Capstone Project Submission

Team Member's Name, Email and Contribution:		
Name	Email	Contribution
Sarthak Arora	Sarthak1611@gmail.com	Coding Summary Ppt making
Jay Nandasana	nandasanajay@gmail.com	ppt making coding Final reviewing
Madhavi Mali	madhavimali1996@gmail.com	Technical documentation coding

Please paste the GitHub Repo link.

Github Link:- https://github.com/Sarthak016/Bike_Sharing_Demand_Predcition_DD

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

Capstone Project Submission

Seoul Bike Sharing Demand Prediction

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. Then checked for any duplicated entries and then diagnosed the data. Then we filtered the data and performed EDA to get various insights of the data. We then found out highly correlated factors and removed one among them. After that we performed feature engineering. We encoded categorical data and tried to find out performance of different models using different types of scalers. 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, XGB Model, Random Forest Regressor and LightGBM and then we compared the results.

Conclusion:

We noted down several conclusion when we performed EDA. The various statistical data analysis showed interesting outcomes in prediction methods

The experimental results show that:

- Most numbers of Bikes were rented in summer, followed by autumn, spring, and winter. May-July is the peak Bike renting Season, and Dec-Feb is the least preferred month for bike renting.

- Majority of the client in the bike rental sector belongs to the Working class. This is evident from EDA analysis where bike demand is more on weekdays, working days in Seoul.

- Temperature of 20-30 Degrees, evening time 4 pm- 8 pm, Humidity between 40%-60% are the most favourable parameters where the Bike demand is at its peak.

- Temperature, Hour of the day, solar radiation, and Humidity are major driving factors for the Bike rent demand.

- When we completed the analysis we found out that these four regression techniques i.e ExtraTreeRegressor , GradientBoostingRegressor , XGBRegressor
  Light-GBM have the highest accuracy. We performed hyperparamter tuning on GradientBoostingRegessor and its R2 score increased to 91.7% from 87%.