

TANMOY BAIDYA

B.Tech, Part-IV, IIT(BHU) Varanasi

Analytics Vidhya Hiring Hackathon

1. Data Preprocessing

1.1 Datetime Object: Year, Months, Quarter, Week , Days, Hour and Day of year are extracted from given datetime object. And then using these newly added features Days difference is calculated from initial date.

1.2 Temperature : As temperature showing seasonality with time, 3 new categorical variable is added to the dataset based on temperature range.Later, those categories are dummy encoded. And outliers are marked with new feature called 'temp_outlier'.

1.3 Var1: As like temperature, var1 is also categorised into 4 new categories and same way it is dummy encoded . Var1 outliers are marked in 'var1_outlier' feature.

1.4 Var2: Var2 feature is dummy encoded.

1.5 Winspeed : Windspeed outliers are marked in 'windspeed_outlier' feature.

1.6 Overall_Outlier: Those rows having combined (temp_outlier+var1_outlier+windspeed_outlier) value 2 or more than 2 are marked as overall outlier.

2. Final Model

Algorithm Used: XGBoost

Parameters: learning_rate=0.1,n_estimators=500,max_depth=10,colsample_bytree=0.5