

Capstone Project Submission

Instructions:

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

Team Member's Name, Email and Contribution:

Rohan A G (rohan.a.g.rohu28@gmail.com):

- Exploratory data Analysis
- Outlier detection
- Outlier treatment
- Handling null values
- Label encoding
- Correlation Heatmap
- Train test split on dataset
- Machine learning model implementation
- Hyper parameter tuning
- Regression evaluation
- Summary, technical documentation

Please paste the GitHub Repo link.

GitHub Link: - <https://github.com/RohanAG2806/Bike-sharing-demand-prediction>

Please write a short summary of your Capstone project and its components.
Describe the problem statement, your approaches and your conclusions. (200400 words)

Currently Rental bikes are introduced in many urban cities for the enhancement of mobility comfort. It is important to make the rental bike available and accessible to the public at the right time as it lessens the waiting time.

Eventually, providing the city with a stable supply of rental bikes becomes a major concern. The crucial part is the prediction of bike count required at each hour for the stable supply of rental bikes.

We imported the dataset and read the csv file, null values were taken care of at first, Distribution was created using histogram for numerical features, later outliers were found out with the help of box plot and treated by log transformation.

Exploratory data analysis was done where few insights were obtained from the data using visualizations. Label encoding of categorical values was done. Correlation heat map was generated to understand the correlation among the variables and removed the features which has high correlation.

The dataset was then split into train and test data and various regression machine learning algorithms like Linear Regression, Polynomial Regression and Random Forest were applied on the data followed by hyper parameter tuning using Grid Search CV on Random Forest model to obtain the best possible results.

Comparison was made on Regression evaluation metrics such as mean squared error, mean absolute error, root mean squared error and r2 score on all the models used to predict the outcome.

Random Forest is the model with best performance for the given dataset having train R2 score of 98.21 % and test R2 score of 86.98 %.