Final project proposal
Sales Forecasting and Optimization

Project Description:

This project builds an end-to-end, sales predictive system that accurately forecasts future sales for an e-commerce business using historical sales data, transforming them into accurate short- and mid-term forecasts that drive inventory planning, marketing spend allocation, and sales execution decisions. The solution covers data collection, cleaning, exploratory data analysis (EDA), feature engineering , time-series model development using best practices, and deployment of a production-ready forecasting application.

Group Members & Roles:

The leader assigns tasks to members, reviews them and choose the best member for each

1. Mariam Ashraf Hashim Mohamed : Data exploration, Further Data cleansing

2. Jana Khaled Salah Abdelwahab : Data cleansing, Data visualization

3. Sama Haitham Ezzat Eisa : Data cleaning, Data Analysis

4. Loay Mohamed Youssef : Data exploration, power BI dashboard

5. Anas Ahmed Mahmoud Khalil : Required data preprocessing, Data Analysis

Team Leader:

Mariam Ashraf Hashim Mohamed

Objectives:

- Deliver a clean, analysis-ready sales dataset with engineered features to capture trends, seasonality, and event effects.

- Build and compare at least three forecasting approaches

- Select the best-performing model via hyperparameter optimization to minimize RMSE/MAE/MAPE, supported by residual diagnostics.

- Deploy as a scalable service for real-time and batch inference with access-controlled UI

- Deploy the solution as a scalable service supporting real-time and batch inference behind an access-controlled user interface.

- Implement the final model within an MLOps pipeline for automated training and prediction.

- Develop monitoring metrics (e.g., forecast drift, retraining triggers) for continuous model performance evaluation.

- Create interactive Power BI dashboard to visualize historical trends, forecast results, and optimization recommendations.

- Deliver comprehensive documentation and a final presentation summarizing methodology, findings, and business impact.

- Provide stakeholder-ready materials and a live demo linking model outcomes to KPIs such as inventory turns, stockout reduction

Tools & Technologies

Python stack: Pandas, NumPy

Visualization: Matplotlib, Seaborn, Plotly

Machine learning: scikit-learn, Facebook Prophet, XGBoost / LightGBM

Model evaluation: Scikit-learn.metrics, MLflow

APIs: FastAPI   -   Interactive apps: Streamlit

Dashboards: Power BI

(Might use) cloud deployment : google cloud

Milestones & Deadlines:

1. Research & Dataset Acquisition   *6 Sep.*
2. Required Data preprocessing   *22 Sep.*
3. Data cleaning & exploration   *30 Sep.*
4. EDA & Feature engineering   *11 oct.*
5. Data Visualization   *17 oct.*
6. Model Development & Optimization   *8 Nov.*
7. MLOps, Deployment & Monitoring   *25 Nov.*
8. Final Documentation & Presentation   *29 Nov.*

KPIs:

1. Data Quality:

    Percentage of missing values handled:100%

    Data accuracy after preprocessing: ≥ 98%

    Dataset diversity: ≥ 80%

2. Model Performance:

    Model prediction speed (Latency): ≤ 300 ms

    Error rate: ≤ 5%

3. Deployment & Scalability:

    API uptime: ≥ 99.5%

    Response time per request: ≤ 500 ms