JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY NOIDA

NOIDA B.TECH V SEMESTER

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING AND INFORMATION TECHNOLOGY

WALMART WEEKLY SALES FORECASTING

SUBMITTED TO :- Dr. Sonal Saurabh

GROUP MEMBERS :

Name	Enroll
Kartik Aggarwal	22103224
Nandini	22103225
Kshitij Gupta	22103236
Ashish Singh	22103231
Om Sharma	22103203

INTRODUCTION

In today’s data-driven world, retail giants like Walmart generate vast amounts of data daily, spanning across sales, inventory, and customer behavior. Analyzing and forecasting this data is crucial for improving operational efficiency, customer satisfaction, and overall business performance. This project leverages advanced data analysis, forecasting techniques, and web scraping to extract meaningful insights from Walmart's historical and real-time sales data.

AIM

The aim of this project is to help a retail store optimize its inventory management and improve revenue by accurately forecasting future sales. By analyzing factors like seasonal trends, promotions, holidays, and economic conditions, the project aims to provide more reliable sales predictions. This will reduce issues like overstocking or stockouts, lower costs, and improve customer satisfaction, ultimately supporting better decision-making and more efficient operations.

Introduction to the Problem Statement

In the retail industry, especially for a large company like Walmart, understanding sales is very important for making good decisions and planning effectively. By studying sales data, Walmart can learn what increases revenue, find seasonal shopping patterns, and discover ways to improve its operations. This helps the company meet customer needs better and work more efficiently.

Forecasting sales is also key. It helps predict future trends so Walmart can plan its inventory, staffing, and finances more accurately. With so many stores and different types of customers, accurate forecasting is crucial for both daily operations and long-term planning.

Looking at weekly sales data from Walmart stores can show important details, like shopping patterns during holidays, sales differences between regions, and the impact of events like promotions or global changes. These insights help Walmart make smarter decisions, use resources wisely, and stay prepared for the future.

Objective

The objective of this project is to:

- **Analyze Sales Trends:** Identify long-term sales patterns and seasonal effects across Walmart's stores to help adapt to market changes and optimize operations.
- **Detect Seasonal and Regional Variations:** Study how seasonal cycles (like holiday spikes) and regional differences impact sales, providing insights for better inventory, staffing, and promotional planning.
- **Develop Predictive Models:** Apply Random Forest to forecast future sales, enabling Walmart to plan resources more accurately.
- **Provide Actionable Insights:** Offer recommendations to improve sales performance, optimize resource allocation, and identify growth opportunities, aligning with Walmart's strategic goals and long-term vision.

Columns in Dataset

The dataset consists of the following columns:

- **Store:**Unique identifier for each Walmart store, enabling analysis at individual and aggregate levels.
- **Date:** Represents the week of sales, suitable for time-series analysis of trends and seasonality.
- **Weekly_Sales:** Total sales generated by the store in a given week, used for trend analysis and forecasting.
- **Holiday_Flag:** Indicates whether the week was a holiday week (1) or not (0), affecting sales performance.
- **Temperature:** Recorded temperature during the sale week, potentially influencing customer behavior.
- **Fuel_Price:** Fuel cost in the region, affecting customer spending and shopping frequency.
- **CPI (Consumer Price Index):** Reflects the price levels in the region, indicating inflation or deflation that may impact consumer spending.
- **Unemployment:** Unemployment rate in the region, which may influence consumer spending during economic downturns.

Holiday Events

Key holidays like the Super Bowl, Labor Day, Thanksgiving, and Christmas drive sales peaks, captured in the dataset. This data aids in analyzing holiday trends and refining predictive models for better sales forecasting and resource planning.

This dataset structure enables a multi-faceted analysis of Walmart's weekly sales, making it possible to identify trends, seasonal patterns, and external factors influencing sales performance. Through exploratory data analysis (EDA), we can uncover potential relationships between sales and economic or environmental conditions, while predictive modeling will help forecast future sales, providing Walmart with actionable insights for improved operational planning and resource allocation.

Tools and technology used

Programming Language and Environment

1. Python:
    - Python is the main programming language used in this project due to its versatility and rich ecosystem for data analysis and machine learning.
2. Jupyter Notebook:
    - Interactive computing environment that allows combining code, output, and documentation in one place.
    - Suitable for exploratory data analysis (EDA) and visualization.

Libraries and Their Uses

1. NumPy
    - Fundamental library for numerical computations.
    - Likely used for mathematical operations, array manipulations, and handling numeric data.
2. Pandas
    - Core library for data manipulation and analysis.
    - Features used include:
        - pd.read_csv(): To load the dataset (likely Walmart.csv).
        - .head(), .info(): For inspecting data.
        - Handling missing values (isnull().sum()).

        - Removing duplicate entries (.duplicated()).
        - Renaming columns for consistency ([col.lower() for col in sales.columns]).
        - Converting string dates to datetime objects (pd.to_datetime).
3. Matplotlib
    - For creating visualizations to understand patterns and trends.
    - %matplotlib inline: Ensures that visualizations are displayed inline within the notebook.
4. Scikit-learn
    - Implements machine learning models like Random Forest, enabling accurate sales predictions and analysis.
5. Beautiful Soup
    - Used for web scraping to dynamically fetch store-related or sales data, enriching the analysis with external sources.
6. SciPy
    - Applied for optimization tasks and statistical computations, enhancing the forecasting process.

Techniques and Features Implemented

1. Data Cleaning and Preprocessing

- Handled missing data by removing or managing incomplete records.
- Detected and managed duplicate rows to ensure data consistency.
- Converted date columns to a datetime format for seamless time-based analysis and feature engineering.
- Renamed columns to lowercase for uniformity and ease of access.

2. Outlier Detection

- Implemented a custom function using the Interquartile Range (IQR) method:
    - Identified numeric columns.
    - Calculated the first quartile (Q1) and third quartile (Q3).
    - Computed the IQR as Q3 - Q1 and flagged outliers beyond 1.5 times the IQR.
    - Summarized outliers in a dictionary for better understanding of data anomalies.

3. Exploratory Data Analysis (EDA)

- Conducted an initial inspection of data structure and feature exploration using methods like .info() and .columns.
- Visualized key features to identify trends, seasonal patterns, and potential correlations.

4. Dataset

- Utilized a CSV file named **Walmart.csv**, containing Walmart's sales data.
- The dataset includes important fields like Date, Weekly_Sales, Holiday_Flag, Temperature, and other economic indicators.
- Performed time-based data manipulations for trend and seasonality analysis.

Potential Extensions

Given the tools and initial implementation, the project could involve:

1. Visualization:
    - Time series plots (e.g., sales trends over time).
    - Histograms, boxplots for data distribution and outliers.
    - Correlation heatmaps to analyze relationships between variables.
2. Feature Engineering:
    - Creating new features like sales growth, day of the week, or holiday effects.
3. Predictive Modeling:
    - Applying machine learning techniques such as linear regression or time series forecasting models (e.g., ARIMA, Prophet).

Exploratory Data Analysis

Now that we have finished the data wrangling phase, we will explore the dataset to extract useful information. Exploratory Data Analysis (EDA) helps in understanding the underlying patterns, trends, and relationships within the data, which can guide our analysis and model selection.

Key Steps in EDA

Descriptive Statistics: Summarize the dataset using measures such as mean, median, mode, standard deviation, and percentiles.

Visualizations: Create visual representations such as histograms, box plots, and scatter plots to better understand the data.

Correlation Analysis: Assess the relationships between numerical variables using correlation coefficients and heatmaps.

Time Series Analysis: Analyze trends over time, particularly focusing on weekly sales.

Conclusion:

This project provides a comprehensive approach to analyzing and forecasting Walmart's sales data. By implementing robust data cleaning, exploratory analysis, and predictive modeling techniques, it identifies key sales trends, seasonal patterns, and regional variations that impact performance. The integration of machine learning models like Random Forest enhances the accuracy of sales forecasts, enabling Walmart to optimize inventory, staffing, and marketing strategies.

Through actionable insights derived from data, this project not only supports efficient operational planning but also helps uncover growth opportunities. It demonstrates the value of data-driven decision-making in the retail sector, showcasing how advanced analytics can transform raw data into meaningful strategies for long-term success.

REFERENCES

•https://www.researchgate.net/publication/345897383_Sales_Prediction_of_Market_using_Machine_Learning

•ttps://www.researchgate.net/publication/334706334_Salesforecasting_of_Retail_Stores_using_Machine_Learning_Techniques

•https://arkieva.com/demand-planning/sales-predictor/

•https://www.getweflow.com/blog/predictive-sales-forecasting