e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 7, July 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707059 |

Beyond Spreadsheets: A Machine Learning Framework for Intelligent Inventory Optimization

Shylesh B C, Suma

Assistant Professor, Department of MCA, Mangalore Institute of Technology & Engineering College, Moodbidre, Karnataka, India

PG Student, Department of MCA, Mangalore Institute of Technology & Engineering College, Moodbidre, Karnataka, India

ABSTRACT: Managing inventory has long been a challenge for businesses of all sizes. Maintaining accurate information regarding different goods are typically a laborious task. Spreadsheets are a common tool used by small organizations to store data in their databases, but they can be corrupted or lost. Managing inventories is crucial. for businesses since it provides a greater variety of production techniques Using a collection of client shopping data, this research creates predictive models to precisely estimate sales prices in the e-fashion industry. The system uses a multiple analytics, such as linear regression and further ML techniques, to provide details regarding the latest profitable, un- interesting, and selling stocks. By providing an all encompassing a resolution for them, this strategy aids inventory managers in optimizing their functioning.

KEYWORDS: LR, RF regression,Decision Tree Regressor, Lasso Regression, Ridge Regression.

I. INTRODUCTION

Effective track of inventory is essential for maintaining operational effectiveness and satisfying consumer needs in the fast-paced corporate climate of today. For retail and ecommerce enterprises, the capacity to precisely forecast sales prices is essential to maximizing inventory levels, pricing policies, and total profitability. The primary objective of this research is to develop and evaluate predictive models that utilize historical data from an online fashion retailer spanning three years to forecast sales prices. Numerous factors, including product categories, sales volumes, geographic locations, and seasonal trends, are included in the information and have an impact on pricing dynamics. This study's main goal is to improve forecasting accuracy by utilizing machine learning techniques. This will help firms make well-informed decisions on inventory stocking and pricing strategies. The procedure includes thorough data preprocessing, the creation of several regression models, and exploratory data analysis (EDA) to find patterns and connections. The accuracy with which these models Reck Regression, LR, Decision Tree, RF regreesor, and Linear Regression can forecast sales prices is the basis for their evaluation. This study will evaluate several methods' predictive ability using measures like the R- squared score and RMSE for the purpose to determine. That have more beneficial technic for predicting sales prices in the context of managing inventories.

The results aid in improving pricing and inventory optimization decision- making processes, which in turn helps companies gain a competitive edge and experience long-term success in the retail industry. Businesses can improve pricing strategies, optimize inventory levels, and obtain insights into customer purchase behavior by utilizing predictive modeling techniques on datasets such as "customer shopping data." This research helps with strategic decision- making related to stock restocking, marketing campaign planning, and overall firm profitability in addition to facilitating accurate sales for ecasting. Businesses can strategically bundle products and improve shelf placements to increase sales and customer satisfaction by recognizing product linkages and customer preferences. By forecasting future sales trends based on historical data, the utilization of linear regression gives the system for managing inventory a new perspective Businesses can more precisely predict demand because to this predictive capabilities, which optimizes inventory levels and lowers carrying costs. Furthermore, incorporating an simple to use UI that is accessible through mobile apps. improves usability and guarantees decision-makers have access to real-time data, which is essential in the fast-paced business world of today. The purpose of this work is to create a comprehensive inventory handling system that uses cutting- edge data analysis techniques to produce insightful findings in addition to automating transaction recording and monitoring of stocks. The suggested solution aims to increase organizational productivity and profitability by lowering manual involvement, decreasing errors, and enhancing operational efficiency. The system's dedication to updating inventory procedures while maintaining dependability and scalability is emphasized by management. In

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707059 |

conclusion, a major step toward resolving the difficulties associated with traditional inventory management is the implementation of a digitalized inventory administration framework that utilizes linear regression. Businesses can attain sustainable growth and a competitive edge in the dynamic marketplace of today by utilizing technology to optimize operations and capitalizing on data analytics.

II. RELATED WORK

An Android app designed for Indian retailers specifically for Tandel et al.'s case study[1] focused on sales forecasting and handling inventory (2020). The study highlights how adding sales prediction algorithms into mobile applications may help owners of smaller enterprises with managing their inventories. In a similarly Sheth et al[2]. (2020) provided a well-thought-out procedure for methodical inventory man- agement, stressing the significance of rationalizing inventory procedures via systematic approaches.By putting in place organized and effective inventory procedures, their research seeks to maximize inventory control. Khurana and Kumar[3] (2017) delved into data analytics in inventory management, exploring how data-driven techniques can enhance inventory management practices. The impact of data analytics on inventory management was explored, and it was stated that ways of analysis might greatly enhance decision-making and inventory control. In a different investigation, Setiawan et al.(2017) applied data mining techniques, specifically Market Basket Analysis, to a sales information system in a stationery company. Their research focuses on identifying patterns in customer purchases to optimize sales strategies and inventory management.

Nemtajela et al. (2016) looked into outcomes of different approaches to inventory control on erratic demand. Their study highlights the requirement for reliable inventory models To be able to manage erratic variations in demand. Cheriyan et al. (2018) investigated the utilization of ML approaches for competent sales predicting. Their study shows how ML algorithms can become accustomed to forecast sales, giving inventory planners an advantage and revealing patterns for future sales. De Santis et al.[7] (2017) studied the previous information. ML for managing inventories item back ordered forecasting. Their research focuses on developing predictive models to foresee backorders, thereby improving inventory management by mitigating the exposure of stockouts. Liu et al[8]. (2012) researched a component design method of inventory management based on business models and cluster algorithms, providing a novel approach to inventory management by integrating business models with clustering techniques.

Maheswari et al. (2017) focused on utilizing ML algorithm for analysis of client behavior and employed SVM classifiers to forecast online shopping behavior. Using machine learning approaches, Krishna et al. (2018) published a paper on pre- dicting sales of one retail store, highlighting the significance of precise sales projections in inventory management. Their study presents a range of machine ML techniques to improve management of inventory and boost sales forecast accuracy by two. Random Forest is utilized by Krishna et al. [10] for sales forecasting in retail stores, achieving a notable accuracy of 93 These studies collectively add to the field of inventory management and sales prediction, employing diverse method- ologies ranging Insights of data and ML to systematic process optimization and clustering algorithms. The incorporation of these advanced techniques in inventory management systems illustrates substantial opportunity to improve inventory control, forecasting accuracy, and overall efficiency in various retail and industrial settings.

III. METHODOLOGY

Accurately predicting sales prices is essential for optimizing controlling inventory and strategic planning in the e-fashion retail industry. This research uses forefront ML methods to create robust predictive models using a detailed dataset (customer Shopping data.csv). The dataset includes various features such as Category, City, Type of items, Quantity, Quarter, Sales Revenue, SKU description, Week, and Year, with the primary goal being the estimation of retail prices. Our methodology Includes 5 key phases aimed at ensuring data quality and model reliability: Data Preprocessing, Exploratory Data Analysis (EDA),Data Splitting, Model Development, and Performance Evaluation. Each phase is carefully crafted to increase predicted classifiers' accuracy and generalizability.

A. Data Preprocessing

The dataset utilized here study, customer shopping data. csv, includes more features including Category, City, Type of items, Quantity, Quarter, Sales Revenue, SKU description, Week, and Year, with the target variable for prediction being price. Using Label Encoder from sklearn preprocessing, data preprocessing stages involved transforming categorized characteristics into numeric numbers. In order for ML algorithms to process and learn from the categorical data, transformation is necessary. Additional preprocessing steps included handling missing values, which were either filled

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707059 |

using the mean, median, or mode for the corresponding feature, or removed if The percent of missing values was significant. Outliers were detected and managed using statistical methods to prevent skewing the model's performance. Features were also normalized or standardized to guarantee that each factor serves equally to the model, especially for distance-based algorithms.

B. Data Cleaning

Initially,a Pandas DataFrame containing the dataset was loaded for inspection. In order To promise the authenticity of the analysis, the initial phase in data cleaning was to remove rows containing null entries To be able to resolve missing values. Additionally, duplicate records were found and eliminated to prevent duplication and distorted findings For the sake of the data types would work with the analyses that came after, they were checked and corrected. To enable precise temporal analysis, for instance, date-related data were transformed to the proper date-time formats.

C. Exploratory Data Analysis

Before model application, an Exploratory Data Analysis (EDA) was conducted to understand the dataset and uncover patterns. EDA involved summarizing key data characteristics, often using visual methods. A key aspect of EDA was generating and analyzing the The value of the coefficient of association among pairs of features. Comprehending these associations is essential for constructing efficacious prediction models. Connect to the correlation matrix, various plots such as histograms, box plots, and scatter plots were created to the allocation and relationships of different features. This helped in identifying any skewness, trends, and anomalies within the data. Furthermore, advanced visualization techniques like pair plots and interactive dashboards were used to provide more all-encompassing perspective of the details dynamics.

D. Data Splitting

The instruction and evaluation subsets of the dataset were separated using train test split from sklearn model selection So that thoroughly assess the model's performance. For training, 80% of it was utilized, and the remaining 20% was put to use for testing. With this split, we can effectively evaluate our models' ability to generalize to new data and obtain a trustworthy approximation of their effectiveness in practical situations. To further ensure robust evaluation, cross- validation tools like k-fold cross validation were employed. Using a distinct subset as the test created and the remaining subset as the training set, this method splits The details entered into k subsets and trains the model k times. This procedure yields a more accurate evaluation of the model's performance and helps to mitigate the variability caused by data splitting.

E. Model Development

Many regression models were created And to guarantee that to forecast the target variable, this is the price. Among these models were the Random Forest (RF) Regressor, the decision trees Regressor, Lasso Regression, Ridge Regression, and Linear Regression. To ascertain its predictive accuracy, every approach was tested on a different subset after being trained on the training subset.

1. **Linear Regression**: To start with, a baseline model for price prediction was created using LR techniques. The independent variables and the target variable in this model are assumed to possess a linear relationship. The R-squared score and RMSE were used to assess performance.
2. **Random Forest Regressor**: An ensemble learning method used to capture non-linear relationships. It builds multiple decision trees and merges their outputs. For this study, 250 estimators were used, and performance was assessed utilizing the R-squared score and RMSE.
3. **Decision Tree Regressor**: A decision tree is a non-linear model that builds a tree structure by dividing data into groups according to feature values. Despite their ease of interpretation and visualization, decision trees are prone to overfitting, which is why ensemble approaches such as Random Forest(RF) are preferred .The significance and relationships of features were understood by using this model. Data is divided into subgroups according to feature values. To avoid overfitting, hyperparameters including maximum leaf nodes, minimum samples split, and maximum depth were adjusted. The RMSE and R-squared score were used to assess performance.
4. **Lasso Regression**: Applied for feature selection and penalizing the magnitude of coefficients, improving model generalization. The penalty term, with the direction of the parameter alpha, helps shrink coefficients of less important features to zero, preventing overfitting and enhancing predictive performance.
5. **Ridge Regression**: employed to add a penalty term, defined by the parameter alpha, to the loss function To be able to handle multicollinearity among features. This contributes to increased stability and a decrease in model variation.

The RMSE and R-squared score were used to assess performance. For each model, hyperparameter adjustment was done to maximize performance. The accuracy and resilience of the prototype were improved by methodically

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

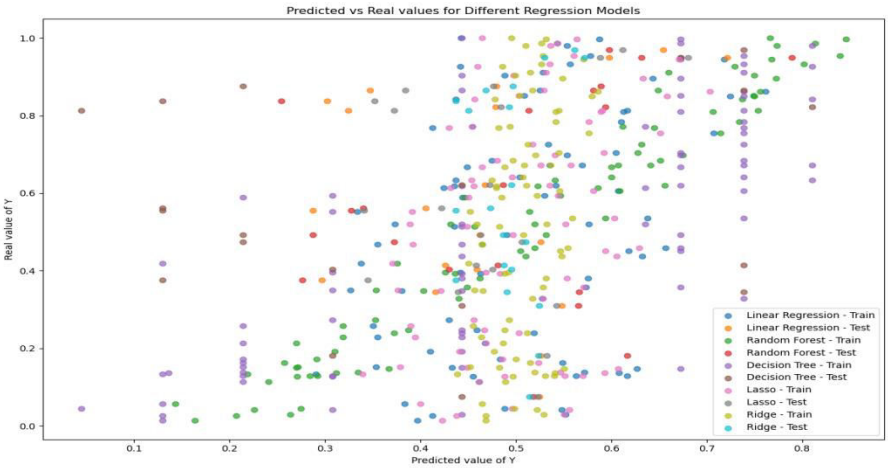
| DOI:10.15680/IJMRSET.2024.0707059 |

searching for the ideal parameters using strategies like Random and Grid Search.

F. Performance Evaluation

The R-squared score and RMSE were used to assess each model's performance. While the RMSE shows the average amount of prediction errors, the R-squared score quantifies the percentage of variance in the target variable explained by the model. To guarantee that models fit training data well and generalized to previously unseen data as well, testing and training datasets were employed. Models were deemed additional competent whenever they had them. lower RMSE values and higher R-squared scores. A thorough appraisal of the model's performance was also provided by the calculation of additional evaluation metrics, such as Mean Absolute Error (MAE) and Mean Squared Error (MSE).

IV. RESULTS AND DISCUSSION




```
Linear Regression - Training R2 Score: 11.45, Test R2 Score: -29.89
Random Forest - Training R2 Score: 83.60, Test R2 Score: -11.52
Decision Tree - Training R2 Score: 55.67, Test R2 Score: -95.08
Lasso - Training R2 Score: 10.71, Test R2 Score: -22.24
Ridge - Training R2 Score: 6.23, Test R2 Score: -13.14
Linear Regression - RMSE: 0.31
Random Forest - RMSE: 0.28
Decision Tree - RMSE: 0.37
Lasso - RMSE: 0.30
Ridge - RMSE: 0.28
```

Fig. 1. Predicted scores for various models of regression

The RF Regressor performed the best in opposition to other regression frameworks Within the analysis of e-fashion store sales price prediction models. This model, in contrast to the others, has the greatest test R-squared score as well as a weak RMSE The RF Regressor's ability to explain a significant amount of the variance in the test data is demonstrated by its high R-squared score and low RMSE, which also demonstrate The way its forecasts line up the actual values. Its high predictive accuracy and good generalization to new data are demonstrated by these metrics. The model's ensemble learning method, which mixes several decision trees to identify intricate patterns in the data and reduce overfitting while enhancing prediction performance, is responsible for its success. Because of this, the Random Forest Regressor is an excellent option for precise and trustworthy sales forecasting in the context of an online fashion retailer.

V. CONCLUSION

Through extensive data preprocessing and EDA, the dataset was transformed and analyzed to uncover important insights and get it ready for modeling. A variety of regression equations, including LR regression, RF Regressor, Decision Leaf Regressor, Lasso Regression, and Ridge Regression, were applied and evaluated based on R-squared score and RMSE metrics. The The intention of this study was to predict sales prices in an e-fashion store using ML techniques. The approach that proved to be the most successful was the RF Regressor, which exhibited exceptional prediction accuracy and resilience. In forecasting sales prices, its capacity to manage nonlinearities and feature interactions proved useful. For e-commerce stakeholders These discoveries provide useful information that will assist

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707059 |

them in making better decisions about pricing and inventory management. Subsequent investigations may augment prognostic capacities by integrating supplementary data sources or optimizing model parameters to accommodate fluctuating market circumstances. Leveraging a ML models— RF in particular Regression, represents a note- worthy progression in the prognostication of e-fashion product sales prices, hence enabling enhanced operational efficiency and profitability within retail settings.

REFERENCES

1. T.Tandel, S. Wagal, N. Singh, R. Chaudhari and V. Badgujar, ”Case Study on an Conferences on Modern Computing and Interaction De- vices, International with Sales Prediction for Local Shopkeepers in India,” 2020 ICACCS, Coimbatore, India, 2020
2. R. Sheth, M. Vora, R. Sharma, M. Thaker and P. Bhavathankar, ”A Proficient Process for Systematic Inventory Management,” 2020 INCET, Belgaum, India, 2020.
3. M. Khurana and D. Kumar, ”The study of data analytics in inventory management,” 2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) ICTUS, Dubai, United Arab Emirates, 2017
4. Setiawan, G. S. Budhi, D. H. Setiabudi and R. Djunaidy, ”Data Mining Applications for Sales Information System Using Market Basket Anal- ysis on Stationery Company,” 2017 ICSIIT, Denpasar, Indonesia, 2017.
5. N. Nemtajela and C. Mbohwa, ”Inventory management models and their effects on uncertain demand,” 2016 IEEE IEEM, Bali, Indonesia, 2016.
6. S. Cheriyan, S. Ibrahim, S. Mohanan and S. Treesa, ”Intelligent Sales Prediction Using Machine Learning Techniques,” 2018 International Convention of,Computing, Electronics Communications Engineering (iCCECE), Southend, UK, 2018.
7. R. B. de Santis, E. P. de Aguiar and L. Goliatt, ”Predicting material backorders in inventory management using machine learning,” 2017 IEEE LA-CCI, Arequipa, Peru, 2017
8. S. Liu, S. Liu and L. Luo, ”Research on component design method of inventory management based on business model and cluster algorithm,” ICSSSM12, Shanghai, China, 2012
9. [9] K. Maheswari and P. P. A. Priya, ”Predicting customer behavior in online shopping using SVM classifier,” 2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), Srivilliputtur, India, 2017
10. Krishna, A. V, A. Aich and C. Hegde, ”Sales-forecasting of Retail Stores using ML Techniques,” 2018 3rd International Convention of CSITSS, Bengaluru, India, 2018

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com