Inventory Management using Machine Learning

A PROJECT REPORT

Submitted by

OVIYA K
(2116210701186)

in partial fulfillment for the award of

the degree of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

RAJALAKSHMI ENGINEERING

COLLEGE ANNA UNIVERSITY,

CHENNAI

MAY 2024

RAJALAKSHMI ENGINEERING COLLEGE, CHENNAI

BONAFIDE CERTIFICATE

Certified that this Thesis titled **"Inventory Management using Machine Learning**" is the bonafide work of "**OVIYA K (2116210701186)"** who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form partof any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE

Dr . S Senthil Pandi M.E.,Ph.D.,

PROJECT COORDINATOR

Professor

Department of Computer Science and Engineering

Rajalakshmi Engineering College

Chennai - 602 105

Submitted to Project Viva-Voce Examination held on_____

Internal Examiner **External Examiner**

Inventory Management Using Machine Learning

Oviya K 210701186

Computer Science Engineering

Rajalakshmi Engineering CollegE

ABSTRACT

Inventory Management is essential for any company since it helps to increase the sales and predicts product demands. Machine learning models help to manage the inventory according to the demand for products at that particular time. It ensures that there is always enough stock to satisfy customer orders and if there is any shortfall, the model recognizes the pattern and issue an appropriate alert. It improves customer experience by reducing out-of-stock situations. It also reduces the overstock situations. Various predictive models are analysed to find best model which has least RMSE value and best r2 score.

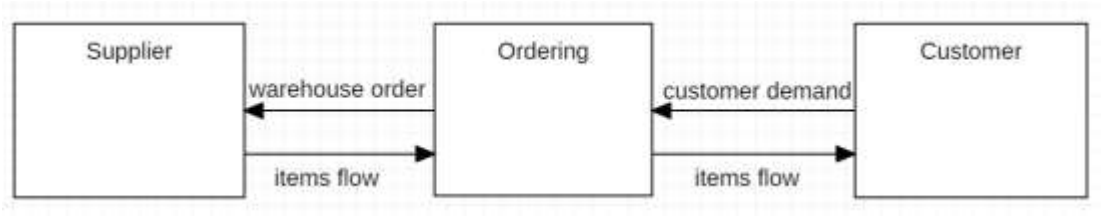
I. INTRODUCTION

In today's technology machine learning is widely used in many areas that improves the way the company operates and manages their process. One area that machine learning has significant advancements is inventory management. Inventory contains raw materials, partially finished goods, components and finished goods a company sells or that is used for production. It can be resources used in hospital like pacemakers, defibrillators, oxygen cylinder that are crucial for hospital to perform emergency department operations. It is a major requirement even for small and medium sized business. Hence Inventory management helps in managing all the details about the stock available which are consumed frequently and which are used less frequently. It predicts the demand of the product which is essential for the company's health. A business system which keeps track of inventory levels, sales and order to make predictions on product will help to reduce the out of stock and over stock situations.

Artificial Intelligence and Machine Learning should be used along with the human experts to determine the demand of the products in the company. Business that uses this model will have two main benefits that is increased sales and customer satisfaction. Traditional process make use of the manual that slow down the process and can make some errors. With the availability of huge amount of real time data, we need a model that manages this data and helps to transform this data into a useful prediction. It makes the company to stay in the competition with other business.

With this model the company knows how much stock they have and it helps to fulfil order from anywhere. It decreases the cost tied up in the inventory and reduce the unsold goods production so that company invests correctly. With the proper inventory management, company spend on inventory that sells and not on unsold goods. This improves the company's cash flow and the cash is always moving through the business. Customer needs can change constantly so that we need a model that finds the trend. Without proper management company finds it hard to refill the stock. AI will increase the inventory monitoring accuracy and decrease the material waste.

This paper proposes certain machine learning algorithms to manage inventories better. Out of all the algorithms, hypertuned XGBoost (Extreme Gradient Boosting) algorithm that uses decision trees has proved to give the highest accuracy.



Customer demands a product and order was placed. If that product is not in the warehouse, Supplier ensures to fulfil the order and items delivered to the customer.

II. LITERATURE SURVEY

When it comes to inventory management, many subjects need to be covered. Because the e-commerce sector is growing so quickly, effective inventory management is becoming more and more important. To modify the current inventory management strategies, more research is required.

A comprehensive machine learning system based on Apache Spark and probabilistic demand forecasting models is described by titans of the e-commerce sector. These enormous internet merchants possess enormous datasets. Time series analysis and machine learning techniques are the two main methods utilized in predictingAn effective case study on inventory management may be found by using a random and organized survey to find variables that impact inventory optimization among small and medium-sized enterprises (SMEs) in the steel industry. The factors fall into the categories of external and internal variables, in that order. Artificial neural networks are intelligent systems that use neural layersANNs excel at fitting problems. To increase forecast accuracy, a review of past research on ANN for inventory management is necessary.

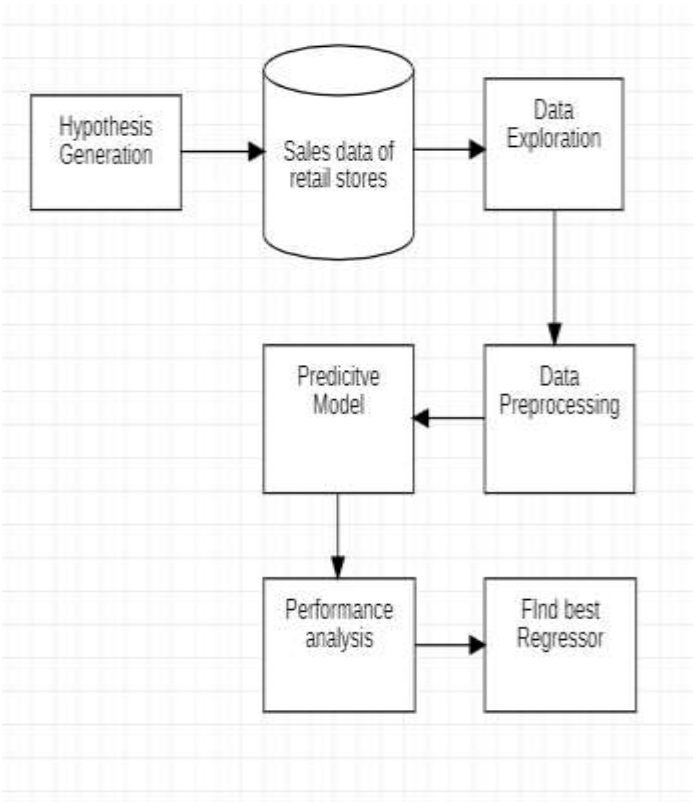
Managing customer data and forecasting customers' purchase habits are two applications of artificial intelligence. AI has the potential to assist in the development of manufacturing plans that precisely take into account variations in demand, including seasonal spikes, and alert companies to the need to replenish supplies.

Another technique to monitor inventory level is decision support system. For inventory analysis that considers a multitude of factors, we can also use decision making with multiple criteria in machine learning which is a hybrid approach. In addition to ABC analyses for class identification, the methodology uses Artificial Neural Networks (ANNs), Bayesian networks, and Support Vector Machines (SVMs) to expect many different classes for inventory items. Dead inventories can be identified using prediction models. Prediction modeling estimates part obsolescence in advance quite well by using a prediction algorithm.

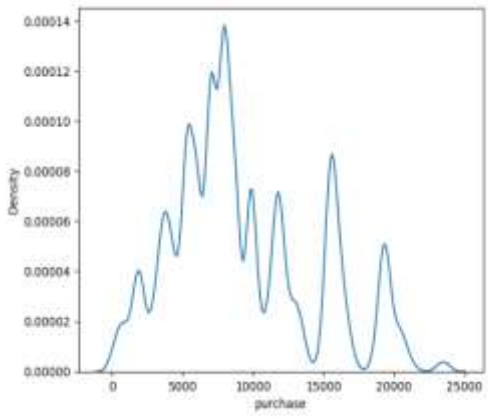
Each inventory item's relevant classes are determined through the use of a categorization approach known as ABC analysis, which provides the demand forecasting process with the ideal foundation. ABC analysis gives an efficient inventory stock classification. Artificial techniques that use neural networks can be used in predicting product demands. Algorithms such as variable learning rate and gradient descent are used to train these kinds of models. JADE platform is used to build software agents. Agents communicate with each other via messages. The solution is designed to replace the manual process with a fully automated one using the inventory management software that is currently in use.

III. PROPOSED METHODOLGY

Several risks such as intensive labor, high error prone systems and other internal as well as external risks are faced in today's organizations. As a result, many of the decisions a firm makes include a high degree of risk and uncertainty. Its sales of goods and services could be less risky if it can anticipate future demand. Demand forecasting is the scientific process of projecting future demand for an item or service that a business provides amongst a range of unpredictable and competitive conditions. In order to forecast demand, this research uses the XGB regression model. XGBoost uses decision trees and is a very useful algorithm.



Company has the product details stored in their database. Sales data of the retail stores is explored and the relevant columns are extracted. Feature engineering takes the important column that contribute for the prediction. Then, Predictive model is generated and its performance are analyzed. From these models, best model is selected for predicting the product demand.



1. Data preprocessing :

Start by exploring the sales data obtained from retail stores, focusing on key information such as product sales volume, pricing, promotions, and seasonal trends. Select relevant columns from the dataset that are deemed important for predicting product demand. These could include historical sales data, product attributes, pricing details, promotional activities, and seasonal indicators.

2. Feature Engineering:

Engage in feature engineering to transform raw data into informative features that can enhance predictive modeling. This process involves creating new features, such as lagged variables to capture temporal patterns, aggregating data at different time intervals, encoding categorical variables, and standardizing numerical features.

3. Predictive Model Generation:

Employ various machine learning algorithms to develop predictive models for forecasting product demand. Experiment with algorithms such as Random Forest, Gradient Boosting, Support Vector Machine, and neural networks to capture diverse patterns present in the data. Train the models using the extracted features from the sales data to learn the underlying relationships between predictor variables and product demand.

4. Performance Analysis:

Using appropriate assessment measures such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), or forecast accuracy metrics such as Mean Absolute Percentage Error (MAPE), evaluate each predictive model's performance. Compare the performance of different models to identify their strengths and weaknesses in terms of accuracy, robustness, and computational efficiency.

Conduct cross-validation to evaluate the generalization ability of the models and mitigate overfitting issues.

5. Model Selection :

Select the most effective model based on the performance analysis conducted. Consider factors such as prediction accuracy, computational efficiency, interpretability of results, and ease of

implementation when choosing the optimal model. Prioritize the model that offers the most accurate and reliable predictions of product demand while aligning with the specific requirements and constraints of the company's inventory management system.

Gathering and preparing data with care is the initial step in the proposed machine learning-based inventory management system. Sales figures, pricing, ratings, and product qualities like category, subcategory, and brand are among the historical data on the inventory that is gathered and meticulously preprocessed. This preprocessing stage consists of coding categorical variables, handling missing values, and occasionally scaling numerical features to ensure optimal performance of the resulting models. Feature engineering plays a critical role in the extraction of important predictors from the data, including ratings, pricing trends, sales history, product attributes, and even seasonal patterns. To extract complicated temporal dynamics from the data, more sophisticated techniques could be applied, including adding moving averages or lagged sales.

V. IMPLEMENTATION

The preprocessing steps are applied to the Big Basket products dataset which contains 10 features –sales prices, category, index, sub category, brands, ratings, descriptions, market prices, products. It has 27555 entries. Then, the dataset is split into two sets, in the ratio of 80:20, this 80% is for training and the 20% is for testing. Various machine learning algorithms are used and performance is analyzed.

1.Linear Regression:
First, linear Regression Model is used that assumes relationship between input variables and output variable. This model has Root Mean Squared Error (RMSE) as 4675.94.

2.KNN Regression:
KNN regression is a non-parametric method that roughly approximates the relationship between independent variables and the continuous result by averaging the observations within the same neighbourhood. The RMSE score of the KNN Regression is 3225.19.

3.Random Forest Regressor:
Random Forest Regressor increases accuracy prediction and it manages over-fitting by fitting several decision tress on different dataset subsets. It has a RMSE value of 3052.894.

4.XGB Regressor:
Extreme Gradient Boosting, or XGB, is a well-liked gradient boosting method for regression and classification applications.
One among the boosting techniques for lowering the bias error of the model. The base estimator for the Gradient Boost method is fixed, or Decision Stump.
You can utilize the gradient boosting algorithm for forecasting categorical and continuous target variables (as a Classifier and Regressor, respectively). When used as a classifier, the cost function is Log loss; when used as a regressor, it is Mean Square Error (MSE). XGB Regressor has RMSE value of 2898.36.
5. Ensemble method:

By Using Ensemble model Voting Regressor, we pass both gradient boosting Regressor and Random forest Regressor. It has RMSE value of 2078.91 and r2 score of 0.76. IT has very good RMSE value and r2 score.

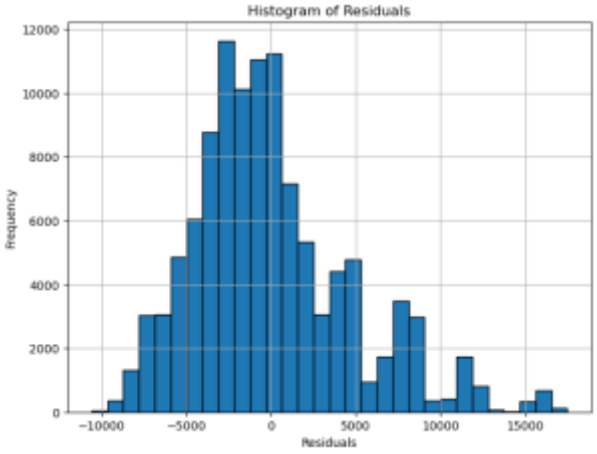
6.Hypertuned XGB Regressor:
An XGB regression model that has undergone hyperparameter tuning optimization is known as a hypertuned XGB Regressor. The values for the hyperparameters that yield the highest performance of the XGB model on a validation set or through cross-validation are chosen after hyperparameter tuning.
In comparison to the default hyperparameters or manually selected values, we hope to increase the predictive accuracy, robustness, and generalization capacity of the XGB Regressor by hypertuning it. It has a RMSE value of 2348.76.

Among these models, Ensemble method which contains voting Regressor for gradient boosting and random forest regressor has the lowest RMSE value and highest r2 score of 0.76.

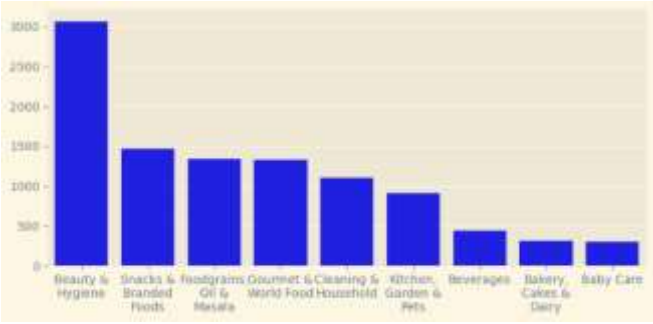
A rigorous process involving model selection, implementation, and data pretreatment must be used to implement the recommended inventory management system. To ensure data integrity, historical inventory data must first go through arduous preparation to manage missing values and encode categorical variables. Feature engineering extracts relevant predictors, and sub setting the dataset into training and testing subsets facilitates subsequent model evaluation. Once the data is prepared, regression techniques such as Linear Regression and Random Forest Regressor are trained and optimized using training data. After the model's performance is assessed using pertinent indicators, the optimal algorithm for forecasting product demand or identifying goods with high demand is selected.

After the optimal model has been determined, it is easily integrated with the existing inventory management systems before being deployed into the production environment. The deployed model's performance is monitored using mechanisms that enable frequent retraining and modifications. Version control and documentation help with system evolution and continual maintenance by ensuring transparency and reproducibility.

The inventory management system may efficiently optimize inventory strategies to meet demand while dynamically adjusting to shifting market conditions by using this iterative deployment, monitoring, and refining process.

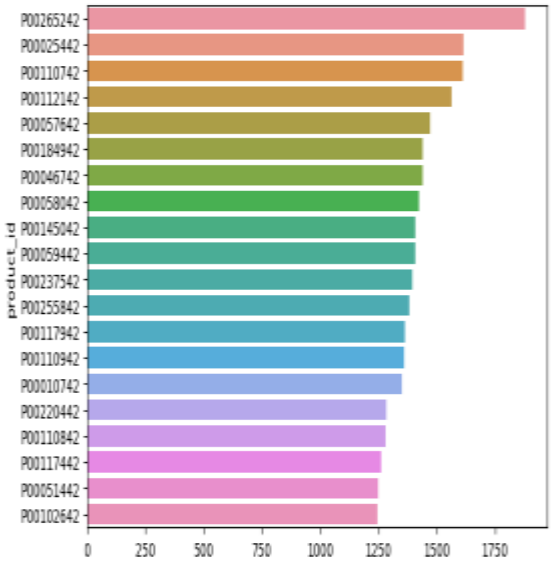


The term "residuals" in regression analysis describes the discrepancies or differences between the observed/actual value and the predicted value. Usually, the plot of residuals over time or across the series of observations is referred to as the "history of residuals". Using this graphic, one can determine whether the residuals show any patterns or trends that go against the regression model's presumptions. In order to show that the regression model is effectively capturing the underlying relationships in the data, the residuals should ideally be randomly distributed about zero with no obvious trends. Trends, cycles, or heteroscedasticity (unequal variance) are common patterns in the history of residuals that may point to issues with the model.
Company has many products brand and categories. Each product can be analysed with respect to their brand and sub categories. Model can predict the correct product that is in demand to be filled in the inventory.



The above graph shows the category of the product and their count so that each category can be analysed and filled accordingly. It is useful for the company to see which product category has maximum demand and they can overcome out of stock situations.

The products that customer purchase the most are found using the products count. The top 20 products are displayed using bar graph.



Top 20 products in demand

Our final goal is to create a machine learning models that reliably predict product according to their attributes, which aids in streamlining decision-making and inventory management procedures.

The ultimate goal is to provide decision-makers with useful information from the predictive model so they can improve supply chain efficiency, satisfy consumer demand efficiently, and optimize inventory management strategies—all of which will contribute to the growth and profitability of the company.

VII. CONCLUSION

In today's world, with growing demand and growth in E-commerce, many of the retail industries find it hard to manage their inventories. They are unaware of whether the stock they have currently is sufficient or surplus, With advancement in technology, many machine learning algorithms can predict every single product's demand and help the industry make better decision on how much to invest. Thus, with this prediction it can be made sure that the products are either over stocked or under stocked. Many algorithms such as Random forest and XGBoost provide good results with high accuracy and certainly help in this field.

Any business can benefit greatly from product demand forecasting since it boosts sales and keeps inventory levels stable while requiring the least amount of manual labor. To obtain best prediction various models like Linear Regression, Random Forest Regressor , XGB Regressor and Ensemble method are used. From these models, Ensemble model has the lowest RMSE values of 2078.91 and the highest r2 score value of 0.76. By utilizing neural networks for very huge datasets in the future, we can increase the accuracy of the model.

Furthermore, deep learning techniques can also be used to enhance the prediction. Techniques like neural networks, specifically, artificial neural networks show excellent predictions. Even more hidden layers as well as epochs can be added to get even better results. Further down the line in the future, industries will be able to get the exact amount of stock they need using accurate predictions

REFERENCES

[1]. International Journal of Engineering Research & Technology (IJERT), Vol. 9 Issue 06, June-2020- "Inventory Management using Machine Learning".

[2]. DONG Hai, ZHENG Hao -" Model Predictive Control for inventory Management in Supply Chain Planning".

[3]. Mahuya Deb, Prabjot Kaur and Kandarpa Kumar Sarma, "Inventory Control Using Fuzzy-Aided Decision Support System", Advances in Computer and Computational Science, Vol 554, Springer, Singapore, 2017, pp. 549-557.

[4]. Yan Tang Demey, and Mikael Wolff -
"SIMISS: A Model based Searching Strategy for Inventory Management Systems ".

[5]. Isabel Fernandez, Paula Chanfreut , Isabel Jurado and Jose Maria Maestre-"A Data-based Model Predictive Decision Support System for Inventory Management in Hospitals ".