ML Regression Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

1. **Abhishek Mishra (Abhishekmishra@gmail.com):**

Exploratory Data Analysis (EDA)

Data frame description
Data frame shape
Understanding label column Rented Bike Count.
Relationship b/w Holiday Seasons and Rented Bike Count
Relationship b/w and bike count.

Feature Engineering

Feature engineering on Hour column
Feature encoding And creation
Features Selections Using Correlation plot
Features Selections Using Variation Inflation(VIF).

Data Pre-processing

Data standrazion
.

Model Implementation

Gradient Boosting Regression
Adaboost Boost Regression
Linear Regression
Lasso with hyper parameter tuning

Creating Data Frame of all Evaluation Matrix with respect of models

Conclusion from Model Training

2. **Kurva Mallesh (kurvamallesh36@gmail.com):**

Exploratory Data Analysis (EDA)

Data frame description
Data frame shape
Relationship b/w seasons humidity and bike count
Relationship b/w Rainfall humidity and bike count
Relationship b/w Snowfall (cm) humidity and bike count

Feature incoding And creation

Features Selections Using Correlation plot
Features Selections Using Variacian Inflation(VIF)

Model Implementation

Ridge with hyperparameter tuning
Polynomial
Using Random Forest Regressor

Creating Data Frame of all Evaluation Matrix with respect of models

Conclusion from Model Training

4. **Arunesh Mishra(Arunesh12mishra@gmail.com):**

Exploratory Data Analysis (EDA)

Data frame description

Data frame shape

Relation of Hour, Rented Bike Count for different Seasons

Relation of Wind speed Rented Bike Count for different Seasons

Feature Engineering

Encoding on Date Column

Creating new feature Week day and weekend

Model Implementation

XGBoost Regression

KNN Regression

Decision Tree

Creating Data Frame of all Evaluation Matrix with respect of models

Conclusion from Model Training

GithubRepository link:

Abhishek Mishra- https://github.com/abhishekmishra-bareilly/ML-Regression-Capistone-project

Kurva Mallesh - https://github.com/kurvamallesh/-ML-Regression-Capstone-project

Arunesh Mishra - https://github.com/kajuun/ML-regression-project

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

Bike sharing market is growing all over the world these years. It is meaningful to do some analysis about it because more and more companies are getting into this business. Since bike sharing is a recent phenomenon, not that many related analyses have been done upon this topic. However, more and more bike sharing companies have started to realize the importance of data driven decision making. One of the major aspects that can be addressed by data analysis is to predict the demand of bikes on any given day. Knowing the demand would help us in creating a better supply and subsequently reduce the gap between supply and demand. Our EDA can make us understand data which variable is very important and check how every variable connected with dependent variable. We make some models to predict the label column based on feathers

After the data wrangling step, we performed EDA by comparing different parameters which are involved in the dataset. EDA helps us to find the different relations among the parameters. It involves the visualization of the data by comparing the different parameters to find out the best among all. In the process of understanding the data we found that Rented Bike count is the label column and rest all are feathers we did some analysis. We explain each one by one.

Seasons are making huge impact on Rented Bike Count. There was very high count in summer (2283234) and very low count in winter (487169) season. We know that there is high Rented Bike Count in summer but during Holiday people like to book the bikes in autumn season more. So, we can say autumn season is best for Holidays. We found there are some rows that have Humidity as 0 so it is not possible so we need to check this. The Temperature (°C) Less than - 10 making huge impacting on Rented Bike Count. We suggest the company to give high preference to 8pm. Because at that time rented Bike Count is very high. We can say that there is not much bike ranted count after Wind speed 5(m/s) We can say that there are very less Rented bike count when Rainfall is greater than 5(mm). Also, we can see Snowfall directly impacting on Rented Bike Count. According to bar plot we can say that Rented Bike Count is very high in June month.

Encoding on date column, Split month from the date. Creating new feature week day and weekend. Do some feature engineering on hour column and make them as evening, morning, noon, night so the model can easily understand. Used correlation plot and variance inflation factor (VIF) for features selections. In data pre-processing we used power transformer, minimum, maximum scalar for standardize the data.