Introduction:

The project "Taxi Tip Prediction:

A Machine Learning Model" aims to create a regression model that can predict the amount of tip paid in taxi rides using a real dataset collected by the NYC Taxi and Limousine Commission. The dataset contains important information about taxi trips, and the goal is to develop a model that can predict the tip amount accurately.

Used Libraries:

The project uses both the Scikit-Learn Python interface and the Python API provided by the Snap Machine Learning (Snap ML) library. Snap ML is preferred for its high-performance implementations of linear and tree-based models, which offer acceleration through system awareness and superior accuracy.

Link to the dataset is

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-ML0101EN-SkillsNetwork/labs/Module%203/data/yellow_tripdata_2019-06.csv

Objectives:

The main objectives of the project include:

- Basic data preprocessing using Scikit-Learn to prepare the dataset for modelling.
- Modelling a regression task using both Scikit-Learn and Snap ML Python APIs.
- Training Decision Tree Regressor models using both libraries to compare their performance.
- Running inference on the trained models and evaluating their quality using evaluation metrics.

Significance:

The project is significant as it has the potential to provide valuable insights for taxi service providers and drivers, allowing them to optimize their operations and improve customer satisfaction. Additionally, the project explores the use of Snap ML, highlighting its efficiency and accuracy in machine learning tasks.

Summary:

Overall, the project contributes to advancing the field of predictive analytics in transportation services while showcasing the capabilities of modern ML techniques and libraries.