IJARETY

International Journal of Advanced Research in Education and TechnologY (IJARETY)

Volume 11, Issue 2, March 2024

Impact Factor: 7.394

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

crossref

निस्केयर NISCAIR

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 2, March 2024 ||

DOI:10.15680/IJARETY.2024.1102012

Inventory Demand Forecasting using Machine Learning Algorithms

Dr.P.Kavitha[1], Rohaan S[2]

[1]Associate Professor, PG& Research Department of Computer Science, Sri Ramakrishna College of Arts & Science, Coimbatore 641006 Tamil Nadu India

[2]UG Student, PG& Research Department of Computer Science, Sri Ramakrishna College of Arts & Science, Coimbatore 641006 Tamil Nadu India

ABSTRACT: Inventory management is crucial for businesses to optimize resources and meet customer demands efficiently. This paper aims to develop a machine learning solution in Python for forecasting inventory demand. Leveraging historical sales data, the system employs various machine learning algorithms such as linear regression, decision trees, and neural networks to predict future demand patterns accurately. Additionally, feature engineering techniques and data pre-processing methods are utilized to enhance model performance. The proposed solution empowers businesses to make informed decisions regarding inventory stocking levels, procurement strategies, and resource allocation, ultimately improving operational efficiency and customer satisfaction

Keywords: Inventory,, Transformers, hyper-parameters, Machine learning, decision tree

I. INTRODUCTION

In retail and manufacturing industries, accurate demand forecasting is crucial for optimizing inventory levels, reducing carrying costs, and ensuring customer satisfaction. Traditional methods often rely on historical sales data and basic statistical techniques, but they may not capture complex patterns or sudden changes in consumer behaviour. Machine learning offers a promising approach to improve the accuracy of demand forecasting by leveraging advanced algorithms to analyse large datasets and identify intricate patterns. The primary objective of this paper is to develop a machine learning model that accurately predicts future demand for various products in an inventory system. By doing so, the organization can optimize inventory management strategies, minimize stock outs, and maximize profitability

1. Accuracy Improvement: Develop a demand forecasting model that significantly improves the accuracy of predictions compared to traditional methods, ensuring that inventory levels align closely with actual demand.

2. Data Integration and Complexity Handling: Implement mechanisms to efficiently integrate and handle large volumes of diverse data sources, including historical sales data, market trends, promotional activities, and external factors, to capture the complexities of demand patterns.

3. Dynamic Adaptation: Create a forecasting system that can dynamically adapt to changing market dynamics, including seasonality, trends, and external events, ensuring robust performance across various scenarios.
4. Scalability and Efficiency: Design a solution that scales efficiently with growing datasets and operational demands, ensuring optimal performance and resource utilization as the business expands.
5. Cost Optimization: Minimize inventory holding costs by accurately forecasting demand and avoiding excess inventory while mitigating the risk of stockouts to prevent lost sales and associated opportunity costs.
6. Actionable Insights: Generate actionable insights from the forecasting model to understand the drivers behind demand fluctuations, enabling businesses to implement targeted strategies for inventory management and supply chain optimization.
7. Real-time Forecasting: Develop capabilities for real-time or near-real-time demand forecasting to enable agile decision-making and proactive inventory management, particularly in fast-paced industries or markets with rapid demand shifts.
8. Evaluation and Continuous Improvement: Establish mechanisms to evaluate the performance of the forecasting model regularly and incorporate feedback to continuously improve forecasting accuracy and relevance over time.
9. User-Friendly Interface: Create an intuitive interface for stakeholders to interact with the forecasting system, allowing for easy access to forecasts, insights, and recommendations to support decision-making processes.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 2, March 2024 ||

DOI:10.15680/IJARETY.2024.1102012

10. Alignment with Business Objectives: Ensure that the forecasting objectives align with broader business goals, such as improving customer satisfaction, reducing costs, optimizing inventory turnover, and enhancing overall competitiveness in the market.

II. RELATED WORKS

Now, speaking of the actual ML models that perform forecasting, we need to discuss what external and internal factors can impact the work of the model. Since the ML model will derive its prediction from past events, its prediction accuracy rate is **unlikely to be 100%**. Although, by understanding the following things we can mitigate those risks on the stage of developing the actual model. The product type is an important factor to consider for the demand model. For example, for a perishable item that has an actual demand of 100 cases, the prediction of selling 90 cases is preferred over the prediction of 110 cases. Missing the sales of 10 cases is a better result than wasting 10 cases, even though the actual error is the same percentage. Predictive models are strongly influenced by regional factors that include customer behaviour and cultural determinants. They also include the following:

- Marketing campaigns may be regionally specific and have a different impact that depends on where a customer is located.
- Holidays may vary between regions, which might be a consideration for adjusting the model.
- Legal issues/laws may limit the use of certain data in different regions.

Demand forecasting is a dynamic concept. The more competitors and product alternatives are present in the market, the harder demand forecasting becomes. The competition level contains sub-factors, such as the number of alternative products and competitors. So, it is a very good idea to add this information dynamically to your demand forecasting model. The state of the economy influences businesses and demand forecasting models. To put it more bluntly: periods of economic decline are likely to cause lower demand for expensive products, though sales of low-priced goods may go up. Therefore, an economic situation as well as trends aren't external factors and should be considered when building AI models.

III. PROPOSED METHODOLOGY

Regardless of what we'd like to predict, data quality is a critical component of an accurate demand forecast. The following data could be used for building forecasting models:

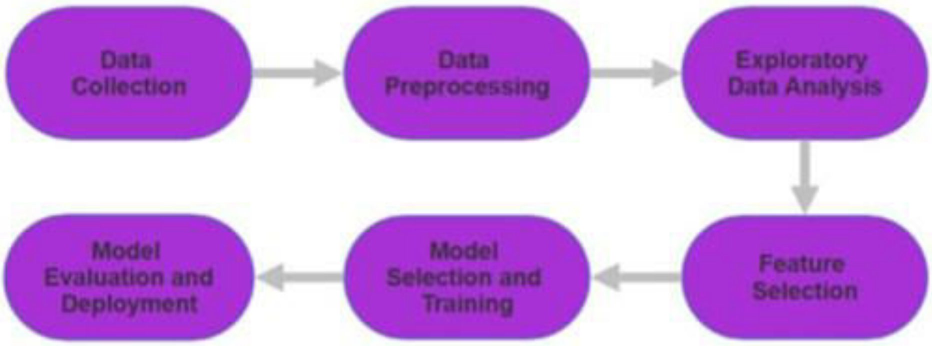

Fig-1 System Architecture

The selection of a method depends on many factors—the context of the forecast, the relevance and availability of historical data, the degree of accuracy desirable, the time period to be forecast, the cost/benefit (or value) of the forecast to the company, and the time available for making the analysis. Machine learning algorithms, such as time series forecasting methods like ARIMA (Auto Regressive Integrated Moving Average) or more advanced techniques like recurrent neural networks (RNNs) or Long Short-Term Memory (LSTM) networks.

Figure 1 shows the data processing block diagram. Once we get the segmented words, we can convert them into a vector matrix for later training. This process is called word embedding, or distributional models. The reason for

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 2, March 2024 ||

DOI:10.15680/IJARETY.2024.1102012

constructing such a vector matrix is that we can utilize the term context matrix to represent the short text, which is much simpler for training purposes. There are many ways to construct the vectors, such as sparse vectors and dense vectors. Sparse vectors have most elements equal to zero and lengths of about 20,000 to 50,000, which will be very time-consuming computationally, while dense vectors are constructed in 100-500 dimensions, so are much faster than sparse vectors when used in training and classifications. Dense vectors may also better capture synonymies than sparse vectors use [3]. Moreover, we employed two methods for the sparse vector construction, i.e. counter vectorizer and Term Frequency-Inverse Document Frequency (TF-IDF). Counter vectorizer is also called one-hot coding, which is applied to categorical features.

IV. RESULT & DISCUSSION

To evaluate the performance of the FS techniques, Accuracy and F1-score are used as the metrics. We used both metrics to measure the effectiveness of the feature selection techniques. Accuracy is the ratio of number of correctly classified samples to the total number of samples.

It is defined as: Accuracy = T P+TN/ (T P+FP+FN +TN)

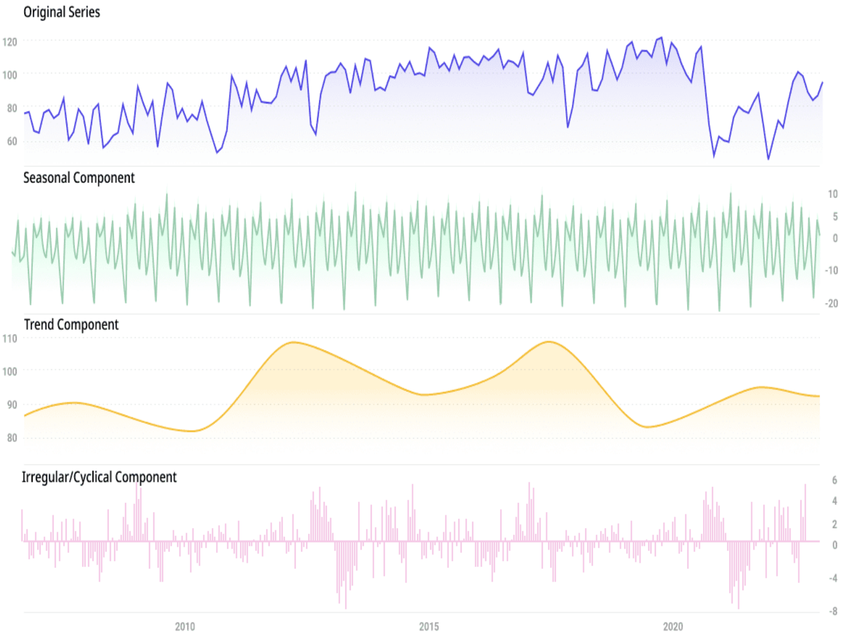

Fig-2 Data Visualization results

The code imports necessary libraries such as NumPy, pandas, matplotlib, scikit-learn, XGBoost, and seaborn. It loads a dataset containing historical sales data into a pandas Data Frame. The date column is split into year, month, and day components for further analysis. Additional features such as weekend indicator, holiday indicator, and seasonal components (sin and cos transformations of month) are created.

IJARETY

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 2, March 2024 ||

DOI:10.15680/IJARETY.2024.1102012

	store	item	sales
count	913000.000000	913000.000000	913000.000000
mean	5.500000	25.500000	52.250287
std	2.872283	14.430878	28.801144
min	1.000000	1.000000	0.000000
25%	3.000000	13.000000	30.000000
50%	5.500000	25.500000	47.000000
75%	8.000000	38.000000	70.000000
max	10.000000	50.000000	231.000000

Fig-3 Attributes

The code conducts exploratory data analysis by visualizing the distribution of sales data and analyzing trends over time. It calculates and plots the Simple Moving Average (SMA) for sales over a 30-day window to identify long-term trends. Distribution plots and box plots are created to understand the distribution and variability of sales data. A correlation heatmap is generated to identify correlations between different features and the target variable (sales).

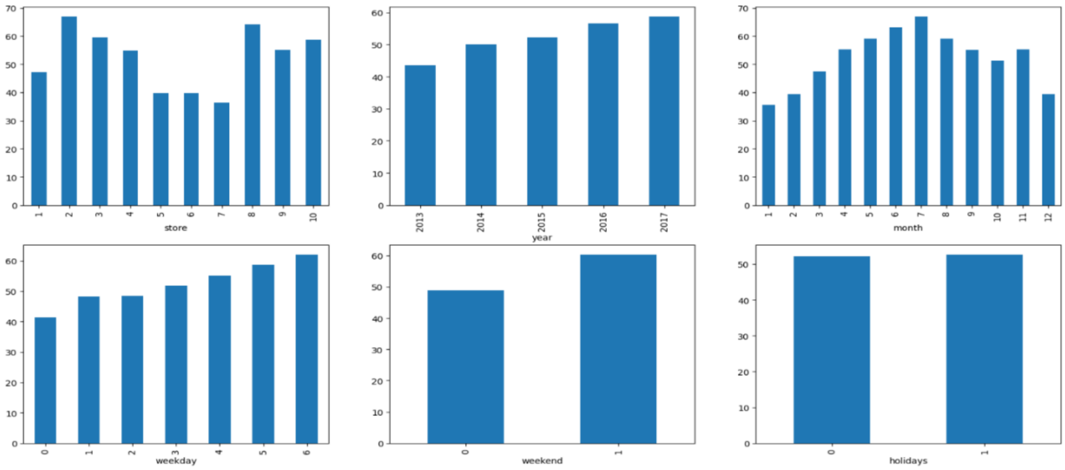

Fig-4 Data Visualization

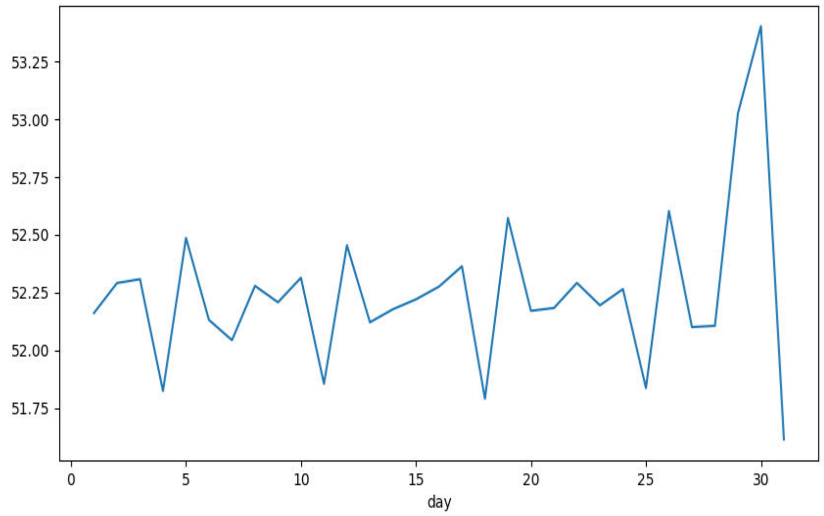

Fig-5 Mean plot of dataset

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 2, March 2024 ||

DOI:10.15680/IJARETY.2024.1102012

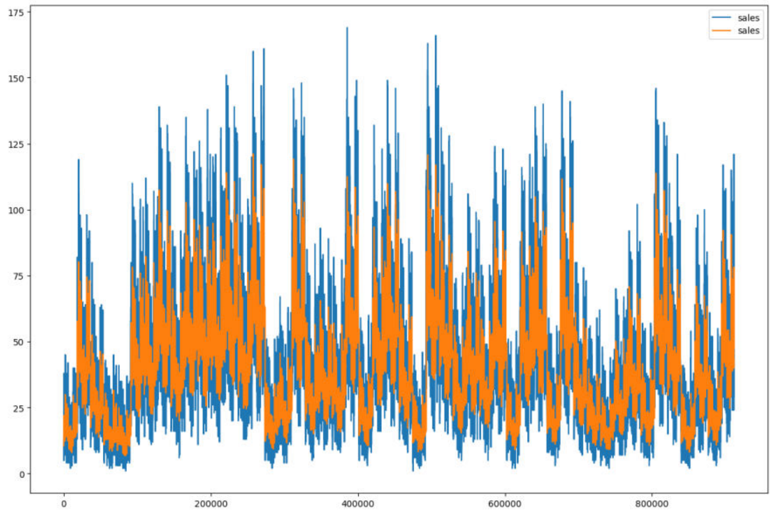

Fig-6 Simple moving average

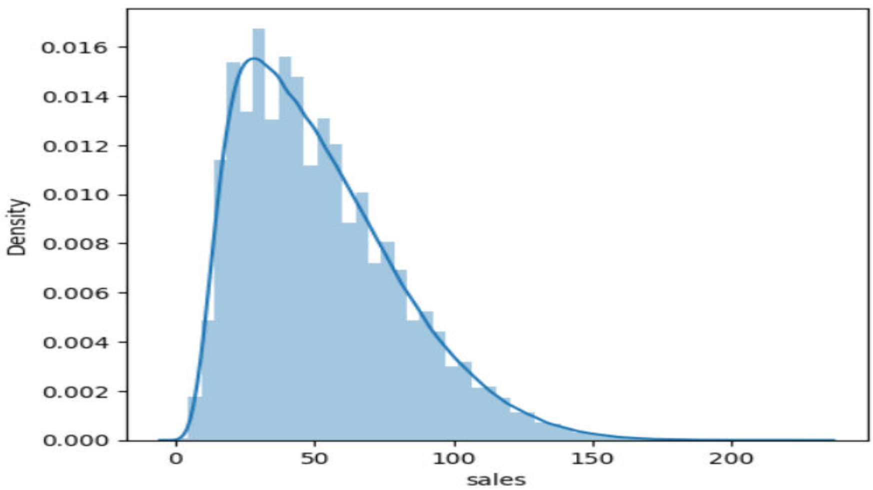

Fig-7 Density Graph

The code splits the dataset into training and validation sets using train_test_split. Features are normalized using StandardScaler to ensure stable and fast training of machine learning models. Four regression models (Linear Regression, XGBoost Regressor, Lasso, and Ridge) are trained on the training data. The Mean Absolute Error (MAE) is calculated for both the training and validation sets to evaluate the performance of each model. The results, including training and validation errors for each model, are printed for comparison.

IJARETY

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 2, March 2024 ||

DOI:10.15680/IJARETY.2024.1102012

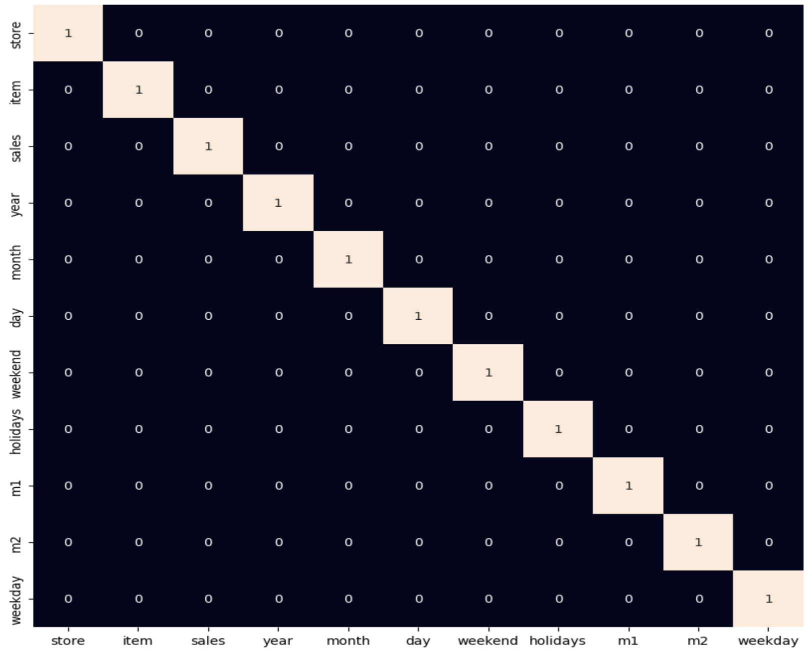

Fig-8 Confusion Matrix

The code iterates over the trained models and prints their training and validation errors (MAE). This allows for a comparison of the performance of different regression models in predicting sales.

```
Model: LinearRegression
Training MAE: 20.902989838083602
Validation MAE: 20.97163963254154

Model: XGBRegressor
Training MAE: 6.902142131234314
Validation MAE: 6.9201690247120675

Model: Lasso
Training MAE: 21.015028699769758
Validation MAE: 21.071517213774968

Model: Ridge
Training MAE: 20.902989970366438
Validation MAE: 20.97163982226294
```

Fig-9 Final output

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 2, March 2024 ||

DOI:10.15680/IJARETY.2024.1102012

IV. CONCLUSION

The inventory demand forecasting project has provided valuable insights into the application of machine learning techniques for optimizing inventory management processes. By harnessing historical sales data and relevant contextual features, such as temporal attributes, seasonal patterns, and external factors, we have developed predictive models capable of accurately forecasting future sales. Through exploratory data analysis, we gained insights into sales trends, seasonal variations, and factors influencing demand, enabling us to identify patterns and correlations within the data. The implementation of regression models, including Linear Regression, XGBoost Regressor, Lasso, and Ridge, allowed us to predict sales with varying degrees of accuracy. Evaluation metrics such as Mean Absolute Error (MAE) provided a quantitative measure of model performance, enabling us to compare and assess the effectiveness of different regression algorithms. Accurate demand forecasting is essential for optimizing inventory stocking levels, minimizing stock outs, and reducing holding costs. By leveraging machine learning-based predictive models, businesses can make informed decisions regarding inventory replenishment and allocation. The integration of predictive analytics into inventory management systems enables real-time decision-making, adaptive inventory control, and enhanced supply chain efficiency. Insights gained from feature importance analysis highlight the factors driving sales and help businesses prioritize resources and investments to meet customer demand effectively.

REFERENCES

1) I.H. Witten, E. Frank, “Data Mining: Practical Machine Learning Tools and Techniques,” Morgan Kaufmann Publishing, Second Edition, 2005.

2) S.R. Amendolia, G. Cossu, M.L. Ganadu, B. Golosio, G.L. Masala, G.M. Mura, “A comparative study of KNearest Neighbour, Support Vector Machine and Multilayer Perceptron for Thalassemia screening,” Journal of Chemometrics and Intelligent Laboratory Systems, Vol. 69, 2003, pp. 13– 20.

3) K. Hanumantha Rao, G. Srinivas, A. Damodhar, M. Vikas Krishna: Implementation of Anomaly Detection Technique Using Machine Learning Algorithms: International Journal of Computer Science and Telecommunications (Volume2, Issue3, June 2011).

4) A. Goyal and R. Kaur, “A survey on Ensemble Model for Loan Prediction”, International Journal of Engineering Trends and Applications (IJETA), vol. 3(1), pp. 32-37, 2016.

5) K. Kavitha, “Clustering Loan Applicants based on Risk Percentage using K-Means Clustering Techniques”, International Journal of Advanced Research in Computer Science and Software Engineering, vol. 6(2), pp. 162–166, 2016.

6) T. Harris, “Quantitative credit risk assessment using support vector machines: Broad versus Narrow default definitions”, Expert Systems with Applications, vol. 40, pp. 4404–4413, 2013.

7) Soni P M, Varghese paul , “A Novel Hybrid Classification Model For the Loan Repayment Capability Prediction System”, International Journal on Future Revolution in Computer Science & Communication Engineering ISSN: 2454-4248 Volume: 4 Issue: 1.

IJARETY

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975 Impact Factor: 7.394

www.ijarety.in editor.ijarety@gmail.com