Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
Suvendu Nayak (kanhasuvendu@gmail.com) **Contribution:** • Defining the problem statement • EDA and data visualization • Data preprocessing • Feature selection • Preparing Dataset for model • Applying model • Model validation and selection
Please paste the GitHub Repo link and drive link.
GitHub Link:-https://github.com/SuvOnGithub/Bike-Sharing-Demand-Prediction---Capstone-Project Drivelink:-https://drive.google.com/drive/folders/1z6cplyueZHrIR11a4gsEFfkPiLD_cT7E?usp=sharing

Please write a short summary of your Capstone project and its components.

Describe the problem statement, your approaches and your conclusions. (200400 words)

Introduction:

Today, bike-sharing systems are blooming across more than 1000 cities around the world, particularly in big or large cities like New York City, Paris, Washington DC, London, Beijing and Barcelona. To complete a short trip renting a bike is a faster way when compared to walking. Moreover, it is eco-friendly and comfortable compared to driving.

Due to global warming,continuous pollution and depletion of sources of energy. Many countries have been focused on using renewable energy which doesn't harm the environment and can be reused as well. South Korea is one of the countries which has adapted to it and their most used service is rented bikes in Seoul. But in order to avoid any difficulties such as waiting time it is necessary to have an estimate of future demand. Our goal here is to build a model that can predict bike sharing demand considering all the factors which have their effects.

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.

Approach:

Here first we imported a data set and performed EDA where we got valuable insights and further we Encoded the Categorical Columns, Feature scaling and fitting into the models. At first we tried with basic linear regression and also with Lasso regularization technique but soon realized we will need a much more complex model and so we then used a Decision tree Regressor, Random Forest Regressor and compared the results.

Conclusion:

The analysis is done with Seoul Bike data. Four regression techniques Linear Regression, Decision Tree, XG Boosting and Random Forest are used to predict the trip duration. This statistical data analysis shows interesting outcomes in prediction methods and also in an exploratory analysis.