Project Initialization and Planning Phase

Date	15 July 2024
Team ID	740049
Project Title	Number Oracle:Big Mart Sales Predictive Analysis
Maximum Marks	3 Marks

Project Proposal (Proposed Solution) template

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	Develop a machine learning model to predict sales for Big Mart stores, providing accurate forecasts based on historical sales data and store attributes.

Scope	Data collection and preprocessing Feature extraction and selection Model development and training Validation, testing, and deployment Documentation and results presentation
Problem Statement	
Description	Predicting sales for Big Mart stores is complex and requires analyzing various factors. An ML model can enhance prediction accuracy by processing extensive historical and real-time data.
Impact	Improves sales prediction accuracy for store managers and analysts, aids in inventory and resource planning, and enhances the overall understanding of sales dynamics.
Proposed Solution	
Approach	Collect historical sales data, store attributes, and product information. Handle missing values, outliers, and data inconsistencies.

Resource Type	**Description**	**Specification/Allocation**
Hardware		
Computing Resources	CPU/GPU specifications, number of cores	e.g., 2 x NVIDIA V100 GPUs
Memory	RAM specifications	e.g., 8 GB

Storage	Disk space for data, models, and logs	e.g., 1 TB SSD
Software		
	Extract relevant features (e.g., product weight, visibility, MRP, store type). Create new features (e.g., visibility bins, store age). Develop and train ML models (e.g., Linear Regression, Decision Trees, Random Forest, XGBoost).	
Key Features	High prediction accuracy Real-time updates Scalability User-friendly interface for predictions	

Resource Requirements

Frameworks	Python frameworks	e.g., Flask
Libraries	Additional libraries	e.g., scikit-learn, pandas, NumPy
Development Environment	IDE, version control	e.g., Jupyter Notebook, Git
Data		
Data	Source, size, format	e.g., Kaggle dataset CSV files, 10,000 recods