**Appendix B.** Selected SML classifiers and hyperparameter

|  |  |
| --- | --- |
| Classifier | Parameters |
| Random Forest (RF) | n\_estimators=150, \*, criterion='gini', max\_depth=None, min\_samples\_split=2, min\_samples\_leaf=1, min\_weight\_fraction\_leaf=0.0, max\_features='sqrt', max\_leaf\_nodes=None, min\_impurity\_decrease=0.0, bootstrap=True, oob\_score=False, n\_jobs=None, random\_state=None, verbose=0, warm\_start=False, class\_weight=None, ccp\_alpha=0.0, max\_samples=None |
| Decision Tree (DT) | criterion='gini', splitter='best', max\_depth=None, min\_samples\_split=3, min\_samples\_leaf=1, min\_weight\_fraction\_leaf=0.0, max\_features=None, random\_state=None, max\_leaf\_nodes=None, min\_impurity\_decrease=0.0, class\_weight=None, ccp\_alpha=0.0 |
| K-Nearest Neighbors (KNN) | n\_neighbors=10, \*, weights='uniform', algorithm='auto', leaf\_size=30, p=2, metric='minkowski', metric\_params=None, n\_jobs=None |
| Gaussian Naive Bayes (GNB) | priors=None, var\_smoothing=1e-09 |
| MultinomialNB (MNB) | alpha=1.0, fit\_prior=True, class\_prior=None |
| ComplementNB (CNB) | alpha=1.0, fit\_prior=True, class\_prior=None, norm=False |
| SGDClassifier (SGDC) | loss='hinge', \*, penalty='l2', alpha=0.0001, l1\_ratio=0.15, fit\_intercept=True, max\_iter=1000, tol=0.001, shuffle=True, verbose=0, epsilon=0.1, n\_jobs=None, random\_state=None, learning\_rate='optimal', eta0=0.0, power\_t=0.5, early\_stopping=False, validation\_fraction=0.1, n\_iter\_no\_change=5, class\_weight=None, warm\_start=False, average=False |
| Bagging (BAG) | base\_estimator= KNeighborsClassifier, n\_estimators=10, \*, max\_samples=1.0, max\_features=1.0, bootstrap=True, bootstrap\_features=False, oob\_score=False, warm\_start=False, n\_jobs=None, random\_state=None, verbose=0 |
| ExtraTreesClassifier (ETC) | n\_estimators=100, \*, criterion='gini', max\_depth=None, min\_samples\_split=2, min\_samples\_leaf=1, min\_weight\_fraction\_leaf=0.0, max\_features='sqrt', max\_leaf\_nodes=None, min\_impurity\_decrease=0.0, bootstrap=False, oob\_score=False, n\_jobs=None, random\_state=None, verbose=0, warm\_start=False, class\_weight=None, ccp\_alpha=0.0, max\_samples=None |
| Adaboost (AB) | base\_estimator=None, \*, n\_estimators=50, learning\_rate=1.0, algorithm='SAMME.R', random\_state=None |
| GradientBoostingClassifier (GBC) | \*, loss='log\_loss', learning\_rate=0.1, n\_estimators=100, subsample=1.0, criterion='friedman\_mse', min\_samples\_split=2, min\_samples\_leaf=1, min\_weight\_fraction\_leaf=0.0, max\_depth=3, min\_impurity\_decrease=0.0, init=None, random\_state=None, max\_features=None, verbose=0, max\_leaf\_nodes=None, warm\_start=False, validation\_fraction=0.1, n\_iter\_no\_change=None, tol=0.0001, ccp\_alpha=0.0 |
| HistGradientBoostingClassifier (HGBS) | loss='log\_loss', \*, learning\_rate=0.1, max\_iter=100, max\_leaf\_nodes=31, max\_depth=None, min\_samples\_leaf=20, l2\_regularization=0.0, max\_bins=255, categorical\_features=None, monotonic\_cst=None, interaction\_cst=None, warm\_start=False, early\_stopping='auto', scoring='loss', validation\_fraction=0.1, n\_iter\_no\_change=10, tol=1e-07, verbose=0, random\_state=None, class\_weight=None |
| Quadratic Discriminant Analysis (QDA) | \*, priors=None, reg\_param=0.0, store\_covariance=False, tol=0.0001 |
| Logistic Regression (LR) | penalty='l2', \*, dual=False, tol=0.0001, C=1.0, fit\_intercept=True, intercept\_scaling=1, class\_weight=None, random\_state=None, solver='lbfgs', max\_iter=100, multi\_class='auto', verbose=0, warm\_start=False, n\_jobs=None, l1\_ratio=None |
| Multi-layer Perceptron (MLP) | hidden\_layer\_sizes=(100,), activation='relu', \*, solver='adam', alpha=0.0001, batch\_size='auto', learning\_rate='constant', learning\_rate\_init=0.001, power\_t=0.5, max\_iter=200, shuffle=True, random\_state=None, tol=0.0001, verbose=False, warm\_start=False, momentum=0.9, nesterovs\_momentum=True, early\_stopping=False, validation\_fraction=0.1, beta\_1=0.9, beta\_2=0.999, epsilon=1e-08, n\_iter\_no\_change=10, max\_fun=15000 |