Capstone Project Submission

Instructions:
i) Please fill in all the required information.
ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
Name : Rajvee Sharma **Email : rajveesharmae100@gmail.com** *I made my project solely.* *Firstly, prepare a google colab notebook for data cleaning, data manipulation, data visualization , applying many ML algorithms, and finalizing the conclusion.* *Make a ppt by making sure all points are covered.* *Prepare a technical documentation on the content of the problem and statement goal of the project.*
Please paste the GitHub Repo link.
Github Link:- https://github.com/Link/to/Repo
Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)
Data science can be summarized into – capture, maintain process, analyze and communicate. In NYC there are a large number of taxis and taxi drivers are present. Because taxi mode of transportation has become a key player in the United States and other countries. Predicting the duration of a taxi trip is important since a user would always like to know precisely how much time it would require him to travel from one place to another. **Different service providers are UBER, OLA, green taxi, yellow taxi and many more. Given the rising population of app based taxi usage through common vendors like OLA, UBER. It can also help drivers to take correct turns which in turn will take lesser time accordingly. Moreover, the transparency about the pricing and trip duration will also attract the users at times when popular taxi app based vendor services apply surge fares. The dataset contains pickup and drop-off date and time, distance of trip, start time and end time, number of passenger count and rate code belonging to different classes of cabs and available such that the rate applied is based on regular and airport basis. Before starting our data science project which is data preparation, data cleaning and analyzing the data I just have to clean this data and perform EDA with it.** **Then I applied many ML algorithms like Linear Regression, Decision Tree, Random Forest. These visualizations showed how the model's predictions are close to test**

data. It was observed that Decision Tree and Random Forest performed well. When I did analysis of model evaluation results with PCA and observed that Decision Tree and Random Forest are well performers. Random Forest provides reduced RMSE and now says that it's a model to be opted for.

Decision Tree with hyper-parameter tuning is really a good performer as it is also providing a reduced RMSE and getting a good fit score for this i.e, close to 1. R2 score must be between 0 to 1, towards 1 is considered a good fit.

Taxi giants such as UBER and OLA can use the same data for analyzing the trends that vary throughout the day in this city. This not only helps in better transport analysis and also helps the concerned authorities in planning traffic control and monitoring.