Python 4.6

O Python é uma linguagem de programação criada em 1990 com multi-paradigma cujo uso abrange desde pequenos projetos e scripts para uso pessoal até softwares grandes e complexos, como parte dos serviços do Youtube, Instagram, Spotify, entre muitos outros. Devido sua rapidez e simplicidade, além do seu vasto acervo de bibliotecas, API’s e frameworks de código aberto e gratuitos para uso, atualmente essa linguagem é uma das principais na comunidade científica da computação, isto é, ela é comumente usada para algoritmos e aplicações que envolvem aprendizagem de máquina, inteligência artificial, big data, deep learning e outras vertentes do Data Science.

SQL 2016

O SQL, acrônimo de “Structured Query Language”, ou ainda, do português, “Linguagem de Consulta Estruturada” é uma linguagem de programação voltada para a criação, manipulação e exclusão de banco de dados relacionais, modelo extremamente difundido na academia e no mercado.  
Embora que nos últimos anos esse modelo tenha perdido espaço mercadológico para os chamados bancos “NoSQL” (bancos não relacionais) e modelos orientados à objetos, ele ainda se apresenta muito perspicaz na facilidade e dinamismo na criação, manipulação e alocação de tabelas (entidades), atributos e bases de dados.

MySQL 8

O MySQL é um SGBD (Sistema de Gerenciamento de Banco de Dados) de código aberto criado em 1995 e atualmente mantido pela empresa Oracle Corporation. Graças à sua versão gratuita denominada Community e seu foco no público de programadores iniciantes e estudantes na área de tecnologia, segundo a pesquisa de 2019 realizada pelo site StackOverflow - maior fórum de programadores do mundo – o MySQL lidera o ranking de popularidade com mais da metade de todos os usuários que utilizam banco de dados para quaisquer fins, superando concorrentes de peso como o SQL Server da Microsoft e o Firebase, mantido atualmente pelo Google.

<https://insights.stackoverflow.com/survey/2019#community>

Anaconda 3

A plataforma open source Anaconda reúne uma série de softwares, bibliotecas e ferramentas úteis para a programação científica a fim de facilitar a integração de diferentes aplicativos e oferecer recursos avançados que não se encontrariam em programadas convencionais de programação. No caso do algoritmo dessa pesquisa, foi utilizada a IDE Spyder, responsável por reunir uma série de pacotes úteis para o processo de crawling/scraping de dados, assim como facilitar a visualização dos resultados obtidos e reduzir a complexidade dos códigos.

Selenium WebDriver 3.141.39

O Selenium é uma ferramenta de automatização de testes funcionais em navegadores de internet, como o Google Chrome e o Mozilla Firefox. Apesar de ser escrito em Java, ele suporta a integração com, além de sua própria linguagem mãe, o JavaScript, Ruby, C#, PHP, Perl e o Python, que é o que se utilizou para o desenvolvimento do crawler da pesquisa. Em síntese, ele é responsável por navegar por links de forma automática sempre executando as funções e comportamentos pré-programados. Para a pesquisa, o Selenium WebDriver serviu apenas para o crawling, ou seja, tanto para entrar nos sites dos veículos de comunicação definidos, como para capturar o título e link das notícias que atendiam todos os parâmetros e pré-requisitos.

Tweepy 3.8.0

Por questões legais dos termos de uso da rede social do Twitter, o único meio que podemos realizar o scraping de dados dos usuários de sua plataforma, como os tweets dos candidatos já citados durante toda a pesquisa, é utilizando a própria API (Application Program Interface) da rede, chamada de Tweepy. Basicamente, essa ferramenta fornece aos programadores da linguagem Python devidamente licenciados no programa Twitter Developer uma série de funcionalidades para a captura dos dados pretendidos, de forma mais intuitiva e simples do que seria feito – por meios ilegais – com scrapers de terceiros.

Outros pacotes Python – NLTK, RE e PYMYSQL

O pacote NLTK (versão 3.4.5), acrônimo de “Natural Language Toolkit”, é escrito em Python e possui o propósito de realizar processamentos qualitativos e quantitativos de textos submetidos a ele pelo programador. Com essa ferramenta, é possível efetuar-se diversos tratamentos em palavras, desde a retirada de termos que não possuem sentido sozinhos (stopwords) do corpo do texto, até uma análise sintética de emoções presentes no discurso do seu locutor. Essa biblioteca foi de suma importância para a obtenção de resultados advindos dos tweets dos candidatos, facilitando a análise para a posterior conclusão.

O pacote RE (nativo do Python) é destinado para se trabalhar com as Regular Expressions, isto é, ele é o responsável por identificar e separar palavras considerando apenas os valores alfanuméricos do texto. Obviamente que ele é extremamente mais amplo e complexo do que isso, mas para a construção do algoritmo, apenas este recurso fora usado.

Por fim, o PYMYSQL (versão 0.7.2) é uma das bibliotecas responsáveis por conectar a aplicação escrita em Python com o banco de dados do MySQL, permitindo que os dados obtidos tanto pelo crawler quanto pelo scraper sejam salvos e passíveis à visualização mesmo após o encerramento da aplicação.