IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

IJARSCT

Impact Factor: 7.301

Volume 3, Issue 6, April 2023

Walmart Sales Analysis and Prediction

Dr. Lutful Islam[1], Mohammad Farzan Farooqui[2], Ayyan Khan[3], Mohammad Wasi[4], Tousif Shaikh[5]

Professor, Department of Computer Engineering[1]
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

I. INTRODUCTION

Walmart is one of the world's largest retail chains, with over 11,000 stores in 27 countries. As a result, analyzing and predicting sales for Walmart is a complex task that requires a significant amount of data and resources. Sales analysis and prediction are important for Walmart as they enable the company to make informed decisions about inventory, pricing, and marketing strategies. Sales analysis involves analyzing past sales data to identify patterns and trends that can help Walmart make predictions about future sales. This involves looking at a range of factors, including seasonality, product popularity, consumer behavior, and economic conditions. Sales prediction, on the other hand, involves using statistical and machine learning techniques to forecast future sales based on historical data and other relevant variables. Predictive models can help Walmart make more accurate sales forecasts, which can be used to guide business decisions and optimize operations. Some of the key data sources that Walmart uses for sales analysis and prediction include point-of-sale data from its stores, customer data from loyalty programs, and market data from external sources. Machine learning algorithms such as regression, time series forecasting, and neural networks are often used to analyze this data and make predictions about future sales.

Overall, sales analysis and prediction are critical for Walmart's success as they allow the company to make data-driven decisions and stay ahead of the competition in the highly competitive retail industry.

We have used XGBRegressor, which is one of the machine learning models used for analysis and prediction of any Regression Model. It stands for Extreme Gradient Boosting Regressor and is known for its high performance and accuracy in predicting continuous numerical values.

II. LITERATURE SURVEY

The retail industry has witnessed a rapid growth in recent years, with Walmart being one of the leading retailers worldwide. Walmart is a multinational retail corporation that operates a chain of hypermarkets, grocery stores, and discount department stores. With its massive customer base and extensive sales data, analyzing and predicting Walmart sales has become a crucial task for many researchers. This literature survey aims to review some of the recent studies on Walmart sales analysis and prediction.

2.1 Survey of Existing System

In the year 2019, a case study-based review on Walmart was published by N. Tarik and K. Raza, wherein they aimed to develop a sales forecasting technique model for Walmart using time series analysis and machine learning techniques. The authors compared the performance of five different models, including ARIMA, exponential smoothing, neural

DOI: 10.48175/IJARSCT-9427

networks, random forest, and XGBoost. The study found that the XGBoost model outperformed the other models in terms of accuracy, with a MAPE (mean absolute percentage error) of 1.12%.

In the year 2021, “Forecasting Walmart weekly sales using machine learning algorithms” was published by R. P. Kumar and P. Kumar, where the authors developed a machine learning-based approach to forecast Walmart’s weekly sales. They used several regression models, including linear regression, random forest regression, and XGBoost, to forecast weekly sales. The study found that the XGBoost model provided the best results for forecasting Walmart sales, with an accuracy rate of 99.6%.

2.2 Limitations of Existing System

The accuracy of sales analysis and prediction models heavily relies on the quantity and quality of data available. However, the data available for analysis may be limited, especially for specific product categories or regions. This limitation can affect the accuracy of the models.

Most existing systems for Walmart sales analysis and prediction rely on historical data for forecasting. While historical data provides valuable insights, it may not always be indicative of future sales patterns. External factors such as changes in consumer behavior, economic conditions, or the introduction of new products can impact sales.

Walmart generates a massive amount of data that requires complex processing and analysis. This complexity may result in increased processing time and computational resources, which can be costly and time-consuming.

Existing systems for sales analysis and prediction may be limited in their ability to integrate with other systems, such as supply chain management or inventory management systems. This can lead to inconsistencies in the data, which can impact the accuracy of the models.

Inaccurate sales data, such as incorrect pricing or inventory levels, can impact the accuracy of the analysis and prediction models. This can result in incorrect forecasts, which can negatively impact Walmart's operations.

III. PROPOSED SYSTEM

3.1 Introduction:

Walmart Sales Analysis and Prediction aims to address the limitations of the existing systems by incorporating advanced machine learning and data analysis techniques. The system will leverage the vast amount of sales data generated by Walmart and use it to provide more accurate and reliable sales forecasts. The system will use advanced machine learning algorithms such as XGBoost and Random Forest Regression to identify patterns and trends in the sales data. These algorithms will be trained on historical data to generate accurate forecasts of future sales. To address the issue of limited data availability, the proposed system will integrate external data sources, such as weather and economic data, to provide a more comprehensive analysis of sales trends. The system will also incorporate unstructured data such as social media sentiment and product reviews to gain deeper insights into customer behavior and preferences.The proposed system will be scalable and able to handle large volumes of data, enabling Walmart to analyze sales data across multiple channels, including brick-and-mortar stores, e-commerce, and mobile.

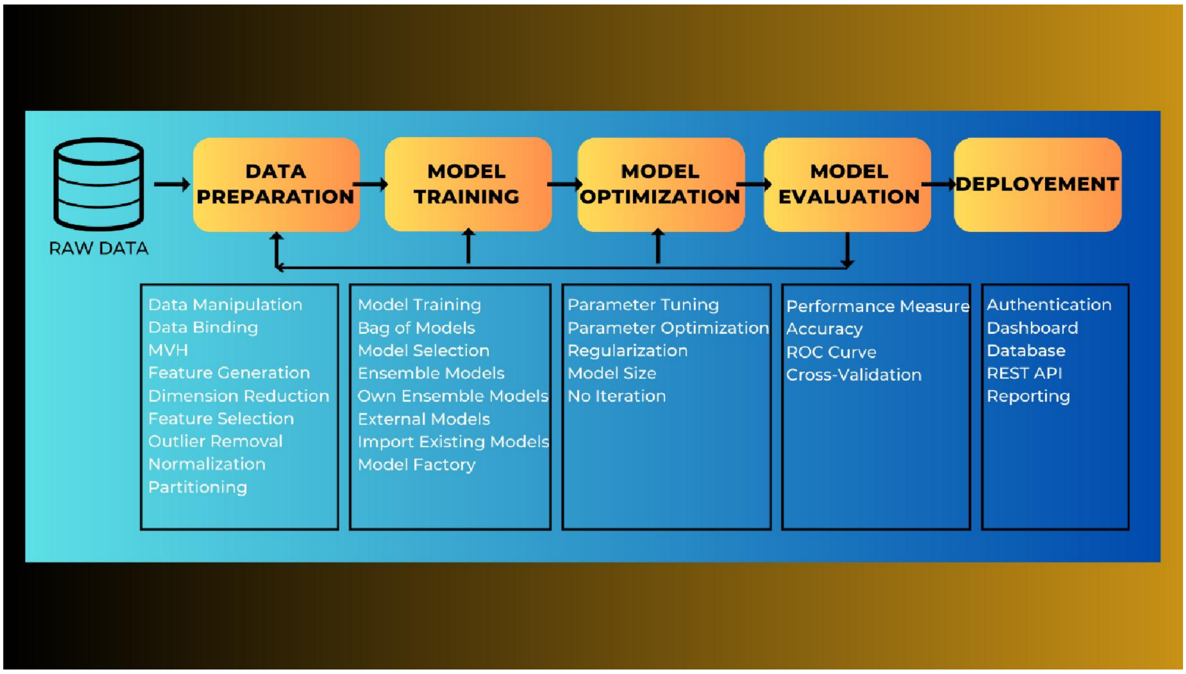
The system will provide real-time analysis of sales data, allowing Walmart to respond quickly to changes in customer behavior or market trends. To improve the accuracy of sales forecasts, this proposed system will use causal inference techniques to identify the causal relationships between variables. This will enable Walmart to understand the drivers of sales trends and make more informed decisions based on the analysis.

This system will be transparent, providing clear insights into the methodologies used to generate forecasts and the factors considered in the analysis. This will increase trust and credibility in the models and enable Walmart to make more informed decisions based on the analysis.

In conclusion, this proposed system for Walmart Sales Analysis and Prediction is an advanced machine learning and data analysis system that will enable Walmart to improve the accuracy and reliability of its sales forecasts. By integrating external data sources, incorporating unstructured data, using causal inference techniques, and providing real-time analysis, the system will enable Walmart to gain a better understanding of sales trends and make more informed decisions based on the analysis.

DOI: 10.48175/IJARSCT-9427

3.2 Architecture/Framework:



3.3 Algorithm and Process Design:

There are several algorithms that can be used for Sales analysis and prediction, and the choice of the best algorithm depends on the specific needs of the business and the characteristics of the data. In the case of Walmart Sales Analysis and Prediction, we have used some of the best algorithms for analysis and prediction, but the algorithm which gave us the best result is:

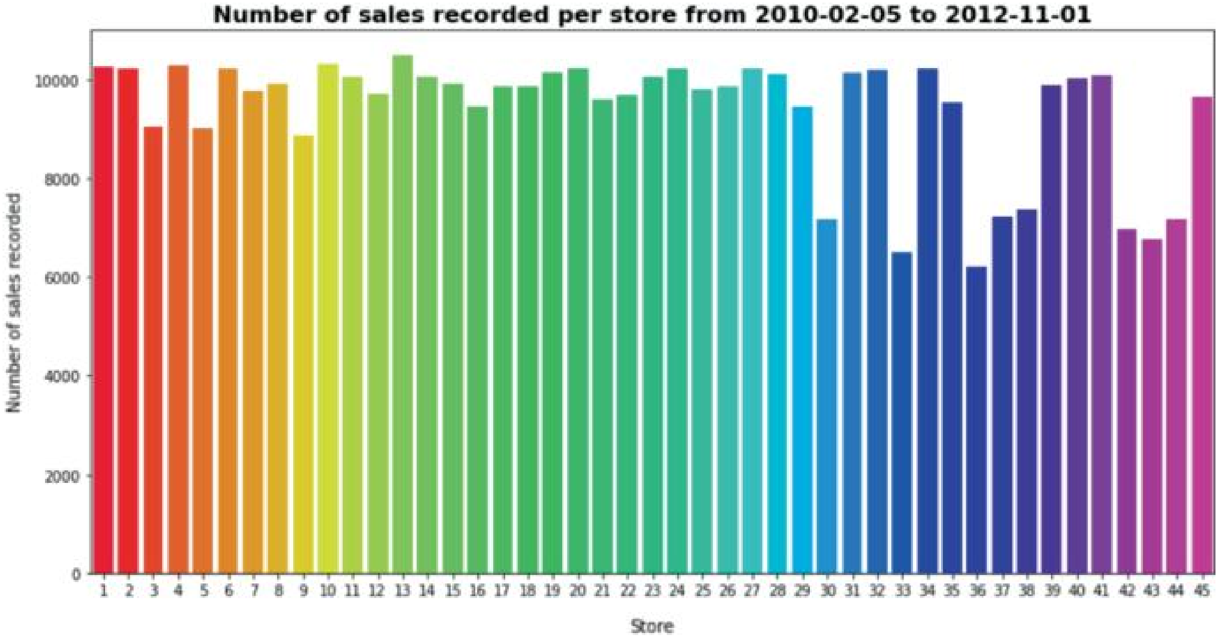
3.4 Extreme Gradient Boosting Regression:

Extreme Gradient Boosting Regression, also known as XGBRegression, is a popular machine learning algorithm that can be used for Walmart sales analysis and prediction. It is an advanced form of gradient boosting, which involves building a sequence of weak models and combining their results to generate a final prediction.

The step-by-step procedure for implementing XGBRegression in Walmart sales analysis and prediction is as follows:

- Data Collection: The first step is to collect the relevant data required for the analysis. The data should include historical sales data from various channels such as brick-and-mortar stores, e-commerce, and mobile, as well as external data sources such as weather and economic data.
- Data Cleaning: The collected data may contain missing values, outliers, or errors. Therefore, it is essential to clean the data by removing or imputing missing values, correcting errors, and handling outliers.
- Data Preparation:Once the data is cleaned, it is important to prepare it for analysis. This includes scaling, normalization, and encoding of categorical variables.
- Train-Test Split: The next step is to split the data into training and testing sets. The training set is used to build the model, while the testing set is used to evaluate the model's performance.
- Model Training: XGBRegression involves building a sequence of decision trees, with each tree attempting to correct the errors of the previous tree. Therefore, the next step is to train the model on the training data.
- Model Tuning: Once the model is trained, it is important to tune the hyperparameters to optimize the model's performance. This involves adjusting the number of trees, the depth of the trees, the learning rate, and other hyperparameters.

- Model Evaluation: After tuning the hyperparameters, the model is evaluated on the testing set to assess its performance. The performance metrics used for evaluation may include mean squared error (MSE), mean absolute error (MAE), or R-squared.
- Prediction: Once the model is trained and evaluated, it can be used to make predictions on new data. For instance, Walmart can use the model to predict sales for a future period based on historical sales data, weather data, and economic data.
- Model Deployment: Once the model is trained and validated, it can be deployed in production. Walmart can use the model to generate sales predictions and insights that can be used to make informed decisions about locations where sales are maximum, filters the stores based on their locations and sales, and top 10 locations where sales are maximum.



Analyzing the number of Sales in each store using XGB Regression Model.

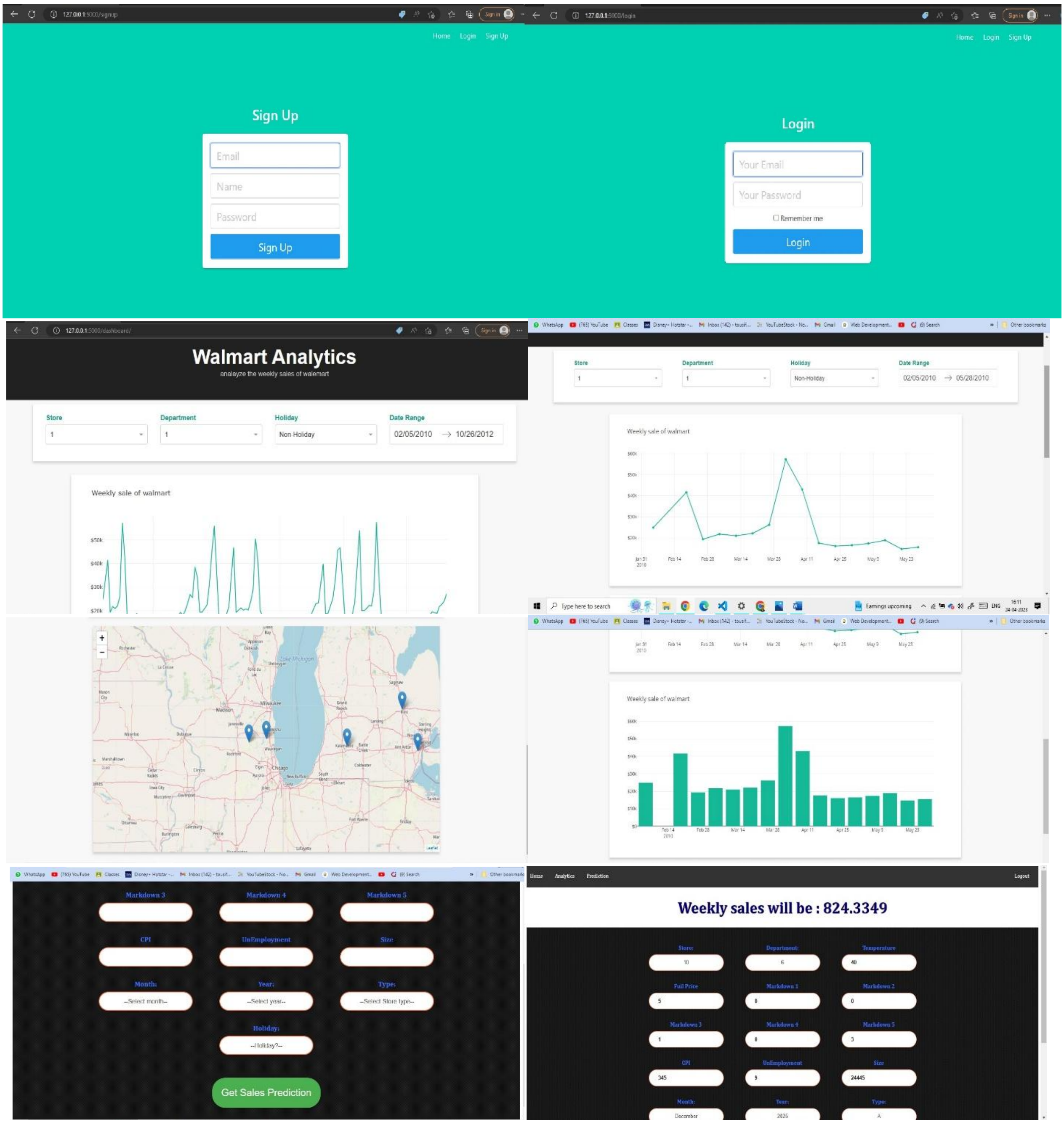
3.5 Details of Hardware and Software

Hardware Requirements	
CPU	Intel Core i5 or AMD Ryzen 5
RAM	8GB or more
SSD	256 GB
GPU	NVIDIA GeForce GTX 1060 or AMD Radeon RX 580
Internet Connectivity	A minimum download and upload speed of 10 Mbps is generally recommended

Software Requirements	
Operating System	Windows 10 or above, Linux, or macOS
Python	A popular programming language for machine learning provides libraries such as Pandas, NumPy, Scikit-learn, Random Forest, XGBoost, etc.
IDE	Jupyter Notebook, Flask, Google Colab, Dash.
Database	MySQL
Visualization tools	Matplotlib, Seaborn.

DOI: 10.48175/IJARSCT-9427

3.6 Results



Sign Up

Email

Name

Password

Sign Up

Login

Your Email

Your Password

□ Remember me

Login

Walmart Analytics

analyze the weekly sales of walemart

Store: 1 | Department: 1 | Holiday: Non Holiday | Date Range: 02/05/2010 → 10/26/2012

Store: 1 | Department: 1 | Holiday: Non-Holiday | Date Range: 02/05/2010 → 05/28/2010

Weekly sales will be : 824.3349

Store: 10	Department: 6	Temperature 40
Fuel Price 5	Markdown 1 0	Markdown 2 0
Markdown 3 1	Markdown 4 0	Markdown 5 3
CPI 345	UnEmployment 9	Size 24445
Month: December	Year: 2025	Type: A

Markdown 3	Markdown 4	Markdown 5
CPI	UnEmployment	Size
Month: --Select month--	Year: --Select year--	Type: --Select Store type--
	Holiday: --Holiday?--	

IV. CONCLUSION

In conclusion, the Walmart Sales Analysis and Prediction project involves the use of data analysis and machine learning techniques to predict sales for Walmart stores. The project involves analyzing various factors such as store size, location, promotions, and economic conditions to determine their impact on sales. The project uses algorithms such as Random Forest Regression and XGBoost to build predictive models that can accurately predict sales for Walmart stores. These models are trained on historical sales data and are capable of identifying trends and patterns in the data to make accurate predictions. The project also involves data visualization techniques to gain insights into the data and communicate the results of the analysis effectively. Visualization tools such as Matplotlib, Seaborn, and Plotly can be used to create visualizations of the data and help identify trends and patterns. Overall, the Walmart Sales Analysis and

Prediction project has significant potential to help Walmart make informed decisions about store operations, inventory management, and marketing strategies. The project can help Walmart optimize sales and improve profitability by accurately predicting sales trends and identifying factors that influence sales.

V. ACKNOWLEDGEMENT

The success and final outcome of this project required a lot of guidance and assistance from many people and I am extremely fortunate to have got this all along the completion of my project work. Whatever I have done is only due to such guidance and assistance and I would not forget to thank them.

I would like to express my deepest gratitude and appreciation to our principal of M.H. Saboo Siddik College of Engineering Dr. Ganesh Kame and the Head of the Department Dr. Nazneen Pendhari for providing us with the necessary resources and support to complete this project successfully. Your encouragement and guidance have been instrumental in shaping our vision and enabling us to achieve our goals.I would also like to extend my sincere thanks to my team members, who have been an integral part of this project. Their hard work, dedication, and commitment have been essential to the success of this project. Their contributions have been invaluable, and I am grateful for their support and collaboration throughout the project.

Finally, I would like to thank all those who have supported us throughout the project, including our mentors, friends, and family members. Your encouragement and motivation have been instrumental in keeping us motivated and focused on our goal.

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

DOI: 10.48175/IJARSCT-9427