**NYC TAXI TRIP TIME PREDICTION**

|  |
| --- |
|  |
| This is an Individual project which is entirely done by me.    Name:D.Ankitha  Mail id : dakithareddy1997@gmail.com |
| **https://github.com/ankitha-dudla/nyc-taxi-trip-time-prediction** |
| Github Link:-https://github.com/ankitha-dudla |
|  |
| **INTRODUCTION :**  In today’s world it has become a race to gain more and more number of customers.In order to gain more number of customers ,companies/ vendors usually try to provide their customers with more comfort to attract them.  So here by using machine learning and Regression analysis ,we will be predicting the time of trip duration ,our customers will take and which algorithm is best suited for that time prediction.  **WORK FLOW:**  The steps followed for our project are :   * **Exploratory Data Analysis** :After loading the dataset we performed this method by comparing our target variable that is trip\_duration with other independent variables. * **Null values Treatment:**In order to get the best results ,we have to remove the Null values at the beginning of our project.Thankfully,there are no Null values present in our project. * **Encoding of categorical columns** :We used One Hot Encoding to produce binary integers of 0 and 1 to encode our categorical features because categorical features that are in string format cannot be understood by the machine and needs to be converted to numerical format. * **Feature Selection :**In these steps we used algorithms like Linear regression,lasso regression,Ridge Regression,GBoost algorithm,XGBoost algorithm to check the results of each feature i.e which feature is more important compared to our model and which is of less importance. * **Standardization of features:**Our main motive through this step was to scale our data into a uniform format that would allow us to utilize the data in a better way while performing fitting and applying different algorithms to it. * **Fitting different models:**For modelling we tried various Regression algorithms like:  1. **Linear Regression** 2. **Decision Tree Regression** 3. **Lasso Regression** 4. **Ridge Regression** 5. **Gradient Boost Algorithm** 6. **XGBoost Algorithm.**  * **Tuning the hyperparameters for better accuracy:**Tuning the hyperparameters of respective algorithms is necessary for getting better accuracy and to avoid overfitting in tree based models . * **Training and Testing the dataset :** In machine learning projects, we generally divide the original dataset into training data and test data.rain and test datasets are the two key concepts of machine learning, where the training data set is used to fit the model, and the test dataset is used to evaluate the model.   **CONCLUSION:**  That's it! We reached the end of our exercise.  Starting with loading the data so far we have done EDA , null values treatment, encoding of categorical columns, feature selection ,training and testing the data and then predict the best one out of it.  In all of these models Gradient Boosting performs well for our data set . |