

Projeto SEGURANÇA CIBERNÉTICA E ENGENHARIA SOCIAL	Consultoria SISTEMAS DISTRIBUÍDOS E APRENDIZAGEM DE MÁQUINA.	Turma 5o Semestre, 2024
Tema Sistema de Arquivo Distribuído – HDFS e PySpark	Título Utilizando PySpark para contar palavras	
Data de Abertura 06.MAI.2023	Data de Fechamento 17.JUN.2024	Complexidade ALTA

PROJETO INTEGRADOR

Projeto 2 (Bimestre Maio e Junho)

O PROJETO

O projeto do 2º bimestre será composto por dois subprojetos a saber:

- 1- Processar conteúdo de sites e redes sociais para contar palavras e identificar a ocorrência de termos específicos
- 2- Apresentar um Dashboard que demonstre os termos que mais ocorrem e de forma positiva e relevante.

2º Projeto – Contando palavras

CONTEXTO

Sua equipe pertence uma empresa que deseja fazer uma ação de marketing em sites e redes sociais, para isso ela deseja identificar os termos que devem ser associados ao seu(s) produto(s)/serviço(s) (nos moldes do AdSense da Google) de forma a maximizar a possibilidade de o anúncio aparecer para seu público-alvo de forma positiva.

A estratégia a ser utilizada é a do Topo de Funil de Vendas:

<https://www.agenciaeplus.com.br/wp-content/uploads/2021/06/funil-de-vendas.jpg>

“Topo de funil é a primeira etapa do funil de vendas, considerada a fase de atração, quando o visitante já sabe que tem um problema (ou uma necessidade) e busca

meios para resolvê-lo. Nessa primeira fase, é preciso despertar o interesse das pessoas para guiá-las pela jornada de compra até que se tornem clientes.” (Nuvemshop)



Considerando que o Topo de Funil é o estágio em que é necessário atrair e conscientizar consumidores, deve-se então criar e distribuir materiais que tenham o poder de concretizar isso na web. Esses materiais podem ser, por exemplo, sites especializados, posts para blogs e redes sociais, e-books, vídeos, lives, webinars, etc., sobre o seu setor de negócios.

Nesse projeto sua equipe **não** vai desenvolver e-books, vídeos, muito menos criar sites, blogs ou redes sociais para seu mercado. Na verdade, sua equipe deverá identificar quais os melhores canais para anunciar seu(s) produto(s).

Esse projeto, que deverá ser feito em equipes, consiste em :

- 1- Identificar uma área de negócio (Ex. Livros, Perfumes, Roupas Esportivas, Produtos Alimentícios, Bebidas)
- 2- Dentro da área de negócio identificar um Produto/Serviço específico ou uma Linha. (Ex. Livros:Um título ou Autor, Roupas Esportivas: Camisas de Time, Bebidas: Energéticos, Sucos, Vinhos)
- 3- Identificar pelo menos 20 sites (páginas dinâmicas)¹ e pelo menos 1(uma) rede social que tenham conteúdo relevantes para divulgação do produto para seu público-alvo.
- 4- Preparar o ambiente para processar os dados dos sites e redes sociais selecionados (PySpark e Jupyter)

¹ Páginas Dinâmicas são páginas de site em que seu conteúdo é gerado dinamicamente a partir de uma fonte de dados. Pode-se citar com exemplo Site de Notícias, Blogs, Rede Sociais em formato WEB, entre outros.

- 5- Processar esses dados (web scraping) contando palavras e apresentar as palavras de relevância com maior frequência. Explicar por que as **palavras** são relevantes para o produto/serviço e público-alvo.
- 6- **(DESAFIO)** Verificar a ocorrência e frequência de **termos** específicos (Ex. "Campeão Paulista", "Yuval Harari", "Quentin Tarantino", "Garrafa Térmica") e verificar se eles de fato são relevantes para anúncio do produto para o público-alvo.
- 7- **(DESAFIO)** Identificar palavras/termos com alta frequência, mas em um contexto negativo para o público-alvo. (análise de sentimento)
- 8- Desenvolver um Dashboard de forma apresentar os dados coletados (com tabelas e gráficos) de forma a indicar quais palavras/termos devem ser associados ao produto/serviço a ser divulgado.

ENTREGÁVEIS

- 1- Uma apresentação (Power Point) descrevendo com a seguinte Agenda (mínimo)
 - Agenda
 - Identificação do grupo
 - Apresentação do negócio, produto e/ou serviço
 - Identificação do público-alvo
 - Palavras Chaves procuradas
 - Sites e Redes Sociais pesquisadas
 - Resultados da pesquisa
 - Frequências de Palavras
 - Frequências de Termos (DESAFIO)
 - Análise de Sentimento
 - Dashboards
 - Análise
 - Conclusão (Tomada Decisão)
 - Sites e Rede Social selecionada
- 2- Notebook (.ipynb)
- 3- Manual de orientação para execução do Notebook

OUTROS CONTEXTOS

A título de exemplo esse tipo de estudo pode também ser utilizado para verificar a audiência QUALITATIVA de um produto ou serviço (streaming). Por exemplo: Qual personagem da série está repercutindo mais nas redes sociais? Como está repercutindo o novo produto/serviço nos canais digitais da crítica especializada? Como está repercutindo a atual campanha de divulgação de um produto/serviço nas mídias digitais? Esses são alguns exemplos que as técnicas necessárias para a realização deste projeto podem ser utilizadas.