forecasting techniques that can handle the inherent variability and seasonality in call data.

Machine learning has emerged as a powerful tool for time series forecasting, offering the ability to model complex, non-linear relationships and capture intricate temporal dependencies. Among the various machine learning methods, Long Short-Term Memory networks and Random Forest Regression are particularly noteworthy for their unique strengths in time series prediction.

LSTM networks, a type of recurrent neural network (RNN), are designed to capture long-term dependencies and sequential patterns in data. Their ability to remember information over extended periods makes them well-suited for modeling time series data with significant temporal dynamics [2, 3]. LSTMs have been shown to outperform traditional statistical methods in capturing complex patterns and trends, especially in cases with strong seasonality and non-linear relationships [4, 5, 6].

On the other hand, Random Forest Regression is an ensemble learning technique that aggregates multiple decision trees to improve prediction accuracy and robustness. Random Forest Regression excels at handling both linear and non-linear relationships and is known for its ability to manage high-dimensional data and prevent overfitting [7, 8]. Its flexibility and ease of implementation make it a popular choice for various forecasting tasks, including those involving time series data [9-12].

This study aims to explore and compare the performance of Long Short-Term Memory and Random Forest Regression in forecasting inbound call volumes for call centers. The comparison is driven by the need to evaluate how each model's strengths contribute to accurate predictions in this specific context. By assessing their performance based on Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2) metrics, we seek to provide actionable insights into the most effective forecasting methods for call center operations [13-16].

2. Problem statement

In the dynamic environment of call center operations, accurately predicting inbound call volumes is essential for efficient resource management and maintaining customer satisfaction. Traditional forecasting methods often struggle to account for the inherent variability and complexity in call volume data, leading to challenges in optimizing agent scheduling and controlling operational costs.

Despite the potential of machine learning models like Random Forest Regression and Long Short-Term Memory networks to address these challenges, their application in the specific context of call center call volume forecasting remains insufficiently explored. There is a critical need to determine the effectiveness of these models in capturing the intricate temporal patterns in call volume data, particularly when forecasting across different time scales, such as daily and half-hour intervals.

This study seeks to address the following key research questions:

- How do Random Forest Regression and LSTM models compare in terms of accuracy and reliability when forecasting inbound call volumes in a call center?
- What are the specific conditions under which each model performs optimally?
- How do different time scales (daily vs. half-hour intervals) and feature engineering strategies impact the performance of these models?

By answering these questions, this research aims to provide a deeper understanding of how machine learning can be effectively leveraged to improve call center operations through more precise call volume forecasting.

3. Methodology

This study employs two machine learning models—Random Forest Regression and Long Short-Term Memory networks—to forecast inbound call volumes in a call center. The methods are applied to two

datasets: one aggregated by day, capturing total daily call volumes, and the other segmented by 30-minute intervals, providing a finer granularity of call activity throughout the day. The methodological approach is divided into several key stages.

3.1. Data preprocessing and feature engineering

The dataset used in this study spans from April 2022 to August 2024 and includes detailed records of inbound call volumes. The datasets were first examined for missing values and outliers. Any anomalies were either corrected or removed to ensure data quality. An anomaly detection step was performed using the Isolation Forest algorithm. This method is particularly effective in identifying outliers in high-dimensional data. The model was trained on the scaled data with a contamination rate set to 0.001, meaning that 0.1% of the data points were considered as anomalies. These identified anomalies were then removed from the dataset, resulting in a cleaner dataset for subsequent modeling.

The plot illustrates the identified anomalies (marked as red points) in the dataset, which were detected using the Isolation Forest algorithm with a contamination rate of 0.001.

To enhance the predictive performance of the machine learning models, we employed a variety of feature engineering techniques based on the available timestamped call data. The raw data provided timestamps for each inbound call, which we transformed into meaningful features aimed at capturing both temporal patterns and the operational dynamics of the call center.
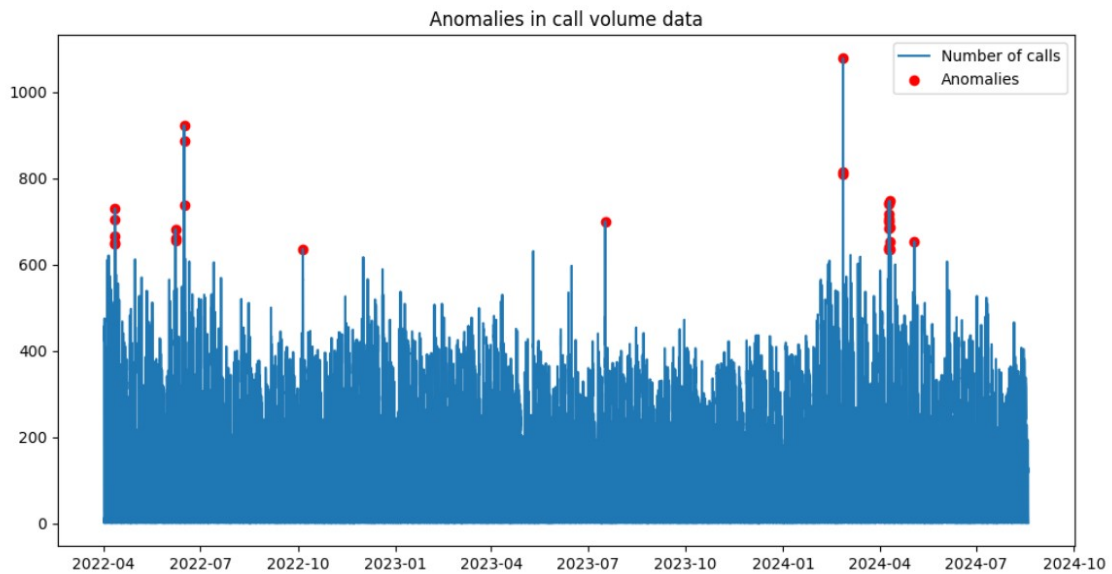
Figure 1: Detection of Anomalies using Isolation Forest.

First, we generated time-based features by extracting key temporal information from the timestamp data. Each call record was assigned its corresponding half-hour of the day (e.g., 9:30, 10:00) and the day of the week (e.g., Monday, Tuesday), thereby allowing the model to account for daily and weekly patterns. We also introduced binary indicators to signal specific periods of interest. For instance, a binary variable was created to mark whether a call occurred during daytime half-hours (8:00 AM to 9:30 PM), which typically see higher call volumes. Similarly, we defined variables for lunch hours (12:00 PM to 1:00 PM) and work hours (8:00 AM to 6:00 PM) to capture fluctuations related to these timeframes.

Categorical features, such as the day of the week, were one-hot encoded to ensure that the model did not mistakenly treat these categories as having an inherent order. This encoding technique

created separate binary variables for each day, allowing the model to capture variations in call volume across different days without assuming any linear relationship between them.

In addition to these temporal features, we incorporated lag features to account for the influence of past call volumes on future predictions. These lag features represent call volumes at previous time steps, and we generated lags from one to ten time steps prior. This allowed the model to recognize autocorrelations within the data, such as whether a high call volume in the previous period would lead to a similar or opposite trend in the next period.

To further capture patterns and trends in the call volume data, we calculated rolling statistics. Rolling mean and rolling standard deviation were computed over various time windows to smooth out short-term fluctuations and reveal longer-term trends. For instance, the rolling mean over the past 8 steps provided the model with information about the general trend in call volume, while the rolling standard over the past 4 step deviation highlighted the variability within those windows.

By thoroughly transforming the timestamp data into informative features, we aimed to improve the model's ability to predict future call volumes by leveraging both short-term dependencies and longer-term trends.

Table 1
Features for Daily and Half-Hourly Call Volume Forecasting Datasets

Feature Type	Dataset Type	Description
Lag Features	Daily and Half-Hourly	Previous call volumes at lagged time steps
Rolling Statistics	Daily and Half-Hourly	Measures of variability and trend over rolling windows
Categorical Variables	Daily and Half-Hourly	Features representing categorical aspects such as time of day, day of week, etc.

3.2. Model training and evaluation

Random Forest Regression model was chosen for its robustness and ability to capture nonlinear relationships in the data. The model was trained using the preprocessed feature set. Hyperparameters such as the number of trees, maximum depth, and minimum samples per split were optimized using RandomizedSearchCV. These hyperparameters were selected to balance model complexity and performance, ensuring that the model generalizes well to unseen data.

LSTM networks were chosen for their strength in modeling sequential data and capturing long-term dependencies, which are crucial for accurate time series forecasting. The LSTM model includes multiple LSTM layers to capture temporal dependencies, followed by dense layers to map the output to the target variable. The model architecture is as follows:

- Input layer with shape matching the feature dimensions;
- The model architecture includes multiple LSTM layers, with a dropout layer to prevent overfitting;
- Dense output layer with a single unit for regression.

Key hyperparameters such as the number of LSTM units, learning rate, and batch size were tuned. Early stopping was used to monitor validation loss and prevent overfitting.

Both models were evaluated using cross-validation to ensure generalizability. The data was split into training and validation sets, with the models trained on the training set and evaluated on the validation set.

The models were assessed using metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) to measure their predictive accuracy and explainability.

A comparative analysis was conducted to evaluate the performance of the Random Forest and LSTM models across different time scales (daily vs. 30-minute intervals) and feature sets. This analysis aimed to identify the conditions under which each model performs best.

3.3. Implementation and tools

The models were implemented using Python, with libraries such as Scikit-learn for the Random Forest model and TensorFlow/Keras for the LSTM model. Data scaling and anomaly detection were performed using StandardScaler and IsolationForest, respectively.

The training and evaluation processes were conducted in a high-performance computing environment to handle the computational demands of model training, particularly for the LSTM network.

Through this methodological approach, the study seeks to provide a robust comparison between traditional ensemble methods and deep learning techniques for forecasting inbound call volumes, offering insights into the most effective strategies for call center management.

4. Results and discussion

4.1. Model Performance

The Random Forest model demonstrated strong performance across both daily and half-hourly datasets. Using the optimized hyperparameters, the model achieved an average Mean Absolute Error (MAE) of 481 for the daily dataset and 14 for the half-hourly dataset. The Root Mean Squared Error (RMSE) was 674 for daily and 26 for half-hourly data. The R-squared (R^2) values were 0.8958 for daily and 0.9686 for half-hourly data, indicating a good fit for both datasets.

The LSTM model, leveraging its capacity to capture long-term dependencies in sequential data, also performed well. The average MAE was 387 for the daily dataset and 13 for the half-hourly dataset. The RMSE was 595 for daily and 23 for half-hourly data. The R-squared (R^2) values were 0.9350 for daily and 0.9720 for half-hourly data. The LSTM model showed superior performance in capturing temporal patterns and trends compared to the Random Forest model.

Table 2
Results of model predictions

| Metric	Random Forest (Daily)	LSTM (Daily)	Random Forest (Half-Hourly)	LSTM (Half-Hourly)
Mean Absolute Error (MAE)	481	387	14	13
Root Mean Squared Error (RMSE)	674	595	26	23
R-squared (R^2)	0.8958	0.9350	0.9686	0.9720

The LSTM model's training dynamics, visualized through loss graphs, indicated efficient learning and convergence, highlighting the model's ability to adapt to sequential data over epochs. This reinforces the LSTM model's suitability for time series forecasting.
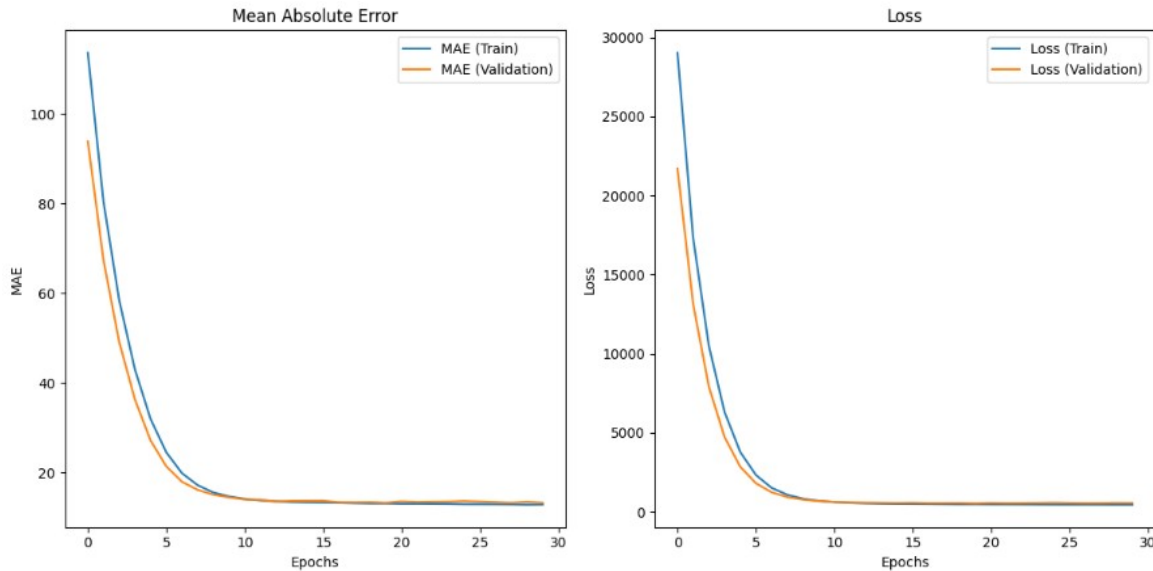
Figure 2: Dynamics of LSTM model training on half-hourly data.

4.2. Comparison across time scales

The comparative analysis revealed that both models performed better with the half-hourly dataset compared to the daily dataset. This improvement is likely due to the finer granularity of the half-hourly data, which allows the models to capture more detailed patterns and trends in call volumes. The LSTM model, in particular, showed a significant advantage in handling the higher resolution of the half-hourly data, reflecting its strength in modeling sequential dependencies.

Figure 3 illustrates the predicted vs. actual call volumes for both models, demonstrating the models' effectiveness in capturing call volume trends.
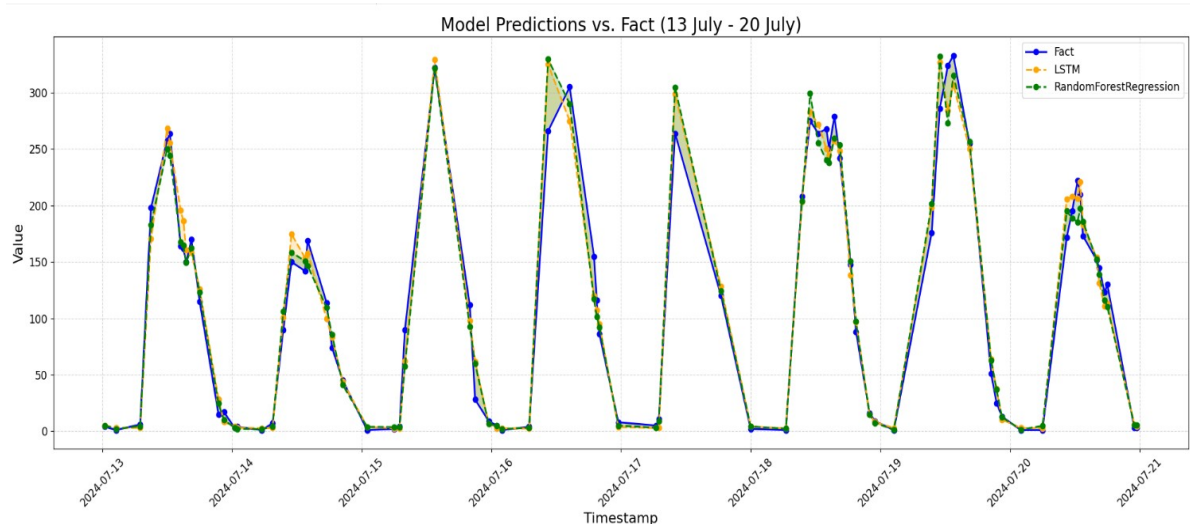
Figure 3: Predicted vs. actual call volumes for both models in the interval from July 13, 2024 to July 20, 2024.

This study highlights the strengths and limitations of Random Forest Regression and LSTM networks in forecasting inbound call volumes. The results indicate that both models offer valuable insights, with LSTM networks showing particular strength in handling sequential data and capturing long-term dependencies. The findings suggest that a combination of these models, along with continuous refinement of features and parameters, can provide robust forecasting solutions for call center operations.
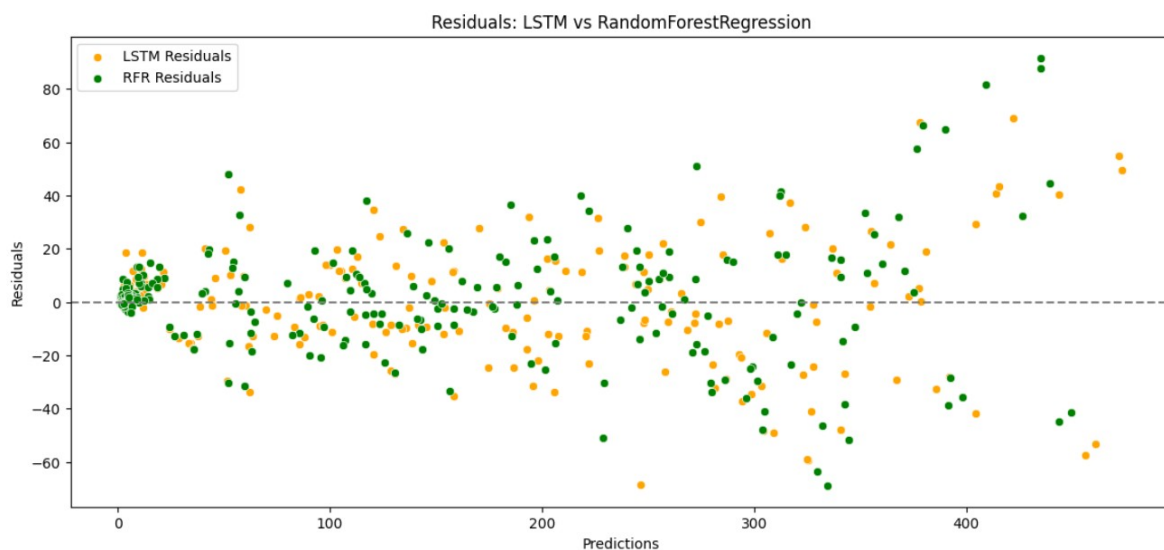
Figure 4: Residuals Distribution Across Predicted Values for Random Forest Regression and LSTM.

The residuals plot displays the difference between actual and predicted values against the predicted values for both Random Forest Regression and Long Short-Term Memory models. It shows a funnel-shaped pattern where residuals widen with higher predicted values, indicating that Random Forest Regression tends to have larger errors for higher call volumes, while LSTM maintains tighter residuals and more consistent predictions. This suggests that LSTM may be more reliable for forecasting larger call volumes compared to Random Forest Regression. Both models perform similarly well for smaller predicted values.

4.3. Feature importance

The feature importance analysis for the Random Forest model indicated that lag features (e.g., Lag-1, Lag-2) and rolling statistics (e.g., Standard Deviation, Rolling Mean) were the most influential in predicting call volumes. This aligns with the expectation that recent call volumes and historical trends play a crucial role in forecasting.

For the LSTM model, the impact of features was less straightforward due to the model's ability to learn complex temporal patterns. However, lagged features and rolling statistics were still important, as they provided essential context for the LSTM's sequential processing.

4.4. Model limitations

While the Random Forest model performed well, it is limited by its inability to capture very long-term dependencies due to its non-sequential nature. Additionally, the model's performance can degrade if the feature set does not adequately capture all relevant temporal patterns.

The LSTM model, despite its strengths, requires significant computational resources and can be sensitive to hyperparameter settings. The performance of the LSTM model also depends on the quality and granularity of the input features.

4.5. Practical implications

The insights from this study have been applied in a call center setting, where machine learning models are now used to forecast call volumes and optimize staffing levels. While the initial implementation has shown promise in improving staffing efficiency and reducing operational costs, we are still in the process of fully evaluating the impact on service levels and customer satisfaction. Early results suggest that these models are aligning staffing more closely with expected call activity, potentially leading to reduced wait times and more effective resource management.

Future enhancements could involve integrating additional factors into the forecasting models, such as external events, promotional campaigns, and technical anomalies. Addressing these aspects could further refine the accuracy of the predictions and improve overall model performance. Exploring other advanced machine learning techniques or hybrid models may also offer further improvements in forecasting capabilities, providing even more robust solutions for call center management and beyond.

5. Conclusion

This study evaluated the effectiveness of Random Forest Regression and Long Short-Term Memory models for forecasting inbound call volumes in call centers. The analysis was conducted using both daily and half-hourly datasets to determine the models' performance across different time scales.

The LSTM model outperformed the Random Forest model in forecasting call volumes. Specifically, the LSTM achieved lower Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), and higher R-squared (R^2) values, demonstrating its superior ability to capture long-term dependencies and sequential patterns in time series data.

The training loss graph for the LSTM model illustrated efficient learning and convergence, reinforcing its suitability for time series forecasting.

While both models showed strong performance, the LSTM's ability to model temporal trends and dependencies provided a significant advantage over the Random Forest model, especially for half-hourly data.

Accurate call volume forecasting is essential for effective call center management. The results suggest that LSTM models can enhance staffing decisions and operational efficiency by providing more precise predictions compared to traditional models like Random Forest Regression.

Further research could explore additional machine learning techniques and integrate external factors to improve forecasting accuracy. Expanding the scope to include varied time series characteristics may also offer deeper insights into call volume prediction.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] Ibrahim, R., Ye, H., L'Ecuyer, P., & Shen, H. (2016). Modeling and forecasting call center arrivals: A literature survey and a case study. International Journal of Forecasting, 32(3), 865–874. doi:10.1016/j.ijforecast.2015.11.012.

[2] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735-1780. doi: 10.1162/neco.1997.9.8.1735.

[3] Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: continual prediction with LSTM. Neural Computation, 12(10), 2451-2471. doi: 10.1162/089976600300015015.

[4] Kumar, I., Tripathi, B. K., & Singh, A. (2023). Attention-based LSTM network-assisted time series forecasting models for petroleum production. Engineering Applications of Artificial Intelligence, 123, 106440. doi: 10.1016/j.engappai.2023.106440.

[5] Graves, A., & Schmidhuber, J. (2005). Framewise Phoneme Classification with Bidirectional LSTM Networks. IEEE Transactions on Neural Networks, 14(5), 993-998.
doi: 10.1109/IJCNN.2005.1556215.

[6] Lindemann, B., Müller, T., Vietz, H., Jazdi, N., & Weyrich, M. (2021). A survey on long short-term memory networks for time series prediction. Procedia CIRP, 99, 650-655.
doi: 10.1016/j.procir.2021.03.088.

[7] Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32.

[8] Liaw, A., & Wiener, M. (2002). Classification and Regression by RandomForest. R News, 2(3), 18-22.

[9] Zhang, G., & Qi, M. (2005). Neural Network Forecasting of Stock Market Time Series. European Journal of Operational Research, 160(2), 570-584. doi: 10.1016/j.ejor.2003.08.037.

[10] Lachaud, A., Adam, M.,, & Miskovic, I. (2023). Comparative Study of Random Forest and Support Vector Machine Algorithms in Mineral Prospectivity Mapping with Limited Training Data. Minerals, 13(8), 1073.

[11] Nurtas, M., Zhantaev, Z., Altaibek, A. (2024), Earthquake time-series forecast in Kazakhstan territory: Forecasting accuracy with SARIMAX. Procedia Computer Science, 231, 353-358.doi: 10.1016/j.procs.2023.12.216.

[12] Aizhan, A., Tokhtakhunov, Il',.. & Nurtas, M. (2024). The Efficacy of Autoencoders in the Utilization of Tabular Data for Classification Tasks. Procedia Computer Science, 238.
doi: 10.1016/j.procs.2024.06.052.

[13] Nurtas, M., Zhantaev, Z., Altaibek, A., Nurakynov, S., Mekebayev, N., Shiyapov, K., Iskakov, B., Ydyrys, A. (2023). Predicting the Likelihood of an Earthquake by Leveraging Volumetric Statistical Data through Machine Learning Techniques. Engineered Science, 26 (1031). doi: 10.30919/es1031.

[14] Nurtas M., Ydyrys A., Altaibek A. (2020). Using of Machine Learning algorithm and Spectral method for simulation of Nonlinear Wave Equation. ICEMIS'20: Proceedings of the 6th

International Conference on Engineering & MIS 2020 doi: https://doi.org/10.1145/3410352.3410778.

[15] Nurtas M., Baishemirov Zh., Zhanabekov Zh. (2020). Convalutional Neural Networks as a method to solve estimation problem of acoustic wave propogation in poroelastic media, News of the National Academy of sciences of the Republic of Kazakhstan. 4(332).
doi: https://doi.org/10.32014/2020.2518-1726.65.

[16] Nurtas M., Baishemirov Zh., Zhanabekov Zh. (2020). Applying Neural Network for predicting cardiovascular disase risk, News of the National Academy of sciences of the Republic of Kazakhstan.4(332). Doi: https://doi.org/10.32014/2020.2518-1726.62.