Rashmi Amardeep, 2023, 11:6
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology

An Open Access Journal

Sales Prediction and Inventory Restocking Using Machine Learning

Asso. Prof. Rashmi Amardeep, Prakhar Ananth,
Priya B Gunali, Riya Raikar, S Deepak Dhore Reddy
Department of Information Science and Engineering,
Dayananda Sagar Academy of Technology and Management, Bengaluru, India

Abstract- In the dynamic landscape of retail, accurate demand forecasting stands as a cornerstone for optimizing inventory, enhancing customer satisfaction, and maximizing profitability. This research addresses the intricate challenges associated with demand forecasting in unorganized retail settings, with a specific focus on super shops. Leveraging machine learning and deep learning techniques, the study explores innovative algorithms to predict future product demand. The methodology begins with comprehensive data collection, overcoming hurdles inherent in unorganized retail data accessibility. Key attributes, including temporal factors (time, month, occasion, and location), promotional activities, and environmental influences, form the basis of the analysis. A statistical approach transforms raw data into a processed dataset, setting the stage for the implementation of three prominent machine learning algorithms—K Nearest Neighbor (KNN), Gaussian Naive Bayes, and Decision Tree Classifier. Results highlight the distinct strengths of each algorithm in capturing local patterns, probabilistic classification, and handling complex decision-making scenarios. Through a meticulous evaluation of performance metrics, K-Nearest Neighbor emerges as the most effective model for demand forecasting in unorganized retail. The research identifies critical factors influencing demand, including temporal variations, promotional impacts, weekend and holiday fluctuations, festival-induced demand shifts, weather conditions, product and market dynamics, shelf availability, brand category influence, and consumption rate dynamics. Past data proves invaluable, serving as a foundation for understanding non-linear demand patterns for individual products. In conclusion, the study contributes to the evolving discourse on demand forecasting in unorganized retail, offering insights into the multifaceted challenges faced by super shops. The identified best model, K-Nearest Neighbor, holds promise for retailers seeking to navigate the complexities of unorganized retail, enabling precise anticipation of consumer demands and informed decision-making. The implications extend to inventory optimization, resource allocation, and strategic planning in the ever-evolving retail sector.

Keywords- Unorganized Retail, Demand Forecasting, Machine Learning, Super Shops, K-Nearest Neighbor, Retail Analytics, Deep Learning, Demand Patterns, Predictive Modeling, Inventory Management

© 2023 Rashmi Amardeep. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited.

I. INTRODUCTION

Customer satisfaction stands as a critical determinant of success in the business realm, with retailers perpetually striving to meet consumer demands. In the dynamic market landscape, retailers engage in fierce competition to maximize product sales and profitability while maintaining prudent investment strategies. Central to this endeavor is the anticipation of product demand, a task that can significantly impact profit margins. This research is dedicated to identifying the most effective algorithm capable of estimating product demand with a high degree of accuracy and minimal error. What sets this study apart is its innovative consideration of temporal factors such as time, month, occasion, and the geographical location of the shop in the prediction process.

The contextual backdrop of this research is Bangladesh, a country where a substantial portion of the population is engaged in entrepreneurial pursuits. Consequently, a plethora of shops, varied retail outlets, and super shops dot different regions, each stocked with diverse models and products to meet consumer demands. The myriad attributes influencing demand prediction include data availability, timing of forecasting, historical records, past demand patterns, location, and special occasions. The interplay of these factors underscores the complexity of forecasting in a multifaceted retail landscape. Effective demand prediction methods not only contribute to minimizing inventory costs but also play a pivotal role in ensuring product availability. The primary objective is to streamline inventory costs, making optimal demand prediction indispensable. Recognizing this, the research emphasizes the critical role of accurate demand forecasting in establishing an efficient supply chain. Employing advanced machine learning techniques, the study focuses on different algorithms to predict future customer demands for specific products.

In the competitive retail landscape, the consequences of stock outs are profound, potentially resulting in lost sales and dissatisfied customers. The challenge lies in accurately predicting customer preferences and demands, especially within the middle and low-level consumer segments where confidentiality is harder to maintain. This research delves into the impact of accurate prediction on retail sales, acknowledging the increasing competition among retailers and the importance of understanding customer behavior across various market segments. Considering market dynamics and financial constraints within the retail sector, accurate predictions become pivotal. The adoption of technology, modern tools, and innovative predictive methods is becoming imperative for retailers aiming to forecast customer demands effectively. This paper addresses the intricate relationship between customers and retailers, shedding light on the nature of the retail shop section and the dynamics at play in this sector. The forecasting algorithms and methodologies discussed herein leverage data from the previous month, highlighting the importance of historical trends in predicting future product demands. The research employs data from three distinct shops, utilizing machine learning and data mining to introduce novel models applicable to time series forecasting. In an era where autonomous decision-making systems are gaining prominence, the study employs K-Nearest Neighbor (KNN), Gaussian Naive Bayes, and Decision Tree Classifier for effective data classification and decision-making. This study endeavors to determine the optimal technique for demand forecasting by examining various features.
The literature review section offers insights into previous forecasting methodologies, setting the stage for the methodology section, where the research approach is delineated. The ensuing performance measurement section elucidates the calculation process, and the performance analysis section scrutinizes the results, ultimately identifying the most effective forecasting method.

II. LITERATURE SURVEY

1. Comparison Study Product Demand Forecasting with Machine Learning for Shop

This study compares existing demand prediction models, including KNN, Random Forest, FNN, ANN, and Holt-Winters, based on Mean Absolute Percentage Error (MAPE). We propose a method employing K Nearest Neighbor, Gaussian Naive

Bayes, and Decision Tree Classifier, considering factors like customer behavior, seasonal weather, time, occasion, month, and product category. In our local market, the Gaussian Naive Bayes algorithm demonstrates the best accuracy at 58.92%. Future implementation in the market will further refine the model, exploring additional decision-making attributes and incorporating diverse datasets from different city segments.

2. A Hybrid Machine Learning Model for Sales Prediction

This paper introduces a novel approach for predicting Wal Mart's sales by combining XG Boost and Light GBM frameworks, emphasizing feature engineering for data processing. The models built on these frameworks outperform traditional methods like Logistic Regression and SVM in experiments. The integrated model, merging XG Boost and Light GBM, demonstrates superior predictive capabilities, showcasing its potential as a robust solution for sales prediction in retail contexts.

3. Sales Prediction Model for Big Mart

This paper explores the application of machine learning algorithms to predict sales at Big Mart. Emphasizing the evolving role of machine learning in handling vast amounts of data, the authors introduce various algorithms, including Linear Regression, K-Nearest Neighbors, Decision Tree, Naïve Bayes Classifiers, Random Tree, and K-means Clustering. The study aims to understand the impact of item properties on sales at Big Mart, formulating a predictive model for each store. The methodology involves preprocessing a dataset from 10 stores and 1559 products, utilizing algorithms like Random Forest and linear regression for accurate sales predictions. Overall, the paper provides a concise overview of using machine learning for sales prediction, specifically in the context of a comprehensive dataset from a large shopping center like Big Mart.

4. Sales Prediction using Machine Learning Algorithms

This paper explores machine learning algorithms for sales prediction, utilizing four algorithms, including an ensemble technique. It aims to predict future sales using clustering models and measures. The research involves decisions from experimental data and insights from data visualization, employing data mining techniques. It outlines a sales prediction system and product recommendation system for retail, leveraging consumer demographics. Deep neural networks are applied for understanding electronic component sales strategy. Bayesian learning, coupled with neural networks, predicts individual retailer sales rates, benefiting a large number of outputs. The study detects suspicious behavior using an automatic prototype, merging various machine learning methodologies for effective user profiling.

5. A Predictive Analysis of Retail Sales Forecasting using Machine Learning Techniques

The paper emphasizes sales forecasting's critical role in information technology chain stores, impacting inventory management, marketing, customer service, and financial planning. It highlights challenges like over and under-forecasting and suggests a blend of human planning and statistical models for accuracy. Utilizing LSTM, ARIMA, Linear Regression, Random Forest, and XG Boost models, it concludes XG Boost as most suitable for the Citadel POS dataset. Future prospects involve leveraging deep learning with larger datasets to enhance accuracy in retail sales forecasting.

6. BMSP-ML: Big Mart Sales Prediction Using Different Machine Learning Techniques

This paper underscores the critical role of accurate sales predictions for businesses in maintaining standards and enhancing profitability. The study formulates hypotheses pertaining to store and product attributes influencing sales, conducts comprehensive data exploration, and employs machine learning models, including linear regression, ridge regression, decision tree, and random forest, to predict store sales. The results indicate improvements over a baseline model, with ridge regression demonstrating superior performance on a public leaderboard. The paper identifies limitations, such as the exclusion of disaster-related factors, and suggests future work, including incorporating these elements and developing an online app for customer reviews and

rankings. Overall, the study contributes to the field by proposing a robust framework for mart sales prediction and evaluating various machine learning models' performance.

III. PROBLEMS IDENTIFIED

Super shops encounter a spectrum of challenges, each contributing to the intricacies of demand forecasting. The demand patterns for fixed items, in particular, reveal constraints associated with store-level quantities, warehouse capabilities, promotional activities, and pricing fluctuations. Traditional demand prediction approaches are evolving, with machine learning and deep learning emerging as transformative tools capable of harnessing extensive datasets to forecast customer demand and behavior.

1. Identified Problems

Temporal Variability in Demand
The study highlights a distinctive pattern in demand fluctuations throughout the month, emphasizing higher demand in the initial portion contrasted with a decline in the latter part. Understanding and predicting these temporal shifts pose a unique challenge.

Impact of Promotional Offers
The influence of promotional offers on customer demand is a crucial factor. These incentives, designed to stimulate purchases, significantly contribute to the volatility of demand patterns.

Weekend and Holiday Demand
Consumer behavior exhibits notable changes during weekends and holidays, with demand witnessing an upswing. Effectively modeling and predicting these variations are vital for accurate demand forecasts.

Festival-Induced Demand Fluctuations
Festivals emerge as pivotal factors affecting customer demand. The study recognizes the substantial impact of cultural and seasonal festivities on consumer behavior, necessitating nuanced forecasting strategies.

Weather Conditions as Demand Determinants
Weather conditions wield influence over customer preferences and purchasing behavior. Incorporating weather-related variables into forecasting models becomes essential to capture demand nuances associated with climatic variations.

Product and Market Dynamics
The intricate interplay of product characteristics and market dynamics significantly contributes to demand variability. Capturing and integrating these intricate relationships into forecasting models pose inherent challenges.

Shelf Availability
The availability of items on the shelf directly impacts customer choices and, consequently, demand. Managing and predicting shelf availability becomes critical aspect of demand forecasting.

Brand Category Influence
Understanding the impact of brand categories on customer preferences and demand patterns adds complexity to forecasting. Distinguishing between brands and their individual influence on demand is a critical consideration.

Consumption Rate Dynamics
The rate at which products are consumed directly influences demand patterns. Predicting consumption rates accurately requires a nuanced understanding of product life cycles and consumer behavior.

2. Importance of Past Data

Past data emerges as a cornerstone for predicting customer demand. The non-linear nature of demand for individual products necessitates a comprehensive understanding of historical patterns. Leveraging past data enables the identification of demand trends and patterns, forming the foundation for more accurate forecasting.

Addressing these challenges requires sophisticated models that not only leverage historical data but also account for the dynamic and multifaceted nature of retail demand.

IV. PROPOSED SYSTEM

The proposed system aims to revolutionize traditional sales forecasting and inventory restocking processes by harnessing the power of machine learning algorithms. By amalgamating historical sales data, market trends, seasonality factors, and other relevant variables, the system will predict future sales trajectories with a higher degree of accuracy. This predictive capability provides businesses with valuable insights into consumer behavior, enabling proactive decision-making and strategic planning.

1. Data Collection

In the initial phase of our research, we focused on gathering relevant data to fuel our machine learning and deep learning models for sales prediction. Recognizing the impact of specific attributes on predictions, we identified key features such as expiration period, seasonal availability, festival season, holidays, temperature, storage management, advertisement/branding/popularity, natural calamity, date of restock, and date of sale. Despite the challenges associated with collecting data from unorganized retail shops, we successfully obtained information from both super shops and various retailers across ten different products. Given the retailers' profit oriented focus and lack of maintained databases, we devised a method involving the analysis of previous sales records, calculating numerical values, and determining the sale and demand data for the selected items.

2. Statistical Analysis

The next step involved the transformation of raw data into processed data to enhance the accuracy of our predictive models. A well-prepared dataset is crucial for achieving reliable predictions. We conducted a comprehensive statistical analysis, establishing relationships between different variables, and addressing missing data by subtracting sales from demand to identify unsold product quantities. The dataset was then partitioned into an 80:20 ratio for training and testing purposes, with shuffling to ensure robust model evaluation. Subsequently, three algorithms—K-Nearest Neighbor (KNN), Gaussian Naive Bayes, and Decision Tree Classifier—were implemented to classify the dataset. The workflow encompassed data collection, evaluation, feature selection, training, testing, and model performance evaluation.

3. Data Preprocessing

Recognizing that machine learning algorithms require formatted and cleaned data, we delved into data preprocessing. This crucial step involved transforming raw data into a suitable format for the algorithms. To ensure the models could effectively address the research problem, we carried out the following activities in data preprocessing:

- Handling missing data: Filling or imputing missing values to maintain dataset completeness.
- Feature selection: Identifying relevant features for model training and discarding unnecessary variables.
- Data formatting: Converting data into a structured format compatible with machine learning algorithms.
- Normalization: Scaling numerical features to a standard range to prevent biases in model training.
- Data cleaning: Addressing outliers, inconsistencies, and errors in the dataset for improved model performance.

In this research, the chosen attributes play a pivotal role in shaping the predictive models. The expiration period and seasonal availability impact product demand, while considerations such as festival seasons, holidays, and temperature influence consumer purchasing behavior. Storage management and natural calamities contribute to stock variability and sales disruptions. The timing of restocking and sales, along with advertisement, branding, and popularity, are essential factors that influence sales patterns. The methodology outlined above underscores the significance of a thorough data collection strategy and subsequent statistical analysis to derive meaningful insights. Our focus on both super shops and smaller retailers ensures a diverse dataset that is reflective of the nuances of unorganized retail. The application of machine learning algorithms, specifically KNN, Gaussian Naive Bayes, and Decision Tree Classifier, further

enhances our ability to make accurate sales predictions. This comprehensive approach, combining data collection, statistical analysis, and machine learning, positions our research at the forefront of addressing challenges in unorganized retail sales prediction. The methodology ensures that the models are trained on quality data, paving the way for reliable predictions and actionable insights for inventory management, resource allocation, and decision-making in the retail sector.

V. CONCLUSION

In the pursuit of optimizing retail operations and maximizing profitability, this research embarked on a comprehensive exploration of demand forecasting methodologies, leveraging machine learning techniques and innovative algorithms. The primary focus was to address the multifaceted challenges inherent in unorganized retail settings, where data availability, variability in product attributes, and the need for accurate demand predictions pose significant hurdles. The research commenced with meticulous data collection, overcoming challenges associated with unorganized retail shops. Key attributes such as expiration period, seasonal availability, festival seasons, holidays, temperature, storage management, advertisement/ branding/ popularity, natural calamity, date of restock, and date of sale were considered. This rich dataset, encompassing diverse features, formed the foundation for subsequent analysis.

A statistical analysis was undertaken, transforming raw data into a processed form conducive to accurate model training.

Feature engineering and dataset partitioning were essential steps to ensure robust model performance. Three prominent machine learning algorithms—K-Nearest Neighbor (KNN), Gaussian Naive Bayes, and Decision Tree Classifier—were implemented to predict future product demand. A. Key Findings:

The study revealed that accurate demand prediction plays a pivotal role in minimizing inventory costs, optimizing supply chain operations, and enhancing customer satisfaction. The exploration of three distinct algorithms highlighted their varied performance across different attributes, emphasizing the need for a nuanced approach to model selection. B. Results and Best Model:

In the analysis of the provided test data, each algorithm demonstrated unique strengths and limitations. K-Nearest Neighbor exhibited commendable accuracy in capturing local patterns, while Gaussian Naive Bayes excelled in probabilistic classification. The Decision Tree Classifier showcased its effectiveness in handling complex decision-making scenarios. However, a comprehensive evaluation of performance metrics, considering accuracy, precision, and recall, identified K Nearest Neighbor as the most suitable model for demand forecasting in unorganized retail. C. Implications for Retailers:

For retailers in regions like Bangladesh, characterized by a plethora of diverse shops and consumer segments, the findings underscore the importance of adopting advanced predictive methods. The identified best model, K-Nearest Neighbor, can serve as a valuable tool for retailers seeking to optimize inventory, allocate resources efficiently, and make informed decisions based on accurate demand forecasts. D. Future Directions:
While this research has made significant strides in advancing the understanding of demand forecasting in unorganized retail, there remains ample room for future exploration.

Considerations for expanding the dataset, refining algorithms, and integrating real-time data streams could enhance the predictive capabilities of models. Additionally, collaboration with retailers and stakeholders can provide valuable insights into the practical implementation of these forecasting techniques.

In conclusion, the integration of machine learning algorithms in demand forecasting for unorganized retail is a dynamic and evolving field. This research, by incorporating a diverse set of attributes, meticulous data preprocessing, and a nuanced evaluation of algorithms, contributes valuable insights to the ongoing discourse. The identified

best model, K-Nearest Neighbor, serves as a practical and effective solution for retailers striving to navigate the complexities of unorganized retail and anticipate consumer demands with precision.

REFERENCES

1. Md. Ariful Islam Arif, Saiful Islam Sany, Faiza Islam Nahin and A.K.M. Shahariar Azad Rabby, "Comparison Study: Product Demand Forecasting with Machine Learning for Shop", Proceedings of the SMART–2019, IEEE Conference ID: 468668th International Conference on System Modeling & Advancement in Research Trends, pp. 171-176, 22nd–23rd November, 2019.
2. Muhammad Sajawal, Sardar Usman, Hamed Sanad Alshaikh, Asad Hayat and M. Usman Ashraf, "A Predictive Analysis of Retail Sales Forecasting using Machine Learning Techniques", LGU Research Journal of Computer Science & IT, Vol. 06 Issue 04, pp. 33-43, October – December 2022, doi:10.54692/lgurjcsit.2022.0604399.
3. Purvika Bajaj, Renesa Ray, Shivani Shedge, Shravani Vidhate and Prof.Dr. Nikhilkumar Shardoor, "Sales Prediction using Machine Learning Algorithm" , International Research Journal Of Engineering & Tech, Vol. 07 Issue 06, pp. 3619-3625, June 2020.
4. Nikita Malik and Karan Singh, "SALES PREDICTION MODEL FOR BIG MART", MSI Janakpuri, New Delhi, Vol. 03, pp. 22-32, July 2020.
5. Rao Faizan Ali, Amgad Muneer, Ahmed Almaghthawi, Amal Alghamdi, Suliman Mohamed Fati and Ebrahim Abdulwasea Abdullah Ghaleb,"BMSP-ML: big mart sales prediction using different machine learning techniques", IAES International Journal of Artificial Intelligence (IJ-AI), Vol. 12, No. 2, pp. 874-883, June 2023, doi:10.11591/ijaiv12.i2.
6. Jingru Wang,"A hybrid machine learning model for sales prediction", International Conference on Intelligent Computing and Human Computer Interaction (ICHCI), pp. 363-366, 2020, doi:10.1109/ICHCI51889.2020.00083.
7. Michael Giering. "Retail Sales Prediction and Item Recommendations Using Customer Demographics at Store Level", ACM SIGKDD Explorations Newsletter, Volume 10, Issue 2 , pp. 84–89, December 2008.
8. Ruiyun Kang,"Sales Prediction of Big Mart Based on Linear Regression, Random Forest, and Gradient Boosting", Proceedings of the 2nd International Conference on Business and Policy Studies Vol. 17,pp. 200207, 13 September 2023.
9. Sai Nikhil Boyapati and Ramesh Mummidi, "Predicting sales using Machine Learning Techniques", DV1478 Bachelor Thesis in Computer Science, this thesis is submitted to the Faculty of Computing at Blekinge Institute of Technology in 2020.
10. Aakanksha Ramesh Jadhav and Dr Ramesh D Jadhav, "Machine Learning for Sales Prediction in Big Mart". Sinhgad Institutes - Sinhgad Institute of Technology and Science, August 3, 2023.