**The Results**

1. **The Train-Test split**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **The best parameters** | **The best parameters** | **The accuracy** | **The recall** | **The ROC score** |
|  |  |  |  |  |
| The Random Forest | Best max\_depth: 5  Best max\_features: log2  Best n\_estimators: 1000 | Accuracy on training set: 0.994  Accuracy on test set: 1.000 | 1.00 | 0.9975490196078431 |
| The Gradient Boosted Trees | Best max\_depth: 5  Best learning\_rate: 0.001  Best n\_estimators: 1000 | Accuracy on training set: 1.000  Accuracy on test set: 1.000 | 1.00 | 0.9975490196078431 |
| The Logistic Regression | Best C: 10  Best penalty: l2 | Accuracy on training set: 0.995  Accuracy on test set: 0.986 | 1.00 | 0.9946107784431139 |
| The Logistic Regression unbalanced | Best C: 1.0  Best penalty: l2 | Accuracy on training set: 0.997  Accuracy on test set: 0.991 | 0.99 | 0.9924038589419623 |
| The SVM | Best C: 10  Best kernel: linear  Best gamma: 1e-05 | Accuracy on training set: 0.974  Accuracy on test set: 0.959 | 0.98 | 0.985948690853587 |

1. **The validation set split**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **The best parameters** | **The best parameters** | **The accuracy** | **The recall** | **The ROC score** |
| The KNN | Best number of neighbors found: 25 | Best score on validation set: 0.8331193838254172  Score on training/validation set: 0.8131621187800963  Score on test set: 0.7940327237728585 |  |  |
| The Random Forest | Best best\_max\_depth found: 20  Best best\_max\_features found: sqrt  Best best\_n\_estimators found: 100 | Score on training/validation set: 1.0  Score on test set: 0.9990375360923965 | 1.00 | 0.9975490196078431 |
| The Gradient Boosted Trees | Best best\_max\_depth found: 5  Best best\_learning\_rate found: 0.2  Best best\_n\_estimators found: 1000 | Best score on validation set: 0.9987163029525032  Score on training/validation set: 1.0  Score on test set: 0.9990375360923965 | 1.00 | 0.9975490196078431 |
| The Logistic Regression | Best best\_C found: 100  Best best\_penalty found: l1 | Best score on validation set: 0.993581514762516  Score on training/validation set: 0.9967897271268058  Score on test set: 0.9980750721847931 | 0.99 | 0.9950980392156863 |
| The Logistic Regression unbalanced | Best best\_C found: 100  Best best\_penalty found: l1 | Best score on validation set: 0.9974337040205303  Score on training/validation set: 0.9993580141236893  Score on test set: 0.9961489088575096 | 0.97 | 0.9863013698630136 |
| The SVM | Best best\_C found: 1.0  Best best\_kernel found: linear  Best best\_gamma found: 1e-05 | Best score on validation set: 0.9897304236200257  Score on training/validation set: 0.9913322632423756  Score on test set: 0.9932627526467758 | 1.00 | 0.9939562052365857 |

1. **The cross validation split**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **The best parameters** | **The best parameters** | **The accuracy** | **The recall** | **The ROC score** |
| The KNN | Best number of neighbors found: 39 | Best average score: 0.8044943820224718  Score on training/validation set: 0.8086677367576244  Score on test set: 0.8017324350336862 |  |  |
| The Random Forest | Best best\_max\_depth found: 5  Best best\_max\_features found: sqrt  Best best\_n\_estimators found: 100 | Best average score: 0.9987158908507224  Score on training/validation set: 0.9990369181380417  Score on test set: 0.9990375360923965 | 1.00 | 0.9975490196078431 |
| The Gradient Boosted Trees | Best best\_max\_depth found: 5  Best best\_learning\_rate found: 0.2  Best best\_n\_estimators found: 1000 | Best average score: 0.997752808988764  Score on training/validation set: 1.0  Score on test set: 0.9990375360923965 | 1.00 | 0.9975490196078431 |
| The Logistic Regression | Best best\_C found: 100  Best best\_penalty found: l1 | Best average score: 0.9961476725521671  Score on training/validation set: 0.9974317817014446  Score on test set: 0.9980750721847931 | 0.99 | 0.9950980392156863 |
| The Logistic Regression unbalanced | Best best\_C found: 100  Best best\_penalty found: l1 | Best average score: 0.9982878539774873  Score on training/validation set: 0.9989300235394821  Score on test set: 0.9955070603337612 | 0.97 | 0.9840182648401826 |
| The SVM | Best best\_C found: 100  Best best\_kernel found: linear  Best best\_gamma found: 1e-05 | Best average score: 0.9916532905296951  Score on training/validation set: 0.9903691813804173  Score on test set: 0.9942252165543792 | 0.97 | 0.9852941176470589 |

1. **The Train- test split + Normalize**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **The best parameters** | **The best parameters** | **The accuracy** | **The recall** | **The ROC score** |
|  |  |  |  |  |
| The Logistic Regression | Best C: 0.1  Best penalty: l1 | Accuracy on training set: 1.000  Accuracy on test set: 1.000 | 1.00 | 0.991442542787286 |
| The Logistic Regression unbalanced | Best C: 0.1  Best penalty: l1 | Accuracy on training set: 1.000  Accuracy on test set: 1.000 | 1.00 | 0.9936424831712789 |
| The SVM | Best C: 100  Best kernel: rbf  Best gamma: 0.0001291549665014884 | Accuracy on training set: 1.000  Accuracy on test set: 1.000 | 1.00 | 0.991442542787286 |

1. **The Train-test split + Standardization**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **The best parameters** | **The best parameters** | **The accuracy** | **The recall** | **The ROC score** |
|  |  |  |  |  |
| The Logistic Regression | Best C: 0.1  Best penalty: l1 | Accuracy on training set: 1.000  Accuracy on test set: 1.000 | 1.00 | 0.991442542787286 |
| The Logistic Regression unbalanced | Best C: 0.1  Best penalty: l1 | Accuracy on training set: 1.000  Accuracy on test set: 1.000 | 1.00 | 0.9936424831712789 |
| The SVM | Best C: 100  Best kernel: rbf  Best gamma: 1e-05 | Accuracy on training set: 1.000  Accuracy on test set: 1.000 | 1.00 | 0.991442542787286 |

1. **The Validation set split + Normalize**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **The best parameters** | **The best parameters** | **The accuracy** | **The recall** | **The ROC score** |
| The KNN | Best number of neighbors found: 2 | Best score on validation set: 0.9948652118100129  Score on training/validation set: 0.9964686998394864  Score on test set: 0.9932627526467758 |  |  |
| The Logistic Regression | Best best\_C found: 100  Best best\_penalty found: l1 | Best score on validation set: 0.993581514762516  Score on training/validation set: 0.9983948635634029  Score on test set: 0.9980750721847931 | 1.00 | 0.9969502172126334 |
| The Logistic Regression unbalanced | Best best\_C found: 1.0  Best best\_penalty found: l1 | Best score on validation set: 0.9991445680068435  Score on training/validation set: 0.9967900706184464  Score on test set: 0.9961489088575096 | 1.00 | 0.9977595220313668 |
| The SVM | Best best\_C found: 100  Best best\_kernel found: rbf  Best best\_gamma found: 0.2782559402207126 | Best score on validation set: 0.9961489088575096  Score on training/validation set: 0.9971107544141252  Score on test set: 0.9980750721847931 | 1.00 | 0.9969502172126334 |

1. **The validation set split + Standardization**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **The best parameters** | **The best parameters** | **The accuracy** | **The recall** | **The ROC score** |
| The KNN | Best number of neighbors found: 2 | Best score on validation set: 0.9922978177150192  Score on training/validation set: 0.9955056179775281  Score on test set: 0.9961501443695862 |  |  |
| The Logistic Regression | Best best\_C found: 100  Best best\_penalty found: l1 | Best score on validation set: 0.9974326059050064  Score on training/validation set: 0.9990369181380417  Score on test set: 0.9980750721847931 | 1.00 | 0.9969502172126334 |
| The Logistic Regression unbalanced | Best best\_C found: 100  Best best\_penalty found: l1 | Best score on validation set: 0.9991445680068435  Score on training/validation set: 0.9991440188315857  Score on test set: 0.9961489088575096 | 0.97 | 0.9863013698630136 |
| The SVM | Best best\_C found: 100  Best best\_kernel found: linear  Best best\_gamma found: 1e-05 | Best score on validation set: 0.9948652118100129  Score on training/validation set: 0.9990369181380417  Score on test set: 0.9980750721847931 | 1.00 | 0.9969502172126334 |

1. **The Cross-validation split + Normalize**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **The best parameters** | **The best parameters** | **The accuracy** | **The recall** | **The ROC score** |
| The KNN | Best number of neighbors found: 1 | Best average score:  0.9932584269662922  Score on training/validation set:  0.78330658105939  Score on test set:  0.9932627526467758 |  |  |
| The Logistic Regression | Best best\_C found: 100  Best best\_penalty found: l1 | Best average score: 0.9967897271268058  Score on training/validation set: 0.9983948635634029  Score on test set: 0.9980750721847931 | 1.00 | 0.9969502172126334 |
| The Logistic Regression unbalanced | Best best\_C found: 100  Best best\_penalty found: l1 | Best average score: 0.9987158904831155  Score on training/validation set: 0.9991440188315857  Score on test set: 0.9961489088575096 | 0.97 | 0.9863013698630136 |
| The SVM | Best best\_C found: 100  Best best\_kernel found: rbf  Best best\_gamma found: 0.2782559402207126 | Best average score: 0.9961476725521671  Score on training/validation set: 0.9971107544141252  Score on test set: 0.9980750721847931 | 1.00 | 0.9969502172126334 |

1. **The Cross-validation split + Standardization**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **The best parameters** | **The best parameters** | **The accuracy** | **The recall** | **The ROC score** |
| The KNN | Best number of neighbors found: 1 | Best average score: 0.9913322632423757  Score on training/validation set: 0.7823434991974317  Score on test set: 0.9961501443695862 |  |  |
| The Logistic Regression | Best best\_C found: 100  Best best\_penalty found: l1 | Best average score: 0.997752808988764  Score on training/validation set: 0.9990369181380417  Score on test set: 0.9980750721847931 | 1.00 | 0.9969502172126334 |
| The Logistic Regression unbalanced | Best best\_C found: 100  Best best\_penalty found: l1 | Best average score: 0.9987161195021127  Score on training/validation set: 0.9991440188315857  Score on test set: 0.9961489088575096 | 0.97 | 0.9863013698630136 |
| The SVM | Best best\_C found: 100  Best best\_kernel found: linear  Best best\_gamma found: 1e-05 | Best average score: 0.9974317817014444  Score on training/validation set: 0.9990369181380417  Score on test set: 0.9980750721847931 | 1.00 | 0.9969502172126334 |