Sales Forecasting System

Final Project Proposal

Project Title: Sales Forecasting System

Group Members & Roles:

- Ahmed Osama – Data Engineer / Data Cleaning
- Rahma Saber – Feature Engineering & Exploratory Data Analysis (EDA)
- Malak Sherif – Model Evaluation & Reporting

Team Leader: Khaled Abdulrahman

Project Description

The Sales Forecasting System aims to build a predictive model that leverages data analysis and machine learning techniques to forecast sales accurately. The project supports managerial decision-making by providing data-driven insights that help anticipate demand, optimize inventory, and plan promotional strategies.

Main Tasks:

- Collect and preprocess sales, customer, and discount data.
- Analyze time series data for trends, seasonality, and correlations.
- Build and compare forecasting models (Prophet, RandomForestRegressor, etc.).
- Evaluate model performance and optimize parameters.
- (Optional) Deploy the model using an API for real-time forecasting.

Objectives

Analyze historical sales data

to uncover patterns and influencing factors.

Support business decision-making

with accurate sales forecasts (monthly, quarterly, yearly).

Study the impact of discounts, promotions, and seasonality

on demand.

Develop a Real-Time Forecasting Model

for fast and automated predictions.

Enhance operational efficiency

and reduce the risks of inaccurate sales estimations.

Tools & Technologies

Programming & Data

- Programming Language: Python (Jupyter Notebook)
- Data Processing: pandas, numpy
- Statistical Analysis: scipy

Modeling & Version Control

- Forecasting Models: Prophet, RandomForestRegressor (scikit-learn)
- Model Saving: joblib
- Version Control: Git / GitHub

Visualization & Deployment

- Visualization: matplotlib, seaborn, plotly
- Deployment (optional): Flask / FastAPI

Milestones & Deadlines

Submission Deadline: 13/11/2025

Task Description	Details	Target Date
Data Collection & Initial Review	Data gathering and structure overview	2025-09-01 - 2025-09-09
Data Cleaning & Preprocessing	Handle missing values, transformations, and feature creation	2025-09-10 - 2025-09-25
EDA & Feature Engineering	Statistical insights, correlation analysis, and visualization	2025-09-26 - 2025-10-08
Model Selection & Training	Prophet and Random Forest training with tuning	2025-10-09 - 2025-10-30
Evaluation & Validation	Metrics calculation and reporting	2025-10-31
Deployment & Documentation	Model saving (joblib) and final report	2025-11-13

KPIs (Key Performance Indicators)

Data Quality

Missing Values Handled

Percentage of missing values handled: 100%

Data Accuracy

Data accuracy after preprocessing: ⩾ 98%

Dataset Diversity

Dataset diversity (representation of different categories): 94.01%

Model Performance

Prophet Model

- Mean Squared Error (MSE): ≈ 5.74×10^{16}
- Root Mean Squared Error (RMSE): ≈ 2.39×10^{8}
- Mean Absolute Error (MAE): ≈ 1.23×10^{8}
- Mean Absolute Percentage Error (MAPE): ≈ 0.08
- Symmetric Mean Absolute Percentage Error (SMAPE): ≈ 0.09
- Coverage: ≈ 0.66

Summary: The Prophet model achieved good forecasting stability with a relatively low MAPE (8%) and SMAPE (9%), indicating reasonable accuracy in long-term trend prediction.

Random Forest Model

- R^2 (Train): 0.7661
- R^2 (Test): 0.6743
- Mean Absolute Error (MAE): 4645.28
- Root Mean Squared Error (RMSE): 6226.21

Summary: The Random Forest model performed well, achieving an R^2 of 0.67 on the test data, demonstrating strong predictive capability and generalization.

Note: Both models are still in the early development phase; additional tuning and feature optimization are planned to further improve accuracy and reduce latency

Deployment & Scalability

API Uptime

~ ....

Response Time

per Request: ≈ 180 ms

Real-Time Processing Speed

Supports near real-time prediction with latency below 200 ms, suitable for live dashboards.

Scalability

The architecture supports horizontal scaling via containerization (Docker) and cloud deployment (AWS).

The deployment pipeline enables seamless updates and ensures high reliability with minimal downtime.

Business Impact & Practical Use

70%

Reduction in Manual Effort

Automated forecasting replaces repetitive analytical work.

45%

Expected Cost Savings

Lower operational costs through faster and more accurate decision-making.

.....

User Satisfaction

Anticipated high satisfaction from data-driven insights.

Overall Summary: Both models – Prophet and Random Forest – have shown strong potential for real-world implementation. Although still in the optimization stage, they already deliver promising forecasting accuracy, reduced human workload, and measurable business value.

FINAL PROJECT STATUS REPORT

Evaluation Plan & Deliverables

Deliverables:

- Documented Jupyter Notebook
- Trained Model File (.joblib / .pkl)
- Final Report (PDF/Word) with methodology, visuals, and results
- (Optional) Simple API Demo

Evaluation Criteria:

- Smooth notebook execution
- Clarity of EDA insights and visuals
- Model accuracy and performance metrics
- Completeness and readability of documentation