

APLICAÇÃO ANALYTICS COM ESTUDO DE CORRELAÇÃO: FATORES QUE INFLUENCIAM NA QUALIDADE DO VINHO

Everton Bublitz Do Prado¹
Igor Andrey Strejevitch²
Leonardo Martinelli Deves³
Leonardo Tasca⁴
Alesio Nicolas Barreiro⁵
Arthur Dos Passos Rosa⁶
Daniel Simone Beira⁷
Matheus Da Silva Cardoso⁸
Anelise Stahnke⁹
Sergio Murilo Schütz¹⁰

RESUMO

Em 8 de maio de 2019 os alunos de todos os semestres de 2019 do curso de Análise e Desenvolvimento de Sistemas participaram do III Aula Aberta. No evento estes foram organizados em equipes com no máximo 10 integrantes, para cada equipe foi entregue um problema de ordem quantitativa. Os professores indicaram um Dataset (conjunto de dados disponibilizados para pesquisa) para cada grupo com dois objetivos a serem alcançados. O primeiro, desenvolver uma rotina automatizada para captura, tratamento e carga dos dados de um site da web para um Banco de Dados Não Relacional; o segundo, extrair os dados, armazená-los em uma planilha e realizar um estudo estatístico de correlação para identificar quais variáveis apresentam relação direta com a variável alvo determinada no problema apresentado. Para desenvolver a rotina automatizada utilizamos a linguagem de programação PYTHON, com as bibliotecas SELENIUM e PANDAS, para armazenar os dados foi utilizado o Banco de Dados MONGODB. O problema apresentado para nosso grupo tem origem nos dados extraídos no [site https://www.kaggle.com/uciml/red-wine-quality-cortez-et-al-2009](https://www.kaggle.com/uciml/red-wine-quality-cortez-et-al-2009) sob o título de *Red Wine Quality* tendo como alvo a identificação das variáveis que tem correlação com a qualidade do vinho. Foi definido um limite arbitrário para sua variável dependente (qualidade do vinho) em 7 ou superior sendo classificado como 'bom / 1' e o restante como 'não bom / 0'. Na rotina desenvolvida foi possível identificar o potencial para a automatização dos processos de captura, extração e análise de dados como forma de identificar soluções para problemas que tenham origem dados disponíveis na internet. A partir da implementação desta rotina e do conhecimento das tecnologias aplicadas os integrantes da equipe identificaram diversas outras aplicações. Após execução da rotina que capturou e armazenou os dados foram identificadas e extraídas as variáveis independentes *fixed acidity*, *volatile acidity*, *citric acid*, *residual sugar*, *chlorides*, *free sulfur dioxide*, *total sulfur dioxide*, *density*, *pH*, *sulfates* e *alcohol* do dataset original e a variável *quality* como dependente. Nos testes realizados nenhuma das variáveis apresentaram correlação significativa, apenas a variável *residual sugar*=0.5709 apresentou certo grau de correlação.

Palavras-chave: Correlação. Analytics. Qualidade do Vinho. Python. Pandas. Seleniun. MongoDB.

¹ Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: everbublitz@gmail.com.

² Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: igorandrey@hotmail.com.

³ Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: leo_lli@hotmail.com.

⁴ Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: ltasca93@gmail.com

⁵ Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: alesionb93@gmail.com

⁶ Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: arthurxrosa@hotmail.com

⁷ Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: dantebeira23@gmail.com

⁸ Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: matheuscardoso18@gmail.com

⁹ Graduanda em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: anelise.gemula2@gmail.com.

¹⁰ Professor Mestre em Ciência da Informação. Faculdade Cesusc. E-mail: sergio.schutz@cesusc.edu.br.