

Instituto Federal do Maranhão – IFMA

Departamento Acadêmico de Informática – DAI

Curso: Sistemas de Informação

Aluno: Maykon Deykon Santa Barros

Matrícula: SI1312002-21

Professor: Josenildo Silva

Inteligência Artificial

Laboratório final

Sumário

- Resumo:.....3
- Preparação:.....3
- Execução:.....4
- Comparação:.....5
- Conclusão:.....7
- Referências:.....7

Resumo:

Laboratório de aprendizagem de máquina com o objetivo de realizar uma comparação entre algoritmos de classificação.

A partir de um dataset sobre doenças no coração, onde o objetivo é tentar identificar possíveis casos com base em uma série de atributos, tais como: idade, sexo, pressão arterial etc.

São utilizados três algoritmos de classificação: Naive Bayes, J48 e BIFReader. Cada um executado 10 vezes, e os dados da execução foram salvos em arquivo csv em anexo, que foram posteriormente usados para plotagem de gráficos comparativos.

Preparação:

Dataset

Statlog (Heart) Data Set – Dataset com 13 atributos, 270 instâncias utilizadas para identificação de doenças no coração.

Parâmetros:

1. *age*
2. *sex*
3. *chest pain type* (4 values)
4. *resting blood pressure*
5. *serum cholestoral in mg/dl*
6. *fasting blood sugar > 120 mg/dl*
7. *resting electrocardiographic results* (values 0,1,2)
8. *maximum heart rate achieved*
9. *exercise induced angina*
10. *oldpeak = ST depression induced by exercise relative to rest*
11. *the slope of the peak exercise ST segment*
12. *number of major vessels (0-3) colored by flourosopy*
13. *thal: 3 = normal; 6 = fixed defect; 7 = reversable defect*

Tipos de atributo

- *Real: 1,4,5,8,10,12*
- *Ordered: 11,*
- *Binary: 2,6,9*
- *Nominal: 7,3,13*

Variáveis de retorno

- *Absence (1) or presence (2) of heart disease*

API

- WEKA v3.8.0 – Coleção de algoritmos de aprendizagem de máquina em java para a tarefa de mineração de dados.

Algoritmos

- Naive Bayes – classificador da família de classificadores probabilísticos simples com base na aplicação de teorema de Bayes com fortes premissas de independência entre as características.
- J48 – implementação de código aberto em Java do algoritmo C4.5 no aplicativo de mineração de dados Weka.
- BIFReader – Variação do algoritmo Bayesiano usando o InterchangeFormat.

Execução:

Criada uma aplicação que utiliza a API do Weka onde se encontram os algoritmos citados e a classe de avaliação com método *crossValidateModel* que realiza a validação cruzada do dataset.

A aplicação realiza 10 iterações em cada algoritmo de classificação e guarda o resultado em um arquivo, denominado “*resultados.csv*”.

Resultados de uma das iterações:

```
Dataset: heart-statlog
Algoritmo: NaiveBayes

Results

Correctly Classified Instances      226           83.7037 %
Incorrectly Classified Instances    44           16.2963 %
Kappa statistic                    0.6683
Mean absolute error                 0.1835
Root mean squared error             0.3598
Relative absolute error             37.163 %
Root relative squared error         72.4003 %
Total Number of Instances          270

Tempo de execução: 174 milisegundos
```

Dataset: heart-statlog
Algoritmo: J48

Results

| | | |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances | 207 | 76.6667 % |
| Incorrectly Classified Instances | 63 | 23.3333 % |
| Kappa statistic | 0.5271 | |
| Mean absolute error | 0.274 | |
| Root mean squared error | 0.4601 | |
| Relative absolute error | 55.4778 % | |
| Root relative squared error | 92.5962 % | |
| Total Number of Instances | 270 | |

Tempo de execução: 178 milisegundos

Dataset: heart-statlog
Algoritmo: BIFReader

Results

| | | |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances | 219 | 81.1111 % |
| Incorrectly Classified Instances | 51 | 18.8889 % |
| Kappa statistic | 0.6172 | |
| Mean absolute error | 0.1982 | |
| Root mean squared error | 0.3666 | |
| Relative absolute error | 40.1338 % | |
| Root relative squared error | 73.7855 % | |
| Total Number of Instances | 270 | |

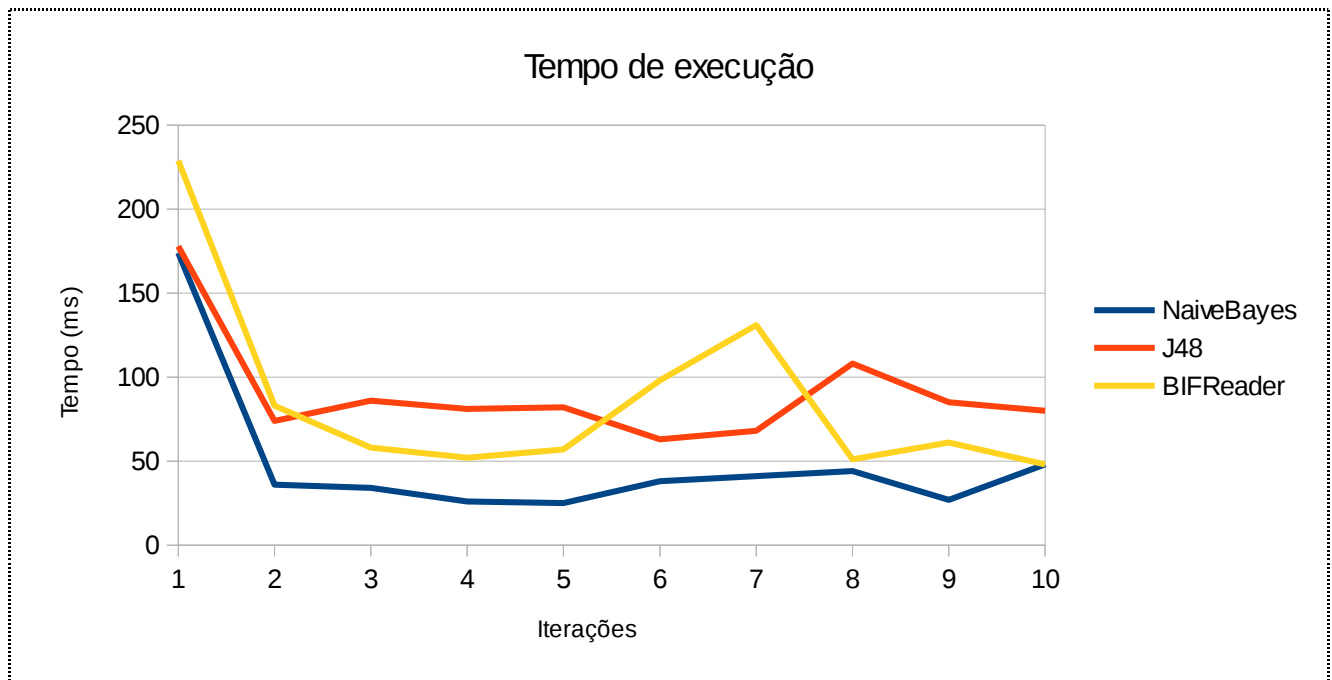
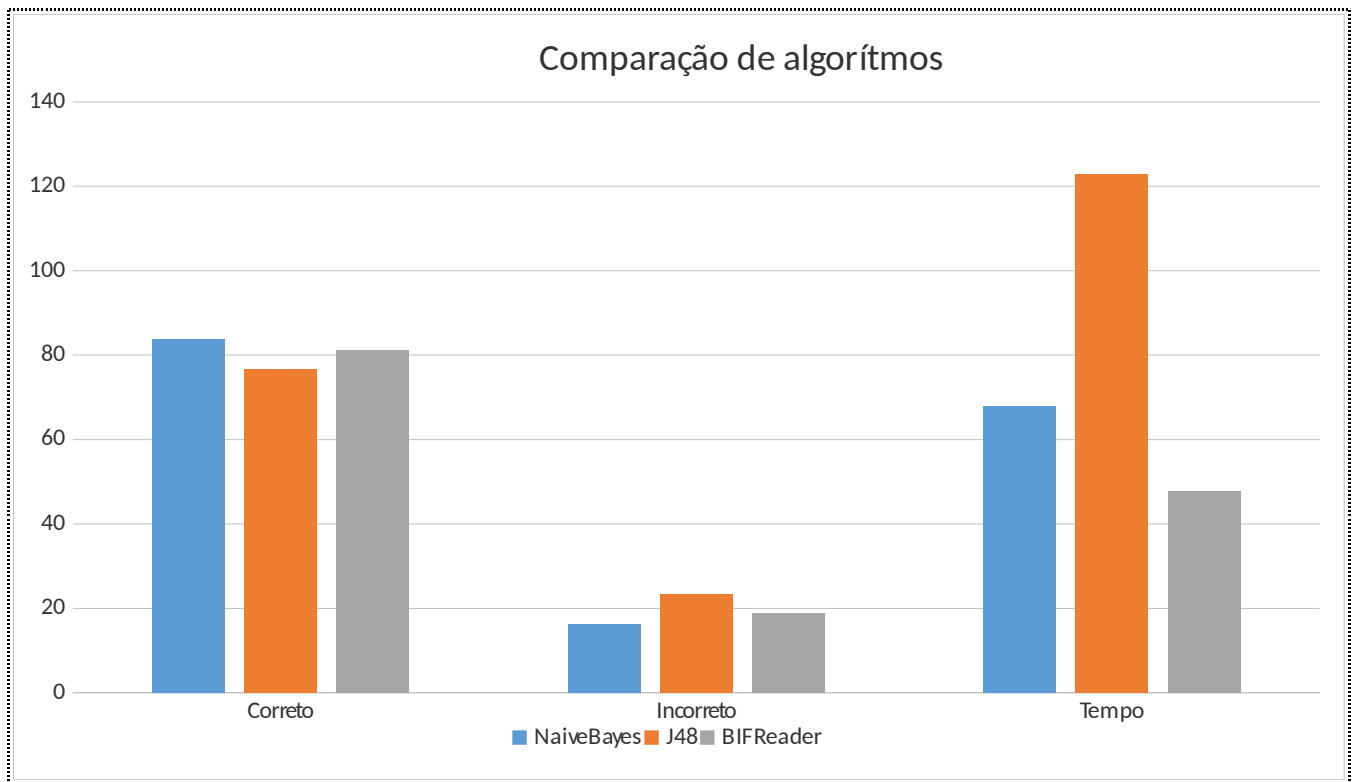
Tempo de execução: 48 milisegundos

Comparação:

Os seguintes dados foram utilizados na comparação dos classificadores:

- A taxa de instâncias corretamente classificadas
- A taxa de instâncias incorretamente classificadas
- O tempo de execução do algoritmo

Plotagem dos dados comparativos:



Conclusão:

Durante a execução do experimento fica evidente, neste dataset, que tanto o Naive Bayes quanto sua variação BIFReader possuem uma taxa de classificação correta muito próxima, e surpreendentemente melhor que o algoritmo J48. Vale notar que a implementação do J48 possui muito mais elementos que o tornariam mais eficiente, enquanto os Bayesianos utilizam como base apenas estatística, daí sua maior agilidade na execução.

O método de avaliação “*crossValidate*” também não proporciona o meio mais eficaz de treinamento para o algoritmo, podendo torná-lo “viciado”.

Referências:

Dataset: [https://archive.ics.uci.edu/ml/datasets/Statlog+\(Heart\)](https://archive.ics.uci.edu/ml/datasets/Statlog+(Heart))

API: <http://www.cs.waikato.ac.nz/ml/weka/>

Avaliador: [https://weka.wikispaces.com/Generating+cross-validation+folds+\(Java+approach\)](https://weka.wikispaces.com/Generating+cross-validation+folds+(Java+approach))

Algoritmos:

<http://weka.sourceforge.net/doc.dev/weka/classifiers/bayes/NaiveBayes.html>

<http://weka.sourceforge.net/doc.stable/weka/classifiers/bayes/net/BIFReader.html>

<http://weka.sourceforge.net/doc.dev/weka/classifiers/trees/J48.html>