* All Binary variables such as Gender, Married, Dependents etc are encoded with OneHot encoder with dummy value trap
* Property\_Area variable is encoded using Ordinal Label Encoder
* Iterative Imputer is used for handling Null Values
  + For Categorical Variables, RandomForestClassifier is used
  + For Continuous variables, RandomForestRegressor is used
* Experimented the data with different algorithms, below are the outputs

|  |  |  |  |
| --- | --- | --- | --- |
| **Algorithm** | **Accuracy** | **Precision** | **Recall** |
| AdaBoostClassifier(random\_state=11) | 0.76 | 0.76 | 0.91 |
| GradientBoostingClassifier(random\_state=11) | 0.77 | 0.77 | 0.93 |
| RandomForestClassifier(random\_state=11) | 0.78 | 0.78 | 0.94 |
| BalancedRandomForestClassifier(random\_state=11) | 0.79 | 0.83 | 0.86 |
| KNeighborsClassifier(n\_neighbors=10) | 0.58 | 0.65 | 0.78 |
| LogisticRegression(random\_state=11) | 0.79 | 0.77 | 0.98 |
| SVC(kernel='linear', random\_state=11) | 0.77 | 0.75 | 0.98 |
| SVC(kernel='poly', random\_state=11) | 0.65 | 0.66 | 0.99 |
| SVC(random\_state=11) | 0.64 | 0.65 | 0.98 |
| SVC(kernel='sigmoid', random\_state=11) | 0.59 | 0.67 | 0.74 |
| XGBClassifier | 0.76 | 0.79 | 0.88 |
| BaggingClassifier(estimator=LogisticRegression()) | 0.79 | 0.77 | 0.98 |

* BalancedRandomForestClassifier is giving good accuracy of 0.79, hence chose that as the final model
* Completed the same feature engineering steps for the test data as well and outputs are available in output.csv file.