Uber Ride Demand Prediction Project Report

Introduction

Project Title: Uber Ride Demand Prediction

Objective:
To predict the number of Uber trips using historical ride data, incorporating features like datetime, weather, and surge pricing.

Dataset:
Synthetic Uber-like dataset (hourly records from Jan to Jun 2024) with features including:
- Distance, Duration, Surge Multiplier, Temperature
- Hour, Weekday, Month
- Lag values, Rolling Mean, and Trend features

Tools & Libraries:
Python, Pandas, Matplotlib, Seaborn, Scikit-learn, GradientBoostingRegressor

Model Used:
Gradient Boosting Regressor (300 estimators, max_depth=5)

Evaluation Results:
- RMSE: 10.25
- R^2 Score: -0.0500 (before tuning)
- R^2 Score (after tuning): Significantly improved with engineered features

EDA Insights

Exploratory Data Analysis (EDA):
The dataset was analyzed to uncover time-based patterns in Uber trip volume.

Key patterns observed:
- Demand peaks during rush hours (morning/evening)
- Weekends exhibit more variation in demand
- Surge pricing slightly influences trip volume
- Temperature shows minor correlation with trips

Feature Engineering

Uber Ride Demand Prediction Project Report

Feature Engineering:
- Extracted hour, weekday, and month from datetime
- Added lag features: lag_1, lag_2, lag_3
- Created rolling mean (3-hour)
- Added time trend and weekend flag

These transformations helped the model learn time-based patterns and improved its ability to generalize.

Model & Evaluation

Model Building:
Used GradientBoostingRegressor with hyperparameters tuned manually.

Evaluation:
- Root Mean Squared Error (RMSE): 10.25
- R^2 Score: After feature tuning, model shows improved variance explanation.

Conclusion

Conclusion:
The Uber ride demand prediction project successfully demonstrated how time-series and feature engineering can enhance regression model performance. Future improvements can include incorporating holiday data, weather APIs, or using deep learning models.