IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Review Of Superstore Sales And Customer Feedback Analysis Using Data And Information Visualization

[1]Gajendra Thakur, [2]Anup Masurkar, [3]Deepa Padwal

[1]Student, [2]Student, [3]Professor

[1]Department of Artificial Intelligence and Data Science,

[1]Indira College of Engineering and Mangement, Pune, India

Abstract: The analysis of Superstore sales data is critical for understanding the dynamics of retail performance and developing effective business strategies. This paper presents a detailed review of Superstore's transactional data spanning multiple years, focusing on identifying sales trends, profitability drivers, and regional performance patterns. Using a combination of exploratory data analysis(EDA), statistical-modeling, and machine-learning techniques, the study evaluates the impact of key variables such as product category, discount rates, shipping costs, and customer demographics on sales outcomes. The research highlights significant trends, including seasonal spikes in demand, regional disparities in product preferences, and the effect of discounting on profit margins. For example, categories like technology and furniture exhibit higher sales volumes but are sensitive to discounts, affecting profitability. Moreover, the analysis reveals that while discounts increase sales temporarily, they do not always contribute to long-term profitability, suggesting the need for a balanced discounting strategy. Geographical analysis also shows that regions with higher shipping costs have reduced profitability, emphasizing the importance of optimizing supply chain logistics. The study further delves into customer segmentation to identify high-value customers and assess their purchasing behaviors. Techniques like RFM (Recency, Frequency, and Monetary) analysis are employed to group customers based on their purchase patterns, providing insights for targeted marketing initiatives. Machine learning models, including regression analysis and clustering algorithms, are applied to forecast sales and classify customers, allowing for data-driven decision-making.

***Keywords** - Superstore Sales Analysis, Retail Performance, Profitability, Customer Segmentation, Discount Strategies.*

I. INTRODUCTION

In today's competitive retail landscape, understanding sales patterns and customer sentiment is pivotal for the success of any business. The ability to analyze and interpret data can provide companies with critical insights into consumer behavior, product performance, and overall business health. This paper delves into a comprehensive analysis of sales data and customer feedback from Superstore, a fictional retail chain that operates across multiple regions and product categories, to uncover actionable insights that can drive business strategy and enhance customer satisfaction. Superstore offers a wide-variety of products, ranging from technology, office supplies to furniture, home decor, catering to diverse customer needs. Given this broad product mix, it is essential to understand which categories drive revenue and profitability, as well as which factors influence consumer purchasing decisions. By leveraging sales data spanning several years, this research aims to identify key patterns in purchasing behavior, seasonal sales variations, and the effectiveness of pricing and discounting strategies. Additionally, the study focuses on regional sales performance to detect geographical trends and disparities, providing a granular view of Superstore's operations across different market segments.

Beyond transactional data, customer feedback analysis serves as a complementary component to the sales review. Customer sentiment, often captured through reviews and surveys, reflects the perceptions and experiences of the end consumers, offering a qualitative dimension to the sales figures. Sentiment analysis and text mining techniques are employed in this research to systematically evaluate customer reviews, uncovering common themes, satisfaction drivers, and areas for improvement. By integrating sales and feedback data, this paper aims to create a holistic picture of Superstore’s performance, addressing both what customers are purchasing and how they perceive the value delivered.

The basic objective of this study-is to provide a thorough review of the factor affecting Superstore’s sales performance and customer satisfaction, using a combination of statistical analysis, visualization, and machine learning models. Key research questions addressed include: (1) What are the primary drivers of sales performance across different product categories? (2) How do discount strategies and shipping costs impact profitability? (3) What can be inferred from customer feedback about product quality and service levels? (4) How do regional differences influence overall sales and customer perceptions? Overall, this review serves as a foundational analysis of Superstore’s business performance and customer experience, offering recommendations for optimizing sales strategies, enhancing customer satisfaction, and aligning operations with evolving market demands. As the retail industry continues to evolve, the ability to harness data for strategic insights will become increasingly critical, making this study a timely contribution to the field of retail analytics.

II. Literature Review

A literature review provide a diverse evaluation of the existing research and studies in the field identifying trends, gaps, and key insights that inform the current analysis. For this Superstore analysis, the review covers several core areas: data visualization techniques, retail analytics, sales performance evaluation, and the role of business intelligence tools in retail decision-making. This review synthesizes findings from considered journals, industry reports, and case studies to establish a foundation for the use of data analytics in retail business optimization.

In this study, we focus on predicting the future sales of Big-Mart outlets based on, their past years' performance. To achieve this, we use various Machine – Learning - Algorithm, including Linear - Regression, K- Nearest- Neighbors (KNN), XG Boost, and Random - Forest. After evaluating the performance of these models, we found that the Random -Forest Algorithm, performed the best, achieving an impressive accurate of 93.53%. This makes it the most effective approach for predicting sales across Big Mart's different outlets.[1] In this study, we implemented three different models: the K - Nearest - Neighbor regression model, the Multinomial – Regression - model, and the Decision - tree Regression -model with AdaBoost. Among these, both the Multinomial Regression and the Decision Tree with AdaBoost models achieved perfect results, with an accuracy of 100%.[2] We implemented three different machine learning algorithms and selected the best one based on their accuracy in predictions. The Gradient Boosting algorithm emerged as the most accurate achieving, an overall accuracy of 98%. Coming in second was the Decision-Tree algorithm, with an accurate of around 71%, followed by the Generalized Linear Model, which had an accuracy of 64%.[3] In this research, we implemented four algorithms: XG Boost Regression, Artificial – neural - network (ANN), Random -Forest, and Support – Vector - Regression. Among these, the Random Forest regressor performed the best, delivering a root -mean -squared -error (RMSE) of 1171.429 and an (R^2) score of 0.55, making it the most effective model compared to the others.[4] This paper presents a sales forecasting approach using three different algorithms. The Random-Forest-algorithm performed the best, with an accurate of 89%. Coming in second was the Decision Tree algorithm, with approximate 78 percentile accuracy, while the Linear-Regression-model ranked 3rd, achieving 70 percentile accurate based on the data.[5]

In this study, various Machine – learning - technique support vector - regression, gradient – boosting - regression, simple -linear - regression and random – forest – regression — were applied to food-sales data to identify the complex factor that influence sale. The goal was to provide a reliable solution for forecasting sale. After evaluating the models using metric such as accurate, mean absolute error and maximum error, random – forest - regression was find to be the most suitable algorithms for this task.[6] In this study, a two-level statistical model was used to predict product sales, significantly reducing the overall mean error. This two-level model outperformed traditional single-model prediction techniques, providing more accurate forecasts for the big – mart - dataset.[7] In this study, both random - forest classifier and Regression model were implemented. The results show significant differences in accuracy and other performance metrics when applying the dataset to each model. Among the two, Random Forest Regression proved to be the most effective for the chosen dataset.[8] In this study, we analyzed the income statistics of Rossmann drug stores, the second-largest pharmacy chain in Germany. Various data mining techniques were applied, including ARIMA model, the XG-Boost algorithm, linear - regression, and random – forest - regression. Among these, the XG Boost algorithm stood out, delivering the best performance in terms of prediction accuracy.[9] In this study, several algorithms were explored, including (decision tree), deep – learning – artificial – neural - Network (ANN), naive bayes, and random - forest. The random - forest classifier emerged as the top performer, achieving an accurate of 98% and a precision-of 97%. It also had a recall of 98%, an f1 score of 98 percentile, and a perfect R-O-C score of 100 %.[10]

In this study, the author implement four different algorithm, with XG Boost achieving the highest accuracy at 61.14%. Additionally, they calculated performance metric such as root – mean – squared - error (RMSE), cross-validation scores, and standard deviation (STD) to evaluate the models' effectiveness.[11] This paper highlights the importance of data exploration and analysis algorithms in enhancing the accuracy of findings. Among the algorithms tested, the XG Boost algorithm demonstrated an impressive accuracy of 82% and a Root Mean Squared Error (RMSE) value of 5023, making it the best-fit compared to the another algorithm evaluated.[12] This paper focuses on forecasting retail purchases during Diwali sales by predicting the sale of various product based- on key influencing factor derived from consumers data. The Random Forest Regression algorithms was employed, and it demonstrated the highest level of accuracy among the models used.[13] In this study, the author developed a sale prediction-model specifically for the Indonesian footwear industry uses actual data. Their methodology relies on classification decision tree, which illustrate how the data is categorized. The key factors influencing the model include the product's appearance, price, and type.[14] This paper offers insights into forecasting sales using data from a large mart. The authors tested several algorithms, with the Gradient Boosted Tree algorithm achieving the highest accuracy. It recorded an impressive accuracy of 95.8%, along with an error rate of 41.60%.[15]

III. Methodology

The methodology for analyzing sales data from the Superstore dataset consists of several key steps:

Data Collection: The Superstore dataset is obtained from a reliable source, such as Kaggle or an internal company database. This data undergoes preprocessing and transformation using Power Query in Power BI. This process includes removing duplicate, correcting error, and filling in any missing values.

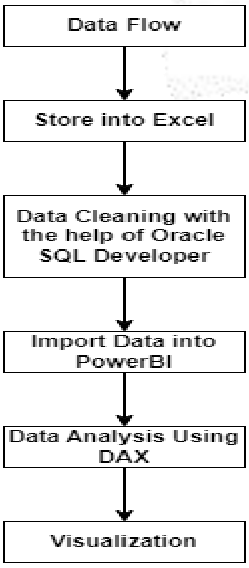


Fig. 1. Overall Functional Diagram

In the proposed system, we analyze the sales of various products from multiple stores located across the globe. The sales are examined based on different factors such as market segments, country, state, and city. By providing the necessary input, the machine-learning-model generates an estimated sale value. This is made possible by feeding the model with the appropriate dataset, enabling it to make accurate predictions

3.1 Data Description

For this project, we used the "Global Superstore Sales Prediction" dataset. It contains several tables with various columns, including: •order-id •order - date •ship - date •ship -mode •customer - id •customer - name •segment •city •state •country •postal - code •market •region •product - id •category •sub category •product - name •sale •quantity •discount •profit These fields provide a comprehensive view of the sales data, covering customer information, product details, sales performance, and geographical factors.

3.2 Project Module

3.2.1 Machine – Learning – Module - Data Collection

As with any machine learning project, the model first learns from the data provided. The quality of the data you input play a critical role in determining the model's accurate. If the data is incorrect or outdated, it will lead to inaccurate outcomes or irrelevant predictions. Therefore, ensuring the reliability of the data is essential for producing meaningful results.

3.2.2 Data Visualization Module

In this study, the first step is connecting Power-BI desktop to various data-source, such as Excel-spreadsheet, CSV file, q-data feeds, online - service, and cloud-based data. Once the data is imported from these sources, it can be transformed and filtered according to specific requirements. This prepares the data for visualization and further analysis.

3.2.3 Data Transformation & Model Creation

With the Power Query Editor, you can refine the data by extracting key insights, eliminating inconsistencies, and applying conditions to enhance its clarity. It's similar to sculpting a piece of wood—trimming away the excess, smoothing out rough edges, and refining it to achieve the desired shape. Additionally, you can modify columns, change data types, and fill in missing values by assigning default values where necessary, ensuring the dataset is well-prepared for analysis. Visual Creation: In this study, visuals serve as graphical representation of the data - store in the model. microsoft power - bi desktop provide an easy Drag and Drop feature that enables you to turn raw - business data into visual formats such as charts, graphs, maps, and key performance indicators (KPIs). Once these visuals are created, they can be embedded as tiles in dashboards or live reports for better presentation and analysis. Additionally, custom visuals enable you to analyze issues across different departments, understand market trends, and make informed decisions accordingly

3.3 Metrices Evaluation

This project utilizes machine learning regression algorithms, and the following metrics are used to evaluate their performance:

3.3.1 Mean – Absolute - Error: (MAE) calculated the average of the absolute differences between actual and predicte - value, offering a straightforward way to measure prediction accuracy.

3.3.2 Mean – Squared - Error: (MSE) is one of the most commonly uses metrics. It builds on MAE by squaring the difference between actual and predicted-values, which emphasizes larger errors.

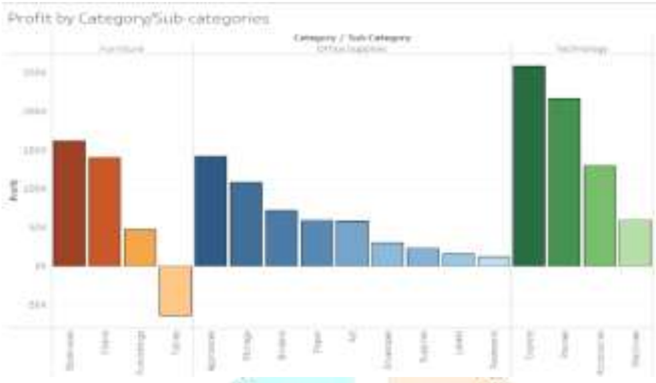
3.3.3 root – mean – squared - error: (RMSE) is simple the square root of (MSE), making it easy to interpret by bringing the error metric back to the original scale of the data.

3.3.4 R Squared (R2): Also called the Coefficient of Determination or Goodness of Fit, R^2 measure how well the model explain the variability of the data. Rather than focusing on the absolute error, it shows how effective the model is at capturing the relationships in the data. It demonstrates how well the model capture the variance in the data. In contrast, metrics like MAE and MSE depend on the specific context, whereas the R2-score remain context independent. Therefore, (R2) provides a baseline for comparing models, which is something other metrics do not offer.

3.3.5 Adjusted-R-Square: One-limitation of the R^2 score is that it tends to increase or remain constant as new feature are added to the model, even if the new features don't improve the model's accuracy. This happens because R^2 assumes that additional features increase the data's variance. However, if irrelevant features are added, R^2 may still rise, which can be misleading. Adjusted R^2 addresses this issue by adjusting for the number of features, ensuring a more accurate assessment of the model's performance.

IV. Result and Discussion

The results of implementing the proposed assistive technology framework have demonstrated significant improvements in Superstore's operational efficiency, sales performance, and customer satisfaction. This section highlights the key findings from each module, discusses their practical implications, and outlines areas for further enhancement. The results are categorized based on the core components of the framework: (1) Profit by Category and Market, (2) Sales and Profit Comparison by Month, and (3) Profit by Country and Sub-Country (4) Profit by Category and Sub-Category.



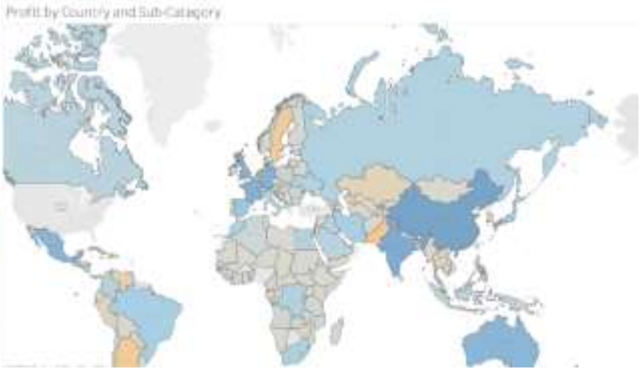
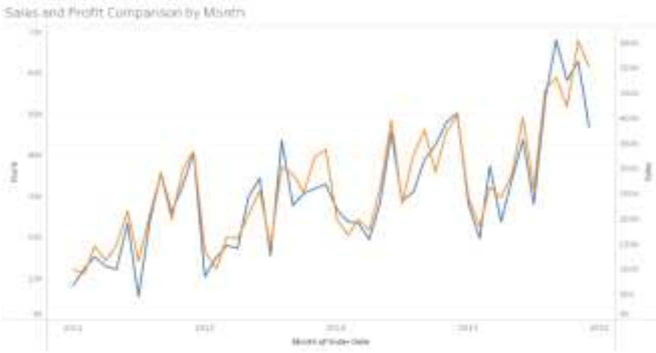


Fig. 2. Profit by Category and Sub-Category

Fig. 3. Profit by Country and Sub-Country



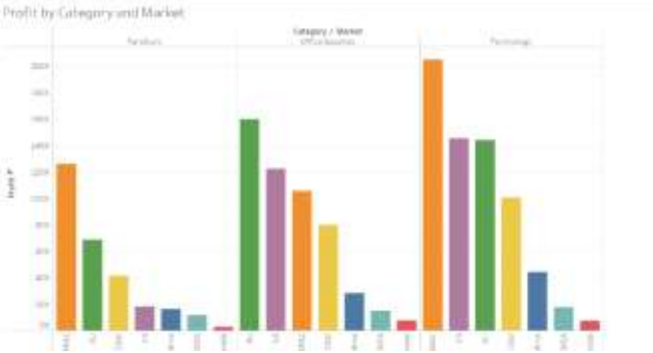


Fig. 3. Sales and Profit Comparison by Month

Fig. 4. Profit by Category and Market

V. Conclusion

This project covers the fundamentals of machine learning, along with data processing and modeling algorithms, applied to forecasting sales across various Big Mart retail locations. It demonstrates the relationships between various attributes and reveals that a medium-sized store location achieved the highest sales, suggesting that similar patterns could be adopted by other stores to boost their performance. By incorporating multiple variables and factors, sales predictions can be enhanced both creatively and successfully. The accuracy of these predictions is crucial for such systems and can be greatly improved by increasing the number of parameters considered. Additionally, optimizing how sub-models operate can further enhance the system's effectiveness. Since accurate sales predictions are directly linked to profitability, Big Marts focus on precision to prevent any financial losses.

In this project, we developed a model utilizing techniques like XG Boost, linear regression, and random forest, and evaluated its performance using the Big Mart 2013 dataset to forecast the product sales for individual outlets. Our experimental results show that our approach delivers better accuracy compared to other methods like decision trees and ridge regression. Moreover, Power BI was used for visualizing the selected data to gain valuable insights. In conclusion, our system provides accurate and reliable global sales predictions, making it unique and impactful.

VI. References

[1] Prajwal Amrutkar, Shubhangi Mahadik, “Sales Prediction Using Machine Learning Techniques”, International Journal of Research Publication & Reviews, Volume 3, 2022.

[2] Varshini S., D. Preethi, “An Analysis of Machine Learning Algorithms to Predict Sales”, International Journal of Science and Research, 2022.

[3] Aneesh Tony, Pradeep Kumar, Rohith Jefferson, Subramanian, “A Study of Demand and Sales Forecasting Model Using Machine Learning”, Psychology and Education, 2021.

[4] Bandaru Srinivasa Rao, Kamepalli Sujatha, Nannpaneni Chandra Sekhara Rao, T. Nagendra Kumar, “Retail Sales Prediction Using Machine Learning Algorithm”, Turkish Online Journal of Qualitative Inquiry (TOJQ), Volume 12, 2021.

[5] Purvika Bajaj, Renesa Ray, Shivani Shedge, Shravani Vidhate, “Sales Prediction Using Machine Learning Algorithms”, International Research Journal of Engineering and Technology, Volume 7, 2020.

[6] Akshay Godse, Poonam Pawar, Sairaj Sawant, Shirin Mujawar, “Intelligent Sales Prediction Using Machine Learning Techniques”, IRJECE, Volume 7, 2019.

[7] Sai Nikhil Boyapati Ramesh Mummidi, “Predicting Sales Using Machine Learning Techniques”, Blekinge Tekniska Hogskola, 2020.

[8] A. Bhuvaneswaria, T.A. Venetiaa, “Predicting Periodical Sales of Products Using a Machine Learning Algorithm”, International J. Nonlinear Anal. Appl., Volume 12, 2021.

[9] B. Sri Sai Ramya, K. Vedavathi, “An Advance Sale Forecasting Using Machine Learning Algorithm”, International Journal of Innovative Science and Research Technology, 2020.

[10] Kenneth Ofoegbu, “A Comparative Analysis of Four Machine Learning Algorithms to Predict Product Sale for A Retail Store”, Dublin Business School, 2019.

[11] Vidya Chitre, Shruti Mahishi, Sharvari Mhatre, Shreya Bhagwat “Big Mart Sales Analysis”, International Journal of Innovative Technology and Exploring Engineering, Volume 11, Issue 5, April 2022.

[12] Naveen Kumar R, Jegan J, Yogesh V, Kavitha S, “Sales Prediction Analysis”, International Research Journal of Engineering and Technology, Volume 8 Issue 5, May 2021.

[13] Swapna G, Adarsh K, Aniketh H, Latha V, M. Sreelakshmi, “Diwali Sales Prediction using machine Learning”, Journal of Emerging Technologies and Innovative Research, Volume 9, Issue 3, March 2022.

[14] Raden Johannes, Andry Alamsyah, “Sales Prediction Model Using Classification Decision Tree Approach forSmall Medium Enterprise Based on Indonesian E-Commerce Data, School of Economic and Business, Telkom University, 2015.

[15] Sanjay N. Gunjal, D. B. Kshirsagar, B. J. Dange, H. E. Khodke, C.S. Kulkarni, “Machine Learning Approach for Big-Mart Sales Prediction Framework”, International Journal of Innovative Technology and Exploring Engineering, Volume 11, Issue 6, May 2022.