Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

Team Member's Role:-

- **Ajinkya Dakhale**

  Email- Ajinkya.dakhale2408@gmail.com
  - Bivariate Analysis
  - VIF
  - linear Regression
  - Ridge Regression
- **Harshjyot Singh**

  Email- hs9158695878@gmail.com
  - Multivariate Analysis
  - Data wrangling
  - Feature Engineering
  - Lasso Regression
- **Suvir Kapse**

  Email- suvirkapse@gmail.com
  - Data understanding
  - Univariate analysis
  - Decision Tress
  - Random forest
  - Hyperparameter tuning

Github Link:-https://github.com/Ajinkya-dak/Bike-Sharing-Demand-Prediction

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.

First, we started with data wrangling, checking all the null values in our data, and then for the outliers. All the independent columns in our data set were in ranges, and for dependent variables we performed a squared root transformation to remove the right skewness.

second step, involves exploratory data analysis(EDA) where we tried to dig insights from the data . It

included univariate ,bivariate and multivariate analysis in which we identified certain trends , relationships, correlation and found out features that had some impact on our dependent variable(Rented_Bike_Count). Bike count rent is highly correlated with 'Hour', which seems obvious. Demand for bike is mostly in morning (7 to 8) and in the evening (3 to 9) when people go to work and returns home. As the temperature increases the more people prefer the rented bike vice versa with the increase in snowfall the rented bike count decrease which clearly indicates the season and weather play role in the demand of rented bike. Using a heat map and VIF, feature selection was done on certain independent features that appear to be highly correlated to one another. Then, on categorical features such as seasons ,hour, month, Holiday_No Holiday, function_day_yes and weekend, dummy encoding was performed. This makes the data suitable for splitting into tests and training before importing them to machine learning algorithms.

We use four machine learning models: linear regression, Ridge and lasso regression, decision tree, and random forest on test and training data and evaluated them on the basis of performance metrics like mean absolute error (MAE), root mean square error (RMSE), and adjusted R squared. We found that decision tree and random forest gave the best results on both train and test data, with random forest's adjusted R-squared value of 0.92 being highest.