Paper ID: 205

Demand Forecasting for Grocery Products with Shorter Shelf Life in a Super Shop Using Machine Learning

Tasfia Kabir[1], Maisha Mushana[2], Md. Ariful Islam[3]
[1]*Department of Industrial and Production Engineering, Rajshahi University of Engineering and Technology, Rajshahi-6204, Bangladesh,*
[2]*Department of Industrial and Production Engineering, Rajshahi University of Engineering and Technology, Rajshahi-6204, Bangladesh,*
[3]*Department of Industrial and Production Engineering, Rajshahi University of Engineering and Technology, Rajshahi-6204, Bangladesh*
E-mail: tasfiakabir32@gmail.com, maishaipe35@gmail.com, arifulislam@ipe.ruet.ac.bd

Abstract

Demand forecasting is very important in managing inventory. This is even more crucial for the products with shorter shelf lives as they tend to get over-stocked or under-stocked which leads to wastage and loss of business. A precise demand forecasting of these products can prevent such problems. To get more accurate demand forecasting for perishable products, this study proposes a data-driven approach. By using machine learning algorithms, and historical sales data, the research aims to create a predictive framework to specify features of perishable products. In this study, we engaged with managers from two super shops, obtaining one month of historical sales data for vegetables, fruits, meat and poultry, fish, milk, and bread. In machine learning, we performed time series analysis using three reliable algorithms: Random Forest, XGBoost, and Catboost. Our research showed that the Random Forest algorithm demonstrated best accuracy in predicting sales of these supermarket goods.

Keywords: Demand forecasting, Shorter shelf life product, Machine learning.

1. Introduction

In today's world, figuring out how much of the fresh stuff to keep on the shelves in the supermarkets can be tricky. In a grocery store, it's essential to know what people want to buy and how much they want. This is how product shortage and wastage can be avoided. The main goal of the case study would be to predict how much perishable items should be stocked in the shelves in order to achieve most accurate sales and in the process avoid wastage and shortage of the perishable items. This goal can be achieved by using various algorithms and forecasting methods. In this era of Artificial Intelligence, Machine Learning and its algorithms have made it possible to easily predict future demand and sales most accurately by analysing historical data. This study explores few of these algorithms and compares them in order to locate the most exact outcome and finally use that trained model algorithm to implement in super shops demand forecasting of products with shorter shelf life.

1.1 Problem Statement

Grocery stores face significant challenges in reducing the waste and managing inventory for products with shorter shelf life, such as fresh vegetables, fruits, meat and poultry, fish, milk, bread and bakery items. The lack of accurate demand forecasting for these perishable goods can result in financial losses due to overstocking or stockouts, leading to customer dissatisfaction and food wastage. This project aims to address the critical problem of demand forecasting for grocery products with shorter shelf life in super shops. With the help of machine learning we will be able to create accurate forecasting systems to reduce waste, optimise inventory levels and enhance the overall shopping experience for customers.

1.2 Objective

To conduct a highly accurate demand forecast for grocery items with a shorter shelf life within a super shop setting in order to minimise product wastage and financial losses using machine learning algorithm.

1.3 Research Gap:

There are still a lot of unresolved issues regarding demand forecasting techniques for perishable products over a variety of industries. Current research often focuses on broad forecasting methods or particular sectors, but it lacks a cohesive strategy which involves outside variables including change in consumer behavior, economic disruptions, and real-time data processing. study has been done on hybrid models that combine machine learning and demand forecasting.

2. Literature Review

This study introduces a novel approach to demand forecasting, leveraging various methodologies such as K-Nearest Neighbour, Support Vector Machine, Gaussian Naive Bayes, Random Forest, Decision Tree Classifier, and regressions. This method's capability to provide precise approximations of product demand offers significant advantages to retailers striving to meet customer needs effectively. [1]

This paper's conceptual framework lays the groundwork for future empirical research, aiming to refine forecasting methodologies within hypermarket retail. Achieving perfect forecasting planning is vital for performance enhancement and inventory accuracy control in the face of fierce competition. Forecasting errors lead to missed sales opportunities and supply chain risks. By examining various phases of the forecasting process, this research contributes to mitigating these risks and fostering sustainable inventory accuracy in hypermarket retail. [2]

Predicting food sales accurately is crucial in the food industry, spanning supermarkets, groceries, restaurants, bakeries, and patisseries. It's imperative to minimise wastage caused by overstocking or expiration while capitalising on sales opportunities. Employing established machine learning methods for food sales forecasting can address this challenge. This involves analysing sales data timelines, choosing input variables, and defining the sales output variable. Numerous machine learning algorithms can be employed for this purpose, and assessing their accuracy involves various evaluation metrics. [3]

Market basket analysis plays a pivotal role in comprehending consumer preferences, optimising product placement, and refining marketing tactics. By scrutinising purchasing patterns, retailers can discern product interactions and strategically arrange similar items in stores, leveraging association rules mining. This analytical approach facilitates informed decisions regarding pricing, product placement, endorsements, and profitability. Through the examination of relationships within a database, companies can gain insights into customer behaviour. Market basket analysis is actively utilised to bolster customer satisfaction and drive revenue growth in practice. [4]

The problem involves sales information on a daily basis from 1C Company, one of the biggest software companies in Russia. Using past data, the objective is to forecast the total sales for every product and retailer for the upcoming month. Here they used a neural network design based on Long Short Term Memory (LSTM) and eXtreme Gradient Boosting (XGBoost) to accomplish this goal. By comparing the actual and anticipated target values, the Root Mean Squared Error (RMSE) statistic was used to assess the performance of the model. As a result XGBoost worked better than LSTM. [5]

Forecasting demand for perishable items such as fruits and vegetables presents a significant challenge for retailers, as it requires effectively managing inventory to avoid both excess stock and shortages. Existing literature offers insights into this complex issue. One commonly employed approach is time series analysis, which uncovers patterns and seasonal trends in historical sales data. Collaborating with supply chain partners to utilise demand detectors, optimization models, and tailored analytics for seasonal demand can enhance prediction accuracy. While metrics like MAE and MSE remain crucial for evaluating forecasting models, industry reports and real-world case studies provide valuable insights into effective forecasting methods. [6]

Sales forecasting is important for maximizing costs, inventories, and production. Especially in sectors that handle perishable items it is even more important. This study shows that ARIMA models are commonly used for time series forecasting and also useful for identifying seasonal patterns and trends. Specifically, the $ARIMA(0,1,1)(0,1,1)_{12}$ model was used to anticipate sales of a perishable orange drink. The model's R^2 score was over 90% which indicates great accuracy. Although its strong forecasting capabilities, ARIMA is unable to adjust for external factors such as industry trends or population expansion. Regression analysis and ARIMA combined may offer a more improved method that enables companies to comprehend and anticipate the factors that influence sales. [7]

Accurate demand forecasting is very important for perishable products. The effectiveness of machine learning algorithm like XGBoost performs well by including some external factors such as weather and calendar events. These multiple variables, dataset-specific models perform better than standard models because they are able to identify individual features of the data and improve accuracy. Studies also show that regular model updates have no significant impact on long-term forecasting. These things shows that machine learning models perform better at predicting demand for perishable goods.[8]

Demand forecasting for perishable products in online retail is important for effective inventory management and profitability. This study highlights some machine learning methods such as Feed-Forward and Autoregressive Recurrent Neural Networks, particularly DeepAR. These are especially effective for products with medium to high demand variability. The coefficient of variation is used to classify products into low, medium, and high variation groups. DeepAR is recommended for products with medium to high variability. More simple models can be applied to the items with low variability for improving overall performance and inventory efficiency. [9]

Retailers have improved demand forecasting through advanced data collection methods. It has helped them make data-driven decisions. Traditional models like ARIMA and Holt-Winters have been widely used but machine learning models such as SVMs and ANNs are now gaining priority due to their ability to handle complex patterns. This paper addresses the lack of research on fish demand forecasting in retail, focusing on reducing waste for mid-sized gilthead bream. Machine learning models like LSTM outperformed traditional statistical models in accuracy and waste reduction. The study highlights the importance of machine learning in creating more sustainable food supply chains by improving demand forecasts and minimizing waste. [10]

The paper explores the challenges in managing pharmaceutical supply chains, focusing on perishable products and delivery scheduling. Previous research has addressed pharmaceutical supply chain network design , but few studies have considered perishability and soft time windows. This paper fills that gap by developing a bi-objective model to minimize costs and delivery penalties. Machine learning techniques like linear and quadratic regression are used to predict demand and improve accuracy. The results show that quadratic regression performs better than linear regression. This approach helps optimize pharmaceutical supply chains by reducing costs, shortages, and waste while increasing customer satisfaction. [11]

3. Methodology

To achieve the set objectives, an informative and quantitative research was conducted in 2 selected Super Stores (Shwapno and Amana Big Bazar) in Rajshahi, Bangladesh. They have provide us the estimated data of the sell and stock of their products for a particular period of time. we found out that machine learning based demand forecasting has not become commonly used in most super shops. So we have three proposed algorithms of machine learning for forecasting the demand of the perishable products. which are XGBoost, CatBoost and random forest. Time series analysis was perform here.

In forecasting, various metrics are used to quantify the accuracy of predictions and measure the errors between predicted values and actual observations. Here are some common error metrics used in forecasting: Mean Absolute Error (MAE), Mean absolute percentage error (MAPE), R^2.

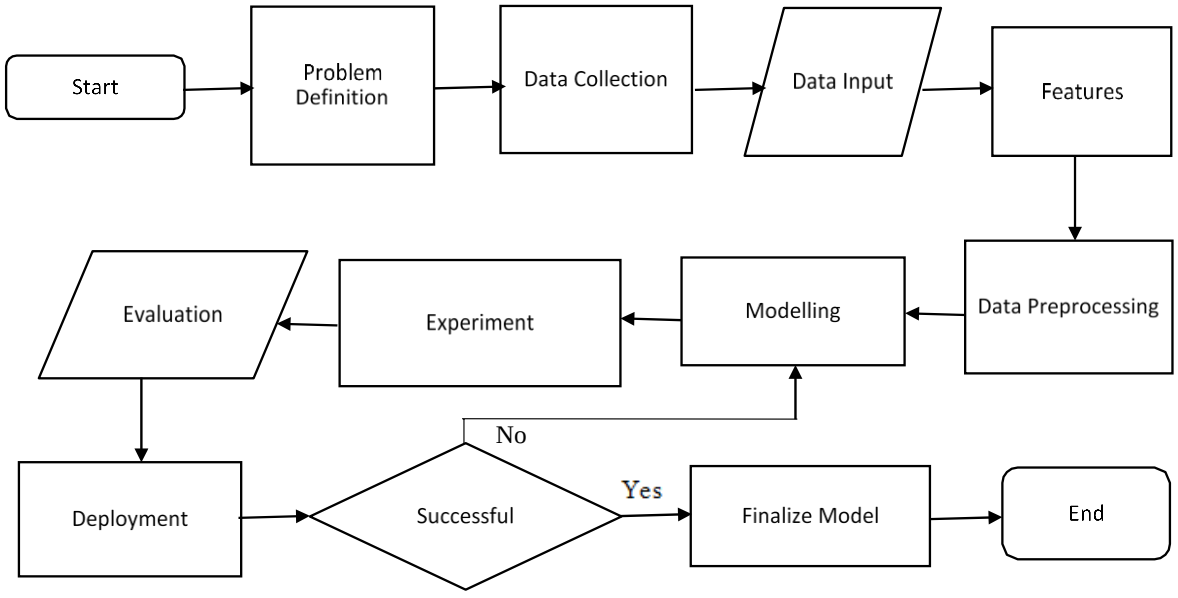


Fig. 1: Steps of Methodology

4. Data Analysis

For this project the datasets were collected from two well-known mega supermarkets, "Shwapno" and "Amana Big Bazar". Data collection took place in the period of October 1 - October 31 in 2023 using face-to-face interviews with predetermined questions. The factors we considered while collecting data from each super shop were Product Name, Restocking schedule, Date, Restock quantity, Restock unit, Quantity sold, Sold unit. They have provided us the estimated dataset.

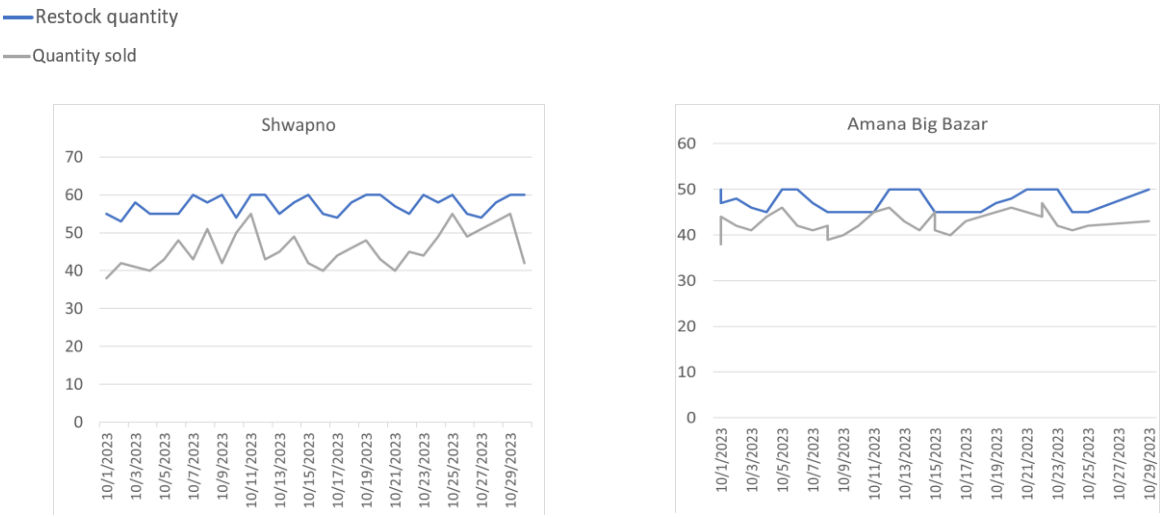


Fig. 2: Restock quantity and quantity sold data of one month from Shwapno and Amana Big Bazar

5. Result and Discussion

For comparing the performance of different models, evaluating them based on Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and R-squared (R^2) on both the training and test sets. Here's a summary of your results for each model:

1.XGBoost Regressor:	2.CatBoost Regressor:	3.Random Forest Regressor:
Training MAE: 0.0485 Test MAE: 3.2685	Training MAE: 0.0339 Test MAE: 4.1497	Training MAE: 0.9740 Test MAE: 3.1915
Training MAPE: 0.0016 Test MAPE: 0.0986	Training MAPE: 0.00077 Test MAPE: 0.1011	Training MAPE: 0.0310 Test MAPE: 0.0766
Training R^2: 0.99999 Test R^2: 0.98698	Training R^2: 0.99998 Test R^2: 0.96279	Training R^2 : 0.99780 Test R^2 : 0.98698

with high R-squared values on both training and validation sets. Though its R-squared values were significantly lower, the Random Forest Regressor likewise showed good performance.

The comparison between true financial loss and predicted financial loss highlights the importance of accurate demand forecasting. The true financial loss is based on actual sales data and the financial loss after prediction estimates the outcomes based on forecasted demand. In the dataset, we see a consistent reduction in financial loss after prediction compared to the true financial loss. This shows that the demand forecasting model is ultimately lowering the overall financial impact very effectively. The comparison table is shown below:

Restock Quantity	Quantity sold	Predicted demand	True financial loss	Financial loss after prediction
15	8	7.68	7	7.32
55	38	42.49	17	12.51
45	35	38.77	10	6.23
15	8	9.26	7	5.74
50	42	44.71	8	5.29
98	80	86.77	18	11.23
100	93	87.31	7	12.69
54	52	52.41	2	1.59
55	53	53.3	2	1.7
50	41	43.31	9	6.69
250	200	185.63	50	64.37
68	65	64.07	3	3.93
48	45	43.25	3	4.75
95	77	86.23	18	8.77
50	46	44.37	4	5.63
40	36	35.58	4	4.42
58	53	49.91	5	8.09
82	81	76.77	1	5.23
56	49	52.57	7	3.43
50	46	42.09	4	7.91
47	41	43.83	6	3.17
60	55	44.13	5	15.87
55	40	43.03	15	11.97
99	83	89.13	16	9.87
85	78	80.89	7	4.11
54	51	48.79	3	5.21
15	9	9.05	6	5.95
12	7	8.35	5	3.65
40	39	35.66	1	4.34
60	43	43.97	17	16.03
12	7	8.49	5	3.51
98	92	86.76	6	11.24
55	53	52.62	2	2.38
10	10	10.06	0	-0.06
38	34	32.3	4	5.7
35	27	30.98	8	4.02
15	7	9.46	8	5.54
70	67	64.44	3	5.56
12	9	8.68	3	3.32
50	40	42.73	10	7.27
10	7	7.2	3	2.8
240	192	185.79	48	54.21
8	7	7.06	1	0.94
18	13	10	5	8
60	51	55.91	9	4.09
50	42	45.59	8	4.41
58	51	53.46	7	4.54

Fig. 3: Comparison of financial loss

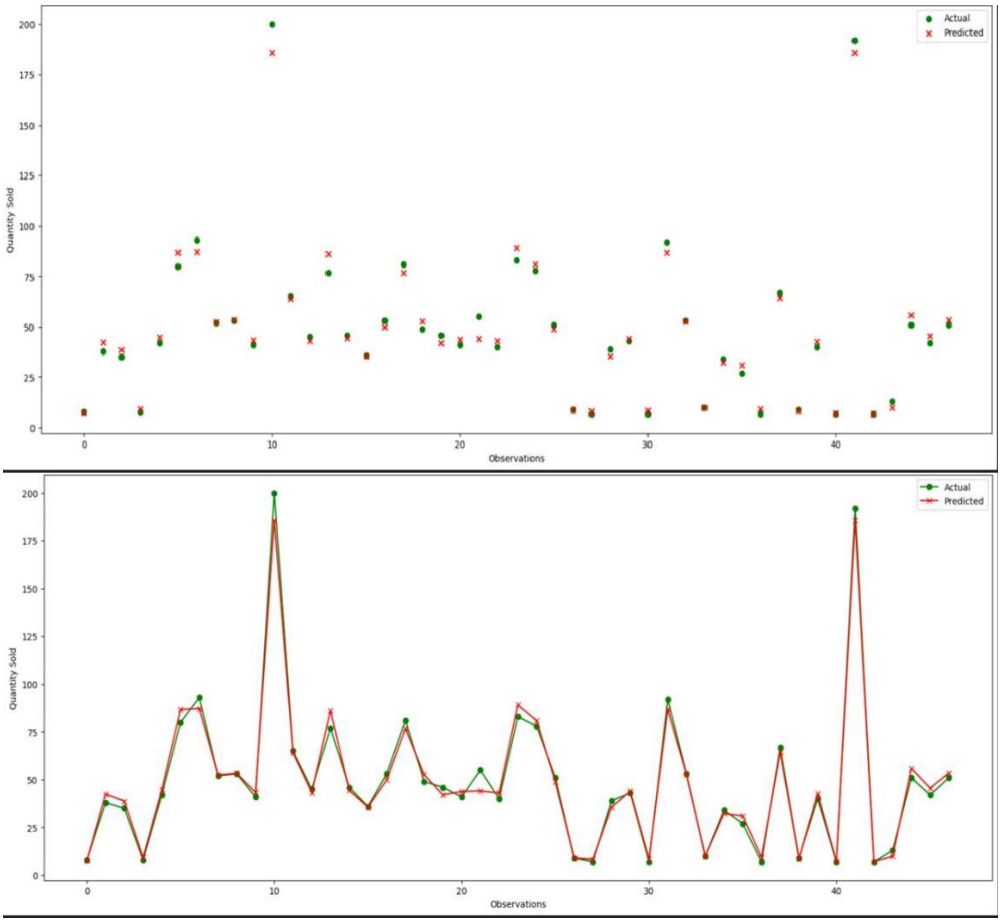


Fig. 4: Plot of actual sales and predicted sales vs time

These are the graphs of the predicted sales data after conducting the demand forecasting time series analysis. The random forest regression analysis's findings for October 2023's replenishment quantities from "Shwapno" and "Amana Big Bazar" generally match the actual data, demonstrating the model's capacity to identify underlying patterns. Significant variances could point to unexpected events or factors that affect restocking. The implications for inventory control show how important accurate forecasts are to maximising stock levels and successfully meeting customer demand. Further optimization or research into different approaches may be considered for better prediction performance.

The possible use of time series forecasting using the Prophet technique was explored, offering directions for future research, while results were discussed and future actions were considered. All things considered offers insightful information for demand forecasting in the grocery shops which have products with shorter shelf life.

6. Conclusion

This case study is performed to conduct demand forecasting in the context of grocery products with shorter shelf life. Upon inspecting our relevant super shops, we found that they use excel to store historical data and perform stock analysis and based on that they stock the product. Due to lack of accuracy in the forecasting, the perishable products can easily be overstocked or under-stocked. This arising problem can cause wastage and customer dissatisfaction and ultimately loss of revenue. Predicting future sales accurately is always a huge concern when it comes to inventory management and overall success of any business. The final result of this study has demonstrated significant advantages for super shops with an extremely high rate of accuracy in predicting future sales using machine learning and the selected algorithm is random forest which provides the most accurate predictions. In future work, analysis can be done by involving more respondents and more perishable products can be added to the analysis over a longer period of time.

7. References

[1] Md. A. I. Arif, S. I. Sany, F. I. Nahin, and A. S. Azad Rabby, “Comparison Study: Product Demand Forecasting with Machine Learning for Shop,” in *Proc. of the SMART–2019, IEEE Conf. ID: 46866, 8th Int. Conf. on System Modeling & Advancement in Research Trends*, Moradabad, India, pp. 171-176, 2019.

[2] S. Mohan; A. Lawrence, “Inaccurate Forecasting Planning Nurture Inventory Wastage (Perishable Items) in the Hypermarket Retail”, *Global Business & Management Research, ISSN 1947-5667,* Vol.11, No.1, pp. 487, 2019.

[3] G. Tsoumakas, "A Survey of Machine Learning Techniques for Food Sales Prediction" *Artif. Intell. Rev.*, Vol. 1, No. 1, pp. 1-20, 2018.

[4] S. Patangia, R. Mohite, K. Shah, G. Kolhe, M. Mokashi, and P. Rokade, "Sales Prediction of Market using Machine Learning," *International Journal of Engineering Research & Technology (IJERT)*, Vol. 9, No. 9, pp. 708-713, 2020.

[5] Devendra Swami, Alay Dilipbhai Shah, and Subhrajeet K B Ray, “Predicting Future Sales of Retail Products using Machine Learning,” *University of Southern California, Los Angeles*, CA, 2020.

[6] Trang T. M. Ho, Ly V. Tran, Hieu M. Tran, and Son V.T. Dao, "Machine Learning in Demand Forecasting," *International Research Journal of Advanced Engineering and Science*, Volume 7, Issue 3, pp. 225-233, 2022.

[7] T. Musora, Z. Chazuka, A. Jaison, J. Mapurisa, and J. Kamusha, “Sales Forecasting of Perishable Products: A Case Study of a Perishable Orange Drink,” *Computer Science & Information Technology (CS & IT)*, Vol. 1, No. 4, pp. 103-116, 2023.

[8] J. Eiglsperger, F. Haselbeck, V. Stiele, C. Guadarrama Serrano, K. Lim-Trinh, K. Menrad, T. Hannus, and D. G. Grimm, “Forecasting Seasonally Fluctuating Sales of Perishable Products in the Horticultural Industry,” *Expert Systems with Applications*, Vol. 249, pp. 123438, 2024.

[9] W. Premrudikul, S. Ahmornahnukul, and A. Pongsathornwiwat, “Developing Optimal Demand Forecasting Models for a Very Short Shelf-Life Item: A Case of Perishable Products in Online’s Retail Business,” *Journal of Information Technology Applications & Management*, Vol. 30, No. 3, pp. 1-13, June 2023.

[10] S. Rekabi, Z. Sazvar, and F. Goodarzian, “Machine learning model with linear and quadratic regression for designing pharmaceutical supply chains with soft time windows and perishable products,” *Decision Analytics Journal*, Vol. 9, pp. 100325, 2023.

[11] V.L. Miguéis, A. Pereira, J. Pereira, and G. Figueira, "Reducing fresh fish waste while ensuring availability: Demand forecast using censored data and machine learning," *Journal of Cleaner Production*, Vol. 359, pp. 1-10, 2022.