

Utilização de Python, Scikit-Learn e Aprendizagem de Máquina para Auxiliar no Controle de Evasão do IFMS

Vitor Oliveira Da Silva¹, Gilson Saturnino Dos Santos¹

¹Instituto Federal De Mato Grosso Do Sul – Coxim MS

vitoro580@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 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 a biblioteca *Scikit-learn* para Python. Os resultados mostram a importância de se realizar tal comparação e os resultados superiores da tecnologia *Scikit-learn*.

Palavras-chave: Aprendizagem de Máquina, *Scikit-learn*, Evasão Escolar.

Metodologia e desenvolvimento

Primeiramente, buscou-se encontrar referências relacionadas a abordagens de processamento de informações referentes aos conceitos de *Machine Learning* (Carvalho, 2012; Mansur, 2008). Por meio da biblioteca *Sci-learn* (Scikit-Learn, 2019) para Linguagem de Programação Python, foram realizadas avaliações de acurácia de diversos classificadores, utilizando bases de dados por Campus do IFMS e geral.

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 Campo Grande, Coxim, Ponta Porã e Três Lagoas com os melhores resultados.

No Gráfico 1 é apresentado uma comparação dos resultados das médias de acurácia dos classificadores por campus, entre duas tecnologias: *Scikit-learn* e Weka (Witten e Frank, 2011; Santos e Santos, 2018). Destaca-se que o *Scikit-learn* obteve melhor resultado das acurácias em um panorama geral, superando o Weka em 7 dos 9 classificadores analisados.

Também foi construído um protótipo de um *Web Service* com auxílio do *framework* Flask em Python (Flask, 2019), que permitirá a utilização das classificações para identificar estudantes que poderão evadir.

Tabela 1. Acurácia de Classificadores.

Campus / Algoritmo	Naive Bayes	MP	SMO	J48
Aquidauana	53,33%	64,44%	58,88%	64,44%
Campo Grande	88,00%	88,00%	88,00%	88,00%
Corumbá	49,04%	61,26%	60,15%	73,49%
Coxim	61,00%	47,22%	44,44%	61,11%
Naviraí	44,40%	66,60%	44,40%	66,60%
Nova Andradina	44,40%	33,00%	44,40%	61,00%
Ponta Porã	78,00%	73,49%	78,25%	73,49%
Três Lagoas	86,60%	75,50%	75,50%	70,00%
IFMS	40,00%	61,00%	76,36%	61,00%

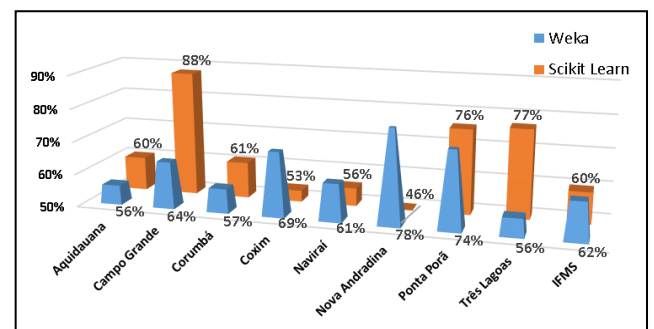


Gráfico 1. Média de Acurácia por Campus.

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

Agradeço à Deus, segundo a todos os envolvidos que prezam pelo desenvolvimento tecnológico, principalmente, em relação a esse trabalho.

Referências

- CARVALHO, D. R. *Mineração de Dados aplicada à fisioterapia*. julho 2012.
- FLASK. *Web Development*. Disponível em: <<http://flask.pocoo.org/>>. Acesso em: 25 abr. 2019.
- MANSUR, M. J. D. B. A. U. *Técnicas de Aprendizado de Máquina Aplicadas na Previsão de Evasão Acadêmica*. Rio de Janeiro, Agosto 2008.
- 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.
- SCIKIT-LEARN. *Machine Learning in Python*. Disponível em: <<https://scikit-learn.org>>. Acesso em: 04 mar. 2019.