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:-

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

Please paste the GitHub Repo link.

Github Link:- https://github.com/SuvirKapse/Seoul-Bike-Sharing-Demand-Prediction/blob/main/Bike_Sharing_Demand_Prediction_Capstone_Project.ipynb

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

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. The data had variables such as date, hour, temperature, humidity, wind-speed, visibility, dew point temperature, solar radiation, rainfall, snowfall, seasons, holiday, functioning day and rented bike count.

Approach :

The problem statement was to build a machine learning model that could predict the rented bikes count required for an hour, given other variables.Explained below is our approach .

- Perform data preparation & cleaning. Explore the number of rows & columns, ranges of values etc.Then we checked for missing values or duplicate values.
- Work on EDA & visualization where we tried to dig insights from the data in hand. It included univariate ,bivariate and multivariate analysis in which we identified certain trends, relationships, correlation and found out the features that had some impact on our dependent variable.
- Perform Feature Engineering for Converting categorical variables into numeric form using one-hot encoding, label encoding etc.
- And lastly modeling where we try various machine learning algorithms on our split and standardized data. We tried different algorithms namely; Linear regression,Regularized linear regression (Ridge and Lasso), Randomforest and Decision tree . We checked their performance using different evaluation metrics to find the best model.

CONCLUSION :

This project taught and helped us understanding a lot about the use of machine learning models .The best score we got in **R2** metric was '**0.92**' in test set for the Random forrest model. In it the feature temperature accounts for the highest importance therefore people of Seoul are more likely to stay home during colder days than on warmer ones. The model performed well in this case but as the data is time dependent values of certain features will not always be consistent. Therefore, there will be scenarios where the model might not perform well. As Machine learning is an exponentially evolving field, we will have to be prepared for all contingencies and also keep checking our model from time to time.

We also made some observation through the project:-

- We noticed high demand of bikes on holidays or non working day.
- Maximum number of bikes is rented between 7 to 9 am and 5 to 7 pm in the evening which indicates office going customer rents the maximum number of bikes.
- Bike demand was least in monsoon and maximum during summer, autumn and spring because of beautiful weather.