Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 8, n. 1, 2021.

Trabalho apresentado no XL CNMAC, Evento Virtual - Co-organizado pela Universidade do Mato Grosso do Sul (UFMS).

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Estimativa da probabilidade de reprovação em disciplinas do curso de Licenciatura em Matemática: Uma aplicação de Regressão Logística

Stella Faria Costa[1]
IFSP, São José dos Campos, SP
Michael Macedo Diniz[2]
IFSP, São José dos Campos, SP

O alto número de reprovações é um dos problemas mais frequentes em cursos de nível superior, sendo ainda mais evidente em disciplinas da área de Ciências Exatas [1]. De acordo com [2], entre os anos 1990 e 1995 o percentual de reprovação em Cálculo Diferencial e Integral na Universidade de São Paulo - USP chegou a 75%. Estudos como [3] mostram que as dificuldades enfrentadas pelos discentes em disciplinas como Cálculo Diferencial e Integral não são facilmente superadas, e acarretam em retenção e futuro abandono do curso.

Diante deste cenário, utilizando a Mineração de Dados Educacionais (Educational Data Mining - EDM), uma área emergente da Mineração de Dados (Data Mining - DM), é possível converter dados provenientes de sistemas educacionais em informações úteis que podem ser utilizadas com o intuito de entender e melhorar o processo de ensino aprendizagem dos alunos. Sendo assim, serão propostos três modelos de regressão logística para estimar a probabilidade de um aluno do curso de Licenciatura em Matemática do IFSP-SJC reprovar em disciplinas da área de exatas.

Com o auxílio da secretaria do campus e do Sistema Unificado de Administração Pública (SUAP), foram obtidas informações sobre os alunos matriculados nos 4 anos de execução do curso no campus, e a partir dessas informações foi construída a base de dados, onde foram selecionados dados importantes e criadas variáveis para ajuste dos modelos. Assim, com o banco de dados definido, iniciamos o pré-processamento dos dados e a construção e avaliação do modelo de regressão logística utilizando o software RStudio.

Foram propostos três modelos de regressão logística, designados da seguinte maneira para identificar, previamente, alunos propensos a reprovação nas disciplinas exatas do curso:

- Modelo 1: Trata apenas das disciplinas de primeiro ano;
- Modelo 2: Trata das disciplinas de segundo, terceiro e quarto ano;
- Modelo 3: Envolve todas as disciplinas.

Para estudar o comportamento da variável dependente (1-Reprovado, 0-Aprovado) e ajustar os modelos propostos, foram analisadas 14 variáveis preditoras: Sexo; Idade; Nota do ENEM; Tipo de escola que cursou o ensino médio (dependência administrativa); Qual semestre do curso; Disciplina de primeiro ano; Quantidade de reprovações por falta na disciplina; Quantidade de reprovações por nota na disciplina; Quantidade de reprovações na disciplina; Quantidade de reprovações por falta no curso; Quantidade de reprovações por nota no curso; Quantidade de reprovações no curso; Índice de Rendimento do Aluno (IRA); Tempo de matriculado no curso na data de início da disciplina

[1]stella.costa@aluno.ifsp.edu.br
[2]michael.diniz@ifsp.edu.br

(em anos). Com essas variáveis foram aplicadas técnicas e procedimentos específicos de regressão logística para que pudéssemos preparar os dados e obter também melhor desempenho dos modelos.

Para construção de cada modelo, foi realizada a divisão do conjunto de treinamento e teste, sendo amostras de 30% para teste e 70% para treino e validação, sendo utilizado o método *K-Fold* de Cross Validation, com *K=10*. Os modelos foram analisados e avaliados com base nas seguintes métricas: acurácia, especificidade, sensibilidade, odds ratio e área sob a curva ROC (AUC). Como resultado, para o Modelo 1 e Modelo 2 foram apontadas 5 variáveis preditoras, já para o Modelo 3 foram apontadas 7 variáveis.

Tabela 1: Variáveis preditoras apontadas como significativas para cada modelo.

Variáveis preditoras	Modelo 1	Modelo 2	Modelo 3
Sexo			
Idade	X	X	X
Nota do ENEM	X		X
Escola que cursou o ensino médio			
Qual semestre do curso	X		X
Disciplina de primeiro ano			
Quantidade de reprovações por falta na disciplina			
Quantidade de reprovações por nota na disciplina			
Quantidade de reprovações na disciplina	X		
Quantidade de reprovações por falta no curso		X	X
Quantidade de reprovações por nota no curso		X	X
Quantidade de reprovações no curso	X		
IRA		X	X
Tempo de matriculado no curso		X	X

Analisando o desempenho dos modelos através do Cross Validation, eles produziram acurácias superiores a 70%, sensibilidade acima de 75%, especificidade acima de 60% e AUC superiores a 0.75, indicando que o Modelo 1 e o Modelo 3 possuem discriminação aceitável e o Modelo 2 discriminação excelente. A partir das amostras de teste, observou-se que os modelos alcançaram valores para acurácia superiores a 65%, sensibilidade acima de 55%, especificidade acima de 74% e AUC superiores a 0.75. Os resultados apresentados a partir do Cross Validation e das amostras de teste se assemelham bastante, o que indica fiabilidade nas predições realizadas pelos modelos.

Com este trabalho é possível que logo no início de uma disciplina sejam identificados, a partir de alguns fatores, alunos que têm maior probabilidade de reprovação, assim, podem ser tomadas algumas medidas e precauções por parte dos professores, tais como direcionar e intensificar a atenção para certos públicos, pensar em metodologias que auxiliem os alunos com idade mais avançada ou até mesmo destinar um horário extra para atender somente as dúvidas destes alunos.

Referências

[1] Barbosa, A. C. de C. e Concordido, C. F. R. Ensino colaborativo em ciências exatas, *Ensino, Saúde e Ambiente*, v. 2, n. 3, 2009.

[2] Barufi, M. C. B. A Construção/Negociação de Significados no Curso Universitário Inicial de Cálculo Diferencial e Integral, Tese de Doutorado, USP, 1999.

[3] Silva, A. C., et al. Análise dos índices de reprovação nas disciplinas de Cálculo I e AVGA do curso de Engenharia Elétrica do Instituto Federal da Bahia de Vitória da Conquista, *XIV International Conference on Engineering and Technology Education*, 2016.