

Table 1: Results for classification with 71.5 Maven Reuse split using RF feature reduction

Algorithm	No. of features	Accuracy	Precision	Recall
SVC(kernel='rbf')	199	0.587	0.68	0.486
SVC(kernel='rbf')	160	0.571	0.633	0.543
SVC(kernel='rbf')	130	0.603	0.667	0.571
SVC(kernel='rbf')	100	0.603	0.679	0.543
SVC(kernel='rbf')	70	0.603	0.667	0.571
SVC(kernel='rbf')	40	0.619	0.677	0.6
SVC(kernel='rbf')	20	0.603	0.656	0.6
SVC(kernel='rbf')	10	0.524	0.586	0.486
SVC(kernel='linear')	199	0.54	0.6	0.514
SVC(kernel='linear')	160	0.587	0.636	0.6
SVC(kernel='linear')	130	0.603	0.679	0.543
SVC(kernel='linear')	100	0.619	0.69	0.571
SVC(kernel='linear')	70	0.635	0.7	0.6
SVC(kernel='linear')	40	0.619	0.69	0.571
SVC(kernel='linear')	20	0.619	0.677	0.6
SVC(kernel='linear')	10	0.635	0.714	0.571
SVC(kernel='poly', degree=3)	199	0.476	0.625	0.143
SVC(kernel='poly', degree=3)	160	0.46	0.571	0.114
SVC(kernel='poly', degree=3)	130	0.476	0.625	0.143
SVC(kernel='poly', degree=3)	100	0.476	0.625	0.143
SVC(kernel='poly', degree=3)	70	0.492	0.667	0.171
SVC(kernel='poly', degree=3)	40	0.508	0.667	0.229
SVC(kernel='poly', degree=3)	20	0.476	0.583	0.2
SVC(kernel='poly', degree=3)	10	0.476	0.583	0.2
SVC(kernel='poly', degree=4)	199	0.492	0.714	0.143
SVC(kernel='poly', degree=4)	160	0.46	0.556	0.143
SVC(kernel='poly', degree=4)	130	0.476	0.625	0.143

SVC(kernel='poly', degree=4)	100	0.492	0.714	0.143
SVC(kernel='poly', degree=4)	70	0.476	0.6	0.171
SVC(kernel='poly', degree=4)	40	0.54	0.714	0.286
SVC(kernel='poly', degree=4)	20	0.492	0.6	0.257
SVC(kernel='poly', degree=4)	10	0.508	0.643	0.257
SVC(kernel='poly', degree=5)	199	0.476	0.667	0.114
SVC(kernel='poly', degree=5)	160	0.476	0.667	0.114
SVC(kernel='poly', degree=5)	130	0.476	0.667	0.114
SVC(kernel='poly', degree=5)	100	0.476	0.667	0.114
SVC(kernel='poly', degree=5)	70	0.508	0.75	0.171
SVC(kernel='poly', degree=5)	40	0.508	0.7	0.2
SVC(kernel='poly', degree=5)	20	0.492	0.636	0.2
SVC(kernel='poly', degree=5)	10	0.492	0.615	0.229
SVC(kernel='sigmoid')	199	0.603	0.727	0.457
SVC(kernel='sigmoid')	160	0.603	0.708	0.486
SVC(kernel='sigmoid')	130	0.571	0.722	0.371
SVC(kernel='sigmoid')	100	0.587	0.696	0.457
SVC(kernel='sigmoid')	70	0.54	0.607	0.486
SVC(kernel='sigmoid')	40	0.571	0.667	0.457
SVC(kernel='sigmoid')	20	0.635	0.731	0.543
SVC(kernel='sigmoid')	10	0.603	0.667	0.571
RandomForestClassifier(n_estimators=10)	199	0.571	0.654	0.486
RandomForestClassifier(n_estimators=10)	160	0.492	0.565	0.371
RandomForestClassifier(n_estimators=10)	130	0.524	0.593	0.457
RandomForestClassifier(n_estimators=10)	100	0.603	0.692	0.514
RandomForestClassifier(n_estimators=10)	70	0.54	0.615	0.457
RandomForestClassifier(n_estimators=10)	40	0.635	0.7	0.6
RandomForestClassifier(n_estimators=10)	20	0.603	0.679	0.543
RandomForestClassifier(n_estimators=10)	10	0.429	0.483	0.4
RandomForestClassifier(n_estimators=20)	199	0.571	0.618	0.6

RandomForestClassifier(n_estimators=20)	160	0.508	0.577	0.429
RandomForestClassifier(n_estimators=20)	130	0.571	0.618	0.6
RandomForestClassifier(n_estimators=20)	100	0.635	0.688	0.629
RandomForestClassifier(n_estimators=20)	70	0.556	0.63	0.486
RandomForestClassifier(n_estimators=20)	40	0.603	0.647	0.629
RandomForestClassifier(n_estimators=20)	20	0.651	0.71	0.629
RandomForestClassifier(n_estimators=20)	10	0.524	0.581	0.514
RandomForestClassifier(n_estimators=50)	199	0.556	0.606	0.571
RandomForestClassifier(n_estimators=50)	160	0.571	0.667	0.457
RandomForestClassifier(n_estimators=50)	130	0.571	0.643	0.514
RandomForestClassifier(n_estimators=50)	100	0.619	0.677	0.6
RandomForestClassifier(n_estimators=50)	70	0.603	0.656	0.6
RandomForestClassifier(n_estimators=50)	40	0.587	0.636	0.6
RandomForestClassifier(n_estimators=50)	20	0.603	0.647	0.629
RandomForestClassifier(n_estimators=50)	10	0.571	0.633	0.543
RandomForestClassifier(n_estimators=100)	199	0.603	0.647	0.629
RandomForestClassifier(n_estimators=100)	160	0.571	0.643	0.514
RandomForestClassifier(n_estimators=100)	130	0.556	0.621	0.514
RandomForestClassifier(n_estimators=100)	100	0.603	0.679	0.543
RandomForestClassifier(n_estimators=100)	70	0.651	0.71	0.629
RandomForestClassifier(n_estimators=100)	40	0.635	0.676	0.657
RandomForestClassifier(n_estimators=100)	20	0.603	0.647	0.629
RandomForestClassifier(n_estimators=100)	10	0.587	0.655	0.543
RandomForestClassifier(n_estimators=300)	199	0.683	0.759	0.629
RandomForestClassifier(n_estimators=300)	160	0.619	0.677	0.6
RandomForestClassifier(n_estimators=300)	130	0.603	0.667	0.571
RandomForestClassifier(n_estimators=300)	100	0.571	0.633	0.543
RandomForestClassifier(n_estimators=300)	70	0.635	0.688	0.629
RandomForestClassifier(n_estimators=300)	40	0.635	0.676	0.657
RandomForestClassifier(n_estimators=300)	20	0.619	0.667	0.629

RandomForestClassifier(n_estimators=300)	10	0.524	0.586	0.486
RandomForestClassifier(n_estimators=500)	199	0.667	0.75	0.6
RandomForestClassifier(n_estimators=500)	160	0.603	0.667	0.571
RandomForestClassifier(n_estimators=500)	130	0.603	0.656	0.6
RandomForestClassifier(n_estimators=500)	100	0.635	0.688	0.629
RandomForestClassifier(n_estimators=500)	70	0.651	0.71	0.629
RandomForestClassifier(n_estimators=500)	40	0.651	0.697	0.657
RandomForestClassifier(n_estimators=500)	20	0.619	0.667	0.629
RandomForestClassifier(n_estimators=500)	10	0.54	0.6	0.514
RandomForestClassifier(n_estimators=700)	199	0.635	0.7	0.6
RandomForestClassifier(n_estimators=700)	160	0.635	0.7	0.6
RandomForestClassifier(n_estimators=700)	130	0.603	0.656	0.6
RandomForestClassifier(n_estimators=700)	100	0.635	0.688	0.629
RandomForestClassifier(n_estimators=700)	70	0.651	0.724	0.6
RandomForestClassifier(n_estimators=700)	40	0.651	0.686	0.686
RandomForestClassifier(n_estimators=700)	20	0.635	0.688	0.629
RandomForestClassifier(n_estimators=700)	10	0.556	0.613	0.543
RandomForestClassifier(n_estimators=1000)	199	0.667	0.733	0.629
RandomForestClassifier(n_estimators=1000)	160	0.619	0.677	0.6
RandomForestClassifier(n_estimators=1000)	130	0.603	0.656	0.6
RandomForestClassifier(n_estimators=1000)	100	0.619	0.677	0.6
RandomForestClassifier(n_estimators=1000)	70	0.635	0.714	0.571
RandomForestClassifier(n_estimators=1000)	40	0.667	0.706	0.686
RandomForestClassifier(n_estimators=1000)	20	0.635	0.688	0.629
RandomForestClassifier(n_estimators=1000)	10	0.524	0.581	0.514
KNeighborsClassifier(n_neighbors=2)	199	0.571	0.722	0.371
KNeighborsClassifier(n_neighbors=2)	160	0.54	0.75	0.257
KNeighborsClassifier(n_neighbors=2)	130	0.508	0.643	0.257
KNeighborsClassifier(n_neighbors=2)	100	0.476	0.562	0.257
KNeighborsClassifier(n_neighbors=2)	70	0.571	0.682	0.429

KNeighborsClassifier(n_neighbors=2)	40	0.571	0.682	0.429
KNeighborsClassifier(n_neighbors=2)	20	0.556	0.684	0.371
KNeighborsClassifier(n_neighbors=2)	10	0.571	0.7	0.4
KNeighborsClassifier(n_neighbors=3)	199	0.587	0.655	0.543
KNeighborsClassifier(n_neighbors=3)	160	0.571	0.654	0.486
KNeighborsClassifier(n_neighbors=3)	130	0.556	0.652	0.429
KNeighborsClassifier(n_neighbors=3)	100	0.54	0.607	0.486
KNeighborsClassifier(n_neighbors=3)	70	0.571	0.633	0.543
KNeighborsClassifier(n_neighbors=3)	40	0.587	0.667	0.514
KNeighborsClassifier(n_neighbors=3)	20	0.524	0.586	0.486
KNeighborsClassifier(n_neighbors=3)	10	0.492	0.548	0.486
KNeighborsClassifier(n_neighbors=5)	199	0.54	0.636	0.4
KNeighborsClassifier(n_neighbors=5)	160	0.476	0.556	0.286
KNeighborsClassifier(n_neighbors=5)	130	0.508	0.583	0.4
KNeighborsClassifier(n_neighbors=5)	100	0.476	0.55	0.314
KNeighborsClassifier(n_neighbors=5)	70	0.508	0.571	0.457
KNeighborsClassifier(n_neighbors=5)	40	0.571	0.654	0.486
KNeighborsClassifier(n_neighbors=5)	20	0.476	0.538	0.4
KNeighborsClassifier(n_neighbors=5)	10	0.492	0.56	0.4
KNeighborsClassifier(n_neighbors=7)	199	0.46	0.529	0.257
KNeighborsClassifier(n_neighbors=7)	160	0.46	0.526	0.286
KNeighborsClassifier(n_neighbors=7)	130	0.429	0.476	0.286
KNeighborsClassifier(n_neighbors=7)	100	0.46	0.526	0.286
KNeighborsClassifier(n_neighbors=7)	70	0.524	0.619	0.371
KNeighborsClassifier(n_neighbors=7)	40	0.54	0.625	0.429
KNeighborsClassifier(n_neighbors=7)	20	0.492	0.56	0.4
KNeighborsClassifier(n_neighbors=7)	10	0.524	0.6	0.429
KNeighborsClassifier(n_neighbors=10)	199	0.476	0.583	0.2
KNeighborsClassifier(n_neighbors=10)	160	0.46	0.545	0.171
KNeighborsClassifier(n_neighbors=10)	130	0.444	0.5	0.171

KNeighborsClassifier(n_neighbors=10)	100	0.444	0.5	0.2
KNeighborsClassifier(n_neighbors=10)	70	0.444	0.5	0.171
KNeighborsClassifier(n_neighbors=10)	40	0.54	0.636	0.4
KNeighborsClassifier(n_neighbors=10)	20	0.476	0.556	0.286
KNeighborsClassifier(n_neighbors=10)	10	0.492	0.571	0.343
KNeighborsClassifier(n_neighbors=15)	199	0.429	0.455	0.143
KNeighborsClassifier(n_neighbors=15)	160	0.397	0.364	0.114
KNeighborsClassifier(n_neighbors=15)	130	0.429	0.462	0.171
KNeighborsClassifier(n_neighbors=15)	100	0.429	0.467	0.2
KNeighborsClassifier(n_neighbors=15)	70	0.46	0.533	0.229
KNeighborsClassifier(n_neighbors=15)	40	0.492	0.571	0.343
KNeighborsClassifier(n_neighbors=15)	20	0.46	0.52	0.371
KNeighborsClassifier(n_neighbors=15)	10	0.556	0.63	0.486
MLPClassifier	199	0.54	0.594	0.543
MLPClassifier	160	0.587	0.655	0.543
MLPClassifier	130	0.54	0.594	0.543
MLPClassifier	100	0.603	0.639	0.657
MLPClassifier	70	0.492	0.543	0.543
MLPClassifier	40	0.556	0.6	0.6
MLPClassifier	20	0.587	0.629	0.629
MLPClassifier	10	0.635	0.676	0.657
DecisionTreeClassifier	199	0.508	0.556	0.571
DecisionTreeClassifier	160	0.603	0.632	0.686
DecisionTreeClassifier	130	0.587	0.61	0.714
DecisionTreeClassifier	100	0.571	0.591	0.743
DecisionTreeClassifier	70	0.667	0.684	0.743
DecisionTreeClassifier	40	0.508	0.559	0.543
DecisionTreeClassifier	20	0.571	0.667	0.457
DecisionTreeClassifier	10	0.587	0.636	0.6
GaussianNB	199	0.492	0.588	0.286

GaussianNB	160	0.492	0.579	0.314
GaussianNB	130	0.476	0.55	0.314
GaussianNB	100	0.444	0.5	0.286
GaussianNB	70	0.444	0.5	0.286
GaussianNB	40	0.444	0.5	0.286
GaussianNB	20	0.444	0.5	0.257
GaussianNB	10	0.476	0.562	0.257

Table 2: Results for classification with 100 Maven Reuse split using RF feature reduction

Algorithm	No. of features	Accuracy	Precision	Recall
SVC(kernel='rbf')	199	0.603	0.619	0.433
SVC(kernel='rbf')	160	0.651	0.682	0.5
SVC(kernel='rbf')	130	0.667	0.68	0.567
SVC(kernel='rbf')	100	0.651	0.682	0.5
SVC(kernel='rbf')	70	0.619	0.65	0.433
SVC(kernel='rbf')	40	0.619	0.636	0.467
SVC(kernel='rbf')	20	0.619	0.636	0.467
SVC(kernel='rbf')	10	0.603	0.619	0.433
SVC(kernel='linear')	199	0.619	0.6	0.6
SVC(kernel='linear')	160	0.651	0.633	0.633
SVC(kernel='linear')	130	0.667	0.655	0.633
SVC(kernel='linear')	100	0.746	0.733	0.733
SVC(kernel='linear')	70	0.683	0.692	0.6
SVC(kernel='linear')	40	0.667	0.68	0.567
SVC(kernel='linear')	20	0.635	0.684	0.433
SVC(kernel='linear')	10	0.603	0.6	0.5
SVC(kernel='poly', degree=3)	199	0.556	0.625	0.167
SVC(kernel='poly', degree=3)	160	0.556	0.625	0.167

SVC(kernel='poly', degree=3)	130	0.556	0.625	0.167
SVC(kernel='poly', degree=3)	100	0.556	0.625	0.167
SVC(kernel='poly', degree=3)	70	0.571	0.667	0.2
SVC(kernel='poly', degree=3)	40	0.635	0.769	0.333
SVC(kernel='poly', degree=3)	20	0.619	0.75	0.3
SVC(kernel='poly', degree=3)	10	0.587	0.643	0.3
SVC(kernel='poly', degree=4)	199	0.571	0.714	0.167
SVC(kernel='poly', degree=4)	160	0.556	0.625	0.167
SVC(kernel='poly', degree=4)	130	0.556	0.625	0.167
SVC(kernel='poly', degree=4)	100	0.556	0.625	0.167
SVC(kernel='poly', degree=4)	70	0.571	0.667	0.2
SVC(kernel='poly', degree=4)	40	0.603	0.727	0.267
SVC(kernel='poly', degree=4)	20	0.635	0.818	0.3
SVC(kernel='poly', degree=4)	10	0.587	0.643	0.3
SVC(kernel='poly', degree=5)	199	0.556	0.667	0.133
SVC(kernel='poly', degree=5)	160	0.556	0.625	0.167
SVC(kernel='poly', degree=5)	130	0.556	0.625	0.167
SVC(kernel='poly', degree=5)	100	0.556	0.625	0.167
SVC(kernel='poly', degree=5)	70	0.571	0.667	0.2
SVC(kernel='poly', degree=5)	40	0.603	0.727	0.267
SVC(kernel='poly', degree=5)	20	0.619	0.8	0.267
SVC(kernel='poly', degree=5)	10	0.571	0.615	0.267
SVC(kernel='sigmoid')	199	0.619	0.667	0.4
SVC(kernel='sigmoid'))	160	0.619	0.688	0.367
SVC(kernel='sigmoid')	130	0.635	0.706	0.4
SVC(kernel='sigmoid')	100	0.587	0.611	0.367
SVC(kernel='sigmoid')	70	0.619	0.65	0.433
SVC(kernel='sigmoid')	40	0.587	0.591	0.433
SVC(kernel='sigmoid')	20	0.571	0.579	0.367
SVC(kernel='sigmoid')	10	0.635	0.667	0.467



RandomForestClassifier(n_estimators=10)	199	0.667	0.737	0.467
RandomForestClassifier(n_estimators=10)	160	0.571	0.552	0.533
RandomForestClassifier(n_estimators=10)	130	0.556	0.542	0.433
RandomForestClassifier(n_estimators=10)	100	0.571	0.56	0.467
RandomForestClassifier(n_estimators=10)	70	0.603	0.6	0.5
RandomForestClassifier(n_estimators=10)	40	0.651	0.7	0.467
RandomForestClassifier(n_estimators=10)	20	0.635	0.606	0.667
RandomForestClassifier(n_estimators=10)	10	0.508	0.483	0.467
RandomForestClassifier(n_estimators=20)	199	0.667	0.68	0.567
RandomForestClassifier(n_estimators=20)	160	0.635	0.63	0.567
RandomForestClassifier(n_estimators=20)	130	0.667	0.68	0.567
RandomForestClassifier(n_estimators=20)	100	0.619	0.615	0.533
RandomForestClassifier(n_estimators=20)	70	0.635	0.621	0.6
RandomForestClassifier(n_estimators=20)	40	0.635	0.667	0.467
RandomForestClassifier(n_estimators=20)	20	0.698	0.677	0.7
RandomForestClassifier(n_estimators=20)	10	0.556	0.533	0.533
RandomForestClassifier(n_estimators=50)	199	0.714	0.75	0.6
RandomForestClassifier(n_estimators=50)	160	0.619	0.625	0.5
RandomForestClassifier(n_estimators=50)	130	0.698	0.72	0.6
RandomForestClassifier(n_estimators=50)	100	0.73	0.724	0.7
RandomForestClassifier(n_estimators=50)	70	0.683	0.679	0.633
RandomForestClassifier(n_estimators=50)	40	0.73	0.783	0.6
RandomForestClassifier(n_estimators=50)	20	0.667	0.645	0.667
RandomForestClassifier(n_estimators=50)	10	0.651	0.625	0.667
RandomForestClassifier(n_estimators=100)	199	0.683	0.708	0.567
RandomForestClassifier(n_estimators=100)	160	0.651	0.654	0.567
RandomForestClassifier(n_estimators=100)	130	0.698	0.72	0.6
RandomForestClassifier(n_estimators=100)	100	0.698	0.704	0.633
RandomForestClassifier(n_estimators=100)	70	0.746	0.792	0.633
RandomForestClassifier(n_estimators=100)	40	0.714	0.75	0.6

RandomForestClassifier(n_estimators=100)	20	0.651	0.625	0.667
RandomForestClassifier(n_estimators=100)	10	0.587	0.553	0.7
RandomForestClassifier(n_estimators=300)	199	0.683	0.708	0.567
RandomForestClassifier(n_estimators=300)	160	0.698	0.72	0.6
RandomForestClassifier(n_estimators=300)	130	0.746	0.75	0.7
RandomForestClassifier(n_estimators=300)	100	0.683	0.692	0.6
RandomForestClassifier(n_estimators=300)	70	0.73	0.76	0.633
RandomForestClassifier(n_estimators=300)	40	0.73	0.741	0.667
RandomForestClassifier(n_estimators=300)	20	0.683	0.667	0.667
RandomForestClassifier(n_estimators=300)	10	0.587	0.556	0.667
RandomForestClassifier(n_estimators=500)	199	0.667	0.696	0.533
RandomForestClassifier(n_estimators=500)	160	0.714	0.75	0.6
RandomForestClassifier(n_estimators=500)	130	0.73	0.741	0.667
RandomForestClassifier(n_estimators=500)	100	0.698	0.72	0.6
RandomForestClassifier(n_estimators=500)	70	0.746	0.769	0.667
RandomForestClassifier(n_estimators=500)	40	0.746	0.792	0.633
RandomForestClassifier(n_estimators=500)	20	0.698	0.69	0.667
RandomForestClassifier(n_estimators=500)	10	0.587	0.553	0.7
RandomForestClassifier(n_estimators=700)	199	0.698	0.72	0.6
RandomForestClassifier(n_estimators=700)	160	0.714	0.75	0.6
RandomForestClassifier(n_estimators=700)	130	0.73	0.741	0.667
RandomForestClassifier(n_estimators=700)	100	0.698	0.72	0.6
RandomForestClassifier(n_estimators=700)	70	0.746	0.769	0.667
RandomForestClassifier(n_estimators=700)	40	0.73	0.76	0.633
RandomForestClassifier(n_estimators=700)	20	0.698	0.69	0.667
RandomForestClassifier(n_estimators=700)	10	0.587	0.553	0.7
RandomForestClassifier(n_estimators=1000)	199	0.714	0.731	0.633
RandomForestClassifier(n_estimators=1000)	160	0.714	0.75	0.6
RandomForestClassifier(n_estimators=1000)	130	0.714	0.731	0.633
RandomForestClassifier(n_estimators=1000)	100	0.714	0.75	0.6

RandomForestClassifier(n_estimators=1000)	70	0.746	0.769	0.667
RandomForestClassifier(n_estimators=1000)	40	0.746	0.769	0.667
RandomForestClassifier(n_estimators=1000)	20	0.667	0.68	0.567
RandomForestClassifier(n_estimators=1000)	10	0.603	0.568	0.7
KNeighborsClassifier(n_neighbors=2)	199	0.667	0.765	0.433
KNeighborsClassifier(n_neighbors=2)	160	0.603	0.667	0.333
KNeighborsClassifier(n_neighbors=2)	130	0.635	0.769	0.333
KNeighborsClassifier(n_neighbors=2)	100	0.635	0.733	0.367
KNeighborsClassifier(n_neighbors=2)	70	0.587	0.611	0.367
KNeighborsClassifier(n_neighbors=2)	40	0.651	0.786	0.367
KNeighborsClassifier(n_neighbors=2)	20	0.556	0.55	0.367
KNeighborsClassifier(n_neighbors=2)	10	0.603	0.619	0.433
KNeighborsClassifier(n_neighbors=3)	199	0.587	0.571	0.533
KNeighborsClassifier(n_neighbors=3)	160	0.619	0.636	0.467
KNeighborsClassifier(n_neighbors=3)	130	0.587	0.571	0.533
KNeighborsClassifier(n_neighbors=3)	100	0.587	0.577	0.5
KNeighborsClassifier(n_neighbors=3)	70	0.524	0.5	0.467
KNeighborsClassifier(n_neighbors=3)	40	0.556	0.536	0.5
KNeighborsClassifier(n_neighbors=3)	20	0.524	0.5	0.533
KNeighborsClassifier(n_neighbors=3)	10	0.651	0.625	0.667
KNeighborsClassifier(n_neighbors=5)	199	0.54	0.524	0.367
KNeighborsClassifier(n_neighbors=5)	160	0.524	0.5	0.3
KNeighborsClassifier(n_neighbors=5)	130	0.571	0.571	0.4
KNeighborsClassifier(n_neighbors=5)	100	0.587	0.577	0.5
KNeighborsClassifier(n_neighbors=5)	70	0.556	0.545	0.4
KNeighborsClassifier(n_neighbors=5)	40	0.508	0.476	0.333
KNeighborsClassifier(n_neighbors=5)	20	0.508	0.483	0.467
KNeighborsClassifier(n_neighbors=5)	10	0.54	0.517	0.5
KNeighborsClassifier(n_neighbors=7)	199	0.508	0.471	0.267
KNeighborsClassifier(n_neighbors=7)	160	0.556	0.562	0.3

KNeighborsClassifier(n_neighbors=7)	130	0.492	0.455	0.333
KNeighborsClassifier(n_neighbors=7)	100	0.508	0.476	0.333
KNeighborsClassifier(n_neighbors=7)	70	0.524	0.5	0.333
KNeighborsClassifier(n_neighbors=7)	40	0.476	0.429	0.3
KNeighborsClassifier(n_neighbors=7)	20	0.492	0.462	0.4
KNeighborsClassifier(n_neighbors=7)	10	0.524	0.5	0.433
KNeighborsClassifier(n_neighbors=10)	199	0.556	0.6	0.2
KNeighborsClassifier(n_neighbors=10)	160	0.587	0.7	0.233
KNeighborsClassifier(n_neighbors=10)	130	0.54	0.545	0.2
KNeighborsClassifier(n_neighbors=10)	100	0.556	0.625	0.167
KNeighborsClassifier(n_neighbors=10)	70	0.556	0.6	0.2
KNeighborsClassifier(n_neighbors=10)	40	0.508	0.429	0.1
KNeighborsClassifier(n_neighbors=10)	20	0.524	0.5	0.2
KNeighborsClassifier(n_neighbors=10)	10	0.524	0.5	0.267
KNeighborsClassifier(n_neighbors=15)	199	0.524	0.5	0.167
KNeighborsClassifier(n_neighbors=15)	160	0.524	0.5	0.2
KNeighborsClassifier(n_neighbors=15)	130	0.54	0.545	0.2
KNeighborsClassifier(n_neighbors=15)	100	0.524	0.5	0.233
KNeighborsClassifier(n_neighbors=15)	70	0.524	0.5	0.2
KNeighborsClassifier(n_neighbors=15)	40	0.556	0.583	0.233
KNeighborsClassifier(n_neighbors=15)	20	0.587	0.667	0.267
KNeighborsClassifier(n_neighbors=15)	10	0.571	0.6	0.3
MLPClassifier	199	0.587	0.562	0.6
MLPClassifier	160	0.619	0.607	0.567
MLPClassifier	130	0.587	0.571	0.533
MLPClassifier	100	0.603	0.581	0.6
MLPClassifier	70	0.683	0.667	0.667
MLPClassifier	40	0.651	0.625	0.667
MLPClassifier	20	0.73	0.71	0.733
MLPClassifier	10	0.587	0.583	0.467

DecisionTreeClassifier	199	0.635	0.621	0.6
DecisionTreeClassifier	160	0.619	0.625	0.5
DecisionTreeClassifier	130	0.619	0.625	0.5
DecisionTreeClassifier	100	0.635	0.64	0.533
DecisionTreeClassifier	70	0.619	0.6	0.6
DecisionTreeClassifier	40	0.635	0.595	0.733
DecisionTreeClassifier	20	0.635	0.595	0.733
DecisionTreeClassifier	10	0.54	0.513	0.667
GaussianNB	199	0.508	0.476	0.333
GaussianNB	160	0.524	0.5	0.267
GaussianNB	130	0.54	0.526	0.333
GaussianNB	100	0.54	0.526	0.333
GaussianNB	70	0.556	0.556	0.333
GaussianNB	40	0.571	0.588	0.333
GaussianNB	20	0.54	0.526	0.333
GaussianNB	10	0.619	0.667	0.4

Table 3: Results for classification with 200 Maven Reuse split using RF feature reduction

Algorithm	No. of features	Accuracy	Precision	Recall
SVC(kernel='rbf')	199	0.714	0.75	0.273
SVC(kernel='rbf')	160	0.714	0.75	0.273
SVC(kernel='rbf')	130	0.714	0.75	0.273
SVC(kernel='rbf')	100	0.714	0.75	0.273
SVC(kernel='rbf')	70	0.714	0.75	0.273
SVC(kernel='rbf')	40	0.714	0.7	0.318
SVC(kernel='rbf')	20	0.635	0.462	0.273
SVC(kernel='rbf')	10	0.635	0.462	0.273
SVC(kernel='linear')	199	0.603	0.435	0.455

SVC(kernel='linear')	160	0.619	0.462	0.545
SVC(kernel='linear')	130	0.619	0.455	0.455
SVC(kernel='linear')	100	0.619	0.467	0.636
SVC(kernel='linear')	70	0.667	0.522	0.545
SVC(kernel='linear')	40	0.683	0.562	0.409
SVC(kernel='linear')	20	0.651	0.5	0.182
SVC(kernel='linear')	10	0.667	0.6	0.136
SVC(kernel='poly', degree=3)	199	0.698	0.714	0.227
SVC(kernel='poly', degree=3)	160	0.698	0.714	0.227
SVC(kernel='poly', degree=3)	130	0.683	0.625	0.227
SVC(kernel='poly', degree=3)	100	0.714	0.75	0.273
SVC(kernel='poly', degree=3)	70	0.698	0.667	0.273
SVC(kernel='poly', degree=3)	40	0.683	0.583	0.318
SVC(kernel='poly', degree=3)	20	0.714	0.833	0.227
SVC(kernel='poly', degree=3)	10	0.683	0.667	0.182
SVC(kernel='poly', degree=4)	199	0.698	0.714	0.227
SVC(kernel='poly', degree=4)	160	0.698	0.714	0.227
SVC(kernel='poly', degree=4)	130	0.683	0.625	0.227
SVC(kernel='poly', degree=4)	100	0.683	0.625	0.227
SVC(kernel='poly', degree=4)	70	0.714	0.75	0.273
SVC(kernel='poly', degree=4)	40	0.667	0.545	0.273
SVC(kernel='poly', degree=4)	20	0.714	0.833	0.227
SVC(kernel='poly', degree=4)	10	0.698	0.8	0.182
SVC(kernel='poly', degree=5)	199	0.698	0.714	0.227
SVC(kernel='poly', degree=5)	160	0.698	0.714	0.227
SVC(kernel='poly', degree=5)	130	0.683	0.625	0.227
SVC(kernel='poly', degree=5)	100	0.683	0.625	0.227
SVC(kernel='poly', degree=5)	70	0.714	0.75	0.273
SVC(kernel='poly', degree=5)	40	0.651	0.5	0.227
SVC(kernel='poly', degree=5)	20	0.714	0.833	0.227

SVC(kernel='poly', degree=5)	10	0.714	0.833	0.227
SVC(kernel='sigmoid')	199	0.667	0.538	0.318
SVC(kernel='sigmoid'))	160	0.667	0.533	0.364
SVC(kernel='sigmoid')	130	0.667	0.533	0.364
SVC(kernel='sigmoid')	100	0.698	0.615	0.364
SVC(kernel='sigmoid')	70	0.683	0.562	0.409
SVC(kernel='sigmoid')	40	0.746	0.875	0.318
SVC(kernel='sigmoid')	20	0.73	0.857	0.273
SVC(kernel='sigmoid')	10	0.714	0.75	0.273
RandomForestClassifier(n_estimators=10)	199	0.698	0.636	0.318
RandomForestClassifier(n_estimators=10)	160	0.667	0.529	0.409
RandomForestClassifier(n_estimators=10)	130	0.683	0.571	0.364
RandomForestClassifier(n_estimators=10)	100	0.683	0.583	0.318
RandomForestClassifier(n_estimators=10)	70	0.651	0.5	0.364
RandomForestClassifier(n_estimators=10)	40	0.619	0.429	0.273
RandomForestClassifier(n_estimators=10)	20	0.635	0.467	0.318
RandomForestClassifier(n_estimators=10)	10	0.603	0.421	0.364
RandomForestClassifier(n_estimators=20)	199	0.603	0.4	0.273
RandomForestClassifier(n_estimators=20)	160	0.651	0.5	0.409
RandomForestClassifier(n_estimators=20)	130	0.714	0.643	0.409
RandomForestClassifier(n_estimators=20)	100	0.714	0.643	0.409
RandomForestClassifier(n_estimators=20)	70	0.635	0.474	0.409
RandomForestClassifier(n_estimators=20)	40	0.698	0.579	0.5
RandomForestClassifier(n_estimators=20)	20	0.667	0.529	0.409
RandomForestClassifier(n_estimators=20)	10	0.571	0.391	0.409
RandomForestClassifier(n_estimators=50)	199	0.651	0.5	0.318
RandomForestClassifier(n_estimators=50)	160	0.683	0.571	0.364
RandomForestClassifier(n_estimators=50)	130	0.683	0.571	0.364
RandomForestClassifier(n_estimators=50)	100	0.698	0.615	0.364
RandomForestClassifier(n_estimators=50)	70	0.651	0.5	0.364

RandomForestClassifier(n_estimators=50)	40	0.73	0.667	0.455
RandomForestClassifier(n_estimators=50)	20	0.698	0.6	0.409
RandomForestClassifier(n_estimators=50)	10	0.635	0.474	0.409
RandomForestClassifier(n_estimators=100)	199	0.651	0.5	0.318
RandomForestClassifier(n_estimators=100)	160	0.698	0.615	0.364
RandomForestClassifier(n_estimators=100)	130	0.698	0.615	0.364
RandomForestClassifier(n_estimators=100)	100	0.683	0.571	0.364
RandomForestClassifier(n_estimators=100)	70	0.714	0.643	0.409
RandomForestClassifier(n_estimators=100)	40	0.698	0.6	0.409
RandomForestClassifier(n_estimators=100)	20	0.714	0.643	0.409
RandomForestClassifier(n_estimators=100)	10	0.587	0.389	0.318
RandomForestClassifier(n_estimators=300)	199	0.683	0.571	0.364
RandomForestClassifier(n_estimators=300)	160	0.698	0.615	0.364
RandomForestClassifier(n_estimators=300)	130	0.714	0.667	0.364
RandomForestClassifier(n_estimators=300)	100	0.683	0.571	0.364
RandomForestClassifier(n_estimators=300)	70	0.714	0.643	0.409
RandomForestClassifier(n_estimators=300)	40	0.714	0.625	0.455
RandomForestClassifier(n_estimators=300)	20	0.683	0.562	0.409
RandomForestClassifier(n_estimators=300)	10	0.571	0.368	0.318
RandomForestClassifier(n_estimators=500)	199	0.698	0.615	0.364
RandomForestClassifier(n_estimators=500)	160	0.714	0.667	0.364
RandomForestClassifier(n_estimators=500)	130	0.698	0.615	0.364
RandomForestClassifier(n_estimators=500)	100	0.667	0.533	0.364
RandomForestClassifier(n_estimators=500)	70	0.714	0.643	0.409
RandomForestClassifier(n_estimators=500)	40	0.698	0.588	0.455
RandomForestClassifier(n_estimators=500)	20	0.698	0.6	0.409
RandomForestClassifier(n_estimators=500)	10	0.571	0.368	0.318
RandomForestClassifier(n_estimators=700)	199	0.714	0.667	0.364
RandomForestClassifier(n_estimators=700)	160	0.714	0.667	0.364
RandomForestClassifier(n_estimators=700)	130	0.714	0.667	0.364



RandomForestClassifier(n_estimators=700)	100	0.667	0.533	0.364
RandomForestClassifier(n_estimators=700)	70	0.714	0.643	0.409
RandomForestClassifier(n_estimators=700)	40	0.698	0.588	0.455
RandomForestClassifier(n_estimators=700)	20	0.698	0.6	0.409
RandomForestClassifier(n_estimators=700)	10	0.571	0.368	0.318
RandomForestClassifier(n_estimators=1000)	199	0.698	0.615	0.364
RandomForestClassifier(n_estimators=1000)	160	0.714	0.667	0.364
RandomForestClassifier(n_estimators=1000)	130	0.714	0.667	0.364
RandomForestClassifier(n_estimators=1000)	100	0.683	0.571	0.364
RandomForestClassifier(n_estimators=1000)	70	0.714	0.643	0.409
RandomForestClassifier(n_estimators=1000)	40	0.698	0.588	0.455
RandomForestClassifier(n_estimators=1000)	20	0.714	0.625	0.455
RandomForestClassifier(n_estimators=1000)	10	0.571	0.368	0.318
KNeighborsClassifier(n_neighbors=2)	199	0.698	0.636	0.318
KNeighborsClassifier(n_neighbors=2)	160	0.73	0.727	0.364
KNeighborsClassifier(n_neighbors=2)	130	0.714	0.75	0.273
KNeighborsClassifier(n_neighbors=2)	100	0.698	0.636	0.318
KNeighborsClassifier(n_neighbors=2)	70	0.73	0.727	0.364
KNeighborsClassifier(n_neighbors=2)	40	0.73	0.667	0.455
KNeighborsClassifier(n_neighbors=2)	20	0.683	0.583	0.318
KNeighborsClassifier(n_neighbors=2)	10	0.667	0.538	0.318
KNeighborsClassifier(n_neighbors=3)	199	0.635	0.471	0.364
KNeighborsClassifier(n_neighbors=3)	160	0.714	0.643	0.409
KNeighborsClassifier(n_neighbors=3)	130	0.698	0.615	0.364
KNeighborsClassifier(n_neighbors=3)	100	0.683	0.562	0.409
KNeighborsClassifier(n_neighbors=3)	70	0.667	0.526	0.455
KNeighborsClassifier(n_neighbors=3)	40	0.667	0.522	0.545
KNeighborsClassifier(n_neighbors=3)	20	0.651	0.5	0.5
KNeighborsClassifier(n_neighbors=3)	10	0.635	0.476	0.455
KNeighborsClassifier(n_neighbors=5)	199	0.667	0.545	0.273

KNeighborsClassifier(n_neighbors=5)	160	0.698	0.667	0.273
KNeighborsClassifier(n_neighbors=5)	130	0.698	0.636	0.318
KNeighborsClassifier(n_neighbors=5)	100	0.683	0.583	0.318
KNeighborsClassifier(n_neighbors=5)	70	0.714	0.667	0.364
KNeighborsClassifier(n_neighbors=5)	40	0.714	0.6	0.545
KNeighborsClassifier(n_neighbors=5)	20	0.714	0.625	0.455
KNeighborsClassifier(n_neighbors=5)	10	0.571	0.381	0.364
KNeighborsClassifier(n_neighbors=7)	199	0.683	0.6	0.273
KNeighborsClassifier(n_neighbors=7)	160	0.635	0.462	0.273
KNeighborsClassifier(n_neighbors=7)	130	0.667	0.545	0.273
KNeighborsClassifier(n_neighbors=7)	100	0.698	0.667	0.273
KNeighborsClassifier(n_neighbors=7)	70	0.73	0.778	0.318
KNeighborsClassifier(n_neighbors=7)	40	0.698	0.588	0.455
KNeighborsClassifier(n_neighbors=7)	20	0.619	0.438	0.318
KNeighborsClassifier(n_neighbors=7)	10	0.619	0.444	0.364
KNeighborsClassifier(n_neighbors=10)	199	0.698	0.714	0.227
KNeighborsClassifier(n_neighbors=10)	160	0.714	0.833	0.227
KNeighborsClassifier(n_neighbors=10)	130	0.698	0.714	0.227
KNeighborsClassifier(n_neighbors=10)	100	0.73	0.857	0.273
KNeighborsClassifier(n_neighbors=10)	70	0.683	0.75	0.136
KNeighborsClassifier(n_neighbors=10)	40	0.667	0.571	0.182
KNeighborsClassifier(n_neighbors=10)	20	0.587	0.25	0.091
KNeighborsClassifier(n_neighbors=10)	10	0.556	0.2	0.091
KNeighborsClassifier(n_neighbors=15)	199	0.683	0.625	0.227
KNeighborsClassifier(n_neighbors=15)	160	0.683	0.625	0.227
KNeighborsClassifier(n_neighbors=15)	130	0.651	0.5	0.227
KNeighborsClassifier(n_neighbors=15)	100	0.683	0.625	0.227
KNeighborsClassifier(n_neighbors=15)	70	0.667	0.571	0.182
KNeighborsClassifier(n_neighbors=15)	40	0.746	0.8	0.364
KNeighborsClassifier(n_neighbors=15)	20	0.619	0.417	0.227

KNeighborsClassifier(n_neighbors=15)	10	0.587	0.375	0.273
MLPClassifier	199	0.571	0.4	0.455
MLPClassifier	160	0.635	0.476	0.455
MLPClassifier	130	0.683	0.545	0.545
MLPClassifier	100	0.635	0.48	0.545
MLPClassifier	70	0.683	0.542	0.591
MLPClassifier	40	0.667	0.52	0.591
MLPClassifier	20	0.698	0.588	0.455
MLPClassifier	10	0.698	0.8	0.182
DecisionTreeClassifier	199	0.587	0.433	0.591
DecisionTreeClassifier	160	0.651	0.5	0.591
DecisionTreeClassifier	130	0.603	0.452	0.636
DecisionTreeClassifier	100	0.651	0.5	0.455
DecisionTreeClassifier	70	0.619	0.455	0.455
DecisionTreeClassifier	40	0.603	0.452	0.636
DecisionTreeClassifier	20	0.635	0.485	0.727
DecisionTreeClassifier	10	0.54	0.36	0.409
GaussianNB	199	0.397	0.34	0.773
GaussianNB	160	0.667	0.538	0.318
GaussianNB	130	0.683	0.571	0.364
GaussianNB	100	0.667	0.538	0.318
GaussianNB	70	0.667	0.538	0.318
GaussianNB	40	0.683	0.583	0.318
GaussianNB	20	0.651	0.5	0.227
GaussianNB	10	0.698	0.714	0.227

Table 4: Results for classification with 71.5 Maven Reuse split using PCA feature reduction

Algorithm	No. of features	Accuracy	Precision	Recall
-----------	-----------------	----------	-----------	--------

SVC(kernel='rbf')	199	0.587	0.68	0.486
SVC(kernel='rbf')	160	0.46	0.514	0.543
SVC(kernel='rbf')	130	0.46	0.517	0.429
SVC(kernel='rbf')	100	0.508	0.567	0.486
SVC(kernel='rbf')	70	0.444	0.5	0.371
SVC(kernel='rbf')	40	0.476	0.55	0.314
SVC(kernel='rbf')	20	0.492	0.588	0.286
SVC(kernel='rbf')	10	0.444	0.5	0.2
SVC(kernel='linear')	199	0.54	0.6	0.514
SVC(kernel='linear')	160	0.556	0.64	0.457
SVC(kernel='linear')	130	0.54	0.636	0.4
SVC(kernel='linear')	100	0.619	0.704	0.543
SVC(kernel='linear')	70	0.683	0.8	0.571
SVC(kernel='linear')	40	0.619	0.762	0.457
SVC(kernel='linear')	20	0.54	0.714	0.286
SVC(kernel='linear')	10	0.476	0.571	0.229
SVC(kernel='poly', degree=3)	199	0.476	0.625	0.143
SVC(kernel='poly', degree=3)	160	0.444	0.0	0.0
SVC(kernel='poly', degree=3)	130	0.444	0.0	0.0
SVC(kernel='poly', degree=3)	100	0.444	0.0	0.0
SVC(kernel='poly', degree=3)	70	0.476	1.0	0.057
SVC(kernel='poly', degree=3)	40	0.46	0.667	0.057
SVC(kernel='poly', degree=3)	20	0.508	1.0	0.114
SVC(kernel='poly', degree=3)	10	0.476	0.667	0.114
SVC(kernel='poly', degree=4)	199	0.492	0.714	0.143
SVC(kernel='poly', degree=4)	160	0.444	0.0	0.0
SVC(kernel='poly', degree=4)	130	0.444	0.0	0.0
SVC(kernel='poly', degree=4)	100	0.444	0.0	0.0
SVC(kernel='poly', degree=4)	70	0.444	0.0	0.0
SVC(kernel='poly', degree=4)	40	0.46	0.667	0.057

SVC(kernel='poly', degree=4)	20	0.476	0.625	0.143
SVC(kernel='poly', degree=4)	10	0.46	0.571	0.114
SVC(kernel='poly', degree=5)	199	0.476	0.667	0.114
SVC(kernel='poly', degree=5)	160	0.444	0.0	0.0
SVC(kernel='poly', degree=5)	130	0.444	0.0	0.0
SVC(kernel='poly', degree=5)	100	0.444	0.0	0.0
SVC(kernel='poly', degree=5)	70	0.444	0.0	0.0
SVC(kernel='poly', degree=5)	40	0.46	0.667	0.057
SVC(kernel='poly', degree=5)	20	0.492	1.0	0.086
SVC(kernel='poly', degree=5)	10	0.476	0.667	0.114
SVC(kernel='sigmoid')	199	0.603	0.727	0.457
SVC(kernel='sigmoid'))	160	0.619	0.739	0.486
SVC(kernel='sigmoid')	130	0.587	0.737	0.4
SVC(kernel='sigmoid')	100	0.587	0.696	0.457
SVC(kernel='sigmoid')	70	0.524	0.667	0.286
SVC(kernel='sigmoid')	40	0.571	0.9	0.257
SVC(kernel='sigmoid')	20	0.492	0.615	0.229
SVC(kernel='sigmoid')	10	0.429	0.471	0.229
RandomForestClassifier(n_estimators=10)	199	0.571	0.654	0.486
RandomForestClassifier(n_estimators=10)	160	0.413	0.464	0.371
RandomForestClassifier(n_estimators=10)	130	0.46	0.522	0.343
RandomForestClassifier(n_estimators=10)	100	0.429	0.478	0.314
RandomForestClassifier(n_estimators=10)	70	0.476	0.531	0.486
RandomForestClassifier(n_estimators=10)	40	0.603	0.692	0.514
RandomForestClassifier(n_estimators=10)	20	0.635	0.714	0.571
RandomForestClassifier(n_estimators=10)	10	0.524	0.571	0.571
RandomForestClassifier(n_estimators=20)	199	0.571	0.618	0.6
RandomForestClassifier(n_estimators=20)	160	0.429	0.484	0.429
RandomForestClassifier(n_estimators=20)	130	0.651	0.76	0.543
RandomForestClassifier(n_estimators=20)	100	0.444	0.5	0.4

RandomForestClassifier(n_estimators=20)	70	0.429	0.481	0.371
RandomForestClassifier(n_estimators=20)	40	0.571	0.682	0.429
RandomForestClassifier(n_estimators=20)	20	0.619	0.667	0.629
RandomForestClassifier(n_estimators=20)	10	0.54	0.583	0.6
RandomForestClassifier(n_estimators=50)	199	0.556	0.606	0.571
RandomForestClassifier(n_estimators=50)	160	0.413	0.469	0.429
RandomForestClassifier(n_estimators=50)	130	0.524	0.609	0.4
RandomForestClassifier(n_estimators=50)	100	0.444	0.5	0.429
RandomForestClassifier(n_estimators=50)	70	0.476	0.531	0.486
RandomForestClassifier(n_estimators=50)	40	0.54	0.615	0.457
RandomForestClassifier(n_estimators=50)	20	0.619	0.667	0.629
RandomForestClassifier(n_estimators=50)	10	0.54	0.579	0.629
RandomForestClassifier(n_estimators=100)	199	0.603	0.647	0.629
RandomForestClassifier(n_estimators=100)	160	0.492	0.56	0.4
RandomForestClassifier(n_estimators=100)	130	0.492	0.565	0.371
RandomForestClassifier(n_estimators=100)	100	0.492	0.552	0.457
RandomForestClassifier(n_estimators=100)	70	0.492	0.548	0.486
RandomForestClassifier(n_estimators=100)	40	0.587	0.645	0.571
RandomForestClassifier(n_estimators=100)	20	0.667	0.706	0.686
RandomForestClassifier(n_estimators=100)	10	0.571	0.611	0.629
RandomForestClassifier(n_estimators=300)	199	0.683	0.759	0.629
RandomForestClassifier(n_estimators=300)	160	0.476	0.542	0.371
RandomForestClassifier(n_estimators=300)	130	0.508	0.583	0.4
RandomForestClassifier(n_estimators=300)	100	0.54	0.588	0.571
RandomForestClassifier(n_estimators=300)	70	0.524	0.586	0.486
RandomForestClassifier(n_estimators=300)	40	0.556	0.606	0.571
RandomForestClassifier(n_estimators=300)	20	0.635	0.676	0.657
RandomForestClassifier(n_estimators=300)	10	0.524	0.571	0.571
RandomForestClassifier(n_estimators=500)	199	0.667	0.75	0.6
RandomForestClassifier(n_estimators=500)	160	0.46	0.522	0.343

RandomForestClassifier(n_estimators=500)	130	0.46	0.52	0.371
RandomForestClassifier(n_estimators=500)	100	0.508	0.562	0.514
RandomForestClassifier(n_estimators=500)	70	0.524	0.586	0.486
RandomForestClassifier(n_estimators=500)	40	0.619	0.667	0.629
RandomForestClassifier(n_estimators=500)	20	0.619	0.667	0.629
RandomForestClassifier(n_estimators=500)	10	0.54	0.588	0.571
RandomForestClassifier(n_estimators=700)	199	0.635	0.7	0.6
RandomForestClassifier(n_estimators=700)	160	0.476	0.542	0.371
RandomForestClassifier(n_estimators=700)	130	0.492	0.56	0.4
RandomForestClassifier(n_estimators=700)	100	0.524	0.581	0.514
RandomForestClassifier(n_estimators=700)	70	0.524	0.586	0.486
RandomForestClassifier(n_estimators=700)	40	0.603	0.647	0.629
RandomForestClassifier(n_estimators=700)	20	0.619	0.667	0.629
RandomForestClassifier(n_estimators=700)	10	0.556	0.6	0.6
RandomForestClassifier(n_estimators=1000)	199	0.667	0.733	0.629
RandomForestClassifier(n_estimators=1000)	160	0.429	0.48	0.343
RandomForestClassifier(n_estimators=1000)	130	0.492	0.56	0.4
RandomForestClassifier(n_estimators=1000)	100	0.492	0.548	0.486
RandomForestClassifier(n_estimators=1000)	70	0.54	0.6	0.514
RandomForestClassifier(n_estimators=1000)	40	0.587	0.636	0.6
RandomForestClassifier(n_estimators=1000)	20	0.619	0.667	0.629
RandomForestClassifier(n_estimators=1000)	10	0.524	0.571	0.571
KNeighborsClassifier(n_neighbors=2)	199	0.571	0.722	0.371
KNeighborsClassifier(n_neighbors=2)	160	0.508	0.577	0.429
KNeighborsClassifier(n_neighbors=2)	130	0.508	0.833	0.143
KNeighborsClassifier(n_neighbors=2)	100	0.46	0.556	0.143
KNeighborsClassifier(n_neighbors=2)	70	0.54	0.8	0.229
KNeighborsClassifier(n_neighbors=2)	40	0.556	0.818	0.257
KNeighborsClassifier(n_neighbors=2)	20	0.476	0.583	0.2
KNeighborsClassifier(n_neighbors=2)	10	0.54	0.714	0.286

KNeighborsClassifier(n_neighbors=3)	199	0.587	0.655	0.543
KNeighborsClassifier(n_neighbors=3)	160	0.476	0.524	0.629
KNeighborsClassifier(n_neighbors=3)	130	0.476	0.571	0.229
KNeighborsClassifier(n_neighbors=3)	100	0.508	0.625	0.286
KNeighborsClassifier(n_neighbors=3)	70	0.524	0.632	0.343
KNeighborsClassifier(n_neighbors=3)	40	0.524	0.647	0.314
KNeighborsClassifier(n_neighbors=3)	20	0.476	0.536	0.429
KNeighborsClassifier(n_neighbors=3)	10	0.571	0.625	0.571
KNeighborsClassifier(n_neighbors=5)	199	0.54	0.636	0.4
KNeighborsClassifier(n_neighbors=5)	160	0.476	0.519	0.771
KNeighborsClassifier(n_neighbors=5)	130	0.524	0.857	0.171
KNeighborsClassifier(n_neighbors=5)	100	0.54	0.714	0.286
KNeighborsClassifier(n_neighbors=5)	70	0.571	0.722	0.371
KNeighborsClassifier(n_neighbors=5)	40	0.54	0.714	0.286
KNeighborsClassifier(n_neighbors=5)	20	0.492	0.552	0.457
KNeighborsClassifier(n_neighbors=5)	10	0.508	0.567	0.486
KNeighborsClassifier(n_neighbors=7)	199	0.46	0.529	0.257
KNeighborsClassifier(n_neighbors=7)	160	0.476	0.519	0.8
KNeighborsClassifier(n_neighbors=7)	130	0.476	0.75	0.086
KNeighborsClassifier(n_neighbors=7)	100	0.587	0.8	0.343
KNeighborsClassifier(n_neighbors=7)	70	0.556	0.769	0.286
KNeighborsClassifier(n_neighbors=7)	40	0.508	0.7	0.2
KNeighborsClassifier(n_neighbors=7)	20	0.556	0.64	0.457
KNeighborsClassifier(n_neighbors=7)	10	0.571	0.633	0.543
KNeighborsClassifier(n_neighbors=10)	199	0.476	0.583	0.2
KNeighborsClassifier(n_neighbors=10)	160	0.508	0.583	0.4
KNeighborsClassifier(n_neighbors=10)	130	0.444	0.0	0.0
KNeighborsClassifier(n_neighbors=10)	100	0.46	0.6	0.086
KNeighborsClassifier(n_neighbors=10)	70	0.508	0.7	0.2
KNeighborsClassifier(n_neighbors=10)	40	0.492	0.8	0.114



KNeighborsClassifier(n_neighbors=10)	20	0.508	0.625	0.286
KNeighborsClassifier(n_neighbors=10)	10	0.508	0.625	0.286
KNeighborsClassifier(n_neighbors=15)	199	0.429	0.455	0.143
KNeighborsClassifier(n_neighbors=15)	160	0.492	0.714	0.143
KNeighborsClassifier(n_neighbors=15)	130	0.444	0.0	0.0
KNeighborsClassifier(n_neighbors=15)	100	0.46	0.667	0.057
KNeighborsClassifier(n_neighbors=15)	70	0.492	0.8	0.114
KNeighborsClassifier(n_neighbors=15)	40	0.492	1.0	0.086
KNeighborsClassifier(n_neighbors=15)	20	0.508	0.625	0.286
KNeighborsClassifier(n_neighbors=15)	10	0.492	0.571	0.343
MLPClassifier	199	0.54	0.594	0.543
MLPClassifier	160	0.571	0.667	0.457
MLPClassifier	130	0.54	0.636	0.4
MLPClassifier	100	0.667	0.769	0.571
MLPClassifier	70	0.619	0.69	0.571
MLPClassifier	40	0.667	0.769	0.571
MLPClassifier	20	0.714	0.84	0.6
MLPClassifier	10	0.46	0.522	0.343
DecisionTreeClassifier	199	0.508	0.556	0.571
DecisionTreeClassifier	160	0.54	0.625	0.429
DecisionTreeClassifier	130	0.524	0.593	0.457
DecisionTreeClassifier	100	0.619	0.677	0.6
DecisionTreeClassifier	70	0.492	0.543	0.543
DecisionTreeClassifier	40	0.619	0.704	0.543
DecisionTreeClassifier	20	0.619	0.649	0.686
DecisionTreeClassifier	10	0.524	0.576	0.543
GaussianNB	199	0.492	0.588	0.286
GaussianNB	160	0.397	0.435	0.286
GaussianNB	130	0.397	0.435	0.286
GaussianNB	100	0.413	0.458	0.314

GaussianNB	70	0.397	0.429	0.257
GaussianNB	40	0.413	0.438	0.2
GaussianNB	20	0.429	0.467	0.2
GaussianNB	10	0.429	0.462	0.171

*Table 5: Results for classification with 100 Maven Reuse split using PCA feature reduction*

<b>Algorithm</b>	<b>No. of features</b>	<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>
SVC(kernel='rbf')	199	0.603	0.619	0.433
SVC(kernel='rbf')	160	0.571	0.588	0.333
SVC(kernel='rbf')	130	0.508	0.48	0.4
SVC(kernel='rbf')	100	0.571	0.556	0.5
SVC(kernel='rbf')	70	0.524	0.5	0.4
SVC(kernel='rbf')	40	0.556	0.556	0.333
SVC(kernel='rbf')	20	0.587	0.625	0.333
SVC(kernel='rbf')	10	0.556	0.583	0.233
SVC(kernel='linear')	199	0.619	0.6	0.6
SVC(kernel='linear')	160	0.556	0.55	0.367
SVC(kernel='linear')	130	0.571	0.571	0.4
SVC(kernel='linear')	100	0.683	0.692	0.6
SVC(kernel='linear')	70	0.698	0.789	0.5
SVC(kernel='linear')	40	0.683	0.778	0.467
SVC(kernel='linear')	20	0.635	0.769	0.333
SVC(kernel='linear')	10	0.571	0.667	0.2
SVC(kernel='poly', degree=3)	199	0.556	0.625	0.167
SVC(kernel='poly', degree=3)	160	0.524	0.0	0.0
SVC(kernel='poly', degree=3)	130	0.524	0.0	0.0
SVC(kernel='poly', degree=3)	100	0.524	0.0	0.0
SVC(kernel='poly', degree=3)	70	0.524	0.0	0.0

SVC(kernel='poly', degree=3)	40	0.54	0.667	0.067
SVC(kernel='poly', degree=3)	20	0.571	0.8	0.133
SVC(kernel='poly', degree=3)	10	0.556	0.667	0.133
SVC(kernel='poly', degree=4)	199	0.571	0.714	0.167
SVC(kernel='poly', degree=4)	160	0.524	0.0	0.0
SVC(kernel='poly', degree=4)	130	0.524	0.0	0.0
SVC(kernel='poly', degree=4)	100	0.524	0.0	0.0
SVC(kernel='poly', degree=4)	70	0.524	0.0	0.0
SVC(kernel='poly', degree=4)	40	0.54	0.667	0.067
SVC(kernel='poly', degree=4)	20	0.556	0.625	0.167
SVC(kernel='poly', degree=4)	10	0.54	0.571	0.133
SVC(kernel='poly', degree=5)	199	0.556	0.667	0.133
SVC(kernel='poly', degree=5)	160	0.524	0.0	0.0
SVC(kernel='poly', degree=5)	130	0.524	0.0	0.0
SVC(kernel='poly', degree=5)	100	0.524	0.0	0.0
SVC(kernel='poly', degree=5)	70	0.524	0.0	0.0
SVC(kernel='poly', degree=5)	40	0.54	0.667	0.067
SVC(kernel='poly', degree=5)	20	0.556	0.75	0.1
SVC(kernel='poly', degree=5)	10	0.556	0.667	0.133
SVC(kernel='sigmoid')	199	0.619	0.667	0.4
SVC(kernel='sigmoid'))	160	0.54	0.526	0.333
SVC(kernel='sigmoid')	130	0.635	0.706	0.4
SVC(kernel='sigmoid')	100	0.619	0.65	0.433
SVC(kernel='sigmoid')	70	0.619	0.75	0.3
SVC(kernel='sigmoid')	40	0.651	0.9	0.3
SVC(kernel='sigmoid')	20	0.587	0.7	0.233
SVC(kernel='sigmoid')	10	0.556	0.6	0.2
RandomForestClassifier(n_estimators=10)	199	0.667	0.737	0.467
RandomForestClassifier(n_estimators=10)	160	0.46	0.417	0.333
RandomForestClassifier(n_estimators=10)	130	0.476	0.421	0.267

RandomForestClassifier(n_estimators=10)	100	0.556	0.556	0.333
RandomForestClassifier(n_estimators=10)	70	0.397	0.25	0.133
RandomForestClassifier(n_estimators=10)	40	0.524	0.5	0.467
RandomForestClassifier(n_estimators=10)	20	0.619	0.625	0.5
RandomForestClassifier(n_estimators=10)	10	0.587	0.571	0.533
RandomForestClassifier(n_estimators=20)	199	0.667	0.68	0.567
RandomForestClassifier(n_estimators=20)	160	0.54	0.526	0.333
RandomForestClassifier(n_estimators=20)	130	0.587	0.6	0.4
RandomForestClassifier(n_estimators=20)	100	0.571	0.579	0.367
RandomForestClassifier(n_estimators=20)	70	0.397	0.3	0.2
RandomForestClassifier(n_estimators=20)	40	0.619	0.6	0.6
RandomForestClassifier(n_estimators=20)	20	0.683	0.708	0.567
RandomForestClassifier(n_estimators=20)	10	0.556	0.528	0.633
RandomForestClassifier(n_estimators=50)	199	0.714	0.75	0.6
RandomForestClassifier(n_estimators=50)	160	0.524	0.5	0.333
RandomForestClassifier(n_estimators=50)	130	0.54	0.524	0.367
RandomForestClassifier(n_estimators=50)	100	0.556	0.538	0.467
RandomForestClassifier(n_estimators=50)	70	0.556	0.538	0.467
RandomForestClassifier(n_estimators=50)	40	0.635	0.606	0.667
RandomForestClassifier(n_estimators=50)	20	0.683	0.667	0.667
RandomForestClassifier(n_estimators=50)	10	0.54	0.515	0.567
RandomForestClassifier(n_estimators=100)	199	0.683	0.708	0.567
RandomForestClassifier(n_estimators=100)	160	0.54	0.526	0.333
RandomForestClassifier(n_estimators=100)	130	0.587	0.6	0.4
RandomForestClassifier(n_estimators=100)	100	0.571	0.556	0.5
RandomForestClassifier(n_estimators=100)	70	0.54	0.52	0.433
RandomForestClassifier(n_estimators=100)	40	0.603	0.586	0.567
RandomForestClassifier(n_estimators=100)	20	0.635	0.621	0.6
RandomForestClassifier(n_estimators=100)	10	0.571	0.548	0.567
RandomForestClassifier(n_estimators=300)	199	0.683	0.708	0.567

RandomForestClassifier(n_estimators=300)	160	0.571	0.579	0.367
RandomForestClassifier(n_estimators=300)	130	0.54	0.524	0.367
RandomForestClassifier(n_estimators=300)	100	0.556	0.536	0.5
RandomForestClassifier(n_estimators=300)	70	0.492	0.464	0.433
RandomForestClassifier(n_estimators=300)	40	0.603	0.586	0.567
RandomForestClassifier(n_estimators=300)	20	0.667	0.645	0.667
RandomForestClassifier(n_estimators=300)	10	0.556	0.529	0.6
RandomForestClassifier(n_estimators=500)	199	0.667	0.696	0.533
RandomForestClassifier(n_estimators=500)	160	0.571	0.579	0.367
RandomForestClassifier(n_estimators=500)	130	0.556	0.545	0.4
RandomForestClassifier(n_estimators=500)	100	0.556	0.536	0.5
RandomForestClassifier(n_estimators=500)	70	0.556	0.533	0.533
RandomForestClassifier(n_estimators=500)	40	0.603	0.586	0.567
RandomForestClassifier(n_estimators=500)	20	0.635	0.613	0.633
RandomForestClassifier(n_estimators=500)	10	0.524	0.5	0.567
RandomForestClassifier(n_estimators=700)	199	0.698	0.72	0.6
RandomForestClassifier(n_estimators=700)	160	0.571	0.588	0.333
RandomForestClassifier(n_estimators=700)	130	0.556	0.545	0.4
RandomForestClassifier(n_estimators=700)	100	0.587	0.571	0.533
RandomForestClassifier(n_estimators=700)	70	0.556	0.536	0.5
RandomForestClassifier(n_estimators=700)	40	0.635	0.621	0.6
RandomForestClassifier(n_estimators=700)	20	0.635	0.613	0.633
RandomForestClassifier(n_estimators=700)	10	0.524	0.5	0.567
RandomForestClassifier(n_estimators=1000)	199	0.714	0.731	0.633
RandomForestClassifier(n_estimators=1000)	160	0.571	0.588	0.333
RandomForestClassifier(n_estimators=1000)	130	0.54	0.524	0.367
RandomForestClassifier(n_estimators=1000)	100	0.603	0.593	0.533
RandomForestClassifier(n_estimators=1000)	70	0.556	0.536	0.5
RandomForestClassifier(n_estimators=1000)	40	0.603	0.593	0.533
RandomForestClassifier(n_estimators=1000)	20	0.635	0.613	0.633

RandomForestClassifier(n_estimators=1000)	10	0.54	0.515	0.567
KNeighborsClassifier(n_neighbors=2)	199	0.667	0.765	0.433
KNeighborsClassifier(n_neighbors=2)	160	0.524	0.5	0.433
KNeighborsClassifier(n_neighbors=2)	130	0.603	1.0	0.167
KNeighborsClassifier(n_neighbors=2)	100	0.54	0.571	0.133
KNeighborsClassifier(n_neighbors=2)	70	0.635	0.889	0.267
KNeighborsClassifier(n_neighbors=2)	40	0.603	0.727	0.267
KNeighborsClassifier(n_neighbors=2)	20	0.603	0.778	0.233
KNeighborsClassifier(n_neighbors=2)	10	0.571	0.615	0.267
KNeighborsClassifier(n_neighbors=3)	199	0.587	0.571	0.533
KNeighborsClassifier(n_neighbors=3)	160	0.429	0.429	0.6
KNeighborsClassifier(n_neighbors=3)	130	0.603	0.857	0.2
KNeighborsClassifier(n_neighbors=3)	100	0.587	0.667	0.267
KNeighborsClassifier(n_neighbors=3)	70	0.571	0.588	0.333
KNeighborsClassifier(n_neighbors=3)	40	0.54	0.529	0.3
KNeighborsClassifier(n_neighbors=3)	20	0.571	0.556	0.5
KNeighborsClassifier(n_neighbors=3)	10	0.603	0.586	0.567
KNeighborsClassifier(n_neighbors=5)	199	0.54	0.524	0.367
KNeighborsClassifier(n_neighbors=5)	160	0.397	0.423	0.733
KNeighborsClassifier(n_neighbors=5)	130	0.571	0.8	0.133
KNeighborsClassifier(n_neighbors=5)	100	0.587	0.7	0.233
KNeighborsClassifier(n_neighbors=5)	70	0.587	0.611	0.367
KNeighborsClassifier(n_neighbors=5)	40	0.571	0.636	0.233
KNeighborsClassifier(n_neighbors=5)	20	0.587	0.577	0.5
KNeighborsClassifier(n_neighbors=5)	10	0.571	0.556	0.5
KNeighborsClassifier(n_neighbors=7)	199	0.508	0.471	0.267
KNeighborsClassifier(n_neighbors=7)	160	0.397	0.426	0.767
KNeighborsClassifier(n_neighbors=7)	130	0.524	0.5	0.033
KNeighborsClassifier(n_neighbors=7)	100	0.587	0.643	0.3
KNeighborsClassifier(n_neighbors=7)	70	0.635	0.818	0.3

KNeighborsClassifier(n_neighbors=7)	40	0.587	0.75	0.2
KNeighborsClassifier(n_neighbors=7)	20	0.651	0.7	0.467
KNeighborsClassifier(n_neighbors=7)	10	0.556	0.542	0.433
KNeighborsClassifier(n_neighbors=10)	199	0.556	0.6	0.2
KNeighborsClassifier(n_neighbors=10)	160	0.492	0.458	0.367
KNeighborsClassifier(n_neighbors=10)	130	0.524	0.0	0.0
KNeighborsClassifier(n_neighbors=10)	100	0.54	0.667	0.067
KNeighborsClassifier(n_neighbors=10)	70	0.571	0.714	0.167
KNeighborsClassifier(n_neighbors=10)	40	0.556	1.0	0.067
KNeighborsClassifier(n_neighbors=10)	20	0.587	0.643	0.3
KNeighborsClassifier(n_neighbors=10)	10	0.54	0.538	0.233
KNeighborsClassifier(n_neighbors=15)	199	0.524	0.5	0.167
KNeighborsClassifier(n_neighbors=15)	160	0.587	0.833	0.167
KNeighborsClassifier(n_neighbors=15)	130	0.524	0.0	0.0
KNeighborsClassifier(n_neighbors=15)	100	0.54	1.0	0.033
KNeighborsClassifier(n_neighbors=15)	70	0.571	0.8	0.133
KNeighborsClassifier(n_neighbors=15)	40	0.571	1.0	0.1
KNeighborsClassifier(n_neighbors=15)	20	0.556	0.583	0.233
KNeighborsClassifier(n_neighbors=15)	10	0.54	0.526	0.333
MLPClassifier	199	0.587	0.562	0.6
MLPClassifier	160	0.556	0.545	0.4
MLPClassifier	130	0.571	0.588	0.333
MLPClassifier	100	0.714	0.773	0.567
MLPClassifier	70	0.683	0.679	0.633
MLPClassifier	40	0.73	0.783	0.6
MLPClassifier	20	0.698	0.72	0.6
MLPClassifier	10	0.524	0.5	0.367
DecisionTreeClassifier	199	0.635	0.621	0.6
DecisionTreeClassifier	160	0.492	0.472	0.567
DecisionTreeClassifier	130	0.524	0.5	0.5

DecisionTreeClassifier	100	0.508	0.485	0.533
DecisionTreeClassifier	70	0.524	0.5	0.567
DecisionTreeClassifier	40	0.571	0.543	0.633
DecisionTreeClassifier	20	0.683	0.667	0.667
DecisionTreeClassifier	10	0.556	0.528	0.633
GaussianNB	199	0.508	0.476	0.333
GaussianNB	160	0.46	0.409	0.3
GaussianNB	130	0.444	0.381	0.267
GaussianNB	100	0.444	0.391	0.3
GaussianNB	70	0.444	0.368	0.233
GaussianNB	40	0.492	0.438	0.233
GaussianNB	20	0.508	0.467	0.233
GaussianNB	10	0.508	0.462	0.2

Table 6: Results for classification with 200 Maven Reuse split using PCA feature reduction

Algorithm	No. of features	Accuracy	Precision	Recall
SVC(kernel='rbf')	199	0.714	0.75	0.273
SVC(kernel='rbf')	160	0.683	1.0	0.091
SVC(kernel='rbf')	130	0.683	1.0	0.091
SVC(kernel='rbf')	100	0.683	0.75	0.136
SVC(kernel='rbf')	70	0.714	0.7	0.318
SVC(kernel='rbf')	40	0.635	0.467	0.318
SVC(kernel='rbf')	20	0.683	0.571	0.364
SVC(kernel='rbf')	10	0.698	0.636	0.318
SVC(kernel='linear')	199	0.603	0.435	0.455
SVC(kernel='linear')	160	0.571	0.353	0.273
SVC(kernel='linear')	130	0.651	0.5	0.318
SVC(kernel='linear')	100	0.667	0.526	0.455



SVC(kernel='linear')	70	0.714	0.643	0.409
SVC(kernel='linear')	40	0.714	0.7	0.318
SVC(kernel='linear')	20	0.714	0.643	0.409
SVC(kernel='linear')	10	0.698	0.667	0.273
SVC(kernel='poly', degree=3)	199	0.698	0.714	0.227
SVC(kernel='poly', degree=3)	160	0.651	0.0	0.0
SVC(kernel='poly', degree=3)	130	0.651	0.0	0.0
SVC(kernel='poly', degree=3)	100	0.651	0.0	0.0
SVC(kernel='poly', degree=3)	70	0.651	0.0	0.0
SVC(kernel='poly', degree=3)	40	0.667	0.667	0.091
SVC(kernel='poly', degree=3)	20	0.698	0.8	0.182
SVC(kernel='poly', degree=3)	10	0.683	0.667	0.182
SVC(kernel='poly', degree=4)	199	0.698	0.714	0.227
SVC(kernel='poly', degree=4)	160	0.651	0.0	0.0
SVC(kernel='poly', degree=4)	130	0.651	0.0	0.0
SVC(kernel='poly', degree=4)	100	0.651	0.0	0.0
SVC(kernel='poly', degree=4)	70	0.651	0.0	0.0
SVC(kernel='poly', degree=4)	40	0.667	0.667	0.091
SVC(kernel='poly', degree=4)	20	0.683	0.667	0.182
SVC(kernel='poly', degree=4)	10	0.667	0.571	0.182
SVC(kernel='poly', degree=5)	199	0.698	0.714	0.227
SVC(kernel='poly', degree=5)	160	0.651	0.0	0.0
SVC(kernel='poly', degree=5)	130	0.651	0.0	0.0
SVC(kernel='poly', degree=5)	100	0.651	0.0	0.0
SVC(kernel='poly', degree=5)	70	0.651	0.0	0.0
SVC(kernel='poly', degree=5)	40	0.667	0.667	0.091
SVC(kernel='poly', degree=5)	20	0.683	0.75	0.136
SVC(kernel='poly', degree=5)	10	0.683	0.667	0.182
SVC(kernel='sigmoid')	199	0.667	0.538	0.318
SVC(kernel='sigmoid'))	160	0.651	0.5	0.136

SVC(kernel='sigmoid')	130	0.635	0.444	0.182
SVC(kernel='sigmoid')	100	0.683	0.625	0.227
SVC(kernel='sigmoid')	70	0.714	0.7	0.318
SVC(kernel='sigmoid')	40	0.746	0.8	0.364
SVC(kernel='sigmoid')	20	0.698	0.636	0.318
SVC(kernel='sigmoid')	10	0.619	0.4	0.182
RandomForestClassifier(n_estimators=10)	199	0.698	0.636	0.318
RandomForestClassifier(n_estimators=10)	160	0.603	0.421	0.364
RandomForestClassifier(n_estimators=10)	130	0.698	0.615	0.364
RandomForestClassifier(n_estimators=10)	100	0.619	0.429	0.273
RandomForestClassifier(n_estimators=10)	70	0.587	0.409	0.409
RandomForestClassifier(n_estimators=10)	40	0.635	0.467	0.318
RandomForestClassifier(n_estimators=10)	20	0.683	0.556	0.455
RandomForestClassifier(n_estimators=10)	10	0.651	0.5	0.409
RandomForestClassifier(n_estimators=20)	199	0.603	0.4	0.273
RandomForestClassifier(n_estimators=20)	160	0.635	0.471	0.364
RandomForestClassifier(n_estimators=20)	130	0.714	0.667	0.364
RandomForestClassifier(n_estimators=20)	100	0.651	0.5	0.364
RandomForestClassifier(n_estimators=20)	70	0.603	0.44	0.5
RandomForestClassifier(n_estimators=20)	40	0.651	0.5	0.409
RandomForestClassifier(n_estimators=20)	20	0.651	0.5	0.409
RandomForestClassifier(n_estimators=20)	10	0.667	0.524	0.5
RandomForestClassifier(n_estimators=50)	199	0.651	0.5	0.318
RandomForestClassifier(n_estimators=50)	160	0.714	0.643	0.409
RandomForestClassifier(n_estimators=50)	130	0.698	0.636	0.318
RandomForestClassifier(n_estimators=50)	100	0.635	0.462	0.273
RandomForestClassifier(n_estimators=50)	70	0.635	0.471	0.364
RandomForestClassifier(n_estimators=50)	40	0.683	0.562	0.409
RandomForestClassifier(n_estimators=50)	20	0.635	0.474	0.409
RandomForestClassifier(n_estimators=50)	10	0.635	0.476	0.455

RandomForestClassifier(n_estimators=100)	199	0.651	0.5	0.318
RandomForestClassifier(n_estimators=100)	160	0.683	0.571	0.364
RandomForestClassifier(n_estimators=100)	130	0.683	0.571	0.364
RandomForestClassifier(n_estimators=100)	100	0.667	0.533	0.364
RandomForestClassifier(n_estimators=100)	70	0.698	0.6	0.409
RandomForestClassifier(n_estimators=100)	40	0.651	0.5	0.409
RandomForestClassifier(n_estimators=100)	20	0.635	0.474	0.409
RandomForestClassifier(n_estimators=100)	10	0.651	0.5	0.455
RandomForestClassifier(n_estimators=300)	199	0.683	0.571	0.364
RandomForestClassifier(n_estimators=300)	160	0.698	0.615	0.364
RandomForestClassifier(n_estimators=300)	130	0.651	0.5	0.409
RandomForestClassifier(n_estimators=300)	100	0.698	0.615	0.364
RandomForestClassifier(n_estimators=300)	70	0.683	0.562	0.409
RandomForestClassifier(n_estimators=300)	40	0.683	0.562	0.409
RandomForestClassifier(n_estimators=300)	20	0.651	0.5	0.455
RandomForestClassifier(n_estimators=300)	10	0.619	0.455	0.455
RandomForestClassifier(n_estimators=500)	199	0.698	0.615	0.364
RandomForestClassifier(n_estimators=500)	160	0.683	0.571	0.364
RandomForestClassifier(n_estimators=500)	130	0.651	0.5	0.409
RandomForestClassifier(n_estimators=500)	100	0.667	0.533	0.364
RandomForestClassifier(n_estimators=500)	70	0.698	0.588	0.455
RandomForestClassifier(n_estimators=500)	40	0.667	0.529	0.409
RandomForestClassifier(n_estimators=500)	20	0.667	0.526	0.455
RandomForestClassifier(n_estimators=500)	10	0.635	0.476	0.455
RandomForestClassifier(n_estimators=700)	199	0.714	0.667	0.364
RandomForestClassifier(n_estimators=700)	160	0.683	0.571	0.364
RandomForestClassifier(n_estimators=700)	130	0.667	0.529	0.409
RandomForestClassifier(n_estimators=700)	100	0.683	0.571	0.364
RandomForestClassifier(n_estimators=700)	70	0.683	0.556	0.455
RandomForestClassifier(n_estimators=700)	40	0.651	0.5	0.409

RandomForestClassifier(n_estimators=700)	20	0.667	0.526	0.455
RandomForestClassifier(n_estimators=700)	10	0.619	0.455	0.455
RandomForestClassifier(n_estimators=1000)	199	0.698	0.615	0.364
RandomForestClassifier(n_estimators=1000)	160	0.683	0.571	0.364
RandomForestClassifier(n_estimators=1000)	130	0.667	0.529	0.409
RandomForestClassifier(n_estimators=1000)	100	0.683	0.562	0.409
RandomForestClassifier(n_estimators=1000)	70	0.683	0.562	0.409
RandomForestClassifier(n_estimators=1000)	40	0.651	0.5	0.409
RandomForestClassifier(n_estimators=1000)	20	0.667	0.526	0.455
RandomForestClassifier(n_estimators=1000)	10	0.619	0.455	0.455
KNeighborsClassifier(n_neighbors=2)	199	0.698	0.636	0.318
KNeighborsClassifier(n_neighbors=2)	160	0.635	0.478	0.5
KNeighborsClassifier(n_neighbors=2)	130	0.683	0.75	0.136
KNeighborsClassifier(n_neighbors=2)	100	0.667	0.6	0.136
KNeighborsClassifier(n_neighbors=2)	70	0.683	0.75	0.136
KNeighborsClassifier(n_neighbors=2)	40	0.698	0.8	0.182
KNeighborsClassifier(n_neighbors=2)	20	0.667	0.6	0.136
KNeighborsClassifier(n_neighbors=2)	10	0.667	0.556	0.227
KNeighborsClassifier(n_neighbors=3)	199	0.635	0.471	0.364
KNeighborsClassifier(n_neighbors=3)	160	0.413	0.308	0.545
KNeighborsClassifier(n_neighbors=3)	130	0.683	0.75	0.136
KNeighborsClassifier(n_neighbors=3)	100	0.667	0.571	0.182
KNeighborsClassifier(n_neighbors=3)	70	0.667	0.6	0.136
KNeighborsClassifier(n_neighbors=3)	40	0.667	0.556	0.227
KNeighborsClassifier(n_neighbors=3)	20	0.683	0.571	0.364
KNeighborsClassifier(n_neighbors=3)	10	0.635	0.471	0.364
KNeighborsClassifier(n_neighbors=5)	199	0.667	0.545	0.273
KNeighborsClassifier(n_neighbors=5)	160	0.333	0.273	0.545
KNeighborsClassifier(n_neighbors=5)	130	0.714	1.0	0.182
KNeighborsClassifier(n_neighbors=5)	100	0.667	0.6	0.136

KNeighborsClassifier(n_neighbors=5)	70	0.698	1.0	0.136
KNeighborsClassifier(n_neighbors=5)	40	0.683	0.75	0.136
KNeighborsClassifier(n_neighbors=5)	20	0.762	0.769	0.455
KNeighborsClassifier(n_neighbors=5)	10	0.603	0.421	0.364
KNeighborsClassifier(n_neighbors=7)	199	0.683	0.6	0.273
KNeighborsClassifier(n_neighbors=7)	160	0.381	0.302	0.591
KNeighborsClassifier(n_neighbors=7)	130	0.651	0.0	0.0
KNeighborsClassifier(n_neighbors=7)	100	0.651	0.5	0.045
KNeighborsClassifier(n_neighbors=7)	70	0.683	1.0	0.091
KNeighborsClassifier(n_neighbors=7)	40	0.651	0.0	0.0
KNeighborsClassifier(n_neighbors=7)	20	0.714	0.667	0.364
KNeighborsClassifier(n_neighbors=7)	10	0.667	0.538	0.318
KNeighborsClassifier(n_neighbors=10)	199	0.698	0.714	0.227
KNeighborsClassifier(n_neighbors=10)	160	0.619	0.0	0.0
KNeighborsClassifier(n_neighbors=10)	130	0.651	0.0	0.0
KNeighborsClassifier(n_neighbors=10)	100	0.651	0.0	0.0
KNeighborsClassifier(n_neighbors=10)	70	0.667	1.0	0.045
KNeighborsClassifier(n_neighbors=10)	40	0.651	0.0	0.0
KNeighborsClassifier(n_neighbors=10)	20	0.698	0.8	0.182
KNeighborsClassifier(n_neighbors=10)	10	0.698	0.667	0.273
KNeighborsClassifier(n_neighbors=15)	199	0.683	0.625	0.227
KNeighborsClassifier(n_neighbors=15)	160	0.683	1.0	0.091
KNeighborsClassifier(n_neighbors=15)	130	0.651	0.0	0.0
KNeighborsClassifier(n_neighbors=15)	100	0.651	0.0	0.0
KNeighborsClassifier(n_neighbors=15)	70	0.651	0.0	0.0
KNeighborsClassifier(n_neighbors=15)	40	0.651	0.0	0.0
KNeighborsClassifier(n_neighbors=15)	20	0.698	0.8	0.182
KNeighborsClassifier(n_neighbors=15)	10	0.698	0.667	0.273
MLPClassifier	199	0.571	0.4	0.455
MLPClassifier	160	0.556	0.333	0.273

MLPClassifier	130	0.587	0.3	0.136
MLPClassifier	100	0.698	0.579	0.5
MLPClassifier	70	0.698	0.615	0.364
MLPClassifier	40	0.683	0.583	0.318
MLPClassifier	20	0.714	0.667	0.364
MLPClassifier	10	0.603	0.412	0.318
DecisionTreeClassifier	199	0.587	0.433	0.591
DecisionTreeClassifier	160	0.556	0.385	0.455
DecisionTreeClassifier	130	0.556	0.385	0.455
DecisionTreeClassifier	100	0.683	0.538	0.636
DecisionTreeClassifier	70	0.603	0.429	0.409
DecisionTreeClassifier	40	0.603	0.429	0.409
DecisionTreeClassifier	20	0.746	0.625	0.682
DecisionTreeClassifier	10	0.571	0.4	0.455
GaussianNB	199	0.397	0.34	0.773
GaussianNB	160	0.603	0.421	0.364
GaussianNB	130	0.603	0.421	0.364
GaussianNB	100	0.587	0.4	0.364
GaussianNB	70	0.603	0.421	0.364
GaussianNB	40	0.619	0.438	0.318
GaussianNB	20	0.635	0.467	0.318
GaussianNB	10	0.635	0.462	0.273

Table 7: Results for regression using RF feature reduction

Algorithm	No. of features	MAE	ME	RMSE	MedAE	ExV	R <sup>2</sup>
SVR(kernel='rbf')	199	360.18	5406.43	1003.77	49.08	0.0	-0.12
SVR(kernel='rbf')	160	360.11	5405.85	1003.61	49.28	0.0	-0.12
SVR(kernel='rbf')	130	360.07	5405.28	1003.59	49.65	0.0	-0.12
SVR(kernel='rbf')	100	360.02	5406.73	1003.86	49.94	0.0	-0.12
SVR(kernel='rbf')	70	359.59	5403.92	1003.07	49.08	0.0	-0.12
SVR(kernel='rbf')	40	359.68	5406.86	1003.3	50.28	0.0	-0.12
SVR(kernel='rbf')	20	360.09	5404.66	1002.94	49.48	0.0	-0.12
SVR(kernel='rbf')	10	359.93	5400.56	1002.3	46.98	0.0	-0.12
SVR(kernel='linear')	199	344.63	5022.1	918.51	83.49	0.12	0.06
SVR(kernel='linear')	160	339.28	5038.0	919.49	79.82	0.13	0.06
SVR(kernel='linear')	130	335.78	5026.59	921.32	74.6	0.13	0.05
SVR(kernel='linear')	100	333.87	5135.63	930.86	73.47	0.11	0.04
SVR(kernel='linear')	70	331.28	5160.94	929.3	76.91	0.12	0.04
SVR(kernel='linear')	40	345.95	5320.09	957.99	77.51	0.07	-0.02
SVR(kernel='linear')	20	353.22	5317.7	967.68	77.88	0.05	-0.04
SVR(kernel='linear')	10	360.45	5341.66	990.92	61.99	0.02	-0.09
SVR(kernel='poly', degree=3)	199	374.61	5409.61	1011.6	54.13	-0.03	-0.14
SVR(kernel='poly', degree=3)	160	378.78	5409.59	1016.58	54.01	-0.05	-0.15
SVR(kernel='poly', degree=3)	130	380.11	5409.38	1018.19	53.94	-0.05	-0.15
SVR(kernel='poly', degree=3)	100	358.99	5410.0	1001.96	53.72	-0.0	-0.12
SVR(kernel='poly', degree=3)	70	359.6	5409.33	999.27	53.75	0.01	-0.11
SVR(kernel='poly', degree=3)	40	354.18	5409.32	980.66	45.22	0.04	-0.07
SVR(kernel='poly', degree=3)	20	359.69	5405.55	995.26	49.23	0.01	-0.1
SVR(kernel='poly', degree=3)	10	360.2	5404.02	1003.06	49.19	0.0	-0.12
SVR(kernel='poly', degree=4)	199	480.7	7877.98	1410.3	54.47	-1.17	-1.21
SVR(kernel='poly', degree=4)	160	559.55	13060.4	1925.47	46.26	-3.12	-3.13
SVR(kernel='poly', degree=4)	130	606.09	16130.67	2263.57	46.29	-4.7	-4.71

SVR(kernel='poly', degree=4)	100	366.94	5409.39	1006.55	46.89	-0.03	-0.13
SVR(kernel='poly', degree=4)	70	366.39	5409.12	1003.34	53.86	-0.02	-0.12
SVR(kernel='poly', degree=4)	40	361.65	5409.52	959.45	49.42	0.07	-0.03
SVR(kernel='poly', degree=4)	20	371.32	5403.82	999.79	50.5	-0.0	-0.11
SVR(kernel='poly', degree=4)	10	368.53	5402.38	1003.89	52.45	0.01	-0.12
SVR(kernel='poly', degree=5)	199	1162.3	50927.22	6493.88	46.69	-45.69	-45.96
SVR(kernel='poly', degree=5)	160	1968.1	101472.42	12823.29	54.09	-180.21	-182.12
SVR(kernel='poly', degree=5)	130	2428.78	130882.33	16519.34	51.39	-299.4	-302.89
SVR(kernel='poly', degree=5)	100	429.63	5409.99	1150.45	53.88	-0.41	-0.47
SVR(kernel='poly', degree=5)	70	422.03	5411.58	1151.86	48.07	-0.42	-0.48
SVR(kernel='poly', degree=5)	40	374.24	5409.51	1055.59	49.18	-0.2	-0.24
SVR(kernel='poly', degree=5)	20	397.34	5402.09	1037.96	48.94	-0.1	-0.2
SVR(kernel='poly', degree=5)	10	375.29	5399.68	1006.12	52.99	0.01	-0.13
SVR(kernel='sigmoid')	199	360.08	5407.56	1003.72	49.44	0.0	-0.12
SVR(kernel='sigmoid')	160	359.97	5406.91	1003.4	49.53	0.0	-0.12
SVR(kernel='sigmoid')	130	359.85	5406.11	1003.19	49.49	0.0	-0.12
SVR(kernel='sigmoid')	100	359.81	5407.15	1003.17	49.58	0.0	-0.12
SVR(kernel='sigmoid')	70	359.55	5405.53	1002.55	49.33	0.0	-0.12
SVR(kernel='sigmoid')	40	359.66	5406.7	1002.42	49.69	0.0	-0.12
SVR(kernel='sigmoid')	20	359.81	5404.37	1002.01	49.63	0.0	-0.12
SVR(kernel='sigmoid')	10	360.4	5404.9	1003.69	49.94	0.0	-0.12
RandomForestRegressor(n_estimators=10)	199	495.55	3753.5	848.97	217.1	0.22	0.2
RandomForestRegressor(n_estimators=10)	160	458.74	3487.4	813.92	164.0	0.28	0.26
RandomForestRegressor(n_estimators=10)	130	463.41	3791.8	844.75	163.4	0.22	0.21
RandomForestRegressor(n_estimators=10)	100	451.13	3637.4	786.12	185.4	0.33	0.31
RandomForestRegressor(n_estimators=10)	70	464.39	2871.7	758.73	214.3	0.38	0.36
RandomForestRegressor(n_estimators=10)	40	478.83	4088.1	894.19	230.9	0.12	0.11
RandomForestRegressor(n_estimators=10)	20	509.25	4336.5	935.36	216.3	0.04	0.03
RandomForestRegressor(n_estimators=10)	10	511.53	3762.4	945.3	236.8	0.03	0.0
RandomForestRegressor(n_estimators=20)	199	481.32	3829.7	835.36	245.2	0.23	0.22



RandomForestRegressor(n_estimators=20)	160	461.99	3574.4	818.71	210.45	0.26	0.25
RandomForestRegressor(n_estimators=20)	130	443.08	3787.9	808.71	204.65	0.28	0.27
RandomForestRegressor(n_estimators=20)	100	459.5	3326.5	774.96	212.4	0.34	0.33
RandomForestRegressor(n_estimators=20)	70	442.71	3513.75	759.0	220.25	0.37	0.36
RandomForestRegressor(n_estimators=20)	40	459.82	4261.9	846.44	240.9	0.21	0.2
RandomForestRegressor(n_estimators=20)	20	462.51	4238.9	904.2	207.75	0.09	0.09
RandomForestRegressor(n_estimators=20)	10	475.74	3895.1	902.7	233.25	0.11	0.09
RandomForestRegressor(n_estimators=50)	199	451.38	3966.62	802.45	214.42	0.29	0.28
RandomForestRegressor(n_estimators=50)	160	462.85	4025.96	811.36	217.96	0.28	0.27
RandomForestRegressor(n_estimators=50)	130	470.9	4014.14	821.9	280.36	0.26	0.25
RandomForestRegressor(n_estimators=50)	100	479.41	3830.48	798.32	250.86	0.31	0.29
RandomForestRegressor(n_estimators=50)	70	459.77	4043.82	793.43	225.18	0.31	0.3
RandomForestRegressor(n_estimators=50)	40	426.31	4203.8	809.7	201.82	0.27	0.27
RandomForestRegressor(n_estimators=50)	20	435.35	4042.08	840.73	201.12	0.21	0.21
RandomForestRegressor(n_estimators=50)	10	442.5	3789.52	863.52	184.78	0.18	0.17
RandomForestRegressor(n_estimators=100)	199	477.35	3850.12	808.83	252.97	0.28	0.27
RandomForestRegressor(n_estimators=100)	160	487.45	4020.86	823.13	236.1	0.26	0.25
RandomForestRegressor(n_estimators=100)	130	487.47	4134.32	836.81	225.1	0.23	0.22
RandomForestRegressor(n_estimators=100)	100	485.08	3890.03	807.26	234.58	0.29	0.27
RandomForestRegressor(n_estimators=100)	70	472.95	3983.06	797.5	238.9	0.31	0.29
RandomForestRegressor(n_estimators=100)	40	440.94	4112.47	817.1	212.44	0.26	0.26
RandomForestRegressor(n_estimators=100)	20	439.84	3983.34	828.77	224.57	0.24	0.24
RandomForestRegressor(n_estimators=100)	10	459.84	3899.06	911.05	201.33	0.08	0.08
RandomForestRegressor(n_estimators=300)	199	490.41	3829.81	830.68	252.8	0.25	0.23
RandomForestRegressor(n_estimators=300)	160	479.16	3747.8	810.42	254.53	0.29	0.27
RandomForestRegressor(n_estimators=300)	130	486.86	3825.47	826.1	236.49	0.26	0.24
RandomForestRegressor(n_estimators=300)	100	487.28	3770.78	817.97	241.84	0.27	0.25
RandomForestRegressor(n_estimators=300)	70	472.14	3759.2	808.09	248.28	0.29	0.27
RandomForestRegressor(n_estimators=300)	40	436.71	3945.45	809.61	174.19	0.27	0.27
RandomForestRegressor(n_estimators=300)	20	446.18	3944.85	842.52	212.21	0.21	0.21

RandomForestRegressor(n_estimators=300)	10	471.44	3906.63	943.26	204.55	0.02	0.01
RandomForestRegressor(n_estimators=500)	199	485.61	3889.56	825.76	267.19	0.26	0.24
RandomForestRegressor(n_estimators=500)	160	476.3	3838.44	810.66	256.02	0.28	0.27
RandomForestRegressor(n_estimators=500)	130	483.36	3841.88	819.08	240.83	0.27	0.25
RandomForestRegressor(n_estimators=500)	100	481.08	3780.51	814.66	245.38	0.28	0.26
RandomForestRegressor(n_estimators=500)	70	469.09	3832.68	808.76	228.59	0.28	0.27
RandomForestRegressor(n_estimators=500)	40	429.95	3933.69	800.94	173.73	0.29	0.29
RandomForestRegressor(n_estimators=500)	20	446.59	3963.28	840.1	217.49	0.22	0.21
RandomForestRegressor(n_estimators=500)	10	465.07	3949.48	936.77	203.65	0.03	0.02
RandomForestRegressor(n_estimators=700)	199	484.65	3833.54	816.06	268.97	0.28	0.26
RandomForestRegressor(n_estimators=700)	160	479.99	3764.75	805.98	261.11	0.29	0.28
RandomForestRegressor(n_estimators=700)	130	482.06	3791.96	814.04	237.0	0.28	0.26
RandomForestRegressor(n_estimators=700)	100	480.53	3740.18	810.75	258.81	0.28	0.27
RandomForestRegressor(n_estimators=700)	70	469.49	3804.37	806.78	239.79	0.29	0.28
RandomForestRegressor(n_estimators=700)	40	427.12	3869.15	793.71	175.78	0.3	0.3
RandomForestRegressor(n_estimators=700)	20	441.83	3936.16	833.09	219.93	0.23	0.23
RandomForestRegressor(n_estimators=700)	10	466.36	3923.11	935.43	215.77	0.03	0.03
RandomForestRegressor(n_estimators=1000)	199	484.4	3882.25	817.33	277.72	0.27	0.26
RandomForestRegressor(n_estimators=1000)	160	476.33	3766.02	803.29	265.23	0.3	0.28
RandomForestRegressor(n_estimators=1000)	130	480.1	3785.97	812.44	237.82	0.28	0.26
RandomForestRegressor(n_estimators=1000)	100	476.63	3753.78	806.72	252.34	0.29	0.28
RandomForestRegressor(n_estimators=1000)	70	467.83	3812.04	806.83	231.84	0.29	0.28
RandomForestRegressor(n_estimators=1000)	40	428.83	3890.08	797.49	190.44	0.29	0.29
RandomForestRegressor(n_estimators=1000)	20	443.68	3931.27	834.15	220.05	0.23	0.23
RandomForestRegressor(n_estimators=1000)	10	463.95	3919.78	938.16	214.06	0.03	0.02
LinearRegression()	199	9136.41	74629.68	16713.1	4326.54	-307.46	-310.06
LinearRegression()	160	3296.92	32807.1	6108.53	1401.8	-40.47	-40.55
LinearRegression()	130	1843.18	9780.52	2907.8	1129.39	-8.18	-8.42
LinearRegression()	100	1382.45	11391.77	2374.56	705.41	-5.11	-5.28
LinearRegression()	70	821.99	6507.46	1313.71	485.91	-0.9	-0.92

LinearRegression()	40	617.5	6227.38	1130.31	364.66	-0.42	-0.42
LinearRegression()	20	499.55	4565.46	949.24	254.94	0.01	-0.0
LinearRegression()	10	517.2	4025.78	856.85	309.04	0.19	0.18
PolynomialRegression(degree=3)	199	1549.0	25520.16	3837.1	675.55	-15.36	-15.4
PolynomialRegression(degree=3)	160	1681.93	28943.72	4365.52	520.8	-20.22	-20.22
PolynomialRegression(degree=3)	130	1780.07	31056.91	4738.33	512.8	-24.0	-24.0
PolynomialRegression(degree=3)	100	1886.92	23855.95	4528.37	598.49	-21.45	-21.84
PolynomialRegression(degree=3)	70	2546.62	40686.51	6404.26	887.79	-44.34	-44.67
PolynomialRegression(degree=3)	40	3012.93	48786.88	8928.27	736.73	-87.27	-87.77
PolynomialRegression(degree=3)	20	5456.96	95968.77	16436.3	718.62	-299.69	-299.85
PolynomialRegression(degree=3)	10	3.87e+12	1.88e+14	2.47e+13	2999.06	-6.7e+20	-6.8e+20

Table 8: Results for regression with Maven Reuse outlier cut-off at >400 using RF feature reduction

Algorithm	No. of features	MAE	ME	RMSE	MedAE	ExV	R <sup>2</sup>
SVR(kernel='rbf')	198	61.71	348.72	105.27	27.9	-0.0	-0.14
SVR(kernel='rbf')	160	61.8	348.38	105.23	27.56	-0.0	-0.14
SVR(kernel='rbf')	130	61.8	348.49	105.24	27.9	-0.0	-0.14
SVR(kernel='rbf')	100	61.77	348.13	105.2	27.23	-0.0	-0.14
SVR(kernel='rbf')	70	61.78	348.47	105.18	27.02	-0.0	-0.14
SVR(kernel='rbf')	40	61.61	348.99	105.09	27.45	-0.0	-0.14
SVR(kernel='rbf')	20	61.88	349.22	104.99	27.83	-0.0	-0.14
SVR(kernel='rbf')	10	61.81	346.37	104.47	31.74	-0.0	-0.13
SVR(kernel='linear')	198	62.64	337.7	105.0	25.94	-0.04	-0.14
SVR(kernel='linear')	160	67.35	335.22	108.19	29.3	-0.14	-0.21
SVR(kernel='linear')	130	64.33	333.25	104.39	28.27	-0.04	-0.12
SVR(kernel='linear')	100	62.92	325.51	104.16	25.9	-0.02	-0.12
SVR(kernel='linear')	70	64.27	333.61	103.94	29.8	-0.02	-0.11
SVR(kernel='linear')	40	62.65	346.2	105.68	27.92	-0.05	-0.15

SVR(kernel='linear')	20	66.03	346.12	108.46	29.9	-0.08	-0.21
SVR(kernel='linear')	10	62.55	342.33	104.78	28.32	-0.01	-0.13
SVR(kernel='poly', degree=3)	198	60.05	347.94	104.26	25.09	0.0	-0.12
SVR(kernel='poly', degree=3)	160	61.4	348.06	104.89	28.0	-0.02	-0.14
SVR(kernel='poly', degree=3)	130	60.96	347.97	104.56	27.92	0.01	-0.13
SVR(kernel='poly', degree=3)	100	61.45	347.98	104.79	27.78	-0.0	-0.13
SVR(kernel='poly', degree=3)	70	62.28	348.26	105.29	28.17	-0.02	-0.14
SVR(kernel='poly', degree=3)	40	67.11	348.91	111.92	28.82	-0.19	-0.29
SVR(kernel='poly', degree=3)	20	62.6	348.66	106.1	28.19	-0.01	-0.16
SVR(kernel='poly', degree=3)	10	63.23	348.49	106.3	29.11	-0.02	-0.17
SVR(kernel='poly', degree=4)	198	64.27	347.53	107.8	27.55	-0.11	-0.2
SVR(kernel='poly', degree=4)	160	69.35	425.99	120.95	29.07	-0.45	-0.51
SVR(kernel='poly', degree=4)	130	62.12	348.03	104.87	28.03	-0.02	-0.13
SVR(kernel='poly', degree=4)	100	64.29	348.13	107.56	28.11	-0.09	-0.19
SVR(kernel='poly', degree=4)	70	73.18	559.04	133.11	28.38	-0.78	-0.83
SVR(kernel='poly', degree=4)	40	107.37	2185.3	335.48	29.07	-10.6	-10.61
SVR(kernel='poly', degree=4)	20	61.88	348.93	105.15	28.91	-0.0	-0.14
SVR(kernel='poly', degree=4)	10	62.62	348.9	105.86	28.87	-0.01	-0.16
SVR(kernel='poly', degree=5)	198	76.28	770.06	152.91	29.04	-1.38	-1.41
SVR(kernel='poly', degree=5)	160	100.47	1888.52	294.52	28.22	-7.95	-7.95
SVR(kernel='poly', degree=5)	130	63.92	348.71	106.72	28.25	-0.05	-0.18
SVR(kernel='poly', degree=5)	100	67.98	348.33	113.49	28.32	-0.22	-0.33
SVR(kernel='poly', degree=5)	70	80.11	884.44	166.59	28.07	-1.84	-1.86
SVR(kernel='poly', degree=5)	40	154.31	3975.4	592.86	29.03	-35.11	-35.27
SVR(kernel='poly', degree=5)	20	66.06	349.08	111.65	28.91	-0.11	-0.29
SVR(kernel='poly', degree=5)	10	64.25	348.1	108.35	28.08	-0.05	-0.21
SVR(kernel='sigmoid')	198	62.05	346.99	105.01	28.15	-0.0	-0.14
SVR(kernel='sigmoid')	160	62.0	346.74	104.89	28.91	-0.0	-0.14
SVR(kernel='sigmoid')	130	62.1	346.87	104.94	28.79	-0.0	-0.14
SVR(kernel='sigmoid')	100	62.13	346.71	104.97	29.74	-0.0	-0.14

SVR(kernel='sigmoid')	70	62.09	346.98	104.95	30.91	-0.0	-0.14
SVR(kernel='sigmoid')	40	62.38	346.55	104.96	31.98	-0.0	-0.14
SVR(kernel='sigmoid')	20	62.35	345.33	104.97	31.81	-0.01	-0.14
SVR(kernel='sigmoid')	10	61.6	344.34	104.69	30.33	0.0	-0.13
RandomForestRegressor(n_estimators=10)	198	73.52	324.5	108.04	44.8	-0.19	-0.2
RandomForestRegressor(n_estimators=10)	160	72.87	332.3	109.65	43.2	-0.23	-0.24
RandomForestRegressor(n_estimators=10)	130	71.46	299.1	108.42	40.9	-0.2	-0.21
RandomForestRegressor(n_estimators=10)	100	77.09	321.1	115.1	43.0	-0.36	-0.37
RandomForestRegressor(n_estimators=10)	70	77.45	356.2	115.91	42.9	-0.38	-0.39
RandomForestRegressor(n_estimators=10)	40	79.43	344.4	116.56	46.0	-0.38	-0.4
RandomForestRegressor(n_estimators=10)	20	79.66	288.8	112.13	58.4	-0.3	-0.3
RandomForestRegressor(n_estimators=10)	10	81.71	274.6	111.94	52.0	-0.29	-0.29
RandomForestRegressor(n_estimators=20)	198	74.09	314.1	105.64	44.9	-0.15	-0.15
RandomForestRegressor(n_estimators=20)	160	74.37	311.9	106.45	48.0	-0.17	-0.17
RandomForestRegressor(n_estimators=20)	130	74.38	309.35	108.56	41.2	-0.21	-0.22
RandomForestRegressor(n_estimators=20)	100	76.89	318.25	112.29	45.75	-0.3	-0.3
RandomForestRegressor(n_estimators=20)	70	75.0	315.85	109.22	54.35	-0.23	-0.23
RandomForestRegressor(n_estimators=20)	40	77.15	316.4	110.99	48.4	-0.27	-0.27
RandomForestRegressor(n_estimators=20)	20	77.12	291.15	107.32	53.25	-0.19	-0.19
RandomForestRegressor(n_estimators=20)	10	80.09	290.15	110.45	56.1	-0.25	-0.26
RandomForestRegressor(n_estimators=50)	198	76.18	295.1	107.21	44.88	-0.19	-0.19
RandomForestRegressor(n_estimators=50)	160	75.29	292.36	107.07	50.74	-0.18	-0.18
RandomForestRegressor(n_estimators=50)	130	76.86	301.1	108.63	53.86	-0.22	-0.22
RandomForestRegressor(n_estimators=50)	100	77.62	303.46	109.82	49.24	-0.24	-0.24
RandomForestRegressor(n_estimators=50)	70	77.06	314.46	109.25	51.12	-0.23	-0.23
RandomForestRegressor(n_estimators=50)	40	77.1	311.88	109.61	48.88	-0.24	-0.24
RandomForestRegressor(n_estimators=50)	20	77.89	298.94	108.5	54.32	-0.21	-0.21
RandomForestRegressor(n_estimators=50)	10	79.34	294.54	109.17	57.64	-0.22	-0.23
RandomForestRegressor(n_estimators=100)	198	74.58	288.72	105.4	49.59	-0.15	-0.15
RandomForestRegressor(n_estimators=100)	160	74.66	285.51	105.8	51.19	-0.15	-0.15

RandomForestRegressor(n_estimators=100)	130	75.51	290.42	106.66	50.3	-0.17	-0.17
RandomForestRegressor(n_estimators=100)	100	77.56	295.34	108.82	51.49	-0.22	-0.22
RandomForestRegressor(n_estimators=100)	70	76.14	300.97	108.11	51.14	-0.21	-0.21
RandomForestRegressor(n_estimators=100)	40	75.86	303.18	107.93	48.81	-0.2	-0.2
RandomForestRegressor(n_estimators=100)	20	77.17	296.92	108.72	54.04	-0.22	-0.22
RandomForestRegressor(n_estimators=100)	10	78.25	296.87	108.84	55.62	-0.22	-0.22
RandomForestRegressor(n_estimators=300)	198	75.19	293.35	105.04	47.57	-0.14	-0.14
RandomForestRegressor(n_estimators=300)	160	75.81	292.32	105.47	48.27	-0.15	-0.15
RandomForestRegressor(n_estimators=300)	130	75.69	295.02	105.18	51.04	-0.14	-0.14
RandomForestRegressor(n_estimators=300)	100	77.66	293.32	107.36	54.18	-0.19	-0.19
RandomForestRegressor(n_estimators=300)	70	76.63	298.72	106.76	50.92	-0.18	-0.18
RandomForestRegressor(n_estimators=300)	40	77.24	299.2	107.66	52.32	-0.2	-0.2
RandomForestRegressor(n_estimators=300)	20	76.84	297.48	108.0	53.16	-0.2	-0.2
RandomForestRegressor(n_estimators=300)	10	78.82	295.76	108.92	58.9	-0.22	-0.22
RandomForestRegressor(n_estimators=500)	198	75.24	294.19	105.45	48.92	-0.15	-0.15
RandomForestRegressor(n_estimators=500)	160	75.29	293.3	105.47	48.52	-0.15	-0.15
RandomForestRegressor(n_estimators=500)	130	75.42	295.81	105.44	51.78	-0.15	-0.15
RandomForestRegressor(n_estimators=500)	100	76.53	295.25	106.86	53.02	-0.18	-0.18
RandomForestRegressor(n_estimators=500)	70	76.35	299.26	106.89	51.72	-0.18	-0.18
RandomForestRegressor(n_estimators=500)	40	77.26	300.2	107.97	51.36	-0.2	-0.2
RandomForestRegressor(n_estimators=500)	20	77.2	295.94	108.33	50.96	-0.21	-0.21
RandomForestRegressor(n_estimators=500)	10	78.47	296.92	108.96	58.08	-0.22	-0.23
RandomForestRegressor(n_estimators=700)	198	74.84	294.32	104.98	47.36	-0.14	-0.14
RandomForestRegressor(n_estimators=700)	160	74.8	294.32	104.97	47.74	-0.14	-0.14
RandomForestRegressor(n_estimators=700)	130	75.05	296.45	105.24	50.23	-0.14	-0.14
RandomForestRegressor(n_estimators=700)	100	75.99	295.05	106.12	51.83	-0.16	-0.16
RandomForestRegressor(n_estimators=700)	70	76.13	299.92	106.57	51.51	-0.17	-0.17
RandomForestRegressor(n_estimators=700)	40	76.72	300.7	107.38	52.15	-0.19	-0.19
RandomForestRegressor(n_estimators=700)	20	76.38	295.92	107.69	48.95	-0.2	-0.2
RandomForestRegressor(n_estimators=700)	10	78.35	296.96	108.96	58.55	-0.22	-0.23

RandomForestRegressor(n_estimators=1000)	198	74.77	295.0	105.11	46.54	-0.14	-0.14
RandomForestRegressor(n_estimators=1000)	160	74.97	294.55	105.07	46.22	-0.14	-0.14
RandomForestRegressor(n_estimators=1000)	130	75.11	295.3	105.43	49.18	-0.15	-0.15
RandomForestRegressor(n_estimators=1000)	100	76.05	295.78	106.29	51.72	-0.17	-0.17
RandomForestRegressor(n_estimators=1000)	70	75.99	300.79	106.56	50.14	-0.17	-0.17
RandomForestRegressor(n_estimators=1000)	40	76.7	299.64	107.33	50.84	-0.19	-0.19
RandomForestRegressor(n_estimators=1000)	20	76.34	297.55	107.44	49.21	-0.19	-0.19
RandomForestRegressor(n_estimators=1000)	10	78.19	296.44	108.56	57.68	-0.21	-0.22
LinearRegression()	198	290.39	1771.48	422.62	196.29	-16.13	-17.43
LinearRegression()	160	513.61	3397.3	812.25	300.5	-66.22	-67.08
LinearRegression()	130	498.21	6788.57	1139.45	221.63	-126.93	-132.97
LinearRegression()	100	280.81	3195.95	684.71	101.22	-43.92	-47.38
LinearRegression()	70	123.03	1078.22	204.74	86.88	-3.22	-3.33
LinearRegression()	40	84.42	458.08	129.27	49.42	-0.71	-0.72
LinearRegression()	20	73.3	319.36	102.81	50.24	-0.09	-0.09
LinearRegression()	10	78.2	310.37	105.62	53.21	-0.15	-0.15
PolynomialRegression(degree=3)	198	133.93	750.31	209.08	67.22	-3.42	-3.51
PolynomialRegression(degree=3)	160	170.61	895.01	278.0	77.57	-6.69	-6.97
PolynomialRegression(degree=3)	130	186.14	1078.57	312.24	103.55	-9.02	-9.06
PolynomialRegression(degree=3)	100	160.19	1041.34	291.72	68.1	-7.66	-7.78
PolynomialRegression(degree=3)	70	174.52	2512.46	417.02	60.25	-16.76	-16.95
PolynomialRegression(degree=3)	40	310.66	2781.83	564.41	142.06	-30.91	-31.87
PolynomialRegression(degree=3)	20	415.07	3222.01	831.7	153.11	-70.22	-70.38
PolynomialRegression(degree=3)	10	3050.5	117431.1	17178.84	198.51	-29642	-30451

Table 9: Results for regression with Maven Reuse outlier cut-off at >200 using RF feature reduction

Algorithm	No. of features	MAE	ME	RMSE	MedAE	ExV	R <sup>2</sup>
SVR(kernel='rbf')	199	35.28	151.54	52.98	21.22	-0.0	-0.18

SVR(kernel='rbf')	160	35.3	151.79	52.92	21.51	-0.01	-0.18
SVR(kernel='rbf')	130	35.31	151.96	52.96	21.28	-0.01	-0.18
SVR(kernel='rbf')	100	35.27	152.23	53.01	20.88	-0.01	-0.18
SVR(kernel='rbf')	70	35.35	153.19	53.32	20.6	-0.01	-0.2
SVR(kernel='rbf')	40	35.36	153.14	53.36	20.7	-0.01	-0.2
SVR(kernel='rbf')	20	35.04	152.42	52.83	20.94	0.01	-0.18
SVR(kernel='rbf')	10	34.87	152.0	52.59	21.0	0.02	-0.16
SVR(kernel='linear')	199	38.8	224.44	63.51	19.37	-0.63	-0.7
SVR(kernel='linear')	160	42.56	205.14	67.17	18.99	-0.87	-0.9
SVR(kernel='linear')	130	44.97	229.32	70.91	22.64	-1.07	-1.12
SVR(kernel='linear')	100	37.98	203.37	61.23	20.81	-0.46	-0.58
SVR(kernel='linear')	70	39.47	207.27	63.73	23.05	-0.59	-0.71
SVR(kernel='linear')	40	33.94	188.44	53.99	17.89	-0.01	-0.23
SVR(kernel='linear')	20	32.85	163.34	50.08	21.17	0.02	-0.06
SVR(kernel='linear')	10	33.56	149.31	50.71	19.79	0.08	-0.08
SVR(kernel='poly', degree=3)	199	39.17	151.87	56.64	23.44	-0.19	-0.35
SVR(kernel='poly', degree=3)	160	40.36	157.75	59.03	23.22	-0.33	-0.47
SVR(kernel='poly', degree=3)	130	41.64	187.97	61.75	24.75	-0.48	-0.61
SVR(kernel='poly', degree=3)	100	36.58	150.05	52.83	23.09	0.02	-0.18
SVR(kernel='poly', degree=3)	70	34.19	150.53	51.48	21.41	0.05	-0.12
SVR(kernel='poly', degree=3)	40	36.0	150.9	53.02	21.36	-0.04	-0.18
SVR(kernel='poly', degree=3)	20	56.02	935.32	156.2	20.52	-9.27	-9.28
SVR(kernel='poly', degree=3)	10	33.79	149.84	51.51	20.52	0.04	-0.12
SVR(kernel='poly', degree=4)	199	46.68	238.07	74.95	22.01	-1.18	-1.37
SVR(kernel='poly', degree=4)	160	49.28	346.81	82.81	23.57	-1.76	-1.89
SVR(kernel='poly', degree=4)	130	52.49	436.28	94.39	23.12	-2.65	-2.75
SVR(kernel='poly', degree=4)	100	40.36	149.91	58.14	23.35	-0.18	-0.42
SVR(kernel='poly', degree=4)	70	40.81	160.98	60.68	21.99	-0.4	-0.55
SVR(kernel='poly', degree=4)	40	46.7	406.85	84.08	22.0	-1.91	-1.98
SVR(kernel='poly', degree=4)	20	143.19	4390.93	696.11	22.02	-199.68	-203.12



SVR(kernel='poly', degree=4)	10	37.51	203.12	59.49	20.66	-0.27	-0.49
SVR(kernel='poly', degree=5)	199	60.13	657.09	126.82	23.92	-5.35	-5.77
SVR(kernel='poly', degree=5)	160	67.81	673.09	145.06	23.11	-7.7	-7.86
SVR(kernel='poly', degree=5)	130	78.6	973.42	191.29	23.01	-14.32	-14.41
SVR(kernel='poly', degree=5)	100	45.21	224.99	67.66	23.0	-0.59	-0.93
SVR(kernel='poly', degree=5)	70	54.83	359.1	93.88	23.0	-2.4	-2.71
SVR(kernel='poly', degree=5)	40	73.22	1065.05	187.82	23.0	-13.85	-13.86
SVR(kernel='poly', degree=5)	20	449.55	16485.03	2607.14	22.07	-2798.43	-2862.17
SVR(kernel='poly', degree=5)	10	33.08	149.12	50.44	20.1	0.07	-0.07
SVR(kernel='sigmoid')	199	35.6	151.72	53.11	23.36	-0.0	-0.19
SVR(kernel='sigmoid')	160	35.66	151.49	53.18	23.17	-0.0	-0.19
SVR(kernel='sigmoid')	130	35.72	151.69	53.23	23.21	-0.01	-0.19
SVR(kernel='sigmoid')	100	35.62	151.96	53.22	23.15	-0.01	-0.19
SVR(kernel='sigmoid')	70	35.62	152.47	53.24	23.36	-0.01	-0.19
SVR(kernel='sigmoid')	40	35.51	153.4	53.29	22.76	-0.01	-0.2
SVR(kernel='sigmoid')	20	35.42	152.8	52.81	22.73	-0.01	-0.17
SVR(kernel='sigmoid')	10	35.32	152.08	52.92	21.95	0.01	-0.18
RandomForestRegressor(n_estimators=10)	199	31.63	134.6	46.8	18.55	0.12	0.08
RandomForestRegressor(n_estimators=10)	160	30.46	134.0	45.27	19.5	0.18	0.14
RandomForestRegressor(n_estimators=10)	130	32.37	139.6	47.99	20.25	0.07	0.03
RandomForestRegressor(n_estimators=10)	100	32.09	128.0	45.91	22.5	0.13	0.11
RandomForestRegressor(n_estimators=10)	70	33.34	150.6	49.75	19.75	-0.01	-0.04
RandomForestRegressor(n_estimators=10)	40	32.79	147.3	48.69	20.65	0.05	0.0
RandomForestRegressor(n_estimators=10)	20	30.58	163.0	45.37	17.85	0.15	0.13
RandomForestRegressor(n_estimators=10)	10	34.27	167.4	51.59	19.05	-0.09	-0.12
RandomForestRegressor(n_estimators=20)	199	30.12	141.65	44.8	17.88	0.18	0.15
RandomForestRegressor(n_estimators=20)	160	29.34	139.65	43.35	20.8	0.25	0.21
RandomForestRegressor(n_estimators=20)	130	30.89	147.55	45.54	20.55	0.15	0.13
RandomForestRegressor(n_estimators=20)	100	29.56	136.15	43.73	21.28	0.22	0.19
RandomForestRegressor(n_estimators=20)	70	29.33	149.15	43.78	16.37	0.21	0.19

RandomForestRegressor(n_estimators=20)	40	29.64	151.35	45.01	18.32	0.18	0.15
RandomForestRegressor(n_estimators=20)	20	29.61	164.15	44.69	16.53	0.19	0.16
RandomForestRegressor(n_estimators=20)	10	31.1	161.8	46.96	15.58	0.1	0.07
RandomForestRegressor(n_estimators=50)	199	29.97	149.26	45.76	20.75	0.15	0.12
RandomForestRegressor(n_estimators=50)	160	28.58	147.36	43.34	16.57	0.25	0.21
RandomForestRegressor(n_estimators=50)	130	29.87	147.2	44.69	18.49	0.19	0.16
RandomForestRegressor(n_estimators=50)	100	28.58	136.5	43.45	17.07	0.24	0.2
RandomForestRegressor(n_estimators=50)	70	29.05	146.64	43.15	17.74	0.24	0.22
RandomForestRegressor(n_estimators=50)	40	29.69	144.54	44.96	17.36	0.19	0.15
RandomForestRegressor(n_estimators=50)	20	30.48	153.04	44.92	17.36	0.19	0.15
RandomForestRegressor(n_estimators=50)	10	31.68	154.0	46.19	19.37	0.14	0.1
RandomForestRegressor(n_estimators=100)	199	29.97	145.84	45.25	16.17	0.17	0.14
RandomForestRegressor(n_estimators=100)	160	29.48	144.33	44.37	18.4	0.21	0.17
RandomForestRegressor(n_estimators=100)	130	29.76	142.4	44.79	19.26	0.19	0.16
RandomForestRegressor(n_estimators=100)	100	29.67	142.72	44.58	17.17	0.2	0.16
RandomForestRegressor(n_estimators=100)	70	29.52	142.44	44.14	17.98	0.21	0.18
RandomForestRegressor(n_estimators=100)	40	29.02	144.18	44.37	17.98	0.21	0.17
RandomForestRegressor(n_estimators=100)	20	30.29	147.7	44.36	16.92	0.21	0.17
RandomForestRegressor(n_estimators=100)	10	31.74	151.41	45.27	19.21	0.18	0.14
RandomForestRegressor(n_estimators=300)	199	30.18	147.41	45.27	16.38	0.17	0.14
RandomForestRegressor(n_estimators=300)	160	29.84	147.1	44.76	16.73	0.19	0.16
RandomForestRegressor(n_estimators=300)	130	29.99	143.75	45.15	17.73	0.18	0.14
RandomForestRegressor(n_estimators=300)	100	29.47	144.49	44.56	15.87	0.2	0.16
RandomForestRegressor(n_estimators=300)	70	29.34	144.11	44.19	15.54	0.21	0.18
RandomForestRegressor(n_estimators=300)	40	28.54	142.23	43.82	16.18	0.23	0.19
RandomForestRegressor(n_estimators=300)	20	30.35	149.64	44.2	17.06	0.22	0.18
RandomForestRegressor(n_estimators=300)	10	32.01	149.32	45.41	19.27	0.17	0.13
RandomForestRegressor(n_estimators=500)	199	30.24	148.02	45.39	17.68	0.17	0.13
RandomForestRegressor(n_estimators=500)	160	30.03	148.18	45.18	17.58	0.18	0.14
RandomForestRegressor(n_estimators=500)	130	30.02	145.87	45.29	17.12	0.17	0.14

RandomForestRegressor(n_estimators=500)	100	29.54	146.1	44.91	15.6	0.19	0.15
RandomForestRegressor(n_estimators=500)	70	29.18	146.02	44.42	15.61	0.2	0.17
RandomForestRegressor(n_estimators=500)	40	28.52	143.07	43.87	16.5	0.23	0.19
RandomForestRegressor(n_estimators=500)	20	30.0	148.66	43.86	16.62	0.24	0.19
RandomForestRegressor(n_estimators=500)	10	32.17	147.06	45.49	19.62	0.17	0.13
RandomForestRegressor(n_estimators=700)	199	30.19	147.98	45.35	17.86	0.17	0.13
RandomForestRegressor(n_estimators=700)	160	30.02	148.52	45.11	18.03	0.18	0.14
RandomForestRegressor(n_estimators=700)	130	30.05	147.06	45.32	17.49	0.17	0.13
RandomForestRegressor(n_estimators=700)	100	29.49	146.26	44.79	16.25	0.19	0.15
RandomForestRegressor(n_estimators=700)	70	29.31	146.14	44.42	16.6	0.2	0.17
RandomForestRegressor(n_estimators=700)	40	28.59	143.25	43.89	17.45	0.23	0.19
RandomForestRegressor(n_estimators=700)	20	29.93	149.48	43.96	16.64	0.23	0.19
RandomForestRegressor(n_estimators=700)	10	32.15	147.8	45.32	19.4	0.18	0.13
RandomForestRegressor(n_estimators=1000)	199	30.12	147.59	45.23	17.88	0.18	0.14
RandomForestRegressor(n_estimators=1000)	160	30.1	147.61	45.19	18.23	0.18	0.14
RandomForestRegressor(n_estimators=1000)	130	30.06	146.73	45.31	18.11	0.17	0.14
RandomForestRegressor(n_estimators=1000)	100	29.59	145.49	44.8	16.91	0.19	0.15
RandomForestRegressor(n_estimators=1000)	70	29.54	145.87	44.58	17.56	0.2	0.16
RandomForestRegressor(n_estimators=1000)	40	28.66	142.73	43.89	17.35	0.22	0.19
RandomForestRegressor(n_estimators=1000)	20	29.95	149.9	44.02	16.96	0.23	0.18
RandomForestRegressor(n_estimators=1000)	10	32.19	147.71	45.41	19.78	0.17	0.13
LinearRegression()	199	155.38	1101.21	279.6	68.68	-30.7	-31.93
LinearRegression()	160	162.66	1624.39	317.94	77.18	-40.5	-41.58
LinearRegression()	130	257.6	2433.26	514.36	101.95	-109.54	-110.44
LinearRegression()	100	124.02	717.86	177.69	77.18	-11.95	-12.3
LinearRegression()	70	77.81	364.38	113.77	56.63	-4.41	-4.45
LinearRegression()	40	53.46	240.14	83.84	26.51	-1.9	-1.96
LinearRegression()	20	34.09	182.37	55.36	18.91	-0.27	-0.29
LinearRegression()	10	30.56	146.22	48.62	16.64	0.06	0.0
PolynomialRegression(degree=3)	199	120.35	801.63	197.6	58.46	-14.53	-15.45

PolynomialRegression(degree=3)	160	122.15	570.04	177.48	79.27	-11.71	-12.27
PolynomialRegression(degree=3)	130	112.87	634.99	179.02	60.13	-12.14	-12.5
PolynomialRegression(degree=3)	100	99.49	520.22	159.68	43.45	-9.46	-9.74
PolynomialRegression(degree=3)	70	117.74	805.81	188.51	63.57	-13.85	-13.97
PolynomialRegression(degree=3)	40	116.79	1512.31	270.49	43.57	-29.63	-29.82
PolynomialRegression(degree=3)	20	417.95	5459.88	1132.71	65.44	-528.73	-539.45
PolynomialRegression(degree=3)	10	754.34	9336.27	1993.91	147.18	-1669.57	-1673.67

Table 10: Results for regression with Maven Reuse outlier cut-off at >100 using RF feature reduction

Algorithm	No. of features	MAE	ME	RMSE	MedAE	ExV	R <sup>2</sup>
SVR(kernel='rbf')	197	17.37	46.66	20.68	17.0	0.01	-0.01
SVR(kernel='rbf')	160	17.38	46.89	20.69	16.83	0.01	-0.01
SVR(kernel='rbf')	130	17.4	47.1	20.68	16.72	0.01	-0.01
SVR(kernel='rbf')	100	17.41	47.04	20.68	16.77	0.01	-0.01
SVR(kernel='rbf')	70	17.44	47.43	20.69	16.65	0.01	-0.01
SVR(kernel='rbf')	40	17.4	47.05	20.67	17.03	0.01	-0.01
SVR(kernel='rbf')	20	17.33	45.93	20.58	16.83	0.02	0.0
SVR(kernel='rbf')	10	17.23	47.22	20.47	16.38	0.03	0.01
SVR(kernel='linear')	197	24.26	66.77	31.18	20.94	-1.28	-1.29
SVR(kernel='linear')	160	24.84	78.17	31.91	23.64	-1.39	-1.4
SVR(kernel='linear')	130	21.89	68.53	28.35	20.41	-0.89	-0.89
SVR(kernel='linear')	100	22.16	59.46	27.47	19.84	-0.78	-0.78
SVR(kernel='linear')	70	18.95	56.17	24.13	16.48	-0.34	-0.37
SVR(kernel='linear')	40	18.65	63.18	23.29	14.77	-0.26	-0.28
SVR(kernel='linear')	20	16.97	57.66	20.84	14.16	0.0	-0.02
SVR(kernel='linear')	10	17.41	62.5	21.49	14.13	-0.07	-0.09
SVR(kernel='poly', degree=3)	197	17.6	48.88	20.86	16.2	0.02	-0.02
SVR(kernel='poly', degree=3)	160	17.43	48.88	20.7	15.98	0.03	-0.01

SVR(kernel='poly', degree=3)	130	17.48	49.25	20.75	16.19	0.03	-0.01
SVR(kernel='poly', degree=3)	100	17.52	49.02	20.82	15.98	0.01	-0.02
SVR(kernel='poly', degree=3)	70	17.43	49.39	20.71	16.04	0.03	-0.01
SVR(kernel='poly', degree=3)	40	17.67	49.92	20.92	16.58	0.0	-0.03
SVR(kernel='poly', degree=3)	20	18.09	54.58	21.31	15.82	-0.04	-0.07
SVR(kernel='poly', degree=3)	10	16.73	66.26	21.28	16.08	-0.02	-0.07
SVR(kernel='poly', degree=4)	197	17.94	49.94	21.25	16.49	0.02	-0.06
SVR(kernel='poly', degree=4)	160	17.76	50.19	21.17	16.49	0.03	-0.06
SVR(kernel='poly', degree=4)	130	17.8	50.25	21.15	16.52	0.03	-0.05
SVR(kernel='poly', degree=4)	100	17.81	50.05	21.2	16.48	0.01	-0.06
SVR(kernel='poly', degree=4)	70	17.64	49.69	20.95	16.59	0.04	-0.03
SVR(kernel='poly', degree=4)	40	17.67	48.99	20.83	17.08	0.02	-0.02
SVR(kernel='poly', degree=4)	20	19.43	83.62	25.1	16.02	-0.47	-0.48
SVR(kernel='poly', degree=4)	10	16.67	58.89	20.68	16.14	0.04	-0.01
SVR(kernel='poly', degree=5)	197	17.77	50.12	21.05	16.5	0.04	-0.04
SVR(kernel='poly', degree=5)	160	17.48	50.21	20.89	14.82	0.07	-0.03
SVR(kernel='poly', degree=5)	130	17.7	50.38	21.06	16.54	0.05	-0.04
SVR(kernel='poly', degree=5)	100	17.82	50.21	21.33	16.5	0.01	-0.07
SVR(kernel='poly', degree=5)	70	17.73	49.99	21.04	16.58	0.03	-0.04
SVR(kernel='poly', degree=5)	40	17.8	50.26	20.96	17.54	0.03	-0.03
SVR(kernel='poly', degree=5)	20	24.66	248.91	47.74	15.97	-4.35	-4.37
SVR(kernel='poly', degree=5)	10	17.21	75.09	22.29	15.97	-0.12	-0.17
SVR(kernel='sigmoid')	197	17.4	49.17	20.75	16.35	-0.0	-0.01
SVR(kernel='sigmoid')	160	17.39	49.15	20.73	16.34	0.0	-0.01
SVR(kernel='sigmoid')	130	17.36	48.81	20.67	16.51	0.01	-0.01
SVR(kernel='sigmoid')	100	17.42	48.81	20.72	16.56	0.0	-0.01
SVR(kernel='sigmoid')	70	17.5	49.08	20.72	16.86	0.0	-0.01
SVR(kernel='sigmoid')	40	17.45	48.3	20.75	16.89	-0.0	-0.01
SVR(kernel='sigmoid')	20	17.4	50.71	20.69	16.12	0.01	-0.01
SVR(kernel='sigmoid')	10	17.4	51.26	20.77	16.52	-0.0	-0.02

RandomForestRegressor(n_estimators=10)	197	19.18	48.4	23.35	17.6	-0.18	-0.28
RandomForestRegressor(n_estimators=10)	160	19.19	43.3	22.33	17.95	-0.1	-0.17
RandomForestRegressor(n_estimators=10)	130	19.33	43.6	22.7	16.0	-0.16	-0.21
RandomForestRegressor(n_estimators=10)	100	18.0	39.1	21.4	16.25	-0.04	-0.08
RandomForestRegressor(n_estimators=10)	70	16.72	41.5	20.2	14.55	0.05	0.04
RandomForestRegressor(n_estimators=10)	40	18.09	42.3	21.55	16.95	-0.03	-0.09
RandomForestRegressor(n_estimators=10)	20	18.23	50.6	22.41	16.65	-0.17	-0.18
RandomForestRegressor(n_estimators=10)	10	17.03	54.5	21.44	13.9	-0.05	-0.08
RandomForestRegressor(n_estimators=20)	197	18.0	40.4	21.47	16.92	-0.06	-0.09
RandomForestRegressor(n_estimators=20)	160	18.44	40.15	21.5	17.55	-0.06	-0.09
RandomForestRegressor(n_estimators=20)	130	18.34	43.5	21.55	15.23	-0.07	-0.09
RandomForestRegressor(n_estimators=20)	100	17.92	45.65	21.54	15.05	-0.08	-0.09
RandomForestRegressor(n_estimators=20)	70	17.09	43.55	20.89	14.45	-0.01	-0.03
RandomForestRegressor(n_estimators=20)	40	17.92	41.1	21.11	16.42	-0.04	-0.05
RandomForestRegressor(n_estimators=20)	20	18.13	45.55	21.93	17.08	-0.12	-0.13
RandomForestRegressor(n_estimators=20)	10	17.64	51.35	21.8	14.12	-0.09	-0.12
RandomForestRegressor(n_estimators=50)	197	17.4	43.56	20.48	16.34	0.02	0.01
RandomForestRegressor(n_estimators=50)	160	17.48	45.84	20.57	17.0	0.01	0.0
RandomForestRegressor(n_estimators=50)	130	17.26	43.66	20.34	15.61	0.03	0.03
RandomForestRegressor(n_estimators=50)	100	16.91	45.36	20.0	14.95	0.06	0.06
RandomForestRegressor(n_estimators=50)	70	16.11	43.34	19.52	13.63	0.11	0.1
RandomForestRegressor(n_estimators=50)	40	17.2	44.44	20.17	18.61	0.04	0.04
RandomForestRegressor(n_estimators=50)	20	17.36	45.74	21.34	14.35	-0.07	-0.07
RandomForestRegressor(n_estimators=50)	10	16.81	45.62	20.93	16.03	-0.02	-0.03
RandomForestRegressor(n_estimators=100)	197	16.69	44.65	20.13	17.2	0.05	0.05
RandomForestRegressor(n_estimators=100)	160	16.35	43.49	19.64	17.38	0.1	0.09
RandomForestRegressor(n_estimators=100)	130	16.62	44.55	19.92	16.81	0.08	0.07
RandomForestRegressor(n_estimators=100)	100	16.32	45.4	19.61	16.3	0.1	0.09
RandomForestRegressor(n_estimators=100)	70	15.89	44.31	19.27	15.6	0.13	0.13
RandomForestRegressor(n_estimators=100)	40	16.89	44.07	19.97	15.4	0.07	0.06

RandomForestRegressor(n_estimators=100)	20	17.07	43.12	21.1	14.49	-0.04	-0.05
RandomForestRegressor(n_estimators=100)	10	17.2	43.97	21.36	16.44	-0.06	-0.07
RandomForestRegressor(n_estimators=300)	197	16.33	47.28	19.73	15.59	0.09	0.08
RandomForestRegressor(n_estimators=300)	160	16.32	46.67	19.59	15.1	0.1	0.1
RandomForestRegressor(n_estimators=300)	130	16.46	45.84	19.62	15.7	0.1	0.09
RandomForestRegressor(n_estimators=300)	100	16.38	46.24	19.73	15.0	0.09	0.08
RandomForestRegressor(n_estimators=300)	70	16.21	45.49	19.56	14.75	0.11	0.1
RandomForestRegressor(n_estimators=300)	40	16.48	46.29	19.89	14.0	0.07	0.07
RandomForestRegressor(n_estimators=300)	20	17.1	42.33	20.95	15.1	-0.03	-0.03
RandomForestRegressor(n_estimators=300)	10	16.83	43.83	21.06	15.84	-0.03	-0.04
RandomForestRegressor(n_estimators=500)	197	16.34	45.91	19.64	15.46	0.1	0.09
RandomForestRegressor(n_estimators=500)	160	16.41	46.19	19.68	14.9	0.09	0.09
RandomForestRegressor(n_estimators=500)	130	16.48	45.3	19.65	15.47	0.1	0.09
RandomForestRegressor(n_estimators=500)	100	16.45	46.33	19.81	14.78	0.08	0.08
RandomForestRegressor(n_estimators=500)	70	16.27	45.26	19.48	14.92	0.11	0.11
RandomForestRegressor(n_estimators=500)	40	16.45	44.9	19.87	14.43	0.07	0.07
RandomForestRegressor(n_estimators=500)	20	17.07	42.35	20.88	14.8	-0.02	-0.03
RandomForestRegressor(n_estimators=500)	10	16.88	43.27	21.04	15.94	-0.03	-0.04
RandomForestRegressor(n_estimators=700)	197	16.3	45.34	19.6	15.21	0.1	0.1
RandomForestRegressor(n_estimators=700)	160	16.37	45.46	19.65	14.68	0.09	0.09
RandomForestRegressor(n_estimators=700)	130	16.44	44.87	19.62	15.22	0.1	0.09
RandomForestRegressor(n_estimators=700)	100	16.38	45.54	19.74	14.42	0.09	0.08
RandomForestRegressor(n_estimators=700)	70	16.24	45.04	19.51	14.96	0.11	0.1
RandomForestRegressor(n_estimators=700)	40	16.47	44.74	19.89	14.31	0.07	0.07
RandomForestRegressor(n_estimators=700)	20	17.01	42.24	20.88	15.07	-0.02	-0.03
RandomForestRegressor(n_estimators=700)	10	16.9	43.55	21.02	16.17	-0.03	-0.04
RandomForestRegressor(n_estimators=1000)	197	16.32	45.56	19.63	15.1	0.1	0.09
RandomForestRegressor(n_estimators=1000)	160	16.35	45.71	19.64	14.67	0.09	0.09
RandomForestRegressor(n_estimators=1000)	130	16.35	45.43	19.59	14.75	0.1	0.1
RandomForestRegressor(n_estimators=1000)	100	16.39	45.74	19.71	14.54	0.09	0.09

RandomForestRegressor(n_estimators=1000)	70	16.1	45.36	19.41	14.62	0.12	0.11
RandomForestRegressor(n_estimators=1000)	40	16.37	44.95	19.78	13.83	0.08	0.08
RandomForestRegressor(n_estimators=1000)	20	16.95	41.81	20.82	15.03	-0.01	-0.02
RandomForestRegressor(n_estimators=1000)	10	16.87	43.82	21.01	16.12	-0.03	-0.04
LinearRegression()	197	38.88	101.62	49.41	32.78	-4.47	-4.75
LinearRegression()	160	43.29	219.03	59.12	40.55	-6.82	-7.23
LinearRegression()	130	62.42	210.18	88.19	34.99	-17.31	-17.31
LinearRegression()	100	369.97	2372.64	645.65	186.3	-893.48	-980.66
LinearRegression()	70	25.53	88.31	33.11	20.07	-1.54	-1.58
LinearRegression()	40	18.31	56.06	23.44	15.1	-0.29	-0.29
LinearRegression()	20	17.37	55.86	21.61	14.42	-0.1	-0.1
LinearRegression()	10	17.4	63.34	21.72	13.93	-0.11	-0.11
PolynomialRegression(degree=3)	197	24.14	89.43	33.78	17.83	-1.68	-1.69
PolynomialRegression(degree=3)	160	26.49	131.52	36.41	19.46	-2.11	-2.12
PolynomialRegression(degree=3)	130	28.19	163.63	40.73	21.47	-2.9	-2.91
PolynomialRegression(degree=3)	100	27.49	117.33	38.13	18.2	-2.41	-2.42
PolynomialRegression(degree=3)	70	26.96	93.39	36.59	19.21	-2.15	-2.15
PolynomialRegression(degree=3)	40	21.95	73.34	28.19	20.37	-0.85	-0.87
PolynomialRegression(degree=3)	20	6.62	172.07	55.02	21.68	-5.95	-6.13
PolynomialRegression(degree=3)	10	275.11	4479.2	811.74	64.31	-1537.51	-1550.69

Table 11: Results for regression using PCA feature reduction

Algorithm	No. of features	MAE	ME	RMSE	MedAE	ExV	R <sup>2</sup>
SVR(kernel='rbf')	199	360.18	5406.43	1003.77	49.08	0.0	-0.12
SVR(kernel='rbf')	160	360.79	5409.72	1005.07	47.38	0.0	-0.12
SVR(kernel='rbf')	130	360.9	5410.69	1005.07	48.51	-0.0	-0.12
SVR(kernel='rbf')	100	360.87	5411.5	1004.9	50.19	-0.0	-0.12
SVR(kernel='rbf')	70	361.09	5412.64	1004.92	50.86	0.0	-0.12



SVR(kernel='rbf')	40	360.93	5410.68	1004.16	49.51	-0.0	-0.12
SVR(kernel='rbf')	20	360.49	5410.24	1004.0	48.38	0.0	-0.12
SVR(kernel='rbf')	10	360.51	5410.04	1004.73	48.29	-0.0	-0.12
SVR(kernel='linear')	199	344.63	5022.1	918.51	83.49	0.12	0.06
SVR(kernel='linear')	160	362.77	5224.58	967.39	114.75	0.03	-0.04
SVR(kernel='linear')	130	364.04	5344.42	983.01	115.03	0.01	-0.08
SVR(kernel='linear')	100	348.08	5335.0	971.06	97.94	0.03	-0.05
SVR(kernel='linear')	70	342.59	5372.13	971.43	74.5	0.04	-0.05
SVR(kernel='linear')	40	338.75	5360.56	975.53	59.87	0.04	-0.06
SVR(kernel='linear')	20	357.39	5390.87	998.71	72.91	-0.01	-0.11
SVR(kernel='linear')	10	353.42	5399.59	997.7	66.56	-0.0	-0.11
SVR(kernel='poly', degree=3)	199	374.61	5409.61	1011.6	54.13	-0.03	-0.14
SVR(kernel='poly', degree=3)	160	360.82	5409.76	1005.22	46.23	0.0	-0.13
SVR(kernel='poly', degree=3)	130	360.78	5409.58	1005.16	46.42	0.0	-0.13
SVR(kernel='poly', degree=3)	100	360.59	5408.61	1004.83	47.38	0.0	-0.12
SVR(kernel='poly', degree=3)	70	360.5	5408.28	1004.7	47.28	0.0	-0.12
SVR(kernel='poly', degree=3)	40	360.11	5406.9	1004.14	45.9	0.0	-0.12
SVR(kernel='poly', degree=3)	20	359.7	5409.52	1005.43	45.52	-0.0	-0.13
SVR(kernel='poly', degree=3)	10	358.19	5408.11	1002.71	62.77	-0.0	-0.12
SVR(kernel='poly', degree=4)	199	480.7	7877.98	1410.3	54.47	-1.17	-1.21
SVR(kernel='poly', degree=4)	160	360.82	5409.74	1005.21	46.26	0.0	-0.13
SVR(kernel='poly', degree=4)	130	360.59	5408.44	1004.78	47.44	0.0	-0.12
SVR(kernel='poly', degree=4)	100	360.52	5408.08	1004.66	47.08	0.0	-0.12
SVR(kernel='poly', degree=4)	70	360.3	5406.9	1004.26	45.9	0.0	-0.12
SVR(kernel='poly', degree=4)	40	360.36	5407.5	1004.34	46.5	-0.0	-0.12
SVR(kernel='poly', degree=4)	20	358.87	5410.1	1003.41	49.1	0.0	-0.12
SVR(kernel='poly', degree=4)	10	353.24	5411.1	1002.07	52.9	-0.01	-0.12
SVR(kernel='poly', degree=5)	199	1162.3	50927.22	6493.88	46.69	-45.69	-45.96
SVR(kernel='poly', degree=5)	160	360.67	5408.87	1004.92	47.13	0.0	-0.12
SVR(kernel='poly', degree=5)	130	360.55	5408.18	1004.69	47.18	0.0	-0.12

SVR(kernel='poly', degree=5)	100	360.26	5406.59	1004.16	45.59	0.0	-0.12
SVR(kernel='poly', degree=5)	70	360.3	5406.9	1004.26	45.9	0.0	-0.12
SVR(kernel='poly', degree=5)	40	360.93	5410.1	1005.23	46.1	-0.0	-0.13
SVR(kernel='poly', degree=5)	20	355.02	5410.82	1004.97	46.82	-0.0	-0.12
SVR(kernel='poly', degree=5)	10	367.96	5411.1	1012.14	50.1	-0.05	-0.14
SVR(kernel='sigmoid')	199	360.08	5407.56	1003.72	49.44	0.0	-0.12
SVR(kernel='sigmoid')	160	360.58	5409.07	1005.03	46.42	0.0	-0.12
SVR(kernel='sigmoid')	130	360.62	5408.9	1005.0	46.26	0.0	-0.12
SVR(kernel='sigmoid')	100	360.41	5408.95	1004.89	46.51	0.0	-0.12
SVR(kernel='sigmoid')	70	360.15	5408.53	1004.57	47.08	0.0	-0.12
SVR(kernel='sigmoid')	40	359.77	5408.01	1004.29	46.93	0.0	-0.12
SVR(kernel='sigmoid')	20	360.15	5409.58	1004.92	46.78	0.0	-0.12
SVR(kernel='sigmoid')	10	360.25	5409.5	1004.79	47.16	-0.0	-0.12
RandomForestRegressor(n_estimators=10)	199	495.55	3753.5	848.97	217.1	0.22	0.2
RandomForestRegressor(n_estimators=10)	160	621.85	5346.2	1067.66	275.5	-0.24	-0.27
RandomForestRegressor(n_estimators=10)	130	660.42	5308.2	1165.09	325.4	-0.48	-0.51
RandomForestRegressor(n_estimators=10)	100	640.24	5272.4	1134.28	300.0	-0.38	-0.43
RandomForestRegressor(n_estimators=10)	70	604.5	5291.4	1056.56	267.5	-0.2	-0.24
RandomForestRegressor(n_estimators=10)	40	600.09	5249.9	1123.94	284.5	-0.37	-0.41
RandomForestRegressor(n_estimators=10)	20	719.07	5886.7	1300.48	283.8	-0.81	-0.88
RandomForestRegressor(n_estimators=10)	10	579.71	5019.3	1077.89	280.8	-0.29	-0.29
RandomForestRegressor(n_estimators=20)	199	481.32	3829.7	835.36	245.2	0.23	0.22
RandomForestRegressor(n_estimators=20)	160	579.52	5333.95	1007.82	295.85	-0.11	-0.13
RandomForestRegressor(n_estimators=20)	130	642.13	5274.8	1136.88	289.3	-0.41	-0.44
RandomForestRegressor(n_estimators=20)	100	639.82	5330.75	1102.79	335.7	-0.31	-0.35
RandomForestRegressor(n_estimators=20)	70	607.18	5330.85	1068.2	255.55	-0.23	-0.27
RandomForestRegressor(n_estimators=20)	40	605.48	5245.1	1118.14	257.05	-0.36	-0.39
RandomForestRegressor(n_estimators=20)	20	699.83	5520.5	1256.31	341.25	-0.69	-0.76
RandomForestRegressor(n_estimators=20)	10	581.3	5162.4	1061.33	264.05	-0.24	-0.25
RandomForestRegressor(n_estimators=50)	199	451.38	3966.62	802.45	214.42	0.29	0.28

RandomForestRegressor(n_estimators=50)	160	566.0	5240.66	1031.6	264.36	-0.17	-0.19
RandomForestRegressor(n_estimators=50)	130	588.36	5264.22	1065.98	236.0	-0.24	-0.27
RandomForestRegressor(n_estimators=50)	100	620.5	5334.44	1113.85	320.78	-0.34	-0.38
RandomForestRegressor(n_estimators=50)	70	601.37	5295.66	1104.99	241.6	-0.33	-0.36
RandomForestRegressor(n_estimators=50)	40	573.6	5289.74	1125.92	266.52	-0.39	-0.41
RandomForestRegressor(n_estimators=50)	20	693.64	5567.96	1256.99	274.76	-0.69	-0.76
RandomForestRegressor(n_estimators=50)	10	574.43	4938.98	1039.7	282.26	-0.18	-0.2
RandomForestRegressor(n_estimators=100)	199	477.35	3850.12	808.83	252.97	0.28	0.27
RandomForestRegressor(n_estimators=100)	160	558.06	5272.77	1029.53	244.16	-0.17	-0.18
RandomForestRegressor(n_estimators=100)	130	576.74	5295.53	1056.82	235.91	-0.23	-0.24
RandomForestRegressor(n_estimators=100)	100	604.21	5315.07	1078.16	336.67	-0.26	-0.29
RandomForestRegressor(n_estimators=100)	70	596.2	5261.75	1105.58	236.45	-0.34	-0.36
RandomForestRegressor(n_estimators=100)	40	594.38	5240.78	1133.91	270.22	-0.41	-0.43
RandomForestRegressor(n_estimators=100)	20	683.09	5202.29	1243.07	278.47	-0.66	-0.72
RandomForestRegressor(n_estimators=100)	10	578.29	5135.98	1049.75	270.69	-0.2	-0.23
RandomForestRegressor(n_estimators=300)	199	490.41	3829.81	830.68	252.8	0.25	0.23
RandomForestRegressor(n_estimators=300)	160	566.51	5225.02	1051.81	262.74	-0.22	-0.23
RandomForestRegressor(n_estimators=300)	130	572.78	5282.38	1061.03	257.08	-0.24	-0.25
RandomForestRegressor(n_estimators=300)	100	608.79	5296.94	1112.02	331.12	-0.34	-0.38
RandomForestRegressor(n_estimators=300)	70	606.49	5265.17	1114.94	286.53	-0.35	-0.38
RandomForestRegressor(n_estimators=300)	40	593.77	5288.65	1149.81	254.93	-0.45	-0.47
RandomForestRegressor(n_estimators=300)	20	663.19	5204.81	1207.54	282.64	-0.57	-0.62
RandomForestRegressor(n_estimators=300)	10	582.32	5186.0	1045.52	263.81	-0.19	-0.22
RandomForestRegressor(n_estimators=500)	199	485.61	3889.56	825.76	267.19	0.26	0.24
RandomForestRegressor(n_estimators=500)	160	562.06	5162.03	1045.73	262.38	-0.2	-0.22
RandomForestRegressor(n_estimators=500)	130	567.59	5293.49	1056.86	263.79	-0.23	-0.24
RandomForestRegressor(n_estimators=500)	100	605.82	5303.82	1099.04	337.28	-0.31	-0.35
RandomForestRegressor(n_estimators=500)	70	598.16	5261.24	1102.28	277.92	-0.32	-0.35
RandomForestRegressor(n_estimators=500)	40	591.29	5272.38	1147.69	280.47	-0.44	-0.47
RandomForestRegressor(n_estimators=500)	20	665.05	5149.63	1219.29	273.58	-0.6	-0.66

RandomForestRegressor(n_estimators=500)	10	588.9	5123.44	1051.68	277.57	-0.2	-0.23
RandomForestRegressor(n_estimators=700)	199	484.65	3833.54	816.06	268.97	0.28	0.26
RandomForestRegressor(n_estimators=700)	160	561.44	5147.49	1039.58	279.52	-0.19	-0.2
RandomForestRegressor(n_estimators=700)	130	570.45	5265.38	1059.08	257.58	-0.23	-0.25
RandomForestRegressor(n_estimators=700)	100	607.24	5305.85	1102.28	341.76	-0.32	-0.35
RandomForestRegressor(n_estimators=700)	70	598.92	5258.52	1101.3	282.69	-0.32	-0.35
RandomForestRegressor(n_estimators=700)	40	588.74	5265.63	1133.6	274.61	-0.41	-0.43
RandomForestRegressor(n_estimators=700)	20	661.75	5141.89	1213.06	277.66	-0.59	-0.64
RandomForestRegressor(n_estimators=700)	10	588.47	5113.22	1046.9	272.75	-0.19	-0.22
RandomForestRegressor(n_estimators=1000)	199	484.4	3882.25	817.33	277.72	0.27	0.26
RandomForestRegressor(n_estimators=1000)	160	564.33	5152.21	1044.04	276.11	-0.2	-0.21
RandomForestRegressor(n_estimators=1000)	130	569.08	5264.46	1057.05	254.07	-0.23	-0.24
RandomForestRegressor(n_estimators=1000)	100	605.24	5295.95	1100.96	340.85	-0.32	-0.35
RandomForestRegressor(n_estimators=1000)	70	601.48	5270.33	1104.18	282.58	-0.33	-0.36
RandomForestRegressor(n_estimators=1000)	40	587.84	5272.18	1131.82	278.25	-0.4	-0.43
RandomForestRegressor(n_estimators=1000)	20	659.91	5155.72	1209.2	269.27	-0.58	-0.63
RandomForestRegressor(n_estimators=1000)	10	586.42	5146.25	1044.17	261.56	-0.18	-0.21
LinearRegression()	199	9136.41	74629.68	16713.1	4326.54	-307.46	-310.06
LinearRegression()	160	4353.07	41201.45	8608.69	1992.75	-81.33	-81.53
LinearRegression()	130	2366.06	15024.67	3629.75	1607.1	-13.67	-13.67
LinearRegression()	100	1981.55	17927.61	3314.01	1168.67	-11.2	-11.23
LinearRegression()	70	1429.58	10428.01	2481.57	518.4	-5.79	-5.86
LinearRegression()	40	1052.3	9767.6	1834.39	554.31	-2.69	-2.75
LinearRegression()	20	569.97	4994.09	1031.31	313.79	-0.18	-0.18
LinearRegression()	10	551.04	5049.09	1041.49	343.17	-0.2	-0.21
PolynomialRegression(degree=3)	199	1549.0	25520.16	3837.1	675.55	-15.36	-15.4
PolynomialRegression(degree=3)	160	473.05	5044.79	950.28	331.53	-0.01	-0.01
PolynomialRegression(degree=3)	130	401.95	5231.29	970.28	174.87	-0.02	-0.05
PolynomialRegression(degree=3)	100	657.18	6457.46	1308.8	353.53	-0.91	-0.91
PolynomialRegression(degree=3)	70	1118.88	14629.75	2338.44	410.65	-4.93	-5.09

PolynomialRegression(degree=3)	40	7494.53	150967.16	21129.39	1884.52	-492.68	-496.17
PolynomialRegression(degree=3)	20	312619.7	5188198	879610.2	20604.83	-858208	-861619
PolynomialRegression(degree=3)	10	4.5E+08	1.12E+10	1.77E+09	4436112	-3.4E+12	-3.5E+12

Table 12: Results for regression with Maven Reuse outlier cut-off at >400 using PCA feature reduction

Algorithm	No. of features	MAE	ME	RMSE	MedAE	ExV	R <sup>2</sup>
SVR(kernel='rbf')	198	61.71	348.72	105.27	27.9	-0.0	-0.14
SVR(kernel='rbf')	160	61.63	347.67	105.07	27.68	0.0	-0.14
SVR(kernel='rbf')	130	61.6	347.58	105.06	27.76	0.0	-0.14
SVR(kernel='rbf')	100	61.58	347.68	105.12	28.05	0.0	-0.14
SVR(kernel='rbf')	70	61.55	347.82	105.22	28.45	0.0	-0.14
SVR(kernel='rbf')	40	61.53	348.21	105.19	28.23	0.0	-0.14
SVR(kernel='rbf')	20	61.52	347.42	105.13	28.18	0.0	-0.14
SVR(kernel='rbf')	10	61.65	348.45	105.01	28.01	-0.0	-0.14
SVR(kernel='linear')	198	62.64	337.7	105.0	25.94	-0.04	-0.14
SVR(kernel='linear')	160	78.62	347.27	110.24	64.13	-0.21	-0.25
SVR(kernel='linear')	130	76.65	330.73	111.93	50.52	-0.2	-0.29
SVR(kernel='linear')	100	69.14	336.14	104.04	43.2	-0.02	-0.12
SVR(kernel='linear')	70	66.63	341.08	105.48	37.98	-0.04	-0.15
SVR(kernel='linear')	40	64.22	351.96	105.91	30.79	-0.03	-0.16
SVR(kernel='linear')	20	63.26	333.39	104.22	32.61	-0.01	-0.12
SVR(kernel='linear')	10	61.95	344.08	103.47	31.85	0.01	-0.1
SVR(kernel='poly', degree=3)	198	60.05	347.94	104.26	25.09	0.0	-0.12
SVR(kernel='poly', degree=3)	160	61.68	347.58	105.04	27.58	-0.0	-0.14
SVR(kernel='poly', degree=3)	130	61.67	347.69	105.08	27.69	-0.0	-0.14
SVR(kernel='poly', degree=3)	100	61.64	347.94	105.17	27.96	-0.0	-0.14
SVR(kernel='poly', degree=3)	70	61.69	347.61	105.05	27.66	-0.0	-0.14
SVR(kernel='poly', degree=3)	40	61.84	346.33	104.51	26.91	0.0	-0.13

SVR(kernel='poly', degree=3)	20	61.59	346.49	104.55	29.77	0.0	-0.13
SVR(kernel='poly', degree=3)	10	61.91	349.51	106.01	28.18	-0.01	-0.16
SVR(kernel='poly', degree=4)	198	64.27	347.53	107.8	27.55	-0.11	-0.2
SVR(kernel='poly', degree=4)	160	61.68	347.59	105.04	27.59	-0.0	-0.14
SVR(kernel='poly', degree=4)	130	61.67	347.73	105.09	27.73	-0.0	-0.14
SVR(kernel='poly', degree=4)	100	61.67	347.67	105.08	27.68	-0.0	-0.14
SVR(kernel='poly', degree=4)	70	61.86	345.89	104.47	27.1	-0.0	-0.13
SVR(kernel='poly', degree=4)	40	62.05	343.56	103.72	29.1	0.0	-0.11
SVR(kernel='poly', degree=4)	20	61.3	348.02	105.05	29.1	-0.0	-0.14
SVR(kernel='poly', degree=4)	10	64.12	349.08	111.65	27.9	-0.12	-0.29
SVR(kernel='poly', degree=5)	198	76.28	770.06	152.91	29.04	-1.38	-1.41
SVR(kernel='poly', degree=5)	160	61.68	347.58	105.04	27.58	-0.0	-0.14
SVR(kernel='poly', degree=5)	130	61.67	347.71	105.08	27.71	-0.0	-0.14
SVR(kernel='poly', degree=5)	100	61.77	346.72	104.74	26.72	-0.0	-0.13
SVR(kernel='poly', degree=5)	70	62.04	344.19	103.91	28.82	-0.0	-0.11
SVR(kernel='poly', degree=5)	40	61.93	345.39	104.24	27.67	0.0	-0.12
SVR(kernel='poly', degree=5)	20	61.47	347.95	105.08	29.06	-0.0	-0.14
SVR(kernel='poly', degree=5)	10	69.32	544.65	129.5	27.9	-0.51	-0.73
SVR(kernel='sigmoid')	198	62.05	346.99	105.01	28.15	-0.0	-0.14
SVR(kernel='sigmoid')	160	61.69	347.96	105.12	27.67	-0.0	-0.14
SVR(kernel='sigmoid')	130	61.69	348.21	105.11	27.8	0.0	-0.14
SVR(kernel='sigmoid')	100	61.64	348.6	105.21	28.37	0.0	-0.14
SVR(kernel='sigmoid')	70	61.78	348.8	105.18	28.9	-0.0	-0.14
SVR(kernel='sigmoid')	40	61.64	351.66	105.58	28.04	-0.0	-0.15
SVR(kernel='sigmoid')	20	61.66	345.62	104.99	28.11	-0.0	-0.14
SVR(kernel='sigmoid')	10	61.56	345.91	104.9	28.49	0.0	-0.14
RandomForestRegressor(n_estimators=10)	198	73.52	324.5	108.04	44.8	-0.19	-0.2
RandomForestRegressor(n_estimators=10)	160	77.89	315.5	105.19	59.7	-0.13	-0.14
RandomForestRegressor(n_estimators=10)	130	75.91	312.3	106.05	48.9	-0.16	-0.16
RandomForestRegressor(n_estimators=10)	100	81.83	330.5	112.83	64.9	-0.31	-0.31

RandomForestRegressor(n_estimators=10)	70	80.08	284.8	109.6	57.4	-0.24	-0.24
RandomForestRegressor(n_estimators=10)	40	82.45	352.5	112.32	55.6	-0.3	-0.3
RandomForestRegressor(n_estimators=10)	20	73.16	306.6	103.5	54.3	-0.11	-0.11
RandomForestRegressor(n_estimators=10)	10	81.76	342.1	109.22	59.1	-0.23	-0.23
RandomForestRegressor(n_estimators=20)	198	74.09	314.1	105.64	44.9	-0.15	-0.15
RandomForestRegressor(n_estimators=20)	160	75.45	292.45	100.31	54.0	-0.02	-0.04
RandomForestRegressor(n_estimators=20)	130	72.37	295.7	102.66	43.85	-0.09	-0.09
RandomForestRegressor(n_estimators=20)	100	78.76	290.6	106.67	59.85	-0.17	-0.17
RandomForestRegressor(n_estimators=20)	70	79.95	283.2	108.39	51.85	-0.21	-0.21
RandomForestRegressor(n_estimators=20)	40	78.04	358.55	107.76	55.7	-0.2	-0.2
RandomForestRegressor(n_estimators=20)	20	72.6	311.1	100.17	50.8	-0.04	-0.04
RandomForestRegressor(n_estimators=20)	10	76.26	340.95	102.75	62.45	-0.09	-0.09
RandomForestRegressor(n_estimators=50)	198	76.18	295.1	107.21	44.88	-0.19	-0.19
RandomForestRegressor(n_estimators=50)	160	71.2	281.64	96.29	51.52	0.05	0.04
RandomForestRegressor(n_estimators=50)	130	69.83	304.78	97.66	46.34	0.02	0.02
RandomForestRegressor(n_estimators=50)	100	75.16	267.36	99.31	57.08	-0.02	-0.02
RandomForestRegressor(n_estimators=50)	70	76.68	266.48	102.32	59.6	-0.08	-0.08
RandomForestRegressor(n_estimators=50)	40	73.51	330.14	100.81	50.96	-0.05	-0.05
RandomForestRegressor(n_estimators=50)	20	71.4	294.56	96.57	51.1	0.04	0.04
RandomForestRegressor(n_estimators=50)	10	75.02	328.12	100.81	55.78	-0.05	-0.05
RandomForestRegressor(n_estimators=100)	198	74.58	288.72	105.4	49.59	-0.15	-0.15
RandomForestRegressor(n_estimators=100)	160	73.93	283.64	99.74	54.26	-0.02	-0.03
RandomForestRegressor(n_estimators=100)	130	73.44	305.84	100.99	50.82	-0.05	-0.05
RandomForestRegressor(n_estimators=100)	100	76.46	276.87	101.21	54.0	-0.05	-0.06
RandomForestRegressor(n_estimators=100)	70	74.5	278.34	100.9	55.17	-0.05	-0.05
RandomForestRegressor(n_estimators=100)	40	74.59	331.39	100.08	53.97	-0.03	-0.03
RandomForestRegressor(n_estimators=100)	20	72.52	299.21	98.66	51.16	-0.0	-0.0
RandomForestRegressor(n_estimators=100)	10	73.74	334.04	99.81	53.24	-0.03	-0.03
RandomForestRegressor(n_estimators=300)	198	75.19	293.35	105.04	47.57	-0.14	-0.14
RandomForestRegressor(n_estimators=300)	160	74.78	302.79	101.41	57.0	-0.06	-0.06

RandomForestRegressor(n_estimators=300)	130	74.61	304.41	101.42	56.6	-0.06	-0.06
RandomForestRegressor(n_estimators=300)	100	77.38	294.29	102.61	56.07	-0.08	-0.09
RandomForestRegressor(n_estimators=300)	70	72.07	280.24	98.12	49.61	0.01	0.01
RandomForestRegressor(n_estimators=300)	40	73.88	329.77	99.84	55.68	-0.03	-0.03
RandomForestRegressor(n_estimators=300)	20	73.77	307.23	98.99	52.54	-0.01	-0.01
RandomForestRegressor(n_estimators=300)	10	75.74	341.01	101.75	51.62	-0.07	-0.07
RandomForestRegressor(n_estimators=500)	198	75.24	294.19	105.45	48.92	-0.15	-0.15
RandomForestRegressor(n_estimators=500)	160	74.88	300.28	100.94	56.22	-0.05	-0.05
RandomForestRegressor(n_estimators=500)	130	75.23	304.62	101.92	54.28	-0.07	-0.07
RandomForestRegressor(n_estimators=500)	100	76.19	294.74	102.06	56.45	-0.07	-0.07
RandomForestRegressor(n_estimators=500)	70	71.93	282.61	98.47	49.08	-0.0	-0.0
RandomForestRegressor(n_estimators=500)	40	73.4	330.11	99.63	56.01	-0.02	-0.02
RandomForestRegressor(n_estimators=500)	20	74.14	308.06	99.23	51.68	-0.02	-0.02
RandomForestRegressor(n_estimators=500)	10	75.74	337.71	101.3	53.74	-0.06	-0.06
RandomForestRegressor(n_estimators=700)	198	74.84	294.32	104.98	47.36	-0.14	-0.14
RandomForestRegressor(n_estimators=700)	160	74.42	301.61	100.67	59.38	-0.04	-0.05
RandomForestRegressor(n_estimators=700)	130	75.78	304.84	102.46	57.99	-0.08	-0.08
RandomForestRegressor(n_estimators=700)	100	75.74	293.77	102.03	56.01	-0.07	-0.07
RandomForestRegressor(n_estimators=700)	70	72.64	280.47	98.81	48.35	-0.01	-0.01
RandomForestRegressor(n_estimators=700)	40	73.46	328.85	100.0	53.62	-0.03	-0.03
RandomForestRegressor(n_estimators=700)	20	74.11	309.98	99.06	51.23	-0.01	-0.01
RandomForestRegressor(n_estimators=700)	10	76.03	338.33	101.69	52.36	-0.07	-0.07
RandomForestRegressor(n_estimators=1000)	198	74.77	295.0	105.11	46.54	-0.14	-0.14
RandomForestRegressor(n_estimators=1000)	160	74.64	301.33	101.33	59.47	-0.06	-0.06
RandomForestRegressor(n_estimators=1000)	130	75.79	301.65	102.67	58.24	-0.09	-0.09
RandomForestRegressor(n_estimators=1000)	100	76.16	297.01	102.57	55.8	-0.08	-0.09
RandomForestRegressor(n_estimators=1000)	70	72.79	282.28	98.99	48.27	-0.01	-0.01
RandomForestRegressor(n_estimators=1000)	40	73.25	330.57	100.03	53.6	-0.03	-0.03
RandomForestRegressor(n_estimators=1000)	20	74.43	312.86	99.39	52.85	-0.02	-0.02
RandomForestRegressor(n_estimators=1000)	10	76.44	338.48	102.14	52.3	-0.07	-0.08



LinearRegression()	198	290.39	1771.48	422.62	196.29	-16.13	-17.43
LinearRegression()	160	160.89	561.09	202.96	126.88	-3.21	-3.25
LinearRegression()	130	976.95	15118.53	2825.24	218.75	-822.37	-822.64
LinearRegression()	100	319.53	3386.05	683.8	114.12	-46.61	-47.25
LinearRegression()	70	181.7	1748.38	341.53	74.21	-10.84	-11.04
LinearRegression()	40	97.38	485.4	132.34	68.41	-0.8	-0.81
LinearRegression()	20	75.13	299.11	103.63	53.41	-0.11	-0.11
LinearRegression()	10	68.53	304.3	95.03	54.3	0.07	0.07
PolynomialRegression(degree=3)	198	133.93	750.31	209.08	67.22	-3.42	-3.51
PolynomialRegression(degree=3)	160	70.32	312.89	98.37	54.24	0.0	0.0
PolynomialRegression(degree=3)	130	65.56	324.18	99.91	43.79	-0.0	-0.03
PolynomialRegression(degree=3)	100	63.49	327.3	98.83	43.18	0.02	-0.01
PolynomialRegression(degree=3)	70	76.87	318.2	109.25	51.54	-0.22	-0.23
PolynomialRegression(degree=3)	40	365.21	4222.79	787.21	162.66	-62.94	-62.94
PolynomialRegression(degree=3)	20	9731.82	176259.8	34209.1	508.63	-117473	-120756
PolynomialRegression(degree=3)	10	4978182	2E+08	29373000	14461.47	-8.7E+10	-8.9E+10

Table 13: Results for regression with Maven Reuse outlier cut-off at >200 using PCA feature reduction

Algorithm	No. of features	MAE	ME	RMSE	MedAE	ExV	R <sup>2</sup>
SVR(kernel='rbf')	199	35.28	151.54	52.98	21.22	-0.0	-0.18
SVR(kernel='rbf')	160	35.2	150.02	52.7	21.48	0.0	-0.17
SVR(kernel='rbf')	130	35.15	149.94	52.63	21.43	0.0	-0.17
SVR(kernel='rbf')	100	35.14	150.25	52.7	21.52	0.0	-0.17
SVR(kernel='rbf')	70	35.16	150.08	52.66	21.65	0.0	-0.17
SVR(kernel='rbf')	40	35.14	150.16	52.59	22.26	0.0	-0.16
SVR(kernel='rbf')	20	35.1	150.48	52.63	22.2	0.0	-0.17
SVR(kernel='rbf')	10	35.28	149.15	52.58	22.39	0.0	-0.16
SVR(kernel='linear')	199	38.8	224.44	63.51	19.37	-0.63	-0.7

SVR(kernel='linear')	160	35.09	142.01	49.83	22.06	-0.01	-0.05
SVR(kernel='linear')	130	49.52	154.49	63.28	38.21	-0.66	-0.69
SVR(kernel='linear')	100	40.83	190.34	56.75	26.46	-0.29	-0.36
SVR(kernel='linear')	70	34.79	156.6	49.72	25.96	0.1	-0.04
SVR(kernel='linear')	40	32.05	148.16	49.19	18.77	0.09	-0.02
SVR(kernel='linear')	20	37.49	149.32	53.26	26.94	-0.05	-0.19
SVR(kernel='linear')	10	34.27	147.85	51.92	20.92	0.04	-0.14
SVR(kernel='poly', degree=3)	199	39.17	151.87	56.64	23.44	-0.19	-0.35
SVR(kernel='poly', degree=3)	160	35.22	150.13	52.74	21.37	-0.0	-0.17
SVR(kernel='poly', degree=3)	130	35.22	150.17	52.75	21.34	-0.0	-0.17
SVR(kernel='poly', degree=3)	100	35.22	150.24	52.77	21.28	0.0	-0.17
SVR(kernel='poly', degree=3)	70	35.18	150.12	52.71	21.39	0.0	-0.17
SVR(kernel='poly', degree=3)	40	34.95	149.7	52.43	21.81	0.0	-0.16
SVR(kernel='poly', degree=3)	20	34.66	150.12	52.57	22.39	0.01	-0.16
SVR(kernel='poly', degree=3)	10	35.34	148.58	52.75	21.83	0.01	-0.17
SVR(kernel='poly', degree=4)	199	46.68	238.07	74.95	22.01	-1.18	-1.37
SVR(kernel='poly', degree=4)	160	35.22	150.15	52.75	21.35	-0.0	-0.17
SVR(kernel='poly', degree=4)	130	35.22	150.15	52.75	21.35	-0.0	-0.17
SVR(kernel='poly', degree=4)	100	35.21	150.13	52.74	21.37	0.0	-0.17
SVR(kernel='poly', degree=4)	70	35.19	150.1	52.71	21.4	0.0	-0.17
SVR(kernel='poly', degree=4)	40	34.78	149.81	52.46	21.62	-0.0	-0.16
SVR(kernel='poly', degree=4)	20	34.43	149.9	52.37	21.65	0.01	-0.16
SVR(kernel='poly', degree=4)	10	36.97	148.9	54.82	22.5	-0.13	-0.27
SVR(kernel='poly', degree=5)	199	60.13	657.09	126.82	23.92	-5.35	-5.77
SVR(kernel='poly', degree=5)	160	35.23	150.17	52.75	21.33	-0.0	-0.17
SVR(kernel='poly', degree=5)	130	35.22	150.14	52.74	21.36	-0.0	-0.17
SVR(kernel='poly', degree=5)	100	35.21	150.11	52.73	21.39	0.0	-0.17
SVR(kernel='poly', degree=5)	70	35.19	150.1	52.71	21.4	0.0	-0.17
SVR(kernel='poly', degree=5)	40	34.73	149.9	52.5	21.55	-0.0	-0.16
SVR(kernel='poly', degree=5)	20	34.43	149.9	52.39	21.67	0.01	-0.16

SVR(kernel='poly', degree=5)	10	50.88	296.98	86.39	22.5	-1.93	-2.14
SVR(kernel='sigmoid')	199	35.6	151.72	53.11	23.36	-0.0	-0.19
SVR(kernel='sigmoid')	160	35.19	149.99	52.69	21.52	0.0	-0.17
SVR(kernel='sigmoid')	130	35.16	150.26	52.7	21.36	0.0	-0.17
SVR(kernel='sigmoid')	100	35.04	150.25	52.65	21.44	0.0	-0.17
SVR(kernel='sigmoid')	70	35.03	150.35	52.45	21.27	0.01	-0.16
SVR(kernel='sigmoid')	40	35.07	150.5	52.49	21.94	0.01	-0.16
SVR(kernel='sigmoid')	20	34.93	150.29	52.54	22.56	0.01	-0.16
SVR(kernel='sigmoid')	10	35.25	150.01	52.79	22.22	0.01	-0.17
RandomForestRegressor(n_estimators=10)	199	31.63	134.6	46.8	18.55	0.12	0.08
RandomForestRegressor(n_estimators=10)	160	36.06	151.0	49.47	28.15	0.01	-0.03
RandomForestRegressor(n_estimators=10)	130	35.83	139.1	49.14	26.75	0.02	-0.02
RandomForestRegressor(n_estimators=10)	100	36.35	129.3	48.78	24.05	0.01	-0.0
RandomForestRegressor(n_estimators=10)	70	34.22	134.8	47.1	22.8	0.08	0.07
RandomForestRegressor(n_estimators=10)	40	30.07	152.5	45.05	20.2	0.18	0.15
RandomForestRegressor(n_estimators=10)	20	34.97	148.8	50.51	25.2	-0.07	-0.07
RandomForestRegressor(n_estimators=10)	10	29.87	137.6	47.05	15.15	0.11	0.07
RandomForestRegressor(n_estimators=20)	199	30.12	141.65	44.8	17.88	0.18	0.15
RandomForestRegressor(n_estimators=20)	160	35.69	144.1	48.98	27.12	0.03	-0.01
RandomForestRegressor(n_estimators=20)	130	35.39	137.1	49.09	25.68	0.01	-0.01
RandomForestRegressor(n_estimators=20)	100	35.49	123.85	47.61	25.75	0.05	0.05
RandomForestRegressor(n_estimators=20)	70	33.97	135.1	46.51	24.95	0.1	0.09
RandomForestRegressor(n_estimators=20)	40	33.96	147.45	48.67	21.0	0.02	0.0
RandomForestRegressor(n_estimators=20)	20	34.09	147.0	49.8	20.78	-0.03	-0.04
RandomForestRegressor(n_estimators=20)	10	31.39	131.95	46.27	18.85	0.13	0.1
RandomForestRegressor(n_estimators=50)	199	29.97	149.26	45.76	20.75	0.15	0.12
RandomForestRegressor(n_estimators=50)	160	35.99	148.04	49.57	24.89	-0.01	-0.04
RandomForestRegressor(n_estimators=50)	130	35.46	139.16	48.73	27.57	0.02	-0.0
RandomForestRegressor(n_estimators=50)	100	34.63	122.24	47.3	22.49	0.07	0.06
RandomForestRegressor(n_estimators=50)	70	33.82	144.3	46.78	23.23	0.09	0.08

RandomForestRegressor(n_estimators=50)	40	34.55	143.96	48.67	20.73	0.02	0.0
RandomForestRegressor(n_estimators=50)	20	33.51	143.4	49.93	23.58	-0.03	-0.05
RandomForestRegressor(n_estimators=50)	10	31.05	137.18	47.06	16.02	0.1	0.07
RandomForestRegressor(n_estimators=100)	199	29.97	145.84	45.25	16.17	0.17	0.14
RandomForestRegressor(n_estimators=100)	160	36.01	144.19	49.79	25.45	-0.02	-0.04
RandomForestRegressor(n_estimators=100)	130	35.2	136.38	48.83	24.84	0.01	-0.0
RandomForestRegressor(n_estimators=100)	100	34.04	130.63	46.79	22.4	0.09	0.08
RandomForestRegressor(n_estimators=100)	70	33.46	144.75	46.68	24.44	0.09	0.08
RandomForestRegressor(n_estimators=100)	40	34.93	142.49	49.81	22.5	-0.02	-0.05
RandomForestRegressor(n_estimators=100)	20	33.33	142.13	49.8	21.54	-0.03	-0.04
RandomForestRegressor(n_estimators=100)	10	32.13	140.22	48.22	17.43	0.05	0.02
RandomForestRegressor(n_estimators=300)	199	30.18	147.41	45.27	16.38	0.17	0.14
RandomForestRegressor(n_estimators=300)	160	35.51	139.46	49.33	23.56	0.0	-0.03
RandomForestRegressor(n_estimators=300)	130	35.37	135.24	48.72	23.8	0.03	0.0
RandomForestRegressor(n_estimators=300)	100	34.35	130.63	47.53	22.86	0.06	0.05
RandomForestRegressor(n_estimators=300)	70	33.19	145.43	46.59	24.53	0.1	0.09
RandomForestRegressor(n_estimators=300)	40	34.24	143.82	50.13	20.15	-0.03	-0.06
RandomForestRegressor(n_estimators=300)	20	33.27	140.57	49.87	22.22	-0.03	-0.05
RandomForestRegressor(n_estimators=300)	10	31.78	132.44	47.68	16.72	0.07	0.04
RandomForestRegressor(n_estimators=500)	199	30.24	148.02	45.39	17.68	0.17	0.13
RandomForestRegressor(n_estimators=500)	160	35.65	139.01	49.61	23.78	-0.01	-0.04
RandomForestRegressor(n_estimators=500)	130	35.47	135.23	48.67	24.47	0.02	0.0
RandomForestRegressor(n_estimators=500)	100	34.41	130.55	47.65	22.4	0.06	0.04
RandomForestRegressor(n_estimators=500)	70	33.13	146.04	46.68	24.59	0.09	0.08
RandomForestRegressor(n_estimators=500)	40	33.94	142.51	49.81	19.6	-0.02	-0.05
RandomForestRegressor(n_estimators=500)	20	33.1	139.15	49.47	22.98	-0.01	-0.03
RandomForestRegressor(n_estimators=500)	10	31.4	132.49	47.37	15.78	0.08	0.05
RandomForestRegressor(n_estimators=700)	199	30.19	147.98	45.35	17.86	0.17	0.13
RandomForestRegressor(n_estimators=700)	160	35.59	139.29	49.63	24.52	-0.01	-0.04
RandomForestRegressor(n_estimators=700)	130	35.51	135.54	48.77	24.08	0.02	-0.0

RandomForestRegressor(n_estimators=700)	100	34.8	130.51	47.82	22.72	0.05	0.04
RandomForestRegressor(n_estimators=700)	70	32.99	145.69	46.4	23.79	0.1	0.09
RandomForestRegressor(n_estimators=700)	40	33.91	143.01	49.54	19.77	-0.01	-0.03
RandomForestRegressor(n_estimators=700)	20	32.73	139.54	49.29	22.55	-0.0	-0.02
RandomForestRegressor(n_estimators=700)	10	31.02	132.37	46.96	16.33	0.09	0.07
RandomForestRegressor(n_estimators=1000)	199	30.12	147.59	45.23	17.88	0.18	0.14
RandomForestRegressor(n_estimators=1000)	160	35.74	138.12	49.79	25.42	-0.02	-0.04
RandomForestRegressor(n_estimators=1000)	130	35.73	133.95	48.94	24.92	0.01	-0.01
RandomForestRegressor(n_estimators=1000)	100	34.75	131.25	47.8	22.29	0.05	0.04
RandomForestRegressor(n_estimators=1000)	70	32.77	144.22	45.94	23.25	0.12	0.11
RandomForestRegressor(n_estimators=1000)	40	33.52	143.01	49.35	19.26	0.0	-0.03
RandomForestRegressor(n_estimators=1000)	20	32.66	141.04	49.35	22.12	-0.01	-0.03
RandomForestRegressor(n_estimators=1000)	10	31.07	133.15	47.0	16.86	0.09	0.07
LinearRegression()	199	155.38	1101.21	279.6	68.68	-30.7	-31.93
LinearRegression()	160	34.74	127.22	48.17	23.18	0.06	0.02
LinearRegression()	130	164.7	511.37	211.31	161.41	-17.01	-17.81
LinearRegression()	100	244.3	1561.44	446.96	80.84	-80.97	-83.15
LinearRegression()	70	148.05	2104.78	399.55	39.35	-65.76	-66.24
LinearRegression()	40	74.81	1068.23	186.87	26.95	-13.59	-13.71
LinearRegression()	20	49.07	394.93	83.26	28.31	-1.91	-1.92
LinearRegression()	10	37.02	135.81	51.09	25.96	-0.08	-0.1
PolynomialRegression(degree=3)	199	120.35	801.63	197.6	58.46	-14.53	-15.45
PolynomialRegression(degree=3)	160	34.68	142.45	50.28	22.16	-0.0	-0.06
PolynomialRegression(degree=3)	130	34.8	143.4	50.56	22.03	-0.0	-0.08
PolynomialRegression(degree=3)	100	34.3	147.69	51.26	20.09	-0.01	-0.11
PolynomialRegression(degree=3)	70	32.14	145.86	45.83	20.6	0.14	0.12
PolynomialRegression(degree=3)	40	123.49	1463.18	262.42	63.61	-27.22	-28.01
PolynomialRegression(degree=3)	20	1466.06	11101.66	2920.99	537.34	-3589.7	-3593.01
PolynomialRegression(degree=3)	10	1871721	29585462	6291781	26673.35	-1.6E+10	-1.7E+10

Table 14: Results for regression with Maven Reuse outlier cut-off at >100 using PCA feature reduction

Algorithm	No. of features	MAE	ME	RMSE	MedAE	ExV	R <sup>2</sup>
SVR(kernel='rbf')	197	17.37	46.66	20.68	17.0	0.01	-0.01
SVR(kernel='rbf')	130	17.51	47.1	20.77	16.02	0.0	-0.02
SVR(kernel='rbf')	100	17.52	47.79	20.78	16.06	0.0	-0.02
SVR(kernel='rbf')	70	17.6	49.32	20.92	16.65	-0.0	-0.03
SVR(kernel='rbf')	40	17.47	48.18	20.86	17.35	-0.01	-0.02
SVR(kernel='rbf')	20	17.59	47.25	20.9	16.7	-0.01	-0.03
SVR(kernel='rbf')	10	17.53	47.58	20.88	16.88	-0.02	-0.03
SVR(kernel='linear')	197	24.26	66.77	31.18	20.94	-1.28	-1.29
SVR(kernel='linear')	130	19.66	51.41	24.12	17.87	-0.25	-0.37
SVR(kernel='linear')	100	22.68	93.86	30.83	17.03	-1.24	-1.24
SVR(kernel='linear')	70	18.6	62.34	22.51	18.28	-0.09	-0.19
SVR(kernel='linear')	40	20.79	69.01	24.56	19.15	-0.19	-0.42
SVR(kernel='linear')	20	20.38	47.2	23.71	19.29	-0.28	-0.32
SVR(kernel='linear')	10	19.82	48.01	22.23	18.59	-0.12	-0.16
SVR(kernel='poly', degree=3)	197	17.6	48.88	20.86	16.2	0.02	-0.02
SVR(kernel='poly', degree=3)	130	17.5	47.0	20.77	16.0	0.0	-0.02
SVR(kernel='poly', degree=3)	100	17.47	46.84	20.75	16.15	0.0	-0.01
SVR(kernel='poly', degree=3)	70	17.5	47.07	20.76	16.0	0.0	-0.01
SVR(kernel='poly', degree=3)	40	17.52	47.59	20.63	16.06	0.02	-0.0
SVR(kernel='poly', degree=3)	20	17.43	48.33	20.33	16.57	0.07	0.03
SVR(kernel='poly', degree=3)	10	19.46	47.9	22.27	19.09	-0.15	-0.17
SVR(kernel='poly', degree=4)	197	17.94	49.94	21.25	16.49	0.02	-0.06
SVR(kernel='poly', degree=4)	130	17.5	47.0	20.77	16.0	-0.0	-0.02
SVR(kernel='poly', degree=4)	100	17.47	46.82	20.75	16.18	0.0	-0.01
SVR(kernel='poly', degree=4)	70	17.55	47.34	20.81	16.0	0.0	-0.02

SVR(kernel='poly', degree=4)	40	17.59	48.1	20.63	16.02	0.03	-0.0
SVR(kernel='poly', degree=4)	20	16.99	49.86	20.13	14.51	0.12	0.05
SVR(kernel='poly', degree=4)	10	18.88	61.72	22.73	14.93	-0.21	-0.22
SVR(kernel='poly', degree=5)	197	17.77	50.12	21.05	16.5	0.04	-0.04
SVR(kernel='poly', degree=5)	130	17.5	47.0	20.77	16.0	0.0	-0.02
SVR(kernel='poly', degree=5)	100	17.47	46.84	20.75	16.16	0.0	-0.01
SVR(kernel='poly', degree=5)	70	17.61	47.64	20.86	16.0	0.0	-0.02
SVR(kernel='poly', degree=5)	40	17.62	48.57	20.6	16.07	0.05	0.0
SVR(kernel='poly', degree=5)	20	16.89	49.43	19.95	13.57	0.12	0.06
SVR(kernel='poly', degree=5)	10	21.77	151.41	33.09	16.8	-1.57	-1.58
SVR(kernel='sigmoid')	197	17.4	49.17	20.75	16.35	-0.0	-0.01
SVR(kernel='sigmoid')	130	17.49	47.23	20.75	16.09	0.0	-0.01
SVR(kernel='sigmoid')	100	17.42	47.38	20.66	16.17	0.01	-0.0
SVR(kernel='sigmoid')	70	17.37	47.62	20.68	16.32	0.01	-0.01
SVR(kernel='sigmoid')	40	17.52	47.0	20.74	16.56	0.0	-0.01
SVR(kernel='sigmoid')	20	17.74	44.8	20.86	17.94	-0.02	-0.03
SVR(kernel='sigmoid')	10	17.68	46.74	20.81	16.72	-0.0	-0.02
RandomForestRegressor(n_estimators=10)	197	19.18	48.4	23.35	17.6	-0.18	-0.28
RandomForestRegressor(n_estimators=10)	130	19.82	41.8	22.16	18.35	-0.11	-0.16
RandomForestRegressor(n_estimators=10)	100	18.03	37.0	20.45	17.0	0.02	0.02
RandomForestRegressor(n_estimators=10)	70	18.34	48.9	22.2	19.1	-0.11	-0.16
RandomForestRegressor(n_estimators=10)	40	18.75	41.7	22.21	19.6	-0.11	-0.16
RandomForestRegressor(n_estimators=10)	20	18.57	43.2	21.93	19.25	-0.12	-0.13
RandomForestRegressor(n_estimators=10)	10	18.69	47.3	22.05	14.95	-0.13	-0.14
RandomForestRegressor(n_estimators=20)	197	18.0	40.4	21.47	16.92	-0.06	-0.09
RandomForestRegressor(n_estimators=20)	130	19.06	43.25	21.43	18.35	-0.03	-0.08
RandomForestRegressor(n_estimators=20)	100	19.15	41.15	21.73	19.0	-0.09	-0.11
RandomForestRegressor(n_estimators=20)	70	17.65	46.85	21.31	17.6	-0.03	-0.07
RandomForestRegressor(n_estimators=20)	40	19.66	49.9	23.48	21.38	-0.23	-0.3
RandomForestRegressor(n_estimators=20)	20	19.32	46.1	22.73	21.05	-0.2	-0.22

RandomForestRegressor(n_estimators=20)	10	18.61	47.35	22.1	18.7	-0.15	-0.15
RandomForestRegressor(n_estimators=50)	197	17.4	43.56	20.48	16.34	0.02	0.01
RandomForestRegressor(n_estimators=50)	130	18.61	47.06	21.52	18.27	-0.06	-0.09
RandomForestRegressor(n_estimators=50)	100	18.64	45.54	21.43	17.57	-0.07	-0.08
RandomForestRegressor(n_estimators=50)	70	18.8	45.08	21.87	18.55	-0.1	-0.13
RandomForestRegressor(n_estimators=50)	40	20.17	52.12	23.35	20.81	-0.26	-0.28
RandomForestRegressor(n_estimators=50)	20	19.16	45.24	22.4	21.96	-0.17	-0.18
RandomForestRegressor(n_estimators=50)	10	18.24	48.58	21.58	19.8	-0.1	-0.1
RandomForestRegressor(n_estimators=100)	197	16.69	44.65	20.13	17.2	0.05	0.05
RandomForestRegressor(n_estimators=100)	130	18.65	42.32	21.48	20.0	-0.07	-0.09
RandomForestRegressor(n_estimators=100)	100	18.46	44.78	21.29	18.63	-0.05	-0.07
RandomForestRegressor(n_estimators=100)	70	18.09	45.28	21.18	17.93	-0.05	-0.06
RandomForestRegressor(n_estimators=100)	40	19.68	52.66	22.59	19.18	-0.18	-0.2
RandomForestRegressor(n_estimators=100)	20	19.0	44.94	21.94	20.74	-0.12	-0.13
RandomForestRegressor(n_estimators=100)	10	17.71	48.45	21.1	19.56	-0.05	-0.05
RandomForestRegressor(n_estimators=300)	197	16.33	47.28	19.73	15.59	0.09	0.08
RandomForestRegressor(n_estimators=300)	130	18.8	42.27	21.66	17.98	-0.09	-0.1
RandomForestRegressor(n_estimators=300)	100	18.64	46.49	21.49	19.96	-0.08	-0.09
RandomForestRegressor(n_estimators=300)	70	18.12	49.35	21.3	17.5	-0.06	-0.07
RandomForestRegressor(n_estimators=300)	40	19.17	54.27	22.32	17.03	-0.16	-0.17
RandomForestRegressor(n_estimators=300)	20	18.21	47.21	21.37	19.54	-0.07	-0.08
RandomForestRegressor(n_estimators=300)	10	17.62	50.79	21.11	19.3	-0.05	-0.05
RandomForestRegressor(n_estimators=500)	197	16.34	45.91	19.64	15.46	0.1	0.09
RandomForestRegressor(n_estimators=500)	130	18.58	44.04	21.64	17.86	-0.09	-0.1
RandomForestRegressor(n_estimators=500)	100	18.41	48.2	21.33	19.2	-0.06	-0.07
RandomForestRegressor(n_estimators=500)	70	17.98	50.88	21.29	16.82	-0.06	-0.07
RandomForestRegressor(n_estimators=500)	40	18.74	55.33	22.05	16.85	-0.14	-0.14
RandomForestRegressor(n_estimators=500)	20	18.11	47.69	21.21	19.29	-0.05	-0.06
RandomForestRegressor(n_estimators=500)	10	17.73	51.34	21.26	18.52	-0.06	-0.06
RandomForestRegressor(n_estimators=700)	197	16.3	45.34	19.6	15.21	0.1	0.1



RandomForestRegressor(n_estimators=700)	130	18.56	43.12	21.64	18.0	-0.09	-0.1
RandomForestRegressor(n_estimators=700)	100	18.48	47.84	21.41	19.59	-0.07	-0.08
RandomForestRegressor(n_estimators=700)	70	18.18	50.42	21.5	17.11	-0.08	-0.09
RandomForestRegressor(n_estimators=700)	40	18.88	55.17	22.18	17.34	-0.15	-0.16
RandomForestRegressor(n_estimators=700)	20	18.13	48.29	21.35	19.45	-0.07	-0.07
RandomForestRegressor(n_estimators=700)	10	17.71	51.45	21.18	18.72	-0.06	-0.06
RandomForestRegressor(n_estimators=1000)	197	16.32	45.56	19.63	15.1	0.1	0.09
RandomForestRegressor(n_estimators=1000)	130	18.53	43.23	21.59	18.24	-0.08	-0.1
RandomForestRegressor(n_estimators=1000)	100	18.41	47.51	21.37	19.26	-0.06	-0.07
RandomForestRegressor(n_estimators=1000)	70	18.16	49.9	21.47	17.44	-0.08	-0.09
RandomForestRegressor(n_estimators=1000)	40	18.87	55.01	22.14	17.27	-0.15	-0.15
RandomForestRegressor(n_estimators=1000)	20	18.17	47.23	21.29	19.87	-0.06	-0.07
RandomForestRegressor(n_estimators=1000)	10	17.7	50.97	21.2	18.89	-0.06	-0.06
LinearRegression()	197	38.88	101.62	49.41	32.78	-4.47	-4.75
LinearRegression()	130	19.25	55.55	23.97	19.09	-0.26	-0.35
LinearRegression()	100	1868.74	26315.95	4890.43	543.84	-55007.1	-56319.4
LinearRegression()	70	56.21	312.17	98.04	24.54	-19.16	-21.63
LinearRegression()	40	26.29	134.3	36.34	22.04	-1.93	-2.11
LinearRegression()	20	21.28	57.61	25.15	20.69	-0.46	-0.49
LinearRegression()	10	19.15	43.32	21.8	18.89	-0.12	-0.12
PolynomialRegression(degree=3)	197	24.14	89.43	33.78	17.83	-1.68	-1.69
PolynomialRegression(degree=3)	130	16.97	41.83	20.77	15.43	0.02	-0.02
PolynomialRegression(degree=3)	100	17.75	50.82	21.44	17.65	-0.08	-0.08
PolynomialRegression(degree=3)	70	17.75	57.15	22.6	14.87	-0.16	-0.2
PolynomialRegression(degree=3)	40	25.07	108.67	33.9	18.57	-1.7	-1.71
PolynomialRegression(degree=3)	20	259.21	3543.03	686.14	42.78	-1072.79	-1107.66
PolynomialRegression(degree=3)	10	35674.71	506956.2	105673.9	202.48	-2.6E+07	-2.6E+07