

**Proceeding Series of the Brazilian Society of Computational and Applied Mathematics**

---

## Aplicação de Modelo Computacional na Análise da Evasão Discente no Curso de Matemática da UFRRJ

Felipe de Andrade Vieira<sup>1</sup>

Discente do Programa de Pós-graduação em Modelagem Matemática e Computacional - UFRRJ

Italo G. do Vale<sup>2</sup>

Discente do Programa de Pós-graduação em Ciência, Tecnologia e Inovação Agropecuária

Robson M. da Silva<sup>3</sup>

Docente do Programa de Pós-Graduação em Modelagem Matemática e Computacional - UFRRJ

### 1 Introdução

A evasão nas instituições de Ensino Superior é um problema mundial. Que afeta o resultado dos sistemas educacionais, gerando prejuízos significativos para o avanço das nações. Os discentes que iniciam seus estudos e não concluem são desperdícios sociais, econômicos e acadêmicos. A busca de suas causas vem sendo objeto de estudos de pesquisadores e órgãos governamentais em diversos países, incluindo o Brasil. Segundo [1], a taxa de evasão no primeiro ano dos cursos de graduação chega a ser duas a três vezes maior do que nos anos posteriores. De acordo com a literatura, a evasão é um objeto complexo, pois, reúne diversas variáveis e muitas inter-relações entre elas que resultam no processo de decisão do aluno em permanecer ou não no curso. Uma proposta que vem despertando interesse dos pesquisadores é a utilização de técnicas de inteligência computacionais [2,3], que vem mostrando bons resultados na predição de evasão discente em Instituições de Ensino Superior. O objetivo desse trabalho é desenvolver um modelo baseado em Máquina de Vetor de Suporte (SVM) para auxiliar nas tomadas de decisão quanto à evasão acadêmica.

### 2 Materiais e Métodos

O modelo SVM consiste na seleção de um hiperplano que minimize o risco estrutural, a partir da resolução de um problema convexo quadrático. Quando associada à função de kernel possibilita a construção de classificadores não-lineares, através do mapeamento dos dados em um espaço de dimensão superior ao original. Estas características garantem ao SVM uma boa capacidade de categorizar corretamente dados não relacionados na amostra

---

<sup>1</sup>lipede88@yahoo.com.br

<sup>2</sup>italovale@ufrj.br

<sup>3</sup>robsonms@ufrj.br.

de treinamento. A base de dados consiste de 490 discentes calouros do período de 2010.1 a 2015.1, obtidos junto a Pró Reitoria de Graduação da UFRRJ. Os atributos utilizados são listados a seguir: Ano de ingresso na instituição, gênero, distância da residência ao campus da UFRRJ, notas ENEM (Linguagem, Códigos e suas Tecnologias; Ciências da Natureza e suas Tecnologias; Redação; Matemática e suas Tecnologias e Ciências Humanas e suas Tecnologias), Notas nas disciplinas cursadas no primeiro período (Cálculo I, Álgebra I e Computação I) e situação referente a evasão. A função de núcleo utilizada foi a RBF (Radial Basis Function) e os parâmetros ( $\gamma$ ) e C (custo) foram definidos utilizando validação cruzada no conjunto de treinamento. O modelo SVM foi implementado utilizando o software R e o pacote kernlab [4], que utiliza o algoritmo Sequential Minimal Optimisation (SMO) para solução do problema dual. Este algoritmo encontra analiticamente a solução global, sem recorrer a métodos numéricos. Apesar de precisar de maior número de iterações para obter a convergência, o SMO é bastante veloz, visto que realiza várias operações de baixo custo computacional. O conjunto de dados foi dividido ao mero acaso em dois conjuntos independentes, conjunto de treino composto com 80% do total das amostras e o teste com os outros 20% restantes. Para avaliar a capacidade de generalização do modelo, foi utilizado o valor da acurácia no conjunto de 50 simulações.

### 3 Resultados e Conclusões

De acordo com a análise dos resultados foi possível evidenciar o desempenho promissor do modelo proposto, visto que o mesmo obteve acurácia média acima de 82,14% no conjunto de teste. A variável que apresentou maior impacto na performance do modelo SVM proposto foi a nota obtida na disciplina Álgebra I. Fato este que nos possibilita concluir que esta variável possui um impacto superior a qualquer outra variável utilizada pelo modelo proposto na determinação da evasão discente no curso de Matemática. Para confirmar a eficácia do modelo, novos testes serão realizados com a utilização de outros algoritmos de classificação (Naive Bayes e k-Nearest neighbor).

### Referências

- [1] R. S. Filho, P. Motejunas, O. Hipólito e M. Lobo. *A Evasão no Ensino Superior Brasileiro*. In: Cadern de Pesquisa, São Paulo, v.37, n. 132, p. 641 – 659, 2007.
- [2] R. Lameira, R. M da Silva. *Aplicação de Rede Bayesiana na identificação de fatores que causam a evasão no curso de Sistemas de Informação da UFRRJ*. In: *Anais da I Escola Regional de Sistemas de Informação do Rio de Janeiro*. p. 13-16. Niterói: SBC, 2014.
- [3] V. R.C. Martinho, Sistema inteligente para a predição de grupo de risco de evasão discente. 2014. 145 f. Tese de doutorado. Universidade Estadual Paulista Júlio de Mesquita Filho, Faculdade de Engenharia de Ilha Solteira, 2014.
- [4] A. Karatzoglou, A. SMOLA, K. Hornik. Kernel-Based Machine Learning Lab, *Journal of Statistical Software*, v.11, n.9, .1-9, 2004.