

## **Um modelo para mineração de dados referente a evasão de alunos do ensino superior**

Rogers Prates de Pelle<sup>1</sup>, Edimar Manica<sup>1</sup>

<sup>1</sup>Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul (IFRS) - *Campus* Ibirubá. Ibirubá, RS, Brasil.

Instituições de ensino superior têm sofrido com o constante aumento nas taxas de evasão em seus cursos. Tanto a redução na quantidade de matrículas como a não conclusão dos cursos pela maioria dos estudantes, podem ser explicadas em partes pelo cenário econômico que o país vive, mas a constatação é que essa interrupção prejudica a sociedade nos aspectos social, econômico e humano. Este trabalho tem por objetivo traçar um perfil dos alunos que fazem parte das estatísticas de evasão, por meio de algoritmos de aprendizado de máquina, aplicados sobre dados socioeconômicos de estudantes dos cursos de graduação. As seguintes informações foram analisadas: faixa etária, sexo, renda familiar per capita, cor, se possui necessidade especial, se utilizou cota, meio de ingresso e ensino médio público ou privado. Para compor a base de dados, foram obtidas essas informações de todos os alunos dos cursos de graduação do Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul – *Campus* Ibirubá, totalizando 618 discentes, cujo os dados se encontravam preenchidos. Estes dados foram submetidos a uma etapa de pré-processamento, onde os campos que poderiam identificar os indivíduos, como nome e CPF, foram completamente removidos, por se tratarem de dados sensíveis. Os dados foram separados em uma classe "sim", que é composta por alunos que não estão mais frequentando o curso. E a classe "não" formada por alunos que estão regularmente matriculados ou já concluíram o curso. Do total de instâncias, 208 pertencem à classe "sim" e 410 à classe não. A ferramenta de mineração de dados escolhida foi o Weka com o algoritmo Naive Bayes, onde foram executados testes utilizando o método de validação cruzada com 10 partições. A tarefa de classificação obteve uma medida F de 80%, sendo o *recall* da classe "sim" igual a 68%. Esse resultado ocorreu porque o conjunto de dados está desbalanceado. Os preditores mais divisivos na classificação foram os dados de faixa etária e sexo. Alunos com idade até 29 anos possuem taxa de evasão de 24%, contra 73% dos estudante com idade superior a 29 anos. Com relação ao gênero, 47% das mulheres evadem, contra 20% dos homens. Esses dados permitem a identificação dos discentes com maior probabilidade de evasão e a adoção de medidas de permanência e êxito mais direcionadas. Como trabalhos futuros destaca-se a inclusão de outros atributos e a expansão da base de dados utilizando dados de todos os estudantes do IFRS.

**Palavras-chave:** Mineração de Dados, Aprendizado de Máquina. Evasão no Ensino Superior.