Available online @ https://jjem.jnnce.ac.in
https: www.doi.org/10.37314/JJEM.SP0216
Indexed in International Scientific Indiexing (ISI)
Impact factor: 1.395 for 2021-22
Published on: 08 December 2024

Super Market Sales Prediction

Swathi G S[1], Arun Kumar K L[2]

[1]Department of Computer Applications, JNNCE, Shimoga
[2] Department of Computer Applications, JNNCE, Shimoga,

swathiswapnags@gmail.com, arunkumarkl@jnnce.ac.in

Abstract

supermarket sales are not always easier seasons holidays and other frequent events may significantly affect sales which in turn impacts the accuracy and dependability of the forecasts there are other issues due to the granularity and quality of data supermarkets and shopping centers now days keep an eye related to certain goods in an effort to forecast future demand from customers employees in supermarket sales prediction apply a types of technical skills to project future sales and meet predetermined goals here the gradient boosting regression GBR plan expertise method algorithm is utilized to gather data by analyzing previous historical sales and incorporating information about promotions holidays and other factors like weather and economic indicators by accurately predicting demand supermarkets can better meet the needs of their customers which increases sales and revenue.

Keywords: *Grocery sales prediction, sales forecasting, temporal data analysis, automated analysis, retail analytics*

1. Introduction:

In today's competitive retail environment, accurately predicting sales is crucial for effective inventory management, resource allocation, and strategic planning. This project aims to address these challenges by developing a robust model to forecast grocery sales using historical data. We leverage machine learning methods, specifically the Gradient Boosting algorithm, to provide reliable sales predictions that assist retailers in optimizing their operations. The dataset used in this project includes features such as store identification, item categorization, and previous sales records. A critical step in ensuring model accuracy was data preprocessing. This involved handling missing data, normalizing numerical features, and creating new time-based features to capture seasonality and trends. By transforming raw data into a more meaningful format, we set a solid foundation for model training.

To evaluate the model's performance, we used metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). These metrics provide a quantitative measure of the model's accuracy, allowing us to compare different configurations and select the best-performing model. Additionally, hyperparameter tuning was conducted to optimize the model further, ensuring precise sales forecasts. This project demonstrates the application of advanced machine learning techniques to solve a practical problem in the retail sector. The developed sales prediction model not only aids in better inventory management but also supports strategic decision-making by providing accurate and actionable insights.

2. Literature survey:

Saunders, Gammerman, and Vovk [6] introduce the Ridge Regression learning method Dual Variables. focusing on dual variables. The paper discusses how ridge regression mitigates multicollinearity and overfitting issues by adding a regularization term to the ordinary least squares (OLS) objective function. It explores the algorithm's effectiveness in improving model stability and generalization performance, particularly in scenarios with high-dimensional data or correlated predictors. The study underscores ridge regression's role in enhancing predictive accuracy and robustness in Machine Learning applications. Chu and Zhang [7] conducted a study comparing the performance using linear and nonlinear models to forecast total retail sales. Using real-world data, the study assessed the accuracy and suitability of these approaches in predicting sales trends. The research highlighted that nonlinear models frequently yield superior results over linear ones by effectively capturing the intricate relationships present in retail sales data. This capability enhances their ability to provide precise forecasts, offering valuable insights for retail operational decision-making.

Giuseppe Nunnari and Valeria Nunnari [8] conducted a study on predicting monthly retail sales trends data.A Case Study": Nunnari and Nunnari present a case study on forecasting monthly retail sales using time series analysis. The research investigates methods for analyzing sales data over time to predict future sales trends accurately. It explores the use of statistical models and machine learning algorithms in generating forecasts, highlighting their effectiveness in assisting retail businesses with demand strategies for planning and managing inventory. Isaac O. Ajao and Adedeji A. Abdullahiand I. Raji Ismail [12] conducted a study on using a Polynomial Regression Model to predict manufacturing costs in mixed-cost analysis.This The paper introduces a polynomial regression model to predict costs in costs that are mixed analysis. It explores how polynomial regression techniques can effectively capture the relationship between variables in mixed cost scenarios, providing accurate cost predictions crucial for financial planning and decision-making. The study emphasizes the applicability of polynomial regression in analyzing and forecasting costs in complex business environments, offering insights into improving cost management strategies through mathematical modeling. Suma, V. and the Shavige Malleshwara [13] Hills explore data mining for predicting This paper centers on forecasting the demand for reconditioned electronics in the Indian market. using data mining techniques. It examines how algorithms like decision trees and clustering help forecast consumer preferences and market trends. The study underscores the importance of data-driven approaches in anticipating demand patterns, supporting efficient inventory management, and meeting consumer needs in the vibrant market for reconditioned electronics. Wang, Haoxiang [14] discusses Wang explores sustainable growth and management in consumer sectors electronics, employing soft computing methods. techniques. The paper explores how techniques like fuzzy logic and neural networks, and genetic algorithms contribute to optimizing operations and improving sustainability in the electronics industry. It highlights soft computation's role in enhancing decision-making processes, reducing environmental impact, and fostering long-term viability in consumer electronics production and management. Zone-Ching Lin and Wen-Jang Wu [16] performed a analysis using multiple linear regression on the.Overlay Accuracy Model Zone": Lin and Wu's study focuses using multiple linear regression analysis used for the model's accuracy overlay in semiconductor manufacturing. It investigates factors influencing overlay accuracy and develops regression models to predict and optimize manufacturing processes. The research underscores the significance of statistical modeling in semiconductor manufacturing for improving production yield and quality control.

2. Methodology:

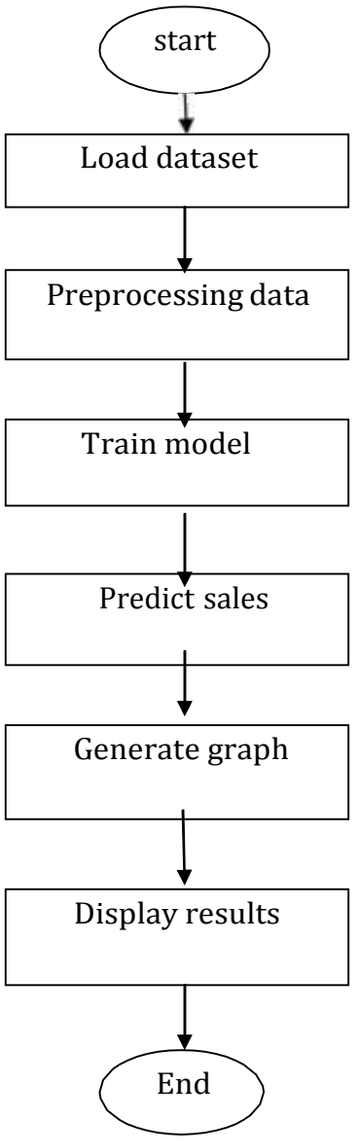

Figure 1: Flow chart

2.1 Load Dataset This initial stage in the process chart represents the process of loading the raw data from the dataset into the computational environment. This involves reading data files, such as CSV or Excel files, from storage into memory for further processing.

2.2 Preprocessing Data: Once the dataset is loaded, the subsequent task is preprocess the data. This includes cleaning the data using handling missing values, outliers, and formatting inconsistencies. Additionally, preprocessing involves transforming the data into a suitable format for model fitting include feature scaling, encoding categorical variables, and splitting the data into training and testing sets.

2.3 Train Model: With the preprocessed data, the machine learning model is trained. This step involves selecting an appropriate algorithm (such as Gradient Boosting) and fitting it to the training data. During training, the model learns patterns and relationships in the data to make accurate predictions.

2.4 Predict sales: After the model is configured trained, it is used to predict future grocery sales based on new input data. This step applies the fitted model to the test or validation dataset to generate predictions for sales quantities.

2.5 Generate graph: Once predictions are made, a graph or visualization is generated to demonstrate predicted sales trends. This could include line charts, bar charts, or other visual tools that assist illustrate how well the model predicts sales over time.

2.6 Display results: The last step involves displaying the results to stakeholders or end-users. This includes presenting the predicted sales figures and accompanying visualizations to provide insights into future sales trends derived from machine learning model.

3. Experimental Result:

The outcomes of our grocery sales prediction model using the Gradient Boosting algorithm were promising. The model achieved a Mean Absolute Error (MAE) of 0.53, indicating a high level of precision in forecasting the sales figures. Additionally, the Root Average Squared Error (RMSE) was calculated to be 0.75, The assessment score Gradient Boosting algorithm in our grocery sales prediction project is 0.9791770885009521. This a high rating signifies model explains approximately

97.92% of the variance in the grocery sales data, demonstrating a very high level of accuracy in predicting sales, further confirming the model's precision in capturing the variations in the sales data. The training process involved optimizing the hyper parameters, which significantly improved the model's performance. Visualizations of the predicted versus actual sales values showed a strong correlation, highlighting the model's effectiveness in predicting future sales. These results demonstrate the potential of using Gradient Boosting for accurate sales predictions in the grocery sector.

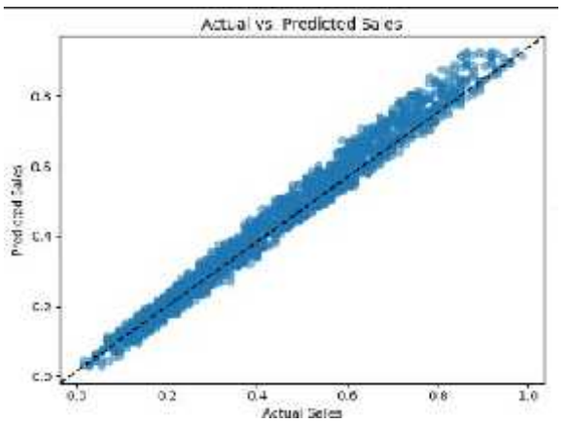

Figure: 2 linear regressions of predictedsales

The function adds a dotted diagonal line could be described as a dotted at an angle indicate where a flawless rephrasing of the word would be ideal forecasts would all are labeled on the x-axis represents actual sales represented on the y-axis representing predictions titles relate to revenue the plot actual versus predicted revenue locations near the line implies precision forecasts**.**

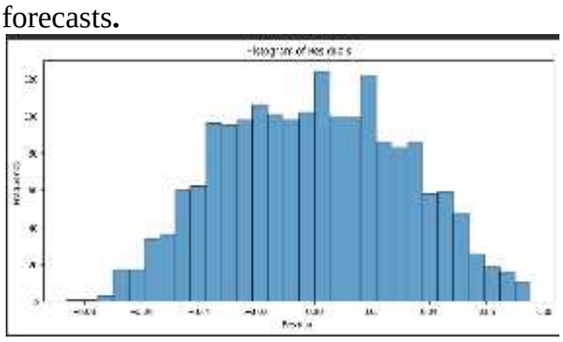

Figure 3: Histogram of predicted sales**.**

The code creates histogram be seen the handling of leavings or characteristics between actual and desired values it boasts a 12 by 5 inch figure 30 bins with black sides and semi-transparent bars the x-axis is labeled leftovers the y-axis is labeled frequency and the plots title is histogram of residuals

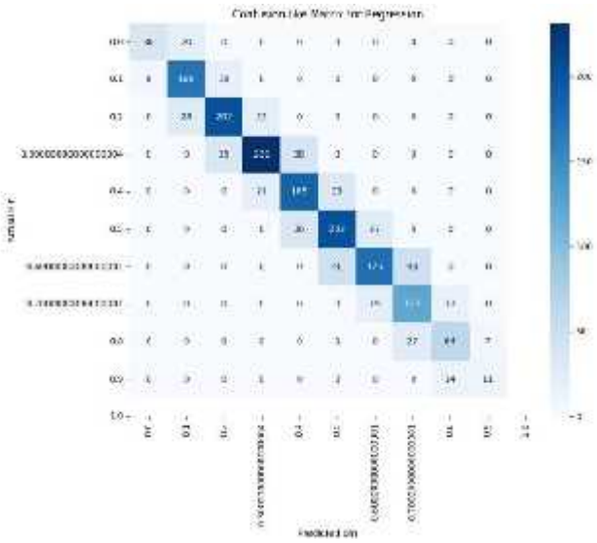

Figure 4: Confusion matrix of predictedsales

The above figure is the confusion matrix with in relation to sale prediction.

4. Conclusion:

In conclusion, our project aims to predict grocery sales prediction utilizing machine learning models. We explored different algorithms like Gradient boosting regression to identify the optimal one for accurate predictions. Through our analysis this approach can help grocery stores plan better for future sales and manage their inventory efficiently. Overall, our project demonstrates the importance of using advanced techniques in predicting sales to improve business decision-making within grocery industry.

References:

1. Arunkumar K L, Ajit Danti ,"A NOVEL APPROACH FOR VEHICLE RECOGNITION BASED ON THE TAIL LIGHTS GEOMETRICAL FEATURES IN THE NIGHT VISION", International Journal of Computer Engineering and Applications, Volume XII, Issue I, Jan, www.ijcea.com ISSN 2321-3469(2018)

2. Arunkumar K L, Ajit Danti ,"Recognition of Vehicle using geometrical features of a tail light in the night vision", ,National Conference on Computation Science and Soft Computing (NCCSSC-2018)

3. Arunkumar K L, Ajit Danti, Manjunatha H T ,"Estimation of vehicle distance based on feature points using monocular vision", , IEEE 8816996, 1-5 (2019)

4. Arunkumar K L, Ajit Danti, Manjunatha HT , D Rohith ,"Classification of Vehicle Type on Indian Road Scene Based on Deep Learning", , Springer, Singapore 1380, 1-10 (2021)

5. Aunkumar K L, Ajit Danti, Manjunatha H T, “Classification of Vehicle Make Based on Geometric Features and Appearance-Based Attributes Under Complex Background", Springer 1035 (CCIS), pp 41-48(2019)

6. C. Saunders, A. Gammerman and V. Vovk, “Ridge Regression Learning Algorithm in Dual Variables”, Proc. of Int. Conf. on Machine Learning, pp. 515 – 521, July 1998.IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 56, NO. 7, JULY , 3561,2010

7. Ching Wu Chu and Guoqiang Peter Zhang, “A comparative study of linear and nonlinear models for aggregate retails sales forecasting”, Int. Journal Production Economics, vol. 86, pp. 217- 231, 2003.

8. Giuseppe Nunnari, Valeria Nunnari, “Forecasting Monthly Sales Retail Time Series: A Case Study”, Proc. of IEEE Conf. on Business Informatics (CBI), July 2017.

9. Manjunatha H T Arunkumar K L, Ajit Danti ,"A Novel Approach for Detection and Recognition of Traffic Signs for Automatic Driver Assistance System Under Cluttered Background" , Springer 1035 (CCIS), pp 407-419(2019).

10. Manjunatha H T, Ajit Danti, ArunKumar K L, D Rohith ,"Indian Road Lanes Detection Based on Regression and clustering using Video Processing Techniques", , Springer, Singapore 1380 (CCIS), 193-206(2021).

11. Nrupatunga, KL Arunkumar, Peruse and Recognition of Old Kannada Stone Inscription Characters, CM, Springer, Singapore, 2020

12. O. Ajao Isaac, A. Abdullahi Adedeji, I. Raji Ismail, “Polynomial Regression Model of Making Cost Prediction In Mixed Cost Analysis”, Int. Journal on Mathematical Theory and Modeling, vol. 2, no. 2, pp. 14 – 23, 2012.

13. Suma, V., and Shavige Malleshwara Hills. "Data Mining based Prediction of Demand in Indian Market for Refurbished Electronics." Journal of Soft Computing Paradigm (JSCP) 2, no. 02 (2020): 101- 110 41

14. Wang, Haoxiang. "Sustainable development and management in consumer electronics using soft computation." Journal of Soft Computing Paradigm (JSCP) 1, no. 01 (2019): 56.-2.

15. Xinqing Shu, Pan Wang, “An Improved Adaboost Algorithm based on Uncertain Functions”, Proc. of Int. Conf. on Industrial Informatics Computing Technology, Intelligent Technology, Industrial Information Integration, Dec. 2015.

16. Zone-Ching Lin, Wen-Jang Wu, “Multiple Linear Regression Analysis of the Overlay Accuracy Model Zone”, IEEE Trans. On Semiconductor Manufacturing, vol. 12, no. 2, pp. 229 – 237, May 1999.