Title

- Project on Pharmaceutical Sales Prediction

- Presented by- Deepak Sharma

Introduction

- This project aims to predict pharmaceutical sales across multiple stores using historical data.

Objective

► The primary objective of this project is to develop a reliable predictive model that can forecast future pharmaceutical sales across multiple store locations. This will assist store managers in making informed decisions regarding stock levels and staffing, ultimately optimizing operational efficiency and customer satisfaction.

Overview of Files

- The project involves a Flask web application and three Jupyter notebooks tasks files focusing on data analysis and prediction modeling.

Technology Stack

- Python, Flask, Pandas, NumPy, Matplotlib, Seaborn, Joblib

Model Overview

- A trained Random Forest model is used for making predictions.

Data Overview

▶ The data consists of sales records from multiple stores, including dates, sales figures, and additional store-related features.

Data Preprocessing

▶ Data cleaning and preprocessing steps are crucial to prepare the data for analysis and modeling.

Exploratory Data Analysis (EDA) 1

- ▶ Given this structure, the 'Store' column has been used to merge the store_df dataset with both the test_df and train_df datasets. This will enrich the test and train datasets with store-specific information, which is crucial for analysis.

EDA 2

- Prediction of sales is the central task in this challenge. We predicted daily sales in various stores up to 6 weeks ahead of time.

Feature Engineering

- Features developed from the data include trends, seasonality components, and store-specific attributes.

Modeling Strategy

▶ The approach includes training multiple models and selecting the best performer based on validation results.

Model Performance

- Evaluation metrics such as RMSE and MAE are used to assess model accuracy.

Predictions

- The model predicts future sales for up to six weeks ahead, aiding in inventory and staff planning.

Web Application

- A Flask app that allows users to input data and receive sales predictions.

User Interface

- The interface is simple and user-friendly, designed to be accessible for non-technical users.

Challenges

- Challenges included dealing with missing data and high variability in sales.

Solutions

- Imputation techniques and robust outlier handling methods were used.

Conclusions

- The project successfully provides accurate sales forecasts that can help in strategic planning.

Future Work

- Further improvements could include real-time data integration and enhanced predictive models.