

APLICAÇÃO ANALYTICS COM ESTUDO DE CORRELAÇÃO: FATORES QUE INFLUENCIAM NO DESEMPENHO DE CONSUMO DE COMBUSTÍVEL

Felipe Cardoso ¹
Fernando Poletto ²
Jeferson Roberto ³
Andres Jaramillo ⁴
André Boa Sorte ⁵
Elcimar Silva ⁶
Paulo André Lima ⁷
Estefania Menendez ⁸
Bruno Loschenkohl ⁹
Roberto Fabiano Fernandes ¹⁰

RESUMO

Em maio de 2019 os alunos de todos os semestres do curso de Análise e Desenvolvimento de Sistemas participaram da 3ª Aula Aberta da edição de 2019/1. No evento estes foram organizados em equipes com no máximo 10 integrantes, onde para cada equipe foi entregue um problema de ordem quantitativa. Os professores indicaram um Dataset (conjunto de dados disponibilizados para pesquisa na web) 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 o 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, ZipFile 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/anderas/car-consume> sob o título de *Car Fuel Consumption* tendo como alvo a identificação das variáveis que influenciam no consumo de combustível dependendo do seu tipo. O dataset possui dois tipos de combustível e os dados foram extraídos em duas estações distintas no ano, primavera e inverno. 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 distância, velocidade, temperatura interna, temperatura externa, tipo de combustível, ar condicionado, chuva e sol do dataset original e a variável “consumo” como dependente. Dentro do processo de testes realizados com as variáveis para a busca da correlação significativa o estudo em questão ainda está em fase de desenvolvimento.

Palavras-chave: Correlação. Analytics. Estudantes. Consumo de Combustível. Python.

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

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

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

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

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

⁶ Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: elcimar8@gmail.com.

⁷ Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: paulo92lima@gmail.com.

⁸ Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: staphanie03@gmail.com.

⁹ Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: bruno.lck@gmail.com.

¹⁰ Professor Doutor em Engenharia do Conhecimento. Faculdade Cesusc. E-mail: [robertofabiano.fernandes@gmail.com].