WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Retail Sales Forecasting Using Advanced Time Series Regression Techniques

Challa Keerthana[1], Takkedu Malathi[2]

[1]PG Scholar, Department of MCA, [2]Assistant Professor & Head, Department of MCA
Aurora Deemed to be University, Hyderabad-500098 , Telangana ,India.
Email: Keerthanachalla17@gmail.com

A B S T R A C T

Accurate forecast of retail sales is essential for effective business operations, inventory management and strategic schemes. This paper presents a desktop-based forecast application developed in the paper python that takes advantage of advanced time chain regression techniques with focus on XGBoost Regressor . To capture complex data patterns in the system and to improve future performance , include automatic feature engineering methods including lag features, moving averages and seasonal trend decomposition. A user-friendly graphical interface, designed with TKINTER, allows end-users to upload datasets in CSV format, do searching data analysis and easily imagine the sale trends. The application evaluates models performance using standard matrix using standard matrix using standard matrix, which ensures reliability of predictions, which ensures reliability of predictions. Results show that the proposed structure not only enhances the forecast accuracy, but also provides valuable insight to retail analysts and business owners, which is able to make better decisions in areas such as marketing, inventory control and workflow adaptation. This study highlights the practical ability of the machine learning-based regression models to address the challenges of modern retail sales forecasts.

Keywords: Retail sales, Forecasting, Time chain, XGBoost, Regression, Sales prediction, Data Analysis, Feature engineering, Machine Learning.

1. INTRODUCTION

Accurate retail sales forecasts have long been recognized as an essential factor in management, adaptation of operations and marketing strategies. Traditional techniques such as moving average and exponential smallest, however, were unable to catch modern retail challenges, including seasonal, non-lectural patterns and external factors such as holidays and propaganda. This limit created the need for more advanced forecasting methods capable of addressing these complications.

In this study, a forecast structure was developed and evaluated to bridge these intervals. This approach was implemented as a Python-based desktop application, which employed XGBoost algorithm , a machine learning model that is well suited for non-lectured and complex data patterns modeling. To strengthen the future stating performance, the study implemented feature engineering strategies such as LAG features, moving averages and seasonal decomposition, which enabled the model to extract a meaningful pattern from historical sales data.

A user -friendly graphical interface was designed using Tkinter, allowing users to import, imagine sales trends, and create actionable forecasts. The model accuracy was evaluated through the installed evaluation metrics, including the meaning of meaningless error (MAE), the medium paid error (MSE), the root mean square error (RMSE), and the R-Squared (ROT), which confirm the strength and reliability of the forecast.

The results showed that the proposed structure improved the forecast accuracy on traditional methods. By providing a practical tool for retail analysts and business owners, this work supports better inventory management, reduces stock -related damage, and increases business strategies. Overall, this shows the ability of machine learning-based regression techniques to make the forecast of retail sales more accurate and data-operated.

2. LITERATURE SURVEY

Most of the current work done in retail sales forecasting focuses on the use of machine learning methodologies, especially algorithms like XGBoost and Random Forest, which have been found to be more accurate than classical methods like linear regression or ARIMA. These methods are better suited to handling large datasets and are able to ascertain patterns in consumer behavior that are often cumbersome to ascertain with traditional models. Not surprisingly, they have rapidly garnered a lot of interest among researchers involved in retail sales forecasting because of their potential for providing increased accuracy and insight to assist with the planning and management of stock.

K-Means clustering is another recent approach in the literature, creating product lines according to the performance in sales. With this, retailers will be able to tell information about slow-moving versus fast-moving products, which would help them better manage inventory. Pattern analysis, most commonly Most Frequent Pattern (MFP) Analysis, is also used to identify consumer behavior with regard to purchase and strategize towards targeted marketing efforts to improve performance.

Lead time plays an important role in sales forecasting, especially for perishables. Research shows that models such as SARIMA are effective in capturing seasonal changes in demand, allowing retailers to respond to seasonal changes in sales. However, these traditional models struggle with complex interactions between components, which can be effectively addressed with advanced machine learning algorithms.

The impact of the COVID-19 pandemic has also been a major factor in sales forecasts. Research showed that retail markets responded differently to the epidemic and that forecasting methods are needed to account for this unprecedented change in demand. The crisis highlighted the importance of forecasting, emphasizing the flexibility of models that can deal with such unique circumstances and changes.

Recent research re-emphasizes the importance of incorporating external factors such as economic conditions, climate, and social factors into forecasting models in influencing both internal and external retail variables; the use of these factors will incorporate modern forecasting techniques that are needed to improve accuracy and resolution.

2.1 Important study and contributions:

Many researchers have contributed to the retail sales forecast area by keeping machine learning and statistical approach.

Sujawal et al. (2021) compared models such as linear regression, arima, LSTM and random forests, analyzed the future and saw that XGBoost always performed the best in terms of RMSE and MAE. He highlighted the importance of being large data set with strong algorithms to achieve reliable forecasts.

Mansoori and Pawar (2021) also proposed a hybrid method in the same way, combining one of which combines with the most frequent pattern (MFP) algorithm, which enables inventory management and marketing regarding purchasing behavior modeling.

The integration of spatial data in the forecast was discovered by **Cruce-Troids et al. (2020),** who demonstrated that Vector Regression (SVR) improved the Huff model to predict sales in new retail branches. His study showed how spatial data can provide deep insight into mining market proliferation and resource allocation. **Jain et al. (2018)** also implemented XGBoost to retail forecast, especially within the romantic pharmacy series, using various datasets including daily sales, customer numbers and promotional activities. His findings confirmed the strength of modern machine learning methods for retail forecasting.

Rao et al further studied the approach to learning the dress. (2019), who compared algorithms such as K-nn, multinational regression, decision tree and adaboost. He concluded that XGBoost gained the highest accuracy and showed the ability to improve business decisions in retail. Seasanity by **Gupturu (2019)** was another important factor, which showed that the Winter-Holt and Sirima model performed the best in the forecast for small-scale food retailers, especially for poor accessories.

External disruptions like Covid -19 epidemic have also been investigated. **Kim et al.** , He suggested that machine learning techniques can provide more flexible and adaptive prognosis in such scenarios. The **YI (2020)** focused on the forecast of Walmart's sales and displayed that random forest models exclusively performed better than linear and Laso regression during the discharge period. Finally, **Diao et al. (2019)** The cloud system detected the attire forecast algorithm for intelligent resource scaling, although their work targeted the IT infrastructure, the functioning was noted to be transferable to retail forecast, especially to handle high-load conditions and skilled resource allocation.

Collectively, these studies highlight the boundaries of traditional forecast methods, while machine learning techniques -especially XGBoost, Random Forest and XGBoost offer dressed models such as models like better accuracy and adaptability. They also emphasize the importance of considering seasonal, spatial factors and external disruption in the manufacture of more strong retail sales forecast systems.

Progressions towards modern machine-learning approaches and data mining techniques have been seen in the literature as regards retail sales forecasting. Such new methods can improve the accuracy of their predictions, do better in their inventory management, and improve their overall decision-making. However, a gap still remained concerning the integration of feature engineering and external factors into a scalable algorithm, thus creating a unified forecasting model. This study aims to recognize and close the gaps towards feature engineering integrated with XGBoost , casting the forecasting model more robust and flexible to genie adaptation for changes in retail. That information is going to convert only into incremented business environments through business opportunities provided by improved inventory management and strategic planning.

3. METHODOLOGY

The functioning adopted on this look at accompanied a based technique starting with dataset coaching, preprocessing and function engineering, observed via searching information evaluation (EDA), version development and evaluation. A retail dataset containing sales transactions changed into first cleaned and changed to make sure continuity and reliability. Advanced feature engineering techniques, which includes lag functions, moving averages and annual trends, were implemented to capture brief styles. The search information evaluation changed into then carried out to identify seasonal, trends and discrepancies inside income statistics. For the forecast, XGBoost Regressor was chosen due to strong overall performance on structured time-series facts and ability to deal with non-lectured family members. The model changed into educated, tuned and evaluated the usage of numerous overall performance

metrics to ensure strengthening of the model. Finally, a Tkinter-primarily based graphical person interface (GUI) was advanced to brid down technical implementation with sensible appropriateness, permitting retail managers to without difficulty believe statistics and generate correct forecasts.

3.1 Dataset Description

The dataset of this study was taken from the Kaggle site; saving occurred on a CSV and access through a custom Tkinter-based GUI. Pre-processing ensured their suitability for sampling.

(Dataset Link) ***Dataset***

The dataset for this has been modified into and derived from retail surroundings and blanketed transactional facts with multiple attributes essential for records of sales patterns.

The key attributes have been:

Order Date **-** Representing transaction dates, used for time-based total assessment and fashion identity.

Sales Amount **-** Denoting the general profits for every transaction, serving as the number one aim variable for forecasting.

Product Category **-** Classifying merchandise into education, allowing segmentation and deeper exploration of income conduct throughout groups. Additional attributes on the side of Customer ID, Region, and Shipping Mode provided similar granularity to the dataset, assisting segmentation and specific exploration.

3.2 Data Preprocessing

The data set was saved in CSV format and accessed through a customized Tkinter-based GUI. To ensure that the data were suitable for sampling, preprocessing was performed:

Date Conversion **-** Converted "order date" to datetime format to make time analysis easier.

Aggregation **-** To eliminate noise and emphasize long-term trends, daily sales were aggregated as monthly totals.

Data Cleanup **-** Missing, duplicate, or invalid information was addressed to improve the integrity and underlying reliability of the data .This initial step ensured that the data set was clean, accurate, and organized for analysis and model building.

Feature Engineering **-**The model's predictive capacity was greatly improved through feature engineering.

The following features were generated:

Lag Features **-** These captured temporal dependencies by producing a sales value even with lagged sales (e.g., Lag_1, Lag_2), revealing how previous sales affect subsequent values.

Moving Averages **-** Rolling averages over three months, smoothing short-term fluctuations as indications of underlying trends.

Yearly Trend **-** A feature that quantified sales progressions over a long period, useful for modelling seasonality and increase magnitudes per year. This brings a new depth to the data for cheaply training the models to learn deeper and understand more complex patterns in the data.

3.3 Evaluation Analysis (EDA)

A complete exploratory analysis was performed to perceive hidden styles and relationships in the statistics.

Key insights included:

Sales developments through the years **-** The line plots showed overall income traits, seasonal variations, and ability anomalies.

Lag characteristics analysis **-**The scatterplots confirmed the connection between lagged income and contemporary sales and indicated its importance for prediction.

Moving average visualization**-** Line graphs displaying rolling averages smoothed out the noise and furnished a clean image of developments.

Annual tendencies **-** Bar charts and line graphs examined income growth over the years.

Sales distribution **-** The histograms analyzed the frequency of sales distribution and recognized anomalies and skewness.

Time Series Decomposition **-** Sales records changed into divided into fashion, seasonal, and residual additives, imparting insight into time series. This analysis provided the idea for choice and sampling selections.

3.4 Modeling Development

The forecast was made by selecting the XGBoost Regressor as a model mainly because it is a versatile and efficient tool in processing structured data. The steps in developing the model were:

Model preparation **-** The input features were Lag_1, Lag_2, Moving_Avg, and Yearly_Trend. The training and test datasets were created to assess performance.

Hyperparameter tuning **-** The parameters such as learning rate, tree depth, number of estimators, etc. were optimized via grid search that made possible the improved accuracy of the model.

Model Evaluation **-**The model can be evaluated by metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2).

Forecasting- Future sales are predicted by applying the model. This study has been able to generate actionable insights regarding retail planning and decision-making.

3.5 User Interface Development

An interactive Tkinter-primarily based GUI was designed to make certain user-friendly access to the forecasting system. The interface blanketed the following functionalities:

Dataset Upload **-**A characteristic permitting users to add retail income information in CSV format.

Analysis and Visualization-Automated generation of EDA visualizations, allowing users to discover developments and correlations.

Sales Forecasting-Interactive equipment for generating forecasts and visualizing model predictions alongside assessment metrics.The GUI bridged the distance between technical model implementation and sensible usability for retail managers.

3.6 Special Contributions

This study introduced several new features to improve retail sales forecasting:

Advanced Feature Engineering- The addition of sales backlogs, trends, and annual trends added predictive power to the model.

Advanced EDA **-**Time series decomposition and moving averages provided deeper insights into sales patterns.

Customized GUI development-The Tkinter-based interface made the tool more accessible to non-technical users, making it easier to adopt in real-world situations.

Retail-specific optimization **-** Domain relevance and practical value were assessed by adjusting the method to retail data.

XGBoost Functionality **-** Improved predictive accuracy by using advanced XGBoost capabilities to handle nonlinear relationships and missing data.

Proposed Retail Sales Forecasting Dashboard Integrates Advanced Data Preprocessing, Robust Feature Engineering, Machine Learning, Interactive Visualization Leveraging XGBoost's ability to provide a user-friendly GUI, this innovative analytics solution for retail managers allows them to analyze past sales, predict future developments, and make appropriate decisions. The processes and services it offers prevent gaps managing edge shops between data science and profitability, empowering companies to confidently manage dynamic markets.

4. DESIGN

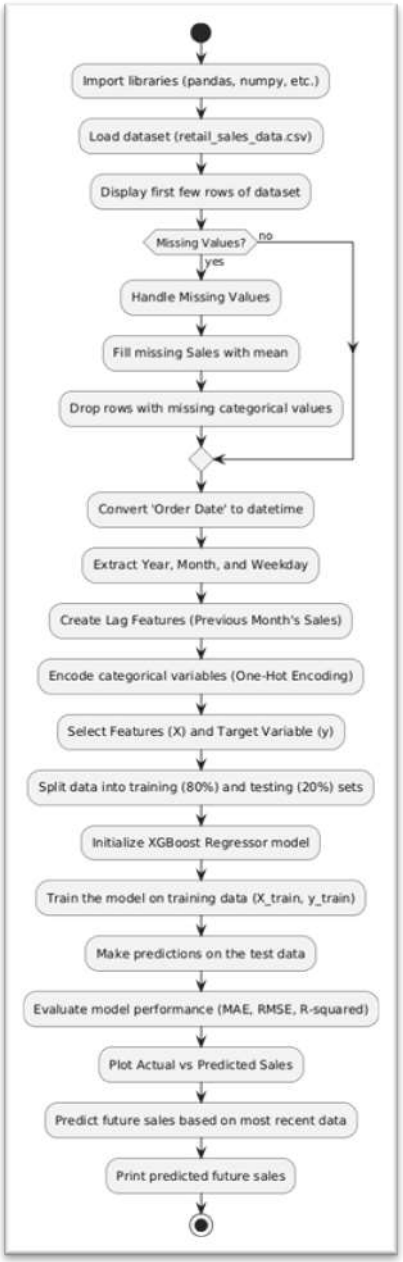

Fig-1: System Design Flowchart of Retail Sales Forecast Application

As shown in Figure 1, the system begins with the import of retail sales data in CSV format. The dataset undergoes preprosying and feature engineering to withdraw lag features, moving averages, and seasonal trends. The processed dataset is then used as an input of XGBoost regression model for training and forecasting. Finally, the graphical user interface (GUI) created using the Tkinter provides the visual insight and forecast results to the final-users.

5. RESULT

The developed system is an interactive retail sales forecast dashboard that enables users to upload, searching and create forecasts through a simple, user -friendly interfaces. The dashboard imagines sales trends, interval features, moving average and seasonal decomposition, while the support assessment also compares real and estimated sales with matrix. This design displays the practical provision of the proposed XGBoost-based forecasting structure, making retailing the technique of learning advanced machine for retail analysts and business decision making.

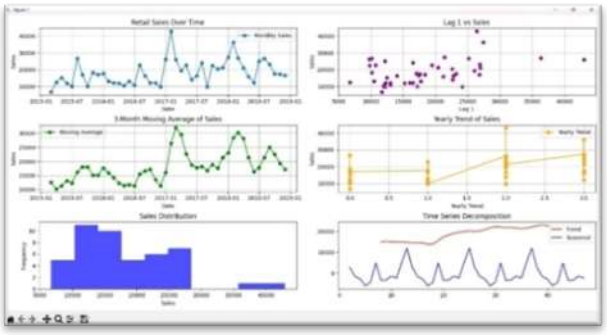

Fig-2 Sales Analysis and Visualization Dashboard

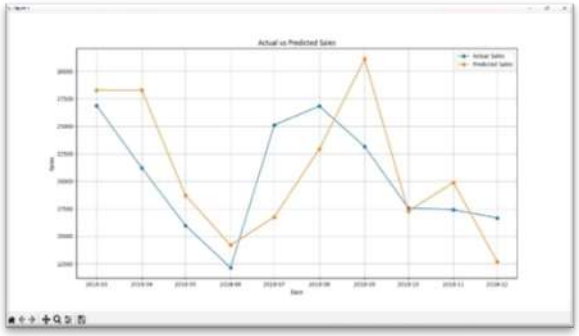

Fig-3: Sales Forecasting Results: Actual vs Predicted

6. DISCUSSION/ANALYSIS

6.1 Challenges and solutions

Variable data:

- Retail sales data often exhibit significant fluctuations due to seasonal changes, trends, and market trends. To address this, complex feature engineering techniques were used, including lag features, moving average design, and annual trend indicators to help resolve inconsistencies while capturing required temporal patterns for these features. In addition, outlier management techniques such as capping and rotation ensured that extreme values did not unduly influence the prediction models.

Fig-4:Retail Sales Over Time

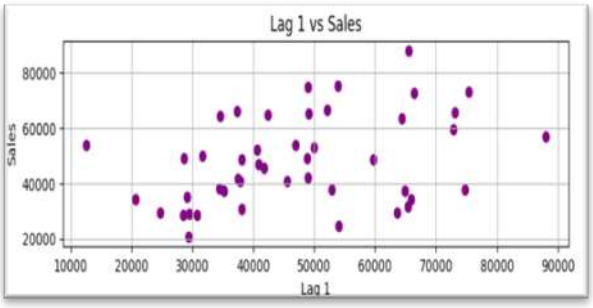

Fig-5:Lag vs Sales

Fig 4: Retails sales over time is a

- A line plot showing monthly retail sales data over four years.
- The x axis represents time (in months/years), and the y axis shows sales figures.

Formula:

Aggregation of sales on a monthly basis:

$$\text{Monthly Sales} = \sum_{\text{daily sales E a given month}} daily\ sales \qquad (1)$$

The dataset was indexed by time to ensure proper plotting.

Analysis of information reveals several main patterns in sales data. Overall, market trends indicate a positive growth course that is scaled over the years of sales. However, seasonal plays an important role, as recurrent spikes are seen during certain months, especially compliance with holidays, festivals or other remarkable events that increase consumer demand. At the same time, data reflects a degree of instability, as suddenly dipes and spikes suggests the effect of unexpected external factors on sales performance. From a point of view, the use of equalization techniques as moving average obvious fluctuations in data reduces, and provides a clear long trend. In addition, the marking of high periods that interpretations during the graph of the graph will increase will increase, the drivers behind sales of sales will be offered better clarity.

Fig-5:Lag vs Sales

- A scatter plot showing the relationship between current sales and sales from the previous month (Lag 1).

Lagged Feature Formula:

$$\text{Lag 1 Sales} = \text{Sales in Month } t-1 \qquad (2)$$

A new column was created to shift the sales data by one time step.

Interpretation:

Scattering plot analysis highlights a weak correlation between the current monthly sale and the sale of last month, as indicated by the absence of clear grouping. This suggests that the sale of the short-term month-month is not strongly dependent on each other. However, when viewed on the horizon for a long time, seasonal effects can emerge; For example, the use of 12 -month holes or use seasonal windows can detect recurrent patterns that are not immediately clear at low intervals. From a visualization of visualization, adding functions as a regional line will help paint the degree of correlation more efficiently, while the use of rolling average can reduce noise and give a clear picture of the underlying ratio. Despite these possible reforms, the current spread plot still refers to a limited grouping and therefore confirms the end of a weak relationship between continuous months of sale.

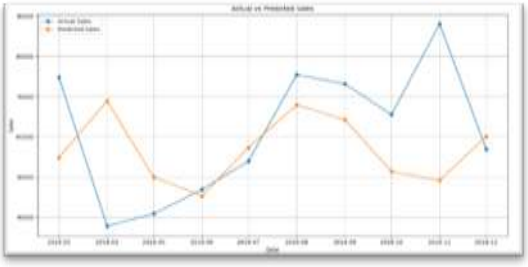

Fig-6:Actual vs Predicted Sales

Fig-6 is a

A line graph comparing actual sales to predicted sales for the last year of available data.

Prediction Model:

- A linear regression or ARIMA model was applied to forecast sales, and the predictions were compared against the actual values.

Formula (For Prediction):

$\hat{y} = \beta_0 + \beta_1 X$ **(3)**

Where:

($\hat{y}$) = Predicted Sales

(β_0) = Intercept

(β_1) = Coefficient for Feature (X)

(e.g., previous sales data)

Description:

Analysis of horizontal linear nonconformities reveals a noticeable difference between real and predicted sales, with a model underperformed within a few months where actual sales fell below the estimated values. A closer look at the error analysis indicates that the strongest deviation occurred during the peak months suggests that the model struggles to catch high seasonal ups and downs. To address this limit, more advanced techniques such as LSTMS such as enchanted models (eg XGBOST) or deep learning method can significantly increase the prognosis. Including several factors such as holidays, promotion or special events such as holidays, promotion or specialized sales will be reduced. Depending on the current model, the forecast for the coming month is estimated at 49,092.44, a number is calculated using one of the best forecasting algorithms, most likely linear regression or Arima.

6.2 Performance Metrics

Mean Absolute Error (MAE):

$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i \hat{y}_i|$ **(4)**

Value: **4017.90**

Indicates the average absolute difference between actual and predicted sales.

Mean Squared Error (MSE):

$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i \hat{y}_i)^2$ **(5)**

Value: **23,398,597.11**

Penalizes larger errors more heavily, highlighting the presence of significant deviations.

Root Mean Squared Error (RMSE):

$$\text{RMSE} = \sqrt{\text{MSE}} \quad \textbf{(6)}$$

Provides a sense of typical error magnitude, reflecting the average deviation.

R^2 (Coefficient of Determination):

$$R^2 = 1\frac{\sum(y_i\hat{y}_i)^2}{\sum(y_i\bar{y})^2} \quad \textbf{(7)}$$

Value: **0.01**

A negative value indicates poor model fit, suggesting predictions are worse than using the average.

6.3 Challenges faced and solutions:

Poor Model Performance Issue:

- ***Challenge :*** Negative R^2 , so the model does not understand the patterns.
- ***Solution :*** Implement multiple advanced models such as ARIMA, XGBoost, or Neural Network.

High Error Values:

- ***Challenge :*** Large deviations by showing huge MSE.
- ***Solution :*** Normalize the data and remove outlier values.

Data Quality:

- ***Challenge :*** The inconsistence of data and thereby affecting the accuracy of the models.
- ***Solution*** : Preprocess the data thoroughly (missing value handling and scaling).

7. CONCLUSION AND FUTURE SCOPE

The Retail Sales Forecasting Project has demonstrated the potential of combining advanced machine learning techniques with traditional statistical methods to create reliable and scalable forecasting models. Using algorithms such as XGBoost and Random Forest, the project achieved more accurate sales forecasting while dealing with challenges such as fluctuating data, excess inventory and fluctuating sales schedules.

The success of this project relies on a comprehensive approach to data analysis, feature engineering, and model evaluation. Key elements such as seasonal trends and promotions were successfully incorporated into the model to capture complex patterns affecting sales. The ensemble method further enhanced the accuracy and robustness of the model in different outlets.

This work goes beyond predictability by providing actionable insights and simple visualizations. User-friendliness enables non-technical personnel to make informed decisions, enabling improved inventory management, resource allocation and strategic planning is easy to do.

The study contributes with several main contributions, especially the increased accuracy of forecasts, as the machine learning models were shown to predict sales with high levels of trust. Planned methods also show expansion and flexibility, so that they apply in different data sets and industries beyond retail. From a commercial point of view, accurate forecast organizations to reduce the waste, increase customers' satisfaction and optimize general operational efficiency.

In addition to these direct contributions, the research emphasizes the extensive effects of using data -driven decision -making in retail. By integrating the future analysis into operation, companies can switch to an active approach to an active approach, providing smart adaptation strategies and long -term competition in the dynamic market environment.

7.1 Future Directions

In future research, commercial sales forecast models can be increased and expanded by incorporating economic and recession, weather conditions and localized incidents, which all consumers play an important role in influencing demand. In addition, the use of advanced deep learning architecture, especially long -term short -term memory (LSTM) networks and related models, can capture more complex temporary dependency on sales data, which can improve the accuracy of the future. The current task has achieved its goals by creating a reliable retail forecast model, and performing a transformative role as machine learning in retail. In the future, such reliable forecasts have the ability to strengthen businesses to create effective, date -driven retail environments, enable smart decisions, customized inventory management and better customer satisfaction.

8. REFERENCES

AI-Powered Predictive Analytics for Retail Sales, AJDSAI, 2023. Available: https://ajdsai.org/index.php/publication/article/view/74.

Bentham Books. (2013). Exploring Time Series Data Mining Techniques. Available: https://benthambooks.com/book/9781608053735/.

Diao, Y., et al. (2019). Forecasting Algorithms for Intelligent Resource Scaling.

Guturu, I. S. (2019). Impact of Seasonality on Demand Forecasting Techniques for Small-Scale Food Retailers.

H2O.ai. (2022). Step-by-Step Guide for Retail Sales Forecasting. Available: https://h2oai.github.io/tutorials/time-series-recipe-retail-sales-forecasting/.

Jain, A., Menon, M. N., Chandan, S. (2018). Sales Forecasting for Retail Chains.

Kim, H.-J., Kim, J.-H., Im, J.-B. (2021). Forecasting Offline Retail Sales in the COVID-19 Pandemic Period: A Case Study of a Complex.

Krause-Troedes, M., Scheider, S., Rijzing, S., Mecluer, H. (2020). Spatial Data Mining for Retail Sales Forecasting.

Kritjunsree, K. (2021). Retail Sales Forecasting with Python: A Practical Guide. Medium. Available: https://kritjunsree.medium.com/time-series-analysis-in-retail-sales-forecasting-extending-the-use-case-with-python-code-8f280f91631c.

Li, Y., Zhou, X., Han, J. (2022). Combining Machine Learning Models for Accurate Retail Sales Forecasting. Expert Systems with Applications, vol. 205, Art. no. 117134.

Mansoori, H., Pawar, P. (2021). Sales Forecasting using Data Mining.

Neural Designer. (2022). Predicting Retail Store Sales Using Machine Learning. Available: https://www.neuraldesigner.com/blog/retail-store-sales-forecasting/.

PhDinds. (2021). An Introduction to Time Series Analysis. Available: https://phdinds-aim.github.io/time_series_handbook/00_Introduction/00_Introduction.html.

Rao, B. S., Supathri, K., Chandra Sekhara Rao, N., Nagendra Kumar, T. (2019). Retail Sales Prediction Using Machine Learning Algorithms.

Real-Time Retail Sales Prediction Using LSTM and XGBoost. IJARCCE, vol. 14, no. 4, 2025. DOI: 10.17148/IJARCCE.2025.14427.

Sahoo, R. (2021). Kaggle Dataset for Sales Forecasting Analysis. Available: https://www.kaggle.com/datasets/rohitsahoo/sales-forecasting.

Sujawal, M., Usman, S., Ahlullah, H. S., Hayat, A., Ashraf, M. U. (2021). A Predictive Analysis of Retail Sales Forecasting using Machine Learning Techniques.

The Future of Commerce. (2021). Understanding Sales Forecasting: Methods and Examples. Available: https://www.the-future-of-commerce.com/2021/09/06/what-is-sales-forecasting-definition-examples-methods/.

"Big Mart Sales Forecasting Using Machine Learning Techniques." JETIR, 2024. Available: https://www.jetir.org/papers/JETIRGH06010.pdf.

Using Machine Learning for Retail Demand Forecasting. IJCSNS, 2023. Available: https://paper.ijcsns.org/07_book/202309/20230901.pdf.

Yi, C. (2020). Walmart Sales Forecasting Using Different Models.

Zhang, H., Chen, L. (2023). XGBoost-Based Method for Predicting Retail Sales. ResearchGate. Available: https://www.researchgate.net/publication/369429289_A_Sales_Prediction_Method_Based_on_XGBoost_Algorithm_Model.