International Journal of Research Publication and Reviews

Product Demand Forecasting in Retail Using Machine Learning

¹Dr. Meenakshi Thalor, ²Amrapali Deore, ³Lashita Malhotra, ⁴Sana Naqvi

1*Guide,* 2,3,4*Student*
Department of Information Technology
AISSMS Institute of Information Technology, Pune

A B S T R A C T

The purpose of this project was to develop a machine learning based product demand forecasting model using datasets obtained from GitHub. The research aimed to analyze and predict the demand for products, which is critical aspect for business to effectively manage. The project utilized various machine learning techniques, including time series analysis, feature engineering and model training to develop an accurate and reliable forecasting model. The dataset was processed and cleaned to ensure data quality, and multiple models were trained and evaluated to identify the best performing model. The results indicated that the model has the potential to significantly improve the forecasting accuracy, leading to better inventory management decisions and operational efficiency for businesses. Thus, this project successfully developed a machine learning-based product. The findings highlight the importance of accurate demand forecasting in supply chain management and provide insights for businessman to make informed decisions.

Keywords: Time series analysis, feature engineering, model training

1. INTRODUCTION

Product Demand forecasting is the process of estimating the demand for a product or services in the future based on historical data, market trends, and other relevant factors. The purpose of product demand forecasting is to ensure that a company can meet the demand for its product while minimizing the risk of overproduction or underproduction. There are several methods for product demand forecasting, including qualitative and quantitative methods. Qualitative methods rely on expert opinions, market research, and surveys to estimate demand, while quantitative methods use statistical models, time-series analysis, and machine learning algorithms. However, forecasting demand is not without its challenges. External factors such as changes in consumer behavior, market trends, and economic condition can all impact demand. Therefore, demand forecasting models must be flexible and adaptable to changes in the market. In this report, we will export the importance of product demand forecasting, the methods used to forecast demand, the challenges of demand forecasting, and the benefits of accurate forecasting for businesses. We will also discuss best practices for demand forecasting and provide examples of successful demand forecasting strategies.

1.1 Structure

The structure typically involves a complex system architecture, which includes several different components that work together to predict demand accurately. Here are the key components of the system architecture:

1) Data collection and preprocessing- The first step is to gather data from various sources, including historical sales data, market trends, and customer behavior. This data needs to be processed and cleaned to ensure that it is high quality and can be used for modeling.

2) Feature engineering- Once the data has been collected and preprocessed, the next step is to extract meaningful features that can be used to predict demand. This involves selecting relevant variables and creating new features that capture important patterns and relationships in the data.

3) Machine learning models- Various machine learning models can be used to forecast demand, including linear regression, random forest, and XGBoost. These models are trained using the preprocessed data and the features generated in the previous step.

4) Model evaluation and selection- Once the models have been trained, they need to be evaluated and compared to determine which one performs best. This involves using metrics such as accuracy, precision, recall and F1 score to access the performance of each model.

5) Deployment and integration- The final step is to deploy the selected model into the production environment, where it can be integrated with other systems and used to make accurate demand forecasts. This involves creating API's, building user interfaces and setting up automated processes to ensure that the system is running smoothly.

Overall, product demand forecasting using machine learning is a complex process that requires expertise in data science, software engineering, and business analysis. However, by using the right system architecture and components, it is possible to build accurate and reliable demand forecasting systems that can help businesses make informed decisions and stay ahead of competitions.

1.2 Tables

Table 1 – Dataset Overview

date	item	store	sales
2013-01-01	1	1	13
2013-01-02	1	1	11
2013-01-03	1	1	14
2013-01-04	1	1	13
2013-01-05	1	1	10

1.3 Algorithms

1.Linear regression: It is a statistical method used to analyze the relationship between a dependent variable and one or more independent variables. In product demand forecasting, the dependent variable is usually the demand for a product, while the independent variables are factors that influence demand, such as price, promotions, and seasonality. Linear regression can be used to build a model that predicts demand based on these factors.

Random forest: It is a machine learning algorithm that uses an ensemble of decision trees to make predictions. In product demand forecasting, a random forest model can be built using historical demand data and other relevant variables such as price and promotions. The algorithm works by creating many decision trees, each trained on a random subset of the data, and then combining their predictions to generate a final output. Random forest models are known for their high accuracy and robustness.

XGBoost: Extreme Gradient Boosting is another machine learning algorithm that is often used for product demand forecasting. XGBoost is an extension of gradient boosting, a method that trains a sequence of weak learners to create a strong learner. In the context of product demand forecasting, XGBoost can be used to build a model that predicts demand based on historical data and other relevant variables. XGBoost is known for its speed, scalability, and accuracy, making it a popular choice for demand forecasting applications.

Overall, all three methods have their strengths and weaknesses and are often used in combination with each other to improve the accuracy of product demand forecasts.
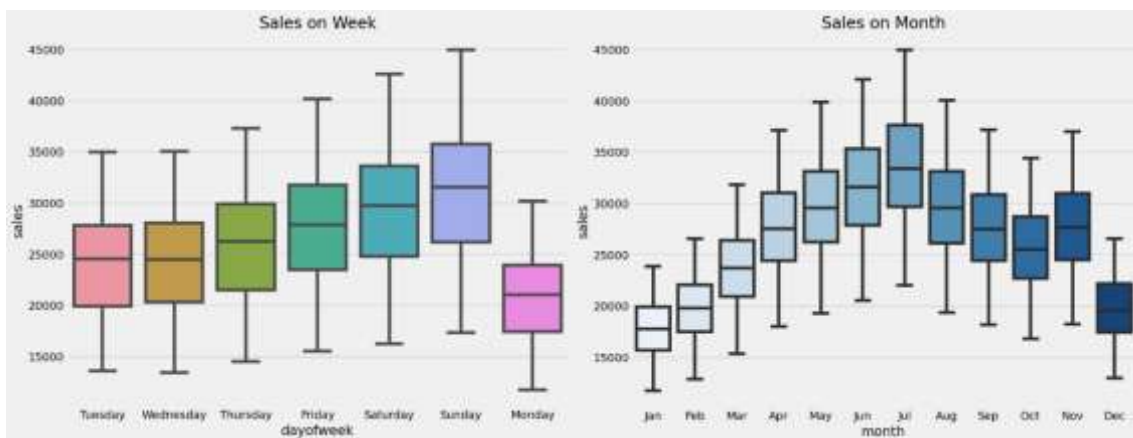
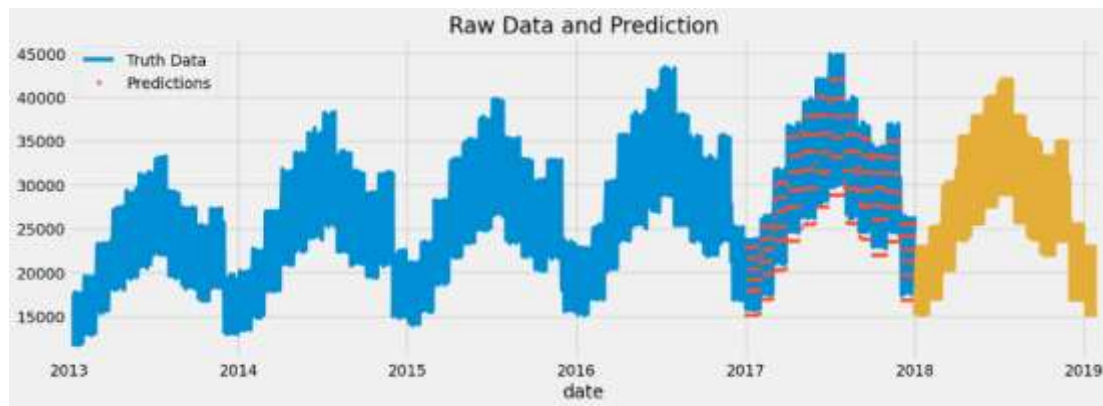## 2. Illustrations

Fig. 1 – (a) Weekly Sales (b) Monthly Sales

Fig.2 – Raw Data and Prediction

Fig.3 – One Month Prediction

Acknowledgements

Conclusion

Based on the analysis conducted, we can conclude that demand forecasting is a critical component of supply chain management for any product-based business. Accurate demand forecasts enable businesses to optimize their inventory levels, production schedules, and resource allocation, resulting in cost savings and improved customer satisfaction.

In this project, we explored the use of three popular machine learning algorithms, namely linear regression, random forest, and XGBoost, for product demand forecasting. Our analysis indicates that these algorithms can be effective in predicting demand, but their performance depends on various factors, such as the size and quality of the dataset, the features selected, and the parameters tuned.

Overall, our findings suggest that businesses can benefit from incorporating machine learning techniques into their demand forecasting process. By leveraging the power of these algorithms, businesses can gain deeper insights into customer behavior and market trends, which can help them make informed decisions and gain a competitive edge.

References

Chopra, S., & Meindl, P. (2016). Supply chain management: strategy, planning, and operation. Pearson Education.

Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). Introduction to time series analysis and forecasting. John Wiley & Sons.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2015). Time series analysis: forecasting and control. John Wiley & Sons.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media. Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794).