Kaggle 2.

Team Name: BulatAleev

Current position:5

MAE result: 92.94030 in 3 Entries made

Methods used:

Firstly I filled the missed data cells with the mean values of the row according to the column.

The first best result showed knn classifier where the parameters of p =2 and n\_neighbours = 2 the result of mae was 154. However this method didn’t show enough result due to the peculiarities of knn method which would be more suitable for clustering.

KNeighborsClassifier(algorithm='auto', leaf\_size=60, metric='minkowski',metric\_params=None, n\_jobs=1, n\_neighbors=1, p=2, weights='distance') where the effective metric was Euclidian .

LinearRegression(copy\_X=True, fit\_intercept=False, n\_jobs=-1, normalize=True) which showed 364.641 mae result. Even after using the scaler that increased the value but not significantly.

I also used Voting classifier as an ensemble to get the better mae result, where as a result I pointed out the gradient boosting from sklearn library. Despite it took longer time to compute the predictions, the principle of gradient boosting helped to achieve better results where the trees built corrected errors made by the previous trained trees.

Features used with weights:

date,hour,holiday,workingday,weekday,weather\_type,temp,ftemp,humidity,wind\_speed

0.197 \* X0 + 0.275 \* X1 + 0.01 \* X2 + 0.042 \* X3 + 0.056 \* X4 + 0.029 \* X5 + 0.106 \* X6 + 0.066 \* X7 + 0.129 \* X8 + 0.091 \* X9 where x0 is date and hour is x1 and others are respectively.

{'n\_estimators': 500, 'max\_depth': 6,

'learning\_rate': 0.1, 'loss': 'huber','alpha':0.85} parameters used and found by grid search.