

# Capstone Project Submission

## Instructions:

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

### **Team Member's Name, Email and Contribution:**

---

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.

### **Problem statement**

We are tasked with predicting the number of bikes rented each hour so as to make a n approximate estimation of the number of bikes to be made available to the public given a particular hour of the day.

---

The dataset contains bike usage patterns with the weather data to forecast bike rental demand. The data set consists of hourly rental data spanning two years.

Approach:

Null values Treatment and Outliers: Dataset contains a no null values to disturb the accuracy, but outliers are present which can disturb the accuracy. So again, we use z-score to remove outliers.

Exploratory data analysis:

After loading and reading the dataset in notebook, we performed EDA. Comparing target variable which is bike rentals counts with other independent variables. This process helped us figuring out various aspects and relationships among the target and the independent variables and also, we observed the distribution of variables. It gave us a better idea that how feature behaves with the target variable.

One hot encoding:

In this dataset some categorical variables:

1. Seasons
  2. Holiday
  3. Function Day,
- we change it with numerical database.

Correlation Analysis:

We plot the heatmap to find the correlation between both dependent variable and independent variables. From the heatmap we observed that :

Temperatures are highly correlated. There is a positive correlation between bike rentals counts and temperature.

We observed a correlation between bike rentals counts and humidity. The more the humidity, the less people prefer to rental bikes.

Bike rentals counts has a weak dependence on wind speed.







Same observation as above where also given by the regression plot as above.

Train test split:

In the train test split we take two variables i.e., X and Y where X contain all the independent variables and Y contain dependent variable. Here the independent variable is bike rentals counts and dependent variables is affecting the bike rentals counts like temperature, weather, seasons etc.

Modeling:

We performed several modeling over the data

-  Linear regression
-  Polynomial regression
-  Decision tree regression without GridSearchCV
-  Decision tree regression with GridSearchCV
-  Random forest regression with Randomized SearchCV
-  GBoost regression with Randomized SearchCV

Conclusion

- Linear Regression and Polynomial Regression performed bad and gave an R2 score of 0.65 and .76 respectively for test dataset.
- Decision tree Regression without gridsearchcv and Decision tree regression without GridSearchCV performed bad and gave an R2 score of 0.79 and .83 respectively for test dataset seen to be over fitting.
- Random forest Regressor with Randomized SearchCV and GB Regressor gives the highest R2 score of 89% and 92% respectively for test dataset.
- Finally, we can say that Ensemble model performed best out of all the models and No overfitting is seen, as we can see the models are performing well with the test data with good results.

Github Link: <https://github.com/oweskhan77/Bike-Sharing-Demand-Prediction>

**Name: - Owes khan**

**Email: - oweskhan77@gmail.com**

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

