BIKE SHARING DEMAND PREDICTION

(Technical Documentation)

BY – AYUSH SHARMA

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.

DATA OVERVIEW

The dataset provided consisted of 8760 rows and 14 columns

we are given the following columns in our data

1. Date : year-month-day
2. Rented Bike count - Count of bikes rented at each hour
3. Hour - Hour of the day
4. Temperature-Temperature in Celsius
5. Humidity - %
6. Wind Speed - m/s

7. Visibility - 10m

8. Dew point temperature - Celsius

9. Solar radiation - MJ/m2

10. Rainfall – mm

11. Snowfall – cm

12. Seasons - Winter, Spring, Summer, Autumn

13. Holiday - Holiday/No holiday

14. Functional Day - No(Non Functional Hours), Yes(Functional hours)

Performing EDA (exploratory data analysis)

A. Exploring head and tail of the data to get insights on the given data.

B. Looking for null values and removing them if it affects the performance of the model.

C. Converting the data into appropriate data types to create a regression model.

D. Creating more columns in our dataset which would be helpful for creating model .

E. Plotting various plots for columns Season , Temperature , Holiday , Month ,Hour with respect to Rented Bikes Count

F. Encoding the string type data to better fit our regression model.

G. Extracting correlation heatmap of variables.

MODELS USED

- **LINEAR REGRESSION -:**

  We first used linear Regression model and fit it on Training set the training accuracy of linear Regression model came out to be 77%. Then we test the model on test set and evaluated the model performance by using evaluation metrics
  MSE – 34.15
  RMSE – 5.84
  R2 – 0.783
  Adjusted R2- 0.773

- **LASSO REGRESSION -:**

- **Next we used LASSO Regression model and fit it on Training set the training accuracy of LASSO Regression model came out to be 77%.**
- **Then we test the model on test set and evaluated the model performance by using evaluation metrics**
- **MSE – 34.15**
- **RMSE – 5.84**
- **R2 – 0.783**

- **Adjusted R2- 0.773**

- ## RIDGE REGRESSION -:

  Then we used RIDGE Regression model and fit it on Training set the training accuracy of RIDGE Regression model came out to be 77%.
  Then we test the model on test set and evaluated the model performance by using evaluation metrics
  MSE – 34.35
  RMSE – 5.86
  R2 – 0.781
  Adjusted R2- 0.771

- ## ELASTICNET-:

  Then we used ElasticNet model and fit it on Training set the training accuracy of ElasticNet model came out to be 62%.
  Then we test the model on test set and evaluated the model performance by using evaluation metrics
  MSE – 59.93
  RMSE – 7.74
  R2 – 0.61
  Adjusted R2- 0.602

- ## DECISION TREE -:

  Then we used Decision Tree model and fit it on Training set the training accuracy of Decision Tree model came out to be 44%.
  Then we test the model on test set and evaluated the model performance by using evaluation metrics
  MSE – 100.92

  RMSE – 10.04

R2 – 0.359
Adjusted R2- 0.329

- **RANDOM FOREST-:**

Then we used Random Forest model and fit it on Training set the training accuracy of Random Forest model came out to be 98%.
Then we test the model on test set and evaluated the model performance by using evaluation metrics
MSE – 17.37
RMSE – 4.16
R2 – 0.889
Adjusted R2- 0.884

conclusion

In the conclusion I want say that from the beginning we did data inspection and also cleaned the data according to our needs. After that we perform EDA and were able to draw relevant conclusions from the given data and then we trained our model on linear regression and other models.

Out of all models used, with Random Forest model we were able to get the r2-score of 0.88. The model which

performed poorly was Decision tree model with r2-score of 0.35.