ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024

PREDICTIVE ANALYSIS FOR BIG MART SALES USING MACHINE LEARNING ALGORITHMS

K. Renuka, P. Srinidhi, B. Sai Pooja, Maganti Divya Smruthi

ASSISTANT PROFESSOR[1], UG SCHOLAR[2,3&4]

DEPARTMENT OF CSE, CMR INSTITUTE OF TECHNOLOGY, KANDLAKOYA VILLAGE, MEDCHAL RD, HYDERABAD, TELANGANA 501401

ABSTRACT - Retail giants like Big Mart utilize sophisticated data tracking to manage individual item sales, aiming to predict consumer demand and optimize inventory. By mining their data warehouses, they can identify anomalies and trends, which are crucial for accurate sales forecasting. Implementing machine learning techniques has proven to enhance the predictive accuracy for future sales volumes. Specifically, a predictive model using Xgboost, Linear Regression, Polynomial Regression, and Ridge Regression has been developed for forecasting sales. Among these methods, the model leveraging Xgboost, known for its gradient boosting framework, has demonstrated superior performance compared to traditional models. Linear Regression, while straightforward and interpretable, sometimes falls short in capturing complex relationships. Polynomial Regression extends linear models to better fit non-linear data but can lead to overfitting. Ridge Regression, a variant of linear regression, addresses multicollinearity through regularization, thereby improving model robustness. The integration of these techniques in a predictive model has resulted in a tool that surpasses existing sales forecasting methods in accuracy and reliability. Retailers can thus leverage this advanced model to make more informed decisions, reduce overstock or stockouts, and ultimately enhance customer satisfaction and operational efficiency. This holistic approach to using machine learning for sales prediction exemplifies the critical role of advanced analytics in modern retail strategies.

I. INTRODUCTION The competitiveness between various shopping centres and large marts is intensifying, driven by the rapid development of global malls and the rise of online shopping. In this fiercely competitive environment, each market is compelled to offer personalized and time-sensitive deals to attract a larger customer base. This dynamic landscape necessitates the accurate estimation of sales volumes for effective stock control, transportation, and logistical services. Machine learning algorithms have become invaluable in this context, offering advanced methods for predicting sales across different organizations. These algorithms help retailers navigate the complexities of market demand, enabling them to optimize their inventory and operational efficiency. One of the most significant advantages of using machine learning for sales prediction is its ability to process vast amounts of data quickly and accurately. Techniques such as Xgboost, Linear Regression, Polynomial Regression, and Ridge Regression have been employed to enhance the predictive accuracy of sales forecasts. Xgboost, in particular, has gained popularity due to its robust performance in handling large datasets and complex patterns. Linear Regression, while simpler, provides a clear and interpretable model but may struggle with non-linear relationships in data. Polynomial Regression extends linear models by including interaction terms and polynomial terms, which can capture more complex trends at the risk of overfitting. Ridge Regression, which adds a

regularization term to the linear regression model, helps mitigate the issue of multicollinearity, thereby improving model stability and prediction accuracy.

The integration of these machine learning techniques allows retailers to develop sophisticated models that outperform traditional methods in predicting future sales volumes. Accurate sales forecasts are crucial for developing effective marketing strategies, ensuring adequate stock levels, and providing a seamless shopping experience for customers. By leveraging predictive analytics, retailers can anticipate market trends, adjust their strategies in real-time, and maintain a competitive edge. This not only helps in attracting and retaining customers but also in optimizing operational costs and improving overall business performance.

In the era of digital transformation, the ability to predict sales accurately is more critical than ever. Retailers who can harness the power of machine learning to gain insights from their sales data are better positioned to respond to market demands, tailor their offerings, and enhance customer satisfaction. As the competition continues to escalate, the role of advanced analytics in retail strategy becomes increasingly prominent. Retailers must invest in cutting-edge technologies and data-driven approaches to stay ahead in the market. By doing so, they can not only survive but thrive in the competitive retail landscape, ultimately driving growth and profitability. Thus, the application of advanced machine learning algorithms in sales prediction is a game-changer for retailers, providing them with the tools needed to navigate the complexities of modern retail and achieve long-term success.

II. LITERATURE SURVEY

A) Pei Chann Chang and Yen-Wen Wang, “Fuzzy Delphi and back propagation model for sales forecasting in PCB industry”, Expert systems with applications, vol. 30,pp. 715-726, 2006.

They proposed a novel approach for predicting sales in the printed circuit board (PCB) industry. They introduce a hybrid model that combines the Fuzzy Delphi Method (FDM) with a back propagation neural network (BPNN) to enhance the accuracy of sales forecasts. The Fuzzy Delphi Method is employed to gather and synthesize expert opinions, converting qualitative assessments into quantitative measures, which helps in addressing uncertainties and subjective judgments inherent in sales forecasting. These quantified expert inputs serve as the training data for the back propagation neural network, which is renowned for its capability to model complex nonlinear relationships within large datasets. By integrating FDM with BPNN, the authors aim to leverage the strengths of both methodologies: FDM’s ability to handle vagueness and ambiguity, and BPNN’s proficiency in learning from historical data to predict future trends. The study demonstrates that this hybrid model outperforms traditional forecasting techniques in terms of accuracy and reliability, providing a more robust tool for decision-makers in the PCB industry. The experimental results, validated with real-world sales data, indicate significant improvements in forecast precision, highlighting the model’s potential for practical applications in inventory management, production planning, and strategic marketing. Chang and Wang's research contributes to the field by offering an innovative solution that addresses the limitations of existing models, thus enhancing the predictive capabilities and operational efficiency of businesses within the PCB sector.

B) R. J. Kuo, Tung Lai HU and Zhen Yao Chen “application of radial basis function neural networks for sales forecasting”, Proc. of Int. Asian Conference on Informatics in control, automation, and robotics, pp. 325- 328, 2009.

It explores the use of radial basis function neural networks (RBFNN) for improving sales forecasting accuracy. The authors highlight the challenges associated with traditional sales forecasting methods, which often struggle with the non-linear and complex nature of market data. To address these challenges, they propose the application of RBFNN, a type of artificial neural network known for its strong capability in approximating complex functions and handling non-linear relationships effectively. The RBFNN model is designed to learn from historical sales data, identifying underlying patterns and making precise future sales predictions. Through a series of experiments, the authors demonstrate that the RBFNN model significantly outperforms conventional forecasting techniques, such as linear regression and moving averages, in terms of accuracy and reliability. The experimental results, validated using real-world sales data, indicate that the RBFNN model can capture the intricate dynamics of sales trends, leading to more accurate and dependable forecasts. This improved forecasting capability is crucial for businesses, as it enables better inventory management, optimized resource allocation, and enhanced strategic planning. The study concludes that the application of RBFNN in sales forecasting represents a substantial advancement over traditional methods, offering a powerful tool for decision-makers seeking to navigate the complexities of modern markets. Kuo, Hu, and Chen's research contributes to the field by providing empirical evidence of the effectiveness of RBFNN in sales forecasting and suggesting its potential for broader applications in various industries.

C) Suresh K and Praveen O, "Extracting of Patterns Using Mining Methods Over Damped Window," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2020, pp. 235-241, DOI: 10.1109/ICIRCA48905.2020.9182893.

They present an innovative approach for pattern extraction in data streams using advanced mining techniques. The study addresses the challenges of identifying significant patterns in dynamic data environments, where data continuously evolves and older data becomes less relevant over time. To tackle this issue, the authors propose a methodology based on damped window models, which assign decreasing weights to older data, ensuring that more recent information has a greater influence on the pattern extraction process. This approach effectively mitigates the impact of obsolete data and enhances the relevance of the extracted patterns. The research leverages various data mining methods to analyze data within these damped windows, focusing on identifying meaningful trends and patterns that can inform decision-making processes.

The authors conducted extensive experiments to evaluate the performance of their proposed model, demonstrating its effectiveness in real-time data stream scenarios. The results show that the damped window-based mining methods outperform traditional static window approaches by providing more accurate and timely insights into the data. This is particularly beneficial in applications requiring continuous monitoring and rapid response, such as stock market analysis, network security, and customer behavior analysis. The paper highlights the practical implications of their findings, suggesting that the proposed model can significantly enhance the ability of organizations to detect and respond to emerging trends and anomalies in their data streams.

Suresh K and Praveen O's research contributes to the field of data mining by offering a robust framework for pattern extraction in ever-changing data environments, emphasizing the importance of adaptive models that prioritize recent information. Their work opens avenues for further research and development in adaptive data mining techniques, potentially impacting various industries that rely on real-time data analysis for strategic decision-making.

IMPLEMENTATION

Modules

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Login, Browse Data Sets and Train & Test, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View All Antifraud Model for Internet Loan Prediction, Find Internet Loan Prediction Type Ratio, View Primary Stage Diabetic Prediction Ratio Results, Download Predicted Data Sets, View All Remote Users.

View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT PRIMARY STAGE DIABETIC STATUS, VIEW YOUR PROFILE.

CONCLUSION

In this study, we evaluated the effectiveness of various regression algorithms on sales data to identify the best-performing model for predicting future sales. By proposing a software tool that utilizes regression approaches, we aimed to enhance the accuracy of sales forecasts based on historical data. Our analysis included Linear Regression, Polynomial Regression, Ridge Regression, and Xgboost Regression. The results demonstrated that Ridge Regression and Xgboost Regression outperformed Linear and Polynomial Regression in terms of accuracy, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). These findings suggest that Ridge Regression, with its ability to handle multicollinearity through regularization, and Xgboost Regression, known for its robustness and performance in managing large datasets and complex patterns, provide superior predictive capabilities. Accurate sales forecasting is crucial for developing effective sales plans, which in turn helps organizations avoid unforeseen cash flow issues and manage production, staffing, and financing needs more efficiently. Our study highlights the importance of choosing the right regression model to achieve better predictive

accuracy, which can significantly impact strategic planning and operational efficiency. In future research, we plan to explore the integration of the ARIMA model, which is adept at handling time series data, to further enhance the forecasting capabilities. By incorporating time series analysis, we aim to capture temporal dependencies and trends that could provide deeper insights into sales patterns. This future direction holds promise for further improving sales forecasts, thereby enabling businesses to make more informed decisions and optimize their operations. Overall, our study underscores the value of advanced regression techniques in sales prediction and sets the stage for future innovations in predictive analytics.

REFERENCES

[1] 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.

[2] Wang, Haoxiang. "Sustainable development and management in consumer electronics using soft computation." Journal of Soft Computing Paradigm (JSCP) 1, no. 01 (2019): 56.- 2. Suma, V., and Shavige Malleshwara Hills. "Data Mining based Prediction of D

[3] 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

[4] Giuseppe Nunnari, Valeria Nunnari, "Forecasting Monthly Sales Retail Time Series: A Case Study", Proc. of IEEE Conf. on Business Informatics (CBI), July 2017.

[5]D. Fantazzini, Z. Toktamysova, Forecasting German car sales using Google data and multivariate models, Int. J. Production Economics 170 (2015) 97-135.

[6] Zone-Ching Lin, Wen-Jang Wu, "Multiple LinearRegression Analysis of the Overlay Accuracy Model Zone", IEEE Trans. on Semiconductor Manufacturing, vol. 12, no. 2, pp. 229 – 237, May 1999.

[7] 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.

[8] 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 2010 3561.

[9] "Robust Regression and Lasso". Huan Xu, Constantine Caramanis, Member, IEEE, and Shie Mannor, Senior Member, IEEE. 2015 International Conference on Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration."An improved Adaboost algorithm based on uncertain

ISSN 2454-9940
www.ijasem.org
Vol 18, Issue 2, 2024

functions".Shu Xinqing School of Automation Wuhan University of Technology.Wuhan, China Wang Pan School of the Automation Wuhan University of Technology Wuhan, China.

[10] 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.

[11] A. S. Weigend and N. A. Gershenfeld, "Time series prediction: Forecasting the future and understanding the past", Addison-Wesley, 1994.

[12] N. S. Arunraj, D. Ahrens, A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting, Int. J. Production Economics 170 (2015) 321-335P