Walmart Sales Analysis for Retail Industry

With Machine Learning

1. Introduction

Sales f Sale forecasting is a crucial process for businesses, especially in the retail industry, as it enables companies to make informed decisions about inventory management, marketing strategies, and overall business planning. For Walmart, one of the largest retail corporations, accurately predicting future sales is essential for optimizing supply chain and customer satisfaction. This project focuses on leveraging historical sales data from 45 Walmart stores to forecast sales and understand the impact of major holidays such as Christmas, Thanksgiving, Super Bowl, and Labor Day on sales trends.

By employing machine learning algorithms such as Random Forest, Decision Tree, XGBoost, and ARIMA, this project aims to identify patterns in sales data and provide Walmart with valuable insights into sales performance around major holidays. The models will be integrated into a Flask application and deployed on IBM Cloud, enabling real-time sales forecasts and supporting Walmart in making data-driven decisions to enhance its operations and maximize profitability.

1.1 Project Overview

The sales forecasting project using Walmart data focuses on predicting future sales and understanding the impact of major holidays (Christmas, Thanksgiving, Super Bowl, and Labor Day) on store sales. It involves data preparation, feature engineering, and model selection using algorithms such as Random Forest, Decision Tree, XGBoost, and ARIMA. After training and validating the models, the best-performing model will be integrated into a Flask application and deployed on IBM Cloud for real-time sales forecasting. The project aims to provide Walmart with insights into sales trends and recommendations to optimize sales strategies around holidays, ultimately supporting better decision-making and maximizing profitability.

1.2 Objectives

The objective of the project is to develop a comprehensive sales analysis tool for Walmart, a leading retail corporation, leveraging machine learning techniques. By analyzing past sales data, store information, and promotional markdown events, the project aims to forecast future sales accurately. Specifically, the project aims to determine the impact of holidays, including Christmas, Thanksgiving, Super Bowl, and Labor Day, on store sales. The ultimate goal is to equip Walmart with a robust forecasting model that aids in informed decision-making, enhances performance predictions, and optimizes resource allocation.

**2. Project Initialization and Planning Phase**

The "Project Initialization and Planning Phase" marks the project's outset, defining goals, scope,

and stakeholders. This crucial phase establishes project parameters, identifies key team members,

allocates resources, and outlines a realistic timeline.

It also involves risk assessment and mitigation planning. Successful initiation sets the foundation

for a well-organized and efficiently executed machine learning project, ensuring clarity,

alignment, and proactive measures for potential challenges.

### 2.1 Define Problem Statement

### The problem at hand is the development of an accurate sales forecasting model for Walmart,

### a well-known retail company that owns a hypermarket chain. Using historical sales data, store

### data and promotional events, the goal is to forecast future sales and estimate the impact of

### holidays, including Christmas, Thanksgiving, the Super Bowl and Labor Day, on store sales.

### The dataset provided contains data from 45 trades every week and the task involves using

### machine learning algorithms such as Random Forest, Decision Tree, XGBoost and ARIMA

### to analyze and model the data. In addition, the project includes integrating the developed

### model into a Flask application and deploying it on IBM Cloud for accessibility.

**Ref. template:** [**Click Here**](https://github.com/SmartInterns-Content/AI-ML-DA-Templates/blob/main/Machine%20Learning%20and%20Natural%20Language%20Processing%20Templates/1.%20Project%20Initialization%20and%20Planning%20Phase/Define%20Problem%20Statements%20Template%20(2).pdf)

**Walmart Sales Analysis Problem Statement Report:** [**Click Here**](https://github.com/SmartInterns-Content/AI-ML-DA-Templates/blob/main/Sample%20Project/1.%20Project%20Initialization%20and%20Planning%20Phase/Problem%20Statements%20Template.pdf)

### 2.2 Project Proposal (Proposed Solution)

The objective of the project is to develop a comprehensive sales analysis tool for Walmart, a leading retail corporation, leveraging machine learning techniques. By analyzing past sales data, store information, and promotional markdown events, the project aims to forecast future sales accurately. Specifically, the project aims to determine the impact of holidays, including Christmas, Thanksgiving, Super Bowl, and Labor Day, on store sales. The ultimate goal is to equip Walmart with a robust forecasting model that aids in informed decision-making, enhances performance predictions, and optimizes resource allocation.

**Ref. template:** [**Click Here**](https://github.com/SmartInterns-Content/AI-ML-DA-Templates/blob/main/Machine%20Learning%20and%20Natural%20Language%20Processing%20Templates/1.%20Project%20Initialization%20and%20Planning%20Phase/Project%20Proposal%20(Proposed%20Solution)%20template.docx.pdf)

**Walmart Sales Analysis Project Proposal Report:** [**Click Here**](https://github.com/SmartInterns-Content/AI-ML-DA-Templates/blob/main/Sample%20Project/1.%20Project%20Initialization%20and%20Planning%20Phase/SL%20Project%20Proposal%20(Proposed%20Solution)%20template%20(1).pdf)

## 2.3. Initial Project Planning

Initial Project Planning involves outlining key objectives, defining scope, and identifying stakeholders for a loan approval system. It encompasses setting timelines, allocating resources, and determining the overall project strategy. During this phase, the team establishes a clear understanding of the dataset, formulates goals for analysis, and plans the workflow for data processing. Effective initial planning lays the foundation for a systematic and well-executed project, ensuring successful outcomes.

**Ref. template:** [**Click Here**](https://github.com/SmartInterns-Content/AI-ML-DA-Templates/blob/main/Machine%20Learning%20and%20Natural%20Language%20Processing%20Templates/1.%20Project%20Initialization%20and%20Planning%20Phase/Project%20Planning%20Template.docx.pdf)

**Walmart Sales Analysis Project Planning Report:** [**Click Here**](https://github.com/SmartInterns-Content/AI-ML-DA-Templates/blob/main/Sample%20Project/1.%20Project%20Initialization%20and%20Planning%20Phase/SL%20Project%20Planning%20Template.docx.pdf)

# 3. Data Collection and Preprocessing Phase

The Data Collection and Preprocessing Phase involves executing a plan to gather relevant

Walmart sales Analysis data from Kaggle, ensuring data quality through verification and

addressing missing values. Preprocessing tasks include cleaning, encoding, and organizing

the dataset for subsequent exploratory analysis and machine learning model development.

## 3.1 Data Collection Plan, Raw Data Sources Identified, Data Quality Report

The dataset for "Walmart Sales Analysis for retail industry with machine learning" is sourced from Kaggle. The provided sample data represents a subset of the collected information, encompassing variables such as store information, department, size, weekly sales, temperature, holidays, markdowns. Data quality is ensured through thorough verification, addressing missing values, and maintaining adherence to ethical guidelines, establishing a reliable foundation for predictive modeling.

**Ref. template:** [**Click Here**](https://github.com/SmartInterns-Content/AI-ML-DA-Templates/blob/main/Machine%20Learning%20and%20Natural%20Language%20Processing%20Templates/2.%20Data%20Collection%20and%20Preprocessing%20Phase_/Raw%20Data%20Sources.pdf)

**Walmart Sales Analysis Data Collection Report:** [**Click Here**](https://github.com/SmartInterns-Content/AI-ML-DA-Templates/blob/main/Sample%20Project/2.%20Data%20Collection%20and%20Preprocessing%20Phase_/SL%20Raw%20Data%20Sources%20And%20Data%20Quality%20Report%20template.pdf)

## 3.2 Data Quality Report

The Data Quality Report will summarize data quality issues from the selected source,

including severity levels and resolution plans. It will aid in systematically identifying and

rectifying data discrepancies. Missing values, handling missing values, dealing with

outliers. Other issues like warnings, etc.

**Ref. template:** [**Click Here**](https://github.com/SmartInterns-Content/AI-ML-DA-Templates/blob/main/Machine%20Learning%20and%20Natural%20Language%20Processing%20Templates/2.%20Data%20Collection%20and%20Preprocessing%20Phase_/Data%20Quality%20Report%20template.docx.pdf)

**Walmart Sales Analysis Data Quality Report:** [**Click Here**](https://github.com/SmartInterns-Content/AI-ML-DA-Templates/blob/main/Sample%20Project/2.%20Data%20Collection%20and%20Preprocessing%20Phase_/SL%20Data%20Quality%20Report.pdf)

## 3.3 Data Exploration and Preprocessing

Data Exploration involves analyzing the Walmart sales dataset to understand patterns, distributions, and outliers. Preprocessing includes handling missing values, scaling, and encoding categorical variables. These crucial steps enhance data quality, ensuring the reliability and effectiveness of subsequent analyses in the sales prediction.

**Ref. template:** [**Click Here**](https://github.com/SmartInterns-Content/AI-ML-DA-Templates/blob/main/Machine%20Learning%20and%20Natural%20Language%20Processing%20Templates/2.%20Data%20Collection%20and%20Preprocessing%20Phase_/Data%20Exploration%20and%20Preprocessing%20template.pdf)

**Walmart Sales Analysis Data Exploration and Preprocessing Report:** [**Click Here**](https://github.com/SmartInterns-Content/AI-ML-DA-Templates/blob/main/Sample%20Project/2.%20Data%20Collection%20and%20Preprocessing%20Phase_/SL%20Data%20Exploration%20and%20Preprocessing%20template.pdf)

# 4. Model Development Phase

The Model Development Phase entails crafting a predictive model for sales prediction. It encompasses strategic feature selection, evaluating and selecting models (Random Forest, Decision Tree, XGB, Arima model), initiating training with code, and rigorously validating and assessing model performance for informed decision-making in the lending process.

## 4.1 Feature Selection Report

The Feature Selection Report outlines the rationale behind choosing specific features (e.g., Store, Department, isHoliday, Month, Year, Size) for the sales prediction model. It evaluates relevance, importance, and impact on predictive accuracy, ensuring the inclusion of key factors influencing the model's ability to discern credible loan applicants.

**Ref. template:** [**Click Here**](https://github.com/SmartInterns-Content/AI-ML-DA-Templates/blob/main/Machine%20Learning%20and%20Natural%20Language%20Processing%20Templates/3.%20Model%20Development%20Phase/Feature%20Selection%20Report%20template.pdf)

**Walmart Sales Analysis Feature Selection Report:** [**Click Here**](https://github.com/SmartInterns-Content/AI-ML-DA-Templates/blob/main/Sample%20Project/3.%20Model%20Development%20Phase/SL%20Feature%20Selection%20Report.pdf)

## 4.2 Model Selection Report

The Model Selection Report details the rationale behind choosing Random Forest, Decision Tree, Arima, and XGB models for sales prediction. It considers each model's strengths in handling complex relationships, interpretability, adaptability, and overall predictive performance, ensuring an informed choice aligned with project objectives.

**Ref. template:** [**Click Here**](https://github.com/SmartInterns-Content/AI-ML-DA-Templates/blob/main/Machine%20Learning%20and%20Natural%20Language%20Processing%20Templates/3.%20Model%20Development%20Phase/Model%20Selection%20Report%20template.pdf)

**Walmart Sales Analysis Model Selection Report:** [**Click Here**](https://github.com/SmartInterns-Content/AI-ML-DA-Templates/blob/main/Sample%20Project/3.%20Model%20Development%20Phase/SL%20Model%20Selection%20Report.pdf)

## 4.3 Initial Model Training Code, Model Validation and Evaluation Report

The Initial Model Training Code employs selected algorithms on the Walmart sales Analysis dataset, setting the foundation for predictive modeling. The subsequent Model Validation and Evaluation Report rigorously assesses model performance, employing metrics like accuracy and precision to ensure reliability and effectiveness in predicting loan outcomes.

**Ref. template:** [**Click Here**](https://github.com/SmartInterns-Content/AI-ML-DA-Templates/blob/main/Machine%20Learning%20and%20Natural%20Language%20Processing%20Templates/3.%20Model%20Development%20Phase/Initial%20Model%20Training%20Code%2C%20Model%20Validation%20and%20Evaluation%20Template.pdf)

**Walmart Sales Analysis Model Development Phase Template:** [**Click Here**](https://github.com/SmartInterns-Content/AI-ML-DA-Templates/blob/main/Sample%20Project/3.%20Model%20Development%20Phase/SL%20Initial%20Model%20Training%20Code%2C%20Model%20Validation%20and%20Evaluation%20Report.pdf)

# 5. Model Optimization and Tuning Phase

The Model Optimization and Tuning Phase involves refining machine learning models for peak

performance. It includes optimized model code, fine-tuning hyperparameters, comparing

performance metrics, and justifying the final model selection for enhanced predictive accuracy

and efficiency.

## 5.1 Hyperparameter Tuning Documentation

The Random Forest model was selected for its superior performance, exhibiting high accuracy during hyperparameter tuning and after cross validation. Its ability to handle complex relationships, minimize overfitting, and optimize predictive accuracy aligns with project objectives, justifying its selection as the final model.

## 5.2 Performance Metrics Comparison Report

The Performance Metrics Comparison Report contrasts the baseline and optimized metrics for various models, specifically highlighting the enhanced performance of the Random Forest model. This assessment provides a clear understanding of the refined predictive capabilities achieved through hyperparameter tuning. Random Forest has the maximum accuracy. With minimum loss of information as compared to other models. The RMSE and MSE score is low for the random forest model making it a optimal model for sales prediction.

## 5.3 Final Model Selection Justification

The Final Model Selection Justification articulates the rationale for choosing Random Forest as the ultimate model. Its exceptional accuracy, ability to handle complexity, overcome the problem of overfitting and successful hyperparameter tuning align with project objectives, ensuring optimal sales predictions.

**Ref. template:** [**Click Here**](https://github.com/SmartInterns-Content/AI-ML-DA-Templates/blob/main/Machine%20Learning%20and%20Natural%20Language%20Processing%20Templates/4.%20Model%20Optimization%20and%20Tuning%20Phase/Model%20Optimization%20and%20Tuning%20Phase%20Template.pdf)**s**

**Walmart Sales Analysis Model Optimization and Tuning Phase Report:** [**Click Here**](https://github.com/SmartInterns-Content/AI-ML-DA-Templates/blob/main/Sample%20Project/4.%20Model%20Optimization%20and%20Tuning%20Phase/SL%20Model%20Optimization%20and%20Tuning%20Phase%20Template.pdf)

# 6. Project Files Submission and Documentation

For project file submission in Github, Kindly click the link and refer to the flow. [**Click Here**](https://github.com/SmartInterns-Content/AI-ML-DA-Templates/blob/main/Machine%20Learning%20and%20Natural%20Language%20Processing%20Templates/Final%20submission.png)

For the documentation, Kindly refer to the link. [**Click Here**](https://github.com/SmartInterns-Content/AI-ML-DA-Templates/blob/main/Machine%20Learning%20and%20Natural%20Language%20Processing%20Templates/ML%20%26%20NLP%20Final%20Project%20Report%20Template.pdf)

# 7. Project Demonstration

# 8. Output Screenshots

# 9. Future Scope

# The future scope of the sales forecasting project using Walmart data includes incorporating additional data sources such as economic indicators, weather patterns, and regional events for a comprehensive understanding of sales trends. Advanced machine learning models like deep learning and ensemble methods can be explored to improve forecasting accuracy. Real-time data processing and forecasting will allow Walmart to quickly respond to market changes. Hyperparameter tuning and model optimization can further enhance accuracy and robustness. Geographical and store-level analysis can offer targeted insights for specific regions or stores. Seasonal and long-term trend analysis can guide strategic planning, while improved user experience and visualization will aid decision-making. Integrating the forecasting system with existing business processes such as inventory management and supply chain planning can lead to efficient operations. Additionally, exploring causal relationships between sales, promotions, and holidays can provide deeper insights into sales fluctuations.

# 