Name : Sridivya Gorantla

Project Title: Predictive Analytics for NYC Taxi Trip Duration

**Problem Statement**:

New York City's taxi network is an important part of transportation in the city. It serves millions of passengers each year. Predicting taxi travel times is essential for optimizing operations. Reducing fuel costs and increase user satisfaction Travel time depends on many factors. This includes traffic patterns, weather conditions, distance and time of day. This project addresses the challenge of accurate forecasting using advanced machine learning (ML) techniques, aiming to provide actionable insights for taxi companies and city planners. This allows for efficient delivery and enhances the driver experience.  
  
**Idea about Model/Methodology:**

The project adopts a structured ML approach, starting with thorough data preprocessing to handle missing values, outliers, and inconsistencies in the NYC Taxi Trip dataset. Feature engineering will extract meaningful variables such as:

* **Haversine Distance**: Measures direct distance between pickup and drop-off points.
* **Temporal Features**: Includes pickup hour, day, and season, capturing time-based patterns.
* **Weather Influence**: Integrates data on precipitation and temperature to account for weather impacts.

Key machine learning algorithms to be employed include:

* **Linear Regression**: Establishes a baseline for model performance.
* **XGBoost**: Utilizes decision tree ensembles for robust predictions and outlier handling.
* **Random Forest**: Provides interpretability with feature importance measures.
* **Deep Neural Networks**: Explores non-linear relationships for enhanced accuracy.
* Cross-validation ensures model reliability, while hyperparameter tuning maximizes performance. Evaluation metrics include Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R² score.

**Dataset Links**

1. **NYC Taxi Trip Dataset** (Kaggle): Provides historical trip data, including pickup and drop-off locations, timestamps, and passenger counts.  
   https://www.kaggle.com/c/nyc-taxi-trip-duration
2. **Weather Data** (NOAA): Supplies detailed weather records for New York City, aiding feature engineering.  
   https://www.ncdc.noaa.gov/

**Literature Review**

1. **Short Hills Tech (2021)**: Achieved 95% accuracy in NYC taxi trip duration prediction using XGBoost, highlighting the importance of feature engineering and hyperparameter tuning.
2. **Liu et al. (2019)**: Introduced traffic-aware models using ensemble learning, emphasizing urban mobility challenges.
3. **Li & Zheng (2017)**: Integrated weather impacts into ML models, achieving improved accuracy in smart transport systems.
4. **Zhang et al. (2018)**: Utilized deep learning for ride-sharing analytics, showing the advantages of non-linear models.
5. **NYC Data Analysis Report (2020)**: Highlighted the critical role of geospatial variables in predictive analytics.
6. **TripTime AI (2021)**: Demonstrated hybrid XGBoost and temporal models for trip duration predictions.
7. **Seoul Bike Study (2021)**: Showcased the value of combining meteorological and geospatial data for mobility predictions.
8. **Isolated XGBoost Models (2022)**: Achieved accurate predictions even under extreme conditions.
9. **Li et al. (2022)**: Explored interpretability of ML models for urban transport systems.
10. **Ahamed et al. (2019)**: Proposed lightweight ML architectures for faster computations and predictions.

**Conclusion:**

This project bridges data science and urban mobility by providing an ML-based framework for predicting taxi trip durations. It combines advanced algorithms, rigorous validation, and meaningful insights to improve urban transportation systems. The findings could directly benefit taxi operations and enhance commuter experiences.