**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| 1. Name: - Avinash Yadav, Email ID: -[avinet1995@gmail.com](mailto:avinet1995@gmail.com)  * Contributed In notebook helped with google diver data connectivity, Data cleaning, EDA, Visualization, Feature creation, feature engineering and feature selection, hyperparameter tuning, model training, model evaluation through different methods like R2 score and RMSE and model selection. * Contributed for the contents of ppt. * Contributed in Technical Documentation in content of problem statement goal of project and steps involved.      1. Name: - Deepika Yadav, Email ID: - [yadavdeepika729@gmail.com](mailto:yadavdeepika729@gmail.com)  * Contributed in notebook for, Data cleaning, EDA, Visualization, Feature creation, feature engineering and feature selection, hyperparameter tuning, model training, model evaluation through different methods like R2 score and RMSE and model selection. * Contributed in presentation PPT with points to be covered and all the images to me it looks good * Contributed in Technical Documentation in content of problem statement goal of project and steps involved. |
| **Please paste the Gather Repo link.** |
| Github Link:- https://github.com/deepika1905/NYC-Taxi-Trip-Time-Prediction |
| **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 five steps: capture, maintain process, analyze, and communicate. New York City taxi rides form the core of the traffic in the city of New York. The many rides taken every day by New Yorkers in the busy city can give us a great idea of traffic times, road blockages, and so on. Predicting the duration of a taxi trip is very important since a user would always like to know precisely how much time it would require of him to travel from one place to another. Given the rising popularity of app-based taxi usage through common vendors like Ola and Uber, competitive pricing has to be offered to ensure users choose them. Prediction of duration and price of trips can help users to plan their trips properly, thus keeping potential margins for traffic congestions. It can also help drivers to determine the correct route which in-turn will take lesser time as accordingly. Moreover, the transparency about pricing and trip duration will help to attract users at times when popular taxi app-based vendor services apply surge fares. Thus in this research study, we used real-time data which customers would provide at the start of a ride, or while booking a ride to predict the duration and fare. This data includes pickup and drop-off point coordinates, the distance of the trip, start time, number of passengers, and a rate code belonging to the different classes of cabs available such that the rate applied is based on a regular or airport basis.  Before we start our Data Science process which is Data Preparing, Data Cleansing and Data Analysis. . With the cleansed data, we have performed Exploratory Data Analysis to understand our dataset.  Hereafter we applied Linear Regression, Decision Trees and Random Forest Algorithms. Visualizations showed us how our model’s predictions were close to Test Data. It was evident that decision tree and Random forest were performing well. When we did analysis on model evaluation result with PCA we observed that decision tree model and random forest model were good performers. As, Random Forest was providing us reduced RMSLE, we can say that it’s a model to be opted for. We got good fit score for decision tree and random forest i.e. close to 1.0. When we did analysis on model evaluation result without PCA we observed that our decision tree with GridsearchCV model were good performers as it was providing us reduced RMSLE. We were getting good fit score for decision tree with GridsearchCV i.e. close to 1.0. R2-score must be between 0 and 1, towards 1 considered as good fit.  Taxi giants such as UBER and OLA can use the same data for analyzing the trends that vary throughout the day in the city. This not only helps in better transport analysis but also helps the concerned authorities in planning traffic control and monitoring. |