

Walmart Sales Analysis and Prediction

Dr. Lutful Islam¹, Mohammad Farzan Farooqui², Ayyan Khan³, Mohammad Wasi⁴, Tousif Shaikh⁵

Professor, Department of Computer Engineering¹

Student, Department of Computer Engineering^{2,3,4,5}

M. H Saboo Siddik College of Engineering, Mumbai, Maharashtra, India

Abstract: *Walmart Sales Analysis and Prediction aims to perform an analysis of Walmart's sales data to gain insights into the performance of the company and to develop a predictive model to forecast future sales. The data includes historical sales figures, promotional activities, and store-specific information for a period of several years. The analysis involves exploratory data analysis, feature engineering, and model selection to identify the most influential factors affecting Walmart's sales. Several machine learning algorithms are used to build a predictive model, and their performances are compared to select the best one. The final model is used to forecast Walmart's sales for the next few quarters. The insights gained from this analysis could help Walmart make informed decisions about inventory management, pricing strategies, and promotional activities.*

Keywords: Machine Learning, XGBoost, Random Forest Regression, Market Trends, Customer Behavior

REFERENCES

- [1]. Bakshi, C. (2020). Random forest regression. <https://levelup.gitconnected.com/random-forest-regression-209c0f354c84>
- [2]. Bari, A., Chaouchi, M., & Jung, T. (n.d.). How to utilize linear regressions in predictive analytics. <https://www.dummies.com/programming/big-data/data-science/how-to-utilize-linear-regressions-in-predictive-analytics/>
- [3]. Crown, M. (2016). Weekly sales forecasts using non-seasonal arima models. <http://mxcrown.com/walmart-sales-forecasting/>
- [4]. Harsoor, A. S., & Patil, A. (2015). Forecast of sales of walmart store using big data applications. International Journal of Research in Engineering and Technology eIS, 04, 51–59. <https://doi.org/https://ijret.org/volumes/2015v04/i06/IJRET20150406008.pdf>
- [5]. Jaccard, J., & Turrisi, R. (2018). Interaction effect in multiple regression second edition. Sage Publications, Thousand Oaks CA.
- [6]. Kumar, A. (2020). Lasso regression explained with python example. <https://vitalflux.com/lasso-ridge-regression-explained-with-python-example/>
- [7]. Sullivan, J. (2019). Data cleaning with r and the tidyverse: Detecting missing values. <https://towardsdatascience.com/data-cleaning-with-r-and-the-tidyverse-detecting-missing-values-ea23c519bc>