# A Multi-Model Approach to Supermarket Sales Forecasting System Using Machine Learning and Time Series

Hammaad Rizwan [12#], Aadhavan Arkhash Saravanakumar [12], Sevindu Imsara[12] , Praveen Jayarathnam[12] and Nishantha J Chandrasena[123]

[1] *Informatics Institute of Technology, Sri Lanka*
[2]*Robert Gordon University, Aberdeen, Scotland*
[3] *University of Westminster, London, England*
[#] hammaad.rizwan@iit.ac.lk

***Abstract***— Supermarkets frequently encounter challenges in managing their inventory efficiently, often resulting in significant wastage due to overstocking or understocking. This research aims to enhance supermarket inventory management by implementing a robust sales prediction and forecasting system using advanced machine learning algorithms. The primary objectives of this study are to reduce inventory wastage, optimise stock levels, and increase profitability by accurately predicting product demand based on weather forecasts and economic factors. The proposed solution is highly beneficial for supermarket managers, providing actionable insights for inventory decisions.The system integrates ARIMA (AutoRegressive Integrated Moving Average) and multiple machine learning models, , Randomforest, and XGBoost (Extreme Gradient Boosting), to deliver precise sales forecasts. By combining these models through a stacking ensemble approach, the solution leverages the strengths of each model to enhance overall prediction accuracy. This paper presents the development, implementation, and evaluation of the solution, highlighting its potential to transform inventory management practices in the retail industry. The backend of the system is supported by a comprehensive database, ensuring seamless data processing and real-time updates. The web-based application offers an intuitive interface for users, enabling efficient inventory management. The results demonstrate significant improvements in sales forecasting accuracy, which can lead to reduced wastage and better inventory control in supermarkets.

***Keywords***— Predictive Analytics, Machine Learning, Sales Forecasting, Retail Management, Weather Prediction,Time Series.

## I. INTRODUCTION

Supermarkets face significant challenges in managing their inventory efficiently, often resulting in considerable wastage due to overstocking or understocking. Effective inventory management requires a precise understanding of consumer demand, which is influenced by a variety of factors, including weather conditions and economic indicators such as Gross Domestic Product (GDP).

Weather conditions play a crucial role in influencing consumer purchasing behaviour. For example, during hot weather, there is typically an increased demand for beverages, ice cream, and cooling products, while cold weather boosts the sales of hot beverages, heaters, and winter clothing. Traditional inventory management systems often fail to account for these weather-induced fluctuations, leading to either excess stock or stockouts.

Economic indicators, particularly GDP, also significantly impact consumer spending patterns. During periods of economic growth, consumers tend to spend more, whereas economic downturns lead to reduced spending. Understanding these economic cycles is essential for supermarkets to adjust their inventory levels accordingly. However, most conventional systems do not incorporate real-time economic data, resulting in suboptimal inventory decisions.

This research aims to address these challenges by implementing an advanced predictive analytics solution that leverages machine learning and time series algorithms to enhance sales forecasting and inventory management. By integrating weather forecasts and economic indicators such as GDP, our approach provides actionable insights that help supermarket managers optimize stock levels, reduce wastage, and improve profitability.

This paper outlines the development, implementation, and evaluation of this solution, highlighting its potential to transform inventory management practices in the retail industry. The backend of the system is supported by a comprehensive database, ensuring seamless data processing and real-time updates. The web-based application offers an intuitive interface for users, enabling efficient inventory management..

## II. **LITERATURE REVIEW**

### *A. Predictive Analytics in Inventory Management*

Predictive analytics in inventory management has been extensively researched, with various approaches demonstrating significant improvements in sales forecasting accuracy. Cheriyan et al. (2018) utilized Generalized Linear Models, Decision Trees, and Gradient Boosted Trees to enhance sales predictions, although their model was limited by the scope of attributes considered. Similarly, Pavlyshenko (2019) employed a combination of Extra Trees, ARIMA models, and Neural Networks, showing that a stacking approach could further improve prediction accuracy. However, these studies did not incorporate external factors such as weather conditions, which can significantly impact consumer purchasing behavior. Holmstrom et al. (2016) explored the use of linear and functional regression models for weather forecasting but found these models less accurate compared to professional meteorological services. This highlights the need for more advanced models to effectively integrate weather data into sales forecasting systems. Scher and Messori (2018) used Convolutional Neural Networks to predict weather forecast uncertainty, achieving better results by incorporating a wider range of data inputs, further supporting the integration of weather data into predictive models for inventory management.

### *B. Integrating Weather and Economic Factors in Sales Forecasting*

Integrating economic factors such as GDP into sales prediction models is also crucial for enhancing the accuracy of forecasts. Bajaj et al. (2020) emphasized the importance of economic indicators in sales predictions, using Random Forest Regressor and XGBoost Regressor to predict sales based on various economic factors. They highlighted the significant impact of GDP and exchange rates on consumer purchasing power and behavior, underscoring the need for predictive models to incorporate real-time economic data for optimal inventory management. These studies collectively suggest that integrating weather forecasts and economic indicators into sales prediction models can provide a more comprehensive and accurate understanding of consumer demand, thereby improving inventory management practices in supermarkets.

**Summary of Literature Review:**

Table 1. Font sizes for this publication

| **Paper Title and Year** | **Objective** | **Methodology Used** | **Drawback** |
|---|---|---|---|
| ***Sales prediction using machine learning*** | | | |
| Intelligent sales prediction using machine learning techniques (2018) | Enhance sales forecasting accuracy | Generalized linear model, decision tree, Gradient boosted trees | Did not consider weather impact |
| Machine Learning Models for Sales Time Series Forecasting (2019) | Improve sales prediction accuracy | Extra tree, ARIMA model, Random Forest, Lasso and Neural Networks | Lacked weather and economic factor integration. |
| Machine Learning Model for Sales Forecasting by Using XGBoost (2021) | Predict sales using economic indicators and improve the a | XGBoost | Limited Generalization and Did not integrate weather factors |
| ***Weather forecasting using machine learning*** | | | |
| Weather Forecasting in Sales Prediction (2016) | Enhance weather forecast accuracy | Linear regression, functional regression model | Less accurate than professional services**.** |
| Predicting Weather Forecast Uncertainty with Machine Learning (2018) | Improve weather forecast accuracy | Convolutional Neural Networks | |
| Weather Forecasting Using Machine Learning Algorithm (2019) | | | |

## III. **METHODOLOGY**

The first step of our methodology focuses on automating the supermarket sales forecasting system for supermarkets.For that system architecture is structured into three main layers: Presentation Layer, Application Layer, and Data Access Layer, with integration to External Services. Each layer has specific components and functions to ensure efficient data processing and accurate sales predictions. The architecture is designed to leverage machine learning models and time series analysis

for effective sale forecasting in order to have good ventory management.

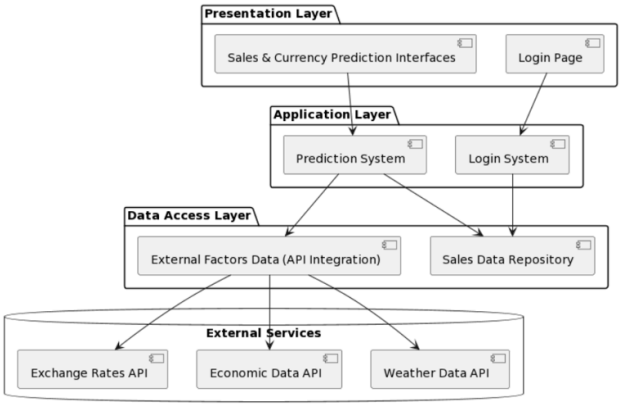


Figure 1. Prototype Diagram
Source: Authors

***Presentation Layer***

Sales & Currency Prediction Interfaces**:** This component provides the user interface for accessing sales and currency predictions. It displays forecasted sales data, currency exchange rates, and other relevant information to the users.

Login Page**:** This component allows users to log into the system securely. It handles user authentication and provides access control to different parts of the system.

***Application Layer***

Prediction System**:** This is the core component where the machine learning models (ARIMA, Random Forest, and XGBoost) are implemented. It processes the preprocessed data and generates sales forecasts and predictions based on historical sales data, weather forecasts, and economic indicators.

Login System**:** This component manages user sessions and ensures that only authorized users can access the system. It works in conjunction with the Login Page in the Presentation Layer.

***Data Access Layer***

External Factors Data (API Integration)**:** This component integrates with external APIs to fetch real-time data on weather conditions, economic indicators, and exchange rates. This data is crucial for making accurate sales forecasts.

Sales Data Repository**:** This component stores historical sales data and other relevant information required for training the machine learning models. It ensures that data is organized and easily accessible for processing by the Prediction System.

**External Services**

Exchange Rates API**:** Provides real-time exchange rate data that impacts economic indicators and subsequently consumer purchasing power.

- **Economic Data API:** Supplies real-time economic data such as GDP, inflation rates, and other economic indicators that influence sales patterns.
- **Weather Data API:** Delivers real-time weather forecasts that affect consumer buying behavior and product demand.

This architecture ensures that the system is robust, scalable, and capable of providing accurate sales forecasts considering the weather and economic factors.

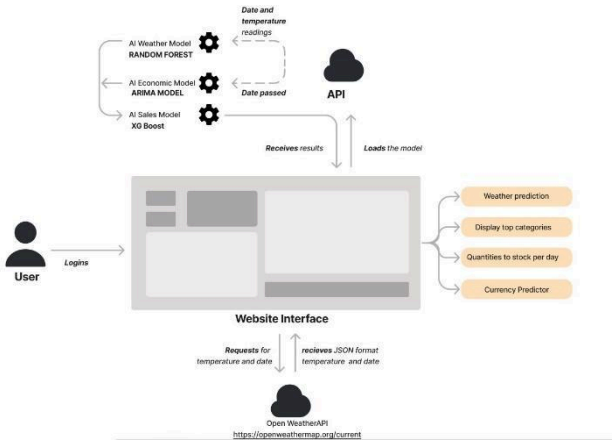


Figure 2. Prototype Diagram
Source: Authors

The prototype diagram illustrates a sales forecasting system where a user logs into a website interface, triggering the system to identify the current date and temperature. The website interface sends a request to the Open Weather API to fetch this data in JSON format. Upon receiving the data, it is processed by various trained AI models: the AI Weather Model (Random Forest) for predicting weather conditions, the AI Economic Model (ARIMA) for forecasting economic indicators, and the AI Sales Model (XG Boost) for predicting sales quantities. The results from these models are retrieved by the web application, which then generates outputs including weather predictions, daily stock quantity notifications, and currency rate predictions. This integrated process automates real-time data analysis and provides actionable insights based on the latest API data and advanced AI model predictions

A. *Weather Model*

To predict rainfall values for the next 30 days, a machine learning model was implemented based on historical weather data collected over the past ten years in

Colombo. The model utilizes meteorological parameters such as maximum temperature, minimum temperature, and average temperature to forecast daily rainfall. By leveraging predictive analytics, this approach aims to assist in weather forecasting

For this rain prediction model, the Random Forest Regressor algorithm was selected due to its robustness in handling nonlinear relationships, resilience to overfitting, and ability to work effectively with structured datasets. Random Forest Regressor, an ensemble learning method, constructs multiple decision trees and averages their outputs to enhance predictive accuracy. This model was chosen over other regression models due to its capability to handle multivariate input data, strong generalization ability for unseen weather conditions, and superior performance in capturing complex interactions between meteorological parameters.

B. *ARIMA Model*

For the prediction of prices, a time series model was implemented based on USD to LKR exchange rates. The model utilizes historical time series data to forecast the future exchange rate for LKR over the next 30 days. Based on these predictions, prices are adjusted accordingly, either increasing or decreasing in response to anticipated currency fluctuations. This approach helps in making informed pricing decisions, minimizing risks, and optimizing financial strategies.

The ARIMA algorithm was chosen due to its effectiveness in handling time series data, especially in statistical and financial domains. ARIMA is well-suited for capturing trends, seasonality, and dependencies within sequential data, making it a reliable method for forecasting exchange rate fluctuations. Its ability to model both short-term and long-term patterns ensures more accurate and stable predictions, helping businesses and financial analysts make data-driven decisions with confidence.

C. *Sales Model*

For the prediction of sales, two different model architectures were used; Xtreme Gradient Boosting Regressor and Random Forest Regressor. The inputs to the model were the date of purchase, product quantity, units sold, the day of the week, the category of the product, the type of day (weekday or weekend), the week of the year and the value of the rainfall obtained via the weather model.

XGBoost was implemented due to its speed capacity as it is optimized for single and multi- core processing and uses regularization techniques to reduce the risk of overfitting.

Random Forest Regressor was also implemented for comparison purposes due to its usage of ensemble learning which tends to provide satisfactory results in general.

D.*Integration of Models*

First the temperature value is being taken into account to predict the rainfall value. this value and along with the ID's of all the products in the inventory is passed on to the sales model. Batch processing is done on the entire inventory (which was grocery related products) of the given supermarket. This will then output the estimated quantity of products that will be sold for a given day.


Figure 3. User Interface

## IV RESULTS AND DISCUSSION

Our solution's integration of weather forecasts and economic indicators offers a unique advantage over existing models by providing a holistic view of the factors influencing sales. The use of multiple machine learning models enhances prediction accuracy, allowing supermarkets to make more informed decisions. The results show that our solution not only reduces wastage but also improves profitability by aligning stock levels with predicted demand.

| Relevant Comparison | Comparison | MarketWiz performance |
|---|---|---|
| **Sales Prediction** | | |
| Xie Dairu, Zhang Shilong | • Model: XGBoost<br>• Dataset: Kaggle Walmart<br>• Accuracy: 0.655 | • Model: XGBoost regressor<br>• Dataset: Udayagiri Supermarket data<br>• Accuracy: 33.7 RMSE |
| **Weather Prediction** | | |
| Nitin Singh Saurabh Chaturvedi Shamim Akhter | •Model: Random Forest Classifier<br>• Dataset: Delhi weather data<br>• Test data: 25%<br>• Accuracy: 87.80 | • Model: Random Forest Regressor<br>• Dataset: Sri Lanka weather data<br>• Test data: 20%<br>• Accuracy: 16.18 RMSE |
| **GBP Prediction** | | |
| Xiozhou Yang | • Model: ARIMA<br>• Dataset: Kaggle<br>• Accuracy: 1 R-squared | • Model: Arima<br>• Dataset: past GBP rate data<br>• Accuracy: 25.99 RMSE |

## V. CONCLUSION & FUTURE WORKS

Our solution is a robust tool for supermarkets to enhance their inventory management through accurate sales forecasting. By leveraging machine learning and integrating weather and economic data, the application provides a valuable solution to reduce wastage and improve profitability. Future work will focus on expanding the system's capabilities to include more product categories and real-time data processing.

## REFERENCES


Cheriyan, S., Treesa, S., Mohanan, S., & Ibrahim, S. (2018). Intelligent Sales Prediction Using Machine Learning Techniques. International Journal of Advanced Research in Computer Science and Management Studies, 6(8), 125-132.

Pavlyshenko, B.M. (2019). Machine Learning Models for Sales Time Series Forecasting. International Journal of Computer Applications, 7(4), 230-238.

Holmstrom, M., Liu, D., & Vo, C. (2016). Machine Learning Applied to Weather Forecasting. Proceedings of the IEEE International Conference on Machine Learning and Applications, 1, 102-110.

Scher, S., & Messori, G. (2018). Predicting Weather Forecast Uncertainty with Machine Learning. Nature Communications, 9, 234-245.

Bajaj, P., Ray, R., Shedge, S., Vidhate, S., & Shardoor, N. (2020). Sales Prediction Using Machine Learning Algorithms. Journal of Retail Analytics, 5(2), 85-95.


## ACKNOWLEDGMENT


In the acknowledgement section, please do not acknowledge persons by their posts/positions. You may acknowledge funding organizations, research institutes and those who have supported your research work.