

Capstone Project Submission

Instructions:

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

Team Member's Name, Email and Contribution:

Team member's role:

- 1) Sahil pardeshi(862387921.sp@gmail.com)

Contribution: 1) Data understanding

2)Data visualization

3)Multivariate analysis

4)lasso regression model

5)Decision Tree model

6) Hyperparameter on decision tree model

- 2) Pravin Bejo(praveen.bejo.pb@gmail.com)

Contribution: 1) Data understanding

2)Data visualization

3)Multivariate analysis

4)Ridge regression model

5)Gradient boosting model

6)Hyperparameter tuning on gradient boosting model

- 3) Kirtesh verma(kirteshverma12345@gmail.com)

Contribution: 1)Data understanding

- 2) Handling null or missing values
- 3) Performing EDA
- 4) Removing Outliers
- 5) Linear regression model
- 6) Random forest model
- 7) Hyperparameter tuning on random forest

Please paste the GitHub Repo link.

Github

Link:<https://github.com/Sahilpardeshi1/Seoul-bike-sharing-demand-prediction>

Google Drive Link:

<https://drive.google.com/drive/u/0/folders/1ku5WDNyGZBB5eOCLoBC3LE5WRau7RCQT>

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

A bike sharing system, bike share program, public scheme, or public bike share scheme, is a shared transport service in which bicycles are made available for shared use to individuals on a short term basis for a price or free.

The bike sharing system can provide urban residents with a convenient and time saving travel mode. We find that bike sharing dramatically decreases traffic, reduces energy consumption, decreases harmful gas emissions, improves public health generally, and promotes economic growth.

continue.....

A new report from the Mineta Transportation Institute sifts through data from bike-share systems in Washington, DC, San Francisco, and Minneapolis. They found that bike-share bikes had lower collision and injury rates than personal bikes in all three cities.

The big problem statement was to build a machine learning model that could predict a rental bike count required for an hour, or given on another variable.

The 1st step was to perform EDA Exploratory data analysis where we have tried to insight data. It includes multivariate and univariate in which we identify certain trends, relationships, correlation and some impact on dependent variables. The 2nd step was to clean data and perform modification models. We have checked missing models and their outliers. We have also cleared dummy variables for some categorical features. The 3rd step was to standardize the learning algorithms. We have tried various algorithms like lasso regression model, ridge regression model, Hyperparameter on decision tree model, Hyperparameter tuning on random forest model, gradient boosting model and technical documentation. In machine learning the best performance was given by gradient boosting and random forest model where the R2 score for training and set was 0.87 and 0.83 for random forest and 0.92 and 0.88 for gradient boosting model respectively

A hyperparameter is a machine learning parameter whose value is chosen before a learning algorithm is trained. The most important features who had a major prediction were hours, temperature, solar radiation, humidity, winter and rainfall. demand of bikes are high during summer season and lower demand during winter season. The model performed well in this case but as the data is time dependent, the value of temperature, wind speed, solar radiation etc. will not also be consistent.

Machine learning is an exponentially model evolving field, we will have to be prepared for all contingency and also keep checking our model from time to time.