**NYC Taxi Fare Prediction using Explainable AI (XAI)**

**Introduction**

Explainable Artificial Intelligence (XAI) is essential in promoting transparency and trust in machine learning applications. As predictive models become more complex, it becomes increasingly important to ensure that these models can be interpreted and understood, especially when deployed in sensitive or high-stakes environments.

In this project, we apply XAI techniques to a fare prediction model trained on NYC yellow taxi trip data. The aim is to predict the taxi fare amount using relevant trip features and, more importantly, to understand how different features influence each prediction. Various model explanation techniques are explored, including SHAP, LIME, Permutation Importance, Partial Dependence Plots, ICE, and TreeInterpreter.

**Data**

The dataset used originates from the NYC Taxi and Limousine Commission, consisting of ride records from January 2015. For processing efficiency, a random sample of 100,000 entries was used. Data cleaning involved:

* Removing rows with missing values,
* Filtering out trips with non-positive fare amounts or passenger counts.

Feature engineering was critical for enhancing the model’s performance and included:

* **Haversine distance** calculation between pickup and drop-off coordinates to represent trip length,
* **Hour of pickup** extracted from the timestamp to account for time-based pricing (rush hours, night rates),
* **Day of the week** to capture weekly travel trends.

These features formed the basis for training the machine learning model.

**Experiment:**

A **Random Forest Regressor** was chosen for its robustness, interpretability, and efficiency with tabular data. The model was trained using the features: passenger count, distance (in kilometers), pickup hour, and day of the week. The model achieved a **Root Mean Squared Error (RMSE)** of approximately **4.15**, demonstrating good predictive performance for a basic engineered dataset.

To interpret the model and ensure transparency, multiple XAI methods were applied:

* **SHAP (SHapley Additive exPlanations):**  
  SHAP was used to evaluate feature influence both globally (entire model) and locally (individual predictions). The summary plot clearly indicated that distance\_km and hour were the most impactful features. The plot was centered at zero, effectively showing the direction (positive or negative) and magnitude of each feature's effect.
* **Permutation Importance:**  
  By shuffling feature values and measuring the change in prediction error, we confirmed the critical role of distance\_km. This method offered a straightforward way to rank features by importance.
* **Partial Dependence Plots (PDP) and ICE (Individual Conditional Expectation):**  
  PDPs showed the **average** effect of a feature (e.g., how fare increases with distance), while ICE plots revealed the **individual-level** variations. These visualizations made it easy to understand how features affected predictions across different rides.
* **LIME (Local Interpretable Model-agnostic Explanations):**  
  LIME generated simplified linear models that approximate the behavior of the Random Forest model for individual predictions. It helped explain the rationale behind specific fare outputs by quantifying how each input feature contributed.
* **TreeInterpreter:**  
  TreeInterpreter provided a decomposition of the prediction into a bias term (average prediction) and additive feature contributions. This approach was especially suitable for tree-based models like Random Forest and served as a cross-validation check for SHAP and LIME insights.

**Conclusion:**

This project demonstrates that combining predictive modeling with Explainable AI techniques offers significant advantages in understanding model behavior and gaining stakeholder trust. Each XAI method contributed a different layer of interpretability:

* **SHAP and Permutation Importance** aligned on feature significance, validating the consistency of global insights.
* **LIME and TreeInterpreter** effectively explained individual predictions, useful for audits or debugging.
* **PDP and ICE** visualized feature effects over a range of values, improving understanding of nonlinear interactions.

Through this analysis, it became clear that **trip distance** and **time of day** are the dominant factors influencing NYC taxi fares. These methods not only confirmed our assumptions but also ensured the model’s transparency and accountability — an essential requirement for deploying models in real-world decision-making systems.