

On Randomization and Reproducibility in Machine Learning for Software Engineering

Supplementary material: full tabular results

In this document, we provide full performance results for all datasets, models, metrics and libraries of interest in our TSE submission. Each datapoint is formatted as “median (range)”, taken over 100 independent runs related to the research question of interest:

- RQ1: randomized data fold partitions for 10-fold cross-validation;
- RQ2: random seeds in the machine learning (ML) training procedure.

We consider datasets from three predictive software engineering tasks: defect prediction (DP), predictive mutation testing (PMT) and code smell detection (CSD).

Results are reported for three ML libraries of interest: Weka, scikit-learn and R. This is done for default configurations for eight ML models: (1) Naive Bayes classifier (NB), (2) Linear Discriminant Analysis (LDA), (3) k -Nearest Neighbor classification (kNN), (4) Logistic Regression (LR), (5) Random Forest (RF), (6) Support Vector Machine (SVM), (7) Decision Tree (DT) and (8) Quadratic Discriminant Analysis (QDA). Further details can be found in our paper.

The exact data files that were fed to the different models, as well as raw outputs for all 100 runs, are provided as part of this supplementary material package.

Task	Dataset	NB	LDA	kNN	LR	RF	SVM	DT	QDA
DP	ant-1.7	0.55 (0.04)	0.51 (0.06)	0.49 (0.10)	0.48 (0.05)	0.54 (0.08)	0.39 (0.06)	0.52 (0.12)	0.55 (0.04)
	camel-1.6	0.31 (0.04)	0.22 (0.08)	0.33 (0.10)	0.15 (0.13)	0.28 (0.11)	0.01 (0.03)	0.29 (0.18)	0.35 (0.05)
	ivy-2.0	0.39 (0.08)	0.35 (0.19)	0.36 (0.13)	0.20 (0.15)	0.28 (0.16)	0.05 (0.17)	0.18 (0.31)	0.30 (0.19)
	jedit-4.3	0.21 (0.18)	0.09 (0.20)	0.00 (0.13)	0.00 (0.00)	0.00 (0.11)	NaN	0.00 (0.00)	0.35 (0.00)
	lucene-2.4	0.53 (0.05)	0.73 (0.04)	0.73 (0.06)	0.74 (0.06)	0.78 (0.05)	0.72 (0.04)	0.72 (0.08)	0.60 (0.06)
	tomcat	0.34 (0.05)	0.37 (0.09)	0.30 (0.10)	0.21 (0.18)	0.23 (0.13)	0.01 (0.03)	0.30 (0.24)	0.29 (0.08)
	equinox	0.57 (0.05)	0.63 (0.06)	0.67 (0.06)	0.66 (0.06)	0.69 (0.09)	0.65 (0.06)	0.68 (0.10)	0.61 (0.05)
	lucene	0.39 (0.04)	0.40 (0.07)	0.32 (0.10)	0.36 (0.15)	0.38 (0.12)	0.09 (0.08)	0.35 (0.26)	0.40 (0.08)
	mylyn	0.32 (0.04)	0.18 (0.05)	0.26 (0.06)	0.12 (0.07)	0.25 (0.09)	0.00 (0.02)	0.20 (0.16)	0.29 (0.04)
	pde	0.34 (0.02)	0.19 (0.06)	0.23 (0.07)	0.16 (0.08)	0.18 (0.10)	0.01 (0.00)	0.17 (0.12)	0.29 (0.04)
	jdt	0.47 (0.02)	0.48 (0.05)	0.45 (0.05)	0.46 (0.07)	0.52 (0.07)	0.29 (0.06)	0.53 (0.12)	0.49 (0.03)
PMT	apns	0.63 (0.03)	NaN	0.80 (0.02)	0.78 (0.03)	0.86 (0.02)	0.78 (0.02)	0.85 (0.03)	NaN
	assj	0.78 (0.02)	NaN	0.94 (0.01)	0.92 (0.00)	0.96 (0.00)	0.91 (0.00)	0.96 (0.00)	NaN
	joda	0.54 (0.00)	NaN	0.90 (0.00)	0.85 (0.00)	0.91 (0.00)	0.85 (0.00)	0.92 (0.00)	NaN
	lajf	0.39 (0.03)	NaN	0.88 (0.01)	0.80 (0.01)	0.90 (0.01)	0.79 (0.00)	0.91 (0.01)	NaN
	lang	0.67 (0.01)	NaN	0.91 (0.00)	0.90 (0.00)	0.92 (0.00)	0.90 (0.00)	0.93 (0.00)	NaN
	msg	0.47 (0.01)	NaN	0.86 (0.01)	0.75 (0.02)	0.90 (0.01)	0.74 (0.01)	0.91 (0.01)	NaN
	uaa	0.52 (0.01)	NaN	0.77 (0.01)	0.66 (0.01)	0.83 (0.01)	0.67 (0.01)	0.84 (0.01)	NaN
	vrapt	0.37 (0.01)	NaN	0.89 (0.01)	0.82 (0.00)	0.94 (0.00)	0.82 (0.00)	0.94 (0.00)	NaN
	wire	0.82 (0.01)	NaN	0.90 (0.01)	0.86 (0.01)	0.92 (0.01)	0.85 (0.01)	0.93 (0.00)	NaN
CSD	data-class	0.76 (0.04)	0.92 (0.02)	0.90 (0.03)	0.97 (0.04)	0.99 (0.01)	0.95 (0.02)	0.98 (0.03)	0.88 (0.05)
	feature-envy	0.80 (0.04)	0.82 (0.08)	0.79 (0.05)	0.93 (0.03)	0.94 (0.04)	0.89 (0.05)	0.93 (0.05)	0.81 (0.08)
	god-class	0.92 (0.03)	0.91 (0.04)	0.89 (0.05)	0.94 (0.04)	0.96 (0.02)	0.94 (0.02)	0.95 (0.04)	0.90 (0.03)
	long-method	0.91 (0.03)	0.93 (0.04)	0.91 (0.05)	0.99 (0.02)	0.99 (0.01)	0.95 (0.03)	0.98 (0.03)	0.90 (0.04)
	long_parameters_list	0.68 (0.07)	0.81 (0.05)	0.61 (0.08)	0.88 (0.06)	0.90 (0.05)	0.80 (0.05)	0.88 (0.06)	0.61 (0.07)
	switch_statements	0.75 (0.03)	0.73 (0.06)	0.67 (0.07)	0.82 (0.06)	0.82 (0.05)	0.74 (0.08)	0.77 (0.10)	0.76 (0.06)

TABLE I

RQ1: PERFORMANCE RESULTS FOR THE F-MEASURE IN WEKA

Task	Dataset	NB	LDA	kNN	LR	RF	SVM	DT	QDA
DP	ant-1.7	0.81 (0.02)	0.82 (0.02)	0.69 (0.09)	0.82 (0.02)	0.83 (0.02)	0.62 (0.02)	0.69 (0.13)	0.79 (0.03)
	camel-1.6	0.67 (0.04)	0.71 (0.03)	0.62 (0.07)	0.68 (0.08)	0.74 (0.04)	0.50 (0.01)	0.60 (0.16)	0.71 (0.04)
	ivy-2.0	0.76 (0.09)	0.78 (0.12)	0.63 (0.17)	0.75 (0.14)	0.79 (0.07)	0.51 (0.05)	0.59 (0.26)	0.64 (0.24)
	jedit-4.3	0.60 (0.31)	0.58 (0.35)	0.65 (0.23)	0.50 (0.19)	0.77 (0.16)	0.50 (0.00)	0.29 (0.32)	0.67 (0.00)
	lucene-2.4	0.73 (0.03)	0.75 (0.03)	0.68 (0.06)	0.75 (0.06)	0.79 (0.04)	0.66 (0.06)	0.67 (0.12)	0.73 (0.04)
	tomcat	0.81 (0.04)	0.80 (0.04)	0.64 (0.10)	0.79 (0.07)	0.82 (0.05)	0.50 (0.01)	0.63 (0.18)	0.77 (0.06)
	equinox	0.80 (0.02)	0.80 (0.03)	0.79 (0.05)	0.81 (0.03)	0.83 (0.04)	0.72 (0.04)	0.72 (0.12)	0.81 (0.02)
	lucene	0.76 (0.05)	0.79 (0.06)	0.65 (0.06)	0.76 (0.08)	0.82 (0.07)	0.52 (0.02)	0.66 (0.23)	0.79 (0.07)
	mylyn	0.66 (0.06)	0.78 (0.02)	0.64 (0.06)	0.77 (0.04)	0.80 (0.03)	0.50 (0.00)	0.65 (0.10)	0.74 (0.04)
	pde	0.69 (0.02)	0.72 (0.03)	0.58 (0.06)	0.70 (0.06)	0.73 (0.03)	0.50 (0.00)	0.61 (0.11)	0.69 (0.05)
	jdt	0.74 (0.03)	0.81 (0.01)	0.66 (0.06)	0.81 (0.03)	0.83 (0.03)	0.58 (0.02)	0.70 (0.09)	0.78 (0.02)
PMT	apns	0.82 (0.01)	NaN	0.85 (0.03)	0.86 (0.01)	0.93 (0.02)	0.79 (0.02)	0.90 (0.03)	NaN
	assj	0.66 (0.02)	NaN	0.85 (0.02)	0.70 (0.01)	0.94 (0.01)	0.50 (0.00)	0.86 (0.03)	NaN
	joda	0.68 (0.00)	NaN	0.86 (0.01)	0.73 (0.00)	0.92 (0.00)	0.50 (0.00)	0.87 (0.01)	NaN
	lafj	0.67 (0.00)	NaN	0.89 (0.01)	0.73 (0.01)	0.94 (0.01)	0.65 (0.00)	0.90 (0.01)	NaN
	lang	0.70 (0.01)	NaN	0.80 (0.01)	0.64 (0.18)	0.87 (0.01)	0.50 (0.00)	0.82 (0.01)	NaN
	msg	0.72 (0.00)	NaN	0.89 (0.01)	0.76 (0.01)	0.94 (0.00)	0.69 (0.01)	0.93 (0.01)	NaN
	uaa	0.75 (0.00)	NaN	0.87 (0.01)	0.79 (0.00)	0.92 (0.01)	0.71 (0.01)	0.91 (0.01)	NaN
	vrapt	0.61 (0.01)	NaN	0.86 (0.01)	0.63 (0.01)	0.94 (0.01)	0.52 (0.00)	0.89 (0.01)	NaN
	wire	0.81 (0.01)	NaN	0.89 (0.01)	0.85 (0.01)	0.94 (0.01)	0.76 (0.01)	0.86 (0.03)	NaN
CSD	data-class	0.92 (0.02)	0.99 (0.00)	0.93 (0.03)	1.00 (0.01)	1.00 (0.00)	0.96 (0.02)	0.98 (0.03)	0.96 (0.02)
	feature-envy	0.93 (0.02)	0.94 (0.04)	0.85 (0.05)	0.99 (0.02)	0.99 (0.01)	0.91 (0.04)	0.95 (0.04)	0.89 (0.07)
	god-class	0.97 (0.01)	0.99 (0.01)	0.91 (0.04)	0.99 (0.02)	1.00 (0.00)	0.95 (0.02)	0.96 (0.05)	0.96 (0.02)
	long-method	0.96 (0.03)	0.99 (0.03)	0.93 (0.05)	1.00 (0.00)	1.00 (0.00)	0.96 (0.03)	0.99 (0.02)	0.94 (0.04)
	long_parameters_list	0.82 (0.06)	0.94 (0.03)	0.71 (0.06)	0.97 (0.02)	0.98 (0.01)	0.84 (0.03)	0.90 (0.07)	0.74 (0.07)
	switch_statements	0.89 (0.02)	0.91 (0.03)	0.76 (0.09)	0.95 (0.02)	0.95 (0.01)	0.80 (0.06)	0.85 (0.10)	0.88 (0.04)

TABLE II

RQ1: PERFORMANCE RESULTS FOR THE AUC IN WEKA

Task	Dataset	NB	LDA	kNN	LR	RF	SVM	DT	QDA
DP	ant-1.7	0.43 (0.04)	0.43 (0.06)	0.36 (0.12)	0.42 (0.05)	0.44 (0.09)	0.37 (0.06)	0.40 (0.13)	0.42 (0.05)
	camel-1.6	0.21 (0.05)	0.19 (0.09)	0.17 (0.11)	0.15 (0.15)	0.22 (0.13)	0.01 (0.12)	0.17 (0.18)	0.26 (0.06)
	ivy-2.0	0.30 (0.10)	0.31 (0.19)	0.28 (0.16)	0.21 (0.20)	0.24 (0.18)	0.09 (0.26)	0.14 (0.37)	0.22 (0.21)
	jedit-4.3	0.20 (0.21)	0.07 (0.21)	-0.02 (0.14)	-0.01 (0.00)	-0.01 (0.10)	NaN	-0.01 (0.01)	0.33 (0.00)
	lucene-2.4	0.30 (0.07)	0.36 (0.09)	0.35 (0.13)	0.35 (0.13)	0.42 (0.13)	0.32 (0.11)	0.32 (0.21)	0.33 (0.12)
	tomcat	0.26 (0.06)	0.32 (0.10)	0.24 (0.11)	0.29 (0.18)	0.24 (0.16)	0.05 (0.12)	0.25 (0.26)	0.22 (0.09)
	equinox	0.42 (0.06)	0.45 (0.09)	0.46 (0.11)	0.47 (0.10)	0.50 (0.14)	0.45 (0.08)	0.47 (0.16)	0.45 (0.07)
	lucene	0.32 (0.05)	0.38 (0.08)	0.26 (0.11)	0.40 (0.15)	0.37 (0.15)	0.17 (0.11)	0.34 (0.24)	0.35 (0.08)
	mylyn	0.21 (0.06)	0.19 (0.06)	0.16 (0.07)	0.16 (0.08)	0.22 (0.10)	-0.01 (0.08)	0.15 (0.14)	0.22 (0.04)
	pde	0.24 (0.03)	0.19 (0.07)	0.12 (0.07)	0.18 (0.09)	0.17 (0.13)	0.06 (0.00)	0.13 (0.14)	0.21 (0.04)
	jdt	0.40 (0.03)	0.42 (0.05)	0.32 (0.07)	0.42 (0.07)	0.44 (0.08)	0.34 (0.07)	0.44 (0.14)	0.41 (0.03)
PMT	apns	0.45 (0.04)	NaN	0.60 (0.05)	0.56 (0.04)	0.70 (0.05)	0.57 (0.03)	0.69 (0.07)	NaN
	assj	0.14 (0.04)	NaN	0.61 (0.03)	0.12 (0.04)	0.76 (0.02)	0.03 (0.00)	0.75 (0.02)	NaN
	joda	0.24 (0.00)	NaN	0.59 (0.01)	0.21 (0.02)	0.65 (0.01)	0.04 (0.00)	0.67 (0.01)	NaN
	lafj	0.13 (0.01)	NaN	0.66 (0.02)	0.35 (0.03)	0.71 (0.02)	0.33 (0.01)	0.73 (0.02)	NaN
	lang	0.22 (0.01)	NaN	0.46 (0.02)	0.07 (0.05)	0.51 (0.01)	NaN	0.54 (0.01)	NaN
	msg	0.27 (0.01)	NaN	0.67 (0.02)	0.39 (0.05)	0.75 (0.01)	0.37 (0.02)	0.78 (0.01)	NaN
	uaa	0.30 (0.01)	NaN	0.59 (0.02)	0.42 (0.02)	0.69 (0.02)	0.41 (0.01)	0.70 (0.02)	NaN
	vrapt	0.10 (0.02)	NaN	0.62 (0.02)	0.14 (0.02)	0.78 (0.01)	0.11 (0.00)	0.78 (0.01)	NaN
	wire	0.48 (0.03)	NaN	0.70 (0.02)	0.55 (0.04)	0.77 (0.02)	0.55 (0.02)	0.79 (0.02)	NaN
CSD	data-class	0.62 (0.07)	0.87 (0.04)	0.84 (0.05)	0.95 (0.06)	0.98 (0.02)	0.92 (0.04)	0.97 (0.04)	0.82 (0.06)
	feature-envy	0.70 (0.06)	0.75 (0.10)	0.70 (0.09)	0.89 (0.05)	0.91 (0.05)	0.84 (0.07)	0.90 (0.08)	0.72 (0.11)
	god-class	0.87 (0.04)	0.88 (0.05)	0.84 (0.07)	0.91 (0.06)	0.94 (0.03)	0.91 (0.02)	0.93 (0.05)	0.86 (0.05)
	long-method	0.86 (0.04)	0.90 (0.06)	0.88 (0.06)	0.98 (0.03)	0.98 (0.02)	0.92 (0.04)	0.98 (0.04)	0.84 (0.06)
	long_parameters_list	0.52 (0.10)	0.73 (0.08)	0.41 (0.11)	0.82 (0.08)	0.86 (0.07)	0.74 (0.06)	0.82 (0.10)	0.39 (0.12)
	switch_statements	0.62 (0.05)	0.61 (0.08)	0.52 (0.11)	0.73 (0.08)	0.73 (0.08)	0.63 (0.12)	0.65 (0.14)	0.63 (0.09)

TABLE III

RQ1: PERFORMANCE RESULTS FOR THE MATTHEWS CORRELATION COEFFICIENT IN WEKA

Task	Dataset	NB	LDA	kNN	LR	RF	SVM	DT	QDA
DP	ant-1.7	0.55 (0.06)	0.50 (0.07)	0.46 (0.06)	0.47 (0.09)	0.53 (0.10)	0.48 (0.06)	0.46 (0.15)	0.54 (0.06)
	camel-1.6	0.32 (0.04)	0.22 (0.09)	0.14 (0.09)	0.16 (0.12)	0.27 (0.11)	0.02 (0.06)	0.34 (0.12)	0.35 (0.06)
	ivy-2.0	0.32 (0.18)	0.31 (0.24)	0.19 (0.20)	0.26 (0.20)	0.25 (0.21)	0.07 (0.20)	0.24 (0.24)	0.27 (0.22)
	jedit-4.3	0.12 (0.21)	0.04 (0.12)	0.00 (0.12)	0.05 (0.10)	0.00 (0.00)	0.00 (0.00)	0.00 (0.13)	0.02 (0.13)
	lucene-2.4	0.52 (0.05)	0.72 (0.04)	0.71 (0.05)	0.73 (0.05)	0.78 (0.05)	0.74 (0.07)	0.70 (0.10)	0.59 (0.06)
	tomcat	0.32 (0.07)	0.35 (0.13)	0.08 (0.14)	0.30 (0.18)	0.21 (0.18)	0.00 (0.03)	0.28 (0.17)	0.20 (0.16)
	equinox	0.55 (0.08)	0.63 (0.09)	0.68 (0.08)	0.63 (0.09)	0.69 (0.07)	0.59 (0.05)	0.61 (0.14)	0.59 (0.08)
	lucene	0.37 (0.08)	0.37 (0.14)	0.24 (0.12)	0.35 (0.17)	0.38 (0.16)	0.29 (0.18)	0.38 (0.22)	0.40 (0.14)
	mylyn	0.31 (0.06)	0.17 (0.06)	0.26 (0.08)	0.16 (0.07)	0.24 (0.09)	0.07 (0.07)	0.30 (0.10)	0.32 (0.05)
	pde	0.33 (0.03)	0.19 (0.07)	0.19 (0.09)	0.18 (0.08)	0.17 (0.10)	0.05 (0.07)	0.28 (0.12)	0.30 (0.06)
	jdt	0.46 (0.06)	0.47 (0.05)	0.43 (0.09)	0.50 (0.05)	0.51 (0.06)	0.45 (0.05)	0.45 (0.09)	0.48 (0.06)
PMT	apns	0.68 (0.02)	0.73 (0.02)	0.86 (0.02)	0.76 (0.02)	0.86 (0.03)	0.56 (0.02)	0.84 (0.04)	0.69 (0.13)
	assj	0.84 (0.07)	0.92 (0.00)	0.96 (0.00)	0.91 (0.00)	0.97 (0.00)	0.91 (0.00)	0.96 (0.00)	0.73 (0.03)
	joda	0.19 (0.01)	0.85 (0.00)	0.91 (0.00)	0.85 (0.00)	0.91 (0.00)	0.85 (0.00)	0.90 (0.00)	0.66 (0.01)
	lafj	0.04 (0.00)	0.78 (0.00)	0.90 (0.00)	0.80 (0.04)	0.90 (0.01)	0.79 (0.00)	0.89 (0.01)	0.42 (0.01)
	lang	0.50 (0.04)	0.90 (0.00)	0.92 (0.00)	0.90 (0.00)	0.92 (0.00)	0.90 (0.00)	0.91 (0.00)	0.72 (0.02)
	msg	0.42 (0.00)	0.72 (0.01)	0.90 (0.01)	0.72 (0.02)	0.90 (0.01)	0.72 (0.00)	0.88 (0.01)	0.62 (0.01)
	uaa	0.52 (0.01)	0.62 (0.01)	0.82 (0.01)	0.65 (0.01)	0.83 (0.01)	0.76 (0.01)	0.81 (0.01)	0.54 (0.01)
	vrapt	0.35 (0.02)	0.82 (0.00)	0.93 (0.00)	0.82 (0.00)	0.94 (0.00)	0.82 (0.00)	0.92 (0.01)	0.46 (0.01)
	wire	0.74 (0.01)	0.84 (0.01)	0.92 (0.01)	0.86 (0.00)	0.93 (0.01)	0.84 (0.00)	0.91 (0.01)	0.62 (0.36)
CSD	data-class	0.70 (0.02)	0.92 (0.05)	0.61 (0.07)	0.83 (0.12)	0.99 (0.02)	0.00 (0.01)	0.96 (0.04)	0.86 (0.08)
	feature-envy	0.72 (0.05)	0.81 (0.09)	0.48 (0.09)	0.65 (0.09)	0.94 (0.04)	0.15 (0.04)	0.90 (0.07)	0.79 (0.08)
	god-class	0.92 (0.03)	0.91 (0.05)	0.58 (0.08)	0.82 (0.05)	0.96 (0.02)	0.22 (0.06)	0.95 (0.05)	0.84 (0.08)
	long-method	0.91 (0.04)	0.93 (0.05)	0.44 (0.09)	0.78 (0.06)	0.99 (0.02)	0.15 (0.05)	0.98 (0.03)	0.89 (0.05)
	long_parameters_list	0.57 (0.14)	0.80 (0.06)	0.55 (0.08)	0.26 (0.13)	0.90 (0.05)	0.00 (0.00)	0.88 (0.09)	0.60 (0.09)
	switch_statements	0.76 (0.05)	0.72 (0.08)	0.28 (0.14)	0.46 (0.13)	0.82 (0.07)	0.00 (0.00)	0.76 (0.13)	0.74 (0.08)

TABLE IV

RQ1: PERFORMANCE RESULTS FOR THE F-MEASURE IN SCIKIT-LEARN

Task	Dataset	NB	LDA	kNN	LR	RF	SVM	DT	QDA
DP	ant-1.7	0.81 (0.03)	0.82 (0.03)	0.77 (0.04)	0.78 (0.03)	0.84 (0.03)	0.75 (0.04)	0.65 (0.09)	0.80 (0.03)
	camel-1.6	0.68 (0.03)	0.71 (0.03)	0.60 (0.06)	0.64 (0.05)	0.73 (0.03)	0.65 (0.06)	0.59 (0.07)	0.72 (0.04)
	ivy-2.0	0.75 (0.15)	0.80 (0.10)	0.76 (0.12)	0.61 (0.18)	0.80 (0.14)	0.64 (0.25)	0.59 (0.19)	0.66 (0.20)
	jedit-4.3	0.61 (0.36)	0.68 (0.31)	0.67 (0.25)	0.56 (0.27)	0.77 (0.30)	0.40 (0.53)	0.67 (0.22)	0.53 (0.12)
	lucene-2.4	0.73 (0.04)	0.75 (0.05)	0.67 (0.07)	0.75 (0.05)	0.79 (0.05)	0.64 (0.18)	0.63 (0.11)	0.74 (0.04)
	tomcat	0.79 (0.06)	0.80 (0.06)	0.67 (0.10)	0.71 (0.09)	0.82 (0.06)	0.55 (0.19)	0.62 (0.11)	0.59 (0.24)
	equinox	0.81 (0.04)	0.81 (0.04)	0.82 (0.05)	0.77 (0.07)	0.83 (0.05)	0.81 (0.04)	0.70 (0.10)	0.81 (0.04)
	lucene	0.77 (0.10)	0.82 (0.07)	0.71 (0.10)	0.72 (0.10)	0.82 (0.09)	0.68 (0.11)	0.64 (0.11)	0.79 (0.10)
	mylyn	0.69 (0.03)	0.79 (0.02)	0.74 (0.05)	0.77 (0.03)	0.80 (0.03)	0.70 (0.05)	0.64 (0.07)	0.75 (0.04)
	pde	0.69 (0.02)	0.73 (0.03)	0.66 (0.05)	0.62 (0.06)	0.73 (0.04)	0.58 (0.06)	0.59 (0.07)	0.70 (0.04)
	jdt	0.74 (0.03)	0.81 (0.02)	0.76 (0.04)	0.80 (0.02)	0.83 (0.03)	0.73 (0.02)	0.66 (0.07)	0.79 (0.03)
PMT	apns	0.83 (0.01)	0.84 (0.01)	0.91 (0.02)	0.85 (0.01)	0.93 (0.02)	0.89 (0.01)	0.87 (0.03)	0.75 (0.09)
	assj	0.62 (0.01)	0.66 (0.00)	0.91 (0.01)	0.60 (0.01)	0.94 (0.01)	0.49 (0.08)	0.89 (0.01)	0.69 (0.01)
	joda	0.67 (0.00)	0.70 (0.00)	0.89 (0.01)	0.65 (0.03)	0.92 (0.00)	0.67 (0.19)	0.86 (0.01)	0.72 (0.00)
	lafj	0.67 (0.00)	0.72 (0.00)	0.91 (0.01)	0.70 (0.06)	0.93 (0.01)	0.76 (0.06)	0.89 (0.01)	0.74 (0.00)
	lang	0.65 (0.00)	0.69 (0.00)	0.84 (0.01)	0.66 (0.01)	0.88 (0.01)	0.58 (0.15)	0.80 (0.01)	0.69 (0.00)
	msg	0.70 (0.00)	0.73 (0.00)	0.92 (0.01)	0.80 (0.02)	0.94 (0.00)	0.88 (0.08)	0.89 (0.01)	0.74 (0.00)
	uaa	0.73 (0.00)	0.77 (0.00)	0.90 (0.01)	0.78 (0.00)	0.92 (0.00)	0.87 (0.00)	0.88 (0.01)	0.78 (0.00)
	vrapt	0.60 (0.00)	0.60 (0.00)	0.92 (0.01)	0.61 (0.01)	0.94 (0.01)	0.75 (0.04)	0.90 (0.01)	0.67 (0.01)
	wire	0.80 (0.01)	0.81 (0.00)	0.92 (0.01)	0.83 (0.01)	0.94 (0.01)	0.80 (0.01)	0.90 (0.01)	0.57 (0.11)
CSD	data-class	0.80 (0.02)	0.99 (0.01)	0.75 (0.05)	0.94 (0.05)	1.00 (0.00)	0.72 (0.11)	0.97 (0.03)	0.96 (0.02)
	feature-envy	0.89 (0.03)	0.95 (0.04)	0.67 (0.06)	0.77 (0.07)	0.99 (0.01)	0.72 (0.06)	0.93 (0.06)	0.90 (0.05)
	god-class	0.98 (0.01)	0.99 (0.01)	0.82 (0.05)	0.95 (0.04)	1.00 (0.01)	0.82 (0.04)	0.96 (0.04)	0.94 (0.04)
	long-method	0.98 (0.02)	0.99 (0.03)	0.71 (0.07)	0.92 (0.04)	1.00 (0.00)	0.76 (0.05)	0.99 (0.02)	0.94 (0.04)
	long_parameters_list	0.77 (0.07)	0.95 (0.03)	0.72 (0.05)	0.51 (0.10)	0.98 (0.02)	0.48 (0.16)	0.91 (0.06)	0.75 (0.07)
	switch_statements	0.90 (0.03)	0.92 (0.03)	0.56 (0.08)	0.67 (0.11)	0.95 (0.02)	0.54 (0.27)	0.83 (0.09)	0.88 (0.05)

TABLE V

RQ1: PERFORMANCE RESULTS FOR THE AUC IN SCIKIT-LEARN

Task	Dataset	NB	LDA	kNN	LR	RF	SVM	DT	QDA
DP	ant-1.7	0.43 (0.06)	0.43 (0.08)	0.34 (0.08)	0.41 (0.10)	0.45 (0.10)	0.41 (0.06)	0.31 (0.17)	0.41 (0.06)
	camel-1.6	0.22 (0.05)	0.19 (0.09)	0.05 (0.11)	0.15 (0.15)	0.22 (0.12)	0.01 (0.11)	0.18 (0.15)	0.26 (0.07)
	ivy-2.0	0.25 (0.20)	0.28 (0.29)	0.15 (0.23)	0.23 (0.20)	0.24 (0.22)	0.07 (0.22)	0.16 (0.29)	0.21 (0.23)
	jedit-4.3	0.11 (0.24)	0.03 (0.14)	0.00 (0.12)	0.04 (0.11)	-0.00 (0.01)	0.00 (0.00)	-0.01 (0.15)	0.02 (0.13)
	lucene-2.4	0.30 (0.06)	0.36 (0.09)	0.25 (0.14)	0.33 (0.10)	0.43 (0.15)	0.00 (0.07)	0.27 (0.21)	0.33 (0.10)
	tomcat	0.25 (0.07)	0.32 (0.13)	0.05 (0.17)	0.29 (0.19)	0.23 (0.20)	-0.01 (0.05)	0.22 (0.18)	0.12 (0.23)
	equinox	0.42 (0.09)	0.45 (0.12)	0.50 (0.11)	0.44 (0.14)	0.50 (0.15)	0.45 (0.07)	0.39 (0.19)	0.43 (0.10)
	lucene	0.32 (0.10)	0.36 (0.15)	0.25 (0.15)	0.37 (0.20)	0.39 (0.20)	0.33 (0.21)	0.33 (0.24)	0.36 (0.14)
	mylyn	0.21 (0.07)	0.19 (0.08)	0.22 (0.09)	0.18 (0.10)	0.22 (0.11)	0.10 (0.11)	0.20 (0.12)	0.24 (0.06)
	pde	0.23 (0.04)	0.18 (0.08)	0.14 (0.10)	0.19 (0.13)	0.17 (0.12)	0.06 (0.10)	0.17 (0.13)	0.21 (0.07)
	jdt	0.40 (0.05)	0.42 (0.05)	0.34 (0.09)	0.43 (0.05)	0.44 (0.08)	0.42 (0.05)	0.32 (0.13)	0.40 (0.07)
PMT	apns	0.50 (0.02)	0.48 (0.03)	0.70 (0.04)	0.52 (0.03)	0.71 (0.05)	0.43 (0.02)	0.67 (0.07)	0.40 (0.16)
	assj	0.06 (0.03)	0.10 (0.06)	0.71 (0.03)	0.00 (0.02)	0.77 (0.02)	0.00 (0.00)	0.72 (0.03)	0.21 (0.02)
	joda	0.11 (0.00)	0.14 (0.01)	0.65 (0.02)	0.00 (0.02)	0.66 (0.01)	0.00 (0.00)	0.63 (0.02)	0.30 (0.01)
	lafj	0.06 (0.01)	0.32 (0.01)	0.70 (0.02)	0.29 (0.19)	0.71 (0.01)	0.00 (0.00)	0.69 (0.02)	0.23 (0.01)
	lang	0.14 (0.02)	0.07 (0.01)	0.50 (0.02)	0.00 (0.00)	0.53 (0.02)	0.00 (0.00)	0.50 (0.02)	0.22 (0.01)
	msg	0.27 (0.00)	0.33 (0.01)	0.75 (0.01)	0.00 (0.10)	0.75 (0.02)	0.00 (0.03)	0.72 (0.02)	0.34 (0.01)
	uaa	0.29 (0.02)	0.35 (0.01)	0.66 (0.02)	0.40 (0.01)	0.69 (0.02)	0.56 (0.01)	0.65 (0.02)	0.34 (0.01)
	vrapt	0.09 (0.02)	0.08 (0.02)	0.76 (0.01)	0.02 (0.04)	0.79 (0.01)	0.00 (0.00)	0.74 (0.02)	0.21 (0.01)
	wire	0.41 (0.02)	0.48 (0.04)	0.76 (0.03)	0.53 (0.01)	0.78 (0.02)	0.43 (0.01)	0.74 (0.03)	0.14 (0.20)
CSD	data-class	0.53 (0.03)	0.88 (0.06)	0.40 (0.10)	0.75 (0.17)	0.99 (0.03)	0.00 (0.00)	0.95 (0.06)	0.81 (0.09)
	feature-envy	0.60 (0.06)	0.75 (0.11)	0.31 (0.12)	0.48 (0.14)	0.91 (0.05)	0.17 (0.08)	0.86 (0.10)	0.71 (0.11)
	god-class	0.88 (0.05)	0.88 (0.06)	0.42 (0.12)	0.73 (0.07)	0.94 (0.03)	0.19 (0.11)	0.93 (0.06)	0.77 (0.11)
	long-method	0.86 (0.05)	0.90 (0.06)	0.31 (0.13)	0.67 (0.09)	0.99 (0.02)	0.18 (0.11)	0.98 (0.04)	0.85 (0.07)
	long_parameters_list	0.38 (0.19)	0.73 (0.09)	0.33 (0.12)	0.07 (0.21)	0.86 (0.07)	0.00 (0.00)	0.82 (0.12)	0.38 (0.12)
	switch_statements	0.65 (0.07)	0.62 (0.10)	0.07 (0.17)	0.23 (0.17)	0.74 (0.08)	0.00 (0.00)	0.65 (0.17)	0.63 (0.11)

TABLE VI

RQ1: PERFORMANCE RESULTS FOR THE MATTHEWS CORRELATION COEFFICIENT IN SCIKIT-LEARN

Task	Dataset	NB	LDA	kNN	LR	RF	SVM	DT	QDA
DP	ant-1.7	0.55 (0.03)	0.51 (0.06)	0.49 (0.08)	0.50 (0.05)	0.55 (0.06)	0.41 (0.08)	0.49 (0.14)	0.55 (0.04)
	camel-1.6	0.32 (0.03)	0.22 (0.08)	0.26 (0.10)	0.19 (0.07)	0.25 (0.07)	0.10 (0.08)	0.25 (0.15)	0.35 (0.05)
	ivy-2.0	0.35 (0.11)	0.35 (0.19)	0.30 (0.12)	0.28 (0.19)	0.29 (0.13)	0.24 (0.17)	0.28 (0.23)	0.31 (0.18)
	jedit-4.3	0.16 (0.17)	0.10 (0.12)	0.17 (0.23)	0.11 (0.02)	NaN	NaN	NaN	NaN
	lucene-2.4	0.52 (0.04)	0.73 (0.04)	0.76 (0.05)	0.74 (0.05)	0.78 (0.04)	0.75 (0.04)	0.74 (0.10)	0.60 (0.06)
	tomcat	0.32 (0.05)	NaN	0.18 (0.15)	0.25 (0.09)	0.23 (0.14)	NaN	0.32 (0.20)	NaN
	equinox	0.56 (0.05)	0.63 (0.07)	0.69 (0.08)	0.65 (0.06)	0.69 (0.05)	0.66 (0.06)	0.64 (0.10)	NaN
	lucene	0.38 (0.04)	0.40 (0.07)	0.28 (0.10)	0.32 (0.10)	0.37 (0.10)	0.33 (0.16)	0.31 (0.20)	NaN
	mylyn	0.32 (0.05)	0.18 (0.05)	0.27 (0.14)	0.15 (0.04)	0.16 (0.07)	0.10 (0.06)	0.18 (0.10)	NaN
	pde	0.33 (0.02)	0.19 (0.06)	0.16 (0.07)	0.18 (0.04)	0.16 (0.07)	0.05 (0.09)	0.24 (0.13)	NaN
	jdt	0.46 (0.03)	0.48 (0.05)	0.52 (0.07)	0.50 (0.04)	0.52 (0.06)	0.46 (0.04)	0.50 (0.12)	NaN
PMT	apns	0.69 (0.02)	0.74 (0.02)	0.83 (0.03)	0.77 (0.02)	0.88 (0.02)	0.80 (0.02)	0.87 (0.04)	NaN
	assj	0.77 (0.03)	0.91 (0.00)	0.95 (0.00)	0.91 (0.00)	0.97 (0.00)	0.93 (0.00)	0.96 (0.00)	0.73 (0.03)
	joda	0.51 (0.00)	0.85 (0.00)	0.90 (0.01)	0.85 (0.00)	0.93 (0.00)	0.87 (0.00)	0.91 (0.00)	0.66 (0.01)
	lafj	0.30 (0.04)	0.79 (0.00)	0.88 (0.01)	0.80 (0.00)	0.91 (0.00)	0.83 (0.00)	0.90 (0.00)	0.40 (0.01)
	lang	0.69 (0.02)	0.90 (0.00)	0.92 (0.00)	0.90 (0.00)	0.94 (0.00)	0.91 (0.00)	0.93 (0.00)	0.70 (0.01)
	msg	0.48 (0.00)	0.72 (0.01)	0.85 (0.01)	0.73 (0.00)	0.91 (0.01)	0.79 (0.01)	0.91 (0.00)	0.62 (0.01)
	uaa	0.52 (0.01)	0.62 (0.01)	0.80 (0.01)	0.64 (0.01)	0.86 (0.01)	0.77 (0.01)	0.84 (0.01)	0.54 (0.00)
	vrapt	0.36 (0.02)	0.82 (0.00)	0.90 (0.01)	0.82 (0.00)	0.94 (0.00)	0.85 (0.00)	0.93 (0.00)	0.46 (0.01)
	wire	0.74 (0.01)	NaN	0.91 (0.01)	0.86 (0.00)	0.94 (0.00)	0.93 (0.01)	0.93 (0.00)	NaN
CSD	data-class	0.76 (0.03)	NaN	0.92 (0.03)	0.90 (0.07)	0.99 (0.01)	0.26 (0.10)	0.97 (0.03)	NaN
	feature-envy	0.78 (0.05)	NaN	0.80 (0.06)	0.84 (0.08)	0.94 (0.03)	0.71 (0.07)	0.93 (0.03)	NaN
	god-class	0.93 (0.02)	NaN	0.91 (0.03)	0.86 (0.07)	0.96 (0.02)	NaN	0.97 (0.01)	NaN
	long-method	0.91 (0.03)	NaN	0.89 (0.05)	0.92 (0.06)	0.99 (0.01)	0.74 (0.04)	0.98 (0.05)	NaN
	long_parameters_list	0.64 (0.09)	NaN	0.69 (0.08)	0.83 (0.09)	0.91 (0.04)	0.42 (0.08)	0.88 (0.04)	NaN
	switch_statements	0.74 (0.06)	NaN	0.71 (0.06)	0.78 (0.09)	0.83 (0.04)	0.10 (0.06)	0.82 (0.09)	NaN

TABLE VII

RQ1: PERFORMANCE RESULTS FOR THE F-MEASURE IN R

Task	Dataset	NB	LDA	kNN	LR	RF	SVM	DT	QDA
DP	ant-1.7	0.80 (0.02)	0.82 (0.02)	0.80 (0.03)	0.81 (0.02)	0.83 (0.02)	0.76 (0.03)	0.77 (0.07)	0.79 (0.03)
	camel-1.6	0.67 (0.04)	0.71 (0.03)	0.69 (0.04)	0.70 (0.03)	0.73 (0.03)	0.65 (0.06)	0.62 (0.10)	0.71 (0.04)
	ivy-2.0	0.74 (0.15)	0.78 (0.12)	0.69 (0.09)	0.75 (0.13)	0.79 (0.05)	0.64 (0.14)	0.65 (0.21)	0.64 (0.24)
	jedit-4.3	0.61 (0.42)	0.58 (0.35)	0.73 (0.25)	0.60 (0.32)	0.77 (0.10)	0.51 (0.51)	0.26 (0.26)	NaN
	lucene-2.4	0.72 (0.03)	0.75 (0.03)	0.76 (0.05)	0.76 (0.03)	0.79 (0.04)	0.76 (0.03)	0.71 (0.09)	0.73 (0.04)
	tomcat	0.78 (0.05)	NaN	0.73 (0.05)	0.80 (0.03)	0.83 (0.03)	0.65 (0.09)	0.72 (0.10)	NaN
	equinox	0.80 (0.02)	0.80 (0.03)	0.82 (0.04)	0.80 (0.03)	0.83 (0.04)	0.79 (0.03)	0.75 (0.06)	NaN
	lucene	0.77 (0.05)	0.79 (0.06)	0.71 (0.08)	0.78 (0.07)	0.82 (0.05)	0.66 (0.11)	0.70 (0.15)	NaN
	mylyn	0.66 (0.06)	0.78 (0.02)	0.73 (0.05)	0.78 (0.02)	0.80 (0.02)	0.66 (0.07)	0.59 (0.11)	NaN
	pde	0.68 (0.02)	0.72 (0.03)	0.65 (0.05)	0.72 (0.02)	0.74 (0.02)	0.60 (0.09)	0.64 (0.09)	NaN
	jdt	0.74 (0.02)	0.81 (0.01)	0.80 (0.03)	0.81 (0.01)	0.83 (0.02)	0.78 (0.02)	0.72 (0.07)	NaN
PMT	apns	0.82 (0.01)	0.84 (0.01)	0.89 (0.02)	0.85 (0.01)	0.94 (0.01)	0.87 (0.01)	0.90 (0.03)	NaN
	assj	0.63 (0.01)	0.65 (0.00)	0.91 (0.01)	0.65 (0.01)	0.96 (0.00)	0.78 (0.03)	0.80 (0.04)	0.67 (0.02)
	joda	0.66 (0.00)	0.70 (0.00)	0.89 (0.00)	0.71 (0.00)	0.93 (0.00)	0.79 (0.00)	0.78 (0.02)	0.72 (0.00)
	lafj	0.67 (0.00)	0.72 (0.00)	0.90 (0.01)	0.73 (0.00)	0.93 (0.00)	0.81 (0.00)	0.79 (0.01)	0.75 (0.00)
	lang	0.66 (0.00)	0.69 (0.00)	0.84 (0.01)	0.69 (0.00)	0.90 (0.00)	0.72 (0.01)	0.61 (0.01)	0.69 (0.00)
	msg	0.69 (0.00)	0.73 (0.00)	0.91 (0.01)	0.74 (0.00)	0.95 (0.00)	0.83 (0.00)	0.86 (0.01)	0.74 (0.00)
	uaa	0.73 (0.00)	0.77 (0.00)	0.90 (0.01)	0.78 (0.00)	0.93 (0.00)	0.87 (0.00)	0.85 (0.01)	0.78 (0.00)
	vrapt	0.59 (0.01)	0.59 (0.01)	0.89 (0.01)	0.60 (0.01)	0.95 (0.01)	0.79 (0.00)	0.84 (0.02)	0.67 (0.01)
	wire	0.80 (0.01)	NaN	0.92 (0.01)	0.83 (0.00)	0.95 (0.01)	0.89 (0.01)	0.85 (0.03)	NaN
CSD	data-class	0.89 (0.02)	NaN	0.98 (0.01)	0.95 (0.03)	1.00 (0.00)	0.58 (0.10)	0.97 (0.03)	NaN
	feature-envy	0.90 (0.03)	NaN	0.92 (0.03)	0.91 (0.06)	0.99 (0.00)	0.79 (0.09)	0.95 (0.02)	NaN
	god-class	0.98 (0.01)	NaN	0.98 (0.02)	0.92 (0.06)	1.00 (0.00)	0.55 (0.12)	0.97 (0.01)	NaN
	long-method	0.97 (0.02)	NaN	0.98 (0.02)	0.98 (0.03)	1.00 (0.00)	0.83 (0.08)	1.00 (0.02)	NaN
	long_parameters_list	0.77 (0.08)	NaN	0.86 (0.04)	0.89 (0.09)	0.98 (0.01)	0.66 (0.08)	0.94 (0.05)	NaN
	switch_statements	0.89 (0.04)	NaN	0.89 (0.03)	0.90 (0.10)	0.95 (0.01)	0.55 (0.09)	0.93 (0.05)	NaN

TABLE VIII

RQ1: PERFORMANCE RESULTS FOR THE AUC IN R

Task	Dataset	NB	LDA	kNN	LR	RF	SVM	DT	QDA
DP	ant-1.7	0.42 (0.04)	0.43 (0.06)	0.37 (0.10)	0.42 (0.05)	0.45 (0.07)	0.35 (0.08)	0.38 (0.17)	0.42 (0.05)
	camel-1.6	0.22 (0.04)	0.19 (0.09)	0.13 (0.11)	0.17 (0.08)	0.21 (0.09)	0.11 (0.11)	0.18 (0.15)	0.26 (0.06)
	ivy-2.0	0.25 (0.12)	0.30 (0.19)	0.26 (0.16)	0.25 (0.20)	0.28 (0.16)	0.26 (0.17)	0.21 (0.26)	0.22 (0.19)
	jedit-4.3	0.16 (0.21)	0.07 (0.21)	-0.01 (0.33)	-0.02 (0.12)	-0.01 (0.01)	-0.01 (0.01)	0.00 (0.00)	NaN
	lucene-2.4	0.30 (0.05)	0.36 (0.09)	0.40 (0.13)	0.37 (0.10)	0.41 (0.10)	0.35 (0.12)	0.32 (0.23)	0.33 (0.12)
	tomcat	0.25 (0.06)	NaN	0.14 (0.15)	0.24 (0.10)	0.27 (0.18)	-0.02 (0.03)	0.28 (0.20)	NaN
	equinox	0.42 (0.05)	0.45 (0.10)	0.50 (0.12)	0.48 (0.08)	0.51 (0.08)	0.46 (0.09)	0.43 (0.16)	NaN
	lucene	0.32 (0.05)	0.38 (0.08)	0.27 (0.13)	0.33 (0.10)	0.39 (0.12)	0.36 (0.15)	0.29 (0.20)	NaN
	mylyn	0.21 (0.06)	0.19 (0.06)	0.21 (0.13)	0.17 (0.06)	0.16 (0.09)	0.13 (0.08)	0.18 (0.13)	NaN
	pde	0.23 (0.02)	0.19 (0.07)	0.10 (0.08)	0.19 (0.06)	0.16 (0.08)	0.08 (0.13)	0.21 (0.13)	NaN
	jdt	0.39 (0.03)	0.42 (0.04)	0.42 (0.08)	0.45 (0.05)	0.46 (0.07)	0.41 (0.04)	0.42 (0.13)	NaN
PMT	apns	0.50 (0.02)	0.49 (0.04)	0.64 (0.05)	0.54 (0.04)	0.75 (0.03)	0.58 (0.04)	0.72 (0.09)	NaN
	assj	0.10 (0.03)	0.00 (0.05)	0.65 (0.03)	0.00 (0.04)	0.78 (0.01)	0.36 (0.02)	0.70 (0.01)	0.17 (0.04)
	joda	0.22 (0.00)	0.14 (0.01)	0.60 (0.02)	0.16 (0.01)	0.70 (0.01)	0.43 (0.01)	0.60 (0.01)	0.30 (0.01)
	lafj	0.12 (0.02)	0.34 (0.01)	0.64 (0.02)	0.35 (0.01)	0.73 (0.01)	0.47 (0.01)	0.70 (0.00)	0.23 (0.01)
	lang	0.18 (0.02)	0.07 (0.01)	0.48 (0.02)	0.06 (0.01)	0.58 (0.01)	0.25 (0.01)	0.46 (0.00)	0.21 (0.01)
	msg	0.27 (0.01)	0.33 (0.01)	0.66 (0.02)	0.37 (0.01)	0.78 (0.01)	0.51 (0.01)	0.78 (0.00)	0.35 (0.01)
	uaa	0.29 (0.02)	0.34 (0.01)	0.63 (0.03)	0.38 (0.01)	0.73 (0.01)	0.59 (0.01)	0.71 (0.01)	0.35 (0.01)
	vrapt	0.09 (0.02)	0.00 (0.00)	0.65 (0.02)	0.00 (0.00)	0.81 (0.01)	0.42 (0.02)	0.77 (0.00)	0.19 (0.01)
	wire	0.42 (0.02)	NaN	0.73 (0.04)	0.56 (0.01)	0.82 (0.02)	0.79 (0.02)	0.79 (0.01)	NaN
CSD	data-class	0.61 (0.06)	NaN	0.87 (0.04)	0.85 (0.11)	0.98 (0.02)	0.32 (0.07)	0.96 (0.04)	NaN
	feature-envy	0.67 (0.08)	NaN	0.72 (0.09)	0.76 (0.12)	0.91 (0.04)	0.56 (0.15)	0.89 (0.04)	NaN
	god-class	0.89 (0.04)	NaN	0.87 (0.05)	0.80 (0.11)	0.94 (0.02)	0.00 (0.00)	0.96 (0.02)	NaN
	long-method	0.87 (0.05)	NaN	0.83 (0.07)	0.88 (0.09)	0.99 (0.02)	0.62 (0.06)	0.98 (0.07)	NaN
	long_parameters_list	0.42 (0.15)	NaN	0.54 (0.11)	0.73 (0.14)	0.87 (0.06)	0.38 (0.10)	0.83 (0.06)	NaN
	switch_statements	0.60 (0.08)	NaN	0.58 (0.08)	0.67 (0.14)	0.74 (0.07)	0.15 (0.10)	0.73 (0.13)	NaN

TABLE IX

RQ1: PERFORMANCE RESULTS FOR THE MATTHEWS CORRELATION COEFFICIENT IN R

Task	Dataset	F-measure			AUC			Matthews		
		RF	SVM	DT	RF	SVM	DT	RF	SVM	DT
DP	ant-1.7	0.51 (0.06)	0.38 (0.01)	0.56 (0.00)	0.83 (0.02)	0.62 (0.00)	0.76 (0.00)	0.42 (0.08)	0.37 (0.02)	0.45 (0.00)
	camel-1.6	0.27 (0.08)	0.00 (0.04)	0.17 (0.00)	0.74 (0.02)	0.50 (0.01)	0.63 (0.00)	0.21 (0.08)	-0.02 (0.13)	0.14 (0.00)
	ivy-2.0	0.28 (0.11)	0.00 (0.05)	0.09 (0.00)	0.79 (0.04)	0.50 (0.01)	0.53 (0.00)	0.25 (0.13)	-0.03 (0.09)	0.09 (0.00)
	jedit-4.3	0.00 (0.00)	NaN	0.00 (0.00)	0.74 (0.14)	0.50 (0.00)	0.23 (0.00)	-0.01 (0.01)	NaN	-0.01 (0.00)
	lucene-2.4	0.78 (0.03)	0.71 (0.00)	0.71 (0.00)	0.78 (0.02)	0.66 (0.00)	0.67 (0.00)	0.42 (0.12)	0.31 (0.01)	0.27 (0.00)
	tomcat	0.25 (0.08)	NaN	0.19 (0.00)	0.82 (0.04)	0.50 (0.00)	0.66 (0.00)	0.28 (0.10)	NaN	0.20 (0.00)
	equinox	0.70 (0.04)	0.65 (0.01)	0.63 (0.00)	0.83 (0.02)	0.72 (0.01)	0.73 (0.00)	0.51 (0.08)	0.45 (0.01)	0.44 (0.00)
	lucene	0.36 (0.08)	0.06 (0.06)	0.28 (0.00)	0.82 (0.03)	0.51 (0.02)	0.68 (0.00)	0.35 (0.09)	0.13 (0.10)	0.27 (0.00)
	mylyn	0.24 (0.05)	0.00 (0.01)	0.07 (0.00)	0.80 (0.01)	0.50 (0.00)	0.70 (0.00)	0.22 (0.06)	-0.01 (0.04)	0.10 (0.00)
	pde	0.18 (0.08)	NaN	0.14 (0.00)	0.72 (0.02)	0.50 (0.00)	0.66 (0.00)	0.16 (0.10)	NaN	0.15 (0.00)
	jdt	0.51 (0.05)	0.28 (0.01)	0.52 (0.00)	0.83 (0.02)	0.58 (0.00)	0.74 (0.00)	0.43 (0.06)	0.32 (0.01)	0.44 (0.00)
PMT	apns	0.86 (0.01)	0.79 (0.00)	0.87 (0.00)	0.93 (0.01)	0.79 (0.00)	0.91 (0.00)	0.70 (0.02)	0.58 (0.00)	0.71 (0.00)
	assj	0.97 (0.00)	0.91 (0.00)	0.96 (0.00)	0.94 (0.00)	0.50 (0.00)	0.92 (0.00)	0.77 (0.01)	0.03 (0.00)	0.75 (0.00)
	joda	0.91 (0.00)	0.85 (0.00)	0.92 (0.00)	0.92 (0.00)	0.50 (0.00)	0.89 (0.00)	0.65 (0.00)	0.04 (0.00)	0.67 (0.00)
	lafj	0.90 (0.00)	0.79 (0.00)	0.91 (0.00)	0.94 (0.00)	0.65 (0.00)	0.90 (0.00)	0.71 (0.01)	0.33 (0.00)	0.72 (0.00)
	lang	0.92 (0.00)	0.90 (0.00)	0.93 (0.00)	0.87 (0.00)	0.50 (0.00)	0.83 (0.00)	0.51 (0.01)	NaN	0.53 (0.00)
	msg	0.90 (0.00)	0.74 (0.00)	0.91 (0.00)	0.94 (0.00)	0.69 (0.00)	0.93 (0.00)	0.75 (0.01)	0.38 (0.00)	0.78 (0.00)
	uaa	0.83 (0.01)	0.67 (0.00)	0.84 (0.00)	0.92 (0.00)	0.71 (0.00)	0.91 (0.00)	0.68 (0.01)	0.42 (0.00)	0.70 (0.00)
	vrapt	0.94 (0.00)	0.82 (0.00)	0.94 (0.00)	0.94 (0.00)	0.52 (0.00)	0.90 (0.00)	0.78 (0.01)	0.11 (0.00)	0.78 (0.00)
	wire	0.92 (0.00)	0.85 (0.00)	0.93 (0.00)	0.94 (0.00)	0.76 (0.00)	0.92 (0.00)	0.77 (0.01)	0.55 (0.00)	0.80 (0.00)
CSD	data-class	0.99 (0.02)	0.94 (0.00)	0.97 (0.00)	1.00 (0.00)	0.95 (0.00)	0.98 (0.00)	0.98 (0.04)	0.91 (0.01)	0.95 (0.00)
	feature-envy	0.94 (0.04)	0.88 (0.00)	0.94 (0.00)	0.99 (0.00)	0.91 (0.00)	0.96 (0.00)	0.91 (0.05)	0.83 (0.00)	0.91 (0.00)
	god-class	0.96 (0.02)	0.94 (0.00)	0.96 (0.00)	1.00 (0.00)	0.95 (0.00)	0.98 (0.00)	0.94 (0.02)	0.91 (0.01)	0.94 (0.00)
	long-method	0.99 (0.01)	0.96 (0.00)	0.96 (0.00)	1.00 (0.00)	0.97 (0.00)	0.98 (0.00)	0.98 (0.02)	0.94 (0.01)	0.94 (0.00)
	long_parameters_list	0.90 (0.03)	0.80 (0.01)	0.91 (0.00)	0.98 (0.01)	0.84 (0.00)	0.96 (0.00)	0.85 (0.05)	0.74 (0.01)	0.87 (0.00)
	switch_statements	0.82 (0.04)	0.73 (0.00)	0.80 (0.00)	0.95 (0.01)	0.80 (0.00)	0.89 (0.00)	0.73 (0.07)	0.62 (0.01)	0.70 (0.00)

TABLE X

RQ2: PERFORMANCE RESULTS FOR ALL METRICS OF INTEREST IN WEKA

Task	Dataset	F-measure			AUC			Matthews		
		RF	SVM	DT	RF	SVM	DT	RF	SVM	DT
DP	ant-1.7	0.52 (0.06)	0.51 (0.00)	0.45 (0.07)	0.84 (0.02)	0.75 (0.00)	0.65 (0.04)	0.43 (0.07)	0.43 (0.00)	0.29 (0.08)
	camel-1.6	0.26 (0.09)	0.04 (0.00)	0.37 (0.07)	0.75 (0.02)	0.64 (0.00)	0.62 (0.04)	0.22 (0.10)	0.05 (0.00)	0.22 (0.08)
	ivy-2.0	0.22 (0.10)	0.03 (0.00)	0.24 (0.12)	0.79 (0.04)	0.65 (0.14)	0.58 (0.07)	0.21 (0.11)	0.02 (0.00)	0.16 (0.13)
	jedit-4.3	0.00 (0.00)	0.00 (0.00)	0.05 (0.07)	0.79 (0.09)	0.36 (0.28)	0.70 (0.06)	-0.01 (0.00)	0.00 (0.00)	0.04 (0.08)
	lucene-2.4	0.78 (0.04)	0.75 (0.00)	0.66 (0.04)	0.79 (0.04)	0.68 (0.26)	0.60 (0.05)	0.43 (0.10)	0.00 (0.00)	0.21 (0.09)
	tomcat	0.21 (0.09)	0.00 (0.00)	0.30 (0.09)	0.82 (0.03)	0.53 (0.10)	0.63 (0.06)	0.23 (0.10)	-0.01 (0.00)	0.24 (0.10)
	equinox	0.69 (0.05)	0.59 (0.00)	0.62 (0.07)	0.83 (0.02)	0.81 (0.00)	0.70 (0.06)	0.51 (0.08)	0.44 (0.00)	0.39 (0.11)
	lucene	0.37 (0.10)	0.28 (0.00)	0.41 (0.10)	0.82 (0.03)	0.70 (0.00)	0.67 (0.06)	0.37 (0.11)	0.30 (0.00)	0.38 (0.12)
	mylyn	0.24 (0.07)	0.08 (0.00)	0.30 (0.05)	0.80 (0.02)	0.70 (0.00)	0.64 (0.03)	0.22 (0.07)	0.09 (0.00)	0.20 (0.06)
	pde	0.17 (0.06)	0.06 (0.00)	0.30 (0.06)	0.72 (0.03)	0.57 (0.00)	0.60 (0.04)	0.16 (0.08)	0.08 (0.00)	0.19 (0.07)
	jdt	0.50 (0.05)	0.45 (0.00)	0.45 (0.07)	0.83 (0.02)	0.73 (0.00)	0.66 (0.05)	0.43 (0.06)	0.43 (0.00)	0.31 (0.09)
PMT	apns	0.87 (0.01)	0.56 (0.00)	0.84 (0.01)	0.93 (0.01)	0.89 (0.00)	0.87 (0.01)	0.73 (0.02)	0.44 (0.00)	0.69 (0.03)
	assj	0.97 (0.00)	0.91 (0.00)	0.95 (0.00)	0.94 (0.01)	0.51 (0.04)	0.89 (0.00)	0.77 (0.01)	0.00 (0.00)	0.71 (0.01)
	joda	0.91 (0.00)	0.85 (0.00)	0.90 (0.00)	0.92 (0.00)	0.66 (0.18)	0.86 (0.00)	0.66 (0.01)	0.00 (0.00)	0.63 (0.00)
	lafj	0.90 (0.00)	0.79 (0.00)	0.89 (0.00)	0.93 (0.00)	0.76 (0.11)	0.89 (0.00)	0.71 (0.01)	0.00 (0.00)	0.69 (0.00)
	lang	0.92 (0.00)	0.90 (0.00)	0.91 (0.00)	0.88 (0.00)	0.60 (0.11)	0.80 (0.00)	0.53 (0.01)	0.00 (0.00)	0.50 (0.01)
	msg	0.90 (0.00)	0.72 (0.00)	0.88 (0.00)	0.94 (0.00)	0.88 (0.08)	0.89 (0.00)	0.75 (0.01)	-0.00 (0.00)	0.72 (0.01)
	uaa	0.83 (0.01)	0.76 (0.00)	0.80 (0.00)	0.92 (0.00)	0.88 (0.00)	0.88 (0.00)	0.69 (0.01)	0.56 (0.00)	0.65 (0.01)
	vrapt	0.94 (0.00)	0.82 (0.00)	0.92 (0.00)	0.94 (0.00)	0.75 (0.00)	0.89 (0.00)	0.79 (0.01)	0.00 (0.00)	0.73 (0.01)
	wire	0.93 (0.00)	0.84 (0.00)	0.91 (0.00)	0.94 (0.00)	0.80 (0.00)	0.89 (0.00)	0.77 (0.01)	0.42 (0.00)	0.74 (0.01)
CSD	data-class	0.99 (0.02)	0.00 (0.00)	0.97 (0.02)	1.00 (0.00)	0.73 (0.26)	0.98 (0.02)	0.99 (0.03)	0.00 (0.00)	0.96 (0.04)
	feature-envy	0.94 (0.03)	0.16 (0.00)	0.90 (0.05)	0.99 (0.00)	0.74 (0.00)	0.93 (0.04)	0.91 (0.05)	0.18 (0.00)	0.86 (0.07)
	god-class	0.96 (0.02)	0.25 (0.00)	0.95 (0.03)	1.00 (0.01)	0.83 (0.00)	0.96 (0.02)	0.94 (0.03)	0.23 (0.00)	0.92 (0.04)
	long-method	0.99 (0.01)	0.15 (0.00)	0.99 (0.01)	1.00 (0.00)	0.77 (0.00)	0.99 (0.01)	0.99 (0.02)	0.18 (0.00)	0.98 (0.02)
	long_parameters_list	0.90 (0.03)	0.00 (0.00)	0.87 (0.04)	0.98 (0.01)	0.49 (0.08)	0.91 (0.03)	0.85 (0.04)	0.00 (0.00)	0.82 (0.06)
	switch_statements	0.82 (0.05)	0.00 (0.00)	0.76 (0.07)	0.95 (0.01)	0.55 (0.20)	0.84 (0.05)	0.73 (0.07)	0.00 (0.00)	0.66 (0.09)

TABLE XI

RQ2: PERFORMANCE RESULTS FOR ALL METRICS OF INTEREST IN SCIKIT-LEARN

Task	Dataset	F-measure			AUC			Matthews		
		RF	SVM	DT	RF	SVM	DT	RF	SVM	DT
DP	ant-1.7	0.52 (0.06)	0.39 (0.04)	0.50 (0.00)	0.83 (0.01)	0.75 (0.01)	0.72 (0.00)	0.42 (0.06)	0.33 (0.04)	0.39 (0.00)
	camel-1.6	0.23 (0.05)	0.10 (0.05)	0.26 (0.00)	0.74 (0.01)	0.64 (0.02)	0.61 (0.00)	0.21 (0.06)	0.12 (0.06)	0.21 (0.00)
	ivy-2.0	0.29 (0.04)	0.27 (0.10)	0.29 (0.00)	0.80 (0.01)	0.70 (0.04)	0.65 (0.00)	0.28 (0.04)	0.30 (0.10)	0.23 (0.00)
	jedit-4.3	NaN	NaN	NaN	0.74 (0.05)	0.49 (0.21)	0.23 (0.00)	-0.01 (0.00)	-0.01 (0.01)	0.00 (0.00)
	lucene-2.4	0.78 (0.02)	0.75 (0.02)	0.75 (0.00)	0.78 (0.01)	0.76 (0.01)	0.69 (0.00)	0.41 (0.07)	0.33 (0.04)	0.34 (0.00)
	tomcat	0.25 (0.06)	NaN	0.31 (0.00)	0.84 (0.01)	0.67 (0.04)	0.68 (0.00)	0.29 (0.09)	-0.02 (0.01)	0.27 (0.00)
	equinox	0.70 (0.03)	0.67 (0.01)	0.66 (0.00)	0.82 (0.01)	0.78 (0.02)	0.76 (0.00)	0.51 (0.04)	0.47 (0.02)	0.45 (0.00)
	lucene	0.36 (0.06)	0.30 (0.16)	0.28 (0.00)	0.81 (0.01)	0.66 (0.03)	0.71 (0.00)	0.37 (0.07)	0.34 (0.13)	0.26 (0.00)
	mylyn	0.15 (0.03)	0.10 (0.04)	0.19 (0.00)	0.80 (0.01)	0.67 (0.03)	0.62 (0.00)	0.15 (0.04)	0.13 (0.06)	0.19 (0.00)
	pde	0.16 (0.03)	0.05 (0.05)	0.25 (0.00)	0.72 (0.01)	0.59 (0.03)	0.62 (0.00)	0.17 (0.04)	0.07 (0.07)	0.22 (0.00)
	jdt	0.52 (0.03)	0.47 (0.01)	0.42 (0.00)	0.83 (0.01)	0.79 (0.01)	0.71 (0.00)	0.46 (0.04)	0.42 (0.01)	0.34 (0.00)
PMT	apns	0.89 (0.01)	0.79 (0.00)	0.87 (0.00)	0.94 (0.00)	0.87 (0.00)	0.90 (0.00)	0.76 (0.02)	0.57 (0.01)	0.72 (0.00)
	assj	0.97 (0.00)	0.93 (0.00)	0.96 (0.00)	0.96 (0.00)	0.78 (0.01)	0.80 (0.00)	0.78 (0.01)	0.36 (0.01)	0.70 (0.00)
	joda	0.93 (0.00)	0.87 (0.00)	0.91 (0.00)	0.93 (0.00)	0.79 (0.00)	0.78 (0.00)	0.70 (0.00)	0.43 (0.00)	0.60 (0.00)
	lafj	0.91 (0.00)	0.83 (0.00)	0.90 (0.00)	0.93 (0.00)	0.81 (0.00)	0.79 (0.00)	0.73 (0.01)	0.47 (0.00)	0.70 (0.00)
	lang	0.94 (0.00)	0.91 (0.00)	0.93 (0.00)	0.90 (0.00)	0.72 (0.00)	0.62 (0.00)	0.58 (0.00)	0.25 (0.00)	0.46 (0.00)
	msg	0.91 (0.00)	0.79 (0.00)	0.91 (0.00)	0.95 (0.00)	0.83 (0.00)	0.85 (0.00)	0.78 (0.00)	0.51 (0.00)	0.78 (0.00)
	uaa	0.86 (0.00)	0.77 (0.00)	0.84 (0.00)	0.93 (0.00)	0.87 (0.00)	0.85 (0.00)	0.73 (0.01)	0.59 (0.00)	0.71 (0.00)
	vrapt	0.94 (0.00)	0.85 (0.00)	0.93 (0.00)	0.95 (0.00)	0.79 (0.00)	0.84 (0.00)	0.81 (0.00)	0.42 (0.00)	0.77 (0.00)
	wire	0.94 (0.00)	0.93 (0.00)	0.93 (0.00)	0.95 (0.00)	0.89 (0.00)	0.85 (0.00)	0.82 (0.01)	0.78 (0.00)	0.79 (0.00)
CSD	data-class	0.99 (0.01)	0.27 (0.05)	0.98 (0.00)	1.00 (0.00)	0.58 (0.04)	0.98 (0.00)	0.99 (0.02)	0.33 (0.04)	0.96 (0.00)
	feature-envy	0.94 (0.01)	0.71 (0.08)	0.93 (0.00)	0.99 (0.00)	0.79 (0.06)	0.95 (0.00)	0.91 (0.02)	0.57 (0.16)	0.89 (0.00)
	god-class	0.96 (0.01)	NaN	0.97 (0.00)	1.00 (0.00)	0.53 (0.07)	0.97 (0.00)	0.94 (0.01)	0.00 (0.00)	0.95 (0.00)
	long-method	0.99 (0.01)	0.74 (0.01)	0.98 (0.00)	1.00 (0.00)	0.83 (0.07)	1.00 (0.00)	0.99 (0.01)	0.61 (0.01)	0.97 (0.00)
	long_parameters_list	0.90 (0.01)	0.42 (0.03)	0.88 (0.00)	0.98 (0.00)	0.64 (0.03)	0.93 (0.00)	0.86 (0.02)	0.37 (0.02)	0.82 (0.00)
	switch_statements	0.82 (0.03)	0.10 (0.03)	0.80 (0.00)	0.95 (0.00)	0.52 (0.03)	0.92 (0.00)	0.73 (0.05)	0.15 (0.04)	0.70 (0.00)

TABLE XII

RQ2: PERFORMANCE RESULTS FOR ALL METRICS OF INTEREST IN R