

FACULDADE DE TECNOLOGIA DE SÃO PAULO

Amanda de Jesus Garcia

Classificação de Sentimento sobre Agentes do Valorant

São Paulo - SP

FACULDADE DE TECNOLOGIA DE SÃO PAULO

Amanda de Jesus Garcia

Classificação de Sentimento sobre Agentes do Valorant

Trabalho submetido como exigência parcial
para a obtenção do Grau de Tecnólogo em Análise e
Desenvolvimento de Sistemas Orientador: Paulo

Roberto Bernice

São Paulo – SP

RESUMO

Este estudo se propõe a investigar a aplicação de técnicas de Processamento de Linguagem Natural (NLP) na análise das percepções dos jogadores acerca dos agentes do renomado jogo Valorant, utilizando como fonte de dados a plataforma social Twitter, que atualmente se chama X mas possui a mesma API que era utilizada no Twitter anteriormente. O principal objetivo é explorar como as opiniões dos jogadores são articuladas em textos curtos e informais nesse ambiente digital e, mais crucialmente, como essas percepções podem ser desvendadas e compreendidas por meio de abordagens de NLP.

Para alcançar esse intuito, a pesquisa seguiu uma metodologia abrangente que abarcou desde a coleta meticulosa de tweets pertinentes até o rigoroso pré-processamento dos dados obtidos. Em seguida, foi conduzida uma análise de sentimentos detalhada, permitindo não apenas capturar as emoções expressas pelos jogadores, mas também discernir a polaridade desses sentimentos em relação a cada agente do jogo Valorant.

Além disso, uma etapa crucial do estudo consistiu na identificação e classificação de tópicos recorrentes nos tweets, visando extrair insights valiosos sobre a recepção dos jogadores em relação aos diferentes agentes do jogo. Essa análise de tópicos proporcionou uma compreensão mais profunda das preferências.

Portanto, ao adotar uma abordagem multidisciplinar que integra conceitos de NLP com a análise de sentimentos e identificação de tópicos, este estudo oferece uma contribuição substancial para o entendimento das percepções dos jogadores sobre os agentes do jogo Valorant, ao mesmo tempo em que destaca o potencial das técnicas de NLP para análise de dados textuais em contextos sociais e informais como o Twitter.

Palavras-chave: Análise de Sentimento, Mídias Sociais, Processamento de Linguagem Natural, Opinião Pública, Big Data, Python.

ABSTRACT

This study aims to investigate the application of Natural Language Processing (NLP) techniques in the analysis of players' perceptions about the agents of the renowned game Valorant, using as a data source the social platform Twitter, which is currently called X and use same API that was used on Twitter previously. The main objective is to explore how players' opinions are articulated in short, informal texts in this digital environment and, more crucially, how these perceptions can be uncovered and understood through NLP approaches.

To achieve this aim, the research followed a comprehensive methodology that ranged from the meticulous collection of relevant tweets to the rigorous pre-processing of the data obtained. Next, a detailed sentiment analysis was conducted, allowing us to not only capture the emotions expressed by players, but also discern the polarity of these feelings towards each agent in the Valorant game.

Furthermore, a crucial stage of the study consisted of identifying and classifying recurring topics in tweets, aiming to extract valuable insights into players' reception towards different game agents. This topic analysis provided a deeper understanding of preferences.

Therefore, by adopting a multidisciplinary approach that integrates NLP concepts with sentiment analysis and topic identification, this study offers a substantial contribution to understanding players' perceptions of Valorant game agents, while also highlighting the potential of NLP techniques for analyzing textual data in social and informal contexts such as Twitter.

Keywords: Sentiment Analysis, Social Media, Natural Language Processing, Public Opinion, Big Data, Python.

SUMÁRIO

1.0 Introdução	2
2.0 Fundamentos Teóricos.....	3
2.1 Redes Sociais.....	3
2.1.1 O Twitter (X).....	5
2.2 Big Data	8
2.2.1 Características do Big Data.....	8
2.2.2 Etapas Big Data.....	11
2.3 Python para dados	13
2.3.1 PySpark.....	14
2.3.2 Numpy e Pandas	16
2.3.3 Matplotlib	18
2.3.4 Scikit Learn.....	19
2.3.5 Tweepy.....	20
2.3.6 NLTK.....	21
2.3.7 Jupyter	22
2.3.8 Seaborn.....	22
2.3.9 Word Cloud.....	23
2.4 Análise de Sentimentos.....	24

2.4.1 Definição de Opinião.....	24
2.4.2 Dicionários Léxicos	26
3.0 Metodologia	26
3.1 Ferramentas Utilizadas.....	27
3.2 Coleta de dados	27
3.3 Pré-processamento.....	28
3.4 Análise de sentiment	30
4.0 Resultados	30
5.0 Conclusão	34
REFERÊNCIAS.....	35

1.0 Introdução

Nos últimos anos, testemunhamos uma revolução sem precedentes no mundo dos jogos eletrônicos, impulsionada pelo avanço contínuo da tecnologia e pela crescente interação entre jogadores em ambientes digitais [1]. No cerne dessa transformação, encontra-se o jogo Valorant, uma criação magistral da Riot Games que segundo o site Esports.net (2024), Valorant manteve sua popularidade com uma estimativa de milhões de jogadores ativos ao longo do ano, estabelecendo-se como um dos principais expoentes do gênero de tiro tático [2].

Valorant transcende o mero entretenimento para se tornar um fenômeno cultural, um mundo virtual onde os jogadores mergulham em batalhas intensas e estratégicas, cada uma moldada pelos agentes únicos que escolhem. Com um elenco diversificado de personagens, cada um com habilidades distintas e personalidades cativantes, Valorant oferece uma experiência de jogo rica em profundidade e complexidade, alimentando discussões apaixonadas e debates fervorosos dentro da comunidade de jogadores.

No entanto, por trás da fachada de diversão e competição, reside um vasto reservatório de dados, uma corrente incessante de conversas, opiniões e insights gerados diariamente pelos jogadores em plataformas sociais como o Twitter. É nesse oceano de informações que encontramos o foco deste estudo: a análise das percepções dos jogadores sobre os agentes do Valorant, através da lente poderosa do Processamento de Linguagem Natural (NLP).

O Twitter surge como o cenário ideal para esta investigação, uma arena virtual onde a comunidade de jogadores se reúne para compartilhar suas experiências, estratégias e, acima de tudo, suas opiniões sobre os agentes do Valorant [3]. Com uma quantidade inimaginável de tweets sendo gerados a cada segundo, o Twitter se revela como um tesouro de dados, uma mina de informações que aguarda ser explorada e compreendida.

Neste estudo, propomos mergulhar nas profundezas desse oceano de dados, utilizando técnicas avançadas de NLP para desvendar os padrões e tendências ocultas nas opiniões dos jogadores. Nosso objetivo é entender a opinião dos jogadores, o mais amado e o menos amado quando se trata dos agentes do Valorant.

Ao integrar o NLP com a análise de dados sociais, buscamos oferecer uma visão holística e abrangente das percepções dos jogadores sobre os agentes do Valorant, contribuindo assim para o avanço do conhecimento e o aprimoramento contínuo da experiência de jogo.

2.0 Fundamentos Teóricos

Neste capítulo, será apresentada a fundamentação teórica essencial para o desenvolvimento e compreensão deste estudo. Primeiramente, exploraremos os conceitos e as características das redes sociais, destacando sua importância e impacto na comunicação e na coleta de dados. Em seguida, discutiremos o conceito de Big Data, abordando suas definições, características e a relevância na análise de grandes volumes de dados gerados pelas interações nas redes sociais.

Também examinaremos o papel da linguagem de programação Python no contexto de análise de dados, detalhando suas principais bibliotecas e ferramentas que facilitam a manipulação e interpretação de grandes conjuntos de dados. Entre essas bibliotecas, destacaremos pandas, NumPy, Matplotlib, e bibliotecas específicas para processamento de linguagem natural (NLP), como NLTK.

Por fim, abordaremos as definições e métodos de Análise de Sentimentos, explicando como essa técnica permite extrair e quantificar emoções e opiniões expressas em textos. Discutiremos os principais algoritmos e abordagens utilizados na Análise de Sentimentos, além de suas aplicações práticas e desafios.

2.1 Redes Sociais

Uma rede social pode ser entendida como um ambiente no qual se estabelece um conjunto de interações e conexões entre indivíduos, grupos ou organizações, que compartilham interesses comuns. Essas interações geralmente ocorrem por meio de plataformas acessíveis na Internet, facilitando a comunicação, troca de informações e a criação de comunidades virtuais.

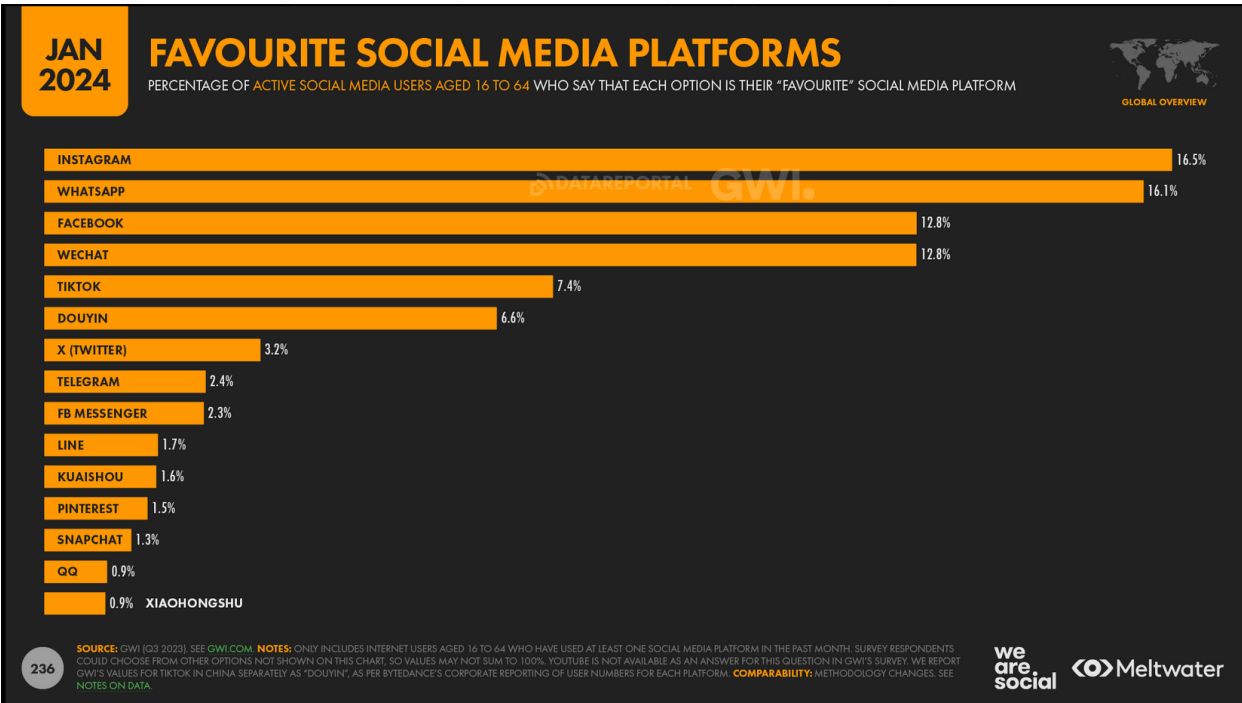
Os Social Networking Sites (SNS), ou sites de redes sociais, são projetados para fornecer aos usuários um espaço próprio onde podem compartilhar informações pessoais, como fotos, interesses, atualizações de status e muito mais. Esses sites permitem que os usuários se conectem e interajam com outros membros da mesma rede, promovendo a formação e manutenção de relacionamentos. [5]

Ao contrário do que acontece em alguns outros tipos de plataformas de encontros ou redes

profissionais, o objetivo principal de um SNS não é facilitar o encontro com estranhos, mas sim reforçar e manter as conexões já existentes. Isso inclui amigos, familiares, colegas de trabalho e conhecidos, criando um ambiente digital onde as relações pessoais e sociais podem ser cultivadas e desenvolvidas. Esses sites frequentemente oferecem diversas funcionalidades, como mensagens privadas, comentários, curtidas e compartilhamentos, que incentivam a interação contínua entre os usuários.[5]

Existem dezenas de redes sociais muito conhecidas no mundo. A Figura 1, mostra as redes sociais mais usadas no mundo em mês de janeiro de 2024:

Figura 1. Redes sociais mais usadas no mundo em janeiro de 2023



Fonte: [4]

Ao longo dos anos, as redes sociais passaram por uma significativa evolução, incorporando novas funcionalidades e métodos inovadores de interação entre os usuários. Inicialmente focadas em conexões básicas, essas plataformas agora permitem uma vasta gama de atividades, incluindo mensagens instantâneas, que possibilitam a comunicação em tempo real. Os usuários podem compartilhar uma variedade impressionante de conteúdos, como textos, fotos, vídeos e até mesmo sua localização geográfica.

Essa capacidade ampliada de compartilhar informações tem transformado as redes sociais em

fontes valiosas de dados. Informações que antes poderiam ser consideradas irrelevantes estão ganhando crescente importância e se tornando foco de estudos detalhados. Por exemplo, os dados das redes sociais são agora amplamente utilizados em pesquisas de opinião sobre produtos e eventos, permitindo uma análise mais precisa e abrangente das percepções públicas.[7]

Além disso, essas plataformas têm facilitado a identificação de comunidades com interesses específicos, permitindo uma segmentação mais eficaz e a criação de estratégias de marketing direcionadas. A opinião pública, que antes era capturada através de métodos tradicionais como chamadas telefônicas ou entrevistas presenciais, agora pode ser monitorada e analisada em tempo real através das interações e conteúdos compartilhados nas redes sociais. Isso oferece uma visão mais imediata e precisa das tendências e sentimentos predominantes entre os usuários, tornando as redes sociais uma ferramenta indispensável para pesquisadores e empresas que desejam compreender melhor seu público-alvo.[8]

2.1.1 O Twitter (X)

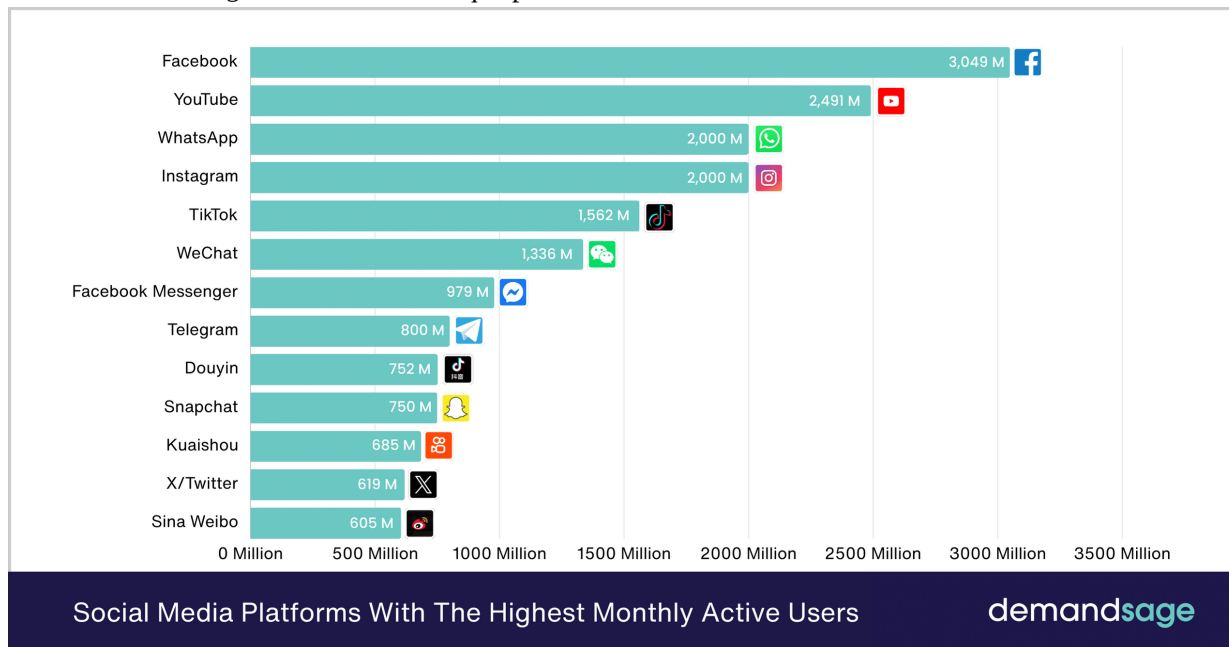
O Twitter, que atualmente se chama X é uma rede social e um serviço de micro blog para comunicação em tempo real usado por milhões de pessoas e organizações. Usuários do Twitter ficam interconectados ao publicar atualizações ao site, conhecidas como “Tweets”, para compartilhar, trocar e descobrir informações. Tweets consistem em 140 ou menos caracteres e podem contar ideias e vários tipos de informação, como fotos, vídeos e links para artigos. Usuários podem acessar essas mensagens no Twitter pela Web ou por um dispositivo móvel com conexão a Internet. Uma vez que o usuário publica um Tweet, a mensagem é publicada em seu perfil e enviada para a página inicial (ou “feeds”) de usuários seguidores que se inscreveram para receber as atualizações daquele Twitter. Usuários podem navegar pelo seu feed para visualizar atualizações de seguidores no Twitter que podem ser desde uma observação pessoal até uma notícia jornalística. Usuários podem seguir ou serem seguidores de amigos, família, colegas de trabalho, organizações, empresas e figuras públicas.[8]

Seguidores comunicam-se com um usuário ao responder um Tweet ou “Retweetar” a publicação para compartilhá-la com seus próprios seguidores. Tweets e respostas são públicas e visíveis a qualquer pessoa, exceto que seja um Tweet protegido. Ao responder e compartilhar tweets, usuários ficam

envolvidos entre si tempo real sobre conteúdos interessantes, assuntos populares e eventos atuais.[8]

A Figura 2, mostra as redes sociais que possuem mais usuários ativos em 2024

Figura 2. redes sociais que possuem mais usuários ativos em 2024



Fonte: [9]

Alguns fatos recentes sobre o Twitter [9]:

- O Twitter tinha 619 milhões de usuários ativos mensais em abril de 2024.
- 60,9% dos usuários do Twitter são homens, enquanto 39,1% são mulheres em 2024.
- A maior parte da audiência do Twitter (38,5%) pertence à faixa etária de 25 a 34 anos.
- Com 105,42 milhões de usuários, os Estados Unidos têm o maior número de usuários do Twitter.
- Um usuário do Twitter nos Estados Unidos passa em média 34 minutos e 6 segundos diariamente na plataforma.
- O Twitter registrou uma receita de US\$ 3,4 bilhões em 2023.
- O Twitter gerou US\$ 1,75 bilhão em receitas nos Estados Unidos, enquanto o resto do mundo foi responsável por US\$ 1,65 bilhão em receitas em 2023

2.1.2 Valorant

Para melhor entendimento do contexto do estudo, este paragrafo aborda o jogo que foi utilizado como objeto de estudo via a rede social Twitter e os principais termos utilizados pela sua comunidade.

O Valorant é um jogo online produzido e distribuído pela Riot Games do gênero FPS (First Person Shooter), o jogo consiste em cinco jogadores em cada time que se alternam entre atacantes e defensores, onde os atacantes devem posicionar uma bomba, chamada de “Spike”, na base inimiga e protegê-la para ganhar um turno sendo que, a partida tem seu fim quando um dos times soma quinze pontos.

Os jogadores antes de cada partida podem escolher seus personagens que comumente são chamados de “Agent” ou Agentes e conseqüentemente realizam distintas funções dentro de uma partida. Essas funções são comumente chamadas de “Role” que pode ser traduzido também como papel que o jogador irá desempenhar.

Como exemplos das funções nos temos:

- **Controlador** que tem como seu objetivo controlar a movimentação dos oponentes. Suas habilidades giram em torno de smokes (bombas de fumaças) e paredes, que são capazes de restringir ou tampar o campo de visão dos inimigos e dificultar eventuais desarmes ou ataques às escuras e até mesmo criar emboscadas. Um personagem dessa função seria o Brimstone.
- **Duelista** são os personagens que são os primeiros a entrarem em uma área e a confrontar os inimigos, um bom exemplo seria a personagem Jett. Suas habilidades consistem em movimentação e dano.
- **Iniciador** é a função de personagens que tem o foco em reconhecimento do mapa, eles buscam informações sobre os oponentes quanto a sua localização e quantidade de inimigos. Suas habilidades têm foco utilitário para o time e normalmente são utilizadas para defender as zonas onde a “Spike” pode ser posicionada. Um ótimo exemplo de personagem que desempenha esta função é o Sova.
- **Sentinelas** são os personagens que desempenham o papel de suporte de suas equipes. Esta função é conhecida pelas suas habilidades de defesa e cura, tendo como sua maior representante a Sage.

Figura 26. Valorant logo



Fonte: [40]

2.2 Big Data

Big Data refere-se ao vasto conjunto de informações armazenadas nos sistemas de servidores e empresas, interligadas entre si e acessíveis através da rede mundial de computadores. Um exemplo clássico é o YouTube, que disponibiliza uma extensa biblioteca de vídeos acessíveis a usuários de todo o mundo.[10]

A análise e otimização de dados sociais em redes sociais em big data desempenham um papel fundamental na compreensão e aproveitamento do potencial das plataformas de redes sociais. Com o crescimento exponencial do número de usuários, cada vez mais aprimoramos as técnicas e estratégias para extrair Insights valiosos dessas plataformas.[11]

Lisa Arthur, colaboradora da Forbes, define Big Data como qualquer dado coletável sobre uma empresa ou assunto, conforme concordam muitos executivos de marketing (CMOs) e de tecnologia da informação (CIOs). Este conceito abrange informações provenientes de diversas fontes, tanto tradicionais quanto digitais, internas e externas à empresa, sendo uma fonte crucial para descobertas e análises contínuas.

Embora a era digital seja dominante na geração de dados, grandes empresas enfatizam a importância de não negligenciar informações provenientes de transações comerciais, registros financeiros e interações não-digitais como call centers e telemarketing. Todas essas fontes de dados são potenciais aliadas no crescimento e desenvolvimento empresarial.

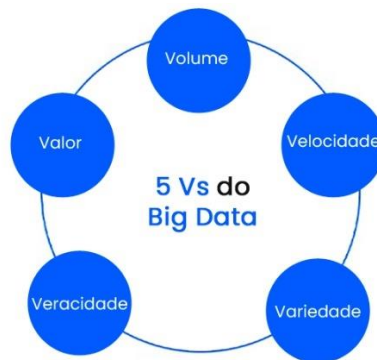
2.2.1 Características do Big Data

Atualmente, podemos dividir o Big Data em 5 V's que formam a base para a implementação do conceito em qualquer empresa.

São eles: Volume, Velocidade, Variedade, Variabilidade e Vínculo.

Vamos explicar em detalhes cada um deles abaixo. Na Figura 3 podemos observar essas características do Big Data.

Figura 3. Os 5 Vs do Big Data.



Fonte: [12]

- **Volume:** Nos últimos anos, a quantidade de dados gerados ultrapassou significativamente tudo o que foi produzido ao longo da história da humanidade, de acordo com o SBG (Social Good Brasil) [13]. Essa produção de dados está dobrando a cada dois anos e estima-se que alcançará 350 zettabytes ainda este ano, equivalente a 35 trilhões de gigabytes. O crescimento exponencial não é mais impulsionado apenas pela atividade humana, pois praticamente todos os dispositivos conectados à internet estão contribuindo para o Big Data. Com o avanço da Inteligência Artificial (IA) e da Internet das Coisas (IoT), espera-se ainda mais expansão na geração de dados.

- **Velocidade:** Velocidade é a taxa mais rápida na qual os dados são recebidos e talvez administrados. Normalmente, a velocidade mais alta dos dados é transmitida diretamente para a memória, em vez de ser gravada no disco. Alguns produtos inteligentes habilitados para internet operam em tempo real ou quase em tempo real e exigem avaliação e ação em tempo real. [14]
- **Variedade:** Refere-se à variedade de tipos e formatos de dados presentes no ambiente de Big Data. Esses dados podem ser classificados como estruturados, semi-estruturados ou não estruturados.
 - Dados estruturados são organizados em tabelas com linhas e colunas, típicos dos bancos de dados tradicionais.
 - Dados semiestruturados, como JSON e XML, possuem algum nível de organização, mas não seguem rigidamente um modelo relacional.
 - Dados não estruturados incluem textos livres, imagens, vídeos e áudios, que não têm uma estrutura fixa.

Lidar com essa diversidade de dados apresenta desafios, mas também oferece uma grande oportunidade para extrair insights valiosos.

- **Veracidade:** A veracidade do Big Data está ligado à confiabilidade e à precisão dos dados. No contexto do Big Data, é crucial assegurar que os dados sejam verídicos, consistentes e confiáveis, pois decisões empresariais significativas podem depender dessas informações. A crescente diversidade de fontes e formatos de dados pode tornar a validação da veracidade dos dados um desafio maior.
- **Valor:** Refere-se à importância e à relevância dos dados para a organização. O principal objetivo do Big Data é extrair valor das enormes quantidades de informações disponíveis. Isso é feito ao identificar insights, padrões e tendências que podem ser utilizados para melhorar os negócios de diversas maneiras. Por exemplo:
 - **Melhorias nos Processos:** Analisar dados para otimizar operações, reduzir ineficiências e aumentar a produtividade.

- **Redução de Custos:** Utilizar dados para identificar áreas onde a empresa pode economizar, como em logística, produção ou marketing.
- **Inovação:** Descobrir novas oportunidades de produtos ou serviços ao analisar as necessidades e comportamentos dos clientes.
- **Aumento de Receita:** Identificar novas fontes de receita e estratégias de mercado, personalizando ofertas para diferentes segmentos de clientes.
- **Vantagem Competitiva:** Tomar decisões mais informadas e ágeis que colocam a empresa à frente da concorrência.

Extrair valor do Big Data significa transformar dados brutos em insights acionáveis que conduzam a decisões estratégicas e operacionais mais eficazes, proporcionando vantagens competitivas no mercado.

2.2.2 Etapas Big Data

A cadeia de valor do Big Data compreende as etapas essenciais para transformar dados em informações valiosas que auxiliam na tomada de decisões. Conforme ilustrado na Figura 4, essa cadeia inclui sete etapas: a coleta de dados, o armazenamento de dados, o processamento de dados, a análise de dados, a visualização dos dados, a implementação e ação e a governança de dados.

Coleta de Dados

- **Fontes de Dados:** Identificação e seleção das fontes de dados, que podem ser internas (como sistemas de CRM e ERP) ou externas (como redes sociais e dados de mercado).
- **Aquisição de Dados:** Coleta dos dados utilizando diversas técnicas, incluindo APIs, web scraping, sensores IoT, logs de servidores, etc.

Armazenamento de Dados

- **Sistemas de Armazenamento:** Escolha da infraestrutura adequada para armazenar grandes volumes de dados, como data lakes, data warehouses ou sistemas de armazenamento em nuvem.
- **Gerenciamento de Dados:** Implementação de políticas e procedimentos para garantir que

os dados sejam armazenados de forma organizada, segura e facilmente acessível.

Processamento de Dados

- **Limpagem e Transformação:** Remoção de dados redundantes ou inconsistentes, e transformação dos dados em um formato adequado para análise.
- **Integração de Dados:** Combinação de dados de diferentes fontes para criar um conjunto de dados coeso.

Análise de Dados

- **Análise Descritiva:** Exploração dos dados para entender o que aconteceu no passado.
- **Análise Diagnóstica:** Investigação das causas por trás dos eventos observados.
- **Análise Preditiva:** Uso de modelos estatísticos e de machine learning para prever eventos futuros.
- **Análise Prescritiva:** Recomendação de ações baseadas nas análises anteriores para otimizar resultados futuros.

Visualização de Dados

- **Criação de Dashboards:** Desenvolvimento de painéis interativos que apresentam as informações de forma clara e acessível.
- **Relatórios e Gráficos:** Geração de relatórios e gráficos que resumem os insights obtidos, facilitando a interpretação dos dados por parte dos tomadores de decisão.

Implementação e Ação

- **Tomada de Decisão:** Utilização dos insights derivados das análises para tomar decisões estratégicas e operacionais.

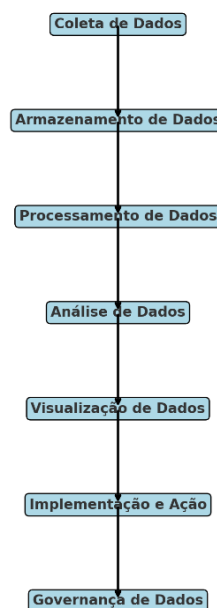
- **Monitoramento e Ajuste:** Monitoramento contínuo dos resultados das ações implementadas e ajustes necessários com base em novos dados e insights.

Governança de Dados

- **Políticas de Dados:** Estabelecimento de políticas para assegurar a qualidade, segurança e privacidade dos dados.
- **Conformidade:** Garantia de que a coleta, armazenamento e uso dos dados estejam em conformidade com regulamentações e leis vigentes

Figura 4. Etapas do Big Data

Fluxograma das Etapas do Big Data



Fonte: Elaborado pela autora

2.3 Python para dados

O Python (Figura 4) é uma linguagem de programação amplamente usada em aplicações da Web, desenvolvimento de software, ciência de dados e machine learning (ML). Os desenvolvedores usam o Python porque é eficiente e fácil de aprender e pode ser executada em muitas plataformas diferentes. O software Python pode ser baixado gratuitamente, integra-se bem a todos os tipos de sistema e agiliza o

desenvolvimento [19]. Sendo uma linguagem de programação de alto nível, interpretada de script, imperativa, orientada a objetos, funcional de tipagem dinâmica e forte. Foi lançada por Guido Van Rossum em 1991.

Figura 5. Python logo.



Fonte: [20]

Python é amplamente adotado tanto na extração quanto na análise de dados devido à sua flexibilidade, acessibilidade e natureza de código aberto. Esses atributos fazem com que seja a linguagem de escolha para profissionais que lidam com dados, beneficiando-se de uma comunidade vibrante e engajada, além de um vasto conjunto de bibliotecas que expandem suas funcionalidades de forma significativa. Essas características não apenas simplificam o processo de desenvolvimento de soluções analíticas, mas também promovem a colaboração e inovação contínua dentro da comunidade de ciência de dados e análise de dados.

2.3.1 PySpark

PySpark (Figura 5) é uma biblioteca que oferece uma interface do Apache Spark para Python, permitindo que desenvolvedores escrevam aplicativos Spark utilizando as APIs do Python. Além disso, fornece o shell PySpark, que permite a análise interativa de dados em um ambiente distribuído. PySpark suporta a maioria dos recursos do Spark, incluindo Spark SQL, DataFrame, Streaming, MLlib (Machine Learning) e Spark Core [17].

Figura 6. Apache PySpark logo.



Fonte: [21]

- **Spark SQL e DataFrame:** Spark SQL é um módulo do Spark voltado para o processamento de dados estruturados. Ele introduz uma abstração de programação chamada DataFrame, que facilita a manipulação de grandes volumes de dados em um formato tabular. Além disso, Spark SQL atua como um mecanismo de consulta SQL distribuído, permitindo consultas complexas em grandes conjuntos de dados de forma eficiente.
- **API do Pandas no Spark:** Para usuários familiarizados com o Pandas, PySpark oferece uma API compatível que permite o uso imediato sem a necessidade de uma curva de aprendizado acentuada. Essa integração garante uma base de código única que funciona tanto com Pandas (ideal para testes e conjuntos de dados menores) quanto com Spark (adequado para grandes conjuntos de dados distribuídos).
- **Streaming:** O recurso de streaming do Apache Spark capacita a criação de aplicativos interativos e análises poderosas tanto em dados de streaming em tempo real quanto em dados históricos. Este recurso herda a facilidade de uso e a robustez contra falhas do Spark, permitindo a análise contínua e em tempo real dos fluxos de dados.

- **MLlib:** MLlib é a biblioteca de aprendizado de máquina escalável do Spark. Ela fornece um conjunto uniforme de APIs de alto nível que ajudam os usuários a criar, ajustar e implementar pipelines de aprendizado de máquina. MLlib inclui ferramentas para tarefas comuns de aprendizado, como classificação, regressão, clustering e filtragem colaborativa, otimizadas para trabalhar em grandes volumes de dados.
- **Spark Core:** O Spark Core é o mecanismo de execução fundamental subjacente à plataforma Spark. Ele é responsável por todas as funcionalidades básicas e fornece a Resilient Distributed Dataset (RDD), uma estrutura de dados distribuída que permite a computação na memória e a manipulação eficiente de grandes conjuntos de dados. O Spark Core gerencia a distribuição de tarefas, a recuperação de falhas e a otimização das operações de computação distribuída.

Essas características tornam o PySpark uma ferramenta poderosa para o processamento e análise de grandes volumes de dados, combinando a facilidade de uso do Python com a eficiência e escalabilidade do Apache Spark.

2.3.2 Numpy e Pandas

NumPy (Figura 5), uma abreviação de Numerical Python, é uma biblioteca de código aberto do Python para computação científica, campo de estudo que utiliza recursos computacionais para entender e resolver problemas. Essa biblioteca permite trabalhar com a manipulação de objetos array multidimensionais, além de seus objetos derivados, como matrizes, sequências, e outros. Além disso, também possui uma grande variedade de operações rápidas com os arrays, incluindo operações matemáticas e lógicas, manipulações de formato, ordenação e seleção, ferramentas de estatística e cálculo, e muito mais.[23]

A biblioteca NumPy foi lançada em 2005, pelo cientista de dados Travis Oliphant, com a proposta de ser uma ferramenta rápida, eficiente e fácil de usar, permitindo assim a realização de cálculos numéricos e matemáticos em larga escala, por meio da funcionalidade chamada vetorização. Por isso,

tornou-se uma das bases fundamentais para análise de dados, aprendizado de máquina (machine learning) e computação científica em geral, estando presente também na construção de várias bibliotecas de ciência de dados, incluindo: Pandas, Matplotlib, Scikit-learn, SciPy, e muitas outras.[23]

Figura 7. NumPy logo.



Fonte: [23]

Pandas (Figura 8) é uma biblioteca para Ciência de Dados de código aberto (open source), construída sobre a linguagem Python, e que providencia uma abordagem rápida e flexível, com estruturas robustas para se trabalhar com dados relacionais (ou rotulados), e tudo isso de maneira simples e intuitiva.[24]

Figura 8. Pandas logo.



Fonte: [23]

De maneira geral, o Pandas pode ser utilizado para várias atividades e processos, entre eles: limpeza e tratamento de dados, análise exploratória de dados (EDA), suporte em atividades de Machine Learning, consultas e queries em bancos de dados relacionais, visualização de dados, webscraping e muito mais. E além disso, também possui ótima integração com várias outras bibliotecas muito utilizadas em

Ciência de Dados, tais como: Numpy, Scikit-Learn, Seaborn, Altair, Matplotlib, Plotly, Scipy e outros.[24]

2.3.3 Matplotlib

Matplotlib (Figura 9) é uma biblioteca de plotagem em Python amplamente utilizada para criar visualizações estáticas, animadas e interativas de dados. Ela é especialmente popular entre cientistas de dados, analistas e pesquisadores devido à sua flexibilidade e capacidade de produzir gráficos de alta qualidade de forma relativamente simples. Aqui estão alguns aspectos importantes do Matplotlib:

Versatilidade: Matplotlib pode ser usado para criar uma ampla variedade de gráficos e visualizações, incluindo gráficos de linhas, gráficos de dispersão, histogramas, gráficos de barras, gráficos de pizza, gráficos de caixas, gráficos de erro, entre outros. A biblioteca é altamente personalizável, permitindo ajustes detalhados na aparência dos gráficos.

Integração com Outras Bibliotecas: Matplotlib integra-se bem com outras bibliotecas populares de análise de dados em Python, como NumPy, Pandas e SciPy. Isso facilita o uso de Matplotlib para visualizar dados que foram processados e analisados com essas bibliotecas.

API Funcional e Orientada a Objetos: Matplotlib oferece uma API funcional simples, bem como uma API orientada a objetos mais avançada. A API funcional, através do módulo pyplot, é fácil de usar para criar gráficos rápidos e simples. A API orientada a objetos proporciona maior controle e flexibilidade, permitindo a criação de gráficos complexos e sofisticados.

Documentação e Comunidade: Matplotlib tem uma documentação extensa e detalhada, com muitos exemplos e tutoriais disponíveis. A comunidade ao redor da biblioteca é grande e ativa, o que significa que há muitos recursos e suporte disponíveis online.

Ferramentas de Interatividade: Embora Matplotlib seja mais conhecida por suas visualizações estáticas, ela também oferece suporte para gráficos interativos. Com a ajuda de bibliotecas adicionais, como mpl_toolkits e mpld3, é possível criar visualizações que permitem zoom, pan e outras formas de interatividade.

Exportação e Publicação: Os gráficos criados com Matplotlib podem ser exportados em diversos formatos, como PNG, PDF, SVG e EPS. Isso torna a biblioteca uma escolha ideal para criar visualizações que serão incluídas em relatórios, apresentações e publicações científicas.

Figura 9. Matplotlib logo



Fonte: [25]

2.3.4 Scikit Learn

O scikit-learn (Figura 10) é uma biblioteca Python projetada especificamente para a aplicação prática de machine learning. Ela oferece ferramentas simples e eficientes para análise preditiva de dados, é reutilizável em várias situações, e é de código aberto, o que a torna acessível a todos. A biblioteca é construída sobre os pacotes NumPy, SciPy e Matplotlib, aproveitando suas funcionalidades robustas para cálculos numéricos, processamento científico e visualização de dados.

Python é uma das melhores linguagens para a aplicação prática de machine learning, em grande parte devido ao scikit-learn. As principais aplicações da biblioteca incluem:

- **Pré-processamento:** Preparação e limpeza dos dados para análise.
- **Classificação:** Atribuição de categorias a dados com base em exemplos.
- **Regressão:** Modelagem da relação entre variáveis para prever valores contínuos.
- **Clusterização:** Agrupamento de dados sem rótulos em grupos com características semelhantes.
- **Redução de Dimensionalidade:** Simplificação de conjuntos de dados complexos mantendo as características essenciais.
- **Ajuste de Parâmetros:** Otimização de parâmetros de modelos para melhorar o desempenho preditivo.

Scikit-learn é amplamente utilizado na comunidade de ciência de dados devido à sua facilidade de uso, documentação abrangente e capacidade de integrar-se bem com outras bibliotecas populares em

Python.

Figura 10. Scikit Learn logo



Fonte: [26]

2.3.5 Tweepy

Tweepy (Figura 11) é uma biblioteca Python que permite interagir com a API do Twitter de forma fácil e eficiente. Esta biblioteca facilita o acesso e a manipulação dos dados disponíveis no Twitter, possibilitando a criação de aplicações que podem buscar, enviar e gerenciar tweets, além de outras funcionalidades relacionadas à plataforma. Aqui estão alguns aspectos importantes sobre o Tweepy:

- **Autenticação:** Tweepy facilita o processo de autenticação na API do Twitter usando OAuth, garantindo que as aplicações possam se conectar de forma segura à plataforma do Twitter.
- **Acesso aos Dados do Twitter:** Com o Tweepy, é possível buscar tweets, perfis de usuários, listas de seguidores, trends e muito mais. A biblioteca fornece métodos para acessar uma ampla gama de endpoints da API do Twitter.
- **Envio de Tweets:** A biblioteca permite enviar tweets, retweetar, responder a tweets, gostar de tweets e enviar mensagens diretas, entre outras ações.
- **Streaming de Dados:** Tweepy oferece uma interface para a API de Streaming do Twitter, permitindo que desenvolvedores escutem em tempo real por tweets que correspondam a determinados critérios, como hashtags ou palavras-chave.

- **Manipulação de Perfis:** A biblioteca permite a manipulação de perfis de usuário, como seguir ou deixar de seguir outros usuários, atualizar informações de perfil e gerenciar listas de amigos e seguidores.
- **Facilidade de Uso:** Tweepy é conhecido por sua simplicidade e facilidade de uso, tornando o desenvolvimento de aplicações relacionadas ao Twitter mais acessível, mesmo para desenvolvedores com menos experiência.

Figura 11. Tweepy logo



Fonte: [29]

2.3.6 NLTK

O Python Nltk (Natural Language Toolkit) é uma biblioteca de processamento de linguagem natural desenvolvida em Python. Ela fornece uma ampla gama de ferramentas e recursos para lidar com texto e linguagem, permitindo que os desenvolvedores implementem soluções avançadas de análise e processamento de texto. O Nltk (Figura 12) é amplamente utilizado em aplicações de aprendizado de máquina, mineração de texto, extração de informações e muito mais.[30]

Figura 12. NLTK logo



Fonte: [31]

2.3.7 Jupyter

Os Jupyter Notebooks são usados para tudo o que acontece na data science, incluindo análise exploratória de dados (EDA), limpeza e transformação de dados, visualização de dados, modelagem estatística, machine learning, deep learning e muito mais.[32]

Figura 12. Jupyter logo



Fonte: [32]

2.3.8 Seaborn

Seaborn é uma biblioteca Python de visualização de dados baseada no Matplotlib. Ela fornece uma interface de alto nível para criar gráficos estatísticos atraentes e informativos. Seaborn é amplamente utilizada em análise de dados exploratória e visualização de dados em Python devido à sua facilidade de uso e à capacidade de produzir gráficos esteticamente agradáveis com poucas linhas de código.

Figura 13. Seaborn logo



Fonte: [33]

2.3.9 Word Cloud

Word cloud, ou nuvem de palavras, é uma representação visual que destaca as palavras mais frequentes em um texto, onde o tamanho de cada palavra é proporcional à sua frequência de ocorrência. Essa técnica é utilizada principalmente para dar uma visão rápida e intuitiva dos temas mais relevantes ou predominantes em um conjunto de dados textual.

Na Figura 14 conseguimos ver um exemplo de como o word cloud fica.

Figura 14. Exemplo Word Cloud



2.4 Análise de Sentimentos

Análise de sentimentos é o processo de analisar um texto digital para determinar se o tom emocional da mensagem é positivo, negativo ou neutro. Atualmente, as empresas têm grandes volumes de dados textuais, como e-mails, transcrições de chat de atendimento ao cliente, comentários em mídias sociais e avaliações. As ferramentas de análise de sentimentos podem escanear esse texto para determinar automaticamente a atitude do autor em relação a um tópico. As empresas usam os insights da análise de sentimentos para melhorar o atendimento ao cliente e aumentar a reputação da marca. [35]

Também conhecida como mineração de dados, a análise de sentimento é uma importante ferramenta de BI (business intelligence) que ajuda empresas na tomada de decisões, a melhorar seus produtos e serviços. Abaixo irei apresentar algumas etapas da análise de sentimento.

Pré-processamento

- Na etapa de pré-processamento, a análise de sentimentos destaca palavras-chave para identificar a mensagem central do texto.
- A tokenização divide a frase em elementos menores, chamados tokens.
- A lematização converte palavras em suas formas raiz. Exemplo: a forma raiz de "sou" é "ser".
- A remoção de palavras irrelevantes elimina termos sem valor significativo, como "com", "para", "em" e "de".

Análise de Palavras-chave

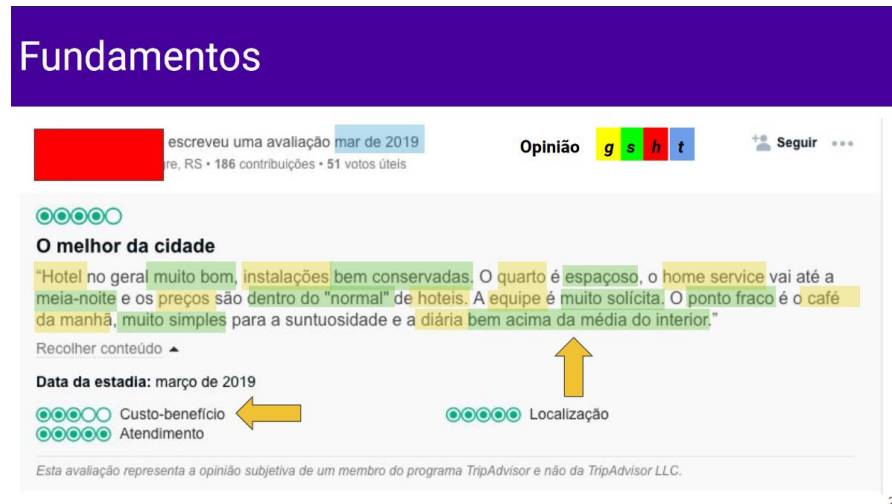
Tecnologias de PLN (Processamento de Linguagem Natural) avaliam palavras-chave extraídas, atribuindo-lhes uma pontuação de sentimento. Essa pontuação mede a emoção expressa no texto. Por exemplo, em avaliações de clientes, "10" pode indicar satisfação e "0" decepção.

2.4.1 Definição de Opinião

Uma opinião pode ser descrita como uma quádrupla como a Figura 15 abaixo (g, s, h, t), onde:

- g é o alvo da opinião,
- s é o sentimento em relação ao alvo,
- h é o autor da opinião,
- t é o tempo em que a opinião foi expressa.

Figura 15. Exemplo de opinião quádrupla



26

Fonte: [36]

No entanto, esta definição não é muito prática porque a descrição completa do alvo frequentemente não está na mesma sentença. Para resolver isso, adiciona-se um quinto elemento: a entidade e, que é descrita por um par (T, W). Aqui, T representa uma hierarquia de partes e subpartes da entidade, e W é o conjunto de atributos dessa entidade.

Portanto, uma opinião pode ser redefinida como uma quintupla $(ei, aij, sijkl, hk, tl)$:

- ei é o nome de uma entidade,
- aij é um aspecto dessa entidade,
- $sijkl$ é o sentimento sobre esse aspecto,
- hk é o autor da opinião,
- tl é o tempo em que a opinião foi expressa.

Neste trabalho, a entidade é o autor dos tweets e os aspectos são as palavras do conteúdo dos tweets. O objetivo é analisar o sentimento de cada aspecto e determinar o sentimento geral do tweet.

O sentimento (*sijkl*) pode ser positivo, negativo, neutro ou ter diferentes níveis de intensidade, como observado em pesquisas de opinião e questionários que oferecem respostas como "discordo" ou "discordo plenamente".

2.4.2 Dicionários Léxicos

Um dicionário léxico é um conjunto de palavras associadas a valores de sentimento positivo ou negativo. Existem três métodos para construir um dicionário léxico: manualmente, com base em dicionários existentes, e a partir de corpus de texto.

Neste trabalho, será utilizado o um dicionário léxico:

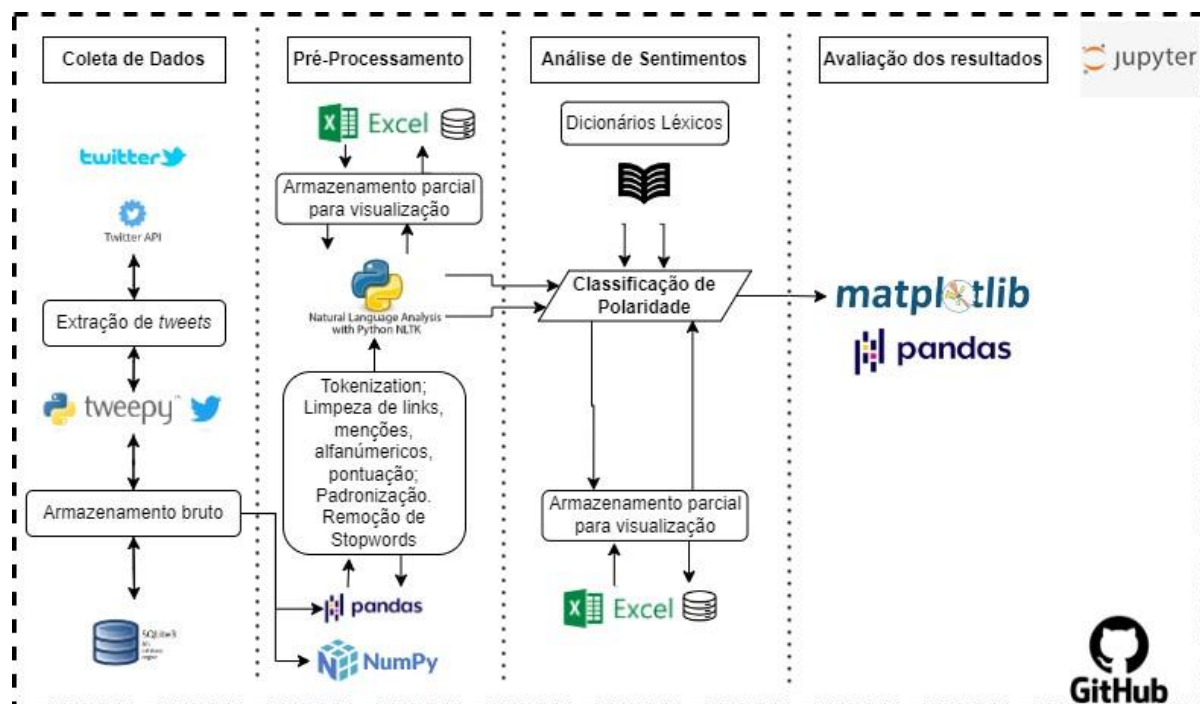
SentiWordNet: um recurso léxico específico para mineração de opinião, que atribui a cada palavra do WordNet três pontuações de sentimento: positiva, negativa ou neutra.[37]

3.0 Metodologia

Neste capítulo será apresentado a metodologia seguida nesse trabalho, assim como os códigos durante o estudo, mostrando a aplicação dos conceitos vistos nos capítulos anteriores.

Segue abaixo na Figura 16 a metodologia seguida nesse trabalho. As etapas seguidas foram: coleta de dados, pré-processamento, análise de sentimento e avaliação dos resultados.

Figura 16. Exemplo de metodologia



Fonte: [38]

3.1 Ferramentas Utilizadas

Este estudo foi conduzido utilizando a linguagem de programação Python em conjunto com a ferramenta Jupyter Notebook, um ambiente web open-source projetado para experimentação interativa. Similar a cadernos de anotações, o Jupyter Notebook permite escrever, executar e salvar código e informações em tempo real.

Para a coleta de tweets, foi empregada a biblioteca Tweepy, que facilita o acesso à API do Twitter. Os dados brutos foram armazenados em um banco de dados relacional SQLite3, onde cada tweet e suas respectivas informações foram registrados como entradas individuais.

No processo de pré-processamento dos dados, foram utilizadas as bibliotecas Pandas e Numpy para manipulação e operações eficientes, além da NLTK (Natural Language Toolkit) para tarefas como tokenização e remoção de stopwords (palavras comuns que não contribuem para a análise de sentimentos).

Para a análise de sentimentos, foi adotado o dicionário léxico SentiWordNet [37]. Por fim, os resultados foram avaliados e visualizados com o auxílio das bibliotecas Matplotlib e Seaborn, que são amplamente utilizadas para a criação de gráficos e visualizações de dados.

3.2 Coleta de dados

Para acessar a API do Twitter, primeiramente é necessário registrar um aplicativo e obter as credenciais de autenticação. Por padrão, o acesso é limitado a informações públicas. No estudo mencionado, foi usada a API v2 com acesso Essencial.

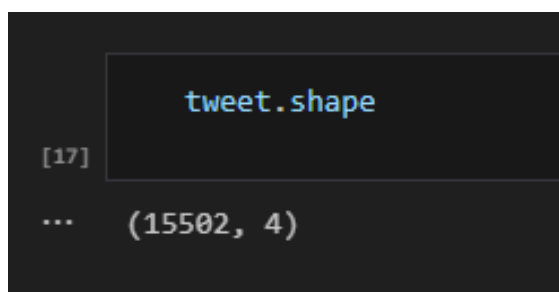
Com a autenticação feita, usa-se o Tweepy para criar uma solicitação com a função “search_recent_tweets”, que aceita vários parâmetros. Neste caso, foi usado:

- query: para definir as palavras-chave da busca;
- tweet_fields: para especificar quais campos dos tweets serão retornados;
- max_results: para limitar o número de tweets obtidos;
- start_time e end_time: para delimitar o intervalo temporal dos tweets buscados.

Para essa análise foi analisado 15502 tweets como mostra na Figura 17, e foi construído quatro colunas para facilitar a análise.

- tweet: a mensagem que foi postada no twitter;
- date: a data em que o tweet foi postado;
- Agent: o agente em que a mensagem se refere;
- Role: posição em que o agente joga, tendo como opção: controller(controlador), duelist(duelista), Initiator(iniciador) e Sentinel(sentinela)

Figura 17. Informações sobre o dataset



Fonte: Elaborado pela Autora

3.3 Pré-processamento

Como é possível ver na Figura 18 os tweets brutos tem divertos problemas em seu texto.

Figura 18. Informações sobre o dataset

@jellycutesenpai Fr FIRE AND BRIMSTONE
@CarolChristof13 @CoGatorKing @FedExField @WashingtonNFL @USARugby @AllBlacks I'm not a psychiatrist
@Lj_era8 @jtuck9 you guys were Fire and Brimstone on that final drive towards glory!! Lions became kittens on 2 plays 100% kissed by God.
@_greenandpurple Yeah
@MarkRPellegrino @whywouldyou_uk @ExposeFakestine @mccormicky @WeinsteinDr @natashahy61225 @ftysugar @terangbulanmlm @HebMacMan @israelies_21
@zero_kgz what ðŸ˜˜-ðŸ˜˜%â€œâ€œ"â€œ, ðŸ˜˜-â€œ is good ðŸ˜˜-ðŸ˜˜%â€œ... for liar ðŸ˜˜-â€œâ€œ brimstone ðŸ˜˜- and fire ðŸ˜˜-â€œ
But the fearful
Fire and brimstone coming down from the skies!
Pork = fire
@Quikwest @NACHOBRED Lmfaoo 2 tweets of fire and brimstone!ðŸ˜˜-ðŸ˜˜-â€œ
@Chortling_Man Describe â€œlateral Hell?â€œ Is it a place of fire and brimstone where the souls of unredeemed sinners and non believers will writhe and burn for eternity?

Fonte: Elaborado pela autora

Após a análise dos problemas no texto, foi definido outros processamentos para o texto:

- Remover linhas duplicadas;
- Remover parte originaria do Retweet;
- Remover menções a usuários;
- Remover links;
- Remover numéricos e caracteres especiais;
- Remover emojis
- Remover pontuações;
- Remover quebras de linha;
- Remover espaços duplos;
- Padronizar todo o texto para letras minúsculas;
- Remover letras duplas (exceto, 'ss' e 'rr').

Figura 19. Resultado Pré-processamento

	tweet	tweet_lower	tweet_punctual	no_emoji	tweet_clean
0	and now i want brimstone in my garden\ni want ...	and now i want brimstone in my garden\ni want ...	and now i want brimstone in my garden\ni want ...	and now i want brimstone in my garden\ni want ...	and now i want brimstone in my garden\ni want ...
1	why half of the brimstone players are literall...	why half of the brimstone players are literall...	why half of the brimstone players are literall...	why half of the brimstone players are literall...	why half of the brimstone players are literall...
2	@DavidLevy101 @Meloniemaher1 That fucker gets ...	@davidlevy101 @meloniemaher1 that fucker gets ...	davidlevy101 meloniemaher1 that fucker gets th...	davidlevy101 meloniemaher1 that fucker gets th...	davidlevy101 meloniemaher1 that fucker gets th...
3	They also did these cute lil' ghosts! \n\nRobo...	they also did these cute lil' ghosts! \n\nrobo...	they also did these cute lil ghosts \n\nrobot ...	they also did these cute lil ghosts \n\nrobot ...	they also did these cute lil ghosts \n\nrobot ...
4	@AuthorDelaneyW @mis_cue @C_Stroop That is not...	@authordelaneyw @mis_cue @c_stroop that is not...	authordelaneyw miscue cstroop that is not true...	authordelaneyw miscue cstroop that is not true...	authordelaneyw miscue cstroop that is not true...

Fonte: Elaborado pela autora

3.4 Análise de sentimento

Nesta etapa que acontece de fato a análise sobre os tweets coletados e pré-processados. Esse é o momento onde os tweets terão suas palavras separadas, e a classificação, onde cada palavra terá um score (que variam de -1 a 1).

Essa pontuação (Figura 20) será totalizada para cada tweet, assim teremos um resultado de polaridade para cada tweet de acordo com o dicionário.

Figura 20. Classificação Final para cada tweet

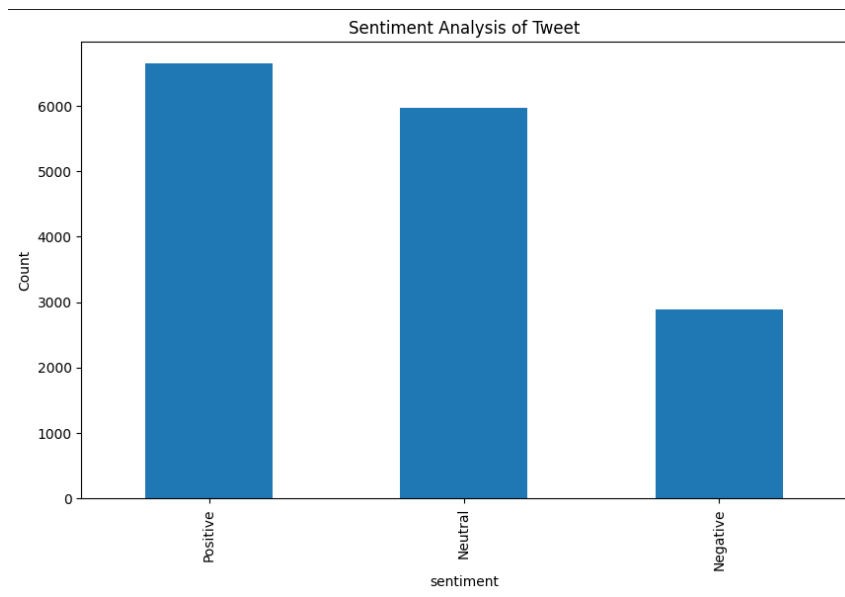
	tweet_clean	subjectivity	polarity	sentiment
0	and now i want brimstone in my garden\ni want ...	0.000000	0.000000	Neutral
1	why half of the brimstone players are literall...	0.166667	-0.166667	Negative
2	davidlevy101 meloniemaher1 that fucker gets th...	0.000000	0.000000	Neutral
3	they also did these cute lil ghosts \n\nrobot ...	0.600000	0.400000	Positive
4	authordelaneyw miscue cstroop that is not true...	0.816667	-0.391667	Negative

Fonte: Elaborado pela autora

4.0 Resultados

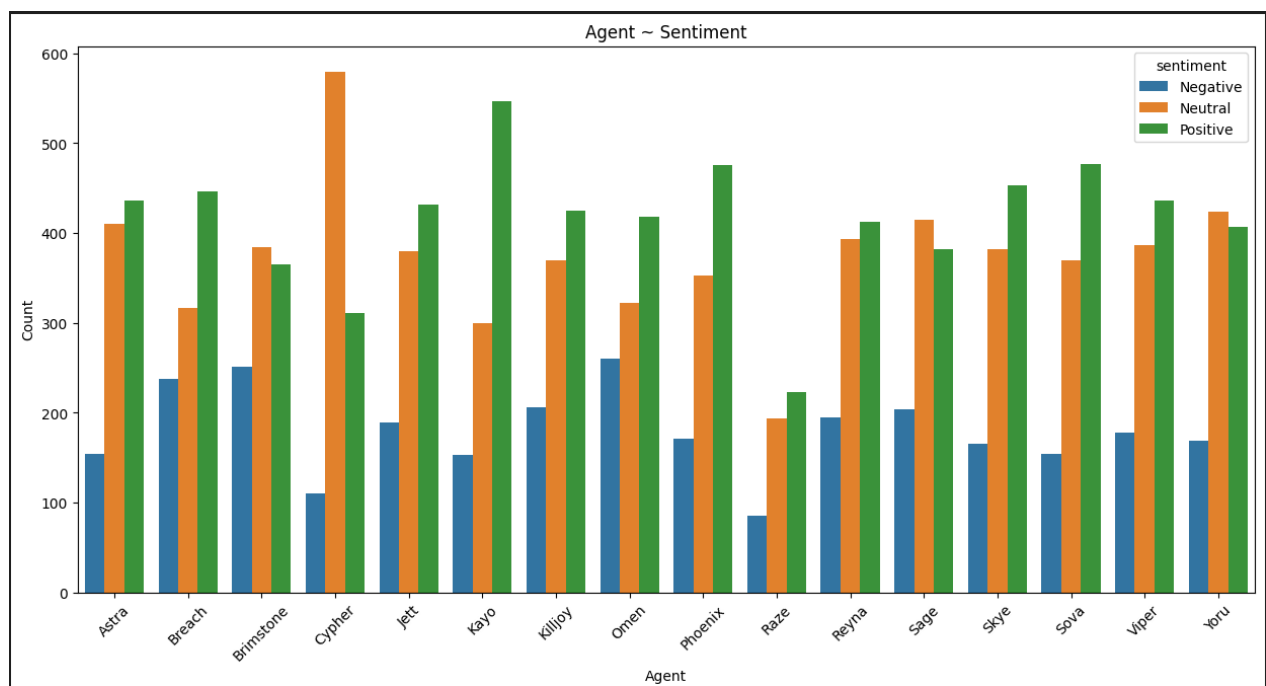
Neste capítulo, serão apresentados os resultados obtidos no trabalho. Lembrando que o número total de tweets classificados é de 15000.

Figura 21. Gráfico polaridade por sentimento (geral)



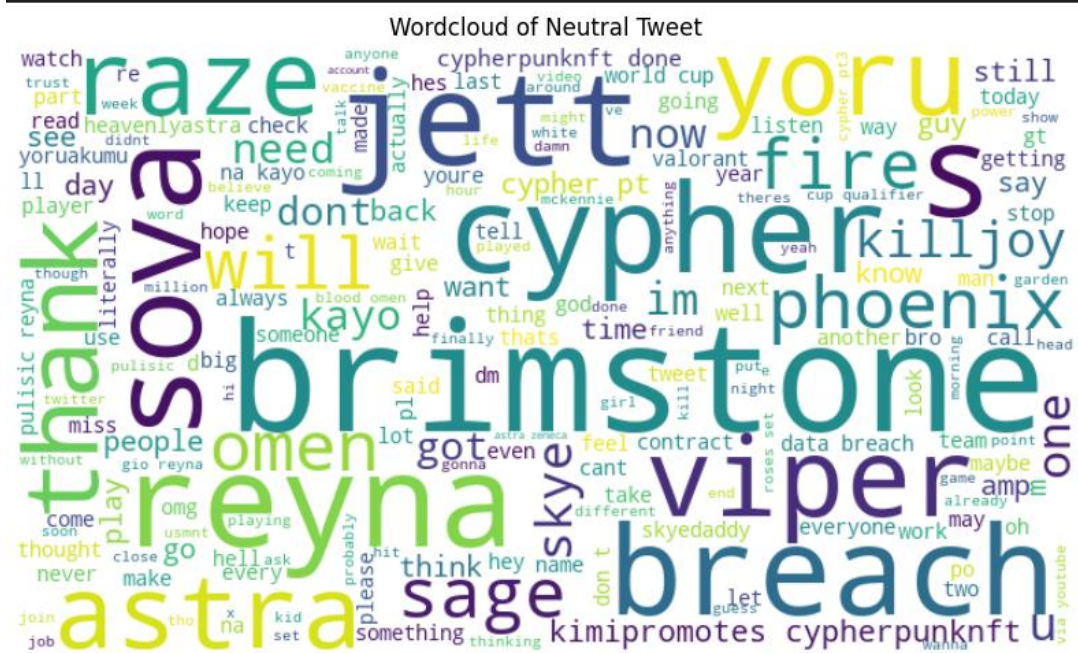
Fonte: Elaborado pela autora

Figura 22. Gráfico polaridade por sentimento para cada agente



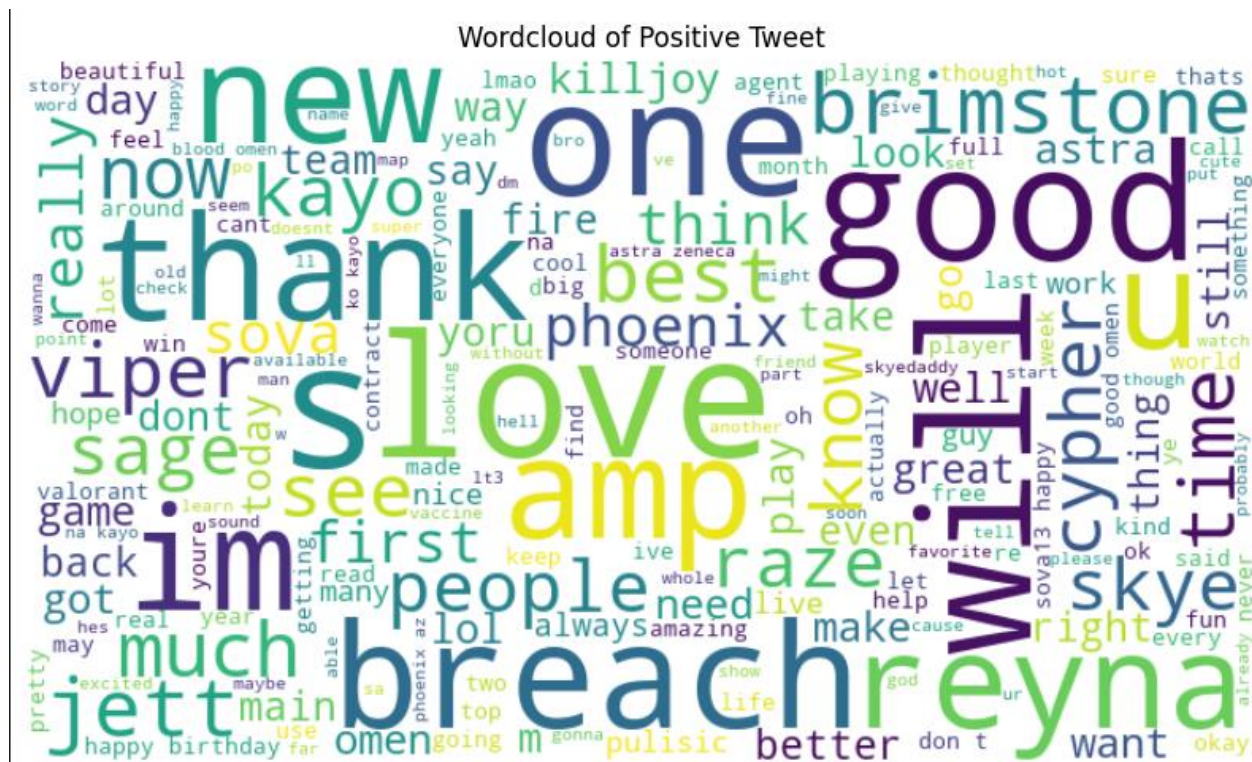
Fonte: Elaborado pela autora

Figura 23. Gráfico polaridade por cada posição



Fonte: Elaborado pela autora

Figura 25. Wordcloud de palavras em tweets positivos



Fonte: Elaborado pela autora

Com base nos resultados obtidos, foi possível identificar insights interessantes sobre a recepção

dos jogadores em relação aos diferentes agentes do Valorant. Observamos, por exemplo, que certos agentes podem despertar emoções mais positivas ou negativas entre os jogadores. Além disso, a análise de tópicos revelou os principais aspectos que os jogadores discutem ao mencionar cada agente.

Essas descobertas fornecem informações valiosas para os desenvolvedores do jogo, que podem utilizar esses insights para ajustar e balancear os agentes de acordo com as demandas e expectativas da comunidade.

5.0 Conclusão

A análise das percepções dos jogadores sobre os agentes do popular jogo Valorant, conduzida por meio de técnicas avançadas de Processamento de Linguagem Natural (NLP), revelou uma série de insights profundos e significativos sobre as preferências e opiniões da vibrante comunidade de jogadores. Ao examinar os dados coletados do Twitter, foi possível identificar padrões intrigantes nas opiniões dos jogadores sobre os diferentes agentes, destacando aspectos essenciais como a popularidade de cada agente e as roles mais frequentemente utilizadas pelos jogadores.

Este estudo exemplifica vividamente o potencial transformador das técnicas de NLP na análise de dados do Twitter para compreender as percepções dos jogadores em relação aos jogos eletrônicos. Além disso, abre portas para uma ampla gama de pesquisas futuras nessa área, promovendo uma compreensão mais profunda e abrangente das interações entre os jogadores, os jogos e as plataformas sociais. Ao unir tecnologia de ponta com a riqueza de dados disponíveis nas redes sociais, este trabalho destaca a importância crescente da análise de dados para informar e aprimorar continuamente a experiência dos jogadores nos jogos eletrônicos modernos.

REFERÊNCIAS

- [1] NEWZOO. Video games in 2023: the year in numbers. Disponível em: <https://newzoo.com/resources/blog/video-games-in-2023-the-year-in-numbers>. Acesso em: 16 maio 2024.
- [2] Esports.net. Quantas pessoas jogaram Valorant em 2024? Disponível em: <https://www.esports.net/br/noticias/valorant/quantas-pessoas-jogaram-em-2024/>. Acesso em: 16 maio 2024.
- [3] GUSMÃO, Amanda. Afinal, para que serve o Twitter? Descubra como usar essa rede social. Rock Content, 20 maio 2022. Disponível em: <https://rockcontent.com/br/blog/para-que-serve-o-twitter/>. Acesso em: 18 maio 2024.
- [4] DATAREPORTAL. Digital 2024: Deep Dive - Social Media is Still Growing. Disponível em: <https://datareportal.com/reports/digital-2024-deep-dive-social-media-is-still-growing?rq=social%20media%2024>. Acesso em: 18 maio 2024.
- [5] INVESTOPEDIA. Social Networking. Disponível em: <https://www.investopedia.com/terms/s/social-networking.asp>. Acesso em: 18 maio 2024.
- [6] SANTOS, José Osmar da Silva; ANDRADE, Vera Lúcia Lopes; SOUZA, Maria Auxiliadora Cárnio. Rede Social: Um Estudo Sobre as Conexões Entre as Bibliotecas Universitárias da Cidade de São Paulo. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 17., 2012, Gramado. Anais [...]. Gramado:SNBU, 2012. Disponível em: http://repositorio.febab.org.br/files/original/49/5868/SNBU2012_007.pdf. Acesso em: 18 maio 2024.
- [7] QUICK. Impacto das Redes Sociais na Sociedade. Disponível em: <https://quick.com.br/impacto-das-redes-sociais-na-sociedade/>. Acesso em: 18 maio 2024.
- [8] LENOVO. O que é Twitter? Disponível em: <https://www.lenovo.com/br/pt/faqs/pc-faqs/que-e-twitter/?orgRef=https%253A%252F%252Fwww.google.com%252F>. Acesso em: 18 maio 2024.
- [9] DEMANDSAGE. Twitter Statistics. Disponível em: <https://www.demandsage.com/twitter-statistics/>. Acesso em: 18 maio. 2024.
- [10] ROCK CONTENT. Big Data: o que é, importância e como usar na sua estratégia. Disponível em:

<https://rockcontent.com/br/blog/big-data/>. Acesso em: 18 maio 2024.

[11] AWARI. Redes sociais em big data: análise e otimização de dados sociais. Disponível em: [https://awari.com.br/redes-sociais-em-big-data-analise-e-otimizacao-de-dados-](https://awari.com.br/redes-sociais-em-big-data-analise-e-otimizacao-de-dados-sociais/?utm_source=blog&utm_campaign=projeto+blog&utm_medium=Redes%20Sociais%20em%20Big%20Data:%20An%C3%A1lise%20e%20Otimiza%C3%A7%C3%A3o%20de%20Dados%20Sociais)

[sociais/?utm_source=blog&utm_campaign=projeto+blog&utm_medium=Redes%20Sociais%20em%20Big%20Data:%20An%C3%A1lise%20e%20Otimiza%C3%A7%C3%A3o%20de%20Dados%20Sociais](https://awari.com.br/redes-sociais-em-big-data-analise-e-otimizacao-de-dados-sociais/?utm_source=blog&utm_campaign=projeto+blog&utm_medium=Redes%20Sociais%20em%20Big%20Data:%20An%C3%A1lise%20e%20Otimiza%C3%A7%C3%A3o%20de%20Dados%20Sociais). Acesso em: 18 maio 2024.

[12] LYNCAS. 5 Vs do Big Data. Disponível em: <https://lyncas.net/5-vs-do-big-data/>. Acesso em: 18 maio 2024.

[13] SOCIAL GOOD BRASIL. Entenda tudo sobre dados para se tornar uma organização data-driven. Disponível em: <https://socialgoodbrasil.org.br/2019/05/15/entenda-tudo-sobre-dados-para-se-tornar-uma-organizacao-data-driven/>. Acesso em: 18 maio 2024.

[14] ORACLE. What Is Big Data? Disponível em: <https://www.oracle.com/br/big-data/what-is-big-data/>. Acesso em: 18 maio 2024.

[15] CHEN, Min; MAO, Shiwen; LIU, Yunhao. Big data: A survey. Mobile Networks and Applications, v. 19, n. 2, p. 171-209, 2014.

[16] MARR, Bernard. Big Data: Using Smart Big Data, Analytics and Metrics to Make Better Decisions and Improve Performance. 1. ed. Chichester: John Wiley & Sons, 2015.

[17] MINELLI, Michael; CHAMBERS, Michele; DWYER, Ambiga. Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses. 1. ed. Hoboken: John Wiley & Sons, 2013.

[18] GARTNER, Inc. IT Glossary. Big Data. Disponível em: <https://www.gartner.com/en/information-technology/glossary/big-data>. Acesso em: 19 maio 2024.

[19] AMAZON WEB SERVICES (AWS). What Is Python? Disponível em: <https://aws.amazon.com/pt/what-is/python/>. Acesso em: 16 jun. 2024.

[20] WIKIPÉDIA. Python logo and wordmark.svg. Disponível em: https://pt.m.wikipedia.org/wiki/Ficheiro:Python_logo_and_wordmark.svg. Acesso em: 19 maio 2024.

- [21] DATASIDE. PySpark: o que é, vantagens e usos. Disponível em: <https://www.dataside.com.br/dataside-community/big-data/pyspark-o-que-e-vantagens-e-usos>. Acesso em: 19 maio 2024.
- [22] DATAGEEKS. PySpark. Disponível em: <https://www.datageeks.com.br/pyspark/>. Acesso em: 19 maio 2024.
- [23] ALURA. NumPy: computação científica com Python. Disponível em: <https://www.alura.com.br/artigos/numpy-computacao-cientifica-com-python>. Acesso em: 19 maio 2024.
- [24] ALURA. Pandas: o que é, para que serve, como instalar. Disponível em: <https://www.alura.com.br/artigos/pandas-o-que-e-para-que-serve-como-instalar>. Acesso em: 19 maio 2024.
- [25] MACHADO, Alysson. Biblioteca Matplotlib do Python. Disponível em: <https://medium.com/@alyssonmachado388/biblioteca-matplotlib-do-python-5e32104439d0>. Acesso em: 19 maio 2024.
- [26] DIDÁTICA TECH. A biblioteca Scikit-Learn: Python para Machine Learning. Disponível em: <https://didatica.tech/a-biblioteca-scikit-learn-pyhton-para-machine-learning/>. Acesso em: 19 maio 2024.
- [27] DADOS AO CUBO. Extraindo dados do Twitter com Python. Disponível em: <https://dadosaocubo.com/extraindo-dados-do-twitter-com-python/>. Acesso em: 19 maio 2024.
- [28] TWEETPY. Tweepy. Disponível em: <https://www.tweepy.org/>. Acesso em: 19 maio 2024.
- [29] MUNENE, Rick. Scraping Tweets with Tweepy (Python). Medium, 2020. Disponível em: <https://rickmunene.medium.com/scraping-tweets-with-tweepy-python-a2364cf8084f>. Acesso em: 19 maio. 2024.
- [30] AWARI. Aprenda Python com o poderoso NLTK: o guia completo para iniciantes. Disponível em: <https://awari.com.br/aprenda-python-com-o-poderoso-nltk-o-guia-completo-para-iniciantes/>. Acesso em: 19 maio.
- [31] LANE, Kelsk. The Natural Language Toolkit: What is it? Medium, 2021. Disponível em: <https://medium.com/@kelsklane/the-natural-language-toolkit-what-is-it-8ca335dce6a9>. Acesso em: 19 maio.
- [32] DATABRICKS. Jupyter Notebook. Disponível em: <https://www.databricks.com/br/glossary/jupyter-notebook>. Acesso em: 19 maio.
- [33] WASKOM, M.. Seaborn: statistical data visualization. GitHub, 2021. Disponível em:

<https://github.com/mwaskom/seaborn>. Acesso em: 19 maio.

[34] RAILAN, Deivid. Tratamento de texto e WordCloud com Python. LinkedIn, 2021. Disponível em: <https://www.linkedin.com/pulse/tratamento-de-texto-e-wordcloud-com-python-railan-deivid/>. Acesso em: 19 maio.

[35] AMAZON WEB SERVICES. What is Sentiment Analysis? Disponível em: <https://aws.amazon.com/pt/what-is/sentiment-analysis/>. Acesso em: 20 maio.

[36] FREITAS, Larissa. Título do Documento. São Paulo: USP, 2023. Disponível em: https://edisciplinas.usp.br/pluginfile.php/8070484/mod_resource/content/1/LarissaFreitas29Nov23.pdf. Acesso em: 26 maio.

[37] SentiWordNet. Disponível em: <https://github.com/Pedro-Thales/SentiWordNet-PT-BR>. Acesso em: 26 maio.

[38] CASTRO, Carlos. Extração de Dados e Análise de Sentimentos. 2021. 100 f. Trabalho de Conclusão de Curso (Graduação em Engenharia Física) - Universidade Federal de São Carlos, São Carlos, 2021. Disponível em: https://repositorio.ufscar.br/bitstream/handle/ufscar/16684/TCC_CarlosCastro628387_revisado.pdf?sequence=1. Acesso em: 30 maio.

[39] GLOBO ESPORTE. Valorant: veja personagens, mapas, armas e tudo sobre o FPS. Globo Esporte, Rio de Janeiro, 23 maio 2023. Disponível em: <https://ge.globo.com/esports/valorant/reportagem/2023/05/23/c-valorant-veja-personagens-mapas-armas-e-tudo-sobre-o-fps.ghtml>. Acesso em: 17 jun. 2024.

[40] RIOT GAMES. Valorant logo - pink color version.svg. Wikimedia Commons, 16 maio 2020. Disponível em: [Ficheiro:Valorant logo - pink color version.svg](#). Acesso em: 17 jun. 2024.

