Insurance_Claims_Fraud_Detection

Data Processing:

Divided the .csv file into the train, test sets and performed different data cleaning techniques to trace and filter the outliers in the dataset.

Training Models and Methodology:

Here we have test our model with different list of ML-Algorithms like

1. Support Vector Classifier (SVC)
2. K-Nearest Neighbors (KNN)
3. Decision Tree Classifier (DTC)
4. Random Forest Classifier (RFC)
5. Ada Boost Classifier (ABC)
6. Gradient Boosting Classifier (GBC)
7. Stochastic Gradient Boosting (SGB)
8. XG Boost Classifier (XGBC)
9. Cat Boost Classifier (CBC)
10. Extra Trees Classifier (ETC)
11. LGBM Classifier (LGBMC)
12. Voting Classifier (VC)

Out of which we have picked Top Three (3) best performing algorithms. They are

1. Ada Boost - 82%,
2. XG Boost - 81%,
3. Random Forest – 77.6%

Hyperparameters:

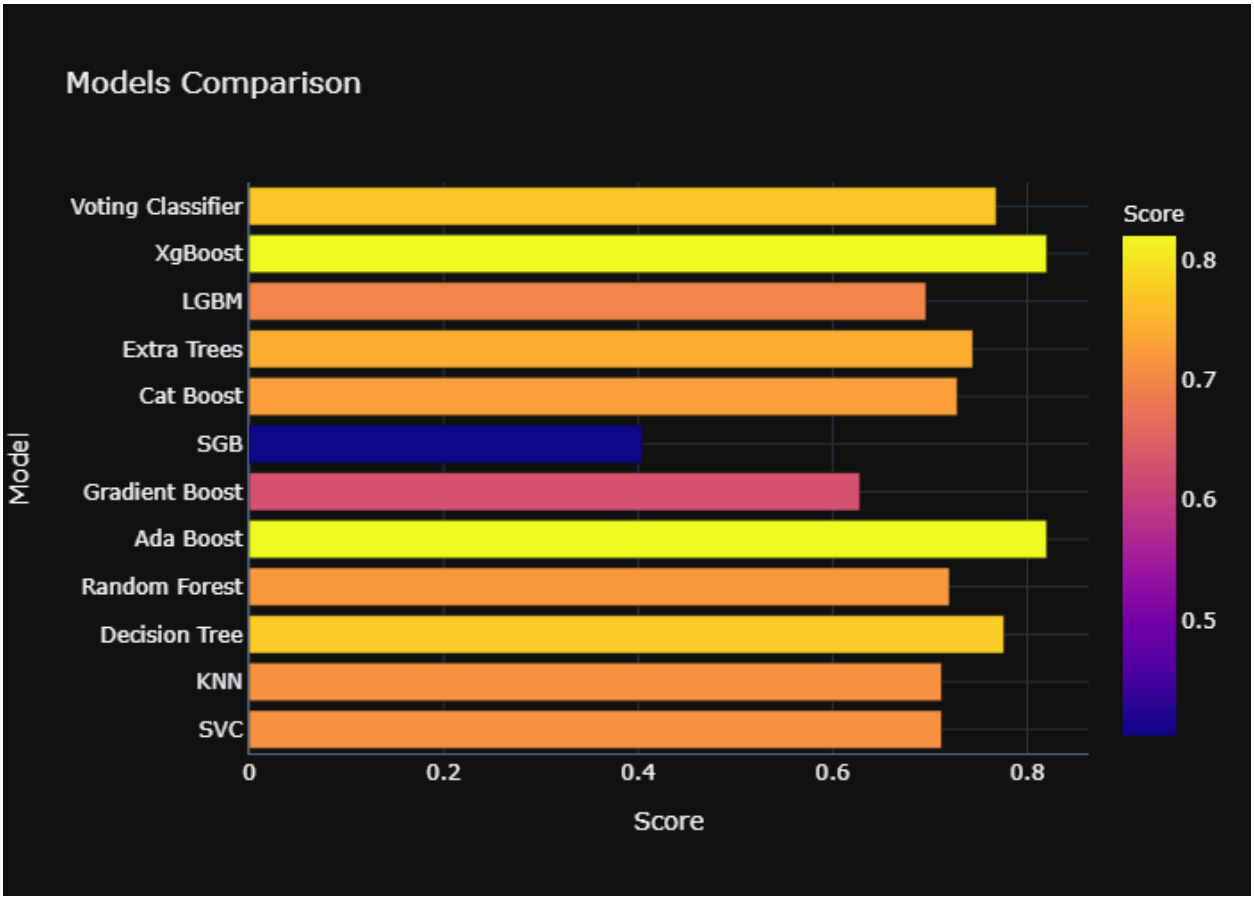
We have used Grid search CV technique for hyper parameter tuning with different values to find the best parameter and achieve high accuracy.

Analysis_Report

1. Ada Boost:

{'algorithm': 'SAMME', 'learning_rate': 0.01, 'n_estimators': 50}

NOTE: Optimal value of each parameter has been selected by trial and error, tried a lot of different values. Accuracy: 0.82, F1 Score: 0.87, Precision Score: 0.89, Recall Score: 0.85.



Note: From the above list of ML-Models Ada Boost gives best accuracy – 82%