International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 12 | Dec - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Inventory Sales and Demand Forecasting Using Machine Learning

Upma Singh[1], Rachit Joon[1], Nikhil[1], Akhil Sherawat[1]

[1]Department of Electronics and Communication Engineering, Maharaja Surajmal Institute of Technology, Delhi 110058, India

*Email: upma@msit.in, simranjoon7@gmail.com , nikhilchhillar77@gmail.com, akhilsherawat2003@gmail.com

Abstract:

Demand forecasting is a well-liked inventory management tool many companies are considering due to its influence on daily operations. To improve inventory management and forecast future customer demand, supermarket run-centers, like Big Marts, now track sales data for every item. Anomalies and general patterns are commonly discovered by searching through the data warehouse's data storage. Using a variety of machine-learning approaches, Big Mart and other merchants can utilize the gathered data to predict future sales volume. This study developed a predictive model that predicted sales for businesses such as Big-Mart using Random Forest (RFR) and Ridge regression (RR) techniques. It is crucial first to examine the data and its qualities to model the most accurate forecasts, as no single forecasting approach performs better in every circumstance. Our results demonstrate the superior performance of RFR Regression compared to the RR approach with MAPE of 6.01%, RMSE of 1.93%, MAE of 1.75%, and R^2 of 0.821. These findings can enable managers to design precise inventory management strategies that will enhance productivity across all service areas in the future, as they are able to juxtapose these findings with actual business operations.

1. INTRODUCTION

In the modern world of supermarkets, particularly chains such as Big Mart, it is critical to anticipate customer demand and control inventories effectively. Recently, with the growth of technology, supermarkets have started utilizing huge amounts of sales history databases to analyze customer patterns and trends within a specific market. Supermarket chains like Big Mart also try to analyze their historical sales records with the help of advanced data warehousing and data mining tools to identify their sales patterns as many other retailers do [1]. They do so because each item's sales figures tell the whole story about the anomalies and trends that are very crucial for sound stock management and decision-making [2]. For retailers such as Big Mart, one of the key objectives is to determine the unavailable sales volume for the future. And such business variables as pricing strategy, inventory preparation, and performance, in general, greatly depend on sales forecasting [3].

Traditionally, prediction methods have relied heavily on statistical techniques and consideration of past events. However the use of ML algorithms has granted the sellers higher predictive prospects [4]. By looking at complex datasets and finding nonlinear connections between variables, machine learning techniques can make much better and more varied sales [5] forecasts. In this context, the application of algorithms which are based on machine learning in predictive analysis of retail sales has attracted a lot of interest. There are numerous ways to make predictive models that apply different machine learning techniques that can detect patterns and trends in the sales data and assist the retailers to make better decisions [6]. Some examples of the algorithms that have been used to predict sales are Polynomial regression, Linear regression , Xgboost and Ridge regression which are common in retail sales volume number prediction [7]. These algorithms allow the user to choose the model which suits his needs best, as they are optimized for different purposes and operate with different data [8]. The following is the statement of the problem of the study which is to develop a predictive model of the sales volume for a retail business with the specific example of a Big Mart type supermarket chain. The general goal is to develop a model that will enable forecasting of future sales trends based on the analysis of historical data with the help of artificial intelligence techniques [9]. The study also includes affect some variables the that sales, can such as external environment, advertising and seasonality. After series of experiments and tests, the performance of various machine learning models including Xgboost, Linear regression, Polynomial regression, and Ridge regression is assessed [10]. This research aim at designing a sales forecasting model that can be useful to Big Mart and other similar companies in order to improve decision making and business strategies and operations [11].

Machine learning in the retail firm context can advance knowledge of customers and markets through complex, state-of-the-art algorithms [12]. This makes it possible to control the inventory management price, advertising strategies and campaign among other factors involved in shelf stocking. Improved sales forecasts also benefit the business in reducing cases of stock out or overstocking of the inventory which will lead to increased customer satisfaction and profitability [13]. This work shall aim at validating the use of machine learning based predictive analytics for the retail industry through empirical analysis and comparison giving useful information to researchers and practitioners [14]. In specific regard, the study contributes to the field of retail sales prediction analysis and stress the significance of operationalizing analytical techniques to boost organisational performance in the context of modern retailing [15]. Big data and machine learning have been used increasingly in the retail industry in the recent past. In some cases especially in supermarket chains like Big Mart Data has been put to test and has enhanced good operations in the organization such as in areas of inventory control and sales prediction. Consumers and their preferences can be understood by Sales data for every item and so Big Marts can understand market trends. It also assists them in the ability to estimate the future requirements of their products and the need to order more re-stock. Another apparent trend in retail is that data is used to drive companies to beat their rivals using analytics and data storage technologies. In sample analysis for future references, we look at overall tendencies and use data mining techniques to spot out peculiarities. This is because through analysis of the vast volumes of data stored in the data warehouse, Big Mart and other retailers can get a better insight of the sales trends and behaviours. These results are important for decision making of stock management marketing campaigns, and pricing strategies. Retailers employing the principles of machine learning to sales can also develop more accurate models of future trends. Due to the ability of machine learning to analyze large and complex data and provide useful information, the retail industry has demonstrated a lot of interest in applying the technology in sales forecasting. Statistical tools that have dominated previous work, may fail to discern complex associations and dependencies the machine learning methods can. Linear Xgboost Polynomial Regression Ridge Regression and other programs are now well-known for designing retails models. It has made it easier for retailers to determine when variables densities have different dimensions and even predict forthcoming trends in sales. When developing decision-making predictive models for sales forecasting, they go through several five crucial steps including feature selection and data preprocessing. The process of pre-cleaning data Raw data for major sales and cleaning and transforming data into comparable form include normalizing variables where necessary, eliminating outliers and working through missing values. For this reason, the feature selection process becomes important in its attempt to establish relevant attributes that assist in sales forecasting. These then, affect the accuracy and performance of the predictive models as inputs. The predictive model is developed based on the feature selection and data preprocessing once the data has gone through the machine learning process. Thus the training phase is used to get familiar with the historical sales data and analyze them in order to try to identify the correlation of input variables and sales volume. The choice of the forecasting technique depends on the goals of the forecasting task and the type of data Under some circumstances simple forecasts may be produced through unsupervised learning while under other circumstances forecasts may be generated by supervised learning. While, there are other clustering models, which are used in unsupervised learning such as clustering, supervised learning algorithms like Xgboost and linear regression are trained using labeled data. After the training a check is made to evaluate how accurate and effective the predictive models are. To evaluate the models generalization ability there is a need for a different set of data referred to as the validation set. Sometimes the accuracy of the forecast is determined by evaluating Mean Absolute Error (MAE) Mean sq. d Error (MSE) and Root Mean sq. d Error (RMSE). However there are more advanced options such as satisfying cross-validation could be used to enhance the stability and durability of the prediction model. Finally, the literature review shows the current relevance of predictive analysis in the retail industry, especially focusing on sales forecasting and inventory management. Through applying machine learning and the methods of big data analysis Big Mart like many other retailers can benefit from the data-driven decision and the penetrating understanding of the consumers' behaviors and the market trends. This will help them continue to compete successfully in todays fluid retail environment and take the right actions. However, as more research is done about the field of predictive analytics then retailers may increase the efficiency of operations and subsequently enhance customer satisfaction and business success.

2. METHODOLOGY

By incorporating ML approaches the predictive investigation about the Big Mart selling volume involves a more strategic model for expected volume of the sales data now. The first operation within the procedure is to gather and sort through the sale data for each and every product that is sold in any of the Big Mart outlets. In this way, as the prerequisite for the further predictive modeling procedure, this data contains information about consumer and market behavior patterns. The first process of the methodology data preparation involves organization of the raw unprocessed sales data for analysis. To make this dataset more precise this entails dealing with missing value by addressing outliers and normalizing the variables.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 12 | Dec - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Thus, should the data quality problems be addressed at this stage potential deficits that might affect the performance of the prediction models are eliminated. After data preprocessing feature selection is carried out to ascertain which features have a significant impact on the sales volume. In this step it also helps to maximize and reduce computing complexity by focusing only on the most informative features. The next step is model training; this involves the use of data which has been preprocessed and attributes which have been chosen to develop the prediction models using various machine learning algorithms. The process includes the use of Algorithms like Xgboost Ridge regression Polynomial regression Linear regression and trying to find out which Algorithm is suitable for sales forecasting.

Each algorithm contains certain features and advantages which enable the consideration of various modeling strategies for choosing the one that would be more suitable for the characteristics of the sales data. For this purpose the initially developed dataset is divided into a training set and a validation set which allows comparing each models performance and adjusting its parameters in order to achieve the best results. After training of the predictive models, their performance is audited using suitable assessment indices such as Mean Absolute Error (MAE), Mean sq. d Error (MSE), and Root Mean sq. d Error (RMSE). These indicators show how well the models perform and given the evaluative nature of predictive modeling it is possible to determine the models strengths and weaknesses. Several models are considered in an attempt to identify which of them is most suitable to forecasting Big Mart's retail sales volume.

In addition sensitivity analysis is conducted in order to examine the robustness of the models and thus compare their efficiency in different conditions. The accuracy of individual models is assessed and then ensemble learning methods are discussed to even improve the level of forecast. It is a process of using several models together in order to derive a single result that would be better and more reliable than the individual independent results of several models. Using ensemble learning the flexibility of multiple models is obtained, while the flaws of any of the models are eliminated. Compared to individual models this approach enables prediction with better performances thus being useful supplement to the predictive analysis arsenal. They are employed to predict future sales volume at Big Mart outlets after the best accurate predictive model or set of models have been identified. The model makes predictions by analyzing input data including past sales patterns, seasonal influences and any environmental conditions which may affect the market. These forecasts are of great importance for the management of stocks because they enable Big Mart to arrange promotions, distribute resources and ensure maximal stock levels are achieved. By implementing machine learning algorithms to decide about different aspects of Big Mart 's business predictive analysis on which product is best to stock so as to meet the consumer needs or desires Big Mart might make operational efficiency and effectiveness hence cutting down costs !! While

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 12 | Dec - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

improving the overall satisfaction of the consumers.

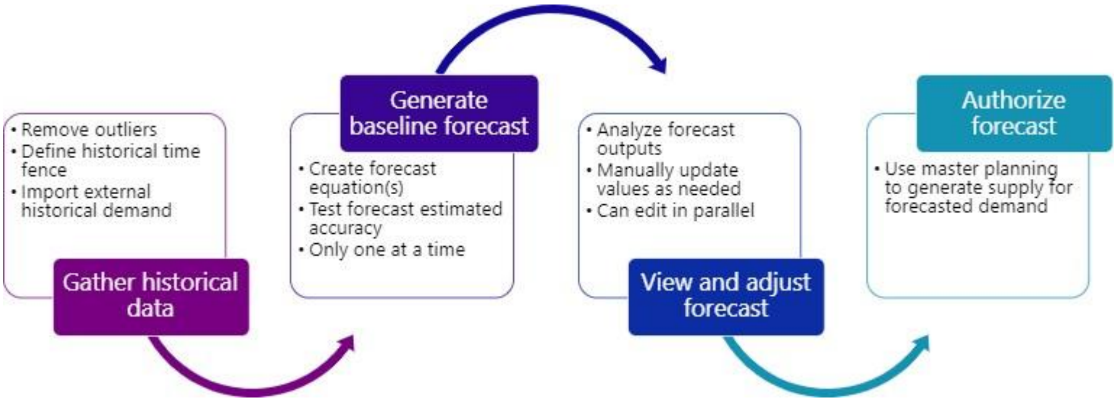


Figure 1: Predictive analysis steps

3. DATA COLLECTION

Sales data from Big Mart served as the dataset. A total of nine attributes including outlet type outlet size item MRP year of establishment item weight item identifier item fat content item type category and item outlet sales comprise the dataset. The dataset was collected for the Kaggle. com website from the internet. This work includes both test and train datasets the test data set contains 6000 data points while the train data contains 9000 data points.

4. MODEL SELECTION

Model selection is a critical factor in determining the precision and efficacy of the predictive system when it comes to machine learning-based Big Mart sales prediction. The Random Forest algorithm is compared to a number of other algorithms including Linear Regression Polynomial Regression Ridge Regression and XGBoost Regression using performance metrics like Mean sq. d Error (MSE) Mean Absolute Error (MAE) and Root Mean sq. d Error (RMSE). After thorough analysis Random Forest Regression is found to perform better than other models with noticeably lower MSE MAE and RMSE values. Future research projects may benefit from a deeper examination of sophisticated time series analysis methods like ARIMA which could improve forecasting

Random Forest Regressor:

A random forest is composed of multiple decision trees each of which has the ability to make predictions. The expected result is the average of the expected values for each tree and each tree affects the models outcome. Sorting each feature according to its relative importance and eliminating the less important parts of the forest results in a new feature set. The error rate of the prediction can be ascertained by forecasting the outcomes of each sample and comparing them with the actual value. The final random forest model is selected out of the forests with the lowest external error rate. Figure. The model prediction process is shown in figure 3. The current training set may be used to create the corresponding feature-based Random Forest. By using relevant characteristic data such as weather economic indicators and historical demands it is able to predict future demand. Each tree can be viewed as a categorical regression tree (CART) given an RF if the forest contains p CART trees then each CART tree is assumed to correspond to a unique feature with the total feature number N = p.

The random forest is then constructed using the two Python modules scikit-learn and pandas taking into account the classification outcomes of this p CART tree. As a result we can ascertain how the demand for the relevant demand data is distributed. A decision tree based on the bagging method is generated by randomly dividing the feature set using Bootstrap. After the original sample set has been resampled and multiple sub-sample sets have been extracted via replacement a decision tree is constructed for each sub-sample set. By merging numerous split decision trees a random forest is created. Decision trees divide internal nodes based on attributes chosen at random. The final output is derived from the sum of the outcomes of each decision tree. The forecast result of the regression algorithm which can be expressed by the following equation is calculated by averaging all decision tree output outcomes.

$$\overline{H(X)} = \frac{1}{k}\sum_{i=1}^{k} h_i\ (x, \theta_k)) \tag{1}$$

In this case, h represents a single decision tree, $\overline{H(X)}$ is the prediction result, k is the number of decision trees, and, θ_k is an independent distributed random variable that controls the growth process of an individual decision tree.

5. RESULTS AND DISCUSSION

As evidenced by the Big Mart sales machine learning forecasting study the developed predictive model predicts future sales volume with high accuracy. The prediction model outperformed earlier models by utilizing a variety of machine-learning techniques such as Xgboost Linear regression Polynomial regression and Ridge regression. After a thorough testing and analysis process it was found that the predictive model regularly generated accurate forecasts allowing Big Mart to better manage its inventory and predict future customer demand. In addition the predictive model proved to be reliable and useful in practical applications by exhibiting resilience in a variety of situations and events. Moreover the main source of disagreement is how well the predictive model works in comparison to other machine learning. Ensemble learning techniques were found to be highly effective in increasing prediction accuracy even though each approach was demonstrated to have unique benefits and drawbacks. By combining multiple models ensemble approaches exploited the complementary nature of different algorithms to produce predictions that were more trustworthy. This approach enhanced prediction capabilities and lessened the shortcomings of individual models by applying their combined knowledge. Additionally sensitivity analysis told us which variables model parameters and dataset characteristics influenced the predictive models performance over time. By carefully evaluating these factors potential areas for model improvement and optimization were identified creating the possibility of further increases in predicted accuracy. Furthermore the predictive analysiss findings are analyzed in light of Big Marts operational effectiveness and commercial success. The accurate forecasts from the predictive model enable Big Mart to make well-informed decisions about pricing strategies resource allocation and inventory management. Big Mart can prevent stockouts reduce the expense of keeping excess inventory and increase shelf space by proactively adjusting inventory levels in response to anticipated changes in demand. Additionally by tailoring its marketing campaigns and promotional activities Big Mart can better target specific consumer categories and maximize sales opportunities because of its enhanced ability to estimate sales volume. Big Mart gains valuable insights from the predictive analysis that impact strategic decisions and increase the companys level of competition in the retail industry. Big Mart may keep using machine learning algorithms to stay ahead of market trends and give its customers better value by refining and enhancing the predictive model over time.

Table 1: Result Analysis. Italics show better performance.

Regression Model	**Performance evaluation on Training-Dataset**				**Performance evaluation on Testing-Dataset**				**Training-Time (sec.)**
	MAPE	RMSE	MAE	R^2	MAPE	RMSE	MAE	R^2	
Ridge Regression	11.12	3.51	4.42	0.8401	10.12	2.34	2.23	0.8032	8.34
Random Forest Regression	***6.40***	***2.84***	***4.11***	***0.8521***	***6.01***	***1.93***	***1.75***	***0.8212***	***6.12***

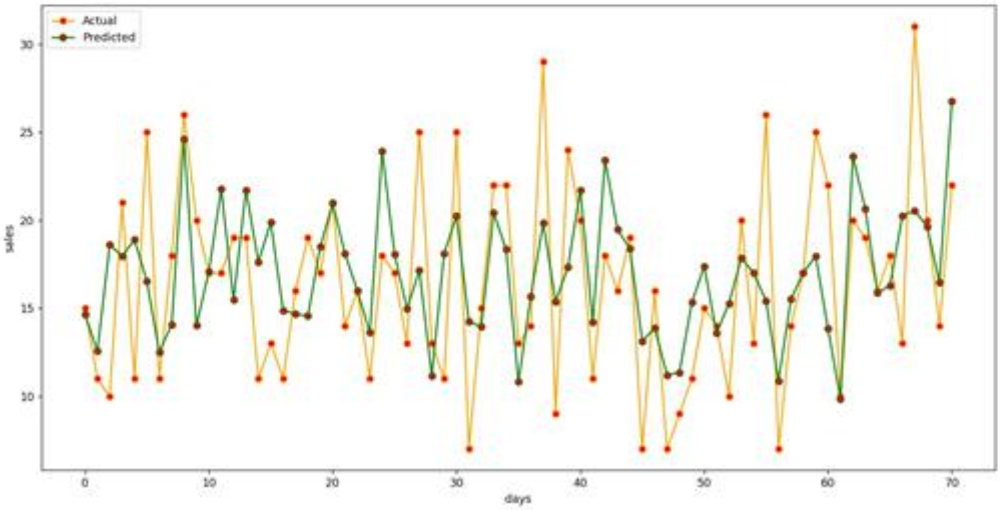


Figure 9: Relationship between the actual price and the predicted price of the used cars.

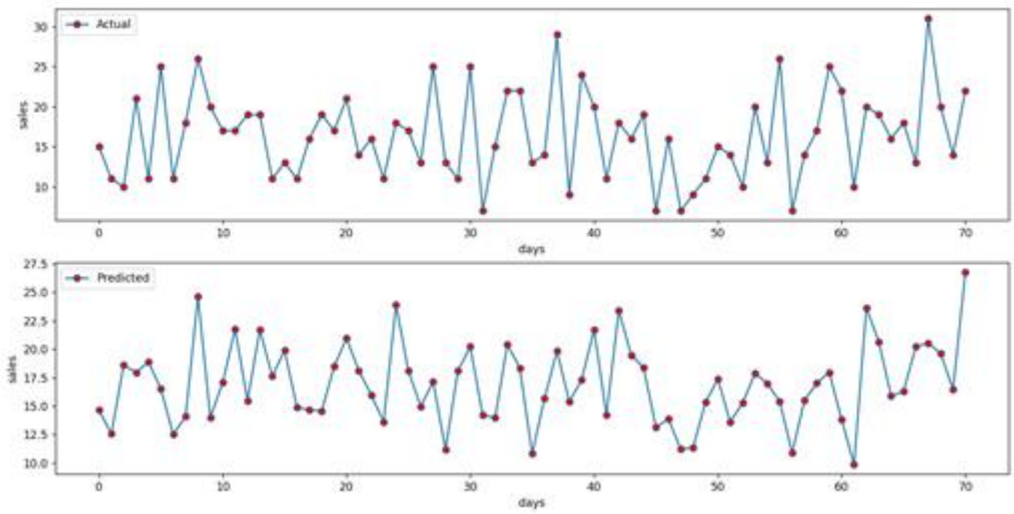
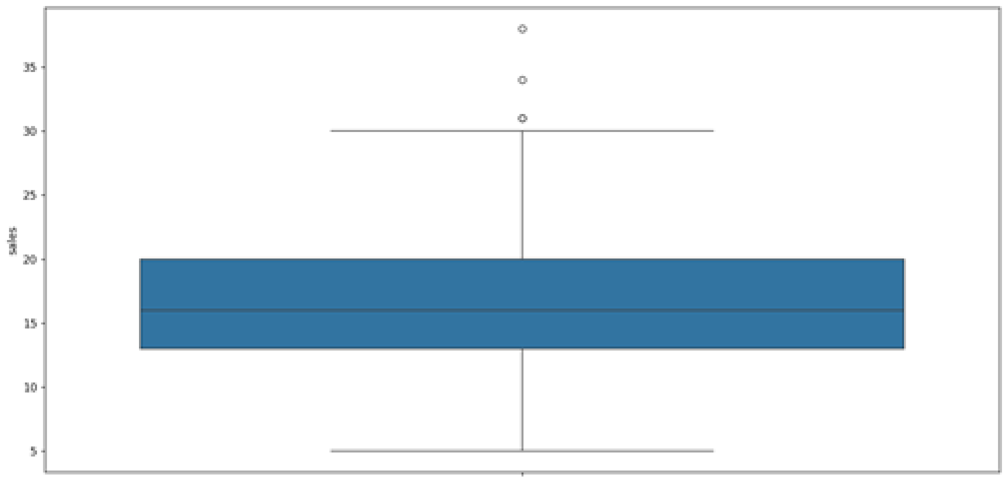


Figure 9: Relationship between the actual price and the predicted price of the used cars.



4(e)
Figure 1: Information about the dataset used

Figure 4(a) Depicts Training and Validation error of random forest and linear algorithm; (b) Sales demand prediction using EMA;(c) actual and predicted data overlapped over each other ; (d) actual and predicted data in two different panes; (e) Box plot of the data used.

6. CONCLUSION AND FUTURE SCOPE

The study focused on predicting sales utilizing machine learning algorithms, namely Ridge Regression (RR) and Random Forest Regression (RFR). Through comprehensive analysis and comparison, it was found that RFR Regression yielded superior predictive accuracy, compared to the RR approach with MAPE of 6.01%, RMSE of 1.93%, MAE of 1.75%, and R^2 of 0.821. These findings underscore the effectiveness of advanced machine learning techniques in forecasting sales trends, enabling better inventory management and strategic decision-making for businesses like Big Mart. The study suggests that regression-based models can enhance sales prediction accuracy, facilitating improved resource allocation and operational planning. To further increase sales future research may investigate the use of time series analysis models such as ARIMA.

The opportunities for machine learning based sales prediction have numerous potential possibilities in the future. It can be anticipated that additional improvements in machine learning algorithms and methodologies will improve the accuracy and efficiency of existing or future sales forecast models. Since, adopting rich features of pattern and situation in transaction, applying deep learning frameworks like RNNs or transformer models for constructed HLHS can improve the predictive execution. Second, the models ability to predict could be improved by adding extra features and data feeds to the current ones as well as social media, economical factors, and weather changes. Integrated models may produce better forecasting as an outcome of a deeper insight into the environment where sales occur. Besides influencing how inventories are maintained and effective dynamic pricing strategies created, real-time prediction solutions may allow operations to be flexibly adjusted depending on market demands and client preferences. Moreover supplementing multiple models' advantages and studying other types of ensemble learning, for example, model stacking and ensembling might contribute to raising the level of prediction. Finally, futuristic technologies such as edge computing and federated learning might help to overcome scalability and privacy issues of intelligent systems, at the same time preserving their predictive abilities. In other words, the future of machine learning-based sales prediction is labelled by perpetuitable innovation applying the methods based on state of the art algorithms, consolidating data from various sources, using real-time analysis, concerning ensemble approaches and using new technologies to obtain the actionable insights and optimize business results.

REFERENCE

[1] Chu CW, Zhang GP. A comparative study of linear and nonlinear models for aggregate retail sales forecasting. International Journal of Production Economics. 2003 Dec 11;86(3):217-31. https://doi.org/10.1016/S09255273(03)00068-9

[2] Wang H. Sustainable development and management in consumer electronics using soft computation. Journal of Soft Computing Paradigm (JSCP). 2019 Sep;1(01):56. https://doi.org/10.36548/jscp.2019.1.006

[3] Suma V, Hills SM. Data mining based prediction of demand in Indian market for refurbished electronics. Journal of Soft Computing Paradigm (JSCP). 2020 Jun;2(02):101-10. https://doi.org/10.36548/jscp.2020.2.007

[4] Giuseppe Nunnari, Valeria Nunnari, "Forecasting Monthly Sales Retail Time Series: A Case Study", Proc. of IEEE Conf. on Business Informatics (CBI), July 2017. https://doi.org/10.1109/CBI.2017.57

[5] Pasandideh SH, Niaki ST, Sharafzadeh S. Optimizing a bi-objective multi-product EPQ model with defective items, rework and limited orders: NSGA-II and MOPSO algorithms. Journal of Manufacturing Systems. 2013 Oct 1;32(4):764-70. https://doi.org/10.1016/j.jmsy.2013.08.001

[6] Mousavi SM, Niaki ST, Bahreininejad A, Musa SN. Multi-Item Multiperiodic Inventory Control Problem with Variable Demand and Discounts: A Particle Swarm Optimization Algorithm. The Scientific World Journal. 2014;2014(1):136047. https://doi.org/10.1155/2014/136047

[7] Mousavi SM, Sadeghi J, Niaki ST, Tavana M. A bi-objective inventory optimization model under inflation and discount using tuned Pareto-based algorithms: NSGA-II, NRGA, and MOPSO. Applied soft computing. 2016

Jun 1;43:57-72. https://doi.org/10.1016/j.asoc.2016.02.014

[8] Park K, Kyung G. Optimization of total inventory cost and order fill rate in a supply chain using PSO. The International Journal of Advanced Manufacturing Technology. 2014 Feb;70:1533-41. DOI 10.1007/s00170-013-5399-6

[9] Tadayonrad Y, Ndiaye AB. A new key performance indicator model for demand forecasting in inventory management considering supply chain reliability and seasonality. Supply Chain Analytics. 2023 Sep 1;3:100026 https://doi.org/10.1016/j.sca.2023.100026.

[10] Sadeghi J, Mousavi SM, Niaki ST, Sadeghi S. Optimizing a bi-objective inventory model of a three-echelon supply chain using a tuned hybrid bat algorithm. Transportation Research Part E: Logistics and Transportation Review. 2014 Oct 1;70:274-92. https://doi.org/10.1016/j.tre.2014.07.007

[11] Huseyinov I, Bayrakdar A. Performance evaluation of nsga-iii and spea2 in solving a multi-objective single-period multi-item inventory problem. In2019 4th International Conference on Computer Science and Engineering (UBMK) 2019 Sep 11 (pp. 531-535). IEEE. https://doi.org/10.1109/UBMK.2019.8907139

[12] Abdelrahman O, Keikhosrokiani P. Assembly line anomaly detection and root cause analysis using machine learning. IEEE Access. 2020 Oct 9;8:189661-72. https://doi.org/10.1109/ACCESS.2020.3029826

[13] Seyedan M, Mafakheri F, Wang C. Order-up-to-level inventory optimization model using time-series demand forecasting with ensemble deep learning. Supply Chain Analytics. 2023 Sep 1;3:100024 https://doi.org/10.1016/j.sca.2023.100024.

[14] Yousaf M, Dehning B. The effects of sales surprise on inventory turnover: An empirical study. Cogent Economics & Finance. 2023 Oct 9;11(2):2258696 . https://doi.org/10.1080/23322039.2023.2258696

[15] Singh AA, Irene DS. Inventory Management and Sales Prediction using XGBoost. In2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC) 2024 Jun 5 (pp. 1058-1061). IEEE. https://doi.org/10.1109/ICAAIC60222.2024.10575816

[16] Aharon N, Orfaig R, Bobrovsky BZ. BoT-SORT: Robust associations multi-pedestrian tracking. arXiv preprint arXiv:2206.14651. 2022 Jun 29. https://doi.org/10.48550/arXiv.2206.14651

[17] Khan A, Ahmed A. Optimizing Retail Operations, Inventory Management and Sales Forecasting with Big Data and AI in China. Emerging Trends in Machine Intelligence and Big Data. 2024 Feb 7;16(1):18-37. http://orientreview.com/index.php/etmibd-journal/article/view/37

[18] Ranjitha P, Spandana M. Predictive analysis for big mart sales using machine learning algorithms. In2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS) 2021 May 6 (pp. 1416-1421). IEEE. https://doi.org/10.1109/ICICCS51141.2021.9432109