Capstone Project Submission

Seoul Bike Sharing Demand Prediction

Contributor's Name, Email :
Contributor   • Gaurav Yogeshwar   • email - yogeshwar.gaurav0@gmail.com

GitHub Repo link.
Github Link:- https://github.com/Gaurav2912/Bike-Sharing-Demand-Prediction.git

Short summary of your Capstone project and its components.

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, XGB Model, Random Forest Regressor and LightGBM compared the results.

Summary
The following topics were covered :

- Downloading a real-world dataset
- Looking for null and missing values
- Try to get some insights from data
- Correlation plot
- Data Cleaning, Preprocessing and Feature engineering
- Preparing a dataset for training
- Training and interpretation with different models
- Overfitting, hyperparameter tuning & regularization
- Predict and calculate sum evaluation metrics for model
- Model Persistence saving and loading the model
- Making predictions on single inputs and explaining it with the help of SHAP.

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.
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 clients in the bike rental sector belong 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 favorable 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.

Feature and Labels had a weak linear relationship; hence the prediction from the linear model was very low. Best predictions are obtained with a LightGBM model with an R2 Score of 0.9523 and RMSE of 156.26