

Universidade Federal de Alagoas

Instituto de Computação  
Exploração e Mineração de Dados

## **vulnerabilidade** **Relatório**

Alunos:

Hyuri S. Maciel

Curso: Ciência da Computação

Professor: Balduino Fonseca

Maceió-AL  
23 de Maio, 2018

# Conteúdo

1	Introdução	1
2	Análise das Correlações	3
3	Testes Estatísticos	4
4	Testes de Modelos	5
5	Conclusão	6

# 1 Introdução

Métricas de software têm sido objeto de pesquisa a setenta anos, e as expectativas eram altas que métricas existem para ajudar na tomada de decisões gerenciais durante o software ciclo da vida. Essencialmente qualquer métrica é uma tentativa de medir ou prever algum atributo (interno ou externo) de algum produto, processo ou recurso. Normalmente, os atributos internos são aqueles que podemos diretamente medir, e os externos aqueles que estamos interessados em descobrir.

Para analisar e tentar solucionar o problema de vulnerabilidade de software utilizamos as seguintes bases de dados :

- glibc
- httpd
- kernel
- mozilla
- xen

Possuindo dois conjuntos das mesmas bases analisadas, possuindo um conjunto de base balanceada e outro não balanceado.

Cada base de dados descrita acima possui as seguinte variáveis:

- AltCountLineCode
- CountInput
- CountLineBlank
- CountLineCodeDec
- CountLineComment
- CountLinePreprocessor
- CountPath
- CountStmt
- CountStmtEmpty
- Cyclomatic

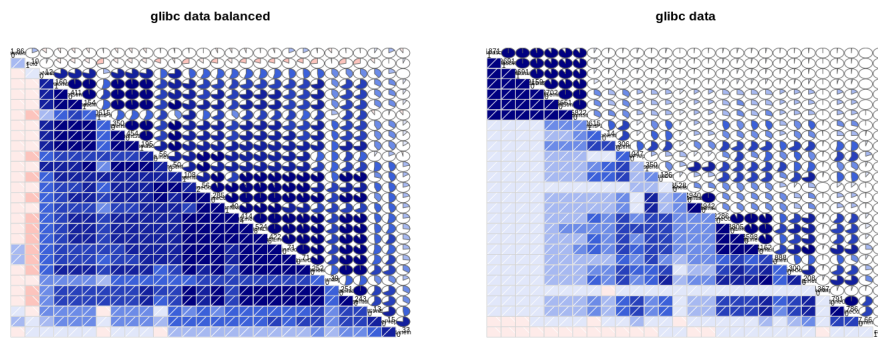
- CyclomaticStrict
- Knots
- MinEssentialKnots
- RatioCommentToCode
- AltCountLineComment
- CountLine
- CountLineCode
- CountLineCodeExe
- CountLineInactive
- CountOutput
- CountSemicolon
- CountStmtDecl
- CountStmtExe
- CyclomaticModified
- Essential
- MaxEssentialKnots
- MaxNesting

Também foram aplicados diferentes testes estatísticos. descrição das variáveis

## 2 Análise das Correlações

As correlações podem ser vistas nas figuras 1 e 2 as outras correlações podem ser encontradas no repositório Git.

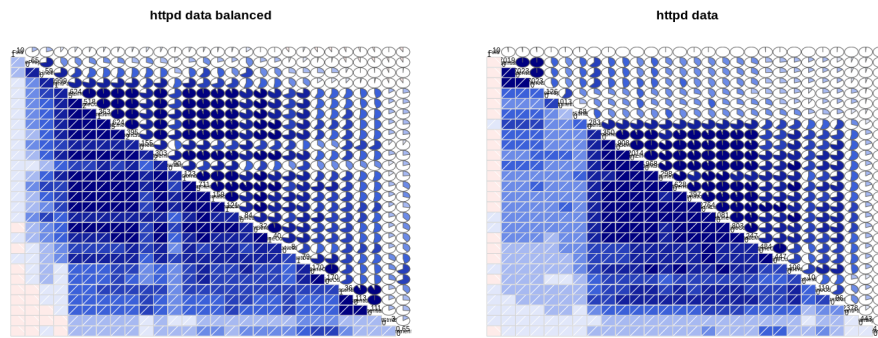
Na maioria as correlações são semelhantes. A análise das correlações é aplicada para todas as bases de dados. É notável que existe uma melhor correlação quando as bases são dos dados balanceados como podemos observar na figura 1(a) e 2(a)



(a) glibc balanced

(b) glibc unbalanced

Figura 1: Correlação glibc.



(a) httpd balanced

(b) httpd unbalanced

Figura 2: Correlação httpd.

### 3 Testes Estatísticos

Analizamos as bases de dados apresentadas as variáveis balanceadas e não balanceadas, os resultados dos **boxplot** foi semelhante. em média as variáveis são diferentes, dependendo ou não da base esta balanceada ou não. Podemos ver nas figuras 3, 4, 5, 6 e 7, algumas dessas relações, as outras estão no link Git.

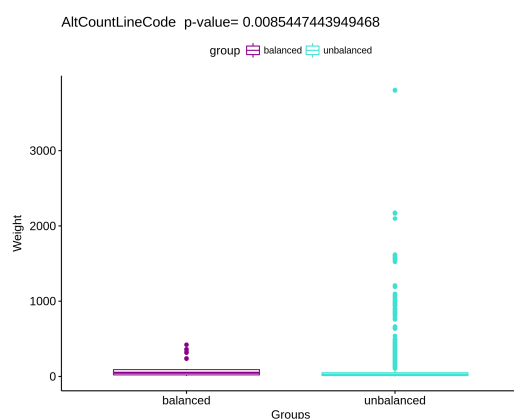


Figura 3: glibc

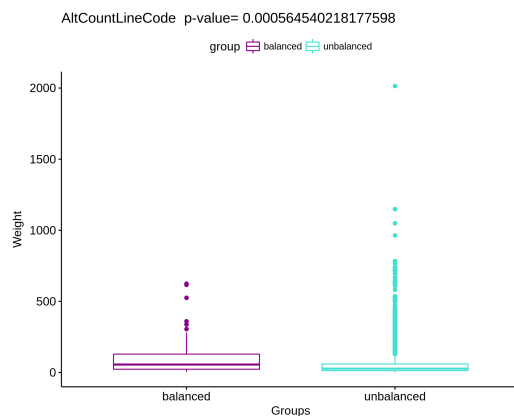


Figura 4: httpd

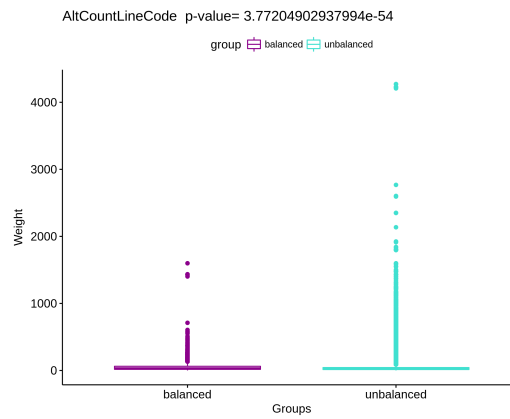


Figura 5: kernel

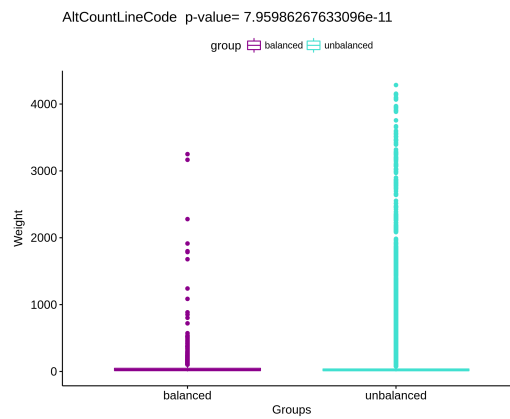


Figura 6: mozilla

## 4 Testes de Modelos

Com intuito de testar e classificar as bases de dados usamos os modelos de classificação: KNN, Naive Bayes e SVM. Para avaliar a efetividade utilizamos as métricas: Accuracy, Precision, Recall e F1-Score.

Entre as técnicas analisadas a que obteve um melhor desempenho foi a SVM, todos os dados estão nas planilhas que estão contidas no link: [Git](#).

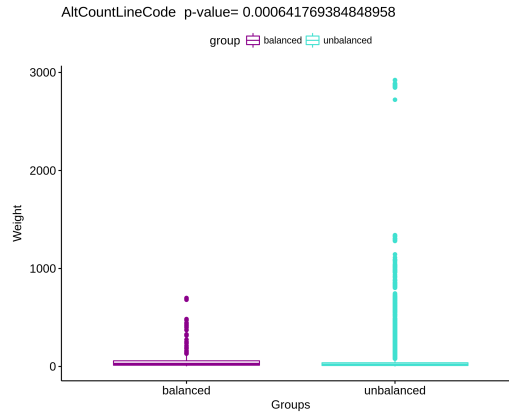


Figura 7: xen

Tabela 1: SVM

	Precision	Recall	F1	Accuracy
glibc	1.00	1.00	1.00	1.00
httpd	1.00	1.00	1.00	0.99
xen	1.00	0.89	0.95	0.99
glibc_balanced	1.00	1.00	1.00	0.35
httpd_balanced	1.00	1.00	1.00	0.48
kernel_balanced	1.00	1.00	1.00	0.62
mozilla_balanced	1.00	1.00	1.00	0.58
xen_balanced	1.00	1.00	1.00	0.55

## 5 Conclusão

Observamos que as bases balanceadas apresentam um melhor desempenho apos observar as métricas analisadas, as desbalanceadas tem em média 98,9% de accuracy.

Observamos que o melhor classificar também é o SVM, para um melhor analise de métricas podemos utilizar essa técnica.

Alguns resultados dao NaN porque podem ser do tipo 0/0. Como o conjunto de dados é pequeno, nem sempre sai as duas classes dentro de fold da validação cruzada. Um tratamento mais refinado nas bases pode ajustar esse problema.

As bases **kernel** e **mozilla** não balanceadas, não foi possível rodar, pois o computador não suportou.