Bike Sharing Demand Prediction

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.

Data Description

season : season (1:spring, 2:summer, 3:fall, 4:winter)

yr : year (0: 2018, 1:2019)

mnth : month (1 to 12)

holiday : weather day is a holiday or not

weekday : day of the week

workingday : if day is neither weekend nor holiday is 1, otherwise is 0

weathersit :

    - 1: Clear, Few clouds, Partly cloudy, Partly cloudy

    - 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist

    - 3: Light Snow+ Thunderstorm + Scattered clouds, Light Rain + Scattered clouds

    - 4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog

EDA

- As compared to season the demands seem to be very down as compared to other season of the year
- There has been a significant grow in demand of bikes from year 2018 to the next year
- Demand for bikes was highest from Jun through Oct months in a whole year.
- Weekday doesn't not seem to be a good driver for understanding any patters for demand in bikes hence is not an important feature.
- In weathersit category C the demand has been very low due to some reason due to thunderstorm and rainy weather.
- In case of holiday we see that the data distribution is more in case of not an holiday however mot much change in case of demand of bikes.
- the data distribution is less in case of the day is neither weekend or holiday but no significant change in demand.

Insights from Correlation

Correlation from target column to independent columns
- There is a positive correlation between the cnt and the atemp column.

Correlation between independent columns
- Workingday has av -ve correlation with holiday feature.

While building the model we need to make sure to treat the multicollinearity features correctly.

Steps to build a model

Scaling the numerical features

Removing multicollinearity

Validating the features which help to explain the variability with the target variable

Validate the accuracy of the model with Adjusted rsquare/Rsquare value.

Here we have used linear regression and xgboost regressor model for reference.

Recommendation

Hyperparameter tuning could be one of the way to increase the accuracy

We could also use a batching mechanism to send data for training which will create a better model performance.

Additionally we could mode into difficult or more better models like xgboosting or stacking regressor.