**Mineração de dados educacionais em dados de pensamento computacional**

Freddy da Paz Ilha¹, Ariel Furtado¹, Gabriel Balota¹

¹Universidade Federal de Pelotas, Centro de Desenvolvimento Tecnológico, Pelotas-RS, Brasil.

{freddy.ilha,gmbalota,afazevedo}@inf.ufpel.edu.br

**Abstract:** This paper show an educational data mining made with the data gathered from a project produced on the Universidade Federal de Pelotas (UFPEL). This data provides the grades from tests made by the students and researchers from the university to stimulate the computacional thought on kids between nine and thirteen years old.

**Keywords:** Computational Thinking, Data Minning, Data Base, Research, Learning

**Resumo:** Este trabalho apresenta uma mineração de dados educacionais, com base em um projeto de extensão desenvolvido na Universidade Federal de Pelotas (UFPel). Esses dados apresentam avaliações feitas pelos estudantes e pesquisadores da UFPel para estimular o pensamento computacional por crianças entre nove e treze anos.

1. **Introdução**

A mineração de dados educacionais tem como principal objetivo desenvolver metodos para que seja possivel a exploração de um determinado conjunto de dados coletados em ambientes educacionais. Essa exploração vem para ajudar a melhorar o desenvolvimento da aprendizagem de estudantes de qualquer idade e em qualquer tipo de ensino.

Com as informações das relações dos dados e a descoberta de informações com essas relações podem ser muito uteis nas atividades de tomada de decisção. Por exemplo, ao minerar os dados de um estoque de supermercado poderia-se descobrir que todas as sextas-feiras uma marca específica de cerveja se esgota nas prateleiras e, portanto, um gerente que obtém esta “nova informação” poderia planejar o estoque do supermercado para aumentar a quantidade de cervejas desta marca as sextas-feiras.

**1.1. Problema**

O problema abordado por esse artigo é minerar uma base de dados com informações sobre um projeto realizado na Universidade Federal de Pelotas em três escolas do Municipio que objetiva fazer uma introdução à lógica em geral para as crianças, estimulando assim seu raciocinio lógico.

Para isso entao, fizeram três atividades em cada escola, Joao Silva, Ferreira Viana, e Pelotense, e cada atividade tinha um objetivo diferente.

* Atividade de Colorir com Numeros
* Atividade com Número Binário
* Atividade com Algoritmos de Ordenação

Com o intuito de analisar os conhecimentos previos de cada aluno, foi feito um pré-teste, que envolvia noções básicas de lógica e matemática. Para avaliar o ensino desenvolvido junto aos estudantes UFPel foi realizado um teste após das atividades.

Entramos assim no problema abordado por esse artigo, minerar essas informações com a finalidade de avaliar os dados obtidos com a ferramenta *Weka*.

**1.2. Motivação**

Uma mineração apurada pode obter informações totalmente relevantes nos dias de hoje é fundamental para o sucesso de muitas aplicações e muitos problemas. Tendo isso em vista, buscamos com essa pesquisa realizar uma mineração de dados na base de dados citada anteriormente com o objetivo de criar relações posteriores com essa informação e assim auxiliar com o projeto de ensinar lógica básica para as escolas do ensino primário

**2. Testes**

Nesta sessão será mostrado os testes realizados, bem como o pré-processamento realizado com a base de dados fornecida para a pesquisa.

**2.1. Pre-Processamento**

Para fazer a mineração dos dados, foi feito um pré processamento para juntar os dados dos três colégios de forma que fosse possível correlacionar os mesmos. Após juntar todos dados foi necessario normalizar os dados por que tinha muitos dados faltando ou irregulares.

Vale ressaltar que as notas após vários testes sem um resultado que se pudesse obter algo, foi feito então uma classe para as notas finais, onde houve o agrupamento da seguinte forma:

* *= 0:* Onde a nota zero, essa classe é discrepante, e pode ser principal motivo dos resultados ruins, já que toda perda de informação na nota, lhe foi atribuido uma nota 0.
* *<= 5:* Onde entra todas notas maior que zero, e igual a cinco, no caso da normalização com intervalo [0,1], então fica 0,5.
* *> 5: Todas notas no mesmo intervalo que são maiores que cinco e menores que um.*
* *= 1:* Notas que são um, no intervalo determinado.

**2.2. Teste 1**

Nesta arvore foram usadas todas as notas das tarefas realizadas nas escolas e podemos perceber que a atividade um não teve influencia na nota do teste final dos alunos, também é possivel notar que vinte alunos que tiraram zero ou não fizeram teste tiveram o mesmo resultado no pos teste

11654062_969805546392672_541082138_o.jpg

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 31 36.9048 %  
Incorrectly Classified Instances 53 63.0952 %  
Kappa statistic 0.307   
Mean absolute error 0.0413  
Root mean squared error 0.1649  
Relative absolute error 72.6123 %  
Root relative squared error 98.086 %  
Total Number of Instances 84

=== Stratified cross-validation ===  
=== Summary ===  
  
Correctly Classified Instances 62 73.8095 %  
Incorrectly Classified Instances 22 26.1905 %  
Kappa statistic 0.5929  
Mean absolute error 0.2092  
Root mean squared error 0.3452  
Relative absolute error 49.6563 %  
Root relative squared error 75.2755 %  
Total Number of Instances 84   
  
=== Detailed Accuracy By Class ===  
  
 TP Rate FP Rate Precision Recall F-Measure ROC Area Class  
 0.732 0.163 0.811 0.732 0.769 0.896 <=0,5  
 0.8 0.172 0.593 0.8 0.681 0.929 0.0  
 0.696 0.066 0.8 0.696 0.744 0.897 >0,5  
Weighted Avg. 0.738 0.138 0.756 0.738 0.741 0.904  
  
=== Confusion Matrix ===  
  
 a b c <-- classified as  
 30 9 2 | a = <=0,5  
 2 16 2 | b = 0.0  
 5 2 16 | c = >0,5

**2.3. Teste 2**

Foram usadas somente as notas finais das atividades finais e uma coisa interessante que é possivel notar é que os alunos que marcaram que gostam de “joguinhos” não tiveram notas ruins mostrando que jogos não são uma causa de nota ruim

teste 2.jpg

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 43 51.1905 %

Incorrectly Classified Instances 41 48.8095 %

Kappa statistic 0.2208

Mean absolute error 0.28

Root mean squared error 0.4256

Relative absolute error 83.6778 %

Root relative squared error 104.2672 %

Total Number of Instances 84

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

0.737 0.435 0.583 0.737 0.651 0.693 >0,5

0.56 0.186 0.56 0.56 0.56 0.62 0.0

0.059 0.149 0.091 0.059 0.071 0.512 <=0,5

0 0 0 0 0 0.266 1.0

Weighted Avg. 0.512 0.282 0.449 0.512 0.476 0.614

=== Confusion Matrix ===

a b c d <-- classified as

28 5 5 0 | a = >0,5

7 14 4 0 | b = 0.0

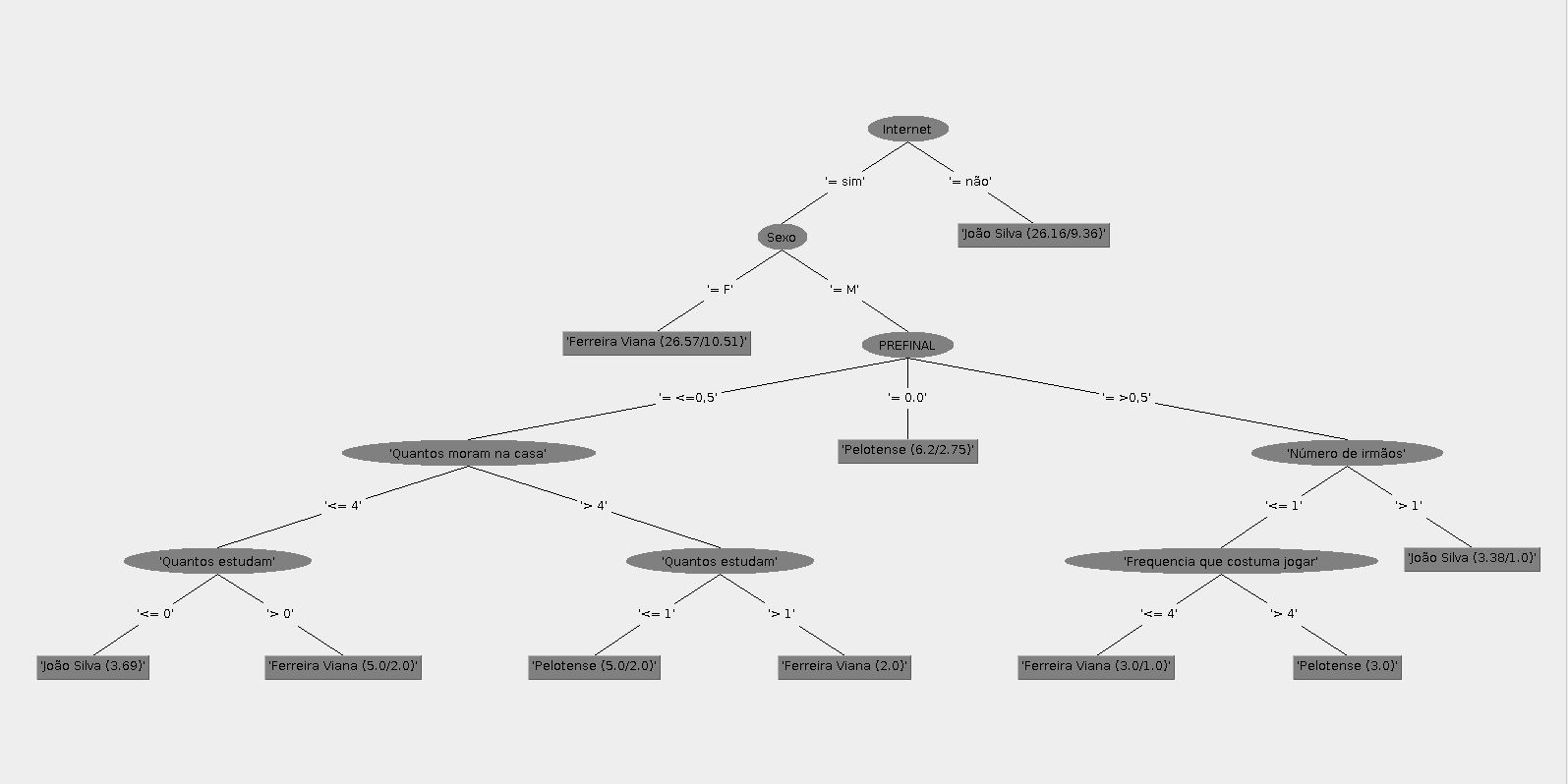
11 5 1 0 | c = <=0,5

2 1 1 0 | d = 1.0

**2.3. Teste 3**

Foi retirado todas as notas para tentar fazer uma analise mais socioeconômica, e assim pode-se notar, utilizando a variável “escola” para teste, que todos os que no possuiam internet estudavam na escola João Silva, isso é um resultado interessante para a pesquisa, pois pode determinar uma futura pesquisa, sobre como a internet influencia nos estudos dos alunos.

Podemos ver também que curiosamente todas as meninas que possuem internet estão na escola Ferreira Viana.



=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 35 41.6667 %

Incorrectly Classified Instances 49 58.3333 %

Kappa statistic 0.1231

Mean absolute error 0.414

Root mean squared error 0.5126

Relative absolute error 93.3699 %

Root relative squared error 108.8248 %

Total Number of Instances 84

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

0.483 0.345 0.424 0.483 0.452 0.596 Ferreira Viana

0.36 0.271 0.36 0.36 0.36 0.548 Pelotense

0.4 0.259 0.462 0.4 0.429 0.579 João Silva

Weighted Avg. 0.417 0.293 0.418 0.417 0.416 0.576

=== Confusion Matrix ===

a b c <-- classified as

14 8 7 | a = Ferreira Viana

9 9 7 | b = Pelotense

10 8 12 | c = João Silva

**3. Conclusão**

Foram realizados três testes de mineração relacionando dados de aprovação e reprovação nas atividades e nos testes executados com os alunos do ensino fundamental do municipio. Com esse testes de mineração foi possivel obter informações nas quais será possivel no futuro adaptar a maneira de ensino nas escolas. Afim de desenvolver pensamento computacional em crianças a partir de 9 anos.

Nota-se que as atividades apesar de serem atividades basicas, alguns alunos possuem imensas dificuldades. Em alguns casos com a exercitação do tipo de raciocinio das atividades, é possivel ver que alunos atingiram médias maiores do que as médias tiradas antes de qualquer atividade de raciocinio.

Com a mineração dessas informações foi possivel ver o quão dificil pode ser essas atividades, e onde é necessário para que a melhoria do ensino para que alunos tenham capacidade de resolver essa atividades.

Vale ressaltar que os resultados não foram razoavelmente bons, não importa o teste tentado, pois houve perda de informação relevante durante a coleta de dados, e isso se tornou um problema para a pesquisa.

Mas também, nota-se o quão importante um projeto dessa magnitude pode significar, sendo que se realizado em escala maior e com mais cuidado e tratado com mais importância, muita coisa pode ser obtida, como por exemplo relevância da internet para um bom raciocinio lógico, quais crianças que possuem internet a utilizam de fato para estudar, ou somente para fins de lazer.