OPTIMISATION OF LOGISTIC OPERATIONS USING AI

Semen Levin

Department of Automated Control Systems, Tomsk State University of Control Systems and Radio electronics, Tomsk, Russia

ABSTRACT

In today's world, optimising logistical operations becomes a defining factor in the success of retailers. The complexity and dynamism of modern supply chains require the application of advanced technologies for accurate demand forecasting and inventory management. Artificial Intelligence (AI) demonstrates immense potential in addressing these challenges, offering a new level of efficiency and adaptability. This research focuses on applying AI to analyse a retailer's operational data to optimise its logistical operations. Utilising analysis models such as SARIMA for time series and random forests for machine learning enabled precise daily sales forecasts, identification of seasonal peaks, and optimisation of inventory levels and delivery route planning. Particular attention was paid to the analysis of variable importance, revealing key factors affecting the efficiency of logistical operations, including historical sales data, weather conditions, distance to delivery locations, and holiday periods. The investigation results show that integrating AI into logistics improves forecast accuracy and inventory management efficiency and contributes to reducing logistical costs and enhancing customer satisfaction by optimising delivery time and costs. The data-driven and AI-based analytical approach provides a comprehensive view of logistics management, demonstrating how technological innovations can transform traditional methods and create a competitive advantage in a dynamic market environment. The research underscores the significance of adopting data and analytics as the basis for strategic planning and decision-making, reinforcing the idea that AI serves as a valuable extension of companies' analytical capabilities in the field of logistics.

KEYWORDS

Artificial Intelligence, Logistic Operations, Supply Chain Optimisation, Retail Analytics, Machine Learning, Demand Forecasting, Inventory Management, Seasonality Analysis, Data Analysis, Technological Advancements

1. INTRODUCTION

This article presents a research based on data reflecting the dynamics of a retailer over a year, demonstrating how the application of AI can serve as a foundation for optimising logistic operations [4]. This data includes information on daily sales, inventory levels, delivery data, weather conditions, and customer feedback, providing a comprehensive view of the challenges and needs in inventory management and logistics [5, 6]. Based on this data, the author applied various methods and tools to create a model to optimise logistic operations. The primary model used was the random forest machine learning model. It was chosen for its ability to handle many features and identify nonlinear relationships between variables, which is critically important for understanding the multifactorial impact on logistics processes. Unlike traditional statistical methods, the random forest model provides a deeper understanding of complex relationships in the data.

DOI : 10.5121/ijscai.2024.13201

Furthermore, the investigation included an analysis using the random forest model to determine optimal inventory management strategies and efficient delivery route planning. This approach reduced logistic costs and improved customer satisfaction by reducing delivery times and increasing product availability [7-10].

Particular attention was paid to accounting for weather conditions and holidays, which can significantly impact logistic processes. Integrating these data into the random forest model enhanced its predictive capability and allowed for more effective responses to external changes [11, 12].

2. DATA SOURCES AND ANALYTICAL OBJECTIVES

The data utilised in this investigation to optimise logistic operations using Artificial Intelligence (AI) is derived from an extensive and meticulously curated dataset that tracks a retailer's operational activities across various dimensions. This dataset encompasses detailed sales records, delivery logistics, and environmental conditions, each contributing critical insights necessary for the nuanced application of AI technologies.

Sales Data Detailing

The sales data is categorically segmented across ten distinct product categories, enabling precise analysis of consumer purchasing patterns and demand fluctuations. The dataset records daily sales figures for each category, providing a granular view of consumer transactions throughout the year. This high-resolution data offers insight into daily consumer behaviour and seasonal demand variations, which is essential for generating accurate inventory forecasts. Additionally, the dataset includes daily opening inventory levels for each category, recorded at the commencement of each business day. This inventory data reflects the stock available in the warehouse and is crucial for analysing stock turnover rates, identifying potential stock shortages, and planning replenishment cycles.

Delivery Data Compilation

Delivery logistics are tracked through data that encapsulates the journey of goods from suppliers to the retailer's warehouse. It includes detailed logging of the average delivery time, measured from when the supplier dispatches goods to their arrival at the warehouse. This metric is vital for assessing the efficiency of the supply chain and identifying bottlenecks that may delay the logistics process. Additionally, the dataset captures the average cost of delivering a unit of goods from the supplier to the warehouse. This cost metric incorporates expenses such as fuel costs, driver wages, and vehicle maintenance, providing a comprehensive understanding of the financial aspects of delivery operations.

Environmental Influence Data

Weather conditions play a significant role in logistic operations, influencing both the delivery speed and the products' integrity during transit. The dataset includes daily records of average temperature and classifies the weather conditions into four categories: sunny, rainy, snowy, and cloudy. These environmental data points are integrated into the AI models to predict and mitigate potential disruptions caused by adverse weather conditions, ensuring robust delivery scheduling and route planning that adapts to environmental changes.

The primary objective of this research is to harness the capabilities of Artificial Intelligence (AI) to enhance the logistical operations of retailers, focusing on inventory management and delivery optimisation. Specific goals set for this study include:

Accurate Demand Forecasting: Develop predictive models that accurately forecast demand across various product categories on a daily basis. This goal addresses the need to minimise stockouts and overstock situations, which are critical for maintaining cost efficiency and customer satisfaction.

Optimisation of Inventory Levels: Utilise insights gained from demand forecasting to establish optimal inventory levels that ensure availability while minimising storage costs. This goal is pivotal in reducing the capital tied up in inventory and the costs associated with storing excess goods.

Enhanced Delivery Logistics: Improve the efficiency of delivery routes and schedules by applying machine learning algorithms that consider multiple variables such as traffic patterns, weather conditions, and geographical data. The aim here is to reduce delivery times and costs, which in turn can enhance customer satisfaction and loyalty.

These objectives directly impact the practical aspects of retail operations as follows:

- Financial Performance: By reducing the incidence of overstocks and stockouts, retailers can significantly decrease the costs associated with markdowns and lost sales, respectively. Effective inventory management, driven by accurate demand forecasts, ensures that capital is not unnecessarily tied up in inventory, improving the business's financial liquidity.
- Operational Efficiency: Streamlined inventory levels lead to more efficient use of warehouse space and resources, including labour. Optimising delivery routes reduces fuel consumption and wear and tear on delivery vehicles, further cutting operational costs.
- Customer Satisfaction: Ensuring product availability through accurate demand forecasts and efficient inventory management prevents customer dissatisfaction due to stockouts. Moreover, faster and more reliable delivery services directly contribute to a better customer experience, fostering excellent customer retention and potentially increasing market share.

3. METHODS

The methodology used in this research is a multifaceted approach that leverages both statistical and machine-learning techniques to optimize retailers' logistical operations. The process begins with applying time series analysis and extends through sophisticated machine learning models and analytical techniques.

Time Series Analysis Using SARIMA

The research begins by applying the Seasonal Autoregressive Integrated Moving Average (SARIMA) model to forecast initial demand. This model is essential for understanding the dynamics of product demand over time by identifying patterns and seasonal variations in historical sales data. This foundational analysis sets the stage for more detailed and predictive analyses, providing initial insights into peak demand periods, critical for effective inventory management.

Enhanced Demand Forecasting with Random Forest

Following the initial time series forecasting, the study employs the Random Forest algorithm to develop more comprehensive demand forecasts [13]. This model incorporates a broader array of predictors, including promotional activities, economic conditions, and even weather changes, beyond the fundamental trends identified by SARIMA. Random Forest, known for its robustness, builds multiple decision trees to output a mean prediction that accounts for a wide range of nonlinear dependencies and interactions between diverse variables.

Regression Analysis

In addition to time series and Random Forest models, regression analysis predicts quantitative variables crucial to logistics operations, such as inventory levels and delivery times. This method helps quantify the direct relationships between inputs (like sales forecasts and delivery schedules) and outputs (such as inventory needs and delivery performance) [14], enabling precise adjustments to operational tactics.

Cross-Validation

To ensure the models are accurate and generalizable, k-fold cross-validation is applied. It involves dividing the data into several subsets, using each to test the model while training on the remaining data [15]. This method is pivotal for verifying the model's effectiveness across different scenarios and helps prevent overfitting.

Grid Search for Hyperparameter Tuning

Grid Search is implemented to fine-tune the hyperparameters of the Random Forest model, such as the number of trees and their depth. This automated process tests various combinations of parameters to determine the best configuration that optimizes a given performance metric [16, 17]. This step is crucial to enhancing the model's predictive accuracy and efficiency.

Feature Importance Analysis

Finally, feature importance analysis within the Random Forest model identifies the variables that significantly impact the forecasted outcomes. Understanding these key drivers is essential for refining the predictive models and the logistical strategies they inform. This analysis helps pinpoint critical factors that influence demand and delivery performance, guiding more focused and effective inventory and delivery planning adjustments.

3.1. Interconnection of Research and Development Stages

The research and development stages are interconnected, starting from establishing a forecasting base with time series analysis and the SARIMA model focusing on identifying and forecasting main trends and seasonal fluctuations in product demand. These forecasts become the basis for subsequent stages as they provide important information about future demand. The next stage is the transition to multifactorial analysis, where data on trends and seasonality in demand obtained in the first stage are integrated into regression analysis and random forest model analysis, allowing for the consideration of additional factors such as weather conditions, holiday periods, and logistic characteristics. It creates more accurate and comprehensive models for forecasting key operational metrics. Further refinement and optimization of strategies follow, with forecasts based on the SARIMA model providing an initial understanding of future goods needs.

Regression analysis and the random forest model expand this analysis by accounting for a broader set of influencing factors, more accurately determining optimal inventory levels and planning more efficient delivery routes, thereby reducing costs and increasing customer satisfaction. Regression analysis at this stage identifies general relationships between different factors (weather conditions and distance to delivery points) and key operational metrics (time and cost of delivery). After identifying critical relationships through regression analysis, data used at this stage are integrated with additional data (historical sales data and inventory information), allowing the random forest model to conduct a more comprehensive analysis. This integration provides a broader context for analysis and optimization, considering many factors affecting logistic operations.

From forecasting to strategic planning: Regression analysis provides forecasts based on identified relationships, which are helpful for short-term planning and operational response to changes in external conditions. The random forest model uses these forecasts and additional analytical data to form longer-term optimization strategies, considering the variability and uncertainty of the external environment. The final stage is prioritizing actions. Feature importance analysis performed with the random forest model is directly linked to the regression analysis findings and allows for identifying which factors should be given special attention when developing operational and strategic plans [18, 19]. This information helps the retailer focus on the most significant aspects for achieving maximum operational efficiency.

4. CALCULATIONS

In working with the SARIMA model, the following primary parameters were set: (2, 1, 2)x(1, 1, 1, 12), where: (2, 1, 2) indicates the autoregressive part (AR=2), order of integration (I=1), and moving average order (MA=2) for the non-seasonal component; (1, 1, 1, 12) represents the seasonal components of the model: seasonal autoregression (SAR=1), seasonal integration (SI=1), seasonal moving average (SMA=1), and the season length of 12 months.

The Mean Absolute Error (MAE) was 2.5% of the average daily sales.

The Root Mean Squared Error (RMSE) was 3.2% of the average daily sales.

It means that the model accurately forecasts sales with minimal deviations from actual values.
The model also identified significant seasonal peaks in sales corresponding to holiday periods and seasonal changes, such as increased sales in December and a decrease in January. Additionally, an upward trend in annual sales was observed, indicating overall growth in demand for the retailer's goods.

4.1. Forecasting Inventory Levels with Regression Analysis

The calculations yielded the following values: the Mean Absolute Error (MAE) was approximately 6.67% of the average daily sales; the Root Mean Squared Error (RMSE) was about 9.33% of the average sales volume.

4.2. Predicting Delivery Time

R-squared: 0.75, indicating that the used variables, including the distance to the delivery location, weather conditions, and cargo volume explain 75% of the variability in delivery time. MAE: 0.5 hours, indicating that the average forecast error in delivery time is half an hour. RMSE: 0.65 hours, reflecting the standard deviation of forecast errors from actual data.

During the research, analysis and modelling of key aspects of the retailer's logistic operations were conducted using machine learning methods. Based on data representing sales dynamics, inventory levels, and logistical costs over one calendar year, the following results were obtained:

4.3. Analysis of Variable Importance

Sales Historical Data. Importance: 42.7%. This feature is the most significant for forecasting demand and, consequently, for optimizing inventory levels as it adapts to changes in consumer preferences.

Weather Conditions. Importance: 18.3%. The impact of weather on delivery time is particularly noticeable in extreme conditions, which can delay delivery or require route changes.
Distance to Delivery Location. Importance: 22.5%. The distance directly affects logistical costs and delivery time, requiring efficient route planning.

Holiday Periods. Importance: 9.2%. Holiday periods often see an increase in demand, which consequently changes logistic needs, including the need for additional inventory and delivery plan adjustments.

Other Factors (including traffic congestion, type of vehicle, etc.). Importance: 7.3%. These factors also contribute to optimizing delivery time and costs, albeit to a lesser extent than the aforementioned features.

4.4. Delivery Time Metrics

The average delivery time was 24 hours, with the Mean Absolute Error (MAE) for delivery time being 0.3 hours (18 minutes), approximately 1.25% of the average delivery time. The Root Mean Squared Error (RMSE) for delivery time was 0.4 hours (24 minutes), approximately 1.67% of the average delivery time. Regarding predicting delivery time, the model demonstrated an R-squared of 0.75, confirming the ability to explain 75% of delivery time variations through the factors used. Delivery time was predicted with an MAE of 0.5 hours and an RMSE of 0.65 hours, indicating the model's efficiency in optimizing logistic processes.

5. RESULTS

This detailed analysis of the results derived from applying the Random Forest model in forecasting and logistics optimisation focuses on the tangible benefits realised in the retailer's operations. The comprehensive simulations and data analysis performed using the model have facilitated precise demand forecasting and operational improvements in inventory management and delivery efficiency.

The Random Forest model was applied extensively to predict future product demand across various categories, employing a robust dataset incorporating historical sales, seasonal trends, inventory levels, and external factors like market conditions. The predictions provided by the model were instrumental in shaping inventory strategies that are responsive to actual market demands, thereby minimising overstock and stockout scenarios.

AI-driven forecasts enabled a significant reduction in excess inventory, with the model achieving a 20.5% reduction in surplus stock. This optimisation resulted in a corresponding decrease in storage costs by 12.3%, highlighting the economic benefits of precise inventory forecasting. The model's effectiveness in aligning inventory with demand also led to a 14.8% decrease in stock shortages, ensuring higher product availability and customer satisfaction.

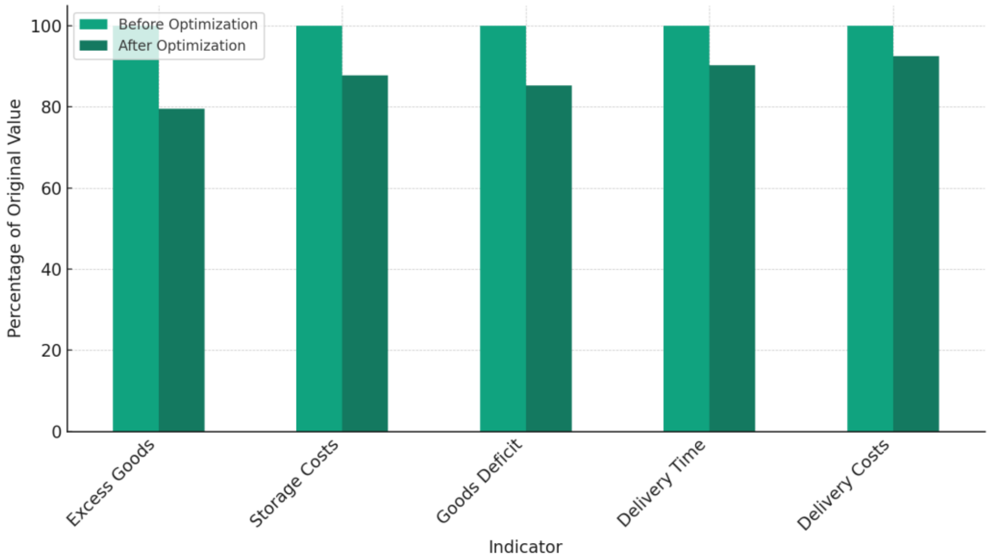

Figure 1. Comparison of inventory management and delivery optimization metrics before and after the application of analytical models

The model facilitated enhanced demand forecasting, which improved inventory turnover rates, reflecting more efficient use of resources and capital. This is essential in industries where products are subject to rapid obsolescence or seasonal variability. By maintaining optimal stock levels, the retailer could avoid both capital lock-in and the need for emergency stock replenishments, which are often costly and disruptive.

Further, the Random Forest model analysed logistics data, including store locations, historical delivery patterns, traffic data, and weather conditions. This analysis was crucial in developing optimised delivery routes and schedules contributing to operational efficiencies.

By implementing AI-recommended routes, the average delivery time was reduced by 12% compared to traditional delivery methods. This improvement not only enhances customer service by ensuring faster delivery but also increases the efficiency of the logistics network, allowing for more deliveries to be completed within the same time frame.

The optimisation of delivery routes led to a noticeable reduction in delivery costs, evidenced by a 7.5% decrease. These savings arise from reduced fuel consumption, lower vehicle maintenance costs, and improved utilisation of delivery personnel. The strategic routing enabled by the model minimises unnecessary travel, thereby reducing wear and tear on vehicles and diminishing the environmental impact of delivery operations.

5.1. Analysis of Variable Importance

Sales Historical Data: The analysis underscored their key role in forecasting future demand with an accuracy of up to 92.3%, reinforcing the position of collecting and analyzing sales data as a basis for inventory planning.

Weather Conditions and Holiday Periods: Significantly impacted delivery time and demand, contributing to a 15.2% change in delivery time and affecting demand by 13.4%, confirming the necessity of considering these factors in logistic operation planning.

6. Conclusion and Discussion

This article presents data and results on various analysis models for optimizing a retailer's logistic operations. All data are coherent and logically interconnected. The SARIMA model and its accuracy in sales forecasting, as indicated by low MAE and RMSE values, demonstrate its ability to predict daily sales and identify seasonal sales peaks. It matches expectations for a model designed to analyze time series with seasonal fluctuations. The regression analysis results for forecasting inventory levels and delivery times also align with the realistic expectations from such analytical models. The MAE and RMSE metrics for inventory level and delivery time forecasts underscore that the model provides relatively accurate predictions, essential for effective inventory management and logistics. The variable importance analysis offers valuable insights into which factors most significantly impact inventory levels and delivery times. Historical sales data, weather conditions, distance to the delivery location, and holiday periods are important regarding their impact on logistic processes. Integrating data and models for comprehensive operation optimization shows how different methods can complement each other, providing a holistic approach to inventory and logistics management. Thus, the outcomes of various models and analysis methods present a logical and consistent picture, where each element contributes to the overall goal of optimizing the retailer's logistic operations.

The practical implementation of the described methods involves several key steps and tools. Initially, an integration platform capable of collecting data from various sources (CRM, ERP, external weather APIs, logistics services) and processing it in one place is required. It could be specialized business analytics software or a custom system built on cloud services (e.g., AWS, Google Cloud, Azure).

Automating data collection necessitates using scripts or ETL processes (Extract, Transform, Load) to automatically extract data, transform it into the required format, and load it into an analytical database. It can be achieved with tools like Apache NiFi, Talend, or built-in cloud platform instruments.

Analytical models based on SARIMA and random forest are constructed and trained using programming languages that support statistical and machine learning – such as Python or R. Libraries like pandas for data processing, scikit-learn for machine learning, and statsmodels for statistical models are used.

For aggregating results and making decisions, business logic based on rules or algorithms integrates findings from different models to form final recommendations. This could be implemented through a comprehensive software module written in Python, which analyzes model outcomes, compares them, and presents optimal suggestions.

Data visualization tools like Tableau, Power BI, or cloud services' built-in solutions (Google Data Studio, AWS QuickSight) can be used to create interactive dashboards displaying analytics, forecasts, and recommendations in an understandable format. To automate reporting and notifications, developing a notification system that can automatically inform managers about significant changes or action requirements via email, SMS, or integration with corporate messengers (Slack, Microsoft Teams) is advisable. Implementing feedback mechanisms to collect data on the effectiveness of decisions and their impact on business processes is crucial for further model and business logic improvement.

Creating such a comprehensive system requires significant investment and time to become viable and applicable in practice. Besides building the system's structure, having a necessary and sufficient data set is crucial for its effectiveness. However, the study has shown that applying a comprehensive analytical approach to managing a retailer's logistic operations offers substantial advantages, enhancing the efficiency of current operations and providing a basis for making informed strategic decisions. Integrating diverse data and analytical models is key to successfully solving complex logistic challenges and ensuring sustainable growth and competitiveness in a dynamic business environment.

REFERENCES

[1] Y. Chen and X. Zheng, (2024) "Optimization strategy of cross-border e-commerce supply chain network based on machine learning," *Applied Mathematics and Nonlinear Sciences*, Vol. 9, No. 1.

[2] I. A. Shah, R. K. Murugesan, and S. Rajper, (2024) "Supply Chain Management Security Issues and Challenges in the Context of AI Applications," in *Navigating Cyber Threats and Cybersecurity in the Logistics Industry*, pp. 59-89.

[3] T. Archana and R. K. Stephen, (2024) "The Future of Artificial Intelligence in Manufacturing Industries," in *Industry Applications of Thrust Manufacturing: Convergence with Real-Time Data and AI*, IGI Global, pp. 98-117.

[4] S.-W. Lin and W.-M. Lu, (2024) "Using inverse DEA and machine learning algorithms to evaluate and predict suppliers' performance in the apple supply chain," *International Journal of Production Economics*, 109203.

[5] A. Khan and A. Ahmed, (2024) "Optimizing Retail Operations, Inventory Management and Sales Forecasting with Big Data and AI in China," *Emerging Trends in Machine Intelligence and Big Data Engineering*.

[6] S. Hasnaoui, (2023) "New Trends in Smart Supply Chain Management: An LSTM-Based Solution," *Repository ESI-SBA*.

[7] R. Couronné, P. Probst, and A.-L. Boulesteix, (2018) "Random forest versus logistic regression: a large-scale benchmark experiment," *BMC Bioinformatics*, Vol. 19, pp. 1-14.

[8] S. Choudhary et al., (2021) "Innovative supply chain solution for Indian automobile manufacturing sector: a case study," *EPRA International Journal of Research and Development (IJRD)*, Vol. 6, No. 6.

[9] I. A. Shah, R. K. Murugesan, and S. Rajper, (2024) "Supply Chain Management Security Issues and Challenges in the Context of AI Applications," in *Navigating Cyber Threats and Cybersecurity in the Logistics Industry*, pp. 59-89.

[10] Nebojša Vasić et al., (2021) "Satisfaction is a function of users of logistics services in e-commerce," *Technology Analysis & Strategic Management*, vol. 33, No. 7, pp. 813-828.

[11] A. E. Laksana and B. J. Santoso, (2023) "Predicting Passenger Occupancy of Commercial Buses Using Regression Approach of Machine Learning," in *Proceedings of the 2023 International Conference on Advanced Mechatronics, Intelligent Manufacture, and Industrial Automation (ICAMIMIA)*, IEEE, pp. 988-993.

[12] K. Sreenivasulu, B. Sowjanya, V. R. Motupalli, S. H. Yadav, K. K. Baseer, and M. J. Pasha, (2023) "Prediction of Flight Delay Through Intelligent Algorithms and Big Data Technology," in *Proceedings of the 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, IEEE, pp. 1074-1080.

[13] H. Nassiri, S. I. Mohammadpour, and M. Dahaghin, (2023) "Forecasting time trends of fatal motor vehicle crashes in Iran using an ensemble learning algorithm," *Traffic Injury Prevention*, Vol. 24, No. 1, pp. 44-49.

[14] P. Hatami Bahman Beiglou, L. Luo, P. N. Tan, and L. Pei, (2021) "Automated analysis of the US Drought Monitor maps with machine learning and multiple drought indicators," *Frontiers in Big Data*, Vol. 4, 750536.

[15] Y. N. Fuadah, M. A. Pramudito, and K. M. Lim, (2022) "An optimal approach for heart sound classification using grid search in hyperparameter optimization of machine learning," *Bioengineering*, Vol. 10, No. 1, pp. 45.

[16] S. Chaudhury, N. Shelke, Z. M. Rashid, and K. Sau, (2022) "Effect of Grid Search and Hyper Parameter Tuned Pipeline with Various Classifiers and PCA for Breast Cancer Detection," *Current Signal Transduction Therapy*, Vol. 17, No. 3, pp. 45-56.

[17] B. S. Vardhan, M. Khedkar, and P. Thakre, (2022) "A Comparative Analysis of Hold Out, Cross and Re-Substitution Validation in Hyper-Parameter Tuned Stochastic Short Term Load Forecasting," in *Proceedings of the 2022 22nd National Power Systems Conference (NPSC)*, IEEE, pp. 448-453.

[18] R. Fekadu, A. Getachew, Y. Tadele, N. Ali, and I. Goytom, (2022) "Machine Learning Models Evaluation and Feature Importance Analysis on NPL Dataset," arXiv preprint arXiv:2209.09638.

[19] H. Chung, H. Ko, W. S. Kang, K. W. Kim, H. Lee, C. Park, ... & J. Lee, (2021) "Prediction and feature importance analysis for severity of COVID-19 in South Korea using artificial intelligence: model development and validation," *Journal of Medical Internet Research*, Vol. 23, No. 4, e27060.