Multi-class

Test-train-split: 90: 10

Metric: balanced accuracy

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Rank | Classifier | Training | Test | Hyperparameters |
| 1 | Gaussian Naïve Bayes | 0.742 | 0.751 | Default |
| 2 | Random Forest Classifier | 0.798 | 0.747 | n\_estimators: 69, max\_depth: 8 |
| 3 | Extra Trees Classifier | 0.871 | 0.745 | n\_estimators: 20, max\_depth: 1 |
| 4 | GradientBoostingClassifier | 0.957 | 0.738 | learning\_rate: 0.1, n\_estimators: 180, max\_depth: 2, init: zero |
| 5 | QuadraticDiscriminantAnalysis | 0.751 | 0.738 | default |
| 6 | LabelSpreading | 0.735 | 0.725 | n\_neighbors: 10, alphas: 0.2 |
| 7 | SVC | 0.695 | 0.724 | C: 1, gamma: scale, shrinking: True, class\_weight: None, decision\_function\_shape: ovo |
| 8 | KNeighborsClassifier | 0.998 | 0.723 | n\_neighbors: 9, leaf\_sizes: 6 |
| 9 | GaussianProcessClassifier | 0.694 | 0.719 | default |
| 10 | LabelPropagation | 0.735 | 0.718 | n\_neighbors: 10 |
| 11 | LogisticRegressionCV | 0.675 | 0.707 | solver: saga, multi-class: auto |
| 12 | LogisticRegression | 0.665 | 0.705 | solver: saga, multi-class: auto |
| 13 | LinearDiscriminantAnalysis | 0.645 | 0.685 | solver: lsqr, shrinkage: auto |
| 14 | DecisionTreeClassifier | 0.682 | 0.681 | max\_depth: 7 |
| 15 | ExtraTreeClassifier | 0.931 | 0.673 | max\_depth: 15 |
| 16 | RadiusNeighborsClassifier | 0.998 | 0.656 | radius: 300, leaf\_size: 100 |
| 17 | LinearSVC | 0.638 | 0.644 | multi\_class: crammer\_singer |
| 18 | NearestCentroid | 0.629 | 0.631 | shrink\_threshold: 1 |
| 19 | PassiveAggressiveClassifier | 0.509 | 0.561 | C: 1, class\_weight: balanced |
| 20 | SGDClassifier | 0.380 | 0.416 | penalty: l2, learning\_rate: adaptive, class\_weight: None, eta0: 0.8 |
| 21 | RidgeClassifierCV | 0.355 | 0.352 | class\_weights: none |
| 22 | RidgeClassifier | 0.301 | 0.310 | alpha: 50000 |
| 23 | Perceptron | 0.250 | 0.246 | penalty: l2, class\_weight: balanced |

5-fold cross-validation

Metric: balanced accuracy

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Rank | Classifier | Training | Test | Hyperparameters |
| 1 | QuadraticDiscriminantAnalysis | 0.750 | 0.736 | default |
| 2 | GaussianNB | 0.746 | 0.735 | default |
| 3 | GradientBoostingClassifier | 0.998 | 0.719 | learning\_rate: 0.1, n\_estimators: 180, max\_depth: 2, init: zero |
| 4 | RandomForestClassifier | 0.998 | 0.712 | n\_estimators: 66, max\_depth: 10 |
| 5 | ExtraTreesClassifier | 0.998 | 0.711 | n\_estimators: 15, max\_depth: 10 |
| 6 | SVC | 0.750 | 0.706 | C: 20, gamma: auto, kernel: rbf, shrinking: False, class\_weight: None, decision\_function\_shape: ovo |
| 7 | LabelPropagation | 0.998 | 0.690 | n\_neighbors: 17 |
| 8 | LabelSpreading | 0.998 | 0.686 | n\_neighbors: 6, alphas: 0.2 |
| 9 | GaussianProcessClassifier | 0.697 | 0.678 | default |
| 10 | LogisticRegressionCV | 0.678 | 0.665 | solver: saga, multi-class: auto |
| 11 | LogisticRegression | 0.666 | 0.658 | solver: saga, multi-class: auto |
| 12 | RadiusNeighborsClassifier | 0.998 | 0.650 | radius: 300, leaf\_size: 100 |
| 13 | LinearDiscriminantAnalysis | 0.652 | 0.642 | solver: lsqr, shrinkage: auto |
| 14 | KNeighborsClassifier | 0.998 | 0.630 | n\_neighbors: 24, leaf\_sizes: 21 |
| 15 | ExtraTreeClassifier | 0.923 | 0.621 | max\_depth: 15 |
| 16 | DecisionTreeClassifier | 0.433 | 0.408 | max\_depth: 10 |
| 17 | PassiveAggressiveClassifier | 0.165 | 0.175 | C: 0.01, class\_weight: balanced |
| 18 | Perceptron | 0.166 | 0.170 | penalty: l2, class\_weight: None |
| 19 | SGDClassifier | 0.163 | 0.166 | penalty: elasticnet, learning\_rate: adaptive, class\_weight: balanced, eta0: 1.1 |

\*less classifiers because estimators without a method for predicting probabilities could not be classified.

Multilabel

Test-train-split: 90: 10

Metric: fbeta macro score (Either accuracy or the f1-score metric with the higher score was shortlisted)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Rank | Classifier | Training | Test | Hyperparameters |
| 1 | ExtraTreesClassifier | 0.319 | 0.927 | n\_estimators: 10, max\_depth: 25, min\_samples\_split: 2, min\_samples\_leaf: 1 |
| 2 | KNeighborsClassifier | 0.996 | 0.907 | n\_neighbors: 8, leaf\_sizes: 1 |
| 3 | RandomForestClassifier | 0.892 | 0.877 | n\_estimators: 300, max\_depth: 6, min\_samples\_split: 10, min\_samples\_leaf: 1 |
| 4 | ExtraTreeClassifier | 0.935 | 0.870 | max\_depth: 14 |
| 5 | DecisionTreeClassifier | 0.991 | 0.843 | max\_depth: 13 |
| 6 | RadiusNeighborsClassifier | 0.996 | 0.128 | radius: 300, leaf\_size: 1 |

5-fold cross-validation

Metric: f1 weighted score

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Rank | Classifier | Training | Test | Hyperparameters |
| 1 | RandomForestClassifier | 0.911 | 0.783 | n\_estimators: 300, max\_depth: 6, min\_samples\_split: 5, min\_samples\_leaf: 1 |
| 2 | ExtraTreeClassifier | 0.966 | 0.730 | max\_depth: 13 |
| 3 | KNeighborsClassifier | 0.996 | 0.723 | n\_neighbors: 6, leaf\_size: 6 |
| 4 | ExtraTreesClassifier | 0.993 | 0.723 | n\_estimators: 15, max\_depth: 15 |
| 5 | DecisionTreeClassifier | 0.988 | 0.714 | max\_depth: 12 |