

EXTRAÇÃO DE CONHECIMENTO DE UMA BASE DE DADOS REAL SOBRE MINERAÇÃO

Divisão Temática

DT 4 - Processos produtivos, tecnologias e tendências para o presente e o futuro

Autores: L. BENITEZ¹; R. RODRIGUES².

**Instituto Federal de Educação, Ciência e Tecnologia de Santa Catarina
(IFSC)**

Resumo:

A tarefa de analisar dados monitorados em uma operação industrial e extrair informações úteis a partir desses dados é muitas vezes realizada manualmente por engenheiros da indústria. Entretanto, com a aumento no volume de dados coletados por indústrias modernas, faz-se necessário automatizar esse processo, permitindo acelerar a tomada de decisões, melhorar a qualidade de produção e diminuir os custos de operação. O objetivo deste artigo é apresentar um caso de aplicação de algoritmos de *Machine Learning* para extrair conhecimento útil de uma base de dados real da indústria de mineração de ferro, focando na interpretabilidade do modelo gerado e na qualidade das informações extraídas. Para isso, utilizou-se uma base de dados pública e processou-se os dados utilizando algoritmos de Árvore de Regressão e *Random Forest*, implementados na linguagem Python. Obteve-se como resultado um Erro Médio Absoluto de 0,092 (equivalente à 3,95% do valor médio da variável), e a árvore gerada pelo algoritmo foi visualizada e interpretada graficamente. Percebeu-se que algumas variáveis possuem pouquíssima contribuição para a concentração final de sílica, o que pode indicar mal funcionamento da máquina ou um algum outro erro no funcionamento do processo.

Palavras-chave: *machine learning*; *big data*; mineração.

¹ Aluno [Engenharia Eletrônica], leonardo.sb@aluno.ifsc.edu.br.

² Professor [Departamento de Eletrotécnica], Rafael@ifsc.edu.br.

Introdução

O objetivo deste artigo é apresentar um caso de aplicação de algoritmos de *Machine Learning* para extrair conhecimento de uma base de dados da indústria de mineração, focando na interpretabilidade do modelo gerado e na qualidade das informações extraídas. Este trabalho utiliza algoritmos baseados em árvore, por gerarem um modelo facilmente interpretável e compreensível por pessoas sem formação na área técnica (FACELI et al., 2010).

Procedimentos metodológicos

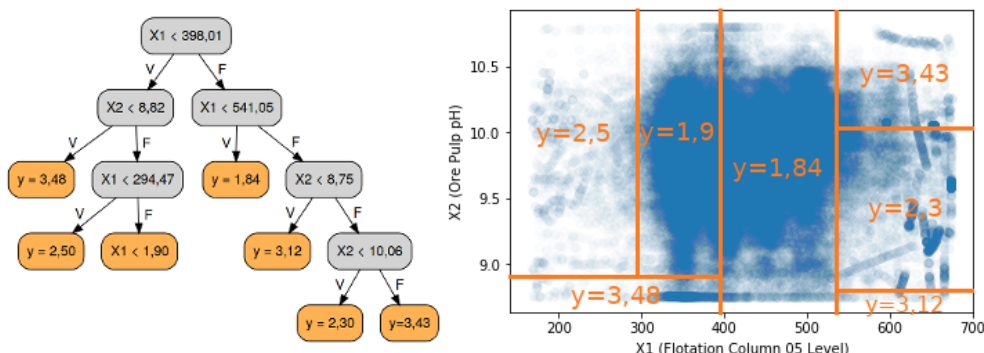
Este trabalho baseia-se numa base de dados pública (OLIVEIRA, 2019), referente à operação de um processo de Flotagem em uma mineradora de ferro. O processo é descrito por 23 variáveis e contém 737 mil exemplos, com o objetivo de prever o valor de uma variável de saída. Trata-se de um problema supervisionado de regressão. O programa foi implementado utilizando a linguagem Python, a biblioteca Scykit Learn e o software jupyter notebook.

Resultados e discussões

As árvores de regressão segmentam sucessivamente o espaço de entrada em regiões mais simples. Para melhorar a interpretabilidade da árvore, deve-se usar uma estratégia de poda (FACELI et al., 2010) e, visando ter o maior controle possível sobre a interpretabilidade da árvore, utiliza-se a estratégia de pré-poda, estabelecendo um limite h para a sua altura. Após vários testes, baseados nas taxas de erro e na facilidade de interpretação, chegou-se ao valor ideal $h=15$.

Uma versão simplificada da árvore gerada pelo algoritmo e a sua interpretação gráfica pode ser vista na Figura 1, com apenas os primeiros níveis das duas principais variáveis de entrada (pH do mineral e nível da coluna de flotagem número 5). A variável de saída é a concentração percentual de sílica, uma impureza no processo que deseja-se prever.

Figura 1 – Árvore gerada pelo algoritmo



Fonte: Autoria própria (2019).

O valor médio da variável de saída é 2,326. Os resultados demonstram que o algoritmo consegue prever o valor desta variável com Erro Absoluto Médio (MAE) igual a 0,614, o que representa prever a variável de saída com 26,38% de erro. Este erro pode ser diminuído deixando de realizar a pré-poda e utilizando uma versão melhorada do algoritmo (*Random Forest*). Neste caso, o MAE é igual a 0,092, significativamente menor que no primeiro caso.

Percebeu-se que algumas variáveis possuem pouquíssima contribuição para a concentração final de sílica, o que pode indicar mal funcionamento da máquina ou um algum outro erro no funcionamento do processo.

Considerações finais

Este artigo é resultado de um projeto de pesquisa financiado pelo CNPq, o qual faz parte de um Acordo Internacional de Cooperação Técnica entre IFSC e Usach (Chile) para otimização de processos de mineração.

Referências

FACELI, Katti et al. **Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina**. Rio de Janeiro: Ltc, 2010.

OLIVEIRA, Eduardo Magalhães. **Quality Prediction in a Mining Process**.

Disponível em: < <https://bit.ly/2XbuEsx> >; Acesso em: 27 mai. 2019.