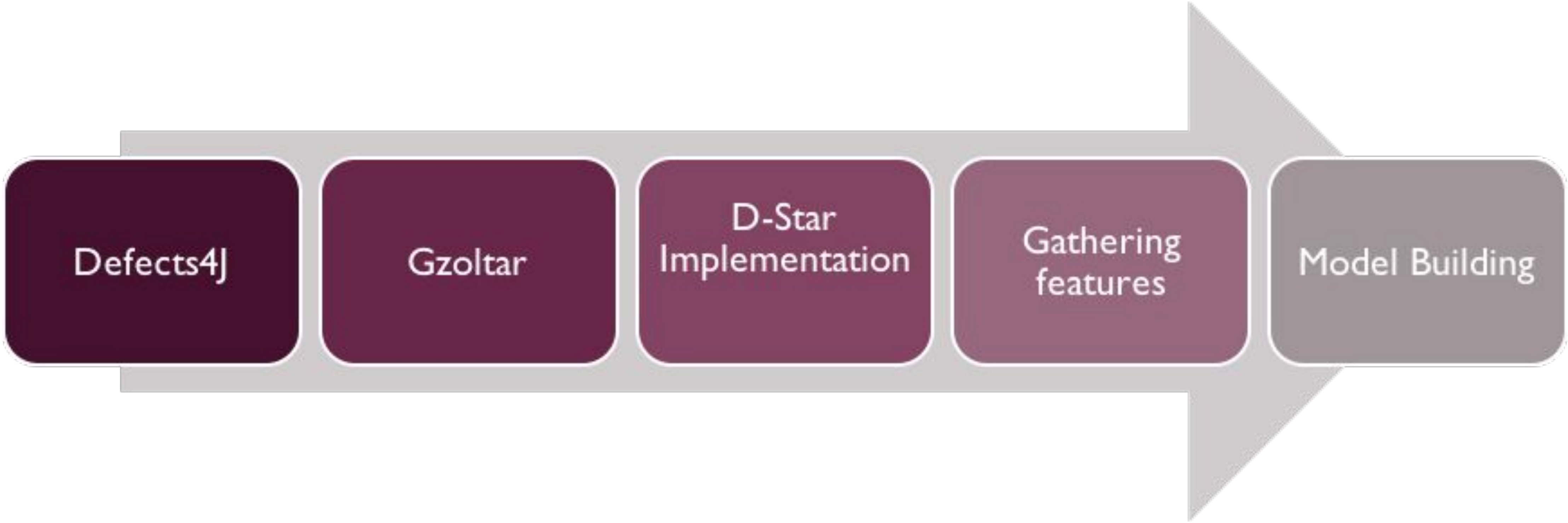
PRAKTIKUM: EXPLORING THE RELATIONSHIP BETWEEN DESIGN METRICS AND SOFTWARE DIAGNOSABILITY USING MACHINE LEARNING

HANDE KARATAY & FURKAN MERT ALGAN

JULY 13, 2018

What have we done?



Defects4J

Gzoltar

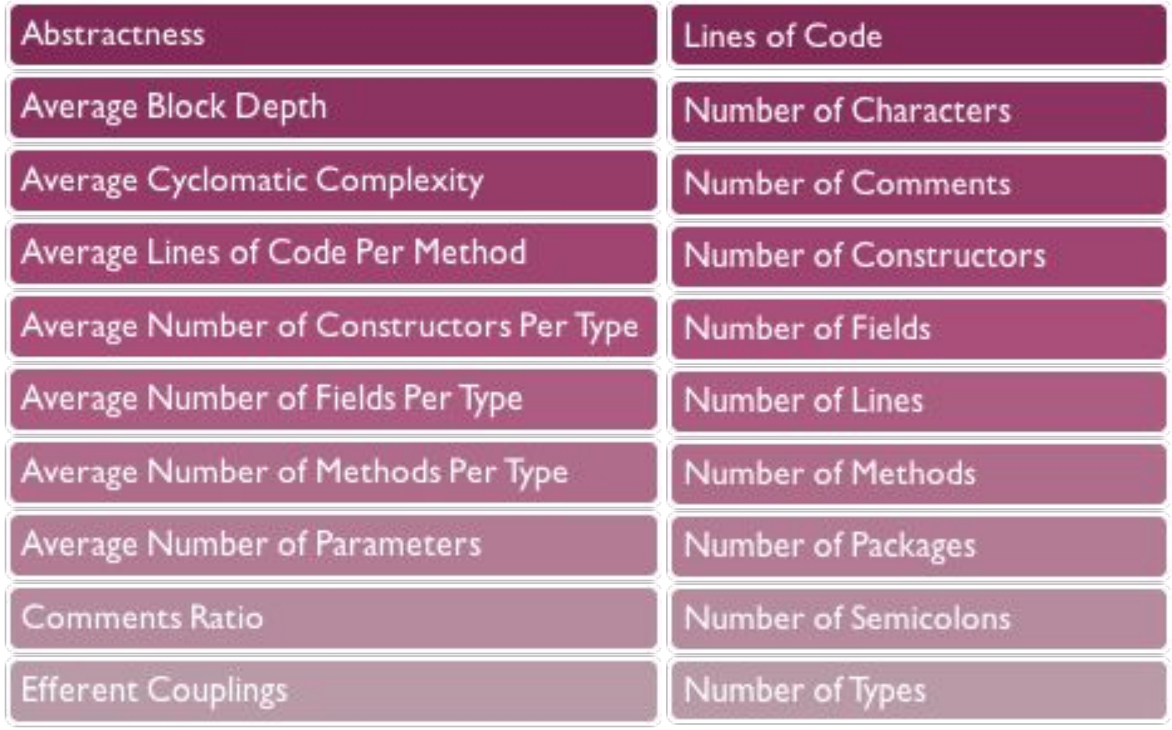
D-Star Implementation

Gathering features

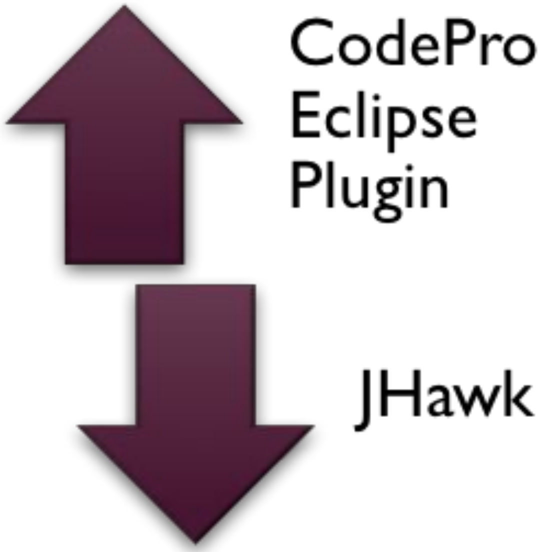
Model Building

Gathering Features: Static Features

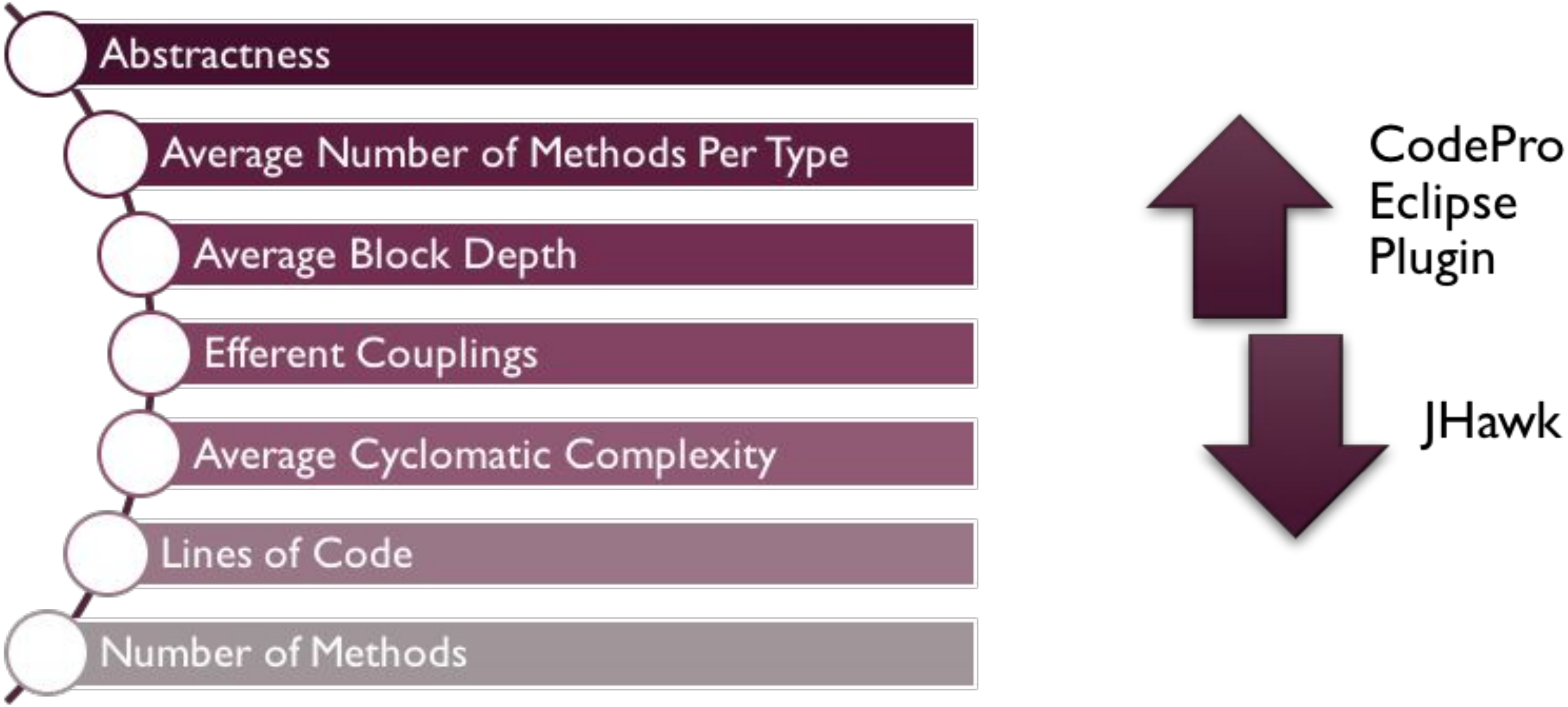
At first...



- Abstractness
- Average Block Depth
- Average Cyclomatic Complexity
- Average Lines of Code Per Method
- Average Number of Constructors Per Type
- Average Number of Fields Per Type
- Average Number of Methods Per Type
- Average Number of Parameters
- Comments Ratio
- Efferent Couplings
- Lines of Code
- Number of Characters
- Number of Comments
- Number of Constructors
- Number of Fields
- Number of Lines
- Number of Methods
- Number of Packages
- Number of Semicolons
- Number of Types



Gathering Features: Static Features

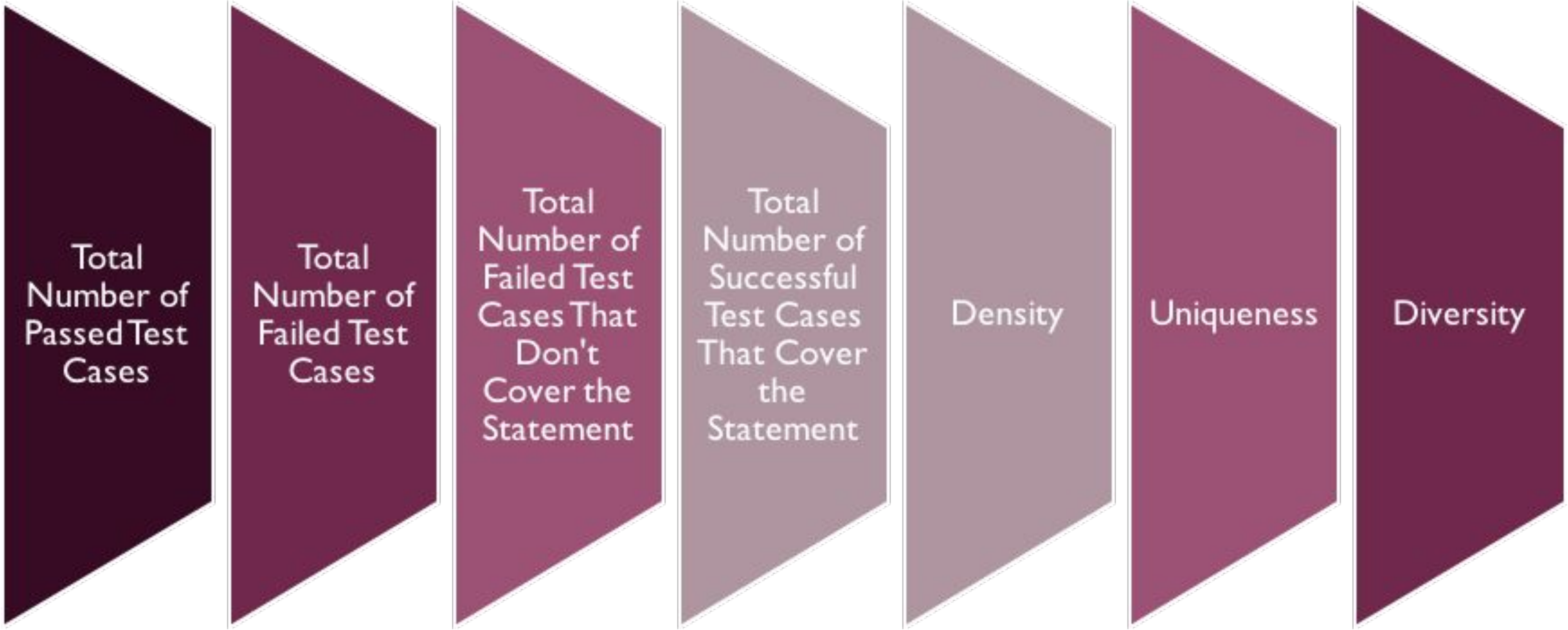


- Abstractness
- Average Number of Methods Per Type
- Average Block Depth
- Efferent Couplings
- Average Cyclomatic Complexity
- Lines of Code
- Number of Methods

CodePro Eclipse Plugin

JHawk

Gathering Features: Test Suite Characteristics



Total Number of Passed Test Cases

Total Number of Failed Test Cases

Total Number of Failed Test Cases That Don't Cover the Statement

Total Number of Successful Test Cases That Cover the Statement

Density

Uniqueness

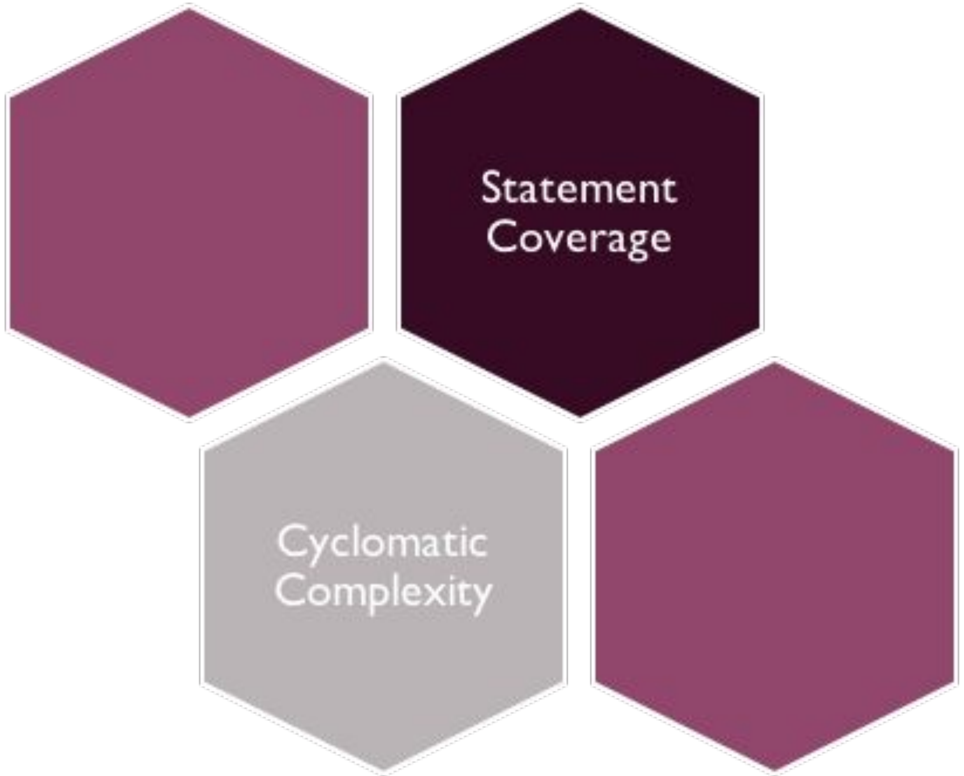
Diversity

Gathering Features: Bug Characteristics

Extracted from the ARP Platform

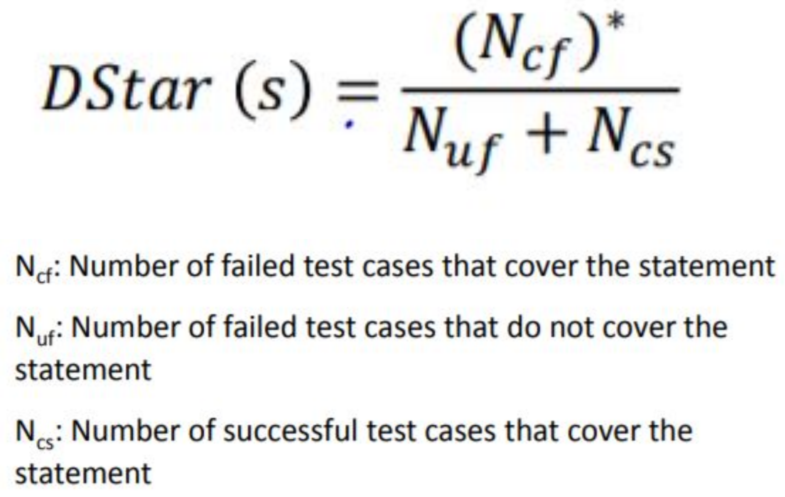
- File Count
- Line Count
- Statement Coverage/Dynamic
- Relevant Test Count
- Triggering Test Count

Gathering Features: Dynamic Features



Our Target Value

- Use **D-Star** algorithm to rank methods
  - 1: Rank 5 between 200 methods
  - 2: Rank 5 between 100 methods



$$DStar\ (s) = \frac{(N_{cf})^*}{N_{uf} + N_{cs}}$$

N_{cf}: Number of failed test cases that cover the statement

N_{uf}: Number of failed test cases that do not cover the statement

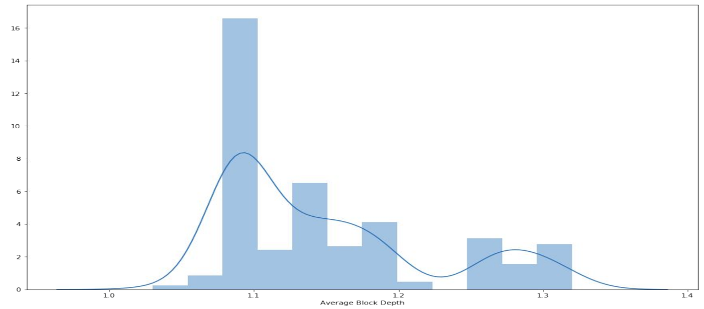
N_{cs}: Number of successful test cases that cover the statement

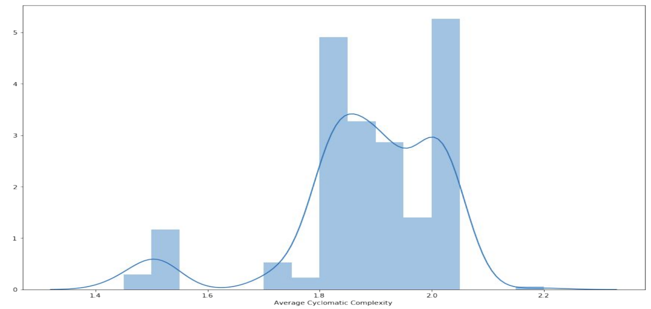
- Proposed Solution: Normalize

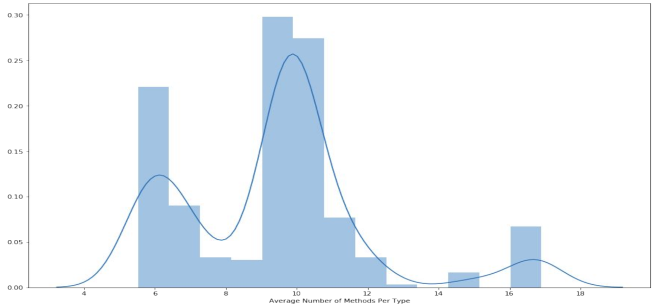
$$target\ value = \frac{Dstar\ rank}{number\ of\ methods\ in\ buggy\ version}$$

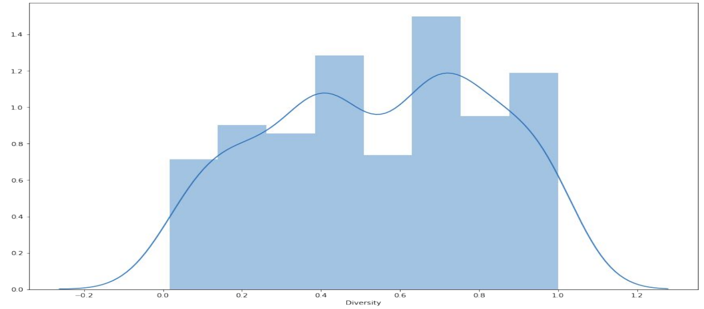
Model Building

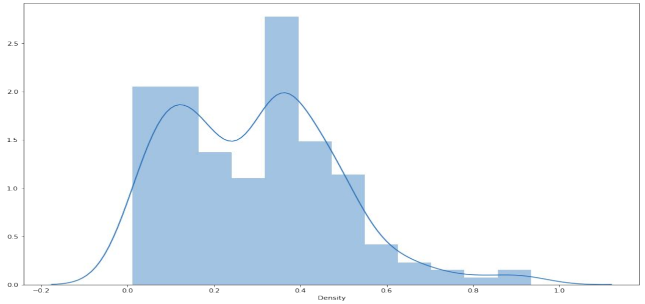
Metric Statistics

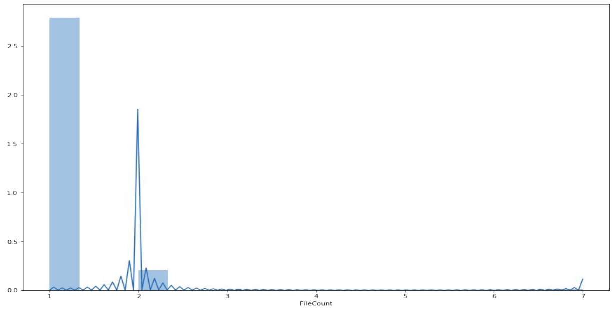


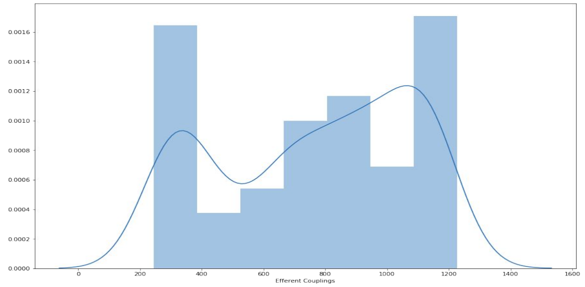


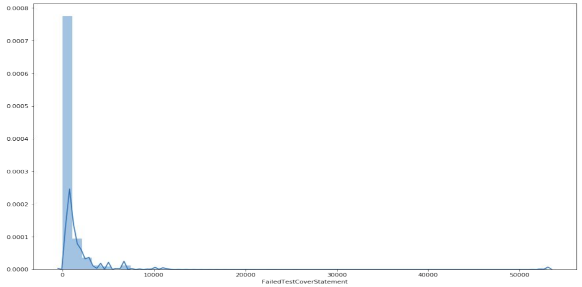


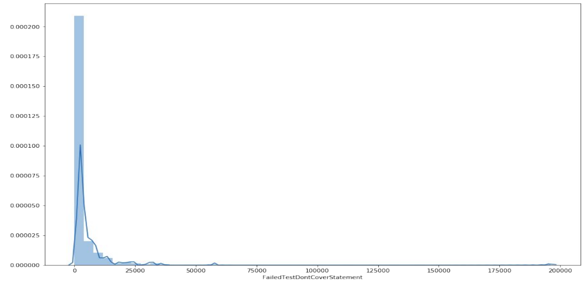


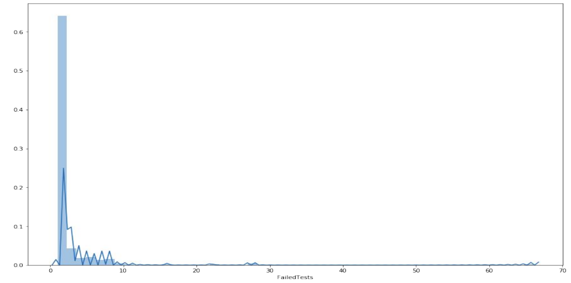




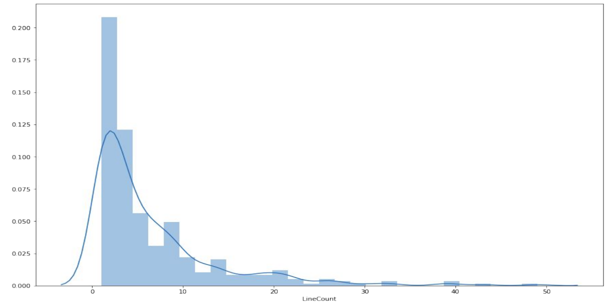


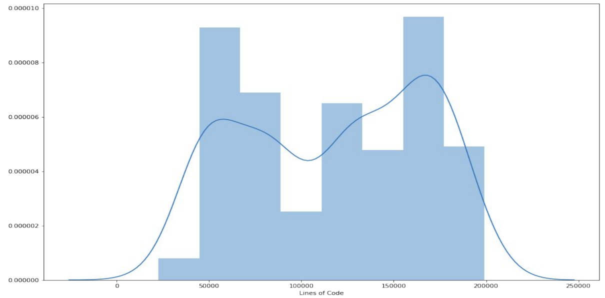


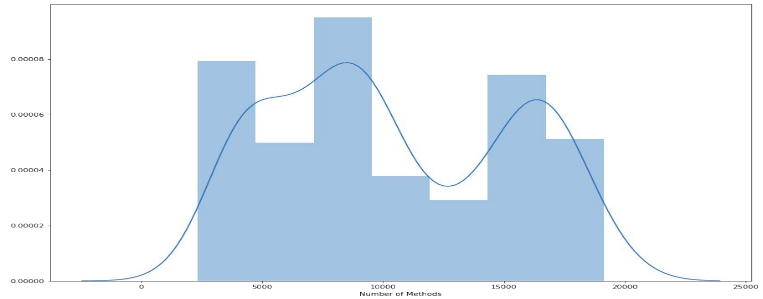


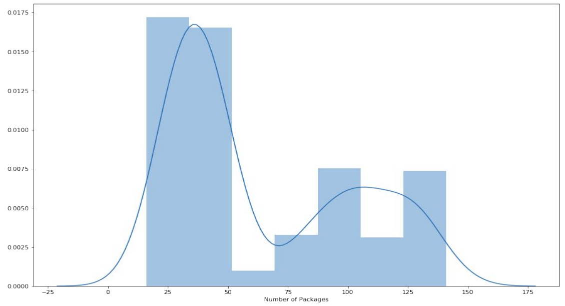


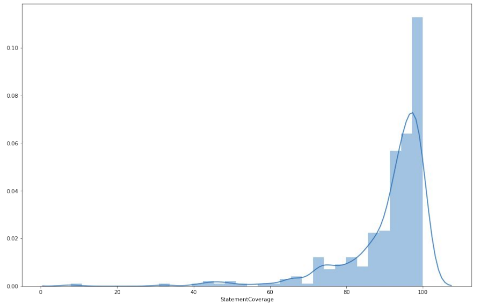
Metric Statistics

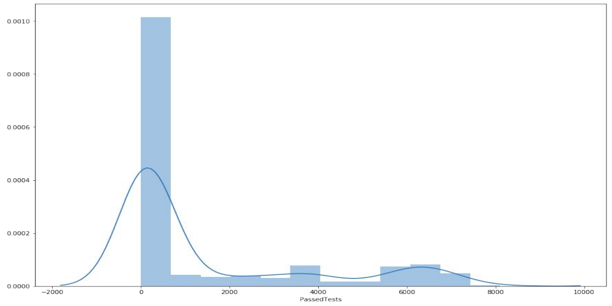


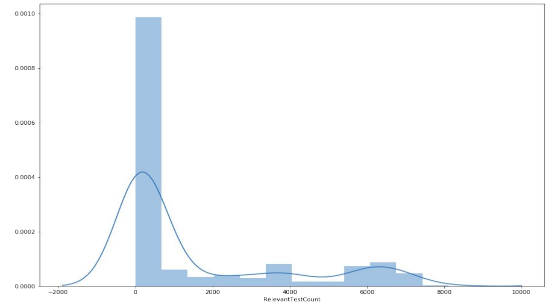


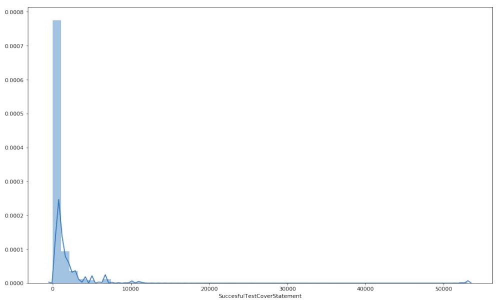


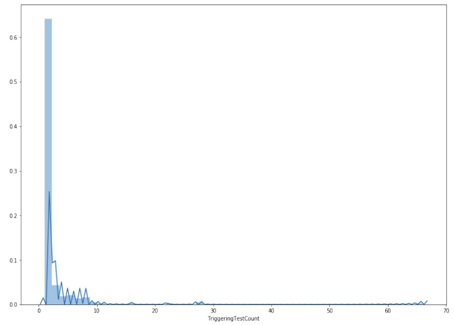


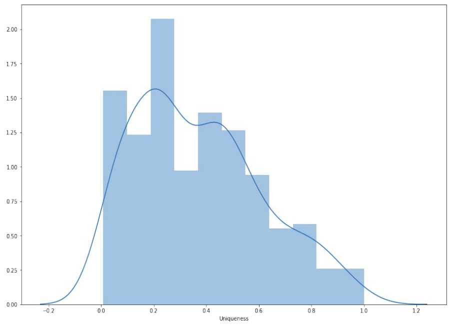


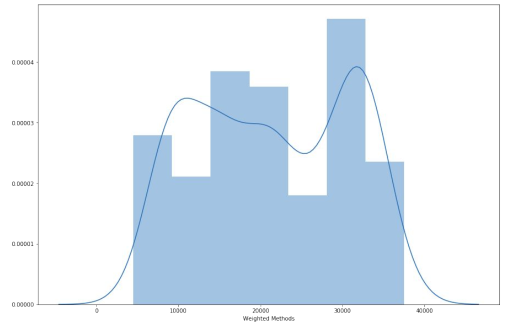






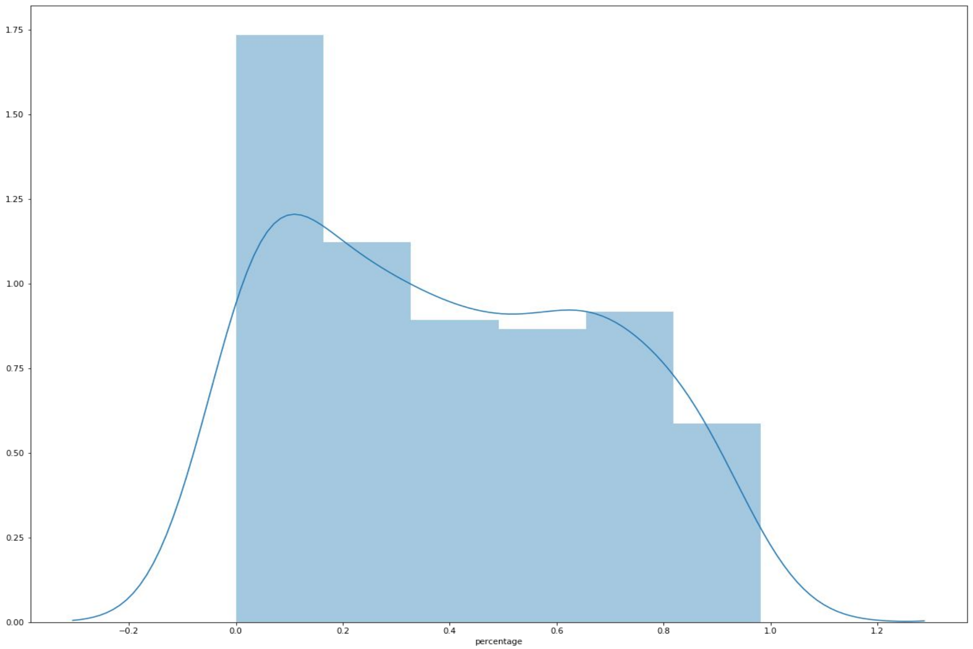




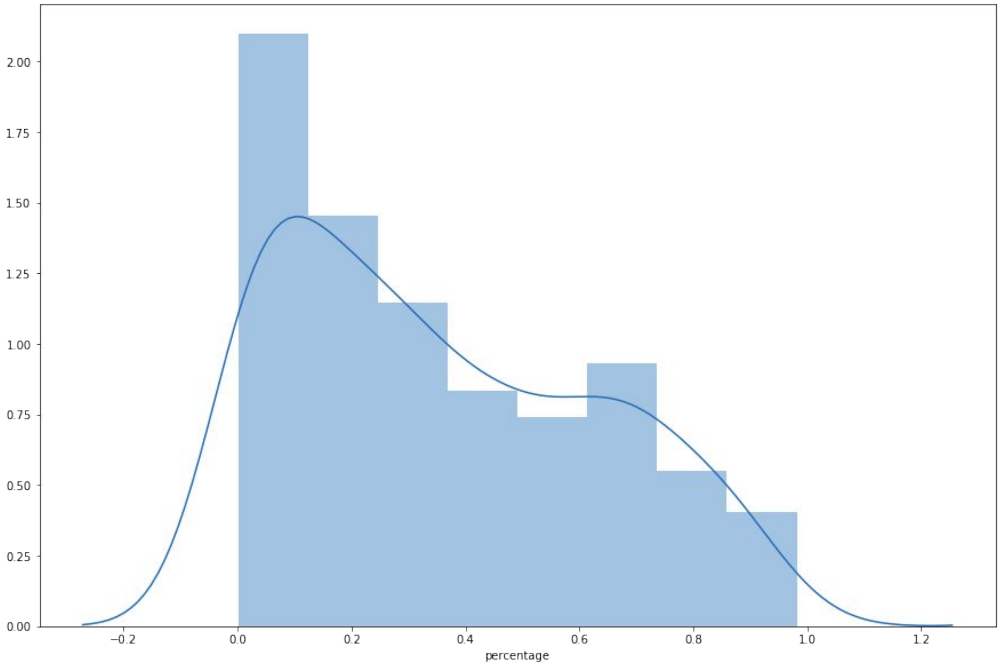


Target Statistics

With More Features

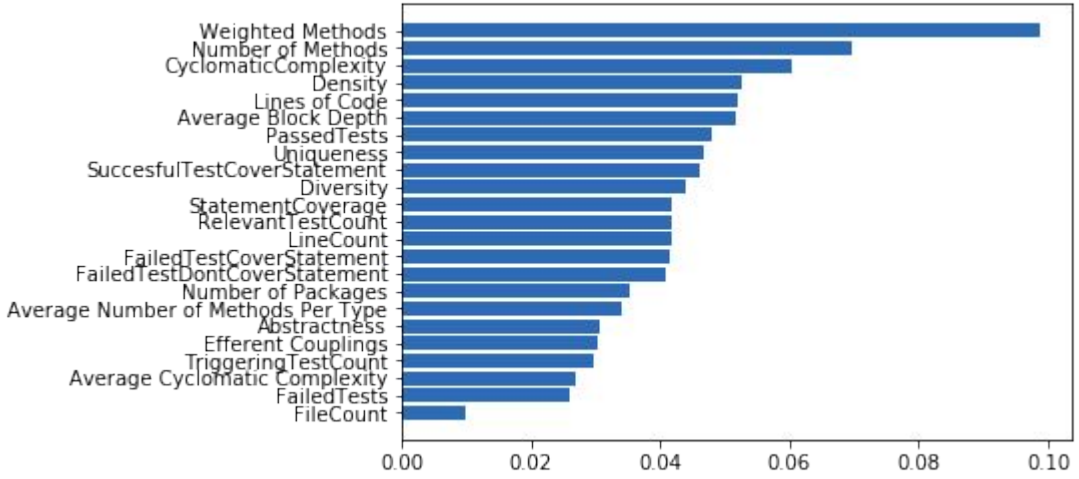


With More Data

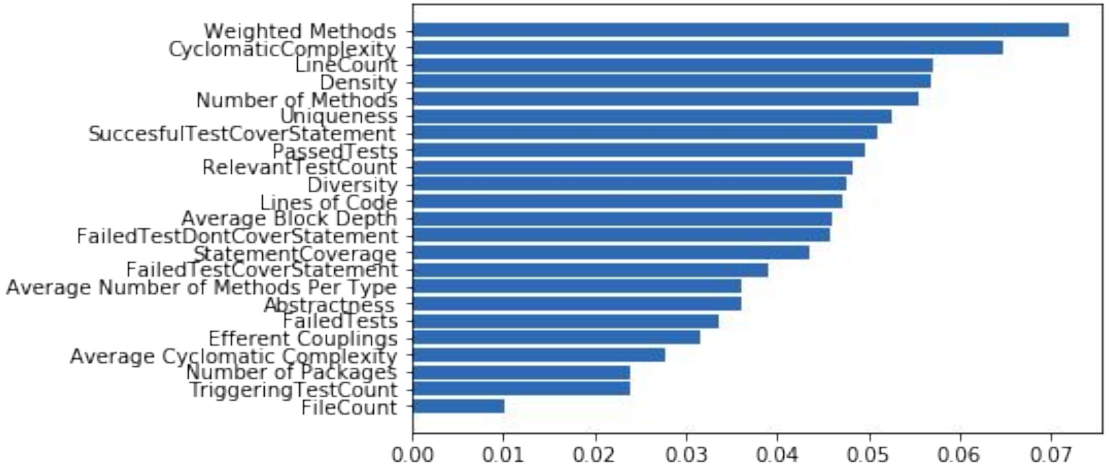


Feature Selection

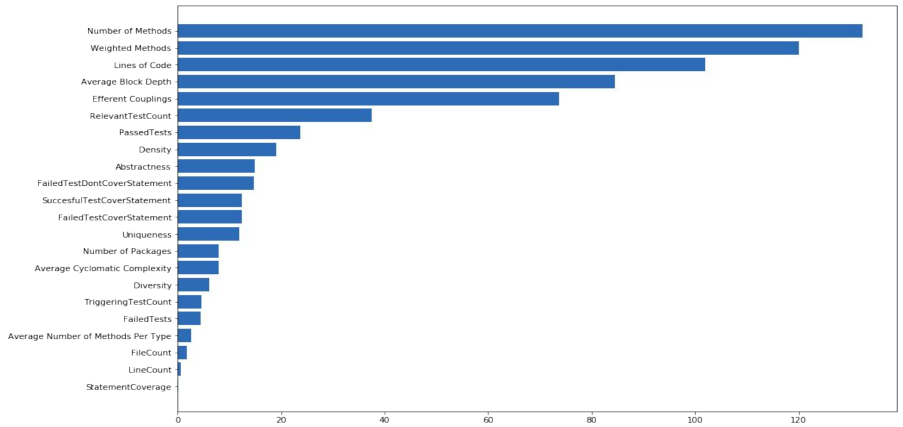
4 classes



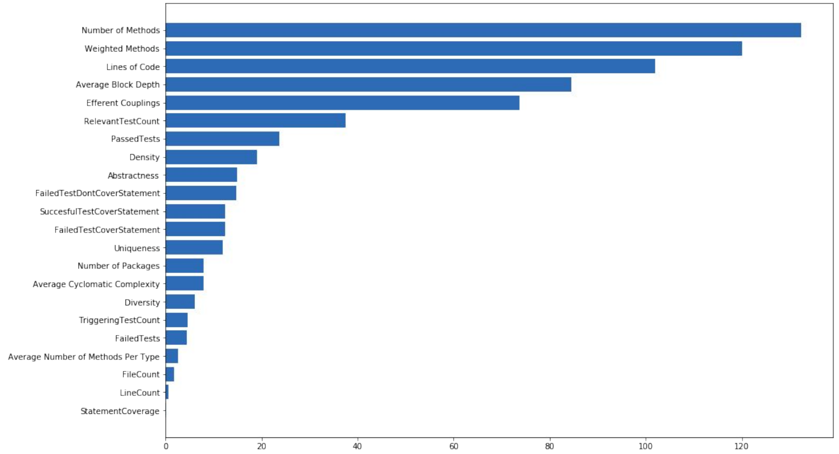
6 Classes



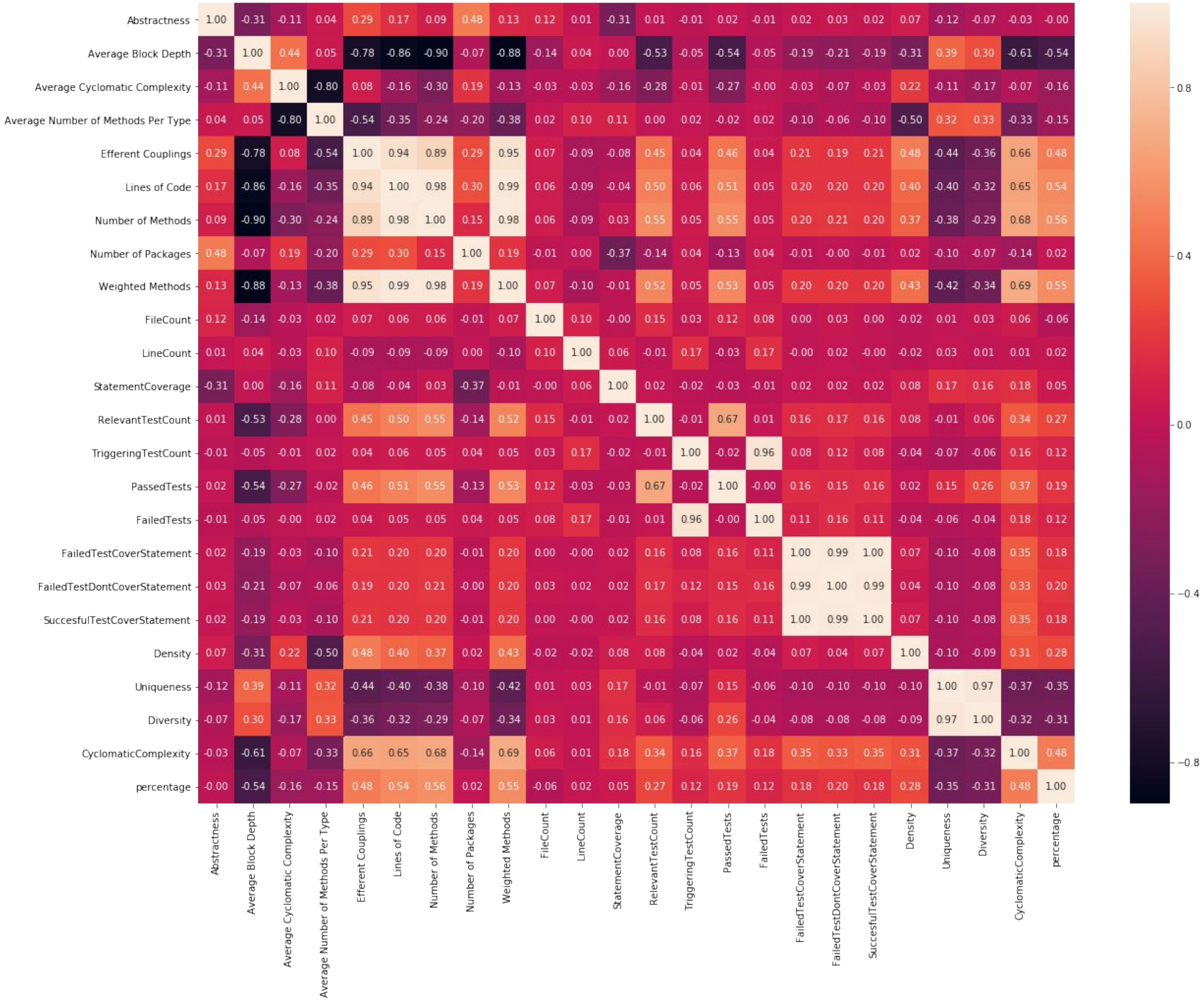
F Test



P Test



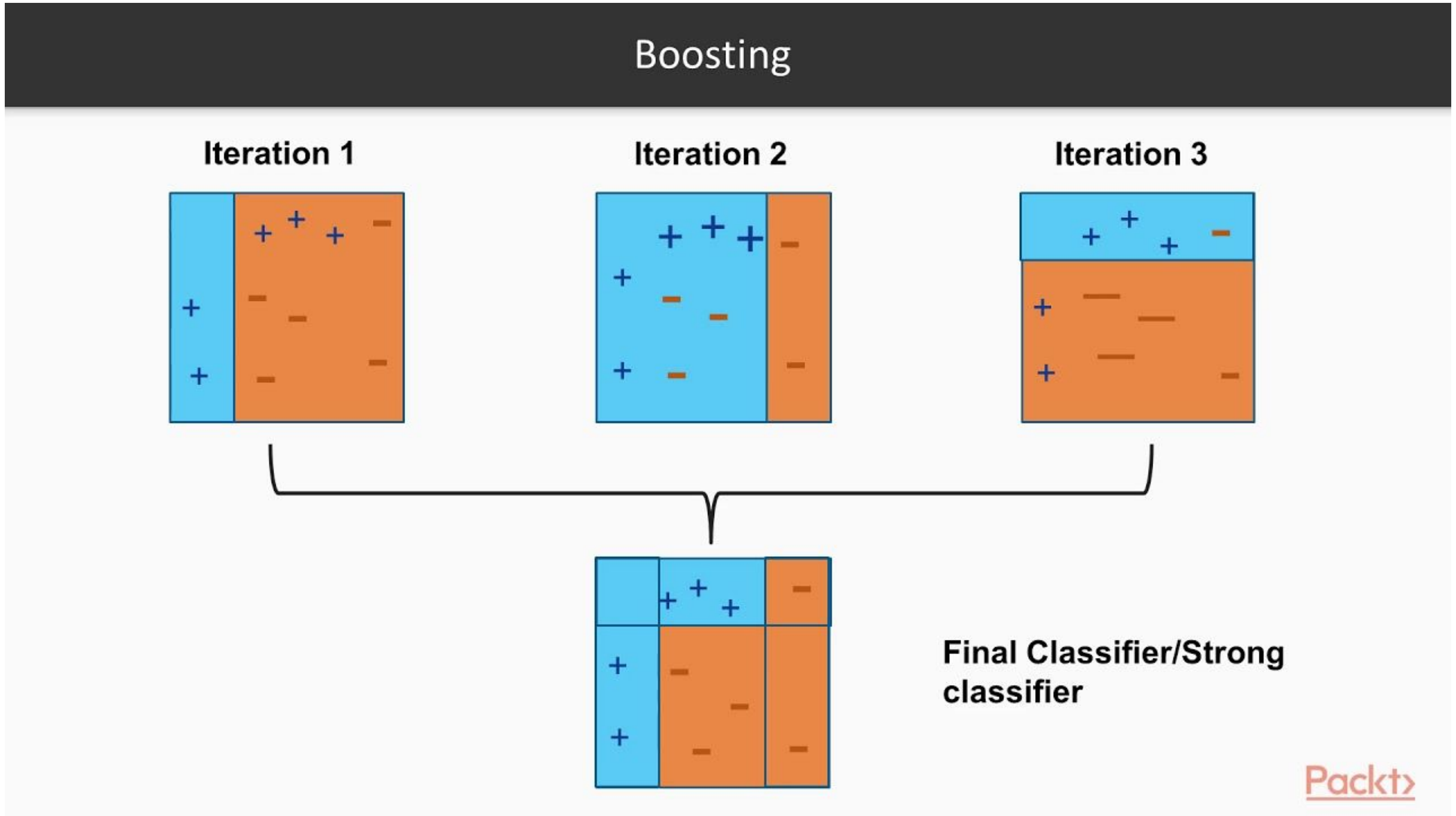
Correlations of Metrics



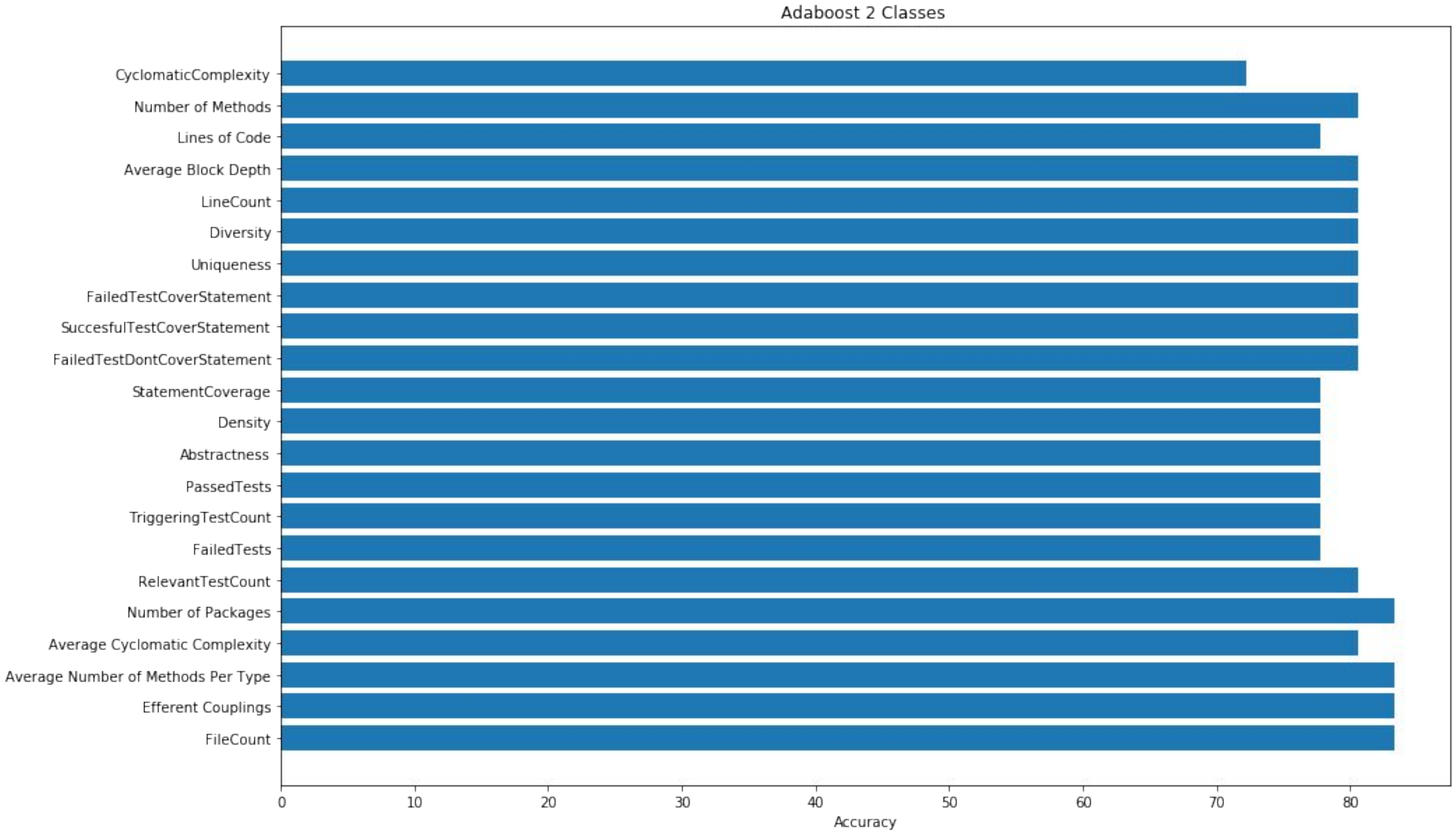
	Abstractness	Average Block Depth	Average Cyclomatic Complexity	Average Number of Methods Per Type	Efferent Couplings	Lines of Code	Number of Methods	Number of Packages	Weighted Methods	FileCount	LineCount	StatementCoverage	RelevantTestCount	TriggeringTestCount	PassedTests	FailedTests	FailedTestCoverStatement	FailedTestDontCoverStatement	SuccesfulTestCoverStatement	Density	Uniqueness	Diversity	CyclomaticComplexity	percentage
Abstractness	1.00	-0.31	-0.11	0.04	0.29	0.17	0.09	0.48	0.13	0.12	0.01	-0.31	0.01	-0.01	0.02	-0.01	0.02	0.03	0.02	0.07	-0.12	-0.07	-0.03	-0.00
Average Block Depth	-0.31	1.00	0.44	0.05	-0.78	-0.86	-0.90	-0.07	-0.88	-0.14	0.04	0.00	-0.53	-0.05	-0.54	-0.05	-0.19	-0.21	-0.19	-0.31	0.39	0.30	-0.61	-0.54
Average Cyclomatic Complexity	-0.11	0.44	1.00	-0.80	0.08	-0.16	-0.30	0.19	-0.13	-0.03	-0.03	-0.16	-0.28	-0.01	-0.27	-0.00	-0.03	-0.07	-0.03	0.22	-0.11	-0.17	-0.07	-0.16
Average Number of Methods Per Type	0.04	0.05	-0.80	1.00	-0.54	-0.35	-0.24	-0.20	-0.38	0.02	0.10	0.11	0.00	0.02	-0.02	0.02	-0.10	-0.06	-0.10	-0.50	0.32	0.33	-0.33	-0.15
Efferent Couplings	0.29	-0.78	0.08	-0.54	1.00	0.94	0.89	0.29	0.95	0.07	-0.09	-0.08	0.45	0.04	0.46	0.04	0.21	0.19	0.21	0.48	-0.44	-0.36	0.66	0.48
Lines of Code	0.17	-0.86	-0.16	-0.35	0.94	1.00	0.98	0.30	0.99	0.06	-0.09	-0.04	0.50	0.06	0.51	0.05	0.20	0.20	0.20	0.40	-0.40	-0.32	0.65	0.54
Number of Methods	0.09	-0.90	-0.30	-0.24	0.89	0.98	1.00	0.15	0.98	0.06	-0.09	0.03	0.55	0.05	0.55	0.05	0.20	0.21	0.20	0.37	-0.38	-0.29	0.68	0.56
Number of Packages	0.48	-0.07	0.19	-0.20	0.29	0.30	0.15	1.00	0.19	-0.01	0.00	-0.37	-0.14	0.04	-0.13	0.04	-0.01	-0.00	-0.01	0.02	-0.10	-0.07	-0.14	0.02
Weighted Methods	0.13	-0.88	-0.13	-0.38	0.95	0.99	0.98	0.19	1.00	0.07	-0.10	-0.01	0.52	0.05	0.53	0.05	0.20	0.20	0.20	0.43	-0.42	-0.34	0.69	0.55
FileCount	0.12	-0.14	-0.03	0.02	0.07	0.06	0.06	-0.01	0.07	1.00	0.10	-0.00	0.15	0.03	0.12	0.08	0.00	0.03	0.00	-0.02	0.01	0.03	0.06	-0.06
LineCount	0.01	0.04	-0.03	0.10	-0.09	-0.09	-0.09	0.00	-0.10	0.10	1.00	0.06	-0.01	0.17	-0.03	0.17	-0.00	0.02	-0.00	-0.02	0.03	0.01	0.01	0.02
StatementCoverage	-0.31	0.00	-0.16	0.11	-0.08	-0.04	0.03	-0.37	-0.01	-0.00	0.06	1.00	0.02	-0.02	-0.03	-0.01	0.02	0.02	0.02	0.08	0.17	0.16	0.18	0.05
RelevantTestCount	0.01	-0.53	-0.28	0.00	0.45	0.50	0.55	-0.14	0.52	0.15	-0.01	0.02	1.00	-0.01	0.67	0.01	0.16	0.17	0.16	0.08	-0.01	0.06	0.34	0.27
TriggeringTestCount	-0.01	-0.05	-0.01	0.02	0.04	0.06	0.05	0.04	0.05	0.03	0.17	-0.02	-0.01	1.00	-0.02	0.96	0.08	0.12	0.08	-0.04	-0.07	-0.06	0.16	0.12
PassedTests	0.02	-0.54	-0.27	-0.02	0.46	0.51	0.55	-0.13	0.53	0.12	-0.03	-0.03	0.67	-0.02	1.00	-0.00	0.16	0.15	0.16	0.02	0.15	0.26	0.37	0.19
FailedTests	-0.01	-0.05	-0.00	0.02	0.04	0.05	0.05	0.04	0.05	0.08	0.17	-0.01	0.01	0.96	-0.00	1.00	0.11	0.16	0.11	-0.04	-0.06	-0.04	0.18	0.12
FailedTestCoverStatement	0.02	-0.19	-0.03	-0.10	0.21	0.20	0.20	-0.01	0.20	0.00	-0.00	0.02	0.16	0.08	0.16	0.11	1.00	0.99	1.00	0.07	-0.10	-0.08	0.35	0.18
FailedTestDontCoverStatement	0.03	-0.21	-0.07	-0.06	0.19	0.20	0.21	-0.00	0.20	0.03	0.02	0.02	0.17	0.12	0.15	0.16	0.99	1.00	0.99	0.04	-0.10	-0.08	0.33	0.20
SuccesfulTestCoverStatement	0.02	-0.19	-0.03	-0.10	0.21	0.20	0.20	-0.01	0.20	0.00	-0.00	0.02	0.16	0.08	0.16	0.11	1.00	0.99	1.00	0.07	-0.10	-0.08	0.35	0.18
Density	0.07	-0.31	0.22	-0.50	0.48	0.40	0.37	0.02	0.43	-0.02	-0.02	0.08	0.08	-0.04	0.02	-0.04	0.07	0.04	0.07	1.00	-0.10	-0.09	0.31	0.28
Uniqueness	-0.12	0.39	-0.11	0.32	-0.44	-0.40	-0.38	-0.10	-0.42	0.01	0.03	0.17	-0.01	-0.07	0.15	-0.06	-0.10	-0.10	-0.10	-0.10	1.00	0.97	-0.37	-0.35
Diversity	-0.07	0.30	-0.17	0.33	-0.36	-0.32	-0.29	-0.07	-0.34	0.03	0.01	0.16	0.06	-0.06	0.26	-0.04	-0.08	-0.08	-0.08	-0.09	0.97	1.00	-0.32	-0.31
CyclomaticComplexity	-0.03	-0.61	-0.07	-0.33	0.66	0.65	0.68	-0.14	0.69	0.06	0.01	0.18	0.34	0.16	0.37	0.18	0.35	0.33	0.35	0.31	-0.37	-0.32	1.00	0.48
percentage	-0.00	-0.54	-0.16	-0.15	0.48	0.54	0.56	0.02	0.55	-0.06	0.02	0.05	0.27	0.12	0.19	0.12	0.18	0.20	0.18	0.28	-0.35	-0.31	0.48	1.00

Adaboost

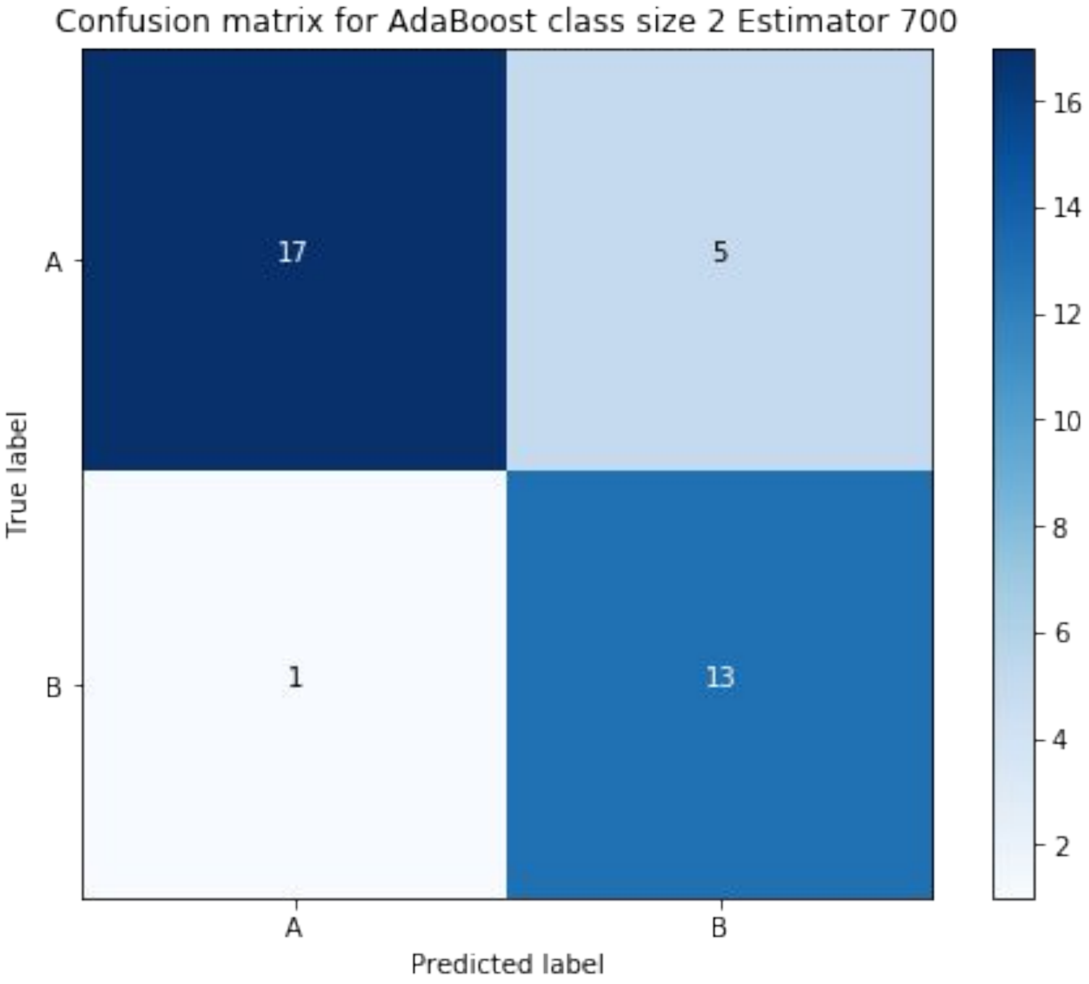
- **Augmentation**
- **Fast**



Adaboost(2 Classes)

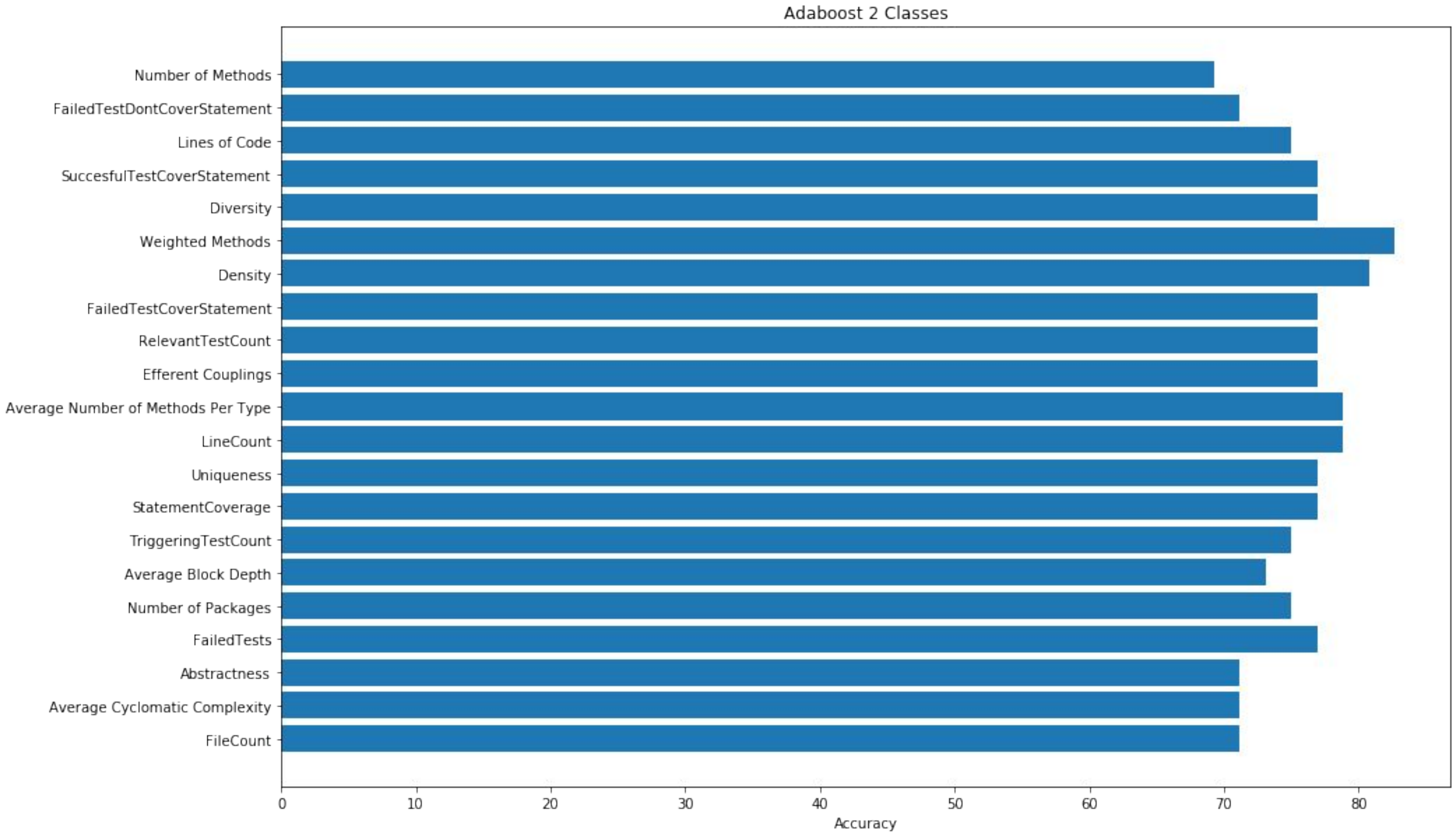


Adaboost

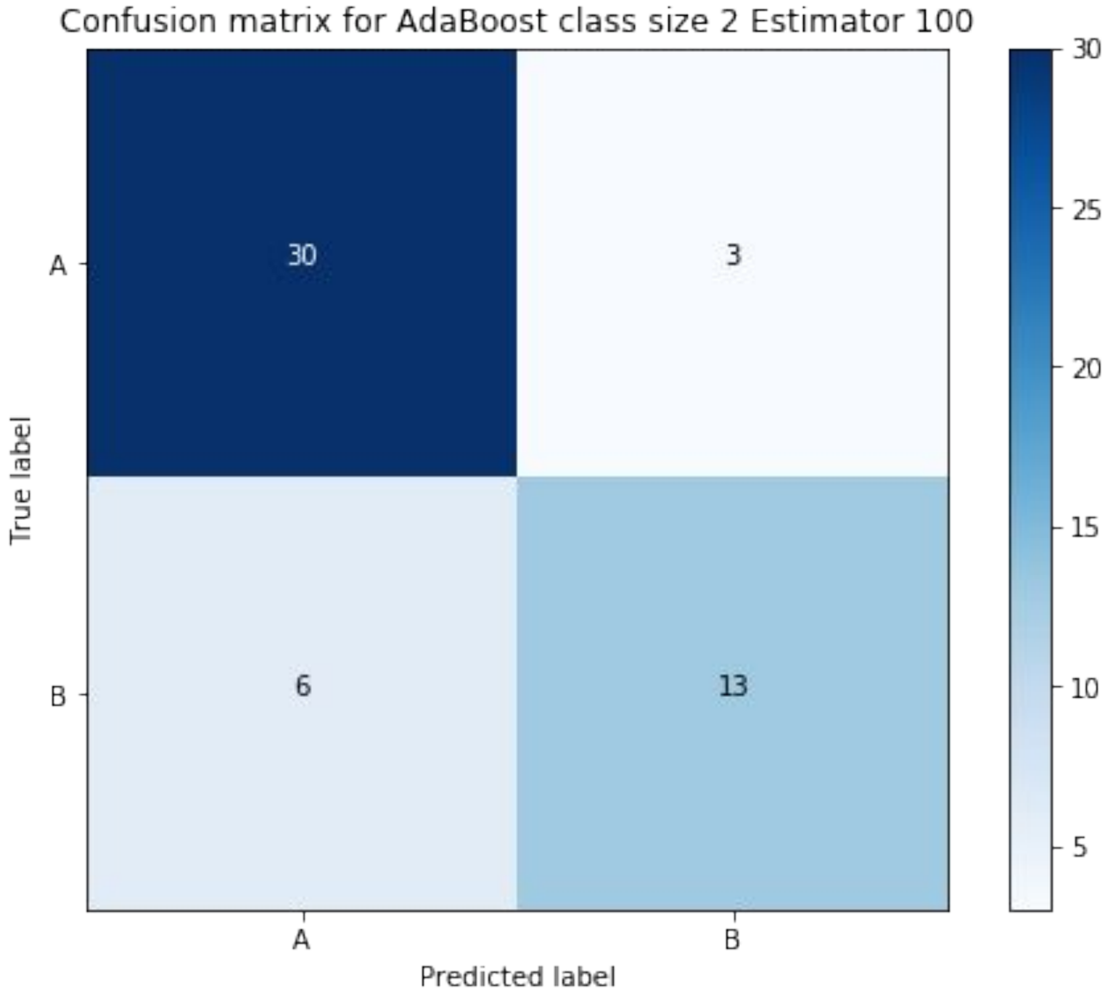


Acc. = 83%

Adaboost with Math



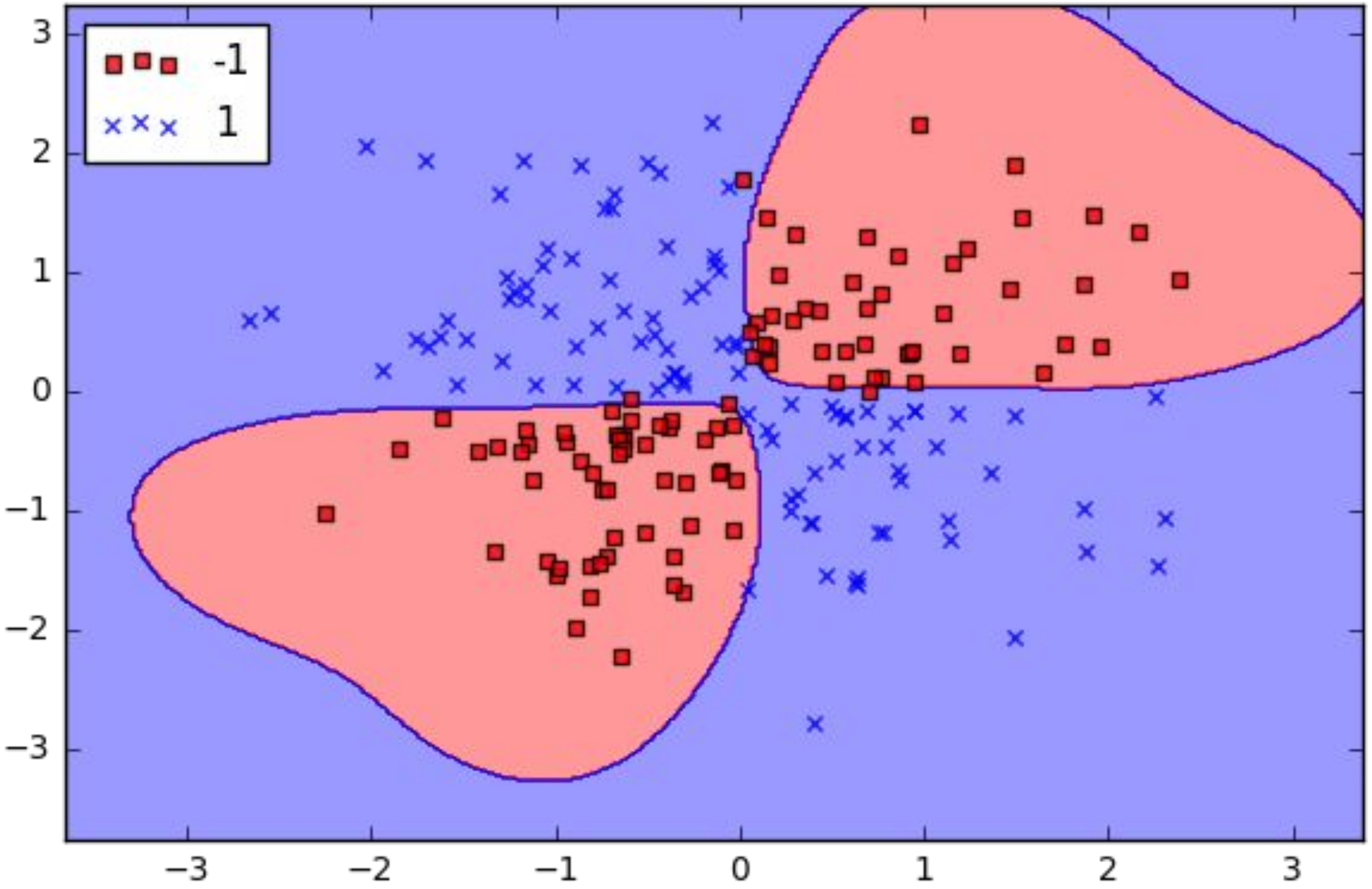
Adaboost with Math



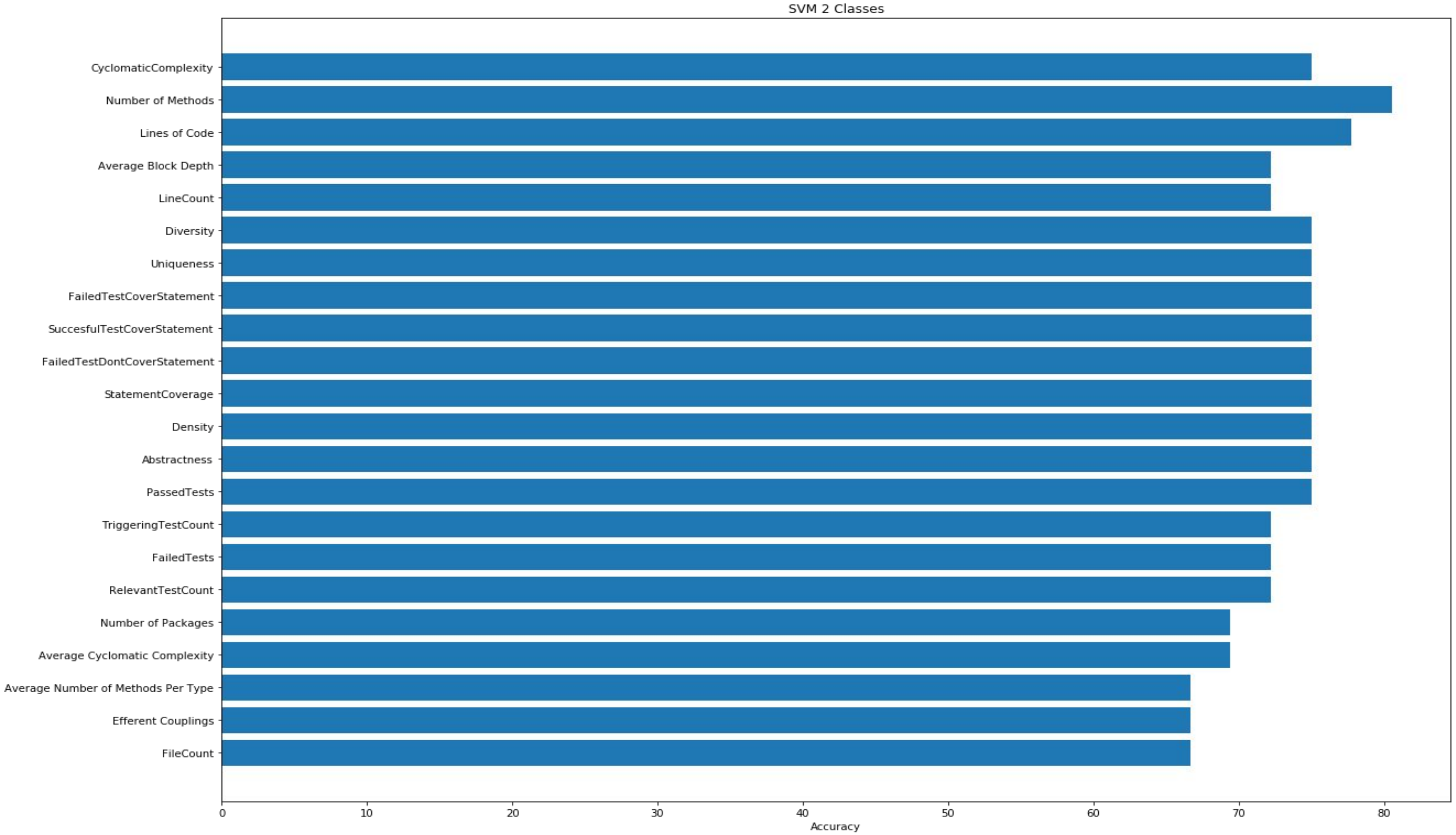
Acc. = 82.6%

SVM

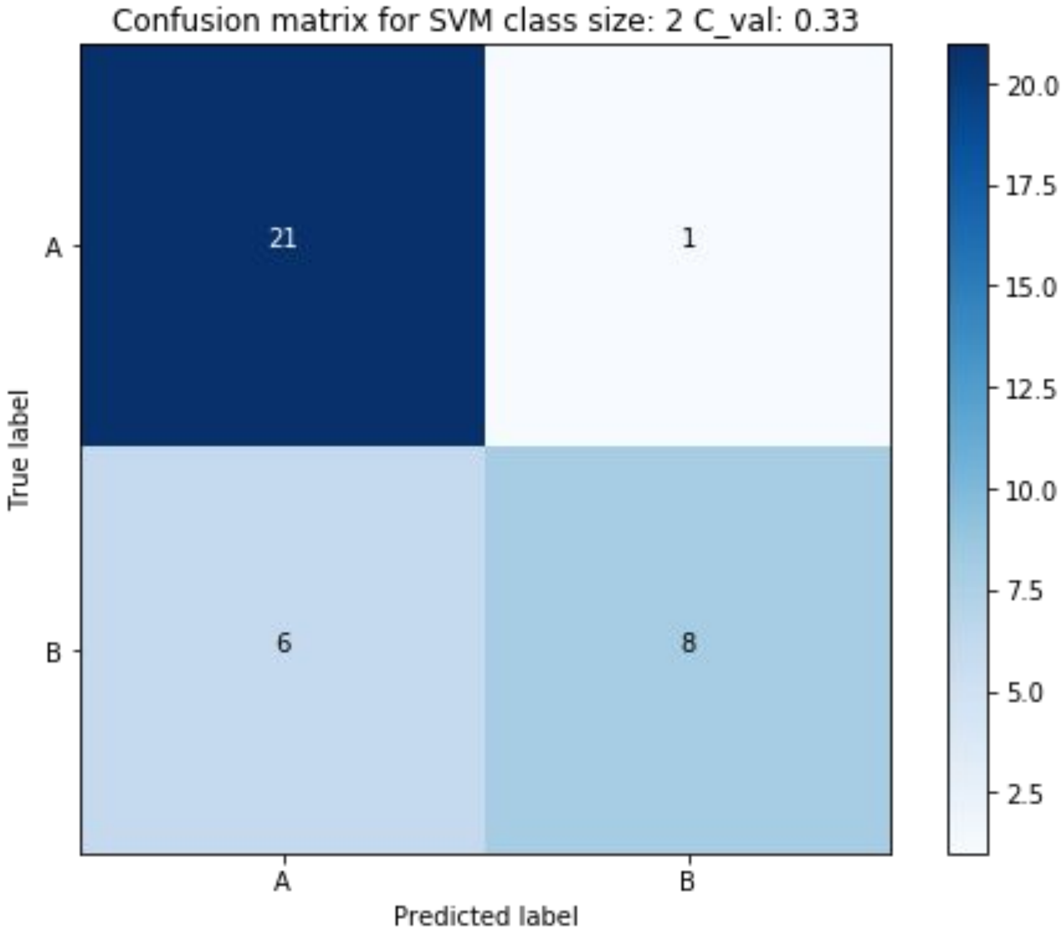
- **Kernel**
- **Robust**



SVM(2 classes)

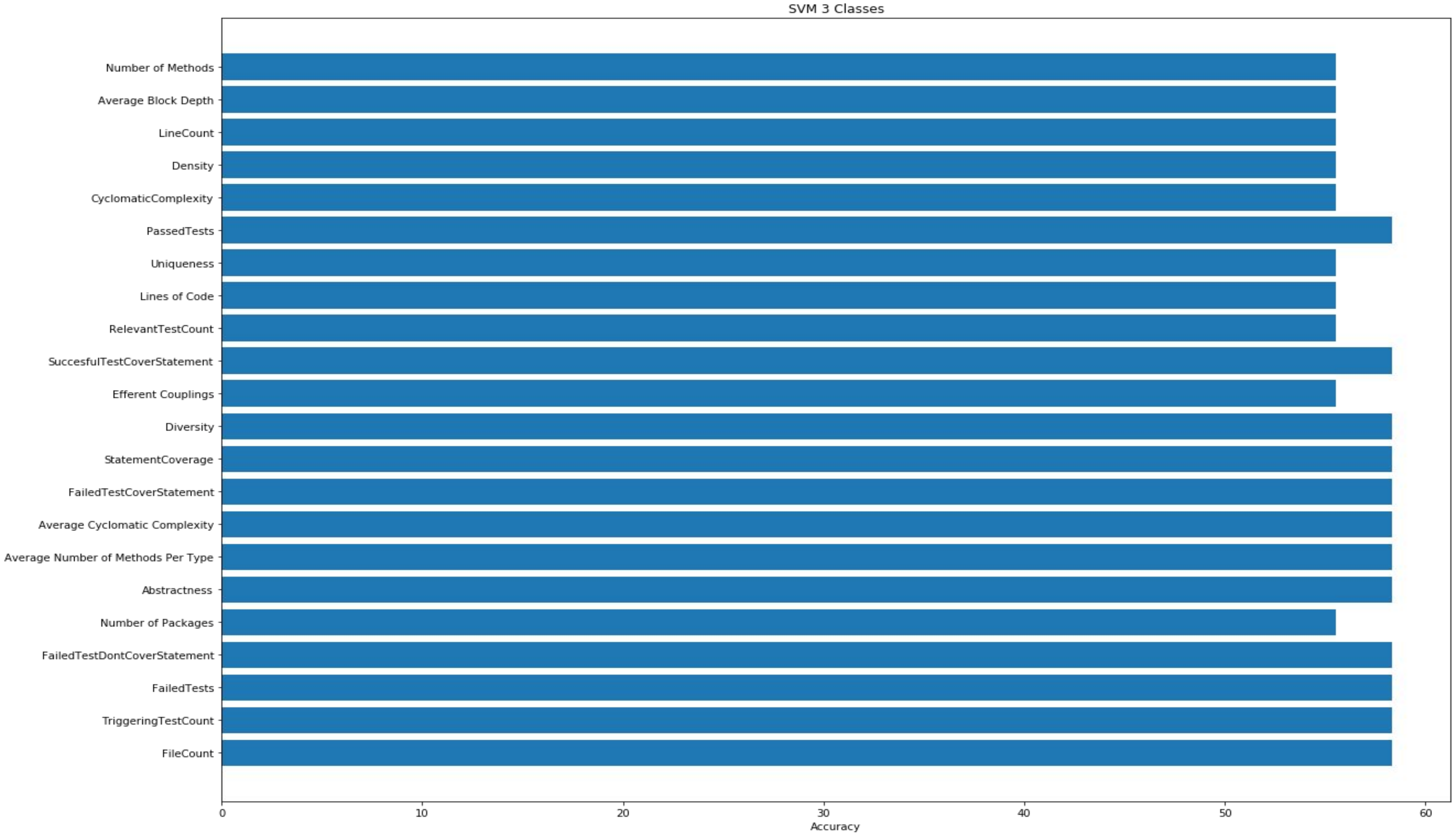


SVM(2 classes)

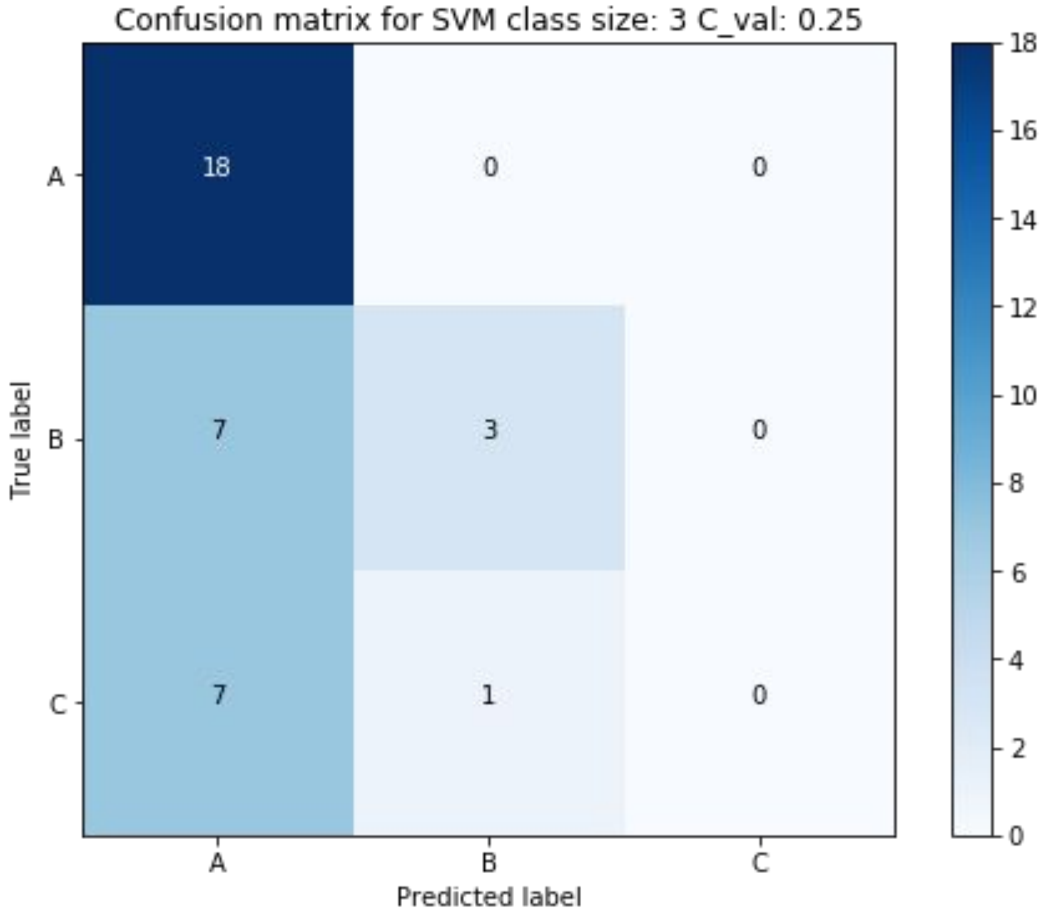


Acc. = 80.5%

SVM (3 classes)

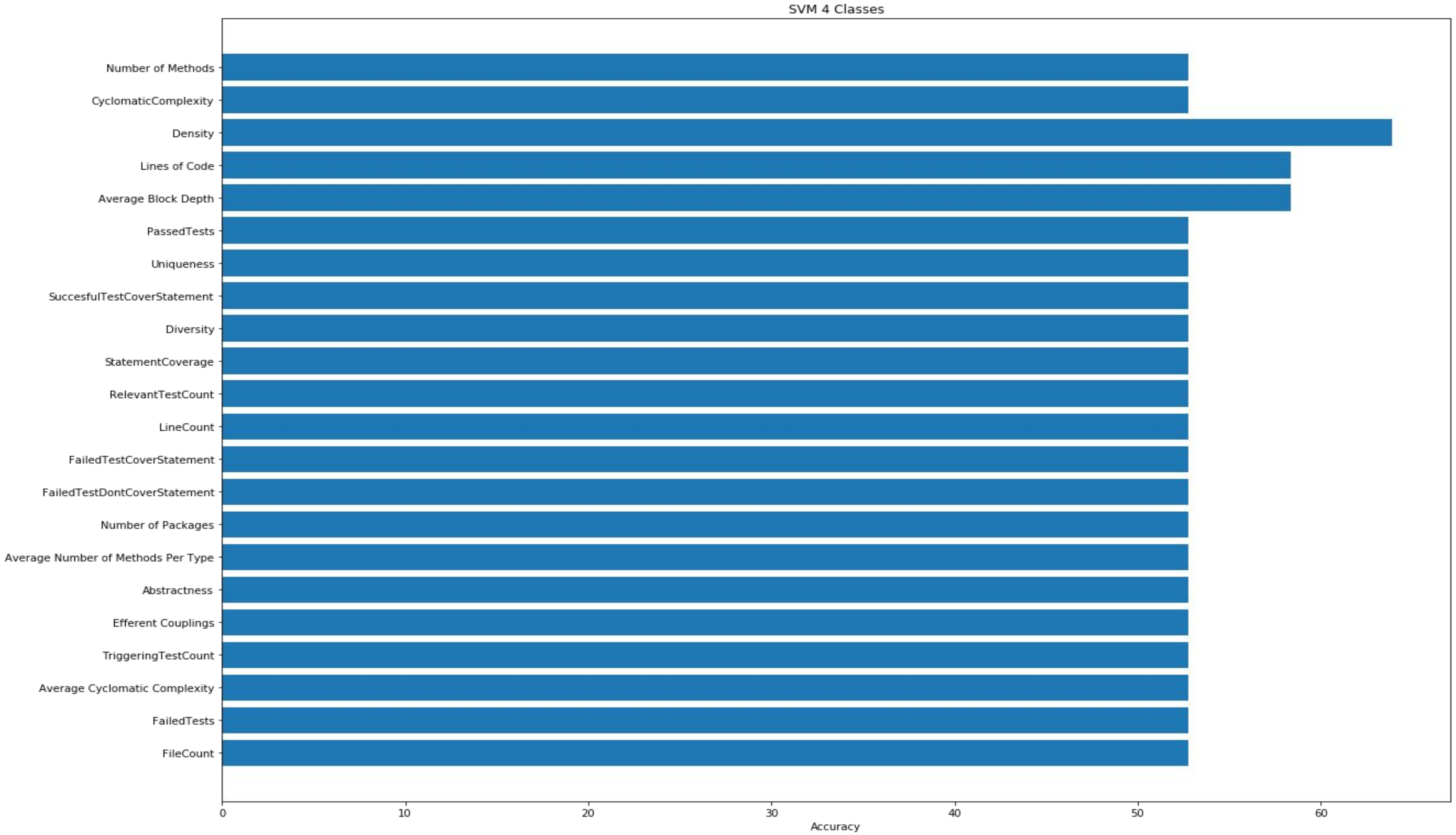


SVM (3 classes)

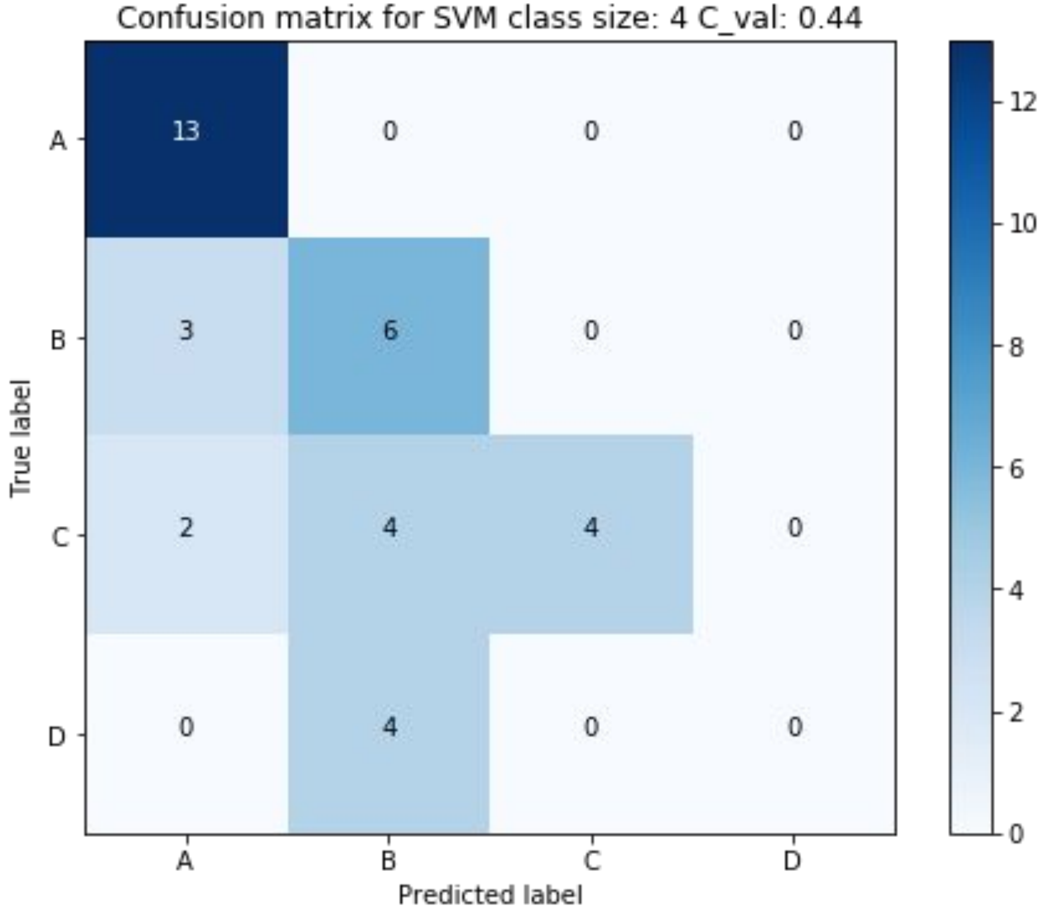


Acc. = 58.3%

SVM (4 classes)



SVM(4 classes)



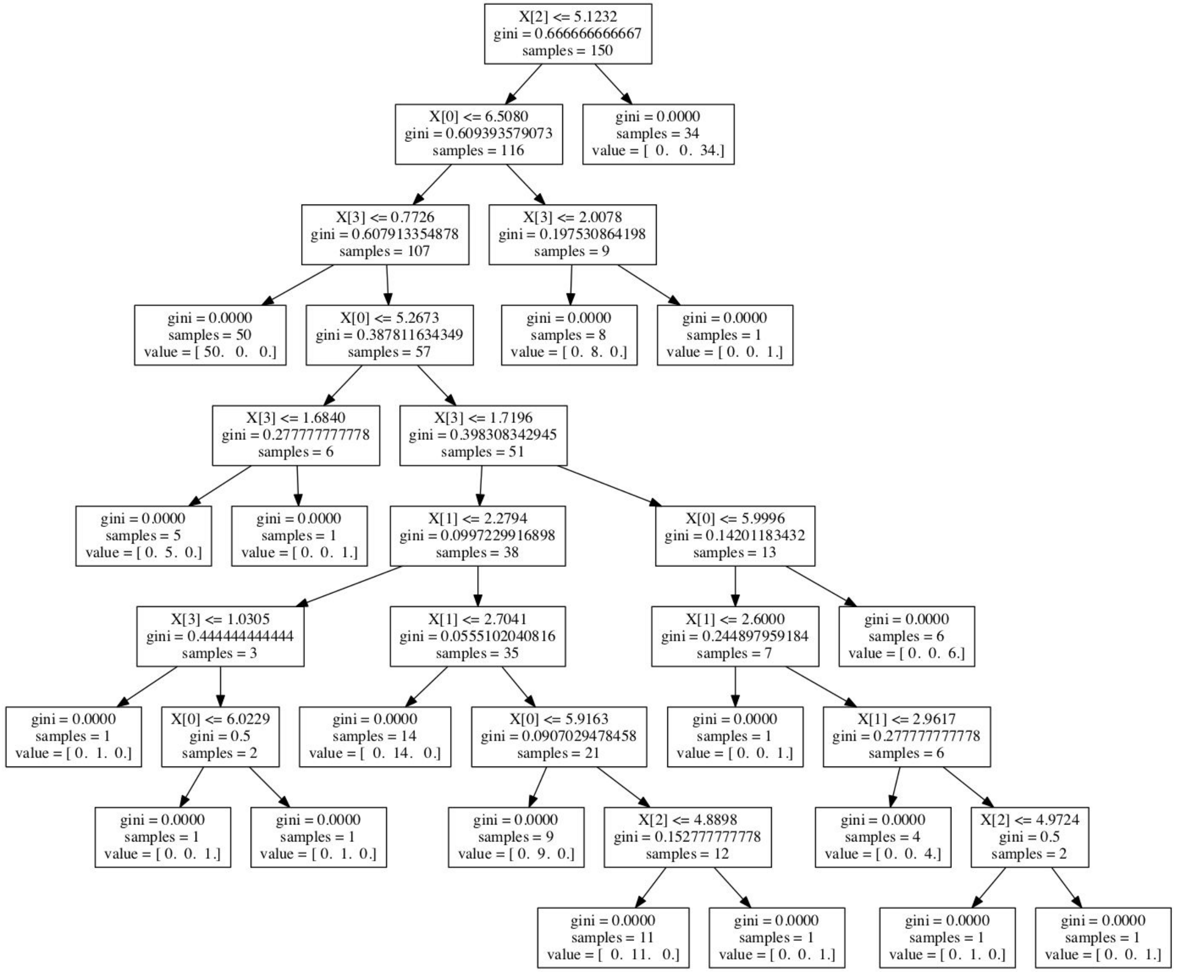
Confusion matrix for SVM class size: 4 C_val: 0.44

True label \ Predicted label	A	B	C	D
A	13	0	0	0
B	3	6	0	0
C	2	4	4	0
D	0	4	0	0

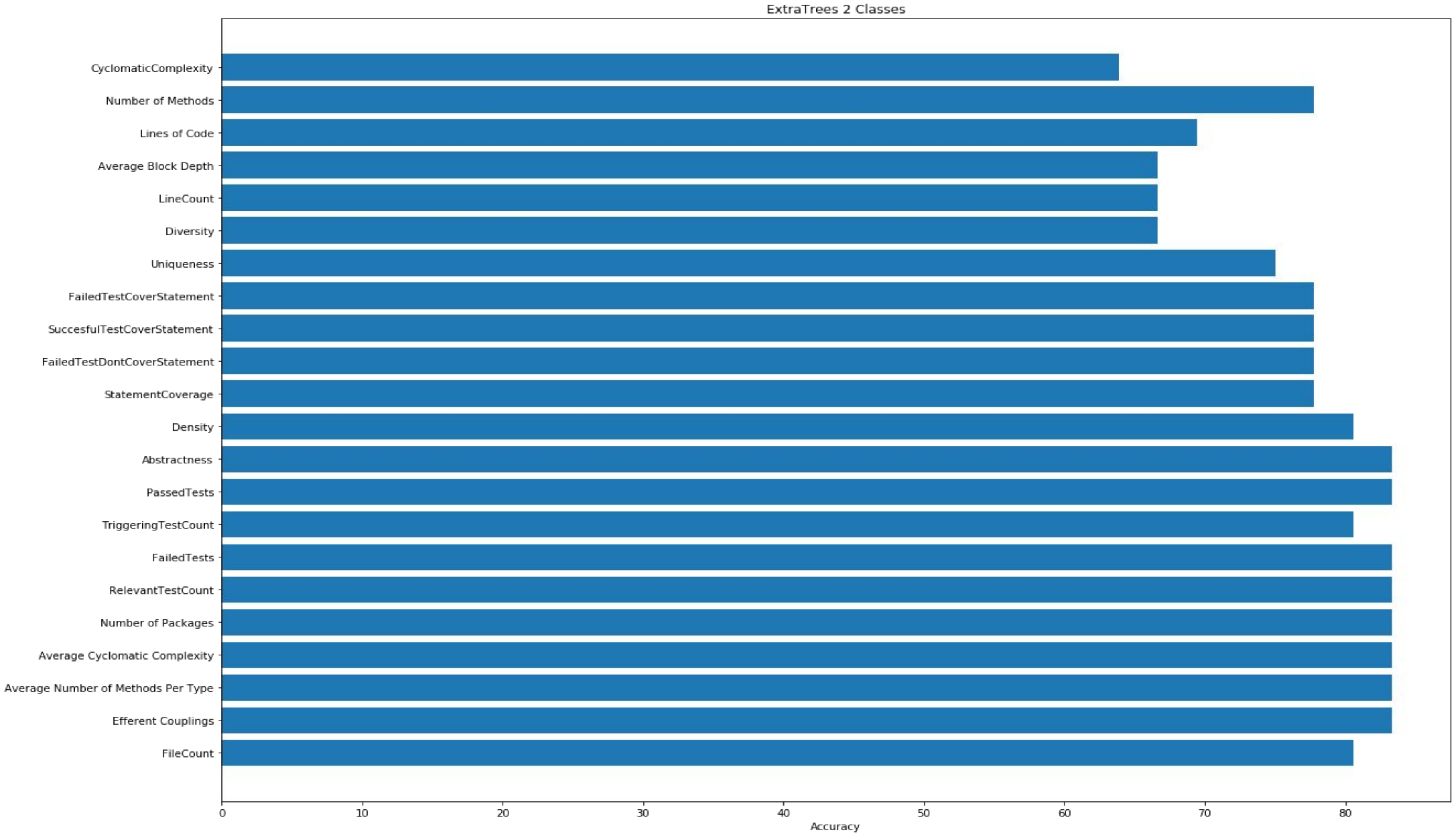
Acc. = 63.88%

ExtraTrees

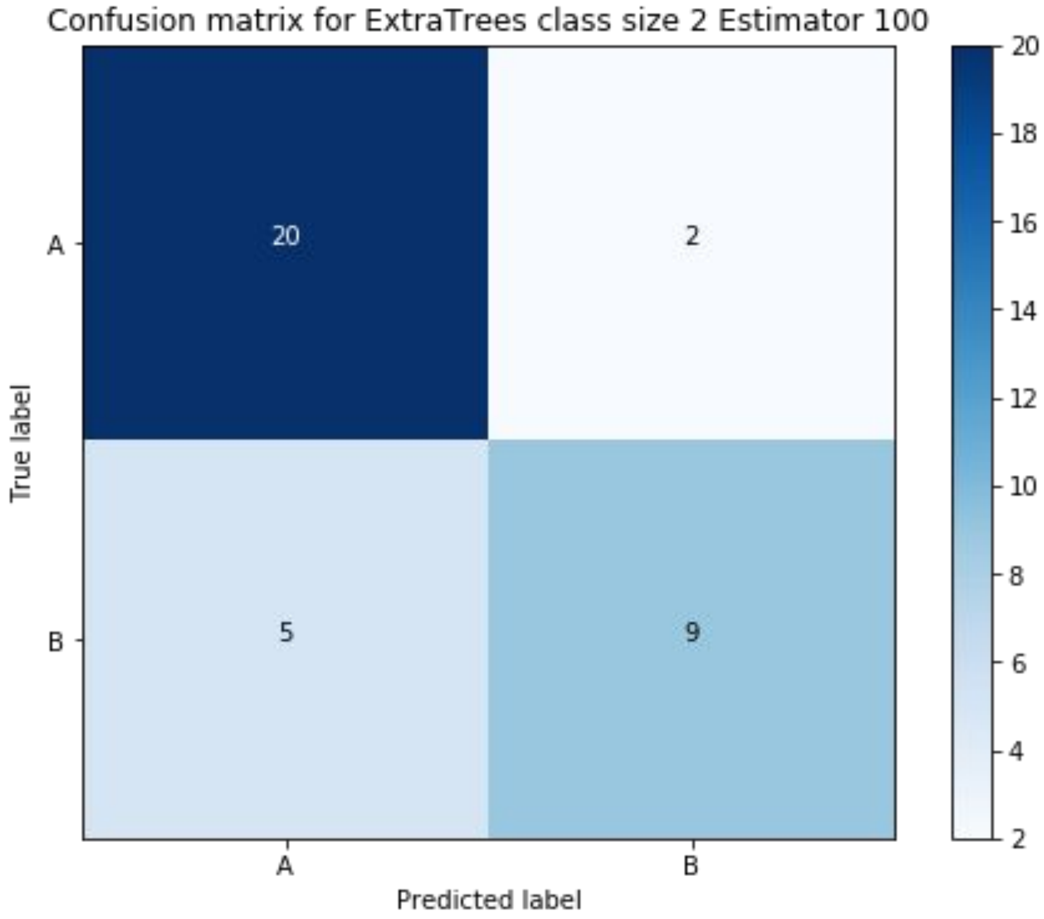
- **ExtRa**



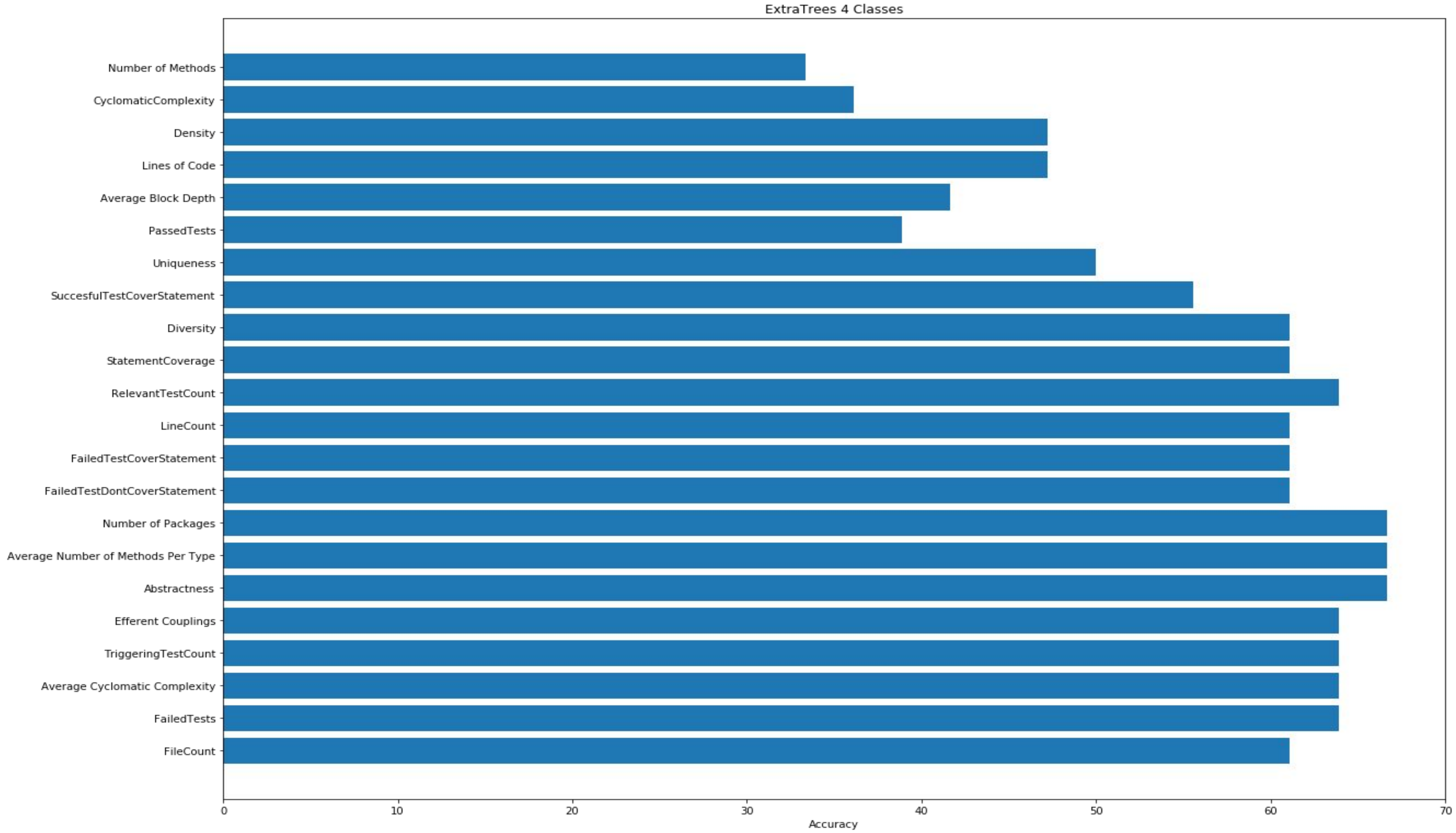
ExtraTrees


ExtraTrees 2 Classes
CyclomaticComplexity
Number of Methods
Lines of Code
Average Block Depth
LineCount
Diversity
Uniqueness
FailedTestCoverStatement
SuccesfulTestCoverStatement
FailedTestDontCoverStatement
StatementCoverage
Density
Abstractness
PassedTests
TriggeringTestCount
FailedTests
RelevantTestCount
Number of Packages
Average Cyclomatic Complexity
Average Number of Methods Per Type
Efferent Couplings
FileCount
0
10
20
30
40
50
60
70
80
Accuracy

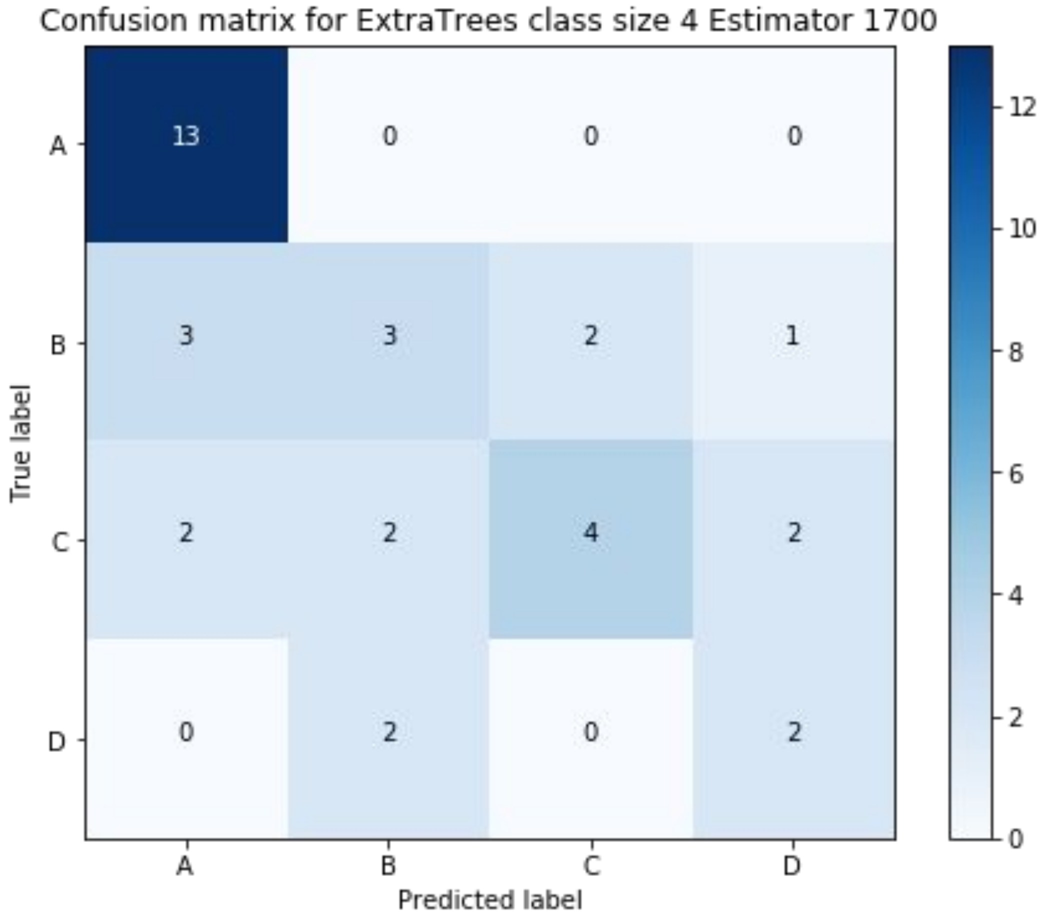
ExtraTrees


Confusion matrix for ExtraTrees class size 2 Estimator 100
True label
A
B
20
2
5
9
Predicted label
A
B
20
18
16
14
12
10
8
6
4
2

ExtraTrees


ExtraTrees 4 Classes
Number of Methods
CyclomaticComplexity
Density
Lines of Code
Average Block Depth
PassedTests
Uniqueness
SuccesfulTestCoverStatement
Diversity
StatementCoverage
RelevantTestCount
LineCount
FailedTestCoverStatement
FailedTestDontCoverStatement
Number of Packages
Average Number of Methods Per Type
Abstractness
Efferent Couplings
TriggeringTestCount
Average Cyclomatic Complexity
FailedTests
FileCount
0
10
20
30
40
50
60
70
Accuracy

ExtraTrees

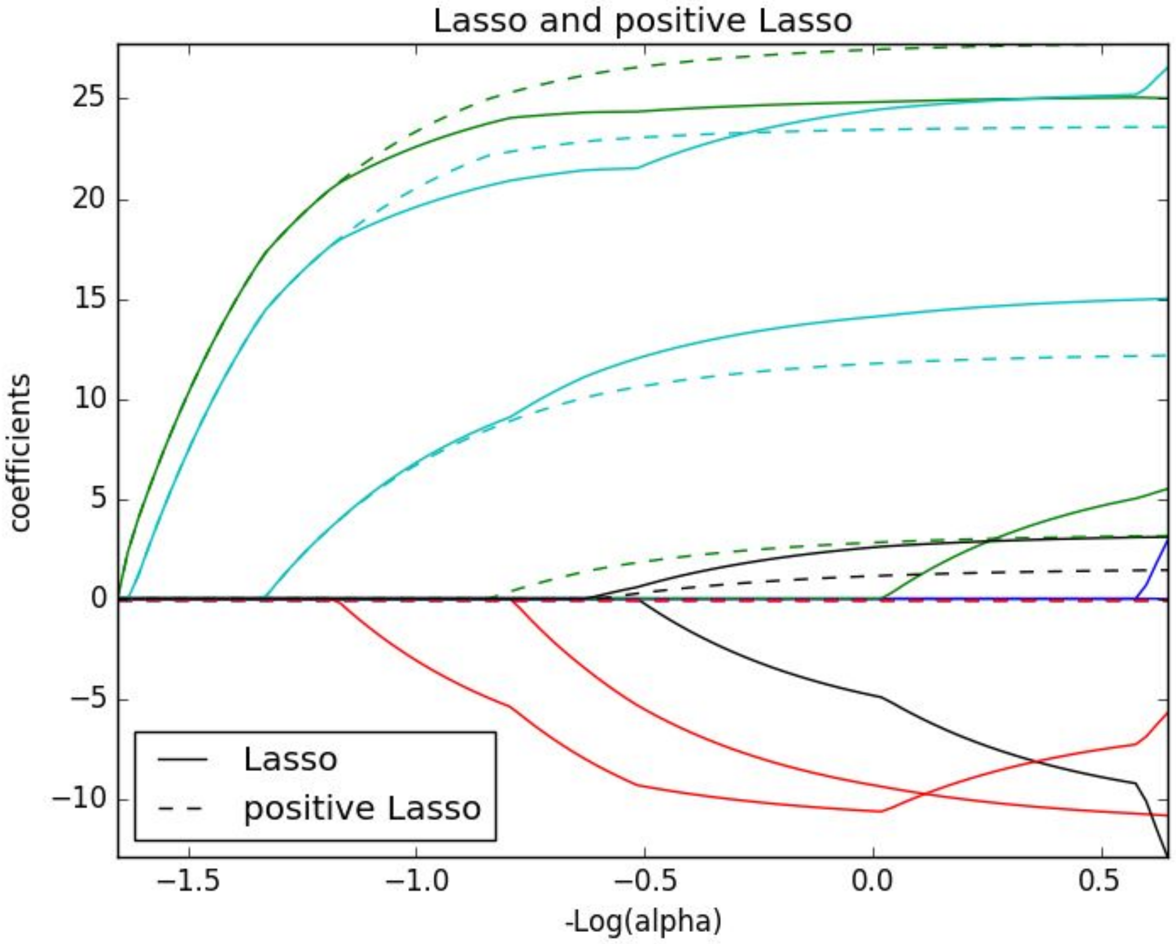


Confusion matrix for ExtraTrees class size 4 Estimator 1700

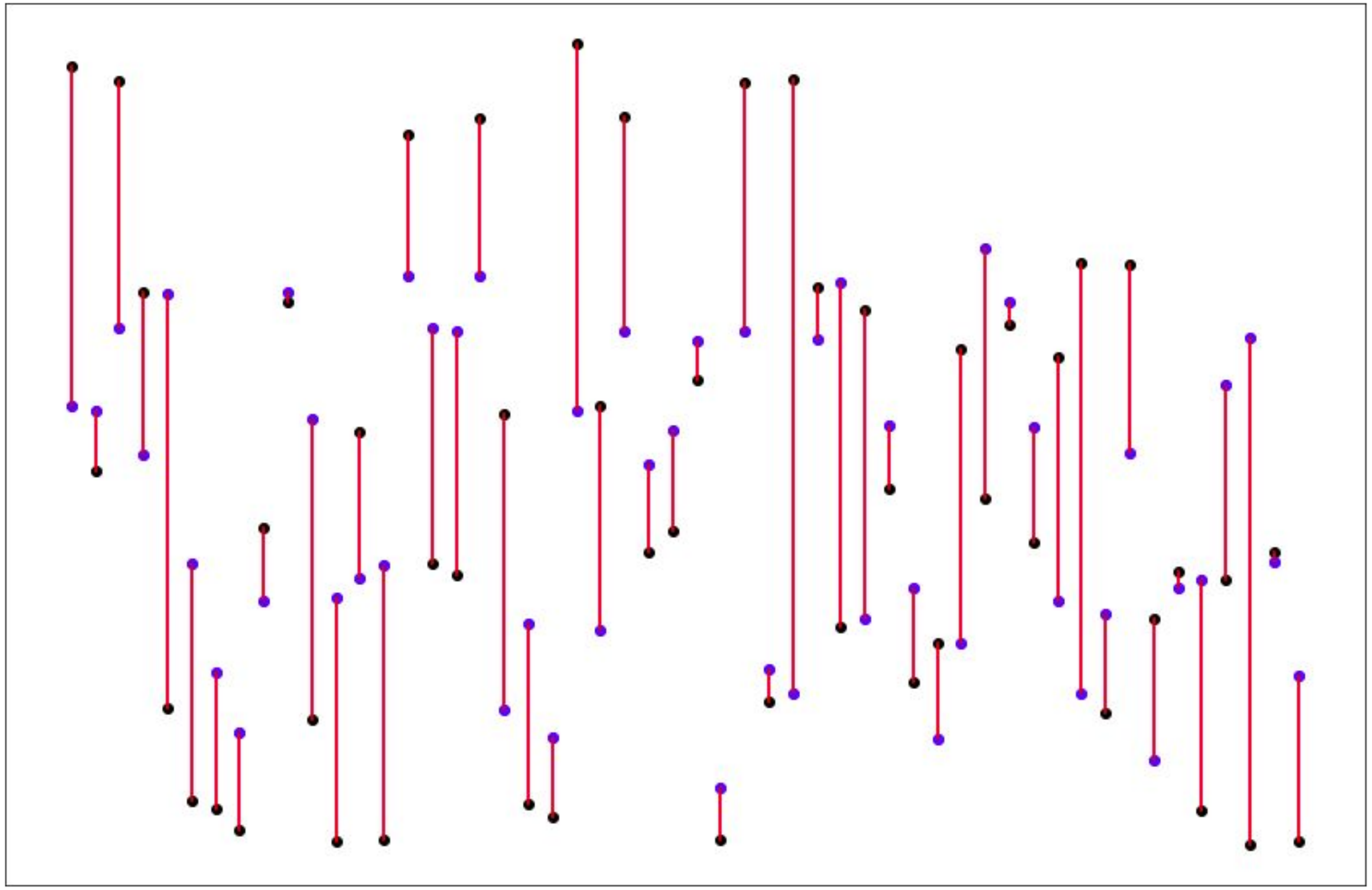
True label \ Predicted label	A	B	C	D
A	13	0	0	0
B	3	3	2	1
C	2	2	4	2
D	0	2	0	2

Bonus: ElasticNet

- **Regression**



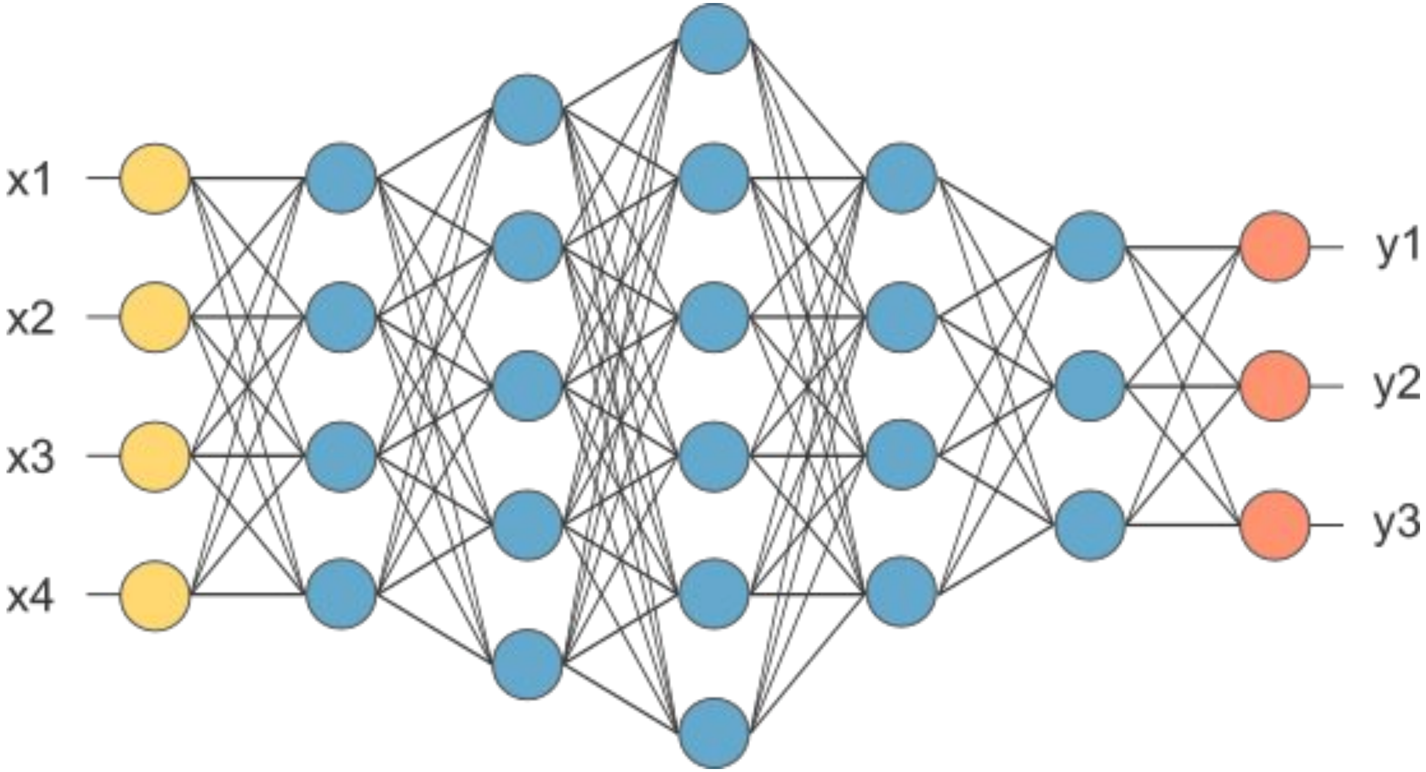
Bonus: ElasticNet



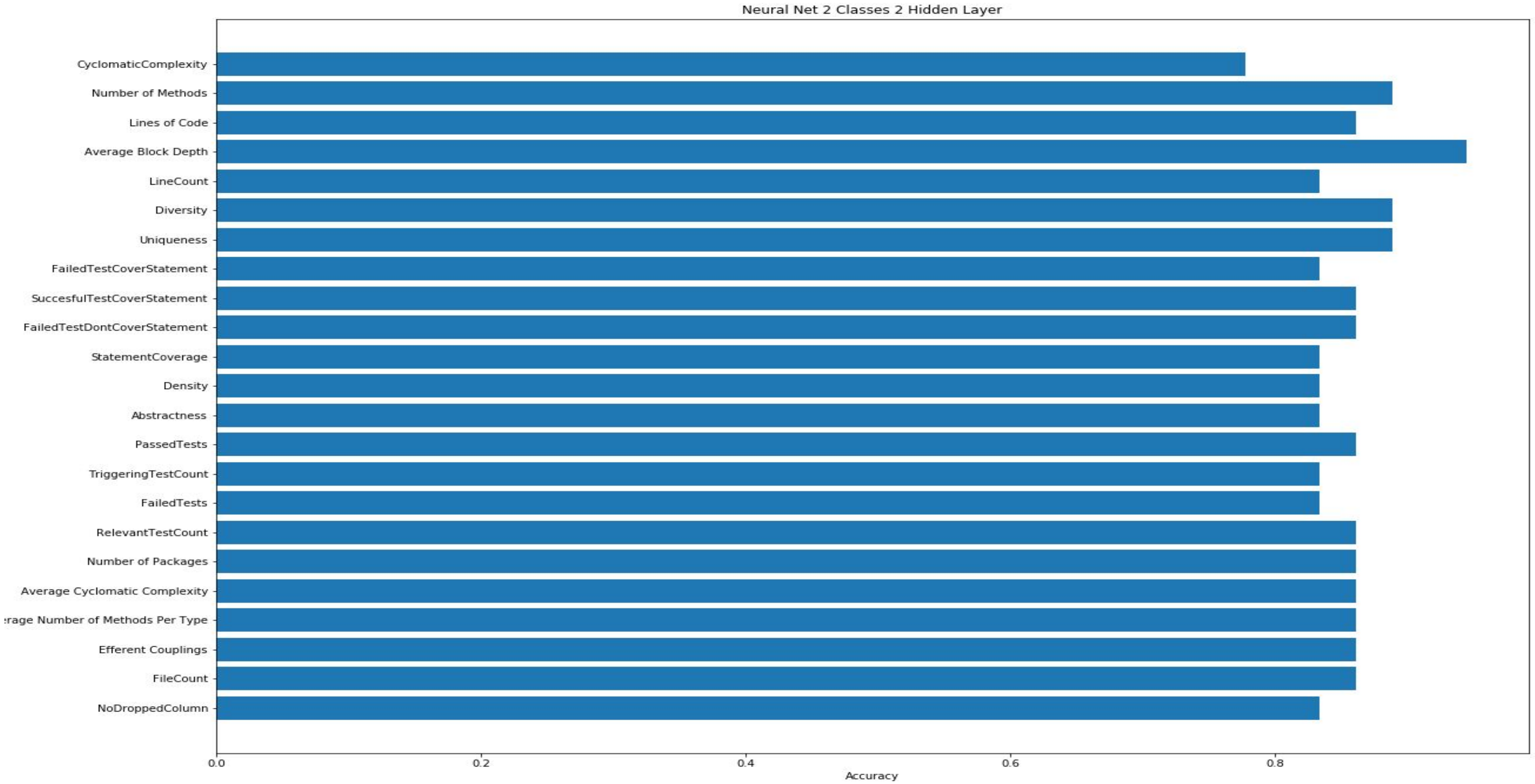
MSE = 0.128

Neural Network - MLP

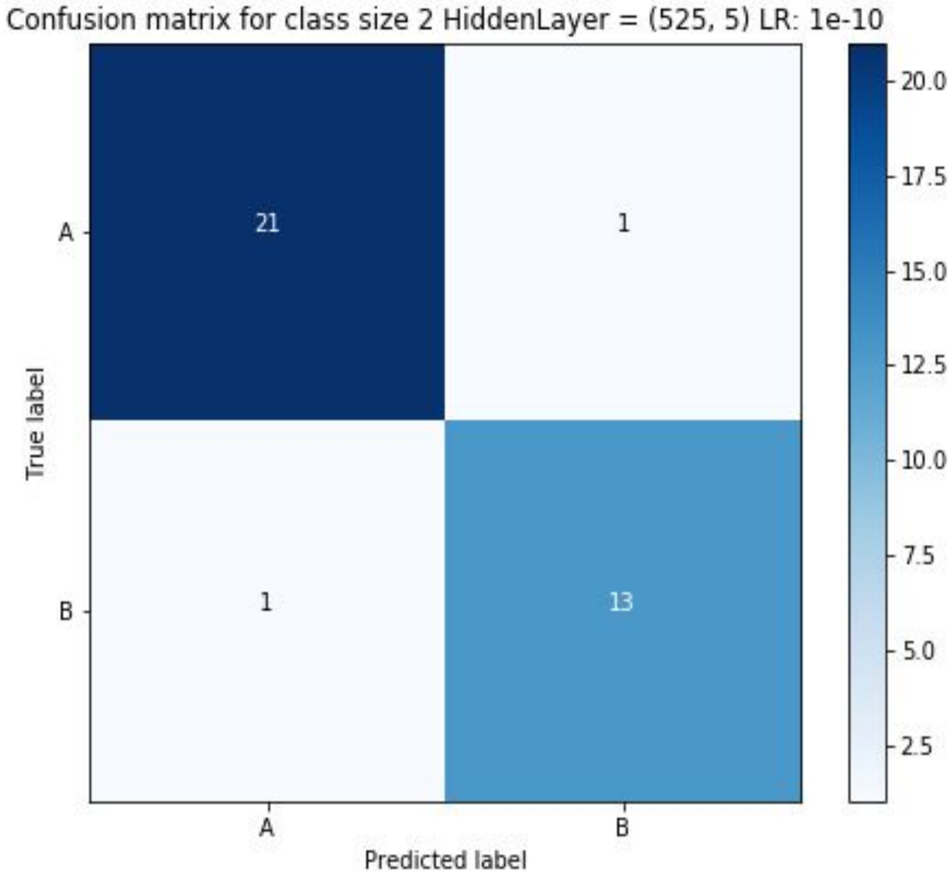
- **Layers**
- **Activation**



Neural Network - MLP (2 hidden layer)



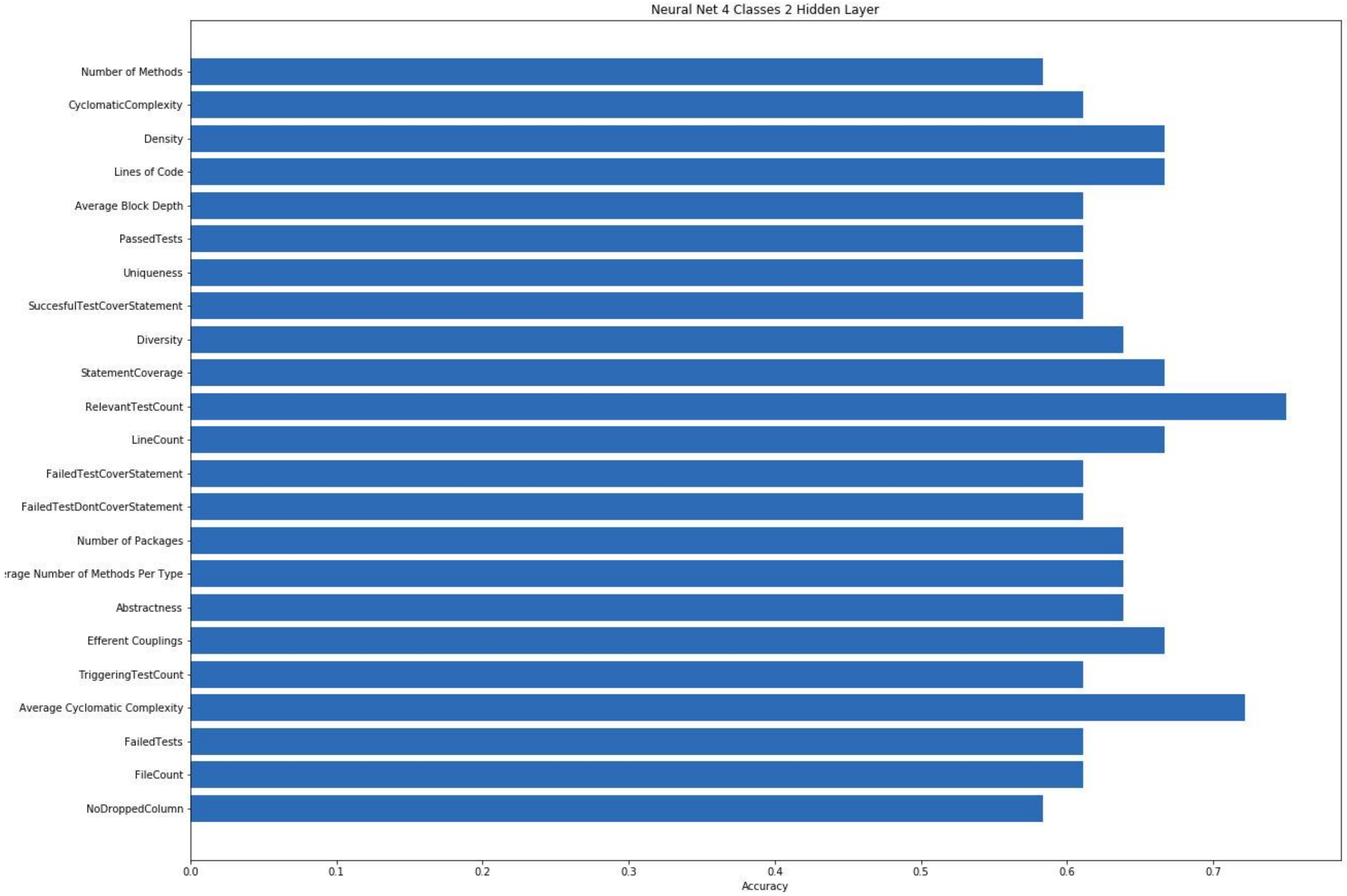
Neural Network - MLP (2 hidden layer)



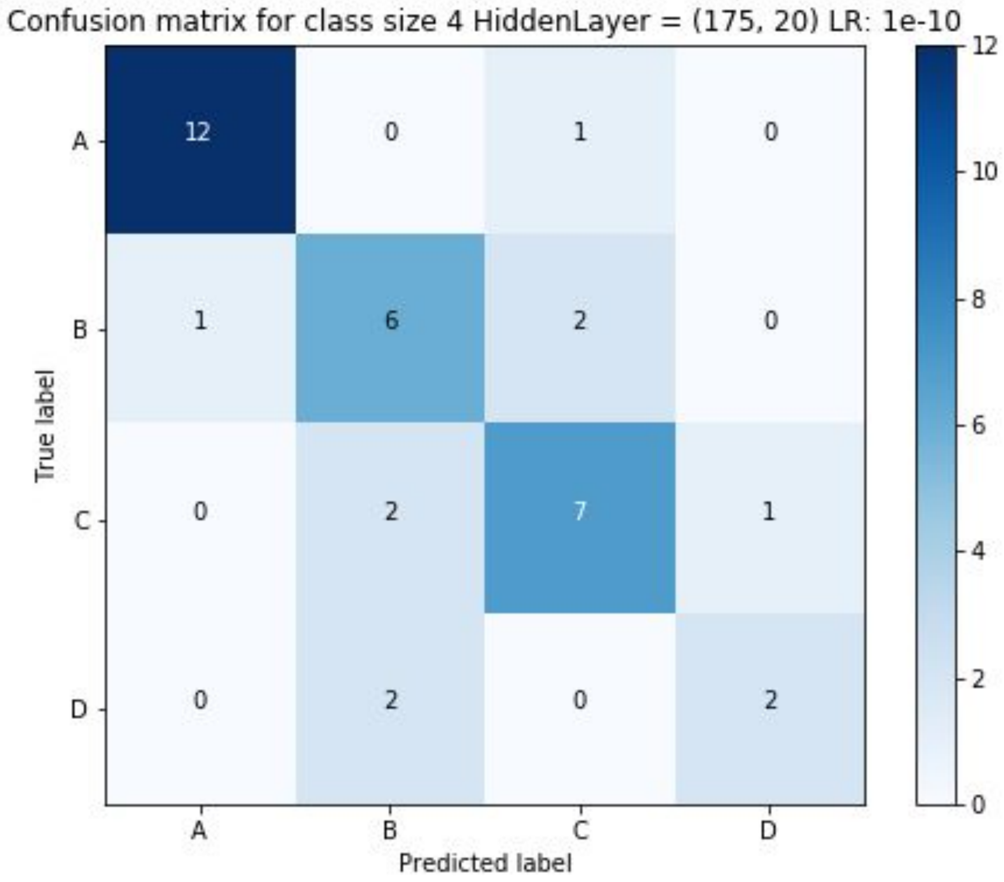
Confusion matrix for class size 2 HiddenLayer = (525, 5) LR: 1e-10

Acc. = 94.4%

Neural Network - (4 classes)



Neural Network - (4 classes)

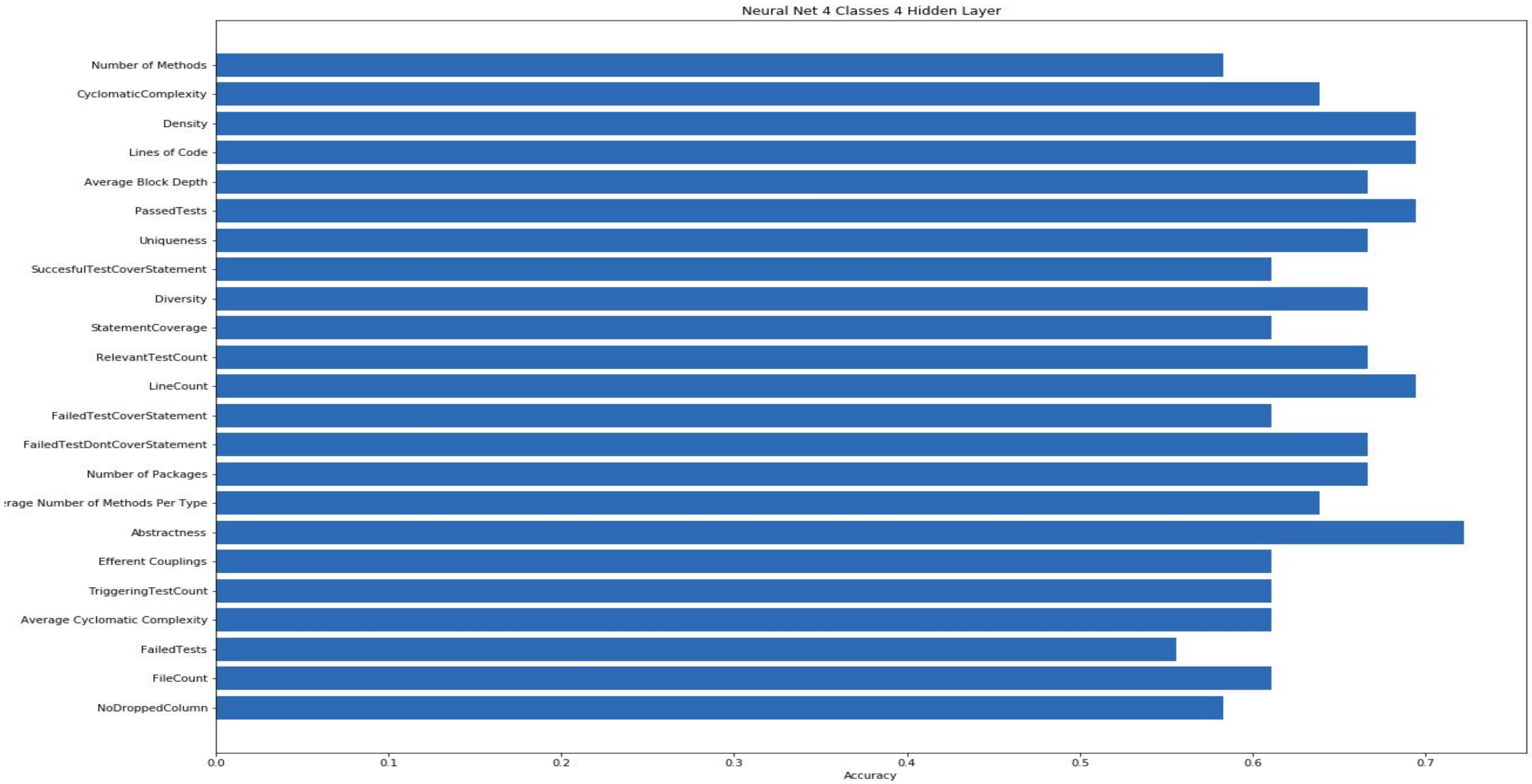


Confusion matrix for class size 4 HiddenLayer = (175, 20) LR: 1e-10

True label \ Predicted label	A	B	C	D
A	12	0	1	0
B	1	6	2	0
C	0	2	7	1
D	0	2	0	2

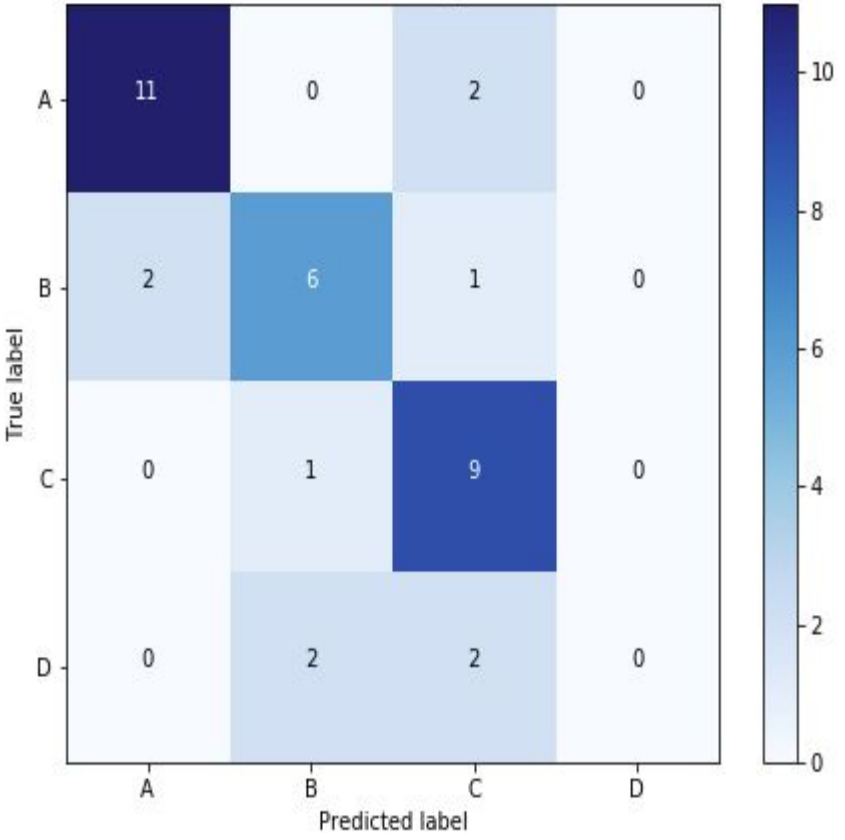
Acc. = 75%

Neural Network - MLP(4 C 4 L)



Neural Network - MLP (4 classes 4 Layer)

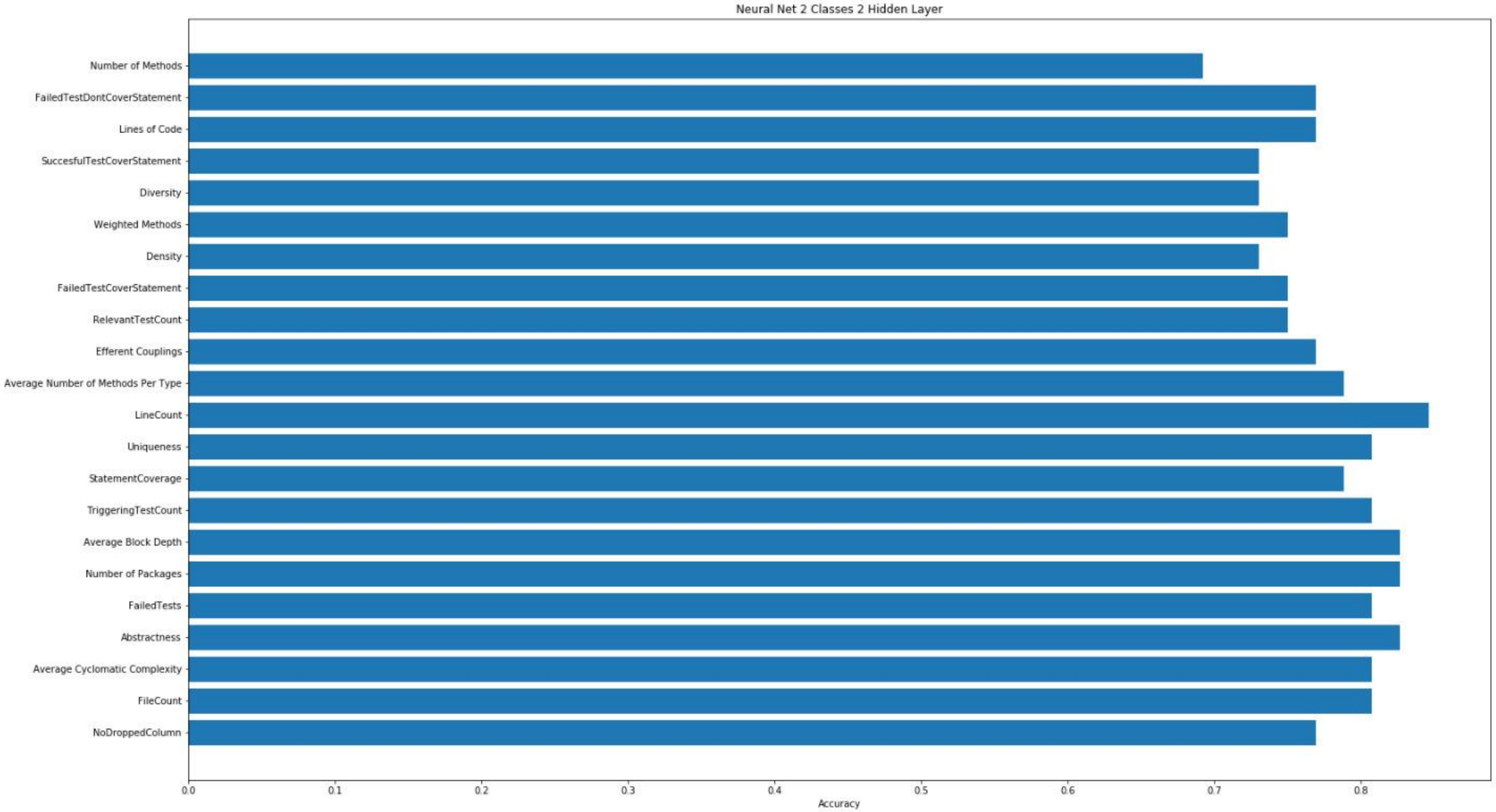
Confusion matrix for class size 4 HiddenLayer = (425, 55, 5, 350) LR: 1e-08



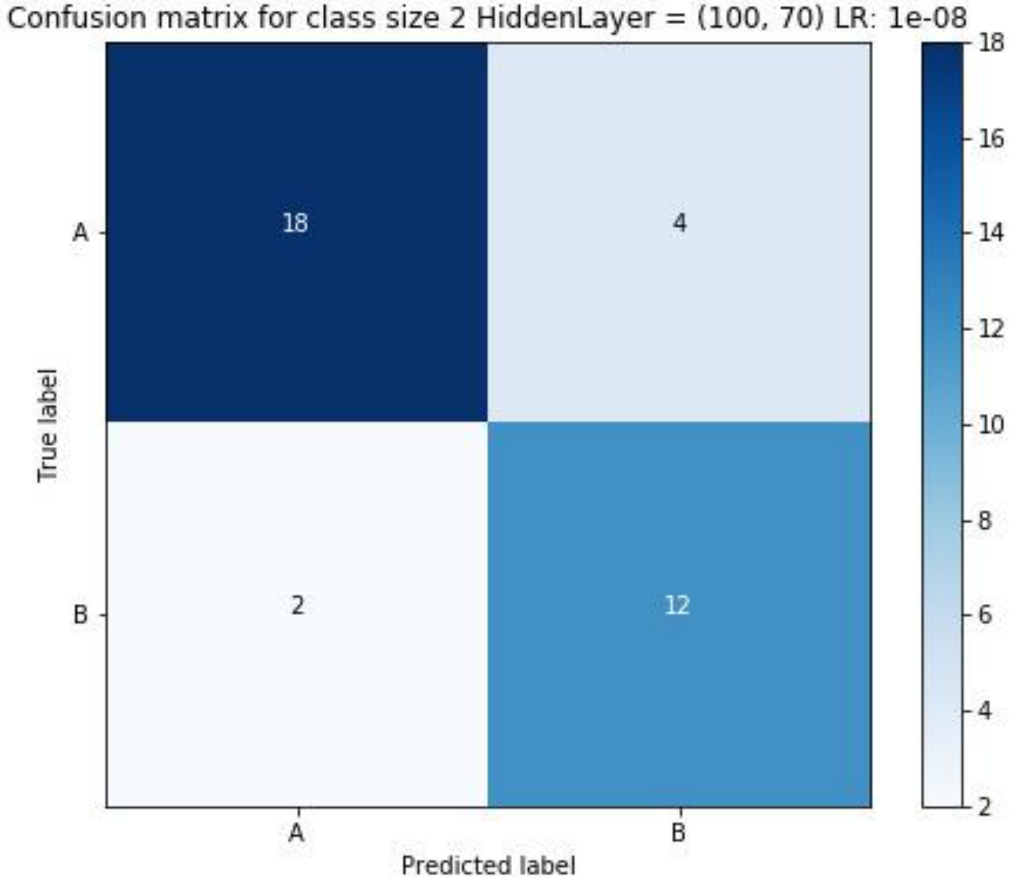
True label \ Predicted label	A	B	C	D
A	11	0	2	0
B	2	6	1	0
C	0	1	9	0
D	0	2	2	0

Acc. = 72%

Neural Network - MLP with Math



Neural Network - MLP with Math



Acc. = 86.6%

Final Results

	2 Class	3 Class	4 Class	5 Class	6 Class
Adaboost	0.83	0.61	0.55	0.5	0.58
SVM	0.8	0.58	0.63	0.58	0.47
ExtraTrees	0.8	0.55	0.61	0.61	0.55
MLP	0.94	0.72	0.75	0.66	0.61

Figure 1: Final accuracy results with less data more features

	2 Class	3 Class	4 Class	5 Class	6 Class
2 Layers	0.94	0.75	0.75	0.61	0.61
3 Layers	0.88	0.72	0.69	0.66	0.61
4 Layers	0.88	0.72	0.72	0.52	0.61
5 Layers	0.88	-	-	-	-

Figure 2: MLP’s accuracy with respect to layers

	2 Class	3 Class
ExtraTrees	0.75	0.57
MLP	0.84	0.71

Figure 3: Final accuracy results with less features more data

Best Features so Far

With More Features

- Weighted Methods
- Cyclomatic Complexity
- Number of Methods
- Line Count
- Density

With More Data

- Number of Methods
- Weighted Methods
- Passed Tests
- Density
- FailedTestsDontCoverStatement

What have we learned?

- Parsing data
- Technologies
- Target representation
- Correlation helps
- Dynamic metrics - JDCall Graph Problems
- Visualization!
- Neural Network
- More data is better than more features?
- Cross validation
- Suggestions

Thank you! Questions?