



**Lambton
College**

Walmart 
**SALES FORECASTING
USING MACHINE LEARNING**

PROJECT PROPOSAL
BDM 2053
Big Data Algorithms and Statistic

SUBMITTED TO:
Meysam Effati, PhD
Lambton College

FEBRUARY, 2024



SUBMITTED BY
Team Everest

BACKGROUND AND MOTIVATION

Walmart Inc. is an American multinational retail corporation. It is one of the largest retail corporations in the world, which was founded by Sam Walton in the year 1962. And, it has revenue of over 485.87 billion dollars recorded in 2016. Walmart runs three types of stores based on several items namely hypermarkets, discount department stores, and grocery stores.

Since it's a major competitor in the retail industry, it is interesting for us to investigate the elements that influence sales. Recently, we discovered Walmart sales data on *Kaggle*. This data was collected for a Walmart recruiting competition. The aim was to forecast weekly sales at retailers in 45 distinct regions. We would estimate store sales by department using this data, which included weekly sales, store size, department code, consumer price index, region-specific unemployment, and promotional markdowns. We have 99 distinct sections here, including dry groceries, sporting goods, frozen goods, and so forth.

The proposed initiative seeks to improve Walmart's sales forecasting skills by leveraging retail analytics. The model will optimize inventory and pricing, schedule promotional events for peak sales days, eliminate overstocking and understocking, properly allocate resources, manage cash flows, and successfully design discounts and incentives.

PROJECT OBJECTIVES

Objective 1: Sales Prediction Analysis

- Analyze historical sales data to understand patterns and trends.
- Implement predictive models, including Random Forest Regressor, Multiple Linear Regression, Lasso Regression, Ridge Regression, and Elastic-Net, for accurate sales forecasting.
- Evaluate model performance using metrics such as Mean Square Error (MSE), Mean Absolute Error (MAE), R2 Score, and Root Mean Squared Error (RMSE).
- Deploy a prediction model web app using Streamlit and host it on Heroku.

Objective 2: Customer Behavior Analysis

- Analyze changes in customer purchasing patterns, preferences, and engagement.
- Evaluate the influence of external factors on customer decisions, such as economic indicators and seasonal trends.

Objective 3: Operational Efficiency Enhancement

- Optimise inventory management and product placement based on predictive models.
- Implement data-driven strategies to improve marketing and promotional campaigns.
- Enhance operational efficiency by aligning business decisions with data-driven insights.
- Redesigning the physical store layout, implementing changes to the company's supply chain structure, and modifying pricing strategies unrelated to sales prediction

DESIGN OVERVIEW

This design overview describes a methodical strategy for addressing the Walmart sales prediction issue, beginning with data preparation and exploration, and progressing to model creation and results interpretation. The final recommendations attempt to help Walmart optimize their sales methods for better success. The steps include:

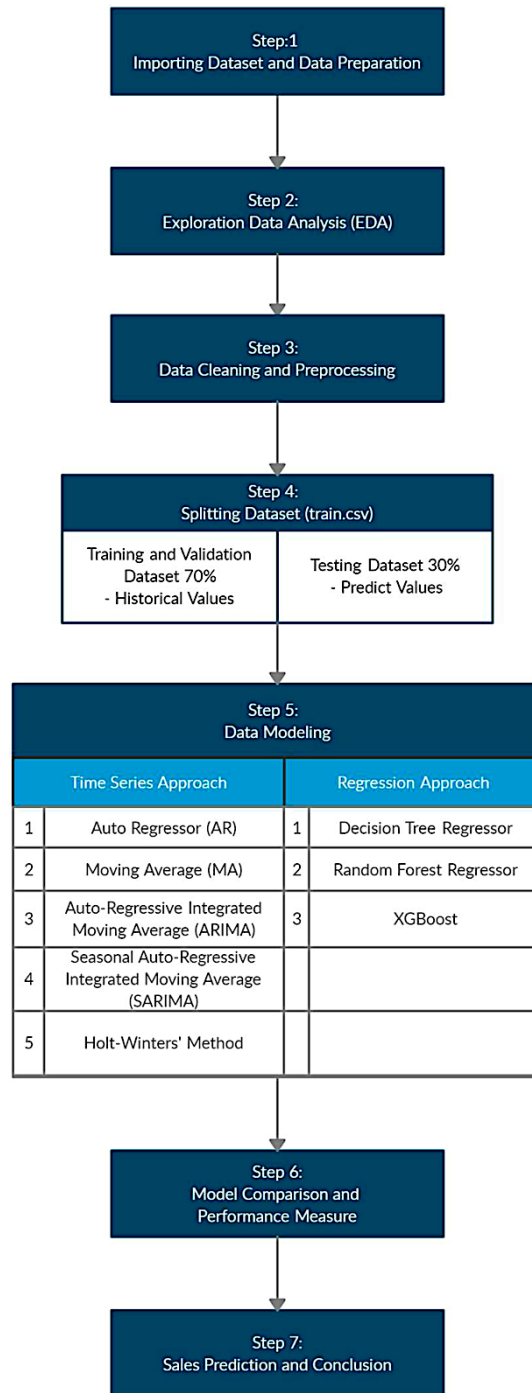


Fig 1: Schematic Illustration of Methodologies

- 1. Importing and Reading Data:** The first step is to import and read the given dataset where the CSV files would be used to load the dataset into the Jupyter Notebook. The review will focus on comprehending the number of variables and cases. The appropriate CSV files would be combined using the inner join.
- 2. Exploratory Data Analysis (EDA):** After merging the CSV file, EDA would be done. Descriptive statistics will be used to analyze the count, mean, minimum, maximum,

interquartile range, and number of NAs. The article will use visual aids to analyze the relationship between sales (the dependent variable) and various independent factors.

3. **Data cleansing and preprocessing:** Based on earlier analyses of dependent and independent factors. To clean the data, replace NAs, Boolean and category types with integers, and use Date as the index. A correlation heat map is used to visualize the relationships between all characteristics. Finally, poor correlation characteristics would be deleted from the dataset.
4. **Feature Engineering:** The step is used to improve the dataset's prediction capabilities. This includes generating new variables or altering existing ones to better capture the relationship between characteristics.
5. **Splitting the dataset:** The CSV file divided historical data into training and testing sets, with a 70:30 split by date.
6. **Data Modelling:** The training set fits all proposed models for time series, regression, and neural network techniques. Traditional time-series approaches included *AR*, *MA*, *ARIMA*, *SARIMA*, and *HW*. In the Regression technique, standard approaches such as *Decision Tree Regressor*, *Random Forest Regressor*, and *XGBoost* would be employed.
7. **Model comparison and performance measurement:** After developing each approach, performance indicators including MAE, MSE, and RMSE were utilized to compare models.
8. **Model Results and Recommendations:** The project concludes with a thorough assessment of model results and the development of suggestions based on the insights gathered. This provides practical tips for Walmart to improve their sales methods.

The first linear model shows that promotional offerings during holidays boost sales, with Walmart Superstores experiencing higher sales than other shops. The second model highlights the significant influence of events like Christmas and Thanksgiving on sales and the impact of the pandemic on sales. The third linear model examines the influence of parameters like unemployment, CPI, and temperature on weekly sales, providing a comprehensive understanding of their impact.

WORKLOAD DISTRIBUTION

TASKS	RESPONSIBILITIES
Brainstorming	<i>All team members</i>
Objective Formulation	<i>All team members</i>
Implementation of Objectives 1,2	<i>Aljesh, Bharat</i>
Implementation of Objectives 3,4	<i>Bijay, Satish</i>
Implementation of Objectives 5,6,7	<i>Satish, Sishir</i>
Create Project Webapp, Screencast	<i>Bijay, Sishir</i>
Every Documentation & Literature Review	<i>Bharat, Aljesh</i>
Final Project Presentation	<i>All team members</i>

PROJECT TIMELINE

The Gantt chart below depicts the expected timeline for major tasks and activities in the Walmart Sales Prediction Project. This graphic depiction shows task timelines, dependencies, durations, milestones, resource allocations, and key routes.

2024	Feb 15	Feb 20	Feb 25	March 5	March 10	March 15
<i>Requirements Gathering</i>						
<i>Data Exploration + Processing</i>						
<i>Data and Target Split</i>						
<i>Feature Scaling, Modeling, Coding</i>						
<i>Data Visualization and Interpretation</i>						
<i>Documentation</i>						

Fig 2: Gantt Chart

REFERENCES

- Franco-Santos, M., & Bourne, M. (2009). The impact of performance targets on behavior: A close look at sales force contexts. *Research Executive Summaries Series*, 5.
- Silverman, D. (2006). *Interpreting Qualitative Data: Methods for Analyzing Talk, Text and Interaction* (3rd ed.). Sage Publications Ltd.
- Walmart recruiting - store sales forecasting. (2014). <https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting/data>
- Walmart’s sales data analysis - a big data analytics perspective. (2017). <https://doi.org/10.1109/APWConCSE.2017.00028>
- Wikipedia, The Free Encyclopedia. (n.d.). Walmart. <https://en.wikipedia.org/w/index.php?title=Walmart&oldid=1001006854>