

Capstone Project Submission

Instructions:

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

Name:- Anurag Taiskar
Email:- art.anu17@gmail.com
Contribution:- Individual

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

The Bike Sharing Demand dataset consist of 8760 observation with 14 features and it is a regression problem because its target column is containing numerical quantity and the main goal is to predict the number of bike demand on the hourly basis.

After loading the dataset, first performed data preprocessing and checking data types, missing values, duplicate values and data description. In this dataset there are neither null values nor duplicate values. After that changed the date type to Date Time which was initially a string object.

After that Exploratory Data Analysis is performed to obtain the insights of our dependent variable Rented Bike Counts. Various graphs are constructed to comparing the Rented Bike Count column with other columns. It contain Univariate Analysis, Bivariate Analysis and Multivariate Analysis . Bike rental count is higher during weekdays than weekend days. The rental bike counts are at its peak at 8 AM in the morning and 6pm in the evening. Highest rental bike count is during Autumn and summer seasons and the lowest in winter season.

After that feature engineering comes in to remove multicollinearity. square root method to normalize the target variable because the skewness is moderate. For scaling independent features, MinMaxScaling is used. Pandas dummies variable is used for encoding the categorical features and map is used to mapping other categorical variable

After that modeling part begins and 7 regression algorithms is used which is, Linear Regression, Ridge Regression, Lasso Regression, Decision Tree Regressor, Random Forest Regressor and Xgbregressor. After fitting the models and evaluating metrics (MAE, MSE, RMSE, R Square) and also hyperparameter tuning. In this observation Random Forest Regressor gives accuracy is 98.79 on training data and 91.05 on test data. So Random Forest Regressor model is the best for predicting the bike rental count on an hourly basis.