WALLMART

Walmart is a multinational retail corporation that operates a chain of discount department stores, grocery stores, and hypermarkets. It was founded in 1962 by Sam Walton in Bentonville, Arkansas, and has since grown to become one of the largest retailers in the world.

The foundation of Walmart's business strategy is its broad selection of low-cost goods. The corporation has more than 2.2 million employees globally and has over 11,000 outlets across 27 countries. Walmart has an online e-commerce platform in addition to its physical shops, enabling customers to buy goods from the convenience of their homes.

INDEPENDENT VARIABLE

Store: The Walmart shop where the sales were registered is identified numerically in this column.

Date: The week's date that the sales were recorded is listed in this column.

Holiday Flag: This column shows if a significant holiday, such Christmas or Thanksgiving, fell within the given week. This can help us comprehend how the timing of Christmas shopping affects sales.

Temperature: For the vicinity of the specified store, the average temperature for the week is shown in this column. Understanding how weather patterns affect sales can be helped by this knowledge.

Fuel Price: For the region of the specified store, this column gives the weekly average gasoline price. Understanding how fuel prices affect sales can be helped by this information.

Consumer Price Index (CPI): This column displays the CPI for the area where the specified store is located. The CPI is a measure of inflation that can be helpful in determining how changes in the cost of living affect sales.

Unemployment: The region of the specified store's unemployment rate is shown in this column. This knowledge can help us comprehend how economic variables like rising employment and consumer confidence affect sales.

DEPENDENT VARIABLE

Weekly Sales: The total sales income for the specified store for the specified week is shown in this column.

PROPOSAL

We propose to analyze Walmart's weekly sales dataset, which contains data on sales revenue, store location, date, and various environmental factors such as temperature, fuel price, CPI, and unemployment. Our objective is to develop a predictive model to forecast weekly sales based on these factors. The analysis will help Walmart to optimize its inventory management, pricing, and marketing strategies and improve overall business performance.

Methodology: Our proposed methodology involves the following steps:

Data cleaning and preparation: We will remove any missing or inconsistent data and convert categorical variables to numerical formats.

Exploratory data analysis: We will conduct an exploratory analysis of the data to identify any patterns or correlations between the various variables.

Feature engineering: We will engineer additional features from the existing data.

Model selection and evaluation: We will train and evaluate several machine learning models, such as linear regression, decision trees, random forests, Support Vector Regression and neural networks. We will use appropriate evaluation metrics such as mean absolute error and mean squared error to compare the performance of these models.

Accuracy of linear regression model is: 0.15

Accuracy of Decision Tree Regression model is: 0.88

Accuracy of Support Vector Regression-0.028

Accuracy of Random forest model is 0.93

Accuracy of Neural network model is: -2.77

This variation may be due to Model Complexity, overfitting. For our research, we tested a number of  models and discovered that the Random Forest model had the best accuracy score (0.93), followed by the Decision Tree Regression model (0.88). The accuracy score for the linear regression model was 0.15, which was much less accurate than the scores for the other models. Poor performance was shown by the Support Vector Regression model's accuracy score of -0.028 and the Neural Network model's accuracy score of -2.77.

We will utilise the Random Forest model for our project since it can handle complicated interactions between characteristics in the dataset and got the highest accuracy score. We will be able to forecast our project outcomes more precisely thanks to this model, which will help us make better decisions and produce better results.

Hyperparameter tuning: We will optimize the hyperparameters of the best-performing models to improve their accuracy.

Prediction and interpretation: We will use the final model to make predictions on new data and interpret the results to provide insights for Walmart's business decisions.

Timeline: Our proposed timeline for completing the project is as follows:

Week 1: Data cleaning and preparation

Week 1: Exploratory data analysis and feature engineering

Week 2: Model selection and evaluation

Week 2: Hyperparameter tuning and final model selection

Week 3: Prediction and interpretation

Week 3: Report writing and presentation.

Deliverables: The deliverables for this project will include:

A detailed report outlining the methodology, results, and insights gained from the analysis.

A presentation summarizing the key findings and recommendations for Walmart's business decisions.

Python code implementing the data cleaning, exploratory analysis, feature engineering, model selection, hyperparameter tuning, and prediction steps.