1. **Random Forests**

This section presents the prediction results of the models built by using Random Forests.

* 1. **Cross-validation (RQ2)**

TABLE I, TABLE II, and TABLE III show the median CEs at different cut-off values (π = 0.1, 0.2, and 1.0) and the median ER-BPPs / ER-AVGs obtained from 10 times 10-fold cross-validation in each project for CM, NM, and AM, respectively. The underlined figure in TABLE II and TABLE III means that NM/AM obtains a lower median value than CM. The figure is followed by “\*”, if the Wilcoxon signed-rank test shows that the difference between NM/AM and CM is significant at the significance level of 0.05. The background colors from the lightest to the darkest indicate the effect sizes from trivial to large.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1. The cross-validation performance of CM | | | | | |
|  |  |  |  | ER-BPP | ER-AVG |
| Firefox | 0.242 | 0.238 | 0.260 | 0.388 | 0.297 |
| Eclipse | 0.345 | 0.332 | 0.355 | 0.513 | 0.528 |
| Ant | 0.007 | 0.044 | 0.091 | 0.088 | -0.043 |
| Camel | 0.220 | 0.213 | 0.330 | 0.389 | 0.362 |
| Ivy | -0.149 | -0.144 | -0.012 | -0.130 | -0.330 |
| JEdit | -0.060 | -0.116 | 0.262 | -1.000 | -0.177 |
| Lucene | 0.657 | 0.653 | 0.677 | 0.540 | 0.579 |
| Poi | 0.547 | 0.542 | 0.562 | 0.404 | 0.424 |
| Tomcat | 0.113 | 0.148 | 0.291 | 0.357 | 0.049 |
| Xalan | 0.687 | 0.671 | 0.717 | 0.656 | 0.675 |
| Xerces | 0.598 | 0.522 | 0.624 | 0.703 | 0.478 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1. The cross-validation performance of NM | | | | | |  | 1. The cross-validation performance of AM | | | | | |
|  |  |  |  | ER-BPP | ER-AVG |  |  |  |  |  | ER-BPP | ER-AVG |
| **Firefox** | 0.416\* | 0.410\* | 0.500\* | 0.627\* | 0.397\* |  | **Firefox** | 0.435\* | 0.432\* | 0.504\* | 0.616\* | 0.426\* |
| **Eclipse** | 0.399\* | 0.396\* | 0.432\* | 0.544\* | 0.510\* |  | **Eclipse** | 0.444\* | 0.435\* | 0.484\* | 0.572\* | 0.560\* |
| **Ant** | 0.020 | 0.034 | 0.137\* | 0.165\* | -0.045 |  | **Ant** | 0.095\* | 0.135\* | 0.180\* | 0.141\* | 0.043\* |
| **Camel** | 0.367\* | 0.388\* | 0.630\* | 0.629\* | 0.442\* |  | **Camel** | 0.384\* | 0.400\* | 0.636\* | 0.644\* | 0.486\* |
| **Ivy** | -0.143 | -0.127 | -0.059 | -0.292 | -0.313 |  | **Ivy** | -0.168 | -0.140 | -0.082\* | -0.510 | -0.360\* |
| **JEdit** | -0.057\* | 0.144\* | 0.726\* | -1.000 | -0.032\* |  | **JEdit** | -0.058\* | 0.075\* | 0.676\* | -1.000 | -0.125\* |
| **Lucene** | 0.671\* | 0.678\* | 0.666 | 0.521 | 0.534\* |  | **Lucene** | 0.662\* | 0.657 | 0.665 | 0.530 | 0.552\* |
| **Poi** | 0.515 | 0.548 | 0.580\* | 0.414 | 0.415\* |  | **Poi** | 0.515\* | 0.547 | 0.562\* | 0.383 | 0.426 |
| **Tomcat** | 0.082 | 0.121 | 0.277 | 0.340 | 0.038 |  | **Tomcat** | 0.099 | 0.153 | 0.274 | 0.405 | 0.012 |
| **Xalan** | 0.722\* | 0.699\* | 0.663\* | 0.624\* | 0.657\* |  | **Xalan** | 0.730\* | 0.705\* | 0.707 | 0.647 | 0.672\* |
| **Xerces** | 0.720\* | 0.689\* | 0.734\* | 0.754\* | 0.445 |  | **Xerces** | 0.718\* | 0.654\* | 0.729\* | 0.740\* | 0.476 |

TABLE IV shows the median values and the mini boxplots of the overall ranking and classification performance under cross-validation, together with the Cliff’s δ of the comparisons between NM/AM and CM. The “\*” following the Cliff’s δ means that the difference between the comparison pair is significant at the significance level of 0.05.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 1. Overall ranking/classification performance under cross-validation | | | | | | |
|  |  |  | median | mini boxplot | Cliff’s *δ* | |
|  |  |  | NM vs. CM | AM vs. CM |
| ranking |  | CM | 0.322 | F:\研二上\工作感知的网络度量缺陷预测\论文\cross_validation\rf.bmp | 0.096\* | 0.121\* |
| NM | 0.405 |
| AM | 0.434 |
|  | CM | 0.312 | F:\研二上\工作感知的网络度量缺陷预测\论文\cross_validation\rf.bmp | 0.121\* | 0.141\* |
| NM | 0.410 |
| AM | 0.428 |
|  | CM | 0.360 | F:\研二上\工作感知的网络度量缺陷预测\论文\cross_validation\rf.bmp | 0.157\* | 0.161\* |
| NM | 0.477 | F:\研二上\工作感知的网络度量缺陷预测\论文\cross_validation\rf.bmp |
| AM | 0.500 | F:\研二上\工作感知的网络度量缺陷预测\论文\cross_validation\rf.bmp |
| classification | ER-BPP | CM | 0.437 | F:\研二上\工作感知的网络度量缺陷预测\论文\cross_validation\rf.bmp | 0.132\* | 0.140\* |
| NM | 0.530 |
| AM | 0.545 |
| ER-AVG | CM | 0.345 | F:\研二上\工作感知的网络度量缺陷预测\论文\cross_validation\rf.bmp | 0.024\* | 0.088\* |
| NM | 0.396 |
| AM | 0.420 |
|  |  |  |  | 0.049 0.690 |  |  |

* 1. **Across-release prediction (RQ3)**

Fig. 1 shows the boxplots which describe the distribution of the CEs at different cut-off values and the ER-BPPs / ER-AVGs obtained from the across-release predictions for CM, NM, and AM. For each model, the boxplot presents the median (the horizontal line within the box), the 25th, and the 75th percentiles (the lower and upper sides of the box) of the prediction results in terms of CEs and ERs. The figures in parentheses indicate the number of prediction values for NM/CM/AM.

|  |  |  |
| --- | --- | --- |
| Firefox(15) | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\Firefox_rf_ce.bmp | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\Firefox_rf_er.bmp |
| Eclipse(10) | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\Eclipse_rf_ce.bmp | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\Eclipse_rf_er.bmp |
| Ant(10) | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\Ant_rf_ce.bmp | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\Ant_rf_er.bmp |
| Jedit(10) | **F:\研二上\工作感知的网络度量缺陷预测\论文\intra\JEdit_rf_ce.bmp** | **F:\研二上\工作感知的网络度量缺陷预测\论文\intra\JEdit_rf_er.bmp** |
| Fig.1. Ranking/Classification performance under across-release prediction | | |

TABLE V shows the Cliff’s δ of comparing NM/AM with CM under across-release prediction. The figure is followed by “\*”, if the p-value obtained from the Wilcoxon signed-rank test is lower than the significance level of 0.05.

|  |
| --- |
| 1. The Cliff’s δ of comparing NM/AM vs. CM under across-release prediction |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **NM vs. CM** | | | | |  | **AM vs. CM** | | | | |
|  | ER-BPP | ER-AVG |  |  |  |  | ER-BPP | ER-AVG |  |  |  |
| **Firefox** | 0.067 | 0.933\* | 0.916\* | 0.804\* | 0.218 |  | 0.173 | 0.933\* | 0.916\* | 0.858\* | 0.333\* |
| **Eclipse** | 0.520\* | 0.100 | 0.180\* | 0.040 | 0.280\* |  | 0.180\* | 0.440\* | 0.080\* | 0.160\* | 0.200\* |
| **Ant** | 0.200\* | 0.360 | 0.180 | 0.300 | 0.180 |  | 0.280\* | 0.320 | 0.220 | 0.360 | 0.140 |
| **JEdit** | 0.100 | 0.200 | 0.160 | 0.060 | 0.080 |  | 0.100 | 0.220 | 0.060 | 0.080 | 0.120 |

TABLE VI shows the median values and the mini boxplots of the overall ranking and classification performance under across-release prediction, together with the Cliff’s δ of the comparisons between NM/AM and CM. The “\*” following the Cliff’s δ means that the difference between the comparison pair is significant at the significance level of 0.05.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1. Overall ranking/classification performance under across-release prediction | | | | | | | |
|  |  |  | median | mini boxplot | Cliff’s *δ* | |
|  |  |  | NM vs. CM | AM vs. CM |
| ranking |  | CM | 0.149 | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\rf.bmp | 0.064 | 0.010 |
| NM | 0.207 |
| AM | 0.211 |
|  | CM | 0.172 | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\rf.bmp | 0.291\* | 0.321\* |
| NM | 0.225 |
| AM | 0.233 |
|  | CM | 0.298 | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\rf.bmp | 0.212 | 0.236 |
| NM | 0.298 |
| AM | 0.329 |
| classification | ER-BPP | CM | 0.427 | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\rf.bmp | 0.177 | 0.227 |
| NM | 0.395 |
| AM | 0.418 |
| ER-AVG | CM | 0.245 | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\rf.bmp | 0.059 | 0.145 |
| NM | 0.425 |
| AM | 0.431 |
|  |  |  |  | 0.087 0.493 |  |  |
|  | | | | | | |

* 1. **Inter-project prediction (RQ4)**

Fig. 2 shows the boxplots which describe the distribution of the CEs at different cut-off values and the ER-BPPs / ER-AVGs obtained from 90 inter-project predictions for CM, NM, and AM.

|  |  |
| --- | --- |
| F:\研二上\工作感知的网络度量缺陷预测\论文\inter\cart_ce.bmp | F:\研二上\工作感知的网络度量缺陷预测\论文\inter\cart_er.bmp |
| Fig 2. Ranking/Classification performance under the inter-project prediction | |

TABLE VII shows the Cliff’s *δ* of comparing NM/AM with CM under inter-project prediction. The figure is followed by “\*”, if the p-value obtained from the Wilcoxon signed-rank test is lower than the significance level of 0.05.

|  |
| --- |
| 1. The Cliff’s δ of comparing NM/AM vs. CM under inter-project prediction |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **ER-BPP** | **ER-AVG** |  |  |  |
| **NM vs. CM** | 0.091 | 0.105 | 0.162\* | 0.154 | 0.142 |
| **AM vs. CM** | 0.071 | 0.124 | 0.138 | 0.140 | 0.109 |

1. **Classification and Regression Trees**

This section presents the prediction results of the models built by using Classification and Regression Trees.

* 1. **Cross-validation (RQ2)**

TABLE VIII, TABLE IX, and TABLE X show the median CEs at different cut-off values (π = 0.1, 0.2, and 1.0) and the median ER-BPPs / ER-AVGs obtained from 10 times 10-fold cross-validation in each project for CM, NM, and AM, respectively.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1. The cross-validation performance of CM | | | | | |
|  |  |  |  | **ER-BPP** | **ER-AVG** |
| Firefox | 0.420 | 0.394 | 0.417 | 0.655 | 0.763 |
| Eclipse | 0.241 | 0.244 | 0.229 | 0.560 | 0.511 |
| Ant | -0.002 | 0.025 | 0.057 | 0.034 | -0.077 |
| Camel | 0.158 | 0.151 | 0.285 | 0.365 | 0.307 |
| Ivy | -0.137 | -0.084 | -0.319 | -0.156 | -0.129 |
| JEdit | -0.060 | -0.116 | -1.000 | -0.500 | -0.023 |
| Lucene | 0.504 | 0.497 | 0.537 | 0.409 | 0.426 |
| Poi | 0.458 | 0.480 | 0.451 | 0.304 | 0.355 |
| Tomcat | 0.069 | 0.087 | 0.035 | 0.296 | -0.035 |
| Xalan | 0.600 | 0.592 | 0.632 | 0.716 | 0.637 |
| Xerces | 0.460 | 0.408 | 0.192 | 0.232 | 0.195 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1. The cross-validation performance of NM | | | | | |  | 1. The cross-validation performance of AM | | | | | |
|  |  |  |  | ER-BPP | ER-AVG |  |  |  |  |  | ER-BPP | ER-AVG | |
| **Firefox** | 0.417 | 0.417\* | 0.476\* | 0.690\* | 0.774\* |  | **Firefox** | 0.420\* | 0.415\* | 0.485\* | 0.696\* | 0.781\* |
| **Eclipse** | 0.307\* | 0.309\* | 0.351\* | 0.599\* | 0.548\* |  | **Eclipse** | 0.293\* | 0.293\* | 0.337\* | 0.552 | 0.537\* |
| **Ant** | 0.029 | 0.051 | 0.114\* | 0.193\* | -0.036 |  | **Ant** | -0.008 | 0.018 | 0.075 | 0.120\* | -0.037 |
| **Camel** | 0.322\* | 0.359\* | 0.503\* | 0.618\* | 0.430\* |  | **Camel** | 0.309\* | 0.364\* | 0.464\* | 0.608\* | 0.401\* |
| **Ivy** | -0.124 | -0.061 | -0.398 | 0.033 | -0.100\* |  | **Ivy** | -0.130 | -0.099 | -0.407 | -0.068 | -0.108 |
| **JEdit** | 0.010 | 0.280 | -0.228 | 0.000 | -0.023 |  | **JEdit** | 0.010 | 0.280 | -0.228 | 0.000 | -0.023 |
| **Lucene** | 0.596\* | 0.581\* | 0.541 | 0.366 | 0.416 |  | **Lucene** | 0.591\* | 0.580\* | 0.541 | 0.384\* | 0.406 |
| **Poi** | 0.458 | 0.493 | 0.488 | 0.275 | 0.337\* |  | **Poi** | 0.436 | 0.457 | 0.402 | 0.233\* | 0.322\* |
| **Tomcat** | 0.151\* | 0.136\* | -0.016 | 0.276 | 0.031 |  | **Tomcat** | 0.079 | 0.072 | -0.102\* | 0.261 | -0.039 |
| **Xalan** | 0.686\* | 0.672\* | 0.660 | 0.719\* | 0.662\* |  | **Xalan** | 0.676\* | 0.666\* | 0.623 | 0.649\* | 0.612\* |
| **Xerces** | 0.655\* | 0.606\* | 0.391\* | 0.392\* | 0.173 |  | **Xerces** | 0.616\* | 0.574\* | 0.495\* | 0.492\* | 0.169 |

TABLE XI shows the median values and the mini boxplots of the overall ranking and classification performance under cross-validation, together with the Cliff’s δ of the comparisons between NM/AM and CM. The “\*” following the Cliff’s δ means that the difference between the comparison pair is significant at the significance level of 0.05.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1. Overall ranking/classification performance under cross-validation | | | | | | | |
|  |  |  | median | mini boxplot | Cliff’s *δ* | |
|  |  |  | NM vs. CM | AM vs. CM |
| ranking |  | CM | 0.286 | F:\研二上\工作感知的网络度量缺陷预测\论文\cross_validation\cart.bmp | 0.127\* | 0.086 |
| NM | 0.359 |
| AM | 0.333 |
|  | CM | 0.275 | F:\研二上\工作感知的网络度量缺陷预测\论文\cross_validation\cart.bmp | 0.152\* | 0.112 |
| NM | 0.372 |
| AM | 0.351 |
|  | CM | 0.279 | F:\研二上\工作感知的网络度量缺陷预测\论文\cross_validation\cart.bmp | 0.114\* | 0.098 |
| NM | 0.373 |
| AM | 0.370 |
| classification | ER-BPP | CM | 0.390 | F:\研二上\工作感知的网络度量缺陷预测\论文\cross_validation\cart.bmp | 0.115\* | 0.072 |
| NM | 0.482 |
| AM | 0.459 |
| ER-AVG | CM | 0.275 | F:\研二上\工作感知的网络度量缺陷预测\论文\cross_validation\cart.bmp | 0.049\* | 0.021 |
| NM | 0.278 |
| AM | 0.276 |
|  |  |  |  | -0.020 0.663 |  |  |
|  | | | | | | |

* 1. **Across-release prediction (RQ3)**

Fig. 3 shows the boxplots which describe the distribution of the CEs at different cut-off values and the ER-BPPs / ER-AVGs obtained from the across-release predictions for CM, NM, and AM.

|  |  |  |
| --- | --- | --- |
| Firefox(15) | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\Firefox_cart_ce.bmp | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\Firefox_cart_er.bmp |
| Eclipse(10) | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\Eclipse_cart_ce.bmp | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\Eclipse_cart_er.bmp |
| Ant(10) | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\Ant_cart_ce.bmp | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\Ant_cart_er.bmp |
| Jedit(10) | **F:\研二上\工作感知的网络度量缺陷预测\论文\intra\JEdit_cart_ce.bmp** | **F:\研二上\工作感知的网络度量缺陷预测\论文\intra\JEdit_cart_er.bmp** |
| Fig. 3. Ranking/Classification performance under across-release prediction | | |

TABLE XII shows the Cliff’s δ of comparing NM/AM with CM under across-release prediction. The figure is followed by “\*”, if the p-value obtained from the Wilcoxon signed-rank test is lower than the significance level of 0.05.

|  |
| --- |
| 1. The Cliff’s δ of comparing NM/AM vs. CM under across-release prediction |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **NM vs. CM** | | | | |  | **AM vs. CM** | | | | |
|  | ER-BPP | ER-AVG |  |  |  |  | ER-BPP | ER-AVG |  |  |  |
| **Firefox** | 0.067\* | 0.067\* | 0.138\* | 0.129\* | 0.582\* |  | 0.209 | 0.022 | 0.076 | 0.173 | 0.458\* |
| **Eclipse** | 0.280 | 0.160 | 0.020\* | 0.040\* | 0.440\* |  | 0.160 | 0.140 | 0.120 | 0.020 | 0.650\* |
| **Ant** | 0.220 | 0.260 | 0.320 | 0.500\* | 0.880\* |  | 0.300 | 0.040 | 0.060 | 0.320\* | 0.580\* |
| **JEdit** | 0.080 | 0.120 | 0.120 | 0.120 | 0.020 |  | 0.220 | 0.060 | 0.120 | 0.220 | 0.020 |

TABLE XIII shows the median values and the mini boxplots of the overall ranking and classification performance under across-release prediction, together with the Cliff’s δ of the comparisons between NM/AM and CM. The “\*” following the Cliff’s δ means that the difference between the comparison pair is significant at the significance level of 0.05.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1. Overall ranking/classification performance under across-release prediciotn | | | | | | | | |
|  |  |  | median | mini boxplot | Cliff’s *δ* | |
|  |  |  | NM vs. CM | AM vs. CM |
| ranking |  | CM | 0.148 | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\cart.bmp | 0.035 | 0.020 |
| NM | 0.156 |
| AM | 0.180 |
|  | CM | 0.146 | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\cart.bmp | 0.153\* | 0.007 |
| NM | 0.160 |
| AM | 0.177 |
|  | CM | 0.196 | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\cart.bmp | 0.003 | 0.058 |
| NM | 0.230 |
| AM | 0.248 |
| classification | ER-BPP | CM | 0.386 | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\cart.bmp | 0.006 | 0.050 |
| NM | 0.409 |
| AM | 0.397 |
| ER-AVG | CM | 0.400 | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\cart.bmp | 0.093 | 0.151 |
| NM | 0.413 |
| AM | 0.400 |
|  |  |  |  | -0.093 0.722 |  |  |

* 1. **Inter-project prediction (RQ4)**

Fig. 4 shows the boxplots which describe the distribution of the CEs at different cut-off values and the ER-BPPs / ER-AVGs obtained from 90 inter-project predictions for CM, NM, and AM.

|  |  |
| --- | --- |
| F:\研二上\工作感知的网络度量缺陷预测\论文\inter\cart_ce.bmp | F:\研二上\工作感知的网络度量缺陷预测\论文\inter\cart_er.bmp |
| Fig. 4. Ranking/Classification performance under the inter-project prediction | |

TABLE XIV shows the Cliff’s *δ* of comparing NM/AM with CM under inter-project prediction. The figure is followed by “\*”, if the p-value obtained from the Wilcoxon signed-rank test is lower than the significance level of 0.05.

|  |
| --- |
| 1. The Cliff’s δ of comparing NM/AM vs. CM under inter-project prediction |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **ER-BPP** | **ER-AVG** |  |  |  |
| **NM vs. CM** | 0.060 | 0.029 | 0.012 | 0.041 | 0.012 |
| **AM vs. CM** | 0.063 | 0.079 | 0.108 | 0.050 | 0.111\* |

1. **Support Vector Machine**

This section presents the prediction results of the models built by using Support Vector Machine.

* 1. **Cross-validation (RQ2)**

TABLE XV, TABLE IX, and TABLE X show the median CEs at different cut-off values (π = 0.1, 0.2, and 1.0) and the median ER-BPPs / ER-AVGs obtained from 10 times 10-fold cross-validation in each project for CM, NM, and AM, respectively.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1. The cross-validation performance of CM | | | | | |
|  |  |  |  | ER-BPP | ER-AVG |
| Firefox | 0.430 | 0.409 | 0.497 | 0.503 | 0.773 |
| Eclipse | 0.240 | 0.256 | 0.325 | 0.494 | 0.536 |
| Ant | -0.016 | -0.008 | -0.039 | 0.009 | -0.226 |
| Camel | 0.141 | 0.166 | 0.314 | 0.430 | 0.308 |
| Ivy | -0.170 | -0.173 | -0.064 | -0.690 | -0.637 |
| JEdit | -0.061 | -0.119 | 0.426 | -1.000 | -0.571 |
| Lucene | 0.605 | 0.595 | 0.637 | 0.646 | 0.542 |
| Poi | 0.505 | 0.515 | 0.533 | 0.569 | 0.436 |
| Tomcat | 0.059 | 0.031 | 0.060 | 0.225 | -0.270 |
| Xalan | 0.656 | 0.650 | 0.699 | 0.758 | 0.685 |
| Xerces | 0.586 | 0.531 | 0.569 | 0.887 | 0.588 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1. The cross-validation performance of NM | | | | | |  | 1. The cross-validation performance of AM | | | | | |
|  |  |  |  | **ER-BPP** | **ER-AVG** |  |  |  |  |  | **ER-BPP** | **ER-AVG** |
| **Firefox** | 0.529\* | 0.506\* | 0.606\* | 0.809\* | 0.789\* |  | **Firefox** | 0.541\* | 0.520\* | 0.590\* | 0.821\* | 0.779\* |
| **Eclipse** | 0.300\* | 0.316\* | 0.387\* | 0.573\* | 0.562\* |  | **Eclipse** | 0.343\* | 0.349\* | 0.382\* | 0.607\* | 0.568\* |
| **Ant** | -0.056\* | -0.027\* | 0.009\* | -0.108 | -0.275 |  | **Ant** | 0.061\* | 0.060\* | 0.044\* | 0.210\* | -0.069\* |
| **Camel** | 0.241\* | 0.258\* | 0.417\* | 0.567\* | 0.454\* |  | **Camel** | 0.258\* | 0.291\* | 0.400\* | 0.633\* | 0.459\* |
| **Ivy** | -0.149 | -0.120\* | -0.043\* | -0.252\* | -0.551\* |  | **Ivy** | -0.154\* | -0.164 | -0.046\* | -1.000 | -0.612\* |
| **JEdit** | -0.061\* | -0.119 | 0.431 | -1.000 | -0.559 |  | **JEdit** | -0.061\* | -0.119 | 0.390 | -1.000\* | -0.655\* |
| **Lucene** | 0.597 | 0.601 | 0.626 | 0.663\* | 0.516\* |  | **Lucene** | 0.641\* | 0.639\* | 0.660\* | 0.463\* | 0.523 |
| **Poi** | 0.477 | 0.536 | 0.543\* | 0.580 | 0.406\* |  | **Poi** | 0.526 | 0.527\* | 0.515\* | 0.399\* | 0.426 |
| **Tomcat** | 0.104\* | 0.100\* | 0.186\* | 0.454\* | -0.114\* |  | **Tomcat** | 0.054 | 0.028 | 0.062 | 0.252 | -0.256 |
| **Xalan** | 0.674\* | 0.665\* | 0.683\* | 0.792\* | 0.686 |  | **Xalan** | 0.687\* | 0.669\* | 0.699\* | 0.688\* | 0.681 |
| **Xerces** | 0.578\* | 0.513\* | 0.552\* | 0.902 | 0.581\* |  | **Xerces** | 0.623\* | 0.561\* | 0.616\* | 0.891 | 0.630\* |

TABLE XVII shows the median values and the mini boxplots of the overall ranking and classification performance under across-release prediction, together with the Cliff’s δ of the comparisons between NM/AM and CM. The “\*” following the Cliff’s δ means that the difference between the comparison pair is significant at the significance level of 0.05.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1. Overall ranking/classification performance under cross-validation | | | | | | | |
|  |  |  | median | mini boxplot | Cliff’s *δ* | |
|  |  |  | NM vs. CM | AM vs. CM |
| Ranking |  | CM | 0.292 |  | 0.046\* | 0.100\* |
| NM | 0.333 |
| AM | 0.371 |
|  | CM | 0.294 |  | 0.053\* | 0.094\* |
| NM | 0.345 |
| AM | 0.375 |
|  | CM | 0.365 |  | 0.070\* | 0.064 |
| NM | 0.431 |
| AM | 0.420 |
| classification | ER-BPP | CM | 0.502 |  | 0.130\* | 0.071\* |
| NM | 0.586 |
| AM | 0.589 |
| ER-AVG | CM | 0.448 |  | 0.015\* | 0.037\* |
| NM | 0.476 |
| AM | 0.490 |
|  |  |  |  | -0.112 0.784 |  |  |
|  | | | | | | |

* 1. **Across-release prediction (RQ3)**

Fig. 5 shows the boxplots which describe the distribution of the CEs at different cut-off values and the ER-BPPs / ER-AVGs obtained from the across-release predictions for CM, NM, and AM.

|  |  |  |
| --- | --- | --- |
| Firefox(15) | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\Firefox_svm_ce.bmp | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\Firefox_svm_er.bmp |
| Eclipse(10) | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\Eclipse_svm_ce.bmp | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\Eclipse_svm_er.bmp |
| Ant(10) | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\Ant_svm_ce.bmp | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\Ant_svm_er.bmp |
| Jedit(10) | **F:\研二上\工作感知的网络度量缺陷预测\论文\intra\JEdit_svm_ce.bmp** | **F:\研二上\工作感知的网络度量缺陷预测\论文\intra\JEdit_svm_er.bmp** |
| Fig. 5. Ranking/Classification performance under across-release prediction | | |

TABLE XIX shows the median values and the mini boxplots of the overall ranking and classification performance under across-release prediction, together with the Cliff’s δ of the comparisons between NM/AM and CM. The “\*” following the Cliff’s δ means that the difference between the comparison pair is significant at the significance level of 0.05.

|  |
| --- |
| 1. The Cliff’s δ of comparing NM/AM vs. CM under across-release prediction |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **`** | **NM vs. CM** | | | | |  | **AM vs. CM** | | | | |
|  | ER-BPP | ER-AVG |  |  |  |  | ER-BPP | ER-AVG |  |  |  |
| **Firefox** | 0.831\* | 0.004 | 0.387\* | 0.404\* | 0.387\* |  | 0.947\* | 0.191 | 0.404\* | 0.422\* | 0.271\* |
| **Eclipse** | 0.520\* | 0.040 | 0.280\* | 0.380\* | 0.100 |  | 0.560\* | 0.020 | 0.320\* | 0.400\* | 0.080 |
| **Ant** | 0.340 | 0.040 | 0.160 | 0.260 | 0.340 |  | 0.240 | 0.000 | 0.100 | 0.280 | 0.300 |
| **JEdit** | 0.080 | 0.080 | 0.100 | 0.200\* | 0.480\* |  | 0.020 | 0.140\* | 0.080 | 0.220 | 0.400\* |

TABLE XX shows the median values and the mini boxplots of the overall ranking and classification performance under across-release prediction, together with the Cliff’s δ of the comparisons between NM/AM and CM. The “\*” following the Cliff’s δ means that the difference between the comparison pair is significant at the significance level of 0.05.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1. Overall ranking/classification performance under across-release prediciotn | | | | | | | | |
|  |  |  | median | mini boxplot | Cliff’s *δ* | |
|  |  |  | NM vs. CM | AM vs. CM |
| Ranking |  | CM | 0.217 | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\svm.bmp | 0.451\* | 0.544\* |
| NM | 0.245 |
| AM | 0.262 |
|  | CM | 0.211 | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\svm.bmp | 0.017\* | 0.051\* |
| NM | 0.259 |
| AM | 0.266 |
|  | CM | 0.364 | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\svm.bmp | 0.189 | 0.211 |
| NM | 0.362 |
| AM | 0.365 |
| classification | ER-BPP | CM | 0.411 | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\svm.bmp | 0.189 | 0.211 |
| NM | 0.549 |
| AM | 0.530 |
| ER-AVG | CM | 0.569 | F:\研二上\工作感知的网络度量缺陷预测\论文\intra\svm.bmp | 0.030 | 0.004 |
| NM | 0.577 |
| AM | 0.569 |
|  |  |  |  | 0.161 0.769 |  |  |

* 1. **Inter-project prediction (RQ4)**

Fig. 6 shows the boxplots which describe the distribution of the CEs at different cut-off values and the ER-BPPs / ER-AVGs obtained from 90 inter-project predictions for CM, NM, and AM.

|  |  |
| --- | --- |
| F:\研二上\工作感知的网络度量缺陷预测\论文\inter\svm_ce.bmp | F:\研二上\工作感知的网络度量缺陷预测\论文\inter\svm_er.bmp |
| Fig. 6. Ranking/Classification performance under the inter-project prediction | |

TABLE VII shows the Cliff’s *δ* of comparing NM/AM with CM under inter-project prediction. The figure is followed by “\*”, if the p-value obtained from the Wilcoxon signed-rank test is lower than the significance level of 0.05.

|  |
| --- |
| 1. The Cliff’s δ of comparing NM/AM vs. CM under inter-project prediction |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **ER-BPP** | **ER-AVG** |  |  |  |
| **NM vs. CM** | 0.020 | 0.043\* | 0.079 | 0.064 | 0.010 |
| **AM vs. CM** | -0.061 | 0.051\* | 0.080 | 0.066 | -0.004 |