

Capstone Project Submission

Instructions:

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

Team Member's Name, Email and Contribution:

1. **Poonam Dattu Shevkar**

Email-id :- dattupoonam@gmail.com

❖ Contribution -

- Histogram plot for dependent & independent variables
- Feature Distribution
- Box plot and Scatter plot with features and Rented bike count.
- Outliers treatment with imputation of mean.
- Multicollinearity to evaluate the VIF factor of the numerical features.
- Model fitting by splitting the dataset into train and test data.
- Model developed for Decision Tree and Gradient Boosting Regressors .
- Hyperparameter Tuning on Gradient Boosting Regressors.

2. **Sanjay Ramkishan Verma**

Email-id :- shankyverma1998@gmail.com

❖ Contribution-

- Checking the presence of NAN values using missing values mno matrix.
- Feature engineering
- Best fitting line between Target & features.
- Correlation matrix
- Normalizing the features of the dataset.
- One hot coding technique on categorical variables.
- Model developed for Linear Regression , Lasso Regression and Random Forest regressors.

Please paste the GitHub Repo link-

Github Link:- <https://github.com/Sara19598/Seoul-Bike-Rental-Prediction---ML-Supervised>

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 rental or bike hire business rents out bicycles for a short period of time, usually for a few hours. Most rentals are provided by bike shops as a sideline to their main businesses of sales and service, but shops specialize in rentals. Most such systems allow a member to sign out a bike from any station for up to half an hour of free use, enough for most commuters to travel to their destination where they can drop the bike at any station bike sharing start after the first free half hour in order to encourage the user to return the bicycle at the end of each trip, and take another bicycle for the next trip.

The Seoul Bike Rental Data sets a view to predict the demand of rental bikes in urban cities to enhance the mobility of comfort. It is important to make the rental bike available at the right time for the public. The dataset contains the information from the year 2016 to 2017 with 8761 rows and 14 columns.

Our main objective is to predict the bike rental count required at each hour for the stable supply of rental bikes using regression techniques.

We perform the data wrangling on the raw data to get the useful data without NAN values and observe the summary statistics of the dataset. We prepared a dataset with feature engineering and feature scaling and also, dropped out the unnecessary columns. In the analysis, we checked the outliers and tried to overcome it by using outlier treatment using imputation of mean. The data was normalized using the standard deviation of the features.

Moreover, VIF factor is evaluated to cross check the multicollinearity in the data set. Also, One hot coding is used to create the new dataframe with dummy variables.

Reviewing all the data preprocessing, the model is fitted with splitting the data into Train and Test set in the ratio 70:30 respectively. Standard Scalar technique is used to scale the data which help to sustain the model prediction rate. Linear Regression, Lasso Regression, Decision Tree regressors, Random Forest Regression and Gradient Boosting Regressors are used to train the model. Cross validation techniques is applied for all fitted models.

The metrics of the model are accuracy , R-squared, MSE, RMSE. We evaluate the Gradient Boosting Regressors with Hyperparameter tuning which in returns gave the result as best fitted model for Seoul Bike Rental data set.

Drive Link:-

https://drive.google.com/drive/folders/192kQl8LcJOhtO9Ppc_Gth1syhTfN_69V