PÓS-GRADUAÇÃO EAD EM COMPUTAÇÃO APLICADA À EDUCAÇÃO

Utilizando a Mineração de Dados como suporte à predição da reprovação em cursos técnicos integrados do CEFET

Gualberto Rabay[1], Carlos Diego Nascimento Damasceno[2], Seiji Isotani[2]
[1]Centro Federal de Educação Tecnológica de MG (CEFET MG),[2] Universidade de São Paulo (USP)

INTRODUÇÃO

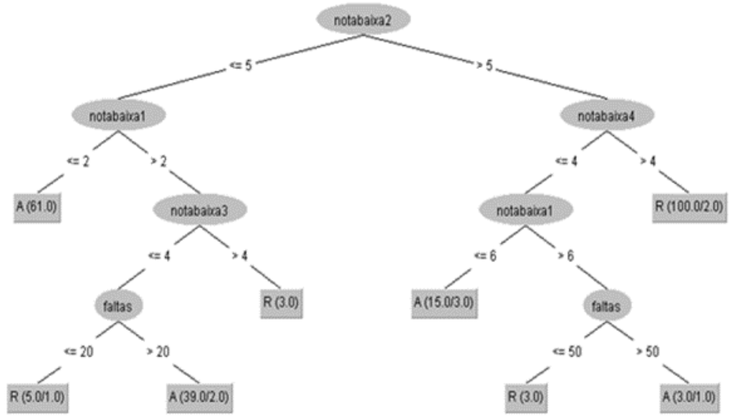
O fracasso escolar no ensino médio é um problema mundial que demanda ações pedagógicas para reduzi-lo de maneira geral e ações para identificação precoce dos estudantes em risco para uma atuação mais pontual. A evasão tem como causa principal a reprovação e, portanto, melhorando o desempenho dos alunos aumentam-se as taxas de permanência. O uso de técnicas de Mineração de Dados Educacionais (MDE) permite identificar mais precocemente os estudantes candidatos à reprovação fornecendo subsídios para que a gestão escolar possa atuar corretivamente nestes casos. O CEFET MG Campus de Nepomuceno apresenta elevados níveis de reprovação e evasão principalmente no primeiro ano do Ensino Médio.

Série	Situação	Quant.	%
1a	Aprovado	115	36,98
	Reprovado	114	36,67
	Evadido	82	26,37
2a	Aprovado	104	80,0
	Reprovado	12	9,2
	Evadido	14	10,8
3a	Aprovado	105	93,75
	Reprovado	3	2,67
	Evadido	4	3,57

Situação dos alunos

OBJETIVOS

Definir um modelo de predição de reprovação para Ensino Médio Integrado do CEFET MG de Nepomuceno com o objetivo de nortear e agilizar as ações dos gestores com o fim de mitigar o fracasso escolar. Para isso utilizaram-se os conceitos de MDE com algoritmos de Aprendizagem de Máquina para analisar dados acadêmicos de alunos dos anos 2018 e 2019 de três curso técnicos



Árvore de Decisão 1o. Ano

MATERIAS E MÉTODOS

Para desenvolver o projeto utilizaram-se os algoritmos Infogain e J48 na ferramenta Weka seguindo-se diversas etapas. Na primeira foi feita a seleção dos dados dos alunos. Na segunda etapa fez-se o pré-processamento filtrando-se dados de diversas fontes e definindo-se os atributos relevantes. Na terceira etapa os dados foram transformados para o formato usado pela Weka. Na quarta etapa primeiramente foi usado o algoritmo *Infogain* para eliminar atributos irrelevantes e por último foi usado o J48 para elaborar as árvores de decisão e a matriz de confusão. Estas etapas foram repetidas para cada um dos 3 anos do ensino médio.

Série	Acurácia	Precisão	Sensibilidade	F1	Área ROC
1a	88,64	0,868	0,918	0,890	0,941
2a	93,10	0,944	0,981	0,962	0,718
3a	97,22	0,972	1,00	0,986	0,714

Parâmetros por ano

RESULTADOS

O primeiro ano é que apresenta a maior taxa de reprovação e evasão merecendo mais atenção. O uso do *Infogain* validou a exclusão de vários atributos irrelevantes na predição. O J48 classificou corretamente 88,6% das instâncias. A matriz de confusão identificou 10 falsos negativos e 16 falsos positivos produzindo os valores de 0,87 para a precisão, 0,91 para a sensibilidade e 0.89 de F1. Pela árvore gerada a nota baixa no segundo bimestre é o atributo principal na predição da reprovação. Para os outros anos onde as taxas de reprovação são baixas a predição não demonstrou trazer benefícios significativos.