CS 6375

ASSIGNMENT -  **Scikit Learn Lab : Parameter Tuning of Classification Models using GridSearch**

Names of students in your group:

Kavin Kuppusamy (KXK190026)

Number of free late days used: \_\_\_\_\_0\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
Note: You are allowed a **total** of 4 free late days for the **entire semester**. You can use at most 2 for each assignment. After that, there will be a penalty of 10% for each late day.

Link to Google Colab :

https://colab.research.google.com/drive/1HifO90oERZlz1RbaPO9WqjNzCXkDJeUN

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Algorithm** | **Best Parameters** | **Avg Precision** | **Avg Recall** | **Avg F1** | **Accuracy\_Score** |
| **Decision Tree** | {'max\_depth': 1000, 'max\_features': 'log2', 'min\_samples\_leaf': 1, 'min\_samples\_split': 10} | 0.94 | 0.94 | 0.94 | 0.93 |
| **Neural Networks** | {'activation': 'tanh', 'alpha': 0.0005, 'hidden\_layer\_sizes': (6,), 'max\_iter': 500} | 0.97 | 0.97 | 0.97 | 0.97 |
| **Support Vector Machine** | {'C': 100, 'gamma': 0.001, 'kernel': 'rbf'} | 0.96 | 0.96 | 0.96 | 0.96 |
| **Naïve Bayes** | {'priors': array([0.1, 0.9])} | 0.90 | 0.90 | 0.90 | 0.89 |
| **Logistic Regression** | {'fit\_intercept': True, 'max\_iter': 10, 'penalty': 'l1', 'tol': 1e-05} | 0.96 | 0.96 | 0.96 | 0.95 |
| **K-Nearest Neighbors** | {'algorithm': 'ball\_tree', 'n\_neighbors': 10, 'p': 1, 'weights': 'uniform'} | 0.96 | 0.96 | 0.96 | 0.95 |
| **Bagging** | {'max\_features': 0.5, 'max\_samples': 0.5, 'n\_estimators': 50, 'random\_state': None} | 0.96 | 0.96 | 0.96 | 0.95 |
| **Random Forest** | {'criterion': 'gini', 'max\_depth': 100, 'max\_features': 0.5, 'n\_estimators': 100} | 0.97 | 0.97 | 0.97 | 0.97 |
| **Ada Boost** | {'algorithm': 'SAMME', 'learning\_rate': 0.8, 'n\_estimators': 200, 'random\_state': None} | 0.98 | 0.98 | 0.98 | 0.98 |
| **Gradient Boosting** | {'learning\_rate': 0.5, 'loss': 'exponential', 'max\_depth': 3, 'max\_features': 'log2', 'n\_estimators': 100} | 0.96 | 0.96 | 0.96 | 0.95 |
| **XGBClassifier** | {'booster': 'gbtree', 'learning\_rate': 0.2, 'max\_delta\_step': 0, 'min\_child\_weight': 1, 'n\_estimators': 250} | 0.98 | 0.98 | 0.98 | 0.98 |

For the Breast Cancer dataset, the classifiers that gave highest accuracy among the eleven-classifier chosen are XGB and Ada Boost because of the fine tuning of the parameters that have been passed to the classifier.

The low accuracy from the other classifier are perhaps due the parameters considered for the training of the model or the parameters should have been fined tuned.