# Cross-validation 10 fold

J48 accuracies

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | prunning | Top 5 | Top 10 | Top 20 | All |
| T | n/a | 1 | T | 81.7441 | 83.4985 | 84.8813 | 84.3344 |
| F | 0.25 | 5 | F | 81.6925 | 82.4252 | 82.9825 | 83.8596 |
| F | 0.5 | 10 | T | 79.1434 | 80.7018 | 81.1042 | 81.8782 |
| F | 0.5 | 20 | T | 76.0991 | 77.5645 | 78.5346 | 79.5459 |
| F | 0.8 | 5 | T | 80.7018 | 82.4974 | 82.9825 | 83.8493 |

Random forest accuracies

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| seed | numI-terations | breakTies-Randoml | maxDepth | Top 5 | Top 10 | Top 20 | All |
| 1 | 100 | F | 0 | 93.5707 | 96.1713 | 96.8111 | 96.7389 |
| 10 | 100 | F | 0 | 93.6223 | 96.0165 | 96.7183 | 96.7905 |
| 10 | 50 | F | 0 | 93.3746 | 95.8308 | 96.2951 | 96.0784 |
| 10 | 100 | T | 10 | 90.4954 | 94.6852 | 95.129 | 92.064 |

# Test data

J48 accuracies

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | prunning | All |
| T | n/a | 1 | T | 69.0291 |
| F | 0.25 | 5 | F | 70.356 |
| F | 0.5 | 10 | T | 70.0647 |
| F | 0.5 | 20 | T | 68.4142 |
| F | 0.8 | 5 | T | 70.1618 |

Random forest accuracies

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| seed | numI-terations | breakTies-Randoml | maxDepth | All |
| 1 | 100 | F | 0 | 66.6667 |
| 10 | 100 | F | 0 | 65.4369 |
| 10 | 50 | F | 0 | 64.0129 |
| 10 | 100 | T | 10 | 65.2104 |

# Test data + 4000

|  |  |
| --- | --- |
| Data set | Number of instances |
| Original test data set | 3090 |
| Old training data | 9690 |
| New test dada + 4000 | 7090 |
| New training data -4000 | 5690 |

4000 instances were extracted with stratify in Python

*X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=* *.4127541017, stratify=y, random\_state=17)*

J48 accuracies

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | prunning | All |
| T | 0.25 | 1 | T | 81.4386 |
| F | 0.25 | 5 | F | 78.8717 |
| F | 0.5 | 10 | T | 77.2638 |
| F | 0.5 | 20 | T | 74.2877 |
| F | 0.8 | 5 | T | 79.055 |

Random forest accuracies

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| seed | numI-terations | breakTies-Randoml | maxDepth | All |
| 1 | 100 | F | 0 | 82.4683 |
| 10 | 100 | F | 0 | 82.5388 |
| 10 | 50 | F | 0 | 81.9182 |
| 10 | 100 | T | 10 | 80.5219 |

# Test data + 9000

|  |  |
| --- | --- |
| Data set | Number of instances |
| Original test data set | 3090 |
| Old training data | 9690 |
| New test dada + 9000 | 12090 |
| New training data -9000 | 690 |

4000 instances were extracted with stratify in Python

*X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.9286967289, stratify=y, random\_state=17)*

J48 accuracies

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | prunning | All |
| T | 0.25 | 1 | T | 61.067 |
| F | 0.25 | 5 | F | 58.7428 |
| F | 0.5 | 10 | T | 60.5707 |
| F | 0.5 | 20 | T | 56.2035 |
| F | 0.8 | 5 | T | 58.7428 |

Random forest accuracies

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| seed | numI-terations | breakTies-Randoml | maxDepth | All |
| 1 | 100 | F | 0 | 63.9868 |
| 10 | 100 | F | 0 | 64.665 |
| 10 | 50 | F | 0 | 62.8122 |
| 10 | 100 | T | 10 | 65.0868 |

## J48 results – top 5 dataset

E1 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| T | 0.25 | 1 | T | T |

Accuracy: 81.7441 %

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.700 0.005 0.746 0.700 0.722 0.717 0.850 0.534 s0

0.823 0.054 0.818 0.823 0.820 0.767 0.873 0.697 s1

0.819 0.048 0.837 0.819 0.827 0.776 0.878 0.688 s2

0.829 0.027 0.841 0.829 0.835 0.807 0.899 0.710 s3

0.859 0.042 0.840 0.859 0.849 0.810 0.906 0.725 s4

0.705 0.007 0.701 0.705 0.703 0.696 0.862 0.517 s5

0.772 0.011 0.728 0.772 0.749 0.740 0.884 0.576 s6

0.738 0.005 0.794 0.738 0.765 0.759 0.865 0.585 s7

0.807 0.013 0.786 0.807 0.796 0.784 0.901 0.651 s8

0.730 0.008 0.719 0.730 0.724 0.716 0.870 0.535 s9

Weighted Avg. 0.817 0.038 0.818 0.817 0.817 0.780 0.886 0.681

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

147 20 5 4 30 1 0 0 2 1 | a = s0

18 1826 158 36 150 4 8 8 6 6 | b = s1

6 181 1842 71 100 12 14 5 11 8 | c = s2

1 38 84 1169 23 13 25 6 29 22 | d = s3

16 140 72 29 1701 2 7 3 7 3 | e = s4

1 2 7 13 1 148 9 7 16 6 | f = s5

0 6 10 18 4 5 278 7 19 13 | g = s6

1 7 10 7 6 6 11 177 10 5 | h = s7

3 8 9 21 7 15 20 8 436 13 | i = s8

4 3 5 22 3 5 10 2 19 197 | j = s9

E2 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| F | 0.25 | 5 | T | F |

Accuracy: 81.6925 %

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.700 0.005 0.742 0.700 0.721 0.715 0.847 0.526 s0

0.821 0.055 0.816 0.821 0.819 0.765 0.880 0.715 s1

0.818 0.049 0.835 0.818 0.826 0.774 0.883 0.705 s2

0.830 0.026 0.844 0.830 0.837 0.809 0.900 0.725 s3

0.856 0.043 0.838 0.856 0.847 0.807 0.906 0.743 s4

0.719 0.007 0.706 0.719 0.712 0.706 0.858 0.516 s5

0.775 0.011 0.728 0.775 0.751 0.741 0.882 0.572 s6

0.738 0.005 0.794 0.738 0.765 0.759 0.865 0.583 s7

0.811 0.012 0.793 0.811 0.802 0.790 0.900 0.651 s8

0.730 0.008 0.719 0.730 0.724 0.716 0.864 0.537 s9

Weighted Avg. 0.817 0.038 0.817 0.817 0.817 0.779 0.888 0.694

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

147 20 5 4 30 1 0 0 2 1 | a = s0

18 1823 160 35 153 4 8 7 6 6 | b = s1

7 181 1840 70 101 12 15 5 11 8 | c = s2

1 38 85 1170 22 13 25 6 28 22 | d = s3

16 145 74 29 1694 2 7 3 7 3 | e = s4

1 2 7 12 1 151 9 7 14 6 | f = s5

0 6 9 18 4 5 279 8 18 13 | g = s6

1 7 10 7 6 6 12 177 9 5 | h = s7

3 8 9 21 7 15 18 8 438 13 | i = s8

4 3 5 21 4 5 10 2 19 197 | j = s9

E3 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| F | 0.5 | 10 | T | T |

Accuracy: 79.1434 %

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.624 0.006 0.701 0.624 0.660 0.654 0.928 0.638 s0

0.798 0.060 0.799 0.798 0.799 0.739 0.931 0.848 s1

0.815 0.060 0.804 0.815 0.809 0.751 0.932 0.851 s2

0.834 0.037 0.792 0.834 0.813 0.780 0.946 0.824 s3

0.838 0.042 0.838 0.838 0.838 0.796 0.947 0.865 s4

0.576 0.004 0.776 0.576 0.661 0.662 0.928 0.648 s5

0.661 0.011 0.702 0.661 0.681 0.669 0.912 0.618 s6

0.683 0.007 0.704 0.683 0.693 0.686 0.898 0.613 s7

0.735 0.016 0.727 0.735 0.731 0.715 0.936 0.758 s8

0.659 0.009 0.667 0.659 0.663 0.653 0.922 0.641 s9

Weighted Avg. 0.791 0.044 0.791 0.791 0.791 0.748 0.935 0.814

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

131 20 5 7 41 0 0 1 4 1 | a = s0

12 1772 210 38 152 4 6 11 13 2 | b = s1

5 201 1833 85 82 4 16 6 11 7 | c = s2

5 36 102 1176 21 7 17 2 21 23 | d = s3

26 157 68 40 1659 0 3 7 14 6 | e = s4

1 2 11 20 2 121 7 8 24 14 | f = s5

0 5 11 41 3 6 238 18 26 12 | g = s6

1 16 13 7 10 3 14 164 9 3 | h = s7

5 6 20 36 6 9 28 12 397 21 | i = s8

1 2 8 34 4 2 10 4 27 178 | j = s9

E4 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| F | 0.5 | 20 | T | T |

Accuracy: 76.0991 %

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.605 0.008 0.623 0.605 0.614 0.605 0.934 0.594 s0

0.779 0.069 0.769 0.779 0.774 0.706 0.923 0.823 s1

0.793 0.065 0.787 0.793 0.790 0.726 0.930 0.828 s2

0.791 0.043 0.760 0.791 0.776 0.737 0.952 0.792 s3

0.816 0.050 0.807 0.816 0.811 0.763 0.948 0.863 s4

0.457 0.005 0.671 0.457 0.544 0.546 0.915 0.544 s5

0.611 0.014 0.634 0.611 0.622 0.608 0.934 0.646 s6

0.604 0.008 0.653 0.604 0.628 0.619 0.917 0.589 s7

0.693 0.017 0.712 0.693 0.702 0.685 0.938 0.695 s8

0.622 0.010 0.634 0.622 0.628 0.618 0.935 0.587 s9

Weighted Avg. 0.761 0.050 0.760 0.761 0.760 0.712 0.935 0.791

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

127 24 4 5 47 0 0 1 2 0 | a = s0

15 1729 217 52 167 10 6 9 8 7 | b = s1

4 228 1784 94 93 7 12 7 10 11 | c = s2

17 41 109 1116 36 5 36 4 25 21 | d = s3

21 191 77 40 1615 0 10 10 8 8 | e = s4

4 3 21 21 1 96 6 9 32 17 | f = s5

0 6 10 52 13 4 220 21 26 8 | g = s6

1 12 21 9 11 1 23 145 13 4 | h = s7

6 9 17 51 11 14 26 11 374 21 | i = s8

9 4 8 28 7 6 8 5 27 168 | j = s9

E5 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| F | 0.8 | 5 | T | T |

Accuracy: 80.7018 %

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.695 0.007 0.692 0.695 0.694 0.687 0.883 0.582 s0

0.821 0.061 0.799 0.821 0.810 0.752 0.919 0.811 s1

0.816 0.052 0.825 0.816 0.821 0.767 0.922 0.816 s2

0.844 0.031 0.822 0.844 0.833 0.804 0.936 0.801 s3

0.838 0.042 0.838 0.838 0.838 0.797 0.935 0.826 s4

0.690 0.005 0.736 0.690 0.713 0.707 0.899 0.653 s5

0.719 0.011 0.717 0.719 0.718 0.708 0.907 0.663 s6

0.683 0.005 0.792 0.683 0.734 0.730 0.884 0.634 s7

0.767 0.013 0.775 0.767 0.771 0.758 0.925 0.735 s8

0.678 0.007 0.744 0.678 0.709 0.702 0.912 0.671 s9

Weighted Avg. 0.807 0.041 0.807 0.807 0.807 0.767 0.923 0.787

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

146 23 1 2 33 0 1 0 1 3 | a = s0

18 1822 168 42 143 5 7 5 6 4 | b = s1

6 197 1837 71 91 10 18 4 11 5 | c = s2

5 44 79 1190 24 3 19 5 28 13 | d = s3

28 154 87 34 1660 0 5 4 7 1 | e = s4

1 4 6 17 1 145 8 5 13 10 | f = s5

0 5 11 35 5 6 259 8 22 9 | g = s6

1 14 14 10 8 5 13 164 9 2 | h = s7

4 12 17 19 12 17 21 8 414 16 | i = s8

2 6 6 27 3 6 10 4 23 183 | j = s9

## Random forest results – top 5 dataset

E1 Settings:

|  |  |  |  |
| --- | --- | --- | --- |
| seed | numIterations | breakTiesRandoml | maxDepth |
| 1 | 100 | F | 0 |

Accuracy: 93.5707 %

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.776 0.000 0.982 0.776 0.867 0.871 0.996 0.950 s0

0.932 0.022 0.928 0.932 0.930 0.909 0.993 0.982 s1

0.944 0.021 0.932 0.944 0.938 0.919 0.996 0.986 s2

0.970 0.014 0.921 0.970 0.945 0.935 0.998 0.990 s3

0.956 0.015 0.941 0.956 0.948 0.935 0.997 0.989 s4

0.833 0.000 0.983 0.833 0.902 0.903 0.999 0.970 s5

0.903 0.002 0.950 0.903 0.926 0.923 0.993 0.957 s6

0.938 0.000 0.987 0.938 0.962 0.961 1.000 0.991 s7

0.939 0.004 0.934 0.939 0.936 0.933 0.998 0.979 s8

0.815 0.001 0.957 0.815 0.880 0.880 0.992 0.926 s9

Weighted Avg. 0.936 0.015 0.936 0.936 0.935 0.922 0.996 0.982

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

163 25 3 0 19 0 0 0 0 0 | a = s0

1 2070 88 14 45 0 0 0 1 1 | b = s1

0 67 2123 18 41 0 0 0 1 0 | c = s2

0 11 20 1367 6 0 3 0 2 1 | d = s3

0 42 29 13 1892 0 1 1 2 0 | e = s4

0 0 6 13 0 175 4 1 10 1 | f = s5

0 5 2 19 2 3 325 0 2 2 | g = s6

0 4 1 3 1 0 1 225 5 0 | h = s7

0 3 4 13 2 0 5 1 507 5 | i = s8

2 4 2 24 2 0 3 0 13 220 | j = s9

E2 Settings:

|  |  |  |  |
| --- | --- | --- | --- |
| seed | numIterations | breakTiesRandoml | maxDepth |
| 10 | 100 | F | 0 |

Accuracy: 93.6223 %

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.795 0.000 0.988 0.795 0.881 0.884 0.997 0.951 s0

0.927 0.020 0.932 0.927 0.929 0.908 0.993 0.982 s1

0.945 0.020 0.933 0.945 0.939 0.921 0.995 0.986 s2

0.970 0.014 0.921 0.970 0.944 0.935 0.998 0.989 s3

0.960 0.016 0.940 0.960 0.950 0.937 0.997 0.989 s4

0.833 0.000 0.983 0.833 0.902 0.903 0.999 0.968 s5

0.894 0.002 0.955 0.894 0.924 0.922 0.995 0.960 s6

0.942 0.000 1.000 0.942 0.970 0.970 1.000 0.992 s7

0.943 0.004 0.929 0.943 0.936 0.932 0.998 0.980 s8

0.819 0.002 0.929 0.819 0.870 0.868 0.994 0.928 s9

Weighted Avg. 0.936 0.015 0.937 0.936 0.936 0.923 0.996 0.982

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

167 19 3 0 21 0 0 0 0 0 | a = s0

1 2058 89 18 52 0 0 0 1 1 | b = s1

0 64 2127 21 36 0 0 0 0 2 | c = s2

0 17 16 1367 6 0 3 0 0 1 | d = s3

0 37 27 11 1900 0 0 0 4 1 | e = s4

0 0 7 14 0 175 3 0 10 1 | f = s5

0 5 2 16 2 3 322 0 6 4 | g = s6

0 4 2 2 0 0 1 226 5 0 | h = s7

0 3 3 12 1 0 5 0 509 7 | i = s8

1 2 3 24 3 0 3 0 13 221 | j = s9

E2 Settings:

|  |  |  |  |
| --- | --- | --- | --- |
| seed | numIterations | breakTiesRandoml | maxDepth |
| 10 | 50 | F | 0 |

Accuracy: 93.3746 %

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.814 0.000 0.988 0.814 0.893 0.895 0.996 0.947 s0

0.923 0.019 0.935 0.923 0.929 0.908 0.992 0.979 s1

0.944 0.023 0.925 0.944 0.934 0.914 0.995 0.983 s2

0.967 0.014 0.922 0.967 0.944 0.935 0.997 0.988 s3

0.957 0.017 0.934 0.957 0.945 0.931 0.996 0.988 s4

0.848 0.000 0.983 0.848 0.910 0.911 0.999 0.964 s5

0.889 0.002 0.947 0.889 0.917 0.914 0.991 0.951 s6

0.938 0.000 0.996 0.938 0.966 0.965 1.000 0.991 s7

0.941 0.004 0.930 0.941 0.936 0.932 0.997 0.977 s8

0.796 0.002 0.927 0.796 0.857 0.855 0.987 0.919 s9

Weighted Avg. 0.934 0.016 0.934 0.934 0.933 0.919 0.995 0.979

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

171 17 2 0 20 0 0 0 0 0 | a = s0

0 2050 91 18 57 0 1 0 0 3 | b = s1

0 64 2123 20 41 0 0 0 0 2 | c = s2

0 15 23 1364 5 0 3 0 0 0 | d = s3

0 35 34 15 1894 0 0 0 2 0 | e = s4

0 0 6 15 0 178 2 0 8 1 | f = s5

0 5 3 12 5 3 320 0 7 5 | g = s6

0 3 3 1 0 0 2 225 6 0 | h = s7

1 2 4 11 1 0 7 0 508 6 | i = s8

1 2 5 24 4 0 3 1 15 215 | j = s9

E3 Settings:

|  |  |  |  |
| --- | --- | --- | --- |
| seed | numIterations | breakTiesRandoml | maxDepth |
| 10 | 100 | T | 10 |

Accuracy: 90.4954 %

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.290 0.000 0.968 0.290 0.447 0.526 0.966 0.647 s0

0.781 0.145 0.616 0.781 0.689 0.589 0.898 0.767 s1

0.756 0.090 0.717 0.756 0.736 0.654 0.922 0.836 s2

0.879 0.090 0.626 0.879 0.731 0.690 0.960 0.853 s3

0.577 0.031 0.826 0.577 0.679 0.629 0.939 0.830 s4

0.157 0.000 1.000 0.157 0.272 0.393 0.963 0.631 s5

0.425 0.010 0.622 0.425 0.505 0.499 0.963 0.626 s6

0.533 0.000 0.977 0.533 0.690 0.717 0.986 0.841 s7

0.624 0.017 0.679 0.624 0.651 0.632 0.974 0.756 s8

0.444 0.006 0.698 0.444 0.543 0.547 0.958 0.631 s9

Weighted Avg. 0.686 0.075 0.715 0.686 0.677 0.622 0.934 0.795

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

61 110 6 10 23 0 0 0 0 0 | a = s0

1 1733 293 75 117 0 0 0 1 0 | b = s1

0 247 1700 232 57 0 0 1 13 0 | c = s2

0 32 106 1240 27 0 1 0 4 0 | d = s3

1 647 141 39 1142 0 0 0 10 0 | e = s4

0 3 41 49 0 33 4 1 54 25 | f = s5

0 8 4 156 1 0 153 1 26 11 | g = s6

0 10 32 28 12 0 17 128 13 0 | h = s7

0 13 34 71 1 0 68 0 337 16 | i = s8

0 11 13 82 3 0 3 0 38 120 | j = s9

## J48 results – top 10 dataset

E1 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| T | 0.25 | 1 | T | T |

Accuracy: 83.4985 %

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.700 0.006 0.731 0.700 0.715 0.709 0.857 0.529 s0

0.840 0.049 0.835 0.840 0.837 0.789 0.885 0.702 s1

0.832 0.049 0.836 0.832 0.834 0.783 0.878 0.715 s2

0.848 0.025 0.850 0.848 0.849 0.823 0.909 0.720 s3

0.881 0.033 0.874 0.881 0.878 0.846 0.920 0.768 s4

0.771 0.005 0.771 0.771 0.771 0.766 0.883 0.600 s5

0.756 0.010 0.751 0.756 0.753 0.744 0.872 0.575 s6

0.796 0.004 0.827 0.796 0.811 0.806 0.894 0.663 s7

0.822 0.011 0.821 0.822 0.821 0.811 0.908 0.684 s8

0.737 0.007 0.740 0.737 0.738 0.731 0.866 0.555 s9

Weighted Avg. 0.835 0.035 0.835 0.835 0.835 0.800 0.894 0.704

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

147 23 5 8 20 2 1 1 2 1 | a = s0

16 1864 162 33 113 5 8 5 7 7 | b = s1

10 167 1871 78 71 11 16 6 13 7 | c = s2

0 36 90 1196 24 7 23 2 16 16 | d = s3

19 103 69 21 1745 2 5 3 6 7 | e = s4

2 4 3 5 6 162 7 6 12 3 | f = s5

1 8 10 34 5 4 272 6 9 11 | g = s6

1 9 11 3 5 4 2 191 13 1 | h = s7

4 11 11 16 5 8 17 7 444 17 | i = s8

1 8 7 13 3 5 11 4 19 199 | j = s9

E2 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| F | 0.25 | 5 | T | F |

Accuracy: 82.4252 %

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.676 0.005 0.743 0.676 0.708 0.703 0.896 0.670 s0

0.841 0.055 0.820 0.841 0.831 0.780 0.922 0.814 s1

0.828 0.052 0.827 0.828 0.827 0.775 0.920 0.810 s2

0.835 0.023 0.859 0.835 0.847 0.821 0.936 0.818 s3

0.872 0.034 0.868 0.872 0.870 0.837 0.952 0.862 s4

0.705 0.006 0.733 0.705 0.718 0.712 0.897 0.698 s5

0.764 0.010 0.741 0.764 0.752 0.743 0.919 0.719 s6

0.742 0.006 0.751 0.742 0.746 0.740 0.912 0.700 s7

0.785 0.013 0.784 0.785 0.784 0.772 0.928 0.719 s8

0.685 0.008 0.714 0.685 0.699 0.691 0.885 0.641 s9

Weighted Avg. 0.824 0.037 0.824 0.824 0.824 0.788 0.927 0.801

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

142 24 4 8 23 1 2 0 3 3 | a = s0

16 1868 156 32 110 5 10 6 12 5 | b = s1

7 188 1863 69 74 7 9 12 12 9 | c = s2

3 36 97 1177 30 10 27 1 16 13 | d = s3

16 110 75 19 1727 3 6 10 10 4 | e = s4

1 5 8 7 7 148 6 10 14 4 | f = s5

1 9 7 29 6 4 275 7 14 8 | g = s6

1 15 14 1 9 10 2 178 8 2 | h = s7

0 9 17 15 2 10 26 11 424 26 | i = s8

4 13 12 13 1 4 8 2 28 185 | j = s9

E3 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| F | 0.5 | 10 | T | T |

Accuracy: 80.7018 %

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.657 0.005 0.762 0.657 0.706 0.702 0.913 0.682 s0

0.818 0.059 0.805 0.818 0.811 0.755 0.934 0.843 s1

0.819 0.063 0.798 0.819 0.808 0.749 0.929 0.842 s2

0.828 0.028 0.835 0.828 0.832 0.803 0.946 0.829 s3

0.861 0.036 0.858 0.861 0.860 0.823 0.956 0.885 s4

0.657 0.005 0.734 0.657 0.693 0.688 0.916 0.722 s5

0.692 0.010 0.722 0.692 0.706 0.695 0.920 0.710 s6

0.725 0.006 0.750 0.725 0.737 0.731 0.914 0.700 s7

0.783 0.015 0.755 0.783 0.769 0.755 0.939 0.744 s8

0.622 0.007 0.712 0.622 0.664 0.657 0.923 0.652 s9

Weighted Avg. 0.807 0.041 0.806 0.807 0.806 0.767 0.937 0.824

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

138 26 10 7 22 0 2 0 4 1 | a = s0

16 1816 197 37 114 7 6 10 14 3 | b = s1

6 191 1842 78 84 8 11 8 14 8 | c = s2

1 41 107 1168 31 9 29 0 14 10 | d = s3

15 130 81 21 1704 5 5 8 5 6 | e = s4

0 7 11 8 4 138 12 11 14 5 | f = s5

1 8 9 39 5 1 249 11 28 9 | g = s6

1 11 18 3 11 7 2 174 12 1 | h = s7

0 15 16 21 5 11 18 6 423 25 | i = s8

3 11 18 16 5 2 11 4 32 168 | j = s9

E4 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| F | 0.5 | 20 | T | T |

Accuracy: 77.5645 %

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.605 0.007 0.672 0.605 0.637 0.630 0.928 0.603 s0

0.785 0.061 0.793 0.785 0.789 0.727 0.931 0.854 s1

0.785 0.069 0.775 0.785 0.780 0.713 0.929 0.834 s2

0.799 0.033 0.803 0.799 0.801 0.768 0.949 0.826 s3

0.847 0.048 0.819 0.847 0.833 0.789 0.954 0.872 s4

0.595 0.006 0.672 0.595 0.631 0.625 0.937 0.691 s5

0.633 0.012 0.669 0.633 0.650 0.638 0.930 0.692 s6

0.696 0.008 0.693 0.696 0.694 0.687 0.932 0.699 s7

0.739 0.017 0.724 0.739 0.731 0.715 0.951 0.742 s8

0.578 0.010 0.617 0.578 0.597 0.586 0.941 0.614 s9

Weighted Avg. 0.776 0.047 0.775 0.776 0.775 0.729 0.939 0.817

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

127 23 5 8 38 0 1 0 5 3 | a = s0

16 1743 221 40 149 14 4 12 16 5 | b = s1

4 205 1767 92 117 13 20 10 9 13 | c = s2

3 44 124 1127 35 7 30 7 16 17 | d = s3

22 126 85 31 1677 5 8 9 10 7 | e = s4

0 4 16 15 5 125 8 6 23 8 | f = s5

2 10 14 41 8 3 228 10 33 11 | g = s6

3 15 13 1 10 8 8 167 13 2 | h = s7

1 17 18 24 5 6 22 17 399 31 | i = s8

11 10 18 24 4 5 12 3 27 156 | j = s9

E5 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| F | 0.8 | 5 | T | T |

Accuracy: 82.4974 %

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.671 0.005 0.758 0.671 0.712 0.707 0.898 0.671 s0

0.841 0.054 0.823 0.841 0.832 0.781 0.922 0.815 s1

0.830 0.052 0.828 0.830 0.829 0.777 0.918 0.805 s2

0.835 0.023 0.859 0.835 0.847 0.821 0.936 0.815 s3

0.875 0.035 0.867 0.875 0.871 0.837 0.951 0.859 s4

0.705 0.006 0.733 0.705 0.718 0.712 0.897 0.698 s5

0.761 0.010 0.741 0.761 0.751 0.741 0.920 0.718 s6

0.742 0.006 0.745 0.742 0.743 0.737 0.911 0.698 s7

0.785 0.013 0.785 0.785 0.785 0.773 0.930 0.719 s8

0.685 0.008 0.714 0.685 0.699 0.691 0.885 0.642 s9

Weighted Avg. 0.825 0.037 0.825 0.825 0.825 0.789 0.927 0.799

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

141 24 4 8 24 1 2 0 3 3 | a = s0

14 1867 158 32 111 5 10 6 12 5 | b = s1

5 185 1867 69 75 7 9 13 11 9 | c = s2

3 36 96 1177 31 10 27 1 16 13 | d = s3

16 106 73 19 1733 3 6 11 9 4 | e = s4

1 5 8 7 7 148 6 10 14 4 | f = s5

1 8 7 29 7 4 274 7 15 8 | g = s6

1 15 14 1 9 10 2 178 8 2 | h = s7

0 9 17 15 2 10 26 11 424 26 | i = s8

4 13 12 13 1 4 8 2 28 185 | j = s9

## Random forest results – top 10 dataset

E1 Settings:

|  |  |  |  |
| --- | --- | --- | --- |
| seed | numIterations | breakTiesRandoml | maxDepth |
| 1 | 100 | F | 0 |

Accuracy: 96.1713 %

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.819 0.000 0.989 0.819 0.896 0.898 0.998 0.969 s0

0.959 0.013 0.956 0.959 0.957 0.945 0.997 0.991 s1

0.970 0.014 0.954 0.970 0.962 0.950 0.998 0.993 s2

0.977 0.006 0.966 0.977 0.971 0.967 0.998 0.994 s3

0.981 0.010 0.962 0.981 0.971 0.964 0.999 0.996 s4

0.938 0.000 0.990 0.938 0.963 0.963 1.000 0.993 s5

0.925 0.001 0.982 0.925 0.953 0.951 0.997 0.973 s6

0.967 0.000 0.991 0.967 0.979 0.978 1.000 0.998 s7

0.963 0.002 0.963 0.963 0.963 0.961 0.999 0.989 s8

0.870 0.001 0.955 0.870 0.911 0.909 0.997 0.955 s9

Weighted Avg. 0.962 0.009 0.962 0.962 0.961 0.953 0.998 0.991

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

172 29 2 0 7 0 0 0 0 0 | a = s0

0 2128 53 4 33 0 0 0 1 1 | b = s1

0 30 2183 9 28 0 0 0 0 0 | c = s2

0 10 16 1377 5 0 0 0 1 1 | d = s3

1 13 18 5 1942 0 0 1 0 0 | e = s4

0 0 4 0 0 197 0 1 7 1 | f = s5

0 5 4 11 1 1 333 0 2 3 | g = s6

0 2 3 1 1 1 0 232 0 0 | h = s7

0 3 2 7 1 0 2 0 520 5 | i = s8

1 5 4 11 1 0 4 0 9 235 | j = s9

E2 Settings:

|  |  |  |  |
| --- | --- | --- | --- |
| seed | numIterations | breakTiesRandoml | maxDepth |
| 10 | 100 | F | 0 |

Accuracy: 96.0165 %

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.824 0.000 0.994 0.824 0.901 0.903 0.997 0.966 s0

0.953 0.013 0.955 0.953 0.954 0.941 0.996 0.990 s1

0.970 0.015 0.950 0.970 0.960 0.948 0.998 0.993 s2

0.972 0.006 0.966 0.972 0.969 0.964 0.998 0.994 s3

0.985 0.010 0.961 0.985 0.973 0.966 0.999 0.996 s4

0.943 0.000 0.990 0.943 0.966 0.965 1.000 0.991 s5

0.914 0.001 0.976 0.914 0.944 0.943 0.998 0.973 s6

0.963 0.000 0.987 0.963 0.975 0.974 1.000 0.998 s7

0.950 0.002 0.959 0.950 0.954 0.952 0.999 0.990 s8

0.889 0.001 0.964 0.889 0.925 0.924 0.996 0.955 s9

Weighted Avg. 0.960 0.010 0.960 0.960 0.960 0.952 0.998 0.991

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

173 25 2 0 10 0 0 0 0 0 | a = s0

0 2116 63 7 32 0 0 0 1 1 | b = s1

0 36 2183 5 26 0 0 0 0 0 | c = s2

0 13 19 1371 5 0 1 0 1 0 | d = s3

0 8 16 4 1950 0 2 0 0 0 | e = s4

0 0 4 0 0 198 0 2 5 1 | f = s5

0 5 5 11 2 1 329 0 4 3 | g = s6

0 2 2 0 1 0 0 231 4 0 | h = s7

0 4 3 12 1 1 2 0 513 4 | i = s8

1 6 1 9 2 0 3 1 7 240 | j = s9

E2 Settings:

|  |  |  |  |
| --- | --- | --- | --- |
| seed | numIterations | breakTiesRandoml | maxDepth |
| 10 | 50 | F | 0 |

Accuracy: 95.8308 %

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.800 0.000 0.988 0.800 0.884 0.887 0.993 0.957 s0

0.954 0.015 0.950 0.954 0.952 0.938 0.996 0.989 s1

0.973 0.015 0.953 0.973 0.963 0.951 0.997 0.992 s2

0.971 0.006 0.964 0.971 0.967 0.962 0.998 0.993 s3

0.983 0.010 0.961 0.983 0.972 0.965 0.999 0.995 s4

0.933 0.001 0.975 0.933 0.954 0.953 1.000 0.989 s5

0.903 0.001 0.973 0.903 0.937 0.935 0.997 0.967 s6

0.954 0.001 0.979 0.954 0.966 0.965 1.000 0.995 s7

0.941 0.003 0.957 0.941 0.949 0.946 0.999 0.987 s8

0.881 0.001 0.960 0.881 0.919 0.918 0.994 0.950 s9

Weighted Avg. 0.958 0.010 0.959 0.958 0.958 0.950 0.997 0.989

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

168 32 1 1 8 0 0 0 0 0 | a = s0

1 2117 59 5 36 0 0 0 1 1 | b = s1

0 32 2189 7 22 0 0 0 0 0 | c = s2

0 13 21 1369 5 0 1 0 1 0 | d = s3

0 13 13 6 1947 0 1 0 0 0 | e = s4

0 1 3 2 0 196 1 2 5 0 | f = s5

0 6 5 9 3 3 325 0 5 4 | g = s6

0 4 2 0 2 0 0 229 3 0 | h = s7

0 7 1 11 1 2 4 1 508 5 | i = s8

1 3 4 10 2 0 2 2 8 238 | j = s9

E3 Settings:

|  |  |  |  |
| --- | --- | --- | --- |
| seed | numIterations | breakTiesRandoml | maxDepth |
| 10 | 100 | T | 10 |

Accuracy: 94.6852 %

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.724 0.000 0.987 0.724 0.835 0.842 0.997 0.948 s0

0.941 0.021 0.931 0.941 0.936 0.917 0.993 0.982 s1

0.957 0.019 0.939 0.957 0.948 0.932 0.996 0.987 s2

0.965 0.010 0.945 0.965 0.955 0.947 0.997 0.990 s3

0.974 0.013 0.951 0.974 0.962 0.953 0.998 0.993 s4

0.924 0.000 1.000 0.924 0.960 0.960 1.000 0.990 s5

0.897 0.001 0.985 0.897 0.939 0.938 0.996 0.966 s6

0.958 0.000 0.996 0.958 0.977 0.976 1.000 0.998 s7

0.950 0.003 0.954 0.950 0.952 0.949 0.999 0.984 s8

0.859 0.001 0.963 0.859 0.908 0.907 0.996 0.943 s9

Weighted Avg. 0.947 0.013 0.947 0.947 0.946 0.935 0.996 0.985

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

152 47 1 0 10 0 0 0 0 0 | a = s0

1 2089 77 13 39 0 0 0 1 0 | b = s1

0 46 2153 16 35 0 0 0 0 0 | c = s2

0 17 22 1361 8 0 1 0 0 1 | d = s3

0 27 19 5 1928 0 0 0 1 0 | e = s4

0 0 7 1 0 194 0 1 6 1 | f = s5

0 5 7 15 2 0 323 0 5 3 | g = s6

0 3 1 0 2 0 0 230 4 0 | h = s7

0 5 3 12 1 0 2 0 513 4 | i = s8

1 4 4 17 2 0 2 0 8 232 | j = s9

## J48 results – top 20 dataset

E1 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| T | 0.25 | 1 | T | T |

Accuracy: 84.8813 %

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.743 0.007 0.709 0.743 0.726 0.720 0.877 0.544 s0

0.856 0.044 0.853 0.856 0.855 0.812 0.896 0.737 s1

0.850 0.041 0.863 0.850 0.857 0.814 0.896 0.743 s2

0.865 0.029 0.837 0.865 0.851 0.825 0.914 0.717 s3

0.886 0.030 0.883 0.886 0.885 0.855 0.922 0.779 s4

0.810 0.004 0.817 0.810 0.813 0.809 0.902 0.665 s5

0.767 0.007 0.812 0.767 0.789 0.781 0.876 0.631 s6

0.863 0.004 0.852 0.863 0.857 0.854 0.929 0.737 s7

0.793 0.011 0.814 0.793 0.803 0.792 0.901 0.663 s8

0.741 0.007 0.758 0.741 0.749 0.742 0.866 0.513 s9

Weighted Avg. 0.849 0.031 0.849 0.849 0.849 0.818 0.903 0.724

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

156 13 7 11 19 0 2 0 2 0 | a = s0

19 1901 118 47 103 3 6 5 15 3 | b = s1

8 140 1913 86 68 6 3 9 10 7 | c = s2

8 37 72 1219 24 3 20 4 15 8 | d = s3

17 99 72 20 1755 1 3 3 9 1 | e = s4

0 4 7 4 3 170 4 7 4 7 | f = s5

1 9 10 20 7 7 276 2 16 12 | g = s6

0 6 6 4 3 6 2 207 4 2 | h = s7

5 17 6 25 6 9 16 4 428 24 | i = s8

6 2 6 20 0 3 8 2 23 200 | j = s9

E2 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| F | 0.25 | 5 | T | F |

Accuracy: 82.9825 %

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.714 0.006 0.739 0.714 0.726 0.721 0.908 0.654 s0

0.842 0.051 0.832 0.842 0.837 0.788 0.924 0.820 s1

0.829 0.048 0.840 0.829 0.835 0.785 0.920 0.822 s2

0.847 0.029 0.835 0.847 0.841 0.814 0.938 0.812 s3

0.876 0.036 0.864 0.876 0.870 0.836 0.948 0.863 s4

0.790 0.007 0.728 0.790 0.758 0.753 0.931 0.745 s5

0.731 0.008 0.783 0.731 0.756 0.747 0.911 0.702 s6

0.821 0.004 0.838 0.821 0.829 0.825 0.951 0.799 s7

0.767 0.012 0.792 0.767 0.779 0.766 0.915 0.716 s8

0.689 0.008 0.724 0.689 0.706 0.698 0.901 0.630 s9

Weighted Avg. 0.830 0.036 0.830 0.830 0.830 0.794 0.929 0.807

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

150 21 6 6 23 0 0 0 3 1 | a = s0

19 1870 143 41 110 7 2 8 10 10 | b = s1

7 157 1866 87 96 6 2 8 14 7 | c = s2

5 54 74 1194 28 3 31 1 11 9 | d = s3

15 103 79 29 1735 4 1 4 9 1 | e = s4

0 3 11 4 2 166 5 6 7 6 | f = s5

1 10 10 19 8 9 263 6 25 9 | g = s6

0 5 12 3 1 11 1 197 6 4 | h = s7

3 16 15 28 6 10 21 3 414 24 | i = s8

3 9 5 19 0 12 10 2 24 186 | j = s9

E3 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| F | 0.5 | 10 | T | T |

Accuracy: 81.1042 %

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.667 0.005 0.749 0.667 0.705 0.700 0.905 0.659 s0

0.838 0.055 0.819 0.838 0.828 0.777 0.936 0.855 s1

0.805 0.056 0.814 0.805 0.809 0.752 0.930 0.842 s2

0.827 0.029 0.827 0.827 0.827 0.797 0.944 0.829 s3

0.854 0.040 0.845 0.854 0.849 0.811 0.956 0.880 s4

0.729 0.007 0.705 0.729 0.717 0.710 0.933 0.775 s5

0.711 0.010 0.731 0.711 0.721 0.711 0.933 0.737 s6

0.783 0.006 0.774 0.783 0.778 0.773 0.956 0.788 s7

0.767 0.013 0.774 0.767 0.770 0.757 0.932 0.742 s8

0.667 0.008 0.717 0.667 0.691 0.683 0.931 0.677 s9

Weighted Avg. 0.811 0.040 0.811 0.811 0.811 0.772 0.939 0.830

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

140 18 7 8 32 1 1 2 1 0 | a = s0

15 1861 150 33 124 6 6 8 11 6 | b = s1

11 184 1811 88 106 11 9 8 14 8 | c = s2

3 39 98 1166 30 5 33 5 17 14 | d = s3

12 123 89 39 1690 6 4 8 7 2 | e = s4

0 5 12 4 4 153 7 12 8 5 | f = s5

1 12 12 18 8 6 256 6 29 12 | g = s6

0 10 8 4 1 13 5 188 9 2 | h = s7

1 16 24 27 4 8 19 5 414 22 | i = s8

4 5 14 23 0 8 10 1 25 180 | j = s9

E4 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| F | 0.5 | 20 | T | T |

Accuracy: 78.5346 %

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.614 0.007 0.648 0.614 0.631 0.623 0.946 0.628 s0

0.802 0.058 0.805 0.802 0.803 0.745 0.930 0.851 s1

0.790 0.068 0.778 0.790 0.784 0.718 0.926 0.838 s2

0.818 0.038 0.786 0.818 0.802 0.767 0.949 0.832 s3

0.844 0.043 0.836 0.844 0.840 0.799 0.957 0.881 s4

0.629 0.004 0.781 0.629 0.697 0.695 0.950 0.737 s5

0.683 0.012 0.695 0.683 0.689 0.677 0.943 0.703 s6

0.704 0.003 0.841 0.704 0.766 0.764 0.962 0.778 s7

0.722 0.020 0.685 0.722 0.703 0.686 0.938 0.681 s8

0.600 0.008 0.686 0.600 0.640 0.632 0.931 0.592 s9

Weighted Avg. 0.785 0.045 0.785 0.785 0.785 0.741 0.940 0.820

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

129 24 3 4 48 0 0 0 2 0 | a = s0

20 1780 195 48 136 6 3 5 19 8 | b = s1

10 179 1777 126 101 7 12 5 22 11 | c = s2

11 42 115 1153 30 1 30 2 15 11 | d = s3

15 128 103 34 1672 4 5 3 8 8 | e = s4

0 3 18 11 4 132 5 6 26 5 | f = s5

2 15 15 31 2 2 246 5 35 7 | g = s6

1 17 11 10 4 6 4 169 16 2 | h = s7

3 13 32 30 4 9 33 4 390 22 | i = s8

8 10 14 20 0 2 16 2 36 162 | j = s9

E5 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| F | 0.8 | 5 | T | T |

Accuracy: 82.9825 %

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.714 0.006 0.732 0.714 0.723 0.717 0.904 0.653 s0

0.842 0.051 0.832 0.842 0.837 0.788 0.924 0.821 s1

0.830 0.047 0.841 0.830 0.835 0.786 0.919 0.826 s2

0.847 0.029 0.833 0.847 0.840 0.812 0.939 0.811 s3

0.876 0.035 0.865 0.876 0.870 0.837 0.947 0.860 s4

0.790 0.007 0.728 0.790 0.758 0.753 0.931 0.747 s5

0.731 0.008 0.783 0.731 0.756 0.747 0.911 0.702 s6

0.821 0.004 0.838 0.821 0.829 0.825 0.951 0.799 s7

0.769 0.012 0.790 0.769 0.779 0.767 0.915 0.715 s8

0.689 0.007 0.727 0.689 0.707 0.699 0.902 0.632 s9

Weighted Avg. 0.830 0.036 0.830 0.830 0.830 0.794 0.929 0.807

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

150 21 6 6 23 0 0 0 3 1 | a = s0

20 1869 144 41 109 7 2 8 10 10 | b = s1

8 155 1867 89 94 6 2 8 14 7 | c = s2

5 54 73 1194 28 3 31 1 12 9 | d = s3

15 105 78 29 1734 4 1 4 9 1 | e = s4

0 3 11 4 2 166 5 6 7 6 | f = s5

1 10 10 19 8 9 263 6 25 9 | g = s6

0 5 12 3 1 11 1 197 6 4 | h = s7

3 16 14 29 6 10 21 3 415 23 | i = s8

3 9 5 19 0 12 10 2 24 186 | j = s9

## Random forest results – top 20 dataset

E1 Settings:

|  |  |  |  |
| --- | --- | --- | --- |
| seed | numIterations | breakTiesRandoml | maxDepth |
| 1 | 100 | F | 0 |

Accuracy: 96.8111 %

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.829 0.000 0.994 0.829 0.904 0.906 0.999 0.979 s0

0.966 0.011 0.962 0.966 0.964 0.953 0.998 0.994 s1

0.972 0.012 0.962 0.972 0.967 0.957 0.998 0.995 s2

0.975 0.005 0.973 0.975 0.974 0.970 0.998 0.994 s3

0.988 0.009 0.966 0.988 0.977 0.971 0.999 0.997 s4

0.952 0.000 0.990 0.952 0.971 0.970 1.000 0.995 s5

0.933 0.001 0.980 0.933 0.956 0.955 0.999 0.980 s6

0.996 0.000 0.992 0.996 0.994 0.994 1.000 0.999 s7

0.969 0.001 0.983 0.969 0.976 0.974 0.999 0.992 s8

0.907 0.001 0.965 0.907 0.935 0.934 0.997 0.966 s9

Weighted Avg. 0.968 0.008 0.968 0.968 0.968 0.961 0.998 0.993

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

174 28 2 0 6 0 0 0 0 0 | a = s0

0 2145 41 5 29 0 0 0 0 0 | b = s1

0 24 2188 10 28 0 0 0 0 0 | c = s2

0 14 17 1375 3 1 0 0 0 0 | d = s3

0 10 11 2 1956 0 0 0 1 0 | e = s4

0 0 3 1 0 200 0 1 2 3 | f = s5

0 4 5 8 2 0 336 0 2 3 | g = s6

0 0 0 0 0 1 0 239 0 0 | h = s7

0 1 5 5 0 0 3 0 523 3 | i = s8

1 4 3 7 1 0 4 1 4 245 | j = s9

E2 Settings:

|  |  |  |  |
| --- | --- | --- | --- |
| seed | numIterations | breakTiesRandoml | maxDepth |
| 10 | 100 | F | 0 |

Accuracy: 96.7183 %

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.824 0.000 0.994 0.824 0.901 0.903 0.998 0.977 s0

0.964 0.012 0.959 0.964 0.962 0.950 0.998 0.993 s1

0.972 0.012 0.961 0.972 0.966 0.956 0.999 0.995 s2

0.973 0.005 0.970 0.973 0.972 0.967 0.999 0.995 s3

0.985 0.009 0.966 0.985 0.976 0.970 0.999 0.997 s4

0.971 0.000 0.981 0.971 0.976 0.976 1.000 0.994 s5

0.928 0.001 0.985 0.928 0.956 0.954 0.997 0.980 s6

0.992 0.000 0.992 0.992 0.992 0.991 1.000 0.998 s7

0.978 0.001 0.976 0.978 0.977 0.976 1.000 0.994 s8

0.904 0.000 0.988 0.904 0.944 0.943 0.997 0.967 s9

Weighted Avg. 0.967 0.008 0.967 0.967 0.967 0.960 0.999 0.993

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

173 33 2 0 2 0 0 0 0 0 | a = s0

0 2141 41 5 32 0 0 0 0 1 | b = s1

0 25 2187 10 28 0 0 0 0 0 | c = s2

0 13 18 1372 5 1 0 0 1 0 | d = s3

0 8 15 5 1951 0 0 0 1 0 | e = s4

0 0 3 0 0 204 0 1 2 0 | f = s5

0 5 5 9 0 2 334 0 3 2 | g = s6

0 1 0 0 0 1 0 238 0 0 | h = s7

0 2 3 4 1 0 2 0 528 0 | i = s8

1 4 2 9 0 0 3 1 6 244 | j = s9

E2 Settings:

|  |  |  |  |
| --- | --- | --- | --- |
| seed | numIterations | breakTiesRandoml | maxDepth |
| 10 | 50 | F | 0 |

Accuracy: 96.2951 %

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.819 0.000 0.994 0.819 0.898 0.901 0.998 0.973 s0

0.960 0.013 0.956 0.960 0.958 0.946 0.997 0.992 s1

0.971 0.013 0.957 0.971 0.964 0.953 0.998 0.993 s2

0.970 0.005 0.968 0.970 0.969 0.964 0.999 0.994 s3

0.984 0.010 0.963 0.984 0.974 0.967 0.999 0.996 s4

0.962 0.000 0.985 0.962 0.973 0.973 1.000 0.994 s5

0.919 0.001 0.968 0.919 0.943 0.941 0.997 0.977 s6

0.988 0.000 0.996 0.988 0.992 0.991 1.000 0.996 s7

0.959 0.002 0.961 0.959 0.960 0.958 0.999 0.992 s8

0.881 0.001 0.975 0.881 0.926 0.925 0.996 0.964 s9

Weighted Avg. 0.963 0.009 0.963 0.963 0.963 0.955 0.998 0.992

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

172 33 3 0 2 0 0 0 0 0 | a = s0

0 2131 46 3 37 0 0 0 2 1 | b = s1

0 23 2185 14 28 0 0 0 0 0 | c = s2

0 15 20 1368 5 0 0 0 2 0 | d = s3

0 11 15 4 1949 0 0 0 1 0 | e = s4

0 0 4 0 1 202 0 1 2 0 | f = s5

0 6 4 9 1 1 331 0 6 2 | g = s6

0 2 0 0 0 1 0 237 0 0 | h = s7

0 3 4 6 0 0 6 0 518 3 | i = s8

1 4 3 9 1 1 5 0 8 238 | j = s9

E3 Settings:

|  |  |  |  |
| --- | --- | --- | --- |
| seed | numIterations | breakTiesRandoml | maxDepth |
| 10 | 100 | T | 10 |

Accuracy: 95.129 %

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.724 0.000 1.000 0.724 0.840 0.848 0.998 0.968 s0

0.945 0.022 0.927 0.945 0.936 0.917 0.995 0.985 s1

0.958 0.018 0.941 0.958 0.950 0.934 0.997 0.991 s2

0.967 0.007 0.961 0.967 0.964 0.958 0.998 0.991 s3

0.970 0.012 0.954 0.970 0.962 0.952 0.998 0.993 s4

0.957 0.000 1.000 0.957 0.978 0.978 1.000 0.995 s5

0.922 0.001 0.982 0.922 0.951 0.950 0.996 0.975 s6

0.983 0.000 1.000 0.983 0.992 0.991 1.000 0.997 s7

0.963 0.002 0.972 0.963 0.967 0.966 0.999 0.991 s8

0.893 0.001 0.980 0.893 0.934 0.933 0.996 0.956 s9

Weighted Avg. 0.951 0.013 0.952 0.951 0.951 0.940 0.997 0.989

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

152 51 2 0 5 0 0 0 0 0 | a = s0

0 2097 70 10 43 0 0 0 0 0 | b = s1

0 39 2156 18 37 0 0 0 0 0 | c = s2

0 19 22 1363 5 0 0 0 0 1 | d = s3

0 36 17 6 1920 0 0 0 1 0 | e = s4

0 0 6 0 0 201 0 0 2 1 | f = s5

0 7 5 9 2 0 332 0 3 2 | g = s6

0 1 3 0 0 0 0 236 0 0 | h = s7

0 5 5 5 0 0 4 0 520 1 | i = s8

0 6 4 7 1 0 2 0 9 241 | j = s9

## J48 results – All dataset

E1 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| T | 0.25 | 1 | T | T |

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.776 0.006 0.738 0.776 0.756 0.751 0.885 0.577 s0

0.846 0.050 0.834 0.846 0.840 0.792 0.895 0.734 s1

0.832 0.046 0.846 0.832 0.839 0.791 0.891 0.731 s2

0.860 0.022 0.868 0.860 0.864 0.841 0.920 0.767 s3

0.879 0.032 0.877 0.879 0.878 0.846 0.923 0.793 s4

0.757 0.005 0.761 0.757 0.759 0.754 0.876 0.581 s5

0.803 0.009 0.767 0.803 0.784 0.776 0.896 0.623 s6

0.867 0.003 0.867 0.867 0.867 0.863 0.932 0.754 s7

0.824 0.009 0.838 0.824 0.831 0.821 0.907 0.700 s8

0.759 0.006 0.777 0.759 0.768 0.761 0.876 0.596 s9

Weighted Avg. 0.843 0.033 0.844 0.843 0.843 0.810 0.904 0.734

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

163 19 4 5 14 0 2 0 2 1 | a = s0

23 1878 148 34 96 9 10 4 13 5 | b = s1

9 168 1872 74 83 7 14 3 13 7 | c = s2

7 43 66 1213 33 6 21 1 11 9 | d = s3

12 109 76 29 1740 3 4 1 4 2 | e = s4

1 4 7 5 6 159 6 6 8 8 | f = s5

0 8 14 14 7 3 289 7 10 8 | g = s6

0 5 6 2 1 6 4 208 5 3 | h = s7

3 12 15 13 5 7 17 7 445 16 | i = s8

3 7 4 9 0 9 10 3 20 205 | j = s9

E2 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| F | 0.25 | 5 | T | F |

Accuracy: 83.8596%

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.714 0.006 0.721 0.714 0.718 0.711 0.905 0.683 s0

0.858 0.058 0.815 0.858 0.836 0.786 0.926 0.807 s1

0.833 0.048 0.840 0.833 0.837 0.788 0.917 0.814 s2

0.856 0.024 0.858 0.856 0.857 0.833 0.937 0.835 s3

0.863 0.032 0.874 0.863 0.869 0.835 0.941 0.839 s4

0.757 0.005 0.787 0.757 0.772 0.767 0.923 0.704 s5

0.786 0.007 0.809 0.786 0.797 0.790 0.921 0.735 s6

0.825 0.003 0.884 0.825 0.853 0.850 0.943 0.834 s7

0.815 0.008 0.851 0.815 0.833 0.823 0.929 0.794 s8

0.741 0.006 0.766 0.741 0.753 0.746 0.916 0.704 s9

Weighted Avg. 0.839 0.036 0.839 0.839 0.839 0.804 0.928 0.809

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

150 25 4 4 20 1 0 1 4 1 | a = s0

18 1905 150 38 81 7 7 2 7 5 | b = s1

9 183 1875 67 87 3 9 1 7 9 | c = s2

9 47 73 1207 31 8 15 1 8 11 | d = s3

12 129 81 36 1709 3 4 0 3 3 | e = s4

3 7 4 6 5 159 3 7 8 8 | f = s5

0 13 13 18 6 3 283 5 11 8 | g = s6

0 8 9 4 1 7 5 198 6 2 | h = s7

5 13 16 14 11 5 13 9 440 14 | i = s8

2 6 6 12 4 6 11 0 23 200 | j = s9

E3 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| F | 0.5 | 10 | T | T |

Accuracy: 81.8782%

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.681 0.006 0.726 0.681 0.703 0.697 0.920 0.672 s0

0.845 0.068 0.787 0.845 0.815 0.759 0.934 0.832 s1

0.793 0.050 0.828 0.793 0.810 0.754 0.928 0.833 s2

0.838 0.024 0.855 0.838 0.846 0.820 0.948 0.861 s3

0.858 0.037 0.856 0.858 0.857 0.821 0.952 0.863 s4

0.738 0.005 0.771 0.738 0.754 0.749 0.934 0.739 s5

0.772 0.008 0.794 0.772 0.783 0.775 0.926 0.747 s6

0.813 0.005 0.819 0.813 0.816 0.811 0.948 0.828 s7

0.802 0.012 0.793 0.802 0.797 0.785 0.940 0.795 s8

0.696 0.007 0.749 0.696 0.722 0.715 0.928 0.698 s9

Weighted Avg. 0.819 0.040 0.819 0.819 0.819 0.780 0.938 0.828

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

143 32 7 1 19 2 1 1 4 0 | a = s0

19 1877 160 46 86 5 5 8 10 4 | b = s1

7 234 1785 65 112 5 10 6 16 10 | c = s2

8 60 80 1181 35 4 16 1 11 14 | d = s3

12 133 81 34 1699 1 7 3 6 4 | e = s4

2 4 7 4 6 155 4 6 17 5 | f = s5

0 10 12 17 6 8 278 6 15 8 | g = s6

1 9 5 2 4 6 6 195 8 4 | h = s7

2 17 14 18 14 7 10 11 433 14 | i = s8

3 8 6 14 3 8 13 1 26 188 | j = s9

E4 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| F | 0.5 | 20 | T | T |

Accuracy: 79.5459%

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.605 0.006 0.676 0.605 0.638 0.632 0.929 0.605 s0

0.815 0.064 0.791 0.815 0.803 0.743 0.937 0.831 s1

0.792 0.061 0.796 0.792 0.794 0.732 0.932 0.843 s2

0.820 0.035 0.798 0.820 0.809 0.776 0.948 0.841 s3

0.844 0.042 0.839 0.844 0.841 0.801 0.955 0.868 s4

0.662 0.005 0.732 0.662 0.695 0.689 0.944 0.721 s5

0.719 0.010 0.736 0.719 0.728 0.717 0.944 0.735 s6

0.750 0.004 0.814 0.750 0.781 0.776 0.956 0.786 s7

0.746 0.014 0.760 0.746 0.753 0.739 0.950 0.782 s8

0.670 0.006 0.751 0.670 0.708 0.702 0.939 0.698 s9

Weighted Avg. 0.795 0.044 0.795 0.795 0.795 0.752 0.943 0.824

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

127 31 7 8 25 2 3 1 6 0 | a = s0

13 1809 177 62 130 8 6 1 11 3 | b = s1

8 204 1783 96 107 9 4 8 14 17 | c = s2

14 54 110 1156 34 2 25 4 7 4 | d = s3

12 135 103 33 1671 3 8 1 8 6 | e = s4

3 4 15 9 2 139 7 8 18 5 | f = s5

2 16 10 35 10 2 259 6 15 5 | g = s6

2 7 9 2 1 8 9 180 16 6 | h = s7

3 21 16 30 11 11 20 11 403 14 | i = s8

4 7 9 18 1 6 11 1 32 181 | j = s9

E5 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| F | 0.8 | 5 | T | T |

Accuracy: 83.8493%

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.714 0.006 0.721 0.714 0.718 0.711 0.905 0.683 s0

0.858 0.058 0.815 0.858 0.836 0.786 0.926 0.806 s1

0.833 0.048 0.840 0.833 0.837 0.787 0.917 0.814 s2

0.858 0.024 0.858 0.858 0.858 0.834 0.938 0.836 s3

0.861 0.032 0.874 0.861 0.868 0.834 0.940 0.840 s4

0.757 0.005 0.787 0.757 0.772 0.767 0.923 0.704 s5

0.786 0.007 0.809 0.786 0.797 0.790 0.921 0.735 s6

0.825 0.003 0.884 0.825 0.853 0.850 0.943 0.834 s7

0.817 0.008 0.851 0.817 0.834 0.824 0.930 0.795 s8

0.741 0.006 0.766 0.741 0.753 0.746 0.916 0.704 s9

Weighted Avg. 0.838 0.036 0.839 0.838 0.838 0.804 0.928 0.809

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

150 25 4 4 20 1 0 1 4 1 | a = s0

18 1905 151 38 80 7 7 2 7 5 | b = s1

9 183 1874 68 87 3 9 1 7 9 | c = s2

9 44 73 1210 31 8 15 1 8 11 | d = s3

12 134 80 36 1705 3 4 0 3 3 | e = s4

3 7 4 6 5 159 3 7 8 8 | f = s5

0 13 13 18 6 3 283 5 11 8 | g = s6

0 8 9 4 1 7 5 198 6 2 | h = s7

5 12 16 14 11 5 13 9 441 14 | i = s8

2 6 6 12 4 6 11 0 23 200 | j = s9

## Random forest results – ALL dataset

E1 Settings:

|  |  |  |  |
| --- | --- | --- | --- |
| seed | numIterations | breakTiesRandoml | maxDepth |
| 1 | 100 | F | 0 |

Accuracy: 96.8111 %

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.814 0.000 1.000 0.814 0.898 0.901 0.999 0.989 s0

0.973 0.016 0.947 0.973 0.960 0.948 0.997 0.991 s1

0.970 0.015 0.952 0.970 0.961 0.949 0.998 0.993 s2

0.965 0.003 0.985 0.965 0.975 0.970 0.997 0.991 s3

0.980 0.006 0.977 0.980 0.979 0.973 0.999 0.997 s4

0.948 0.000 1.000 0.948 0.973 0.973 1.000 0.993 s5

0.944 0.000 0.988 0.944 0.966 0.965 0.999 0.986 s6

0.983 0.000 0.996 0.983 0.990 0.989 1.000 0.999 s7

0.978 0.001 0.981 0.978 0.980 0.978 1.000 0.997 s8

0.948 0.000 0.988 0.948 0.968 0.967 0.998 0.986 s9

Weighted Avg. 0.967 0.009 0.968 0.967 0.967 0.960 0.998 0.993

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

171 34 3 0 2 0 0 0 0 0 | a = s0

0 2161 45 4 10 0 0 0 0 0 | b = s1

0 36 2183 9 22 0 0 0 0 0 | c = s2

0 14 31 1360 5 0 0 0 0 0 | d = s3

0 19 19 2 1940 0 0 0 0 0 | e = s4

0 1 5 0 1 199 0 1 3 0 | f = s5

0 6 3 3 3 0 340 0 4 1 | g = s6

0 0 1 0 1 0 0 236 2 0 | h = s7

0 6 0 0 1 0 3 0 528 2 | i = s8

0 6 3 3 0 0 1 0 1 256 | j = s9

E2 Settings:

|  |  |  |  |
| --- | --- | --- | --- |
| seed | numIterations | breakTiesRandoml | maxDepth |
| 10 | 100 | F | 0 |

Accuracy: 96.7905%

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.810 0.000 1.000 0.810 0.895 0.898 1.000 0.992 s0

0.973 0.018 0.941 0.973 0.957 0.944 0.997 0.991 s1

0.972 0.012 0.961 0.972 0.966 0.956 0.998 0.993 s2

0.962 0.004 0.976 0.962 0.969 0.964 0.997 0.990 s3

0.981 0.004 0.983 0.981 0.982 0.978 0.999 0.997 s4

0.933 0.000 0.995 0.933 0.963 0.963 1.000 0.991 s5

0.942 0.000 0.988 0.942 0.964 0.963 0.997 0.986 s6

0.988 0.000 0.996 0.988 0.992 0.991 1.000 1.000 s7

0.981 0.001 0.983 0.981 0.982 0.981 1.000 0.997 s8

0.959 0.001 0.981 0.959 0.970 0.969 0.999 0.988 s9

Weighted Avg. 0.968 0.008 0.968 0.968 0.968 0.961 0.998 0.993

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

170 36 4 0 0 0 0 0 0 0 | a = s0

0 2161 38 7 14 0 0 0 0 0 | b = s1

0 37 2187 16 10 0 0 0 0 0 | c = s2

0 21 27 1357 5 0 0 0 0 0 | d = s3

0 20 14 3 1943 0 0 0 0 0 | e = s4

0 1 4 0 1 196 0 1 4 3 | f = s5

0 8 2 5 1 1 339 0 3 1 | g = s6

0 0 0 0 1 0 1 237 1 0 | h = s7

0 7 0 0 1 0 1 0 530 1 | i = s8

0 6 0 2 0 0 2 0 1 259 | j = s9

E2 Settings:

|  |  |  |  |
| --- | --- | --- | --- |
| seed | numIterations | breakTiesRandoml | maxDepth |
| 10 | 50 | F | 0 |

Accuracy: 96.0784%

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.824 0.000 1.000 0.824 0.903 0.906 0.999 0.981 s0

0.971 0.023 0.927 0.971 0.949 0.933 0.996 0.987 s1

0.965 0.015 0.950 0.965 0.957 0.944 0.996 0.990 s2

0.951 0.004 0.975 0.951 0.963 0.957 0.997 0.987 s3

0.968 0.005 0.981 0.968 0.974 0.968 0.999 0.995 s4

0.924 0.000 0.995 0.924 0.958 0.958 0.999 0.984 s5

0.939 0.001 0.983 0.939 0.960 0.959 0.997 0.981 s6

0.983 0.000 0.987 0.983 0.985 0.985 1.000 0.999 s7

0.981 0.001 0.983 0.981 0.982 0.981 1.000 0.995 s8

0.944 0.000 0.985 0.944 0.964 0.963 0.999 0.985 s9

Weighted Avg. 0.961 0.010 0.962 0.961 0.961 0.952 0.997 0.990

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

173 33 3 0 1 0 0 0 0 0 | a = s0

0 2156 41 7 16 0 0 0 0 0 | b = s1

0 53 2171 14 12 0 0 0 0 0 | c = s2

0 25 38 1341 6 0 0 0 0 0 | d = s3

0 40 19 5 1916 0 0 0 0 0 | e = s4

0 1 7 0 1 194 0 2 3 2 | f = s5

0 6 4 5 1 1 338 0 3 2 | g = s6

0 1 0 0 1 0 1 236 1 0 | h = s7

0 5 0 1 0 0 3 1 530 0 | i = s8

0 6 2 3 0 0 2 0 2 255 | j = s9

E4 Settings:

|  |  |  |  |
| --- | --- | --- | --- |
| seed | numIterations | breakTiesRandoml | maxDepth |
| 10 | 100 | T | 10 |

Accuracy: 92.064%

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.424 0.000 1.000 0.424 0.595 0.647 0.998 0.965 s0

0.942 0.048 0.853 0.942 0.895 0.864 0.986 0.963 s1

0.924 0.030 0.902 0.924 0.913 0.886 0.991 0.977 s2

0.917 0.008 0.953 0.917 0.935 0.924 0.993 0.978 s3

0.932 0.012 0.953 0.932 0.942 0.928 0.995 0.984 s4

0.895 0.000 1.000 0.895 0.945 0.945 1.000 0.991 s5

0.933 0.001 0.980 0.933 0.956 0.955 0.998 0.985 s6

0.971 0.000 0.987 0.971 0.979 0.978 1.000 0.998 s7

0.969 0.002 0.974 0.969 0.971 0.970 1.000 0.995 s8

0.904 0.000 0.988 0.904 0.944 0.943 0.999 0.978 s9

Weighted Avg. 0.921 0.022 0.924 0.921 0.919 0.902 0.992 0.977

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

89 116 4 0 1 0 0 0 0 0 | a = s0

0 2091 78 11 40 0 0 0 0 0 | b = s1

0 100 2078 37 35 0 0 0 0 0 | c = s2

0 37 70 1293 10 0 0 0 0 0 | d = s3

0 82 48 4 1846 0 0 0 0 0 | e = s4

0 1 9 0 2 188 1 3 4 2 | f = s5

0 6 4 6 2 0 336 0 6 0 | g = s6

0 1 1 0 1 0 1 233 3 0 | h = s7

0 8 2 2 1 0 3 0 523 1 | i = s8

0 10 9 4 0 0 2 0 1 244 | j = s9

## J48 results – All dataset with TEST set

E1 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| T | 0.25 | 1 | T | T |

69.0291 %

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.233 0.007 0.400 0.233 0.295 0.295 0.613 0.111 s0

0.758 0.093 0.713 0.758 0.735 0.652 0.830 0.589 s1

0.743 0.090 0.725 0.743 0.734 0.647 0.808 0.583 s2

0.702 0.047 0.717 0.702 0.709 0.660 0.821 0.499 s3

0.708 0.070 0.733 0.708 0.720 0.646 0.811 0.540 s4

0.650 0.013 0.494 0.650 0.561 0.557 0.818 0.328 s5

0.533 0.013 0.545 0.533 0.539 0.526 0.758 0.261 s6

0.217 0.016 0.210 0.217 0.213 0.197 0.597 0.061 s7

0.607 0.014 0.684 0.607 0.643 0.627 0.787 0.434 s8

0.467 0.013 0.519 0.467 0.491 0.477 0.724 0.252 s9

Weighted Avg. 0.690 0.068 0.689 0.690 0.688 0.623 0.803 0.512

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

14 8 3 2 31 0 2 0 0 0 | a = s0

5 546 83 7 51 2 4 12 6 4 | b = s1

1 77 557 56 44 1 3 8 1 2 | c = s2

1 39 45 316 27 0 7 0 9 6 | d = s3

11 86 49 26 467 4 9 2 3 3 | e = s4

0 3 0 9 2 39 4 1 0 2 | f = s5

0 0 9 1 1 1 48 0 14 16 | g = s6

0 1 11 13 1 17 2 13 1 1 | h = s7

2 5 7 6 10 13 9 2 91 5 | i = s8

1 1 4 5 3 2 0 24 8 42 | j = s9

E2 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| F | 0.25 | 5 | T | F |

Accuracy: 70.356%

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.217 0.012 0.271 0.217 0.241 0.229 0.643 0.195 s0

0.788 0.096 0.713 0.788 0.749 0.669 0.879 0.704 s1

0.788 0.095 0.727 0.788 0.756 0.675 0.867 0.700 s2

0.718 0.043 0.741 0.718 0.729 0.684 0.871 0.687 s3

0.671 0.066 0.733 0.671 0.701 0.625 0.842 0.643 s4

0.600 0.006 0.679 0.600 0.637 0.632 0.875 0.493 s5

0.522 0.010 0.610 0.522 0.563 0.553 0.843 0.468 s6

0.350 0.016 0.309 0.350 0.328 0.315 0.735 0.147 s7

0.593 0.010 0.748 0.593 0.662 0.651 0.805 0.546 s8

0.489 0.011 0.571 0.489 0.527 0.516 0.764 0.508 s9

Weighted Avg. 0.704 0.068 0.702 0.704 0.701 0.638 0.852 0.642

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

13 8 2 4 32 0 1 0 0 0 | a = s0

9 567 74 8 50 0 7 2 1 2 | b = s1

0 72 591 36 39 1 2 8 0 1 | c = s2

2 41 47 323 24 0 4 0 7 2 | d = s3

24 95 58 27 443 0 6 1 2 4 | e = s4

0 3 0 13 2 36 0 5 0 1 | f = s5

0 0 12 2 1 0 47 0 14 14 | g = s6

0 1 16 15 1 2 1 21 1 2 | h = s7

0 7 8 5 10 11 8 5 89 7 | i = s8

0 1 5 3 2 3 1 26 5 44 | j = s9

E3 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| F | 0.5 | 10 | T | T |

Accuracy: 70.0647%

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.350 0.010 0.412 0.350 0.378 0.368 0.736 0.336 s0

0.797 0.103 0.701 0.797 0.746 0.665 0.902 0.792 s1

0.747 0.079 0.752 0.747 0.749 0.669 0.888 0.785 s2

0.696 0.034 0.775 0.696 0.733 0.692 0.892 0.721 s3

0.712 0.084 0.697 0.712 0.705 0.623 0.871 0.645 s4

0.450 0.006 0.587 0.450 0.509 0.506 0.770 0.420 s5

0.589 0.013 0.570 0.589 0.579 0.566 0.876 0.540 s6

0.333 0.019 0.253 0.333 0.288 0.274 0.788 0.191 s7

0.553 0.011 0.716 0.553 0.624 0.613 0.849 0.602 s8

0.489 0.006 0.698 0.489 0.575 0.574 0.788 0.541 s9

Weighted Avg. 0.701 0.068 0.704 0.701 0.700 0.636 0.876 0.697

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

21 12 0 6 21 0 0 0 0 0 | a = s0

10 574 59 9 49 1 1 14 2 1 | b = s1

0 81 560 33 63 0 1 9 2 1 | c = s2

5 44 45 313 32 0 6 0 5 0 | d = s3

14 95 46 16 470 1 9 6 1 2 | e = s4

0 0 0 4 26 27 0 3 0 0 | f = s5

0 3 7 3 0 1 53 0 14 9 | g = s6

0 3 11 9 2 2 10 20 3 0 | h = s7

1 6 10 11 10 8 13 2 83 6 | i = s8

0 1 7 0 1 6 0 25 6 44 | j = s9

E4 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| F | 0.5 | 20 | T | T |

Accuracy: 68.4142%

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.400 0.005 0.615 0.400 0.485 0.488 0.802 0.455 s0

0.736 0.088 0.717 0.736 0.727 0.642 0.910 0.767 s1

0.757 0.104 0.700 0.757 0.728 0.637 0.864 0.765 s2

0.707 0.045 0.726 0.707 0.716 0.669 0.861 0.732 s3

0.735 0.078 0.720 0.735 0.727 0.652 0.879 0.646 s4

0.400 0.018 0.300 0.400 0.343 0.332 0.751 0.384 s5

0.522 0.008 0.662 0.522 0.584 0.577 0.878 0.541 s6

0.200 0.023 0.146 0.200 0.169 0.152 0.631 0.149 s7

0.427 0.011 0.660 0.427 0.518 0.512 0.850 0.532 s8

0.467 0.006 0.712 0.467 0.564 0.566 0.778 0.411 s9

Weighted Avg. 0.684 0.071 0.689 0.684 0.683 0.618 0.867 0.682

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

24 9 3 8 16 0 0 0 0 0 | a = s0

4 530 87 12 54 14 3 14 2 0 | b = s1

6 70 568 42 51 3 1 8 0 1 | c = s2

2 36 47 318 32 0 6 0 9 0 | d = s3

3 80 50 32 485 2 5 1 0 2 | e = s4

0 0 0 4 26 24 0 4 2 0 | f = s5

0 3 14 1 1 4 47 3 8 9 | g = s6

0 8 18 9 0 13 0 12 0 0 | h = s7

0 3 19 10 9 16 9 15 64 5 | i = s8

0 0 5 2 0 4 0 25 12 42 | j = s9

E5 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| F | 0.8 | 5 | T | T |

Accuracy: 70.1618

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.217 0.012 0.271 0.217 0.241 0.229 0.643 0.195 s0

0.785 0.096 0.712 0.785 0.747 0.666 0.876 0.705 s1

0.783 0.097 0.722 0.783 0.751 0.668 0.866 0.701 s2

0.718 0.042 0.746 0.718 0.732 0.687 0.873 0.702 s3

0.671 0.068 0.727 0.671 0.698 0.621 0.843 0.633 s4

0.600 0.006 0.679 0.600 0.637 0.632 0.875 0.492 s5

0.522 0.010 0.610 0.522 0.563 0.553 0.837 0.468 s6

0.350 0.016 0.309 0.350 0.328 0.315 0.735 0.146 s7

0.593 0.010 0.748 0.593 0.662 0.651 0.805 0.546 s8

0.489 0.011 0.571 0.489 0.527 0.516 0.764 0.508 s9

Weighted Avg. 0.702 0.068 0.700 0.702 0.699 0.636 0.851 0.643

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

13 8 2 4 32 0 1 0 0 0 | a = s0

9 565 76 8 50 0 7 2 1 2 | b = s1

0 72 587 35 44 1 2 8 0 1 | c = s2

2 41 47 323 24 0 4 0 7 2 | d = s3

24 95 59 26 443 0 6 1 2 4 | e = s4

0 3 0 13 2 36 0 5 0 1 | f = s5

0 0 12 2 1 0 47 0 14 14 | g = s6

0 1 16 15 1 2 1 21 1 2 | h = s7

0 7 8 5 10 11 8 5 89 7 | i = s8

0 1 6 2 2 3 1 26 5 44 | j = s9

## Random forest results – ALL dataset with TEST set

E1 Settings:

|  |  |  |  |
| --- | --- | --- | --- |
| seed | numIterations | breakTiesRandoml | maxDepth |
| 1 | 100 | F | 0 |

Accuracy: 66.6667%

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.017 0.000 1.000 0.017 0.033 0.128 0.860 0.310 s0

0.794 0.124 0.660 0.794 0.721 0.630 0.924 0.824 s1

0.671 0.113 0.656 0.671 0.663 0.554 0.867 0.736 s2

0.620 0.072 0.596 0.620 0.608 0.540 0.863 0.629 s3

0.735 0.066 0.751 0.735 0.743 0.674 0.905 0.815 s4

0.500 0.002 0.811 0.500 0.619 0.631 0.981 0.663 s5

0.378 0.007 0.618 0.378 0.469 0.472 0.951 0.404 s6

0.250 0.001 0.789 0.250 0.380 0.439 0.977 0.419 s7

0.667 0.021 0.621 0.667 0.643 0.625 0.974 0.746 s8

0.456 0.009 0.594 0.456 0.516 0.508 0.801 0.524 s9

Weighted Avg. 0.667 0.083 0.676 0.667 0.656 0.586 0.898 0.727

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

1 34 11 0 14 0 0 0 0 0 | a = s0

0 572 100 11 37 0 0 0 0 0 | b = s1

0 75 503 117 55 0 0 0 0 0 | c = s2

0 64 82 279 23 0 2 0 0 0 | d = s3

0 99 38 38 485 0 0 0 0 0 | e = s4

0 0 0 2 28 30 0 0 0 0 | f = s5

0 6 5 1 0 0 34 4 25 15 | g = s6

0 1 19 5 3 7 5 15 0 5 | h = s7

0 6 6 15 1 0 14 0 100 8 | i = s8

0 10 3 0 0 0 0 0 36 41 | j = s9

E2 Settings:

|  |  |  |  |
| --- | --- | --- | --- |
| seed | numIterations | breakTiesRandoml | maxDepth |
| 10 | 100 | F | 0 |

Accuracy: 65.4369%

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.033 0.000 1.000 0.033 0.065 0.181 0.863 0.366 s0

0.767 0.134 0.634 0.767 0.694 0.595 0.914 0.800 s1

0.659 0.107 0.664 0.659 0.661 0.553 0.872 0.737 s2

0.633 0.077 0.583 0.633 0.607 0.537 0.871 0.648 s3

0.721 0.071 0.733 0.721 0.727 0.654 0.902 0.818 s4

0.467 0.002 0.800 0.467 0.589 0.605 0.994 0.832 s5

0.356 0.010 0.525 0.356 0.424 0.418 0.964 0.430 s6

0.200 0.003 0.571 0.200 0.296 0.331 0.949 0.335 s7

0.667 0.018 0.649 0.667 0.658 0.640 0.970 0.742 s8

0.456 0.008 0.631 0.456 0.529 0.524 0.870 0.539 s9

Weighted Avg. 0.654 0.085 0.662 0.654 0.644 0.571 0.899 0.728

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

2 38 10 0 10 0 0 0 0 0 | a = s0

0 552 113 18 37 0 0 0 0 0 | b = s1

0 77 494 120 59 0 0 0 0 0 | c = s2

0 65 81 285 17 0 2 0 0 0 | d = s3

0 115 28 40 476 0 0 0 0 1 | e = s4

0 0 0 3 28 28 0 1 0 0 | f = s5

0 4 3 4 0 2 32 8 20 17 | g = s6

0 3 5 2 21 5 10 12 0 2 | h = s7

0 8 3 17 1 0 17 0 100 4 | i = s8

0 8 7 0 0 0 0 0 34 41 | j = s9

E2 Settings:

|  |  |  |  |
| --- | --- | --- | --- |
| seed | numIterations | breakTiesRandoml | maxDepth |
| 10 | 50 | F | 0 |

Accuracy: 64.0129%

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.050 0.000 0.750 0.050 0.094 0.191 0.808 0.231 s0

0.750 0.141 0.619 0.750 0.678 0.572 0.907 0.769 s1

0.659 0.123 0.632 0.659 0.645 0.528 0.866 0.723 s2

0.589 0.073 0.580 0.589 0.584 0.513 0.869 0.646 s3

0.718 0.077 0.717 0.718 0.718 0.641 0.895 0.802 s4

0.483 0.003 0.763 0.483 0.592 0.601 0.989 0.763 s5

0.300 0.008 0.540 0.300 0.386 0.390 0.951 0.375 s6

0.167 0.002 0.625 0.167 0.263 0.317 0.936 0.252 s7

0.633 0.019 0.629 0.633 0.631 0.612 0.964 0.709 s8

0.456 0.006 0.707 0.456 0.554 0.557 0.888 0.526 s9

Weighted Avg. 0.640 0.091 0.643 0.640 0.629 0.552 0.893 0.705

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

3 31 9 0 17 0 0 0 0 0 | a = s0

0 540 126 15 39 0 0 0 0 0 | b = s1

0 79 494 113 64 0 0 0 0 0 | c = s2

0 80 88 265 16 0 0 0 1 0 | d = s3

1 112 37 35 474 0 0 0 1 0 | e = s4

0 0 0 4 26 29 0 0 1 0 | f = s5

0 6 6 5 0 4 27 6 23 13 | g = s6

0 2 5 2 24 5 11 10 0 1 | h = s7

0 15 6 18 1 0 12 0 95 3 | i = s8

0 8 11 0 0 0 0 0 30 41 | j = s9

E4 Settings:

|  |  |  |  |
| --- | --- | --- | --- |
| seed | numIterations | breakTiesRandoml | maxDepth |
| 10 | 100 | T | 10 |

Accuracy: 65.2104%

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.000 0.000 ? 0.000 ? ? 0.885 0.201 s0

0.789 0.146 0.621 0.789 0.695 0.595 0.916 0.809 s1

0.624 0.103 0.660 0.624 0.642 0.531 0.863 0.723 s2

0.651 0.075 0.597 0.651 0.623 0.556 0.859 0.661 s3

0.705 0.073 0.724 0.705 0.714 0.638 0.898 0.809 s4

0.333 0.000 0.952 0.333 0.494 0.559 0.994 0.858 s5

0.411 0.009 0.578 0.411 0.481 0.475 0.957 0.447 s6

0.300 0.001 0.818 0.300 0.439 0.490 0.974 0.457 s7

0.687 0.021 0.628 0.687 0.656 0.638 0.973 0.754 s8

0.478 0.007 0.683 0.478 0.562 0.561 0.823 0.548 s9

Weighted Avg. 0.652 0.087 ? 0.652 ? ? 0.894 0.728

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

0 41 10 0 9 0 0 0 0 0 | a = s0

0 568 101 8 43 0 0 0 0 0 | b = s1

0 84 468 134 64 0 0 0 0 0 | c = s2

0 71 63 293 21 0 2 0 0 0 | d = s3

0 117 42 36 465 0 0 0 0 0 | e = s4

0 0 0 3 27 20 0 3 6 1 | f = s5

0 10 5 2 0 0 37 1 24 11 | g = s6

0 3 10 2 11 1 11 18 0 4 | h = s7

0 10 4 13 2 0 14 0 103 4 | i = s8

0 10 6 0 0 0 0 0 31 43 | j = s9

## J48 results – All dataset with TEST4000 set

E1 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| T | 0.25 | 1 | T | T |

Correctly Classified Instances 5774 81.4386 %

Incorrectly Classified Instances 1316 18.5614 %

Kappa statistic 0.7729

Mean absolute error 0.0376

Root mean squared error 0.1923

Relative absolute error 22.9184 %

Root relative squared error 67.225 %

Total Number of Instances 7090

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.732 0.012 0.765 0.732 0.748 0.735 0.858 0.577 s8

0.617 0.011 0.621 0.617 0.619 0.608 0.806 0.397 s9

0.837 0.056 0.822 0.837 0.829 0.775 0.888 0.728 s2

0.866 0.039 0.855 0.866 0.860 0.823 0.908 0.733 s4

0.815 0.033 0.806 0.815 0.810 0.778 0.884 0.642 s3

0.835 0.047 0.845 0.835 0.840 0.791 0.877 0.716 s1

0.695 0.007 0.676 0.695 0.685 0.679 0.841 0.477 s0

0.693 0.005 0.732 0.693 0.712 0.707 0.844 0.514 s5

0.760 0.009 0.760 0.760 0.760 0.752 0.876 0.586 s6

0.599 0.005 0.702 0.599 0.646 0.642 0.797 0.429 s7

Weighted Avg. 0.814 0.039 0.814 0.814 0.814 0.776 0.881 0.676

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

267 21 11 9 10 13 4 6 15 9 | a = s8

25 121 18 0 10 4 0 2 6 10 | b = s9

7 3 1404 77 71 95 6 5 7 3 | c = s2

4 9 35 1288 34 82 20 4 11 1 | d = s4

9 9 85 31 837 41 1 5 6 3 | e = s3

9 7 121 75 33 1383 13 2 9 5 | f = s1

3 1 14 15 0 10 98 0 0 0 | g = s0

0 9 2 2 29 0 0 104 0 4 | h = s5

22 14 2 1 8 6 1 4 187 1 | i = s6

3 1 17 9 7 3 2 10 5 85 | j = s7

E2 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| F | 0.25 | 5 | T | F |

Correctly Classified Instances 5592 78.8717 %

Incorrectly Classified Instances 1498 21.1283 %

Kappa statistic 0.7411

Mean absolute error 0.0457

Root mean squared error 0.1881

Relative absolute error 27.8182 %

Root relative squared error 65.7525 %

Total Number of Instances 7090

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.721 0.015 0.717 0.721 0.719 0.703 0.899 0.677 s8

0.663 0.014 0.570 0.663 0.613 0.603 0.848 0.570 s9

0.815 0.068 0.788 0.815 0.801 0.738 0.902 0.797 s2

0.810 0.042 0.836 0.810 0.823 0.777 0.915 0.749 s4

0.797 0.030 0.821 0.797 0.809 0.777 0.904 0.765 s3

0.824 0.068 0.786 0.824 0.805 0.744 0.906 0.762 s1

0.645 0.004 0.765 0.645 0.700 0.697 0.843 0.607 s0

0.653 0.005 0.726 0.653 0.688 0.682 0.850 0.649 s5

0.687 0.007 0.768 0.687 0.725 0.717 0.884 0.644 s6

0.592 0.004 0.771 0.592 0.669 0.669 0.821 0.584 s7

Weighted Avg. 0.789 0.047 0.790 0.789 0.788 0.744 0.900 0.745

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

263 22 21 7 13 10 3 5 17 4 | a = s8

15 130 21 0 17 2 0 7 3 1 | b = s9

16 8 1367 73 66 138 2 0 7 1 | c = s2

6 9 63 1206 41 136 17 2 7 1 | d = s4

12 12 93 23 819 53 0 5 5 5 | e = s3

7 16 137 79 26 1365 3 13 7 4 | f = s1

3 0 12 16 3 12 91 1 1 2 | g = s0

3 10 3 28 4 0 0 98 1 3 | h = s5

28 15 4 4 9 10 1 2 169 4 | i = s6

14 6 14 7 0 10 2 2 3 84 | j = s7

E3 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| F | 0.5 | 10 | T | T |

Correctly Classified Instances 5478 77.2638 %

Incorrectly Classified Instances 1612 22.7362 %

Kappa statistic 0.722

Mean absolute error 0.0541

Root mean squared error 0.1898

Relative absolute error 32.941 %

Root relative squared error 66.3224 %

Total Number of Instances 7090

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.688 0.012 0.758 0.688 0.721 0.708 0.905 0.715 s8

0.704 0.021 0.489 0.704 0.577 0.573 0.873 0.543 s9

0.800 0.064 0.794 0.800 0.797 0.734 0.930 0.815 s2

0.804 0.047 0.820 0.804 0.812 0.763 0.916 0.794 s4

0.797 0.038 0.781 0.797 0.789 0.753 0.922 0.829 s3

0.797 0.067 0.785 0.797 0.791 0.727 0.918 0.816 s1

0.603 0.008 0.590 0.603 0.596 0.588 0.833 0.537 s0

0.567 0.001 0.924 0.567 0.702 0.719 0.855 0.631 s5

0.642 0.011 0.669 0.642 0.656 0.644 0.885 0.621 s6

0.577 0.006 0.661 0.577 0.617 0.611 0.818 0.526 s7

Weighted Avg. 0.773 0.048 0.777 0.773 0.773 0.726 0.913 0.778

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

251 44 14 11 6 7 4 2 20 6 | a = s8

19 138 15 5 4 4 0 0 5 6 | b = s9

11 7 1342 81 79 136 8 0 12 2 | c = s2

5 8 74 1197 56 113 26 1 7 1 | d = s4

9 17 72 22 819 68 3 1 14 2 | e = s3

5 30 133 97 47 1321 13 1 7 3 | f = s1

1 1 12 22 5 14 85 0 0 1 | g = s0

4 14 1 8 24 4 0 85 4 6 | h = s5

17 20 12 10 7 6 1 0 158 15 | i = s6

9 3 15 6 2 10 4 2 9 82 | j = s7

E4 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| F | 0.5 | 20 | T | T |

Correctly Classified Instances 5267 74.2877 %

Incorrectly Classified Instances 1823 25.7123 %

Kappa statistic 0.6862

Mean absolute error 0.0654

Root mean squared error 0.196

Relative absolute error 39.8239 %

Root relative squared error 68.491 %

Total Number of Instances 7090

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.605 0.021 0.612 0.605 0.609 0.588 0.925 0.635 s8

0.628 0.020 0.469 0.628 0.537 0.528 0.899 0.481 s9

0.751 0.059 0.797 0.751 0.773 0.706 0.920 0.822 s2

0.776 0.055 0.788 0.776 0.782 0.725 0.932 0.805 s4

0.777 0.052 0.716 0.777 0.745 0.701 0.925 0.786 s3

0.798 0.069 0.780 0.798 0.789 0.723 0.931 0.838 s1

0.447 0.005 0.643 0.447 0.527 0.528 0.877 0.472 s0

0.627 0.002 0.870 0.627 0.729 0.734 0.954 0.764 s5

0.618 0.019 0.545 0.618 0.579 0.564 0.903 0.560 s6

0.563 0.007 0.625 0.563 0.593 0.586 0.862 0.447 s7

Weighted Avg. 0.743 0.052 0.747 0.743 0.743 0.692 0.923 0.773

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

221 52 4 14 31 12 4 0 15 12 | a = s8

31 123 12 1 16 4 2 0 3 4 | b = s9

31 13 1260 105 105 126 3 0 23 12 | c = s2

13 6 43 1154 68 146 12 12 30 4 | d = s4

11 16 100 15 798 52 5 2 27 1 | e = s3

5 3 131 117 56 1322 6 0 10 7 | f = s1

9 0 7 36 8 17 63 0 0 1 | g = s0

3 13 11 8 5 3 0 94 11 2 | h = s5

23 29 9 6 17 5 0 0 152 5 | i = s6

14 7 4 8 10 8 3 0 8 80 | j = s7

E5 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| F | 0.8 | 5 | T | T |

Correctly Classified Instances 5605 79.055 %

Incorrectly Classified Instances 1485 20.945 %

Kappa statistic 0.7433

Mean absolute error 0.0458

Root mean squared error 0.1883

Relative absolute error 27.8564 %

Root relative squared error 65.8264 %

Total Number of Instances 7090

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.721 0.013 0.745 0.721 0.733 0.718 0.899 0.677 s8

0.663 0.014 0.570 0.663 0.613 0.603 0.848 0.570 s9

0.815 0.066 0.793 0.815 0.804 0.742 0.901 0.795 s2

0.815 0.044 0.831 0.815 0.823 0.777 0.915 0.748 s4

0.800 0.030 0.818 0.800 0.809 0.777 0.903 0.763 s3

0.825 0.068 0.787 0.825 0.806 0.745 0.906 0.769 s1

0.645 0.004 0.765 0.645 0.700 0.697 0.843 0.607 s0

0.653 0.005 0.726 0.653 0.688 0.682 0.850 0.649 s5

0.687 0.007 0.768 0.687 0.725 0.717 0.880 0.641 s6

0.592 0.004 0.771 0.592 0.669 0.669 0.821 0.584 s7

Weighted Avg. 0.791 0.047 0.791 0.791 0.790 0.746 0.899 0.746

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

263 22 20 8 13 10 3 5 17 4 | a = s8

15 130 21 0 17 2 0 7 3 1 | b = s9

13 8 1368 74 67 138 2 0 7 1 | c = s2

3 9 60 1213 43 133 17 2 7 1 | d = s4

6 12 95 23 822 54 0 5 5 5 | e = s3

7 16 128 86 26 1367 3 13 7 4 | f = s1

3 0 12 16 3 12 91 1 1 2 | g = s0

3 10 3 28 4 0 0 98 1 3 | h = s5

26 15 5 4 10 10 1 2 169 4 | i = s6

14 6 14 7 0 10 2 2 3 84 | j = s7

## Random forest results – ALL dataset with TEST4000 set

E1 Settings:

|  |  |  |  |
| --- | --- | --- | --- |
| seed | numIterations | breakTiesRandoml | maxDepth |
| 1 | 100 | F | 0 |

Correctly Classified Instances 5847 82.4683 %

Incorrectly Classified Instances 1243 17.5317 %

Kappa statistic 0.784

Mean absolute error 0.0799

Root mean squared error 0.1737

Relative absolute error 48.6118 %

Root relative squared error 60.6975 %

Total Number of Instances 7090

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.855 0.008 0.855 0.855 0.855 0.847 0.992 0.930 s8

0.724 0.003 0.882 0.724 0.796 0.794 0.942 0.804 s9

0.818 0.060 0.808 0.818 0.813 0.755 0.964 0.913 s2

0.861 0.043 0.842 0.861 0.851 0.812 0.971 0.934 s4

0.810 0.030 0.819 0.810 0.814 0.783 0.959 0.893 s3

0.878 0.071 0.789 0.878 0.831 0.778 0.974 0.936 s1

0.468 0.000 1.000 0.468 0.638 0.681 0.948 0.745 s0

0.740 0.001 0.917 0.740 0.819 0.821 0.998 0.933 s5

0.732 0.003 0.896 0.732 0.805 0.803 0.988 0.861 s6

0.676 0.000 0.970 0.676 0.797 0.807 0.990 0.825 s7

Weighted Avg. 0.825 0.045 0.829 0.825 0.823 0.785 0.970 0.911

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

312 2 4 0 21 14 0 0 12 0 | a = s8

22 142 18 1 0 13 0 0 0 0 | b = s9

0 0 1372 85 117 104 0 0 0 0 | c = s2

0 0 56 1281 36 115 0 0 0 0 | d = s4

1 0 73 34 832 83 0 0 4 0 | e = s3

0 0 139 56 7 1455 0 0 0 0 | f = s1

0 0 11 16 0 48 66 0 0 0 | g = s0

7 1 1 30 0 0 0 111 0 0 | h = s5

23 13 16 1 2 7 0 1 180 3 | i = s6

0 3 7 17 1 4 0 9 5 96 | j = s7

E2 Settings:

|  |  |  |  |
| --- | --- | --- | --- |
| seed | numIterations | breakTiesRandoml | maxDepth |
| 10 | 100 | F | 0 |

Correctly Classified Instances 5852 82.5388 %

Incorrectly Classified Instances 1238 17.4612 %

Kappa statistic 0.7851

Mean absolute error 0.0801

Root mean squared error 0.174

Relative absolute error 48.7311 %

Root relative squared error 60.8093 %

Total Number of Instances 7090

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.855 0.008 0.846 0.855 0.850 0.842 0.988 0.927 s8

0.730 0.004 0.831 0.730 0.777 0.773 0.943 0.804 s9

0.800 0.058 0.811 0.800 0.805 0.745 0.960 0.905 s2

0.874 0.036 0.864 0.874 0.869 0.834 0.972 0.941 s4

0.814 0.033 0.806 0.814 0.810 0.778 0.951 0.885 s3

0.884 0.073 0.787 0.884 0.833 0.780 0.975 0.936 s1

0.461 0.000 1.000 0.461 0.631 0.675 0.968 0.745 s0

0.767 0.001 0.927 0.767 0.839 0.840 0.998 0.938 s5

0.715 0.003 0.889 0.715 0.793 0.791 0.983 0.848 s6

0.690 0.001 0.933 0.690 0.794 0.799 0.995 0.854 s7

Weighted Avg. 0.825 0.044 0.830 0.825 0.824 0.786 0.968 0.910

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

312 0 4 1 21 11 0 0 16 0 | a = s8

32 143 10 0 0 10 0 0 0 1 | b = s9

0 0 1342 74 133 129 0 0 0 0 | c = s2

0 0 44 1301 34 109 0 0 0 0 | d = s4

0 0 82 23 836 83 0 0 3 0 | e = s3

0 0 138 48 7 1464 0 0 0 0 | f = s1

0 0 14 14 0 48 65 0 0 0 | g = s0

1 1 0 29 2 0 0 115 0 2 | h = s5

23 19 11 2 4 3 0 4 176 4 | i = s6

1 9 10 13 0 3 0 5 3 98 | j = s7

E2 Settings:

|  |  |  |  |
| --- | --- | --- | --- |
| seed | numIterations | breakTiesRandoml | maxDepth |
| 10 | 50 | F | 0 |

Correctly Classified Instances 5808 81.9182 %

Incorrectly Classified Instances 1282 18.0818 %

Kappa statistic 0.7774

Mean absolute error 0.08

Root mean squared error 0.1752

Relative absolute error 48.6882 %

Root relative squared error 61.2318 %

Total Number of Instances 7090

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.855 0.009 0.834 0.855 0.844 0.836 0.982 0.915 s8

0.735 0.004 0.832 0.735 0.780 0.776 0.922 0.784 s9

0.797 0.066 0.790 0.797 0.793 0.729 0.956 0.895 s2

0.868 0.041 0.849 0.868 0.858 0.820 0.970 0.935 s4

0.819 0.033 0.806 0.819 0.813 0.781 0.949 0.887 s3

0.868 0.069 0.793 0.868 0.829 0.775 0.971 0.926 s1

0.454 0.000 1.000 0.454 0.624 0.670 0.959 0.714 s0

0.767 0.000 0.991 0.767 0.865 0.870 0.996 0.933 s5

0.695 0.004 0.877 0.695 0.776 0.774 0.980 0.837 s6

0.655 0.000 0.969 0.655 0.782 0.793 0.993 0.813 s7

Weighted Avg. 0.819 0.046 0.824 0.819 0.818 0.778 0.964 0.901

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

312 2 2 3 17 13 0 0 16 0 | a = s8

32 144 9 0 0 11 0 0 0 0 | b = s9

0 0 1337 91 123 127 0 0 0 0 | c = s2

0 0 56 1292 36 104 0 0 0 0 | d = s4

0 0 88 25 841 71 0 0 2 0 | e = s3

0 0 158 49 11 1439 0 0 0 0 | f = s1

0 0 20 15 2 40 64 0 0 0 | g = s0

2 1 0 29 3 0 0 115 0 0 | h = s5

24 17 13 3 10 4 0 1 171 3 | i = s6

4 9 10 15 0 5 0 0 6 93 | j = s7

E4 Settings:

|  |  |  |  |
| --- | --- | --- | --- |
| seed | numIterations | breakTiesRandoml | maxDepth |
| 10 | 100 | T | 10 |

Correctly Classified Instances 5709 80.5219 %

Incorrectly Classified Instances 1381 19.4781 %

Kappa statistic 0.7601

Mean absolute error 0.0959

Root mean squared error 0.1917

Relative absolute error 58.4086 %

Root relative squared error 67.0179 %

Total Number of Instances 7090

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.855 0.009 0.834 0.855 0.844 0.836 0.992 0.930 s8

0.724 0.003 0.866 0.724 0.789 0.787 0.937 0.800 s9

0.773 0.062 0.794 0.773 0.783 0.717 0.946 0.887 s2

0.841 0.040 0.848 0.841 0.845 0.804 0.963 0.924 s4

0.810 0.039 0.778 0.810 0.794 0.758 0.949 0.878 s3

0.873 0.086 0.756 0.873 0.810 0.750 0.964 0.917 s1

0.376 0.000 1.000 0.376 0.546 0.609 0.953 0.715 s0

0.673 0.000 0.971 0.673 0.795 0.805 0.999 0.961 s5

0.728 0.004 0.877 0.728 0.796 0.793 0.991 0.867 s6

0.669 0.001 0.950 0.669 0.785 0.794 0.993 0.850 s7

Weighted Avg. 0.805 0.050 0.812 0.805 0.803 0.761 0.960 0.897

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

312 2 5 1 22 12 0 0 11 0 | a = s8

30 142 14 0 0 10 0 0 0 0 | b = s9

0 0 1297 80 162 139 0 0 0 0 | c = s2

0 0 54 1252 35 147 0 0 0 0 | d = s4

0 0 83 25 832 84 0 0 3 0 | e = s3

0 0 143 58 10 1446 0 0 0 0 | f = s1

0 0 12 11 1 64 53 0 0 0 | g = s0

9 4 3 31 0 0 0 101 0 2 | h = s5

23 9 15 2 5 8 0 2 179 3 | i = s6

0 7 8 16 2 2 0 1 11 95 | j = s7

## J48 results – All dataset with TEST9000 set

E1 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| T | 0.25 | 1 | T | T |

Correctly Classified Instances 7383 61.067 %

Incorrectly Classified Instances 4707 38.933 %

Kappa statistic 0.5247

Mean absolute error 0.078

Root mean squared error 0.2786

Relative absolute error 47.295 %

Root relative squared error 97.1765 %

Total Number of Instances 12090

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.667 0.107 0.655 0.667 0.661 0.557 0.790 0.527 s2

0.494 0.014 0.514 0.494 0.504 0.490 0.740 0.268 s9

0.649 0.116 0.595 0.649 0.621 0.517 0.769 0.470 s4

0.632 0.108 0.636 0.632 0.634 0.525 0.762 0.487 s1

0.302 0.010 0.390 0.302 0.341 0.331 0.646 0.133 s0

0.403 0.012 0.446 0.403 0.423 0.411 0.696 0.194 s7

0.567 0.017 0.416 0.567 0.480 0.473 0.775 0.245 s5

0.426 0.028 0.463 0.426 0.443 0.413 0.699 0.228 s8

0.646 0.051 0.685 0.646 0.665 0.610 0.798 0.495 s3

0.452 0.011 0.598 0.452 0.514 0.504 0.720 0.290 s6

Weighted Avg. 0.611 0.084 0.611 0.611 0.610 0.528 0.766 0.447

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

1881 9 376 350 19 15 19 19 121 9 | a = s2

19 168 7 8 5 11 3 61 49 9 | b = s9

217 7 1625 384 46 5 42 34 132 12 | c = s4

353 7 421 1757 17 20 25 66 102 13 | d = s1

34 1 78 50 78 0 3 0 2 12 | e = s0

24 5 35 7 2 112 55 24 10 4 | f = s7

27 16 2 12 3 16 144 9 18 7 | g = s5

88 64 21 55 7 38 29 278 41 32 | h = s8

194 12 143 127 21 21 6 68 1144 34 | i = s3

36 38 25 11 2 13 20 42 51 196 | j = s6

E2 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| F | 0.25 | 5 | T | F |

Correctly Classified Instances 7102 58.7428 %

Incorrectly Classified Instances 4988 41.2572 %

Kappa statistic 0.495

Mean absolute error 0.0871

Root mean squared error 0.26

Relative absolute error 52.8482 %

Root relative squared error 90.694 %

Total Number of Instances 12090

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.660 0.094 0.680 0.660 0.670 0.571 0.812 0.604 s2

0.568 0.024 0.411 0.568 0.477 0.465 0.817 0.348 s9

0.627 0.120 0.578 0.627 0.602 0.493 0.790 0.527 s4

0.627 0.132 0.587 0.627 0.606 0.484 0.816 0.603 s1

0.151 0.003 0.520 0.151 0.234 0.273 0.688 0.130 s0

0.273 0.002 0.809 0.273 0.409 0.464 0.789 0.333 s7

0.551 0.019 0.387 0.551 0.455 0.448 0.823 0.300 s5

0.309 0.031 0.365 0.309 0.335 0.301 0.721 0.298 s8

0.629 0.066 0.620 0.629 0.624 0.559 0.822 0.567 s3

0.382 0.013 0.524 0.382 0.442 0.430 0.794 0.381 s6

Weighted Avg. 0.587 0.090 0.592 0.587 0.583 0.499 0.801 0.528

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

1859 13 306 431 0 5 10 31 131 32 | a = s2

10 193 3 9 0 1 28 21 64 11 | b = s9

316 1 1571 396 28 0 11 35 137 9 | c = s4

278 1 455 1743 3 1 25 136 123 16 | d = s1

2 0 76 90 39 0 0 6 45 0 | e = s0

20 22 42 34 0 76 47 9 20 8 | f = s7

18 31 4 22 0 0 140 11 24 4 | g = s5

45 93 36 79 0 6 65 202 91 36 | h = s8

175 27 195 145 1 2 34 43 1113 35 | i = s3

12 89 30 22 4 3 2 59 47 166 | j = s6

E3 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| F | 0.5 | 10 | T | T |

Correctly Classified Instances 7323 60.5707 %

Incorrectly Classified Instances 4767 39.4293 %

Kappa statistic 0.5174

Mean absolute error 0.0959

Root mean squared error 0.2448

Relative absolute error 58.1906 %

Root relative squared error 85.3664 %

Total Number of Instances 12090

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.639 0.102 0.656 0.639 0.648 0.543 0.824 0.615 s2

0.409 0.009 0.563 0.409 0.474 0.467 0.847 0.336 s9

0.661 0.112 0.607 0.661 0.633 0.533 0.825 0.561 s4

0.668 0.094 0.680 0.668 0.674 0.578 0.852 0.627 s1

0.411 0.015 0.381 0.411 0.396 0.382 0.799 0.197 s0

0.273 0.002 0.768 0.273 0.403 0.451 0.833 0.351 s7

0.343 0.010 0.418 0.343 0.377 0.366 0.849 0.209 s5

0.413 0.042 0.362 0.413 0.386 0.350 0.763 0.249 s8

0.689 0.090 0.568 0.689 0.623 0.554 0.864 0.623 s3

0.253 0.005 0.667 0.253 0.367 0.399 0.869 0.314 s6

Weighted Avg. 0.606 0.085 0.612 0.606 0.601 0.524 0.836 0.546

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

1802 1 356 391 1 6 17 50 191 3 | a = s2

28 139 7 9 3 3 13 95 43 0 | b = s9

201 1 1656 220 105 0 2 59 242 18 | c = s4

288 0 380 1857 21 1 1 88 143 2 | d = s1

6 0 20 85 106 0 0 0 41 0 | e = s0

27 0 49 8 0 76 58 25 26 9 | f = s7

31 8 60 1 3 0 87 44 19 1 | g = s5

105 50 12 40 30 6 15 270 113 12 | h = s8

216 0 166 105 6 4 1 42 1220 10 | i = s3

41 48 20 14 3 3 14 72 109 110 | j = s6

E4 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| F | 0.5 | 20 | T | T |

=== Summary ===

Correctly Classified Instances 6795 56.2035 %

Incorrectly Classified Instances 5295 43.7965 %

Kappa statistic 0.4662

Mean absolute error 0.1052

Root mean squared error 0.2515

Relative absolute error 63.8108 %

Root relative squared error 87.7219 %

Total Number of Instances 12090

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.598 0.094 0.659 0.598 0.627 0.522 0.822 0.575 s2

0.000 0.000 ? 0.000 ? ? 0.870 0.209 s9

0.681 0.127 0.583 0.681 0.628 0.524 0.823 0.572 s4

0.634 0.106 0.641 0.634 0.637 0.530 0.840 0.597 s1

0.314 0.034 0.169 0.314 0.220 0.208 0.749 0.085 s0

0.439 0.023 0.310 0.439 0.364 0.352 0.819 0.186 s7

0.000 0.000 ? 0.000 ? ? 0.866 0.186 s5

0.458 0.063 0.293 0.458 0.357 0.321 0.785 0.232 s8

0.578 0.069 0.588 0.578 0.583 0.512 0.838 0.546 s3

0.276 0.010 0.519 0.276 0.361 0.363 0.858 0.278 s6

Weighted Avg. 0.562 0.088 ? 0.562 ? ? 0.829 0.508

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

1684 0 376 317 61 20 0 116 233 11 | a = s2

37 0 2 22 13 76 0 150 38 2 | b = s9

209 0 1704 253 162 3 0 77 79 17 | c = s4

306 0 420 1762 63 3 0 129 91 7 | d = s1

0 0 43 108 81 0 0 5 21 0 | e = s0

34 0 16 44 0 122 0 34 28 0 | f = s7

20 0 59 13 2 115 0 27 17 1 | g = s5

65 0 17 67 23 33 0 299 103 46 | h = s8

183 0 264 137 70 3 0 63 1023 27 | i = s3

17 0 20 26 4 18 0 122 107 120 | j = s6

E5 Settings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Binary split | Confidence Factor | minNumObj | Multiple branching | prunning |
| F | 0.8 | 5 | T | T |
|  |  |  |  |  |

=== Summary ===

Correctly Classified Instances 7102 58.7428 %

Incorrectly Classified Instances 4988 41.2572 %

Kappa statistic 0.495

Mean absolute error 0.0871

Root mean squared error 0.26

Relative absolute error 52.8482 %

Root relative squared error 90.694 %

Total Number of Instances 12090

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.660 0.094 0.680 0.660 0.670 0.571 0.812 0.604 s2

0.568 0.024 0.411 0.568 0.477 0.465 0.817 0.348 s9

0.627 0.120 0.578 0.627 0.602 0.493 0.790 0.527 s4

0.627 0.132 0.587 0.627 0.606 0.484 0.816 0.603 s1

0.151 0.003 0.520 0.151 0.234 0.273 0.688 0.130 s0

0.273 0.002 0.809 0.273 0.409 0.464 0.789 0.333 s7

0.551 0.019 0.387 0.551 0.455 0.448 0.823 0.300 s5

0.309 0.031 0.365 0.309 0.335 0.301 0.721 0.298 s8

0.629 0.066 0.620 0.629 0.624 0.559 0.822 0.567 s3

0.382 0.013 0.524 0.382 0.442 0.430 0.794 0.381 s6

Weighted Avg. 0.587 0.090 0.592 0.587 0.583 0.499 0.801 0.528

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

1859 13 306 431 0 5 10 31 131 32 | a = s2

10 193 3 9 0 1 28 21 64 11 | b = s9

316 1 1571 396 28 0 11 35 137 9 | c = s4

278 1 455 1743 3 1 25 136 123 16 | d = s1

2 0 76 90 39 0 0 6 45 0 | e = s0

20 22 42 34 0 76 47 9 20 8 | f = s7

18 31 4 22 0 0 140 11 24 4 | g = s5

45 93 36 79 0 6 65 202 91 36 | h = s8

175 27 195 145 1 2 34 43 1113 35 | i = s3

12 89 30 22 4 3 2 59 47 166 | j = s6

## Random forest results – ALL dataset with TEST9000 set

E1 Settings:

|  |  |  |  |
| --- | --- | --- | --- |
| seed | numIterations | breakTiesRandoml | maxDepth |
| 1 | 100 | F | 0 |

Correctly Classified Instances 7736 63.9868 %

Incorrectly Classified Instances 4354 36.0132 %

Kappa statistic 0.5536

Mean absolute error 0.1219

Root mean squared error 0.2314

Relative absolute error 73.983 %

Root relative squared error 80.693 %

Total Number of Instances 12090

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.683 0.149 0.582 0.683 0.628 0.506 0.877 0.718 s2

0.568 0.009 0.652 0.568 0.607 0.598 0.927 0.648 s9

0.627 0.060 0.733 0.627 0.676 0.602 0.890 0.757 s4

0.743 0.157 0.586 0.743 0.655 0.543 0.894 0.750 s1

0.116 0.000 0.938 0.116 0.207 0.327 0.919 0.389 s0

0.392 0.002 0.832 0.392 0.533 0.565 0.978 0.678 s7

0.335 0.000 1.000 0.335 0.501 0.574 0.984 0.696 s5

0.662 0.012 0.766 0.662 0.710 0.697 0.983 0.826 s8

0.630 0.062 0.637 0.630 0.633 0.571 0.904 0.724 s3

0.484 0.004 0.827 0.484 0.610 0.623 0.961 0.705 s6

Weighted Avg. 0.640 0.093 0.665 0.640 0.634 0.560 0.903 0.729

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

1924 0 169 455 1 0 0 0 269 0 | a = s2

56 193 3 30 0 2 0 29 26 1 | b = s9

344 0 1571 424 0 0 0 0 165 0 | c = s4

536 0 88 2067 0 0 0 0 90 0 | d = s1

17 0 3 205 30 0 0 0 3 0 | e = s0

55 17 28 19 0 109 0 31 4 15 | f = s7

37 36 46 12 1 8 85 24 2 3 | g = s5

32 31 17 79 0 4 0 432 38 20 | h = s8

246 0 204 200 0 0 0 0 1115 5 | i = s3

61 19 14 35 0 8 0 48 39 210 | j = s6

E2 Settings:

|  |  |  |  |
| --- | --- | --- | --- |
| seed | numIterations | breakTiesRandoml | maxDepth |
| 10 | 100 | F | 0 |

=== Summary ===

Correctly Classified Instances 7818 64.665 %

Incorrectly Classified Instances 4272 35.335 %

Kappa statistic 0.5619

Mean absolute error 0.1212

Root mean squared error 0.23

Relative absolute error 73.5452 %

Root relative squared error 80.2348 %

Total Number of Instances 12090

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.700 0.143 0.598 0.700 0.645 0.529 0.888 0.749 s2

0.535 0.009 0.621 0.535 0.575 0.565 0.922 0.644 s9

0.611 0.057 0.737 0.611 0.669 0.596 0.891 0.758 s4

0.770 0.156 0.596 0.770 0.672 0.566 0.899 0.757 s1

0.120 0.000 0.939 0.120 0.213 0.332 0.916 0.386 s0

0.432 0.002 0.822 0.432 0.566 0.589 0.974 0.683 s7

0.335 0.000 0.988 0.335 0.500 0.571 0.989 0.790 s5

0.631 0.009 0.792 0.631 0.702 0.692 0.983 0.844 s8

0.638 0.065 0.628 0.638 0.633 0.569 0.903 0.728 s3

0.493 0.003 0.863 0.493 0.628 0.643 0.955 0.691 s6

Weighted Avg. 0.647 0.091 0.672 0.647 0.641 0.569 0.907 0.741

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

1972 0 135 430 2 0 0 0 279 0 | a = s2

55 182 1 46 0 2 0 22 32 0 | b = s9

336 0 1531 460 0 0 0 0 177 0 | c = s4

480 0 75 2142 0 0 0 0 84 0 | d = s1

25 0 0 202 31 0 0 0 0 0 | e = s0

52 17 30 13 0 120 0 36 5 5 | f = s7

31 46 45 16 0 11 85 13 1 6 | g = s5

50 20 23 80 0 6 0 412 46 16 | h = s8

236 0 225 172 0 1 0 0 1129 7 | i = s3

59 28 11 33 0 6 1 37 45 214 | j = s6

E2 Settings:

|  |  |  |  |
| --- | --- | --- | --- |
| seed | numIterations | breakTiesRandoml | maxDepth |
| 10 | 50 | F | 0 |

=== Summary ===

Correctly Classified Instances 7594 62.8122 %

Incorrectly Classified Instances 4496 37.1878 %

Kappa statistic 0.5384

Mean absolute error 0.1218

Root mean squared error 0.2324

Relative absolute error 73.8639 %

Root relative squared error 81.0461 %

Total Number of Instances 12090

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.688 0.159 0.569 0.688 0.623 0.498 0.866 0.694 s2

0.512 0.008 0.644 0.512 0.570 0.563 0.918 0.618 s9

0.591 0.066 0.702 0.591 0.642 0.561 0.880 0.730 s4

0.751 0.157 0.589 0.751 0.660 0.549 0.891 0.736 s1

0.120 0.000 0.886 0.120 0.212 0.322 0.878 0.317 s0

0.450 0.003 0.758 0.450 0.564 0.576 0.967 0.653 s7

0.339 0.000 0.966 0.339 0.501 0.568 0.979 0.703 s5

0.609 0.010 0.771 0.609 0.681 0.670 0.976 0.800 s8

0.603 0.063 0.623 0.603 0.613 0.548 0.896 0.705 s3

0.470 0.003 0.872 0.470 0.611 0.631 0.936 0.667 s6

Weighted Avg. 0.628 0.097 0.652 0.628 0.622 0.546 0.894 0.707

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

1940 0 171 442 2 0 0 0 263 0 | a = s2

63 174 1 46 0 5 0 23 28 0 | b = s9

406 0 1481 442 1 0 0 0 173 1 | c = s4

518 0 80 2088 0 0 0 0 95 0 | d = s1

23 0 3 200 31 0 0 1 0 0 | e = s0

45 14 24 23 0 125 1 33 4 9 | f = s7

28 40 52 8 1 16 86 20 1 2 | g = s5

70 27 15 86 0 5 1 398 39 12 | h = s8

258 0 261 175 0 3 0 0 1067 6 | i = s3

60 15 23 36 0 11 1 41 43 204 | j = s6

E4 Settings:

|  |  |  |  |
| --- | --- | --- | --- |
| seed | numIterations | breakTiesRandoml | maxDepth |
| 10 | 100 | T | 10 |

=== Summary ===

Correctly Classified Instances 7869 65.0868 %

Incorrectly Classified Instances 4221 34.9132 %

Kappa statistic 0.5676

Mean absolute error 0.1218

Root mean squared error 0.2309

Relative absolute error 73.919 %

Root relative squared error 80.5253 %

Total Number of Instances 12090

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.694 0.137 0.607 0.694 0.647 0.533 0.884 0.754 s2

0.538 0.008 0.661 0.538 0.593 0.586 0.938 0.663 s9

0.622 0.057 0.742 0.622 0.677 0.605 0.889 0.759 s4

0.766 0.155 0.596 0.766 0.671 0.564 0.897 0.749 s1

0.105 0.000 1.000 0.105 0.189 0.320 0.916 0.422 s0

0.432 0.003 0.789 0.432 0.558 0.577 0.972 0.654 s7

0.386 0.000 0.961 0.386 0.551 0.604 0.991 0.826 s5

0.660 0.010 0.791 0.660 0.720 0.708 0.983 0.843 s8

0.646 0.066 0.625 0.646 0.635 0.572 0.906 0.739 s3

0.514 0.003 0.854 0.514 0.642 0.654 0.956 0.715 s6

Weighted Avg. 0.651 0.090 0.675 0.651 0.645 0.574 0.906 0.745

=== Confusion Matrix ===

a b c d e f g h i j <-- classified as

1956 0 140 429 0 0 0 0 293 0 | a = s2

49 183 5 46 0 6 0 19 30 2 | b = s9

314 0 1557 449 0 0 0 0 184 0 | c = s4

472 0 75 2131 0 0 0 0 103 0 | d = s1

22 0 1 208 27 0 0 0 0 0 | e = s0

62 11 27 12 0 120 3 35 4 4 | f = s7

35 28 41 10 0 12 98 22 1 7 | g = s5

38 27 17 74 0 4 1 431 42 19 | h = s8

216 0 226 179 0 0 0 0 1143 6 | i = s3

60 28 10 36 0 10 0 38 29 223 | j = s6