Análise do desempenho acadêmico de alunos do Ensino Médio

Rafael Pinto Mesquita   
Pontifícia Universidade Católica de Minas Gerais  
Belo Horizonte  
rafael.mesquita.1276212@sga.pucminas.br

Marco Túlio de Oliveira Valeriano   
Pontifícia Universidade Católica de Minas Gerais  
Belo Horizonte  
marco.valeriano@sga.pucminas.br

Claudio Manoel Jansen de Oliveira  
Pontifícia Universidade Católica de Minas Gerais  
Belo Horizonte  
email@sga.pucminas.br

Felipe Amaral   
Pontifícia Universidade Católica de Minas Gerais  
Belo Horizonte  
email@sga.pucminas.br

Victor Hugo de Sá  
Pontifícia Universidade Católica de Minas Gerais  
Belo Horizonte  
email@sga.pucminas.br

Victor Hugo de Sá  
Pontifícia Universidade Católica de Minas Gerais  
Belo Horizonte  
vhsparreiras@sga.pucminas.br

RESUMO

A educação é uma das principais coisas presentes nas vidas de jovens, com isso, a mineração de dados nessa área é de suma impotância. Sabe-se que essa área não é nada simples de ser estudada, por isso esse trabalho tem como meta usar a mineração de dados para ajudar no estudo de problemas relacionados a instrução de alunos.  
 Para isso se utilizará de diferentes modelos de aprendizado para prever resultados e aumentar nosso conhecimento a cerca desse contexto. O trabalho ira utilizará algoritmos aprendidos em sala de aula para assim trazer resultados mais concretos e mais próximos da realidade.

KEYWORDS

Machine Learning, base de dados

1 Objetivo

A base de dados utilizada retrata diversos fatores que podem influenciar nas notas dos estudantes, como relação famíliar, renda financeira, estrutura escolar,etc. E sabe que tudo isso influencia diretamente em seu desempenho escolar, mas muitas vezes é ignorado pela escola e família.   
No artigo Fatores que facilitam e dificultam a aprendizagem públicado em 2018 pela revista Educação Publica é retratado como fatores socioecômicos, físicos e mentais são incomodos para o aprendizado real do aluno influenciando assim em suas notas, um exemplo usado pelo autor foi que se um aluno tem algum problema mental, ele terá mais dificuldade no aprendizado, visto que, em muitos casos o própio professor não está preparado para ensinar esse aluno.

Dessa maneira, esse trabalho tem como objetivo mostrar o desempenho academico levbando em consideração também outros fatores do aluno , fatores externos que podem atrapalha-lo em suas notas escolares.

2 Base da dados

A base de dados utilizada nesse trabalho foi retirado site UCI Machine Learning Repository. Ela é uma base que contêm 33 atributos, 16 númericos e 16 categóricos e foi feita a partir de um estudo nos anos de 2005 e 2006 nas escolas Gabriel Pereira e Mousinho da Silveira, em Portugal.

3 Pré-Processamento

Para realizar limpeza e adequação da base de dados para conseguir realizar um trabalho limpo e sem problemas, alguns atributos não foram utilizados, contudo foi nescessarioadicionar alguns outros a base. Com isso a base ficou com 53 atributos.  
Para construir novos atributos foi nescessário a utilização de alguns métodos, como o Pearson que tem o objetivo de indicar como as duas variáveis estão associadas estão entre si. Obtemos os resultados e assim tivemos que realizar outro procedimento, pois possivelmente a base iria dar problemas para executar algoritmos lineares.

Resolver esse problema é dificil, mas para isso criamos uma nova variável que englobava as duas que estavam em conflito, sabendo que elas tinham valores muito semelhantes. Os atributos eram relacionados ao consumo de alcool, um semanal, outro aos finais de semana.  
Foi descartados alguns atributos que não tem muita nescessidade de uso, como o G3 que é calculado com base nas notas g1 e g2.

4 Resultados Iniciais

Para realizar os testes iniciais, foram utilizados alguns algortimos, que são: Árvore de decisão, os resultados obtidos foram: para o Grid Search foi encontrado 4 no melhor valor para o ‘max\_depth’ e 54 para o ‘max\_features’.O MAE médio foi de 0.384 para o conjuento de treinamento, já o MAE médio dos testes foi de 2.71, e sabe-se que houve diretos erros no modelo. O valor do resultado do r² foi 0.87 e de teste foi 0.43, e com isso percebe-se que isso não é um resultado bom , descartando a possibilidade de usar a arvore gerada.

**REFERENCES**

[1] Messias, Manuel (2018).Fatores que facilitam e dificultam a aprendizagem. https://educacaopublica.cecierj.edu.br/artigos/18/14/fatores-que-facilitam-e-dificultam-a-aprendizagem

[2] UCI Machine Learning. 2016. Student Alcohol Consumption. https://archive.ics.uci.edu/ml/datasets/Student+Performance