

APLICAÇÃO ANALYTICS COM ESTUDO DE CORRELAÇÃO: FATORES QUE INFLUENCIAM O CONSUMO DE ÁLCOOL DOS ESTUDANTES

Pedro Henrique Medeiros Tomasi¹
Thyago de Souza Pereira²
Lean Enzo Constantino Lima³
Rafael Alexandre Izidoro Hessim⁴
Lucas Góes⁵
Isadora Fagundes Gonçalves⁶
Mariah Rodrigues⁷
Fabio Luiz Piggosso⁸
Aline Jessica Rodrigues da Silva⁹
André de Oliveira Leite¹⁰

RESUMO

Em maio deste ano os alunos de todos os semestres do curso de Análise e Desenvolvimento de Sistemas participaram da 3ª Aula Aberta da edição 2019. No evento este 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 abertos disponibilizados para pesquisa via 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 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 csv, json e pandas e 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/uciml/student-alcohol-consumption> sob o título de *Student Alcohol Consumption* tendo como alvo a identificação das variáveis que tem correlação com a performance dos alunos e o consumo de álcool. O dataset foi obtido em uma pesquisa com estudantes de cursos de matemática e português no ensino médio. 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 sexo, idade, status de coabitação dos pais, guardião, relacionamentos, notas, qualidade das relações familiares e tempo livre do dataset original e as variáveis consumo de álcool diário e consumo de álcool no final de semana como dependentes. No processo de testes, os objetivos dos valores das variáveis ainda estão sendo levantados para futuras correlações significativas.

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

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

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

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

⁴ Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: [faculdade@rafaelizidoro.com.br].

⁵ Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: [lucaasgoes@live.com].

⁶ Graduanda em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: [isaa.fg@hotmail.com].

⁷ Graduanda em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: [rodriguesmariah5@gmail.com].

⁸ Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: [fabiojog@hotmail.com].

⁹ Graduanda em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: [rodriguealine@gmail.com].

¹⁰ Professor Especialista em Gestão da Segurança da Informação. Faculdade Cesusc. E-mail: [andreredes@gmail.com].