Capstone Project Submission

**Instructions:**

1. Please fill in all the required information.
2. Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| 1. Ankur Vishwakarma   Email :- [Ankurvish1920@gmail.com](mailto:Ankurvish1920@gmail.com)   * + Data inspection   + Exploratory Data Analysis     - Checking distribution of features.     - Checking relation of target feature with independent feature.     - Used multiple graphs and did analysis on dataset.   + Feature Engineering     - Checking null values.     - Handling outliers.     - Encoding categorical features.   + Feature Selection     - Variance Threshold   + Model training     - Linear regression     - Xg boost regressor     - Xg boost regressor with hyperparametric tuning     - Random forest     - Random forest with hyperparametric tuning |

|  |
| --- |
|  |
| **Please paste the GitHub Repo link. https://github.com/ankurvish1920** |
|  |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| **PROBLEM STATEMENT:**  The goal of a project on public transport demand prediction is to develop a system or model that can accurately forecast the number of passengers who will use a public transportation system over a given period of time. This requires collecting and analyzing historical data on passenger numbers and other relevant factors, as well as implementing machine learning algorithms and other techniques to make predictions.  The project may involve developing a proof-of-concept prototype, a full-scale system for use by a public transportation provider, or a research study to evaluate the effectiveness of different prediction methods. Some of the key challenges in this type of project include dealing with data quality and missing data, accounting for changing patterns of demand over time, and selecting appropriate models and parameters to achieve accurate predictions.  Overall, the aim of a project on public transport demand prediction is to help transportation providers optimize their services, reduce costs, and improve the overall passenger experience by anticipating and meeting the needs of their customers.  **APPROACH:**   * The first step includes loading of dataset and then inspecting the data through which we get to know the summary or description of data, shape and size of data, null value count, and duplicates values in the data and about the data types of column. * On the basis of univariate, bivariate, and multivariate analysis, we have carried out several visualisations. First, we performed a Univariate analysis since we needed to comprehend each feature or column's individual significance and the insights it would add to our study. Second, we used bivariate analysis to examine how one column or characteristic affects another, as well as the direction these discoveries may take us. * Next Step we find our dependent variable which name is number of ticket. * Following data visualisation, we utilize onehotencoder and label encoding to perform encoding, which converts categorical data to numerical data. We then performed feature |

|  |
| --- |
| selection We then divide the data by 80:20 using train test split. 20% for model testing and 80% for model training.   * Then, various models are applied. We used , * Linear regression * Xg boost regressor * Xg boost regressor with hyperparametric tuning * Random forest * Random forest with hyperparametric tuning   **CONCLUSION:**  With 99% Training Accuracy and 99% Testing Accuracy, Random Forest and Xgoost has proven to be the most efficient model out of the algorithms used in our model, including Linear Regression, Lasso Regression, Ridge Regression, Decision Tree, and Random Forest. Whereas, Linear Regression, Lasso and Rigde are not fitting well into the data points. |