

Utilização de R e Aprendizagem de Máquina para Auxiliar no Controle de Evasão no IFMS

Gabrielle Helpis dos Santos¹, Gilson Saturnino dos Santos¹

¹Instituto Federal do Mato Grosso do Sul – Coxim - MS

gabrielle.helpis@gmail.com, gilson.santos@ifms.edu.br

Resumo

A Evasão é um problema crescente nas Instituições de Ensino Superior. Desse modo, torna-se cada vez mais necessário um plano de ação para combatê-la. Este trabalho teve por objetivo a criação de classificadores utilizando Aprendizagem de Máquina, para identificar alunos que possam evadir de cursos superiores do IFMS, a partir das respostas do questionário socioeconômico. Após a criação dos classificadores, foram realizadas comparações entre duas tecnologias, por meio da comparação de porcentagem de acurácia entre os classificadores criados com o Weka e com o R. Os resultados mostram a importância de se realizar tal comparação e os resultados superiores da tecnologia R.

Palavras-chave: Aprendizagem de Máquina, R, Evasão Escolar.

Metodologia e desenvolvimento

Inicialmente foi realizada uma pesquisa para a atualização nas referências sobre evasão escolar, onde foi possível perceber que o problema persiste principalmente nos cursos de graduação (FREITAS; ALVES; COSTA, 2017).

Após, foi realizado uma pesquisa sobre a linguagem de programação R, utilizada normalmente para fins estatísticos e de análise de dados, e também sobre o RStudio, um ambiente de desenvolvimento de software livre, integrado para R (AMARAL, 2016).

O banco de dados do questionário socio econômico respondido pelos alunos do IFMS foi utilizado para gerar os classificadores, passando antes por uma etapa de pré-processamento no RStudio para se encaixar em uma formatação de dados possível de ser analisada pelo R.

A etapa de análise de dados, dada após o pré-processamento do banco de dados, foi realizada com 4 algoritmos: *Naive Bayes* (NB), *Multilayer Perceptron* (MP), *Support Vector Machine* (SVM) e *J48*, utilizando a metodologia *cross-validation* com 10 partes. A base de dados total possui 2698 exemplos e 50 atributos incluído o atributo classe.

Resultados e Considerações Finais

Na Tabela 1 é apresentado a acurácia dos 36 classificadores que foram gerados. Destaca-se em azul os Campus Coxim, Nova Andradina e Ponta Porã, que além de obterem os melhores resultados também tiveram a mesma porcentagem de acurácia para cada algoritmo/campus.

No Gráfico 1 é apresentado uma comparação dos resultados das médias de acurácia dos classificadores por campus, entre duas tecnologias: R e Weka (Witten e Frank, 2011; Santos e Santos, 2018). Destaca-se que o R obteve um melhor resultado das acurácias em um panorama geral.

Esses resultados podem ser em consequência de diferenças na implementação dos algoritmos realizadas para o R.

Tabela 1. Acurácia de Classificadores.

Campus / Algoritmo	Naive Bayes	MP	SMO	J48
Aquidauana	50,12%	52,75%	59,12%	64,16%
Campo Grande	57,35%	58,21%	61,37%	65,09%
Corumbá	52,26%	53,63%	58,82%	54,98%
Coxim	74,17%	74,17%	74,17%	74,17%
Naviraí	66,43%	66,43%	66,43%	67,68%
Nova Andradina	83,19%	83,19%	83,19%	83,19%
Ponta Porã	79,48%	79,48%	79,85%	79,48%
Três Lagoas	60,99%	60,99%	60,99%	63,81%
IFMS	64,93%	64,93%	65,65%	64,89%

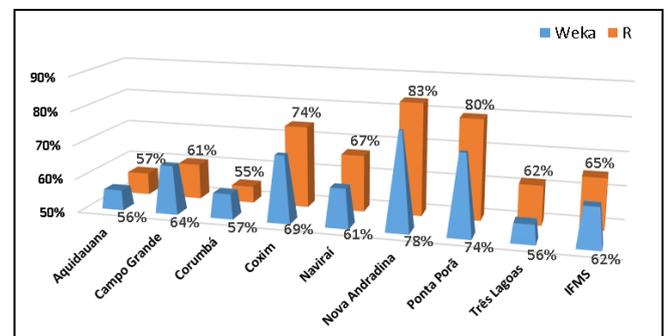


Gráfico 1. Média de Acurácia por Campus.

Portanto, o trabalho revela a importância de comparar as tecnologias para encontrar os melhores resultados. Os resultados permitirão utilizar a tecnologia com melhores resultados para construção de aplicação Web que auxilie o controle de evasão no IFMS.

Agradecimentos

Programa de Iniciação Científica e Tecnológica do IFMS.

Referências

- AMARAL, Fernando. **Introdução à Ciência de Dados: Mineração de dados e BIG DATA**. Rio de Janeiro: Alta Books, 2016. 320 p.
- FREITAS, Bruno Andrade; CANUTO ALVES, Érika Carla; COSTA, Cláudio Pereira. Fatores da evasão discente no curso de Engenharia Civil da Universidade Estadual da Paraíba. **Revista Principia - Divulgação Científica e Tecnológica do IFPB**, [S.l.], n. 34, p. 69-76, jun. 2017. ISSN 2447-9187. Disponível em: <<http://periodicos.ifpb.edu.br/index.php/principia/article/view/1340>>. Acesso em: 11 Maio. 2019. doi:<http://dx.doi.org/10.18265/1517-03062015v1n34p69-76>.
- Santos, Gabrielle Helpis; Santos, Gilson Saturnino. **Aprendizado de Máquina com Dados Socioeconômicos para Apoiar o Controle de Evasão**. Seminário de Iniciação Científica do IFMS, 2018.
- WITTEN, Ian H.; FRANK, Eibe; MARK, A. Hall. 2011. **Data mining: Practical machine learning tools and techniques**, 2011.