

**PREDIÇÃO À EVASÃO ESCOLAR: Estudo de caso aplicado no IFSULDEMINAS –  
Campus Passos**

**Carla Fernandes da SILVA<sup>1</sup>; Clayton Silva MENDES<sup>2</sup>.**

**RESUMO**

A evasão escolar é um dos principais desafios a ser superado pelas instituições educacionais, pois é o principal fator que contribui para a frustração do processo educacional. Identificar esses alunos a tempo de impedir o desligamento definitivo é uma tarefa complexa e o desenvolvimento de ferramentas computacionais pode auxiliar nesta tarefa. O presente trabalho apresenta uma análise para prever alunos com risco de evasão nos cursos de técnicos presenciais do IFSULDEMINAS - Campus Passos e determinar os principais fatores que levam à evasão. O trabalho foi realizado utilizando coleta de dados físicos e o processamento e testes com classificadores computacionais. Em dados obtidos até o presente momento, o classificador mais vantajoso foi o *Naives Bayes* e o uso desta ferramenta é um importante passo para traçar o perfil do aluno com tendência a abandono do curso, extrair métricas de evasão por curso e auxiliar gestores na elaboração das políticas educacionais.

**Palavras-chave:** Evasão escolar; Mineração de dados; Políticas Educacionais.

**1. INTRODUÇÃO**

A evasão escolar é um tema constante no contexto educacional e tem lugar de destaque no mesmo, pois se apresenta como um desafio a ser superado pelas instituições, pela sociedade e pelo Estado, pois contribui para frustração do processo educacional em virtude do afastamento do aluno da escola.

Nos cursos técnicos, ela representa um problema complexo e multifatorial. Existe a preocupação dos governos e das instituições em diminuir e controlar os índices de evasão nestes cursos. Os debates e reflexões deste tema na educação brasileira ocupam um espaço de relevância no cenário das políticas públicas educacionais. As dificuldades de acesso e permanência na educação básica são marcas significativas do sistema educacional brasileiro (SCHWARTZMAN et al, 1993).

A evasão escolar pode ser classificada em três tipos: a) econômica – a impossibilidade de manutenção do vínculo por questões socioeconômicas; b) vocacional – escolha inadequada aos interesses do aluno; c) institucional – associada ao fracasso nas disciplinas iniciais,

---

<sup>1</sup> USP - Universidade de São Paulo - Ribeirão Preto/SP - Email: carlanandess@usp.br

<sup>2</sup> UNESP - Universidade Estadual Paulista - Franca/SP - Email: claytonaux@hotmail.com

deficiências prévias de conteúdo, inadequação nos métodos de estudo e dificuldades de relacionamento em geral (BARROSO e FALCÃO, 2004).

Identificar alunos em risco de evasão é um dos problemas mais complexos da educação e motivo de preocupação nos vários níveis e modalidades de ensino. Nesse contexto, é importante o desenvolvimento de métodos e instrumentos eficientes e eficazes para predição, avaliação e acompanhamento de estudantes em risco de evasão (MARTINHO, 2013).

Os estudos de evasão no Brasil estão voltados na sua grande maioria para ensino básico e conforme Cravo (2012) não existem muitos estudos sobre o tema em ensino técnico, enquanto em outros países é um tema de estudo bastante frequente. É com base nessa lacuna sobre evasão em ensino técnico que este trabalho visa identificar precocemente alunos que apresentem risco de evasão utilizando sistemas computacionais, através de técnicas de aprendizagem de máquinas, aplicados no Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas (IFSULDEMINAS), campus de Passos.

Atualmente a identificação dos alunos evadidos na instituição é feita manualmente pela equipe pedagógica da escola e é uma tarefa muito demorada, árdua e tardia, realizada muitas vezes sem tempo para reversão da situação dos alunos. Portanto, a adoção de mecanismos automatizados fornecem informações para as tomadas de decisão e identificação de grupos de riscos de evasão e eventualmente minimizar este problema.

## **2. MATERIAL E MÉTODOS**

A princípio, o universo de interesse nesta pesquisa foram os estudantes matriculados no ano de 2013 nos quatro cursos técnicos subsequentes presenciais ofertados no referido ano. Os cursos têm duração média de um ano e meio a dois anos, turno noturno e ingresso anual, portanto são turmas que já concluíram o período de integralização do curso. Foram analisados os seguintes cursos técnicos: Informática, Comunicação Visual, Produção Moda e Enfermagem. Os atributos selecionados para a predição e identificação da evasão foram: sexo, idade, etnia, estado civil, renda familiar, procedência escolar e as notas das disciplinas do primeiro semestre dos cursos. Esses dados foram coletados manualmente pela ficha da matrícula e planilhas adicionais fornecidas pelo setor de registro acadêmico do campus. Os alunos tiveram a sua identificação preservada.

A base de dados ficou composta por informações de 54 alunos que concluíram os cursos e mais 45 alunos que não concluíram. A base foi fracionada em duas classes distintas e bem definidas, mantida a separação por curso. A primeira classe, composta por alunos que completaram todos os requisitos para aprovação e conclusão do curso e a segunda classe, composta por alunos que não concluíram o curso (evadiram, reprovaram, etc.). A Tabela 1 apresenta o número de alunos para cada situação.

**Tabela 1. Alunos em cada curso**

<b>Curso</b>	<b>Concluintes</b>	<b>Evadidos</b>	<b>Total de alunos</b>
<b>Produção em moda</b>	9	10	19
<b>Enfermagem</b>	12	16	28
<b>Comunicação visual</b>	15	9	24
<b>Informática</b>	18	10	28

Foram selecionados três classificadores de árvore de decisão: *SimpleCart*, *J48* e *RandomForest* e um classificador probabilístico simples baseado na aplicação do teorema de *Bayes* chamado *Naive Bayes* (DEKKER; PECHENIZKIY; VLEESHOUWERS, 2009). Para facilitar esta tarefa foi utilizada a ferramenta *Weka* (*Waikato Environment for Knowledge Analysis*). A ferramenta foi escolhida para este trabalho por ser uma ferramenta livre e a presença de várias versões de algoritmos empregados (HALL et al., 2009).

A base de dados foi dividida por curso e definimos para fins de utilização da mineração de dados duas classes distintas de alunos a serem pesquisadas: a) alunos que concluíram o curso; b) alunos evadidos. Os classificadores aplicados foram executados 10 vezes, valor padrão de configuração do ambiente através do método de validação cruzada (10 *folds cross-validation*).

### **3. RESULTADOS E DISCUSSÕES**

Os resultados obtidos até o momento mostraram que a acurácia (proximidade entre o valor obtido experimentalmente e o valor real) varia em média em torno de 75% a 94%. A tabela 2 mostra a acurácia média dos em diferentes classificadores:

**Tabela 2. Acurácia média dos classificadores**

Classificadores	Acurácia média por curso			
	Informática	Com. Visual	Produção Moda	Enfermagem
<i>SimpleCart</i>	64,28%	54,16%	89,47%	92,86%
<i>J48</i>	64,28%	66,66%	94,75%	82,14%
<i>RandomForest</i>	75%	79,16%	89,47%	85,71%
<i>Naive Bayes</i>	67,86%	79,16%	89,47%	85,71%

#### 4. CONCLUSÕES E CONSIDERAÇÕES FINAIS

O trabalho comparou o desempenho de quatro classificadores amplamente utilizados (*Naive Bayes*, *Simpcart*, *Randomflorest* e *J48*) para verificar qual apresentaria o melhor resultado para o problema em questão e os números obtidos foram bastante próximos, com uma pequena vantagem para o *Naive Bayes* analisando a matriz de confusão, outras configurações deverão ser testadas e uma análise estatística deverá ser realizada. A pesquisa mostrou até o momento que a utilização de softwares no controle da evasão escolar é uma importante ferramenta para a percepção e tomada de decisões estratégicas e estabelecimento de políticas educacionais no combate ao abandono dos estudos. O próximo passo do estudo é investigar o peso de cada um dos atributos no processo de evasão, traçando o perfil do aluno.

#### REFERÊNCIAS BIBLIOGRÁFICAS

- BAKER, R.; ISOTANI S.; CARVALHO A. **Mineração de Dados Educacionais: Oportunidades para o Brasil**. Revista Brasileira de Informática na Educação. Vol. 19, n. 02, p. 03, 2011.
- BARROSO, M. F.; FALCÃO, E. B. M. **Evasão Universitária: O Caso do Instituto de Física da UFRJ**. IX Encontro Nacional de Pesquisa em Ensino de Física. Jaboticatubas, 2004.
- CRAVO, A. C. **Análise das causas da evasão escolar do curso técnico de informática em uma faculdade de tecnologia de Florianópolis**. Revista Gestão Universitária na América Latina. Florianópolis, 2012.
- DEKKER, G.W.a; PECHENIZKIY, M.b; VLEESHOUWERS, J.M.a. Predicting students drop out: A case study. **EDM'09 - Educational Data Mining 2009: 2nd International Conference on Educational Data Mining** p. 41–50, 2009. Disponível em <<http://www.educationaldatamining.org/EDM2009/uploads/proceedings/dekker.pdf>>.
- HALL, Mark et al. The WEKA data mining software. **SIGKDD Explorations Newsletter** v. 11, n. 1, p. 10, 2009. Disponível em: <<http://portal.acm.org/citation.cfm?doid=1656274.1656278>>.

MARTINHO, V. R. C. **Sistema Inteligente para a Predição de Grupo de Risco de Evasão Discente.** UNESP, 2013.

SCHWARTZMAN, S.; DURHAM, E. R.; GOLDEMBERG, J. **A educação no Brasil em uma perspectiva de transformação.** Universidade de São Paulo. São Paulo, 1993.