

## Project Initialization and Planning Phase

Date	4 June, 2024
Team ID	SWTID1719938571
Project Title	Walmart Sales Analysis for Retail Industry with Machine Learning
Maximum Marks	3 Marks

### Project Proposal (Proposed Solution)

This project proposal outlines a solution to address a specific problem. With a clear objective, defined scope, and a concise problem statement, the proposed solution details the approach, key features, and resource requirements, including hardware, software, and personnel.

Project Overview	
Objective	The objective of this project is to utilize machine learning algorithms and historical sales data from Walmart to enhance sales performance and decision-making in the retail industry. The project aims to address key challenges such as accurately forecasting product demand, effectively segmenting customers, and optimizing pricing strategies. By developing predictive models for demand, analyzing customer purchase behavior for better segmentation, and determining optimal pricing strategies, the project seeks to improve inventory management, tailor marketing efforts, and maximize revenue. The ultimate goal is to provide actionable insights that lead to increased operational efficiency, customer satisfaction, and overall sales performance for retailers.
Scope	The scope of this project includes analysing sales data from 45 Walmart stores, incorporating store-specific details and monthly sales metrics. The analysis aims to identify patterns and trends, with a focus on assessing the effects of promotional markdown events and holidays on sales performance. The project will implement machine learning algorithms like Random Forest, Decision Tree, XGBoost, and ARIMA to develop and test a reliable forecasting model. Furthermore, the project involves integrating the model into a Flask application.
Problem Statement	
Description	This project involves a comprehensive analysis of sales data from 45 Walmart stores to enhance sales performance and decision-making in the

	retail sector. The analysis will identify patterns and trends, particularly examining the impact of promotional markdown events and holidays on sales. Machine learning algorithms such as Random Forest, Decision Tree, XGBoost, and ARIMA will be employed to develop and test a robust demand forecasting model. The project also includes integrating this model into a Flask application.
Impact	The impact of this project is significant for retail management, as it provides actionable insights to optimize sales performance and decision-making processes. By accurately forecasting product demand, the project will enhance inventory management, reducing both excess inventory and stockouts. Effective customer segmentation will enable targeted marketing and personalized promotions, leading to increased customer satisfaction and loyalty. Optimized pricing strategies, based on comprehensive analysis, will help maximize revenue and maintain competitiveness in the market. Overall, this data-driven approach to retail management can drive sustained growth and improve operational efficiency, making it a valuable asset for stakeholders.
<b>Proposed Solution</b>	
Approach	The project will adopt a structured approach, starting with data preprocessing to manage outliers and ensure data quality. Historical sales data will then be used to train and test the forecasting model with machine learning algorithms such as Random Forest, Decision Tree, XGBoost, and ARIMA. The analysis will specifically assess the impact of holidays on sales, assigning greater weight to holiday weeks. The resulting model will be integrated into a Flask application to facilitate user interaction and ease of use. Finally, the Flask application will be deployed on an appropriate platform to ensure accessibility and scalability.
Key Features	<ul style="list-style-type: none"> <li>□ <b>Data Preprocessing:</b> Handling outliers and ensuring data quality for accurate analysis.</li> <li>□ <b>Machine Learning Algorithms:</b> Utilizing Random Forest, Decision Tree, XGBoost, and ARIMA to develop and test forecasting models.</li> <li>□ <b>Sales Trend Analysis:</b> Identifying patterns and trends in historical sales data.</li> <li>□ <b>Holiday Impact Evaluation:</b> Assessing the effect of holidays on sales, with higher weighting for holiday weeks.</li> <li>□ <b>Customer Segmentation:</b> Analyzing customer behavior for effective segmentation and targeted marketing.</li> <li>□ <b>Pricing Optimization:</b> Developing strategies to optimize pricing and maximize revenue.</li> <li>□ <b>Model Integration:</b> Incorporating the developed model into a Flask application for user interaction.</li> <li>□ <b>Platform Deployment:</b> Ensuring accessibility and scalability by deploying the Flask application on a suitable platform.</li> </ul>

## Resource Requirements

Resource Type	Description	Specification/Allocation
<b>Hardware</b>		
Computing Resources	CPU/GPU specifications, number of cores	2 x NVIDIA V100 GPUs
Memory	RAM specifications	8 GB
Storage	Disk space for data, models, and logs	1 TB SSD
<b>Software</b>		
Frameworks	Python frameworks	Flask
Libraries	Additional libraries	scikit-learn, pandas, numpy
Development Environment	IDE, version control	Jupyter Notebook, Git
<b>Data</b>		
Data	Source, size, format	Kaggle dataset ( <a href="https://www.kaggle.com/competitions/walmart-recruiting-store-sales-forecasting/data">https://www.kaggle.com/competitions/walmart-recruiting-store-sales-forecasting/data</a> )