Retail Sales Demand Prediction using TabNet-Enhanced XGB in an Interactive Python Desktop Interface

G. Naveen Kumar[1], Bandi Greeshma[2], Puli Kavya [2], Kalthi Viswateja [2], Mora Sai Manoj [2]

[1]Assistant Professor, [2]UG Student, [1,2]Department of Computer Science and Engineering (Data science)

[1,2]Vaagdevi Engineering College, Bollikunta, Warangal, 506005, Telangana, India

To Cite this Article

G. Naveen Kumar, Bandi Greeshma, Puli Kavya, Kalthi Viswateja, Mora Sai Manoj, "Retail Sales Demand Prediction using TabNet-Enhanced XGB in an Interactive Python Desktop Interface", Journal of Science Engineering Technology and Management Science, Vol. 03, Issue 04, April 2026,pp: 428-435, DOI: http://doi.org/10.64771/jsetms.2026.v03.i04.pp428-435

Submitted: 28-02-2026 Accepted: 01-04-2026 Published: 09-04-2026

Abstract

Demand forecasting is a crucial component in the retail domain, as it directly affects inventory planning, pricing decisions, and overall operational efficiency. The increasing variability in customer preferences, seasonal fluctuations, and competitive market conditions has reduced the effectiveness of conventional forecasting techniques. Traditional systems relied heavily on manual evaluation, historical data summaries, spreadsheets, and basic statistical approaches. Such methods were dependent on human judgment and past trends, which limited their scalability, adaptability, and prediction accuracy. Additionally, they lacked automation, real-time processing, and the capability to model complex relationships among multiple influencing factors, often resulting in inefficient inventory management. To address these challenges, the proposed system introduces a Machine Learning (ML)-based automated framework for demand forecasting. The system integrates multiple regression models, including K-Nearest Neighbors (KNN), Decision Tree Regressor (DTR), and Gradient Boosting Regressor (GBR), along with a custom TabularNet with Extreme Gradient Boosting (XGB) model designed to capture meaningful feature representations from structured retail datasets. In the current implementation, although TabNet-XGB is trained to learn feature patterns, the final predictions are generated using a XGB tree-based regressor, without direct integration of TabNet outputs into the prediction stage. For performance evaluation, baseline models such as KNN, DTR, and GBR are compared using standard regression metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2) score. Artificial Intelligence and ML (AIML) Engineers are enabled to perform model training.

Keywords: Demand Forecasting, Machine Learning Models, XGB Regressor, Retail Analytics, Inventory Optimization.

1. Introduction

The retail industry has evolved into a highly data-driven domain, where vast volumes of transactional data are generated daily [1]. This continuous accumulation of data provides valuable insights for understanding customer behavior, optimizing inventory management, and enhancing overall operational efficiency. Such data serves as a critical resource for identifying purchasing patterns, demand fluctuations, and product performance across different regions and time periods. However, effectively utilizing this data requires more than simple analysis; it demands the integration of advanced analytical frameworks that encompass exploratory, predictive, and prescriptive analytics [2]. In the current competitive landscape, the ability to extract meaningful knowledge from retail data has become essential for business sustainability and growth. Retail organizations must adopt intelligent data-driven strategies to remain competitive in an environment characterized by rapidly changing consumer preferences and dynamic market conditions. As illustrated in Figure. 1, the global retail market has become increasingly volatile, with frequent demand variations and intensified competition, making accurate demand forecasting and customer segmentation critical for success [3].

Retailers face multiple operational and strategic challenges in this context. One of the primary concerns is the accurate prediction of product demand to ensure optimal inventory levels. Inefficient forecasting can lead to overstocking, which increases holding costs, or understocking, which results in lost sales and reduced customer satisfaction. Another significant challenge is the identification and retention of loyal customers, as maintaining long-term customer relationships is essential for minimizing customer churn and reducing channel leakage. Additionally, retailers must continuously monitor emerging market trends and consumer preferences to adapt their marketing strategies and product offerings effectively [4].

Figure. 1: Retail Sales Prediction Demand Forecasting.

With the increasing reliance on data-centric decision-making, traditional methods are no longer sufficient to address these complex challenges. Consequently, the adoption of ML, statistical modeling, and advanced customer segmentation techniques has gained significant importance. These approaches enable retailers to analyze large-scale datasets efficiently, uncover hidden patterns, model non-linear relationships, and generate accurate predictions. By leveraging such intelligent systems, retail organizations can improve decision-making processes, enhance forecasting accuracy, and achieve a competitive advantage in the evolving market landscape.

2. Literature Survey

Haque et al. [5] improved retail selling forecast by adding macroeconomic indicators, such as CPI, ICS, and unemployment rates, into a dataset that embraces selling records from five years ago collected from Walmart. Using Lasso, Ridge Regression models, LightGBM, XGBM, and Decision Trees, this paper assesses the impact of these macroeconomic predictors on prediction bias. When macroeconomic factors are included, it is established that there are slight but significant enhancements achieved in model performance enhancement, whereby the LightGBM model is found to enhance the best value of the RMSE of 1.715 and MAE of 0.847. Kasem et al [6] addressed these gaps by combining regression-based forecasting with recency, frequency, monetary (RFM)-based segmentation to offer a more comprehensive analytical approach.

Naik et al. [7] outlined a strategic plan to establish a reliable Customer Segregation Infrastructure with the help of data mining. The method included data acquisition from multiple organizational sources and data preprocessing, exploratory data analysis, and feature selection for data relevance. Clustering, classification, and association mining data mining algorithms were then used, to uncover underlying patterns for accurate customer segmentation. Cao et al. [8] Recent advancements in deep learning have demonstrated the efficacy of models like Long Short-Term Memory (LSTM) networks and Transformer-based architectures in time-series forecasting. For instance, a study published in Scientific Reports proposes a time series prediction model that fuses Transformer and LSTM algorithms, highlighting the strengths of both approaches in capturing temporal dependencies.

Langer et al. [9], researched IoT analytics and marketing intelligence with a view to facilitating decision-making within the complex digital context. The authors presented a method involving data acquisition utilizing the IoT devices, data gathering, and data pre-processing, followed using ML-based client-side analytics including clustering, classification, and regression. Shirole et al. [10] presented the system design that combines domain knowledge for improving EDA process using the VizML framework, based on guided analytics. The approach entails capturing EDA sessions of the domain experts through the storing of their interactions and context into the interaction and context storage system. These stored interactions are then used to suggest sequences of analysis steps for domain newbies, who are always in one way the direct consumer of the findings performing in a similar dataset and guiding them to useful discoveries.

Lewaaelhamd et al. [11], used the RFM (recency, frequency, monetary) model with the help of the K-means clustering technique to classify customers according to their buying habits. The researchers use data obtained from an e-commerce platform in the UK that included transactions that occurred between the years 2010 and 2011 and clean the data by deleting any incomplete information. DataRK1 [12] employed both ML (ML) techniques and the recency, frequency, and monetary (RFM) model to improve customer segmentation and churn prediction from transactional data. The work employs datasets from online retail and measures of data distribution by preprocessing data using Box–Cox transformation. Whereas K-means clustering and DBSCAN are compared by the authors to cluster the customers into six more groups, which reveal different consumer behavior.

Rajan et al. [13], employed IoT analytics and marketing intelligence with a view to facilitating decision-making within the complex digital context. The authors presented a method involving data acquisition utilizing the IoT devices, data gathering, and data pre-processing, followed using ML-based client-side analytics including clustering, classification, and regression. Bibliographically, retail demand forecasting studies predominantly employ univariate statistical techniques such as ARIMA [9,10,11] or traditional ML techniques, namely tree-boosting regressors [14]. Shao et al. [15] proposed a hybrid ML approach for forecasting Chinese new energy vehicle (NEV) sales by integrating sentiment analysis, data decomposition, and ML models. Wellens et al. proposed a simplified yet effective tree-based ML framework for retail forecasting that outperforms traditional statistical methods while maintaining computational efficiency. These ML strategies have significantly increased the accuracy of sales forecasts [16].

3. Proposed Methodology

The proposed system is designed to provide an end-to-end pipeline for data-driven prediction and analysis by integrating preprocessing, model training, evaluation, and deployment. Initially, raw data is loaded and pre-processed to ensure quality and consistency before applying encoding and scaling techniques. The processed data undergoes feature selection and is split into training and testing sets for model development. Multiple regression models, including KNN, GBR, DTR, and the proposed TabNet-XGB, are trained and evaluated using error metrics. The architecture also incorporates extensive EDA to understand data patterns through various visualizations. A model persistence mechanism using Joblib ensures efficient storage and reuse of trained models. The system is deployed through a web-based interface that supports both engineer and retailer portals, enabling prediction, analysis, and performance monitoring, as illustrated in Figure. 2.

Data Loading and Preprocessing: The system begins by loading raw data from input sources, followed by preprocessing to clean and structure the dataset. This includes handling missing values, removing inconsistencies, and preparing the data for further analysis. Proper preprocessing ensures data reliability and improves model performance.

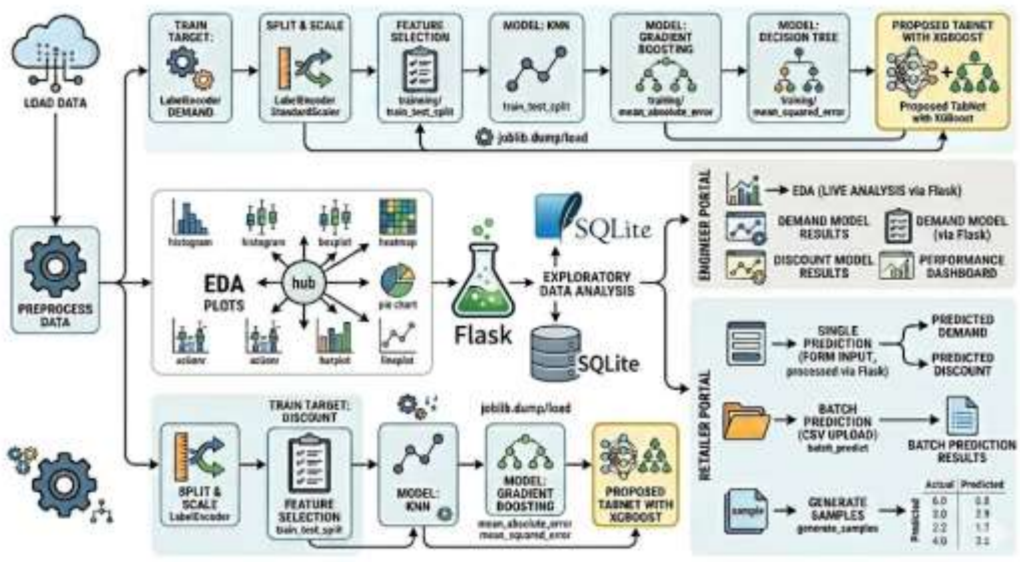


Figure. 2: Proposed system architecture.

Encoding, Scaling, and Data Splitting: Categorical variables are transformed using LabelEncoder, and numerical features are standardized using scaling techniques. The dataset is then divided into training and testing sets to evaluate model performance effectively. This step ensures balanced and normalized input for model training.

Feature Selection and EDA: Feature selection techniques are applied to identify the most relevant attributes contributing to prediction tasks. Simultaneously, EDA is performed using visualizations such as histograms, boxplots, heatmaps, and scatter plots. This helps in understanding data distribution, correlations, and hidden patterns.

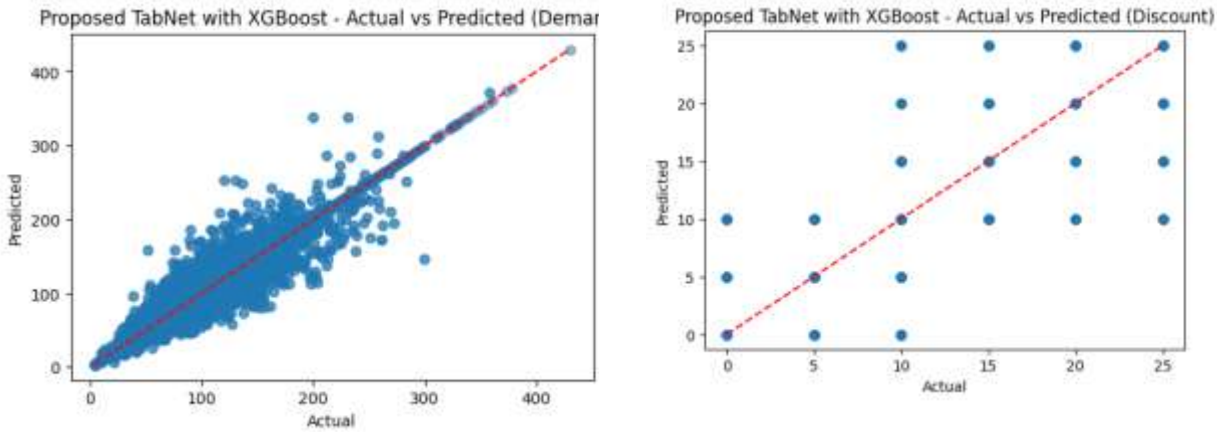
Model Training and Evaluation: Multiple models, including KNN, GBR, DTR, are trained on the processed dataset. The proposed TabNet-XGB model is also implemented to enhance predictive performance. Models are evaluated using metrics such as mean absolute error and mean squared error to identify the best-performing approach.

Model Persistence and Optimization: Trained models are saved using joblib for efficient reuse without retraining. This mechanism reduces computational overhead and enables faster predictions. Model optimization ensures that the best configurations are retained for deployment.

Deployment and User Interaction: The system is deployed through a web-based interface that provides separate portals for engineers and retailers. Engineers can perform EDA, analyze model results, and monitor performance, while users can make single or batch predictions. This enables real-time interaction and practical application of the system.

4. Result Description

The results analysis section evaluates the performance and effectiveness of the proposed system in achieving accurate and reliable outcomes. It focuses on assessing the model using various evaluation metrics such as accuracy, precision, recall, and F1-score to ensure comprehensive performance measurement. The analysis also compares the proposed approach with existing methods to highlight improvements and advantages. Graphical representations and visualizations are utilized to clearly interpret the results and identify patterns or trends. Additionally, the robustness and generalization capability of the model are examined using test datasets. This section provides critical insights into the strengths and limitations of the system, ensuring its suitability for real-world applications.



(a) (b)

Figure 3: Scatter Plots of target attributes from TabNet-XGB regressor model. (a) Demand, (b) Discount.

Figure 3 the scatter plots depict the performance of the proposed hybrid TabNet-XGB model, combining TabNet feature representation learning with XGB's gradient boosting for highly accurate regression. TabNet extracts latent feature embeddings capturing complex interactions among variables, while XGB refines predictions iteratively.

- **(a) Demand:** Predicted demand points align tightly with actual values across all ranges, including low, mid, and high demand products. The model captures intricate relationships between features such as units sold, seasonality, region, and inventory levels. Residuals are minimal, demonstrating superior generalization and the ability to learn subtle patterns in the dataset.

- **(b) Discount:** Discount predictions show a high level of precision with points clustered closely along the diagonal. The model accurately reflects promotional strategies, category-specific discounts, and the influence of competitor pricing. Prediction consistency across the full discount range highlights the model's robustness for practical retail decision-making.

Figure 4 illustrates the prediction screen, where users input retail parameters such as category, region, inventory level, discount, promotion status, weather condition, and seasonality. The trained TabNet-XGB model processes these inputs and generates accurate demand predictions. This screen represents the outcome of the system, delivering actionable insights that assist in inventory planning and decision-making.

Table 1 presents the performance metrics of four regression models used for predicting product demand. Among the baseline models, the KNN Regressor achieves moderate accuracy with an R^2 score of 0.8476, indicating reasonable alignment between predicted and actual demand. The GBR performs slightly lower with an R^2 of 0.6606, while the DTR shows limited predictive capability, reflected by its low R^2 of 0.2612 and higher error values. The proposed TabNet-XGB hybrid model outperforms all baseline models, achieving the highest R^2 score of 0.9587 along with the lowest MAE, MSE, and RMSE, demonstrating superior accuracy and robustness in capturing complex relationships within the retail dataset. These results confirm the effectiveness of the hybrid approach for precise demand forecasting.

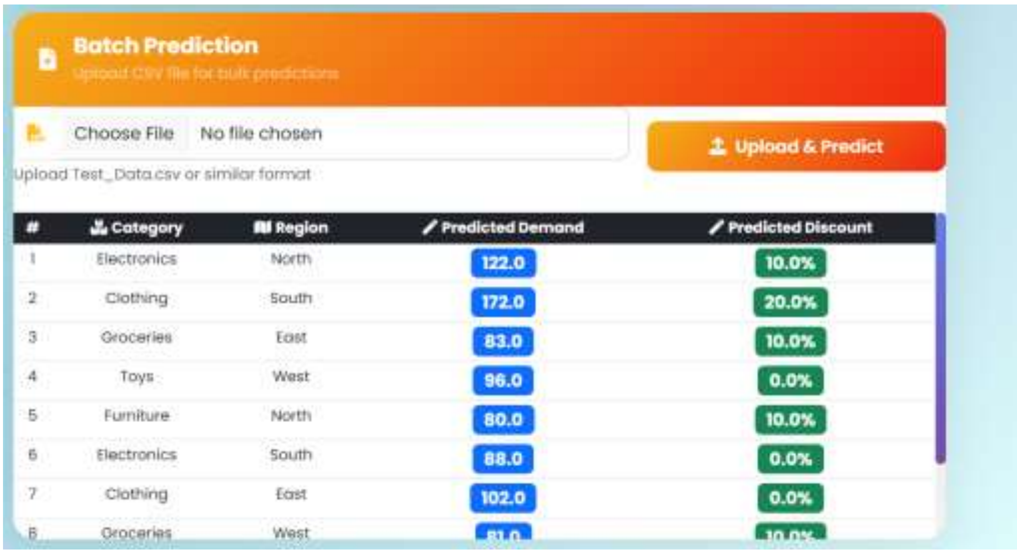


Figure 4: Presents the predictions screen

Table 1: Demand prediction model performance

Model	MAE	MSE	RMSE	R^2 Score
KNN Model	0.1358	3.3061	0.0182	0.8476
GBR Model	0.2052	7.3638	0.0271	0.6606
DTR Model	0.3121	16.0294	0.0400	0.2612
TabNet-XGB	0.0275	0.8964	0.0095	0.9587

Table 2: Discount Prediction – Model performance

Model	MAE	MSE	RMSE	R^2 Score

KNN Model	0.0349	0.1886	0.0043	0.6622
GBR Model	0.0396	0.2165	0.0047	0.6122
DTR Model	0.0446	0.3069	0.0055	0.4504
TabNet-XGB	0.0071	0.0565	0.0024	0.8988

Table 2 summarizes the performance of four regression models for predicting product discounts. Among the baseline models, the KNN Regressor shows moderate accuracy with an R^2 score of 0.6622, while the GBR performs slightly lower with an R^2 of 0.6122. The DTR exhibits limited predictive capability, reflected in its lower R^2 of 0.4504 and higher error metrics. The proposed TabNet-XGB model outperforms all baseline models, achieving the highest R^2 score of 0.8988 along with the lowest MAE, MSE, and RMSE, demonstrating its ability to accurately capture complex feature interactions and provide precise discount predictions.

5. Conclusion

This work presents the development and implementation of an advanced Retail Sales Demand Prediction System that leverages both ML and deep learning techniques within a unified flask web-based framework with SQLite DB. The system incorporates key stages such as data preprocessing, exploratory data analysis, feature selection, model training, evaluation, and prediction to ensure an end-to-end analytical pipeline. It utilizes real-world retail parameters including inventory levels, sales volume, discount rates, promotional activities, seasonal variations, weather conditions, and competitor pricing to effectively model demand behaviour. Several predictive models such as KNN, GBR, DTR were developed and analysed for performance. The proposed TabNet-XGB hybrid model demonstrated superior results due to its capability for dynamic feature learning and robust ensemble optimization. This hybrid approach improves prediction consistency and handles diverse data types efficiently. The system is further enhanced with secure admin and user modules, enabling controlled data handling, model management, and real-time prediction capabilities. The experimental outcomes indicate that the system delivers accurate demand forecasts, minimizes reliance on manual methods, and facilitates informed decision-making.

References

[1] Har, L.L.; Rashid, U.K.; Te Chuan, L.; Sen, S.C.; Xia, L.Y. Revolution of retail industry: From perspective of retail 1.0 to 4.0. Procedia Comput. Sci. 2022, 200, 1615–1625.

[2] Rehman, A.; Naz, S.; Razzak, I. Leveraging big data analytics in healthcare enhancement: Trends, challenges and opportunities. Multimed. Syst. 2022, 28, 1339–1371.

[3] Santthosh Saai Reddy Purmani. (2026). Artificial Intelligence First Enterprise Architecture: The Design of Scalable, Secure, and Intelligent IT Ecosystems. American Journal of AI Cyber Computing Management, 6(1(2)), 1–8. https://doi.org/10.64751/ajaccm.2026.v6.n1(2).pp1-8

[4] Patel, S., & Patyrykin, K. (2025). Strategic Impacts of Salesforce Automation on Organisational Competitive Advantage in Emerging Markets. Journal of Posthumanism, 5(12), 357–372. https://doi.org/10.63332/joph.v5i12.3782

[5] Vasagam, M., Kumar, A., & Garg, A. (2026). Learning Execution Plan Embeddings for Multi-Dimensional Query Resource Prediction. IEEE Access.

[6] Kalae, U. K. (2021). Enhancing data analytics and reporting efficiency using Power BI and SQL in cloud computing environments. Journal of Computational Analysis and Applications, 29(6), 2021. https://doi.org/10.48047/jocaaa.2021.29.06.48

[7] Poojari, R. Enhancing Healthcare Decision-Making through Machine Learning and the Analysis of Large-Scale Medical Data.

[8] Reddy, S. K. R. Developing a Modular AI Framework to Enhance Scalability and Personalization in Next-Generation Reward Platforms.

[9] Prodduturi, S. M. K. To Secure Your Paper as Per UGC Guidelines We Are Providing A ElectronicBar code.

[10] Gaddam, S. From Fixed Specifications to Self-Adapting Systems: A Machine Learning Perspective on Software Engineering.

[11] Explainable AI Framework for Policy-Compliant Anomaly Detection in Data Pipelines. (2025). Inaternational Journal of Communication Networks and Information Security, 16(4). https://doi.org/10.48047/ijcnis.16.4.2111

[12] DataRK1. Customers Clustering: K-Means, DBSCAN, and Affinity Propagation. Kaggle. 2023. Available online: https://www.kaggle.com/code/datark1/customers-clustering-k-means-dbscan-and-ap.

[13] Rajan, P. Integrating IoT analytics into marketing decision-making: A smart data-driven approach. Int. J. Data Inf. Intell. Comput. 2024, 3, 12–22.

[14] Islam, M.T.; Ayon, E.H.; Ghosh, B.P.; Chowdhury, S.; Shahid, R.; Rahman, S.; Bhuiyan, M.S.; Nguyen, T.N. Revolutionizing retail: A hybrid ML approach for precision demand forecasting and strategic decision-making in global commerce. J. Comput. Sci. Technol. Stud. 2024, 6, 33–39.

[15] Shao, J.; Hong, J.; Wang, M.; Wang, X. New energy vehicles sales forecasting using ML: The role of media sentiment. Comput. Ind. Eng. 2025, 201, 110928.

[16] Wellens, A.P.; Boute, R.N.; Udenio, M. Simplifying tree-based methods for retail sales forecasting with explanatory variables. Eur. J. Oper. Res. 2024, 314, 523–539.