Results

Total Rows: 356255

Logistic Regression

|  |  |  |  |
| --- | --- | --- | --- |
| Comments | ROC/ Accuracy | Hyperparameters | Algorithm |
| model.fit(train\_x,train\_y,class\_weight = 'balanced') | 65% | Class\_weight = ‘balanced’  Stratified split  One hot encoding | Logistic Regression |
| Modifed the feature set little bit | 67%/ 66 | Class\_weight = ‘balanced’  Stratified split  One hot encoding | Logistic Regression |
| Same feature set |  |  | SVM |
| Used imblearn to balance the dataset before applying the algorithms  Split the data instead of using stratified sampling | 57/56 |  | Logistic Regression |
| K-Stratified Shuffle  Class-weight = “Balanced” | .91/.76 |  | Light gbm |
| 100 estimators  Class-weight = “Balanced” | /.74 |  | Random Forest |
| Neighbors = 6  Algorithm: ball\_tree |  |  | KNN |
| With scaling all above algorithms |  |  | Light gbm |
|  |  |  | Xgboost |
|  |  |  | Logistic regression |
| Ensembling With stacking | /.777 |  |  |
|  |  |  |  |

1. Flag\_Document features not at all useful