

APLICAÇÃO ANALYTICS COM ESTUDO DE CORRELAÇÃO: FATORES QUE INFLUENCIAM O CONSUMO DE PRODUTOS ORGÂNICOS EM UM SUPERMERCADO

Christian Berkenbrock¹
Mitzrael Tassinari²
Allan Mainardes³
Guilherme Lisboa⁴
Afonso Rossoni⁵
Pedro Gabriel Godoy⁶
Gisele Zommer Rossi⁷

RESUMO

Em 8 de maio de 2019 os alunos de todos os semestres 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, e 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 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 o framework scrapy que importa diversas bibliotecas nativas, para armazenar os dados foi utilizado o Banco de Dados MongoDB. O problema apresentado para o grupo tem origem nos dados extraídos no site <https://www.kaggle.com/papercool/organics-purchase-indicator> sob o título de *Organics Purchase Indicator* tendo como alvo a identificação das variáveis que tem correlação com de clientes a consumir produtos orgânicos. O conjunto de dados contém 13 variáveis e mais de 22.000 observações. 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 A = Affluence Grade, B = Loyalty Status do dataset original e a variável Organic Purchase Count como dependente. Nos testes realizados as variáveis que apresentaram correlação significativa foram A = 0,5700 e B = 0,03.

Palavras-chave: Correlação. Analytics. Estudantes. Produtos Orgânicos. Python. Scrapy. MongoDB.

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

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

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

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

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

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

⁷ Professora Mestre em Ciência da Computação. Faculdade Cesusc. E-mail: giselezrossi@gmail.com.