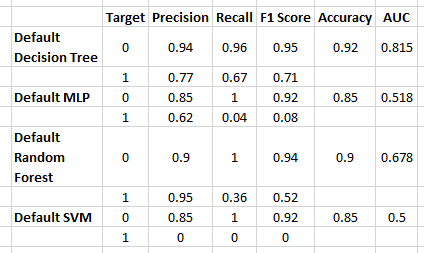
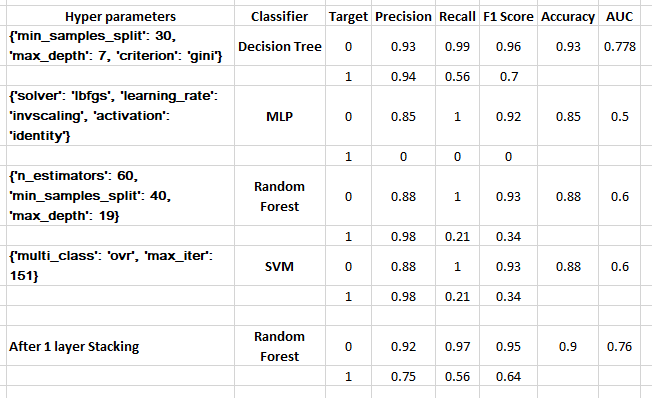
Assignment – 6 – Mahindra Venkat Lukka

1. Data is imbalanced
2. Data contains 6 Nan values and data is cleaned
3. Performed one hot encoding
4. Performed Default decision tree, default random forest, default SVM, default MLP and with hyper parameter tuning.
5. As the data is imbalanced, I performed smote with ratio = 0.5
6. After observing the results, I considered MLP, Random Forest and Decision tree classifiers with hyper parameters for 1st layer of stacking. Accuracy score achieved = **0.902**
7. After stacking, as given in the problem, I used random forest classifier on the stacked model.
8. Results from the above classifiers are shown below,





1. Surprisingly, decision tree got very high accuracy and AUC score.
2. So, we can see that our results are almost same for default and hyper parameter tuned classifiers due to imbalanced data.
3. But, after considering this and performing SMOTE and stacking, we can see the increase of accuracy, and improvement in Precision, Recall, F1 scores for both 0 and 1 Target values. This is mainly because of SMOTE.
4. So, SMOTE and 1st layer of stacking helped a lot in this problem. To get even more accuracy we can even try 2nd layer of stacking which reduces the probability of error.