



APLICABILIDADE E POTENCIALIDADES NO USO DO *ORANGE DATA MINING* NA DIFUSÃO DA APRENDIZAGEM DE MÁQUINA

Ricardo Nunes de Souza¹
Luiz Gustavo Pereira da Silva²
Gustavo Henrique R. Magalhães³
Danieli A. Duarte⁴
Rafael Fernandes Garcia⁵
Márcio Alves de Araújo⁶
Rodney Costa Machado⁷

RESUMO

Introdução: O aprendizado de máquina, tradução literal do inglês *machine learning*, é um campo da inteligência artificial dedicado à construção de modelos analíticos, com base na identificação de padrões. Apesar de não ser considerado algo novo, está em ampla expansão e isto ocorre devido as suas inúmeras aplicabilidades, que vão desde a identificação de possíveis fraudes bancárias até a criação de desafios mais complexos em jogos de vídeo game. Apesar dos avanços notórios, a difusão e a popularização do uso de técnicas de *machine learning* ainda enfrentam desafios, especialmente relacionados à exigência dos operadores possuírem conhecimentos em programação. **Objetivo:** Busca-se avaliar o efeito de um *software* livre e interativo, com uma interface *drag and drop*, na difusão da aprendizagem de máquina entre pessoas que não sabem codificar. Por se tratar de um *software* com código aberto, procura-se, também, avaliar as potencialidades para programadores desenvolverem pacotes de ferramentas que melhorem a experiência dos demais usuários. **Metodologia:** Utilizando diferentes conjuntos de dados quantitativos, o pacote básico do *software Orange Data Mining* (versão 3.36.2), que inclui os módulos “Data”, “Transform”, “Visualize”, “Model”, “Evaluate” e “Unsupervised”, foi utilizado para realizar previsões (a partir de técnicas de aprendizagem supervisionada), agrupamentos (empregando técnicas de aprendizagem não supervisionada) e

¹ Mestre em Engenharia Civil – UFRJ – E-mail: ricardonunes@unifucamp.edu.br

² Mestre em Educação Profissional Tecnológica – IFTM – E-mail: luizpereirasilva@unifucamp.edu.br

³ Mestrando em Ciências da Computação – UFU – E-mail: gustavoribeiro@unifucamp.edu.br

⁴ Mestre em Educação Profissional Tecnológica – IFTM – E-mail: danieliduarte@unifucamp.edu.br

⁵ Especialista em Segurança da Informação – PITÁGORAS – E-mail: rafaelgarcia@unifucamp.edu.br

⁶ Especialista em Gestão de Riscos e Cibersegurança – FOCUS – E-mail: marcioalves@unifucamp.edu.br

⁷ Mestrando em Engenharia de Software – UNIPAMPA – E-mail: rodnneymachado@unifucamp.edu.br

visualização de dados. **Resultados:** O *Orange* é um programa visual que funciona a partir da conexão entre diferentes ícones, criando um fluxograma de análise. Apesar da grande biblioteca de ícones e da disponibilidade de ícones adicionais, disponíveis por meio de extensões (*e.g. Image Analytics, Network e Geo*), pode-se mencionar pontos de melhoria, principalmente nas etapas de pré-processamento, transformação de dados e visualizações um pouco mais complexas. São etapas em que ainda é necessário o uso de *scripts* em *Python*, e que, o operador que não é apto a fazê-lo, normalmente recorre a outros programas. **Conclusão:** O *software* proporciona uma experiência de uso através de uma interface de usuário extremamente amigável e intuitiva, dispensando a necessidade de qualquer conhecimento prévio em programação. Esta característica é essencial para ampliar o alcance do aprendizado de máquina, atingindo diversos públicos, independentemente da faixa etária ou setor de atuação. Com um módulo para programação em *Python* e a possibilidade de criar extensões, o programa também oferece possibilidades para usuários avançados. Uma vez que, em pouco tempo de uso, foi perceptível a falta de alguns recursos que, gradualmente, podem ser incorporados ao *software*.

Palavras-chave: Popularização da aprendizagem de máquinas. *Softwares* interativos e de código aberto. Desenvolvimento de módulos