

Capstone Project Submission

Instructions:

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

Team Member's Name, Email and Contribution:

1). Meenu Benny

E-mail: meenubenny555@gmail.com

- Data Visualization
- Data Analysis
- Linear Regression
- Ridge Regression
- Lasso Regression
- Random Forest Regression

2). T. Ompriya Subudhi

E-mail: ompriyasubudhi2000@gmail.com

- Data Analysis
- ElasticNet Regularization Model
- Cross Validation on ElasticNet Regularization
- Polynomial Regression
- Technical Documentation
- Project summary Template

3). Thasniya Mansoor

E-mail: thasniyamansoor91@gmail.com

- Data Visualization
- Data Analysis
- Decision Tree Model
- PPT Presentation

Problem Statement:

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

The dataset contains weather information (Temperature, Humidity, Windspeed, Visibility, Dewpoint, Solar radiation, Snowfall, Rainfall), the number of bikes rented per hour and date information.

The main objective of this project is to deal with the data provided and to find out how the dependent attribute varies with respect to independent attribute also to find out which model is best fit to our given dataset.

Conclusion:

So here we come at the end of our project which is Seoul Bike Sharing Demand Prediction. Let's take a short recap of what we have done. We have found our dataset info where we have found there are 8760 rows and 14 columns with 13 features as independent and one as dependent according to our problem statement that is Rented bike count on which we have done our prediction. We didn't find any null values or duplicates on the given dataset.

Then we go for exploratory data analysis, in this section, we add some column and drop unnecessary column as per analysis of data like column 'Date'. In addition to this we divided the dataset into numerical and categorical feature. Then we analyze the dataset by plotting graph for each independent feature with dependent feature

From this we are concluded that Less demand of rented bike counts on winter season, Slightly Higher demand during Nonholidays, Almost no demand on non functioning days.

Now here comes the feature engineering, in which we got to know that which feature gives more or less importance to dependent feature. Also in this part of study we got some conclusions. They are listed below: The demand for rented bike from month 5 to 10 is higher than other months and these months fall in summer, the demand for rented bike is less on winter seasons, highest number of bike have rented in summer season, almost equal percentage of bike rented in spring and autumn, slightly, higher demand during Nonholidays, high increase of bikes rented between 8:00 am to 9:00 pm means people prefer rented bike during peak hours. We can clearly see that the demand peaks at 8:00 AM and 6:00 PM, so we can say that there is a very high demand during the opening and closing hours of the office

Using box plot we found out there are some outliers in rented bike count which one is the dependent feature, then by using square root transformation we free it from outliers.

In correlation analysis, we got that there is high correlation between Temperature and Dew Point Temperature.

Last but not the least we go for model training in which we split the dataset into train dataset (for hyperparameter tuning) and test data set (for model evaluation) in proportion of 75% train dataset and 25% test dataset.

We have performed 8 models using these train and test data set out of which we concluded that Random Forest Regressor Model is best fit model to our dataset with mean squared error on test dataset is 14.98 and mean absolute error is 2.75 and r2score is 0.9.

16. References

- GeeksforGeeks
- Kaggle
- Analytics Vidya
- Statistics

Please paste the GitHub Repo link.

Github Link:-

<https://github.com/OmpriyaSubudhi/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. (200-400 words)