



Bayesian Modeling for Software Quality

By

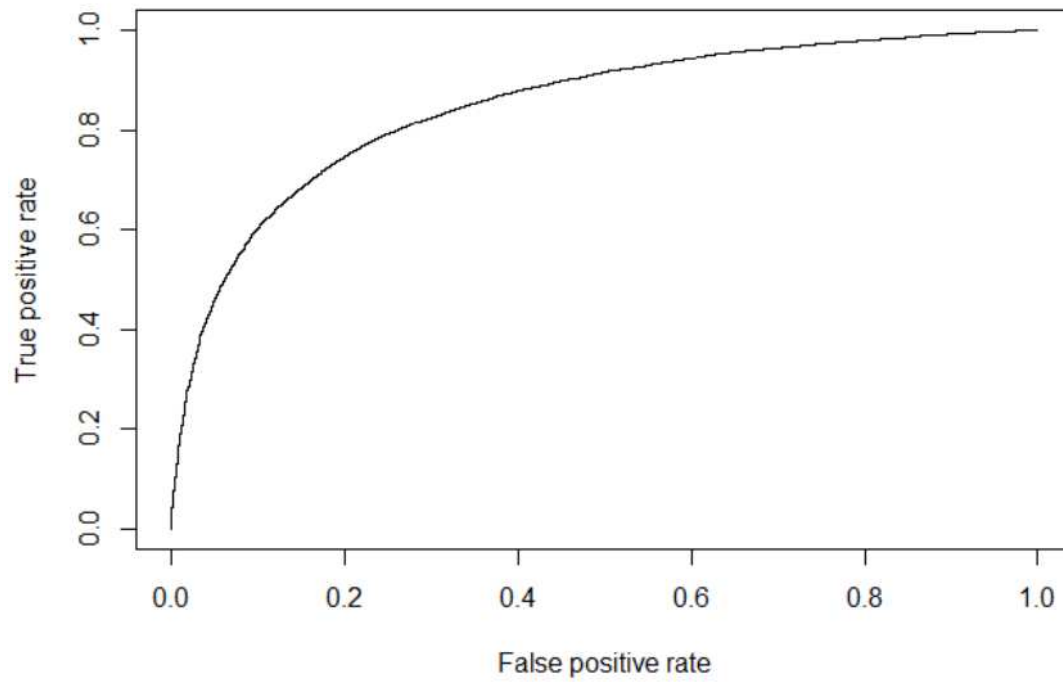
Mohammad Sharif Hanif

Department of Computer Science

Brunel University London

22 June 2015

Text Mining



```
> print(attributes(performance(eval, 'auc'))$y.values[[1]])  
[1] 0.8478918
```

```
> #distribution of bug and not-bug  
> table(report.raw$category)
```

0	1
15954	13539

sparse matrix with most frequent words that appear at 10 times in the corpus.

Eclipse Defect Mining

Relation: eclipse-jhawke																																				
No.	id String	class String	jClass String	method String	jMethod String	sLine Numeric	eLine Numeric	COMP Numeric	NOCL Numeric	NOS Numeric	HLTH Numeric	HVOC Numeric	HEFF Numeric	HBUG Numeric	CREF Numeric	XMET Numeric	LMET Numeric	NLOC Numeric	NOC Numeric	NOA Numeric	MOD Numeric	HDIF Numeric	VDEC Numeric	EXCT Numeric	EXCR Numeric	CAST Numeric	TDN Numeric	HVOL Numeric	NAND Numeric	VREF Numeric	MDN Numeric	NOPR Numeric	NEXP Numeric	LOOP Numeric	defective Nominal	
1	1000	org.eclipse.jdt.cor...	org.e...	execute	execute	30.0	60.0	9.0	5.0	25.0	246.0	63.0	3205...	0.49	8.0	9.0	1.0	31.0	5.0	0.0	1.0	21.8	5.0	1.0	2.0	1.0	10.0	1470....	109.0	19.0	4.0	137.0	40.0	1.0	0	
2	1001	org.eclipse.jdt.cor...	org.e...	check...	checkC...	62.0	71.0	3.0	0.0	10.0	64.0	33.0	3449.3	0.11	3.0	3.0	0.0	10.0	0.0	1.0	1.0	10.68	2.0	0.0	0.0	0.0	2.0	322.84	29.0	5.0	2.0	35.0	9.0	1.0	0	
3	1002	org.eclipse.jdt.cor...	org.e...	setFile	setFile	73.0	75.0	1.0	0.0	2.0	13.0	12.0	163.12	0.2	1.0	0.0	0.0	3.0	0.0	1.0	1.0	3.5	0.0	0.0	0.0	0.0	0.0	46.6	7.0	2.0	0.0	6.0	1.0	0.0	0	
4	1003	org.eclipse.jdt.cor...	org.e...	setPro...	setPro...	77.0	79.0	1.0	0.0	2.0	13.0	12.0	163.12	0.2	1.0	0.0	0.0	3.0	0.0	1.0	1.0	3.5	0.0	0.0	0.0	0.0	0.0	46.6	7.0	2.0	0.0	6.0	1.0	0.0	0	
5	1004	org.eclipse.jdt.cor...	org.e...	execute	execute	51.0	71.0	4.0	9.0	15.0	183.0	68.0	1624...	0.37	13.0	13.0	1.0	20.0	5.0	0.0	1.0	14.58	7.0	1.0	3.0	1.0	4.0	1114.1	81.0	23.0	2.0	102.0	35.0	0.0	1	
6	1005	org.eclipse.jdt.cor...	org.e...	setupJ...	setupJ...	74.0	327.0	40.0	108.0	116.0	843.0	126.0	2158...	1.96	13.0	23.0	3.0	162.0	71.0	0.0	1.0	36.7	12.0	1.0	3.0	3.0	59.0	5881....	347.0	120.0	3.0	496.0	116.0	0.0	1	
7	1006	org.eclipse.jdt.cor...	org.e...	addExt...	addExt...	336.0	356.0	5.0	9.0	18.0	121.0	46.0	1057...	0.22	5.0	11.0	0.0	20.0	3.0	2.0	1.0	15.83	4.0	0.0	0.0	0.0	5.0	668.35	54.0	16.0	2.0	67.0	18.0	1.0	0	
8	1007	org.eclipse.jdt.cor...	org.e...	(init)	AntAda...	33.0	35.0	1.0	1.0	1.0	4.0	4.0	12.0	0.0	0.0	0.0	0.0	2.0	1.0	0.0	1.0	1.5	0.0	0.0	0.0	0.0	0.0	8.0	1.0	0.0	0.0	3.0	0.0	0.0	0	
9	1008	org.eclipse.jdt.cor...	org.e...	getStri...	getString	37.0	43.0	2.0	0.0	4.0	29.0	17.0	1016.3	0.4	2.0	1.0	0.0	7.0	0.0	1.0	2.0	8.57	0.0	0.0	1.0	0.0	2.0	118.54	12.0	3.0	1.0	17.0	3.0	0.0	0	
10	1010	org.eclipse.jdt.cor...	org.e...	(init)	Classp...	0.0	38.0	1.0	1.0	2.0	15.0	12.0	134.44	0.2	2.0	0.0	0.0	3.0	1.0	2.0	0.0	2.5	0.0	0.0	0.0	0.0	0.0	53.77	10.0	4.0	0.0	5.0	3.0	0.0	0	

#	Act...	Align	Column...	Description	Co...	Warning	Danger	Class	Type
0	<input checked="" type="checkbox"/>	Left	Name	Name of method	N/A	0.0	0.0	com.virtualma...	Stri...
1	<input checked="" type="checkbox"/>	Right	COMP	Cyclomatic Complexity	>	6.0	8.0	com.virtualma...	Inte...
2	<input checked="" type="checkbox"/>	Right	NOCL	Number of comment lines	>	0.0	0.0	com.virtualma...	Inte...
3	<input checked="" type="checkbox"/>	Right	LOOP	Number of loops	>	0.0	0.0	com.virtualma...	Inte...
4	<input checked="" type="checkbox"/>	Right	MDN	Maximum depth of nesting	>	0.0	0.0	com.virtualma...	Inte...
5	<input checked="" type="checkbox"/>	Right	NLOC	Number of Lines of Code in the method	>	0.0	0.0	com.virtualma...	Inte...
6	<input checked="" type="checkbox"/>	Right	EXCT	Number of exceptions thrown	>	0.0	0.0	com.virtualma...	Inte...
7	<input checked="" type="checkbox"/>	Right	NOA	Number of arguments	>	0.0	0.0	com.virtualma...	Inte...

Data

jhawke

data

11123 obs. of 35 variables

```
> table(jhawke$defective)
```

0	1
10686	437

ProblemsJavadocDeclarationConsoleJHawk Metrics

Dashboard GaugesDashboard TablesSystem DetailsSystem PackagesPackage DetailsClass DetailsAll ClassesAll MethodsExport

Save in CSV FormatSave in HTML FormatSave in XML Format

Packages

Classes

Generate Details for

Generate Details for each

Select Active/All Metrics....

Include in CSV line

Select Sort Field for

Export data in raw format

CSV Separator

Save to fileC:\Users\sharifan\Desktop\eclipse-falut-jhawk-v3.csvBrowse...

Save CSV

Naive Bayes

```
␣
Time taken to build model: 0.1 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      10381           93.3291 %
Incorrectly Classified Instances     742           6.6709 %
Kappa statistic                     0.2566
Mean absolute error                  0.0664
Root mean squared error              0.2551
Relative absolute error              87.8351 %
Root relative squared error          131.2927 %
Total Number of Instances           11123

=== Detailed Accuracy By Class ===

                TP Rate   FP Rate   Precision   Recall   F-Measure   ROC Area   Class
                0.957     0.652     0.973     0.957     0.965     0.799     0
                0.348     0.043     0.25      0.348     0.291     0.798     1
Weighted Avg.   0.933     0.628     0.944     0.933     0.939     0.799

=== Confusion Matrix ===

      a      b  <-- classified as
10229  457 |      a = 0
   285  152 |      b = 1
```

Logistic

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      10692           96.1251 %
Incorrectly Classified Instances     431           3.8749 %
Kappa statistic                     0.1374
Mean absolute error                  0.0656
Root mean squared error              0.1836
Relative absolute error              86.8586 %
Root relative squared error          94.5036 %
Total Number of Instances          11123

=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.997	0.915	0.964	0.997	0.98	0.81	0
	0.085	0.003	0.544	0.085	0.147	0.81	1
Weighted Avg.	0.961	0.879	0.947	0.961	0.947	0.81	

```
=== Confusion Matrix ===

  a    b  <-- classified as
10655  31 |    a = 0
  400   37 |    b = 1
```

LibSVM

```
=== Classifier model (full training set) ===

LibSVM wrapper, original code by Yasser EL-Manzalawy (= WLSVM)

Time taken to build model: 86.69 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      10685           96.0622 %
Incorrectly Classified Instances     438           3.9378 %
Kappa statistic                     0.0916
Mean absolute error                  0.0394
Root mean squared error              0.1984
Relative absolute error              52.1053 %
Root relative squared error          102.1408 %
Total Number of Instances          11123

=== Detailed Accuracy By Class ===

                TP Rate    FP Rate    Precision    Recall    F-Measure    ROC Area    Class
                0.998      0.945      0.963      0.998      0.98         0.526      0
                0.055      0.002      0.49       0.055      0.099        0.526      1
Weighted Avg.    0.961      0.908      0.944      0.961      0.945        0.526

=== Confusion Matrix ===

      a      b  <-- classified as
10661    25 |      a = 0
   413    24 |      b = 1
```

Eclipse Defect Mining

Relation: eclipse-jhawk																																			
No.	id	class	jClass	method	jMethod	sLine	eLine	COMP	NOCL	NOS	HLTH	HVOC	HEFF	HBUG	CREF	XMET	LMET	NLOC	NOC	NOA	MOD	HDIF	VDEC	EXCT	EXCR	CAST	TDN	HVOL	NAND	VREF	MDN	NOPR	NEXP	LOOP	defective
	String	String	String	String	String	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Nominal
1	1000	org.eclipse.jdt.cor...	org.e...	execute	execute	30.0	60.0	9.0	5.0	25.0	246.0	63.0	3205...	0.49	8.0	9.0	1.0	31.0	5.0	0.0	1.0	21.8	5.0	1.0	2.0	1.0	10.0	1470....	109.0	19.0	4.0	137.0	40.0	1.0	0
2	1001	org.eclipse.jdt.cor...	org.e...	check...	checkC...	62.0	71.0	3.0	0.0	10.0	64.0	33.0	3449.3	0.11	3.0	3.0	0.0	10.0	0.0	1.0	1.0	10.68	2.0	0.0	0.0	0.0	2.0	322.84	29.0	5.0	2.0	35.0	9.0	1.0	0
3	1002	org.eclipse.jdt.cor...	org.e...	setFile	setFile	73.0	75.0	1.0	0.0	2.0	13.0	12.0	163.12	0.2	1.0	0.0	0.0	3.0	0.0	1.0	1.0	3.5	0.0	0.0	0.0	0.0	0.0	46.6	7.0	2.0	0.0	6.0	1.0	0.0	0
4	1003	org.eclipse.jdt.cor...	org.e...	setPro...	setPro...	77.0	79.0	1.0	0.0	2.0	13.0	12.0	163.12	0.2	1.0	0.0	0.0	3.0	0.0	1.0	1.0	3.5	0.0	0.0	0.0	0.0	0.0	46.6	7.0	2.0	0.0	6.0	1.0	0.0	0
5	1004	org.eclipse.jdt.cor...	org.e...	execute	execute	51.0	71.0	4.0	9.0	15.0	183.0	68.0	1624...	0.37	13.0	13.0	1.0	20.0	5.0	0.0	1.0	14.58	7.0	1.0	3.0	1.0	4.0	1114.1	81.0	23.0	2.0	102.0	35.0	0.0	1
6	1005	org.eclipse.jdt.cor...	org.e...	setupJ...	setupJ...	74.0	327.0	40.0	108.0	116.0	843.0	126.0	2158...	1.96	13.0	23.0	3.0	162.0	71.0	0.0	1.0	36.7	12.0	1.0	3.0	3.0	59.0	5881....	347.0	120.0	3.0	496.0	116.0	0.0	1
7	1006	org.eclipse.jdt.cor...	org.e...	addExt...	addExt...	336.0	356.0	5.0	9.0	18.0	121.0	46.0	1057...	0.22	5.0	11.0	0.0	20.0	3.0	2.0	1.0	15.83	4.0	0.0	0.0	0.0	5.0	668.35	54.0	16.0	2.0	67.0	18.0	1.0	0
8	1007	org.eclipse.jdt.cor...	org.e...	(init)	AntAda...	33.0	35.0	1.0	1.0	1.0	4.0	4.0	12.0	0.0	0.0	0.0	0.0	2.0	1.0	0.0	1.0	1.5	0.0	0.0	0.0	0.0	8.0	1.0	0.0	0.0	3.0	0.0	0.0	0	
9	1008	org.eclipse.jdt.cor...	org.e...	getStri...	getString	37.0	43.0	2.0	0.0	4.0	29.0	17.0	1016.3	0.4	2.0	1.0	0.0	7.0	0.0	1.0	2.0	8.57	0.0	0.0	1.0	0.0	2.0	118.54	12.0	3.0	1.0	17.0	3.0	0.0	0
10	1010	org.eclipse.jdt.cor...	org.e...	(init)	Classp...	0.0	38.0	1.0	1.0	2.0	15.0	12.0	134.44	0.2	2.0	0.0	0.0	3.0	1.0	2.0	0.0	2.5	0.0	0.0	0.0	0.0	53.77	10.0	4.0	0.0	5.0	3.0	0.0	0	

#	Act...	Align	Column...	Description	Co...	Warning	Danger	Class	Type
0	<input checked="" type="checkbox"/>	Left	Name	Name of method	N/A	0.0	0.0	com.virtualma...	Stri...
1	<input checked="" type="checkbox"/>	Right	COMP	Cyclomatic Complexity	>	6.0	8.0	com.virtualma...	Inte...
2	<input checked="" type="checkbox"/>	Right	NOCL	Number of comment lines	>	0.0	0.0	com.virtualma...	Inte...
3	<input checked="" type="checkbox"/>	Right	LOOP	Number of loops	>	0.0	0.0	com.virtualma...	Inte...
4	<input checked="" type="checkbox"/>	Right	MDN	Maximum depth of nesting	>	0.0	0.0	com.virtualma...	Inte...
5	<input checked="" type="checkbox"/>	Right	NLOC	Number of Lines of Code in the method	>	0.0	0.0	com.virtualma...	Inte...
6	<input checked="" type="checkbox"/>	Right	EXCT	Number of exceptions thrown	>	0.0	0.0	com.virtualma...	Inte...
7	<input checked="" type="checkbox"/>	Right	NOA	Number of arguments	>	0.0	0.0	com.virtualma...	Inte...

Ongoing task: feature selection for text mining and eclipse defects

Next meeting: In 1-2 weeks. Result Expected: ?