

**CENTRO ESTADUAL DE EDUCAÇÃO TECNOLÓGICA PAULA SOUZA**  
**Faculdade de Tecnologia de Jundiaí – “Deputado Ary Fossen”**  
**Curso Superior de Tecnologia em Gestão da Tecnologia da Informação**

Natália Maika de Toledo  
Rhuan Ferreira Nascimento  
Thays Aleksa Afonso  
Vanessa Barros Barbosa da Silva

**BAFAFÁ – ANÁLISE DE SENTIMENTOS DE COMENTÁRIOS EM REDES  
SOCIAIS COMO FERRAMENTA DE SUPORTE A DECISÃO (SAD)**

**Jundiaí**  
**2024**

**Natália Maika de Toledo**  
**Rhuan Ferreira Nascimento**  
**Thays Aleksa Afonso**  
**Vanessa Barros Barbosa da Silva**

**BAFAFÁ – ANÁLISE DE SENTIMENTOS DE COMENTÁRIOS EM REDES  
SOCIAIS COMO FERRAMENTA DE SUPORTE A DECISÃO (SAD)**

Trabalho de Graduação apresentado à Faculdade de Tecnologia de Jundiaí - “Deputado Ary Fossen” como requisito parcial para a obtenção do título de Tecnólogo em Gestão da Tecnologia da Informação, sob a orientação da Professora, Mestre, Adani Cusin Sacilotti.

**Jundiaí**  
**2024**

NATÁLIA MAIKA DE TOLEDO  
RHUAN FERREIRA NASCIMENTO  
THAYS ALEKSA AFONSO  
VANESSA BARROS BARBOSA DA SILVA

**BAFAFÁ – ANÁLISE DE SENTIMENTOS DE COMENTÁRIOS EM REDES  
SOCIAIS COMO FERRAMENTA DE SUPORTE A DECISÃO (SAD)**

Trabalho de Conclusão de Curso para  
obtenção do título de Tecnólogo em  
Gestão da Tecnologia da Informação  
apresentado à Graduação apresentado à  
Faculdade de Tecnologia de Jundiaí -  
“Deputado Ary Fossen”

Jundiaí, \_\_\_\_ de \_\_\_\_\_ de 2024.

**Banca Examinadora:**

---

Professora Mestra Adani Cusin Sacilotti

---

Professor Mestre Carlos Eduardo Schuster

---

Professor Paulo Jacobsen Pelizari Pinto

---

Professora Mestra Rita de Cássia Félix

Dedicamos este trabalho à nós mesmos,  
Aos esforços que empenhamos,  
Mesmo com todas as dificuldades.  
Obrigado a nós,  
Por não desistir  
De nós.

## **AGRADECIMENTOS**

Gostaríamos de expressar nossa profunda gratidão a todos que desempenharam um papel crucial na realização bem-sucedida do nosso trabalho de graduação. Cada um de vocês contribuiu de maneira significativa e gostaria de reconhecê-los aqui:

À Fatec Jundiaí, meu sincero agradecimento por fornecer o ambiente e recursos necessários para esta conquista acadêmica.

À nossa orientadora, Adani Cusin Sacilotti, por suas valiosas considerações, paciência e apoio contínuo que moldaram este trabalho.

Aos colegas e amigos que apoiaram nossa pesquisa, agradeço por suas discussões estimulantes e retorno construtivo.

As nossas famílias e amigos, nosso profundo agradecimento pelo amor e apoio constantes ao longo dessa jornada.

Este trabalho não teria sido possível sem a colaboração de todos vocês. Obrigado por tornarem este momento possível.

Com gratidão,

Natália Maika de Toledo

Rhuan Ferreira Nascimento

Thays Aleksa Afonso

Vanessa Barros Barbosa da Silva

A ignorância do outro é a minha morte no futuro

Rita Von Hunty

## RESUMO

TOLEDO, Natália e NASCIMENTO, Rhuan e ALEKSA, Thays e BARROS, Vanessa. **BAFAFÁ – ANÁLISE DE SENTIMENTOS DE COMENTÁRIOS EM REDES SOCIAIS COMO FERRAMENTA DE SUPORTE A DECISÃO (SAD)**. 80 p. Trabalho de Conclusão de Curso de Tecnólogo em Gestão da Tecnologia da Informação. Faculdade de Tecnologia de Jundiaí - “Deputado Ary Fossen”. Centro Estadual de Educação Tecnológica Paula Souza. Jundiaí. 2024.

Diariamente, produzimos uma quantidade significativa de dados por meio das mídias sociais. Esses dados, quando usados adequadamente, podem nos dar uma vantagem competitiva sobre nossos concorrentes através de informações valiosas que lá estão. Este trabalho de conclusão de curso enfatiza a importância da inovação nos setores organizacionais em um ambiente competitivo e em constante crescimento. A opinião do cliente, seja ela expressa como crítica ou elogio, influencia diretamente o desempenho financeiro de uma empresa. Isso ocorre porque estamos sempre conectados às mídias sociais e, muitas vezes, baseamos nossas compras em recomendações de pessoas próximas, pessoas conhecidas ou influenciadores digitais. Nesse contexto, o Bafafá surgiu como um *software* de sistema de apoio à decisão (SAD) capaz de analisar, interpretar e sentimentalizar comentários por meio de análise lexical e técnicas de aprendizado de máquina. O principal objetivo do sistema é ajudar gerentes, diretores e executivos de diversos setores de negócios no Brasil a tomar decisões estratégicas mais fundamentadas e informa também informações que podem ser utilizadas no futuro, através de um histórico que pode ser adicionado em um banco de dados. Em suma uma ferramenta que pode ser utilizado de forma OLAP e OLTP (informações de histórico e informações atualizadas).

**Palavras-chave:** Redes Sociais. Análise de Sentimentos. Aprendizado de Máquina. Apoio à Decisão.

## ABSTRACT

TOLEDO, Natália e NASCIMENTO, Rhuan e ALEKSA, Thays e BARROS, Vanessa. **BAFAFÁ – SENTIMENT ANALYSIS OF COMMENTS ON SOCIAL NETWORKS AS A DECISION SUPPORT TOOL (DSS)**. 80 p. End-of-course paper in Technologist Degree in Information Technology Management. Faculdade de Tecnologia de Jundiaí - “Deputado Ary Fossen”. Centro Estadual de Educação Tecnológica Paula Souza. Jundiaí. 2024.

On a daily basis, a considerable amount of data is produced through social media. When used in an appropriate manner, this data can provide a competitive advantage over competitors by offering valuable insights. This term paper emphasizes the significance of innovation in organizational sectors in a competitive and constantly evolving environment. Customer opinion, whether expressed as criticism or praise, directly affects a company's financial performance. This is because we are always connected to social media and often base our purchases on recommendations from people close to us, people we know, or digital influencers. In this context, Bafafá has emerged as a decision support system (DSS) software capable of analyzing, interpreting, and sentimentalizing comments through lexical analysis and machine learning techniques. The system's primary objective is to assist managers, directors, and executives from various business sectors in Brazil in making more informed strategic decisions. Additionally, it provides information that can be utilized in the future, through the addition of historical data to a database. In essence, it is a tool that can be employed in both OLAP and OLTP (historical and updated information).

**Keywords:** Social Networks. Sentiment Analysis. Machine Learning. Decision Support.



## LISTA DE ILUSTRAÇÕES

Figura 1. Processo de aprendizagem do Aprendizado de Máquina.....	19
Figura 2. Diagrama do Banco de Dados .....	35
Figura 3. Diagrama de fluxo de dados do Bafafá.....	37
Figura 4. Diagrama de estado para dados de redes sociais .....	44
Figura 5. Diagrama de estado para previsão de vendas .....	44
Figura 6. Diagrama de estado para análise de sentimentos .....	45
Figura 7. Diagrama entidade de relacionamento .....	46
Figura 8. Diagrama de classe do Bafafá .....	49
Figura 9. Diagrama de sequência de análise de sentimentos.....	51
Figura 10. Diagrama de comunicação do Bafafá .....	52
Figura 11. Tela inicial do Visual Studio Code.....	60
Figura 12. Tela de importação do Visual Studio Code.....	61
Figura 13. Tela código fonte do X do Bafafá .....	62
Figura 14. Como informar a palavra-chave para pesquisa.....	62
Figura 15. Como executar a pesquisa da palavra-chave na API do X.....	63
Figura 16. Arquivo .xlsx dos comentários coletados no X .....	64
Figura 17. Tela código fonte do Instagram do Bafafá .....	64
Figura 18. Como informar a palavra-chave para pesquisa.....	65
Figura 19. Como executar a pesquisa da palavra-chave na API do Instagram.....	66
Figura 20. Arquivo .xlsx dos comentários coletados no Instagram .....	66
Figura 21. Tela código fonte do VADER do Bafafá.....	67
Figura 22. Como executar a sentimentalização dos comentários.....	68
Figura 23. Arquivo .csv com os comentários tratados .....	68
Figura 24. Arquivo Power BI do Bafafá .....	69
Figura 25. Como atualizar os dados do Power BI do Bafafá.....	70
Figura 26. Resultado visual dos comentários sentimentalizados.....	71

## LISTA DE TABELAS

Tabela 1. Modelo da estruturação com nome/total de colunas dos <i>data frames</i> por palavra-chave.....	23
Tabela 2. Entidade: dados de redes sociais.....	40
Tabela 3. Tabela final após aplicação do VADER .....	41
Tabela 4. Caso de uso .....	48

## LISTA DE ABREVIATURAS E SIGLAS

AL	Análise Léxica
AM	Aprendizado de Máquina
API	Application Programming Interface
BD	Banco de dados
DD	Dicionário de Dados
DER	Diagrama Entidade Relacionamento
DFD	Diagrama de Fluxo de Dados
DS	Diagrama de Sequência
FATEC	Faculdade de Tecnologia
IA	Inteligência Artificial
IoT	Internet das Coisas
LGPD	Lei Geral de Proteção de Dados
PLN	Processamento de Linguagem Natural
PYPI	Python Package Index
RD	Raspagem de dados
SAD	Sistema de Apoio à Decisão
SQL	Structured Query Language
SSMS	SQL Server Management Studio
TCC	Trabalho de Conclusão de Curso
TG	Trabalho de Graduação
VADER	Valence Aware Dictionary and Sentiment Reasoner
VSC	Visual Studio Code
WS	Web Scraping
X	Antigo Twitter

# SUMÁRIO

1.	INTRODUÇÃO .....	14
1.1.	Análise de Sentimentos.....	16
1.2.	Análise Léxica .....	17
1.3.	Análise de Sentimentos por Aprendizado de Máquina.....	18
2.	ESPECIFICAÇÃO DO SOFTWARE .....	19
2.1.	Escopo .....	21
2.2.	Entrada .....	21
2.3.	Coleta de Dados .....	21
2.4.	Armazenamento dos Dados.....	22
2.5.	Pré-processamento da Base de Dados Final.....	23
2.6.	Processamento .....	24
2.7.	Remoção de Pontuação e Acentuação.....	24
2.8.	Lematização.....	24
2.9.	Padronização do Texto.....	24
2.10.	Sentimentalização dos Comentários .....	25
2.11.	Dicionário Léxico .....	25
2.12.	Ressentimentalização dos Comentários .....	26
3.	SAÍDAS.....	26
3.1.	Base de Dados e Saída de dados.....	26
3.2.	Visão Front-end .....	27
3.3.	Clientes do Software.....	27
4.	REQUISITOS DO SISTEMA .....	28
4.1.	Requisitos Funcionais.....	29
4.2.	Requisitos Não Funcionais .....	30
4.3.	Requisitos de Domínio .....	32
5.	DEFINIÇÃO DO PROJETO.....	33
5.1.	Linguagem de Programação.....	34
5.2.	Banco de Dados .....	34
5.3.	Ferramentas de Auxílio .....	36
5.4.	Diagramas de Fluxo de Dados.....	36
5.5.	Dicionário de Dados .....	39
5.6.	Descrição Lógica dos Processos.....	42
5.7.	Diagramas de Transição de Estado .....	43

5.8.	Diagramas Estruturais.....	45
5.9.	Diagrama Entidade-Relacionamento .....	46
5.10.	Caso de uso:.....	47
5.11.	Diagramas de Classe .....	49
5.12.	Diagramas de Sequência .....	50
5.13.	Diagramas de Comunicação.....	52
6.	ARQUITETURA DA SOLUÇÃO.....	53
7.	CONSIDERAÇÕES FINAIS .....	53
	REFERÊNCIAS.....	55
	APÊNDICE A – MANUAL DO USUÁRIO .....	60
1.	UTILIZAÇÃO E MANUSEIO DO SOFTWARE .....	60
1.1.	Iniciando o software Bafafá no Visual Studio Code.....	60
1.2.	Alterando a palavra-chave a ser pesquisada:.....	61
1.3.	Executando a pesquisa da palavra-chave nas redes sociais: .....	63
2.	AGRUPANDO OS ARQUIVOS EXPORTADOS.....	67
3.	VISUALIZAÇÃO – POWER BI.....	69
	APÊNDICE B – CÓDIGO FONTE DO INSTAGRAM.....	72
	APÊNDICE C – CÓDIGO FONTE DO X .....	75
	APÊNDICE D – CÓDIGO FONTE DO VADER.....	77

## 1. INTRODUÇÃO

Devido a intensa concorrência de mercado, inovar é fundamental. Até o ano de 2021 muito se falava da “Era dos Dados”, da Internet das Coisas (IoT) e de como a conectividade está mudando o mundo dos negócios para melhor (Rebouças, 2021). Atualmente o foco tem se voltado para tecnologias como Aprendizado de Máquina (em inglês *Machine Learning*), Inteligência Artificial (IA), e tantas outras que empregam o uso de dados estruturados e não estruturados (*Big Data*) para atender a necessidade das empresas em inovar (Borovkova, 2020).

Independente das tecnologias aplicadas, a realidade das empresas em 2023 é diferente dos últimos anos. A inovação, bem como qualquer outra iniciativa no âmbito empresarial depende da gestão eficiente das corporações, sejam elas públicas ou privadas. A gestão eficiente por sua vez é dependente de planejamento, pois as decisões estratégicas ou operacionais de uma empresa requerem uma exploração constante do relacionamento entre os elementos que compõe a realidade em que a empresa está inserida (Passari, 2003), ou seja, a corporação e todas as variáveis internas e externas a mesma.

Para auxiliar no apoio de decisões corporativas, as empresas têm investido cada dia mais em seu *Digital Branding*, através de estratégias, táticas e ferramentas para impulsionar o seu negócio na era digital. Com o *Big Data* e a facilidade de acesso à informação assim como o poder das redes sociais se torna inquestionável.

Segundo Klein, Guidi Neto e Tezza (2017), a cada minuto, milhares de pessoas ao redor do mundo compartilham publicamente grande volume de informações tais como comentários, pensamentos, críticas, elogios e suas mais diferentes opiniões a respeito de notícias, marcas e produtos.

A correlação entre os rumores nas mídias sociais e o mundo real tem sido comprovada em diferentes países (Hempel, 2014). De acordo com Antunes *et al.* (2014), um dos pontos fortes do monitoramento de redes sociais é fornecer uma grande quantidade de informação, proveniente de diversificados números de fontes.

Desta forma, entende-se que as redes sociais podem desempenhar um papel muito importante para o negócio e uma gestão eficiente, a exemplo, ao corroborar com dados quase que em tempo real sobre a opinião de consumidores. Ignorar ou

subestimar a importância das redes sociais, neste contexto, seria um erro estratégico grave, pois, as empresas que compreenderem a relevância destes canais de comunicação, terão a oportunidade de se conectar de maneira mais profunda e genuína com seu público-alvo.

A justificativa deste trabalho se fundamenta na necessidade que as empresas têm de inovar e conseqüentemente, faturar, em um mercado cada vez mais digitalizado e competitivo. Tecnicamente, a ferramenta surge com a capacidade de “dar sentimento” aos comentários coletados nas redes sociais através de técnicas de Processamento de Linguagem Natural (PLN) e aprendizado de máquina. Segundo Khan e Das (2018, p. 6, tradução própria): "A transição de problemas teóricos de aprendizado de máquina para implementação prática ajudou empresas de internet a aproveitar o aprendizado de máquina para expandir seus negócios".

O *software* pode ser compreendido como um Sistema de Apoio à Decisão (SAD) que tem como objetivo principal oferecer percepções valiosas as partes interessadas que podem direcionar melhor as estratégias de *marketing*, desenvolvimento de novos produtos, atendimento ao cliente entre etc.

Desta forma, o sistema parte da hipótese, que a sua utilização será de uma ferramenta para dar sentimento à comentários de redes sociais. Seu uso ajuda a melhorar a capacidade das empresas na tomada de decisões estratégicas, aproveitando os sentimentos e opiniões dos consumidores em relação às marcas, produtos entre outros, o que, por sua vez, pode impulsionar a sua vantagem competitiva em um ambiente de mercado digitalizado e altamente competitivo.

Os autores deste trabalho acabaram de escrever sobre a introdução e nesta mesma divisão abordarão a análise de sentimentos e análise léxica. No capítulo 2 abordaremos a especificação do *software* (escopo, entrada) coleta de dados, armazenamento dos dados, processamento da base de dados, remoção de pontuação e acentuação, lematização, padronização de texto, sentimentalização, dicionário léxico e ressentimentalização de comentários. Já no capítulo 3 será abordada a base de dados/saída de dados, visão *front-end* e clientes do *software*. E no subsequente capítulo, os requisitos do sistema, requisitos funcionais, requisitos não funcionais, e requisitos de domínio. Na divisão 5 serão abordados os tópicos definição do projeto, linguagem de programação, banco de dados, ferramentas de auxílio, diagrama de fluxo de dados, dicionário de dados, descrição lógica do

processo, transição de estado, entidade de relacionamento, caso de uso, diagrama de classe, diagrama de comunicação, no capítulo 6 será abordada a arquitetura da solução e, por fim, as considerações finais.

Na próxima seção será exposta a análise de sentimentos.

## 1.1. Análise de Sentimentos

A análise de sentimentos nas mídias sociais tem como objetivo extrair textos, comentários, pensamentos, opiniões, publicados na Internet a cerca de um determinado assunto, evento, produto entre outros, classificando e quantificando as opiniões em valores positivos, negativos ou neutros (Rosa, 2015).

A comunicação na internet tem como característica principal a informalidade, com uso de gírias e várias emoções expressadas por meio de emojis, símbolos, desenhos e memes. Esta característica torna a classificação de opiniões em textos extraídos das redes sociais ainda mais complexa.

Segundo Benevenuto, Ribeiro e Araújo (2015, p. 32):

O principal objetivo da análise de sentimentos é definir técnicas automáticas capazes de extrair informações subjetivas em textos em linguagem natural, como opiniões e sentimentos, a fim de criar conhecimento estruturado que possa ser utilizado por um sistema de apoio ou tomador de decisão.

A palavra *sentimento*<sup>1</sup>, conforme o dicionário *Oxford Languages*, é o estado ou condição psicológica, e suas manifestações, originadas das pulsões de afeto ou aversão. Esses sentimentos podem expressar o grau de satisfação ou insatisfação de um usuário perante um produto, serviço ou evento (Rosa, 2015).

De maneira geral, a análise de sentimentos pode ser realizada por uma abordagem léxica, ou seja, um acervo (dicionário) de palavras portadoras de sentido; ou através de técnicas de Aprendizado de Máquina (AM). Segundo Yoon *et al.* (2013, *apud* Oliveira, 2019, p. 240): "A análise de sentimento é um subcampo do

---

Mais informações consultar:

<sup>1</sup> SENTIMENTO. *In*: Oxford Languages. Google, c2023. Disponível em: <https://www.google.com/search?client=opera-gx&q=sentimento&sourceid=opera&ie=UTF-8&oe=UTF-8>. Acesso em: 08 set. 2023.



processamento de linguagem natural voltada a extrair, classificar e analisar opiniões sobre diversos temas em grandes volumes de dados textuais".

O tema seguinte discorre sobre a análise léxica.

## 1.2. Análise Léxica

Na análise de sentimentos pela abordagem léxica, os vocábulos são classificados através de uma base de dados que podem ser construídas de forma orgânica ou automatizada com uma lista de palavras pré-definidas em *softwares* já existentes, como por exemplo SentiLex-PT<sup>2</sup> (Carvalho; Silva 2015), PropBank.Br<sup>3</sup>Duran; Aluísio, 2012) e VADER<sup>4</sup> (Hutto; Gilbert 2014) para a língua portuguesa. E o SentiWordNet<sup>5</sup> no caso da língua inglesa, sendo esse, o *software* mais utilizado nesse idioma (Baccianella; Esuli; Sebastiani, 2010).

Naturalmente a abordagem clássica de análise de sentimentos é baseada no uso de análises léxicas, isto é, utiliza-se dicionários léxicos com palavras previamente escolhidas e suas relativas polaridades.

Desta forma, podemos destacar o processo realizado na ferramenta LIWC<sup>6</sup>: um *software* de dicionário léxico que, nesse estudo, se baseia na classificação das palavras de acordo com categorias “psicologicamente significativas”. Dentre uma ampla gama de categorias e subcategorias, as mais utilizadas são: Emoções Positivas (feliz, bonito, bom), Emoções Negativas (ódio, inimigo, raiva), Ansiedade (nervoso, medo, tenso) (Tausczik; Pennebaker 2009).

Tratado os tópicos anteriores, será abordada a análise de sentimentos por aprendizado de máquina no capítulo a seguir.

---

Mais informações consultar:

<sup>2</sup> SENTILEX-PT. Disponível em: <https://github.com/sillasgonzaga/lexiconPT>. Acesso em: 29 set. 2023.

<sup>3</sup> PROPBANK.BR. Disponível em: <http://143.107.183.175:21380/portlex/index.php/pt/downloads>. Acesso em: 29 set. 2023.

<sup>4</sup> VADER. Disponível em: <https://github.com/cjhutto/VADERSentiment>. Acesso em: 29 set. 2023.

<sup>5</sup> SENTIWORDNET. Disponível em: <https://github.com/aesuli/SentiWordNet>. Acesso em: 29 set. 2023.

<sup>6</sup> LIWC. Disponível em: <https://www.liwc.app/>. Acesso em: 29 set. 2023.

### 1.3. Análise de Sentimentos por Aprendizado de Máquina

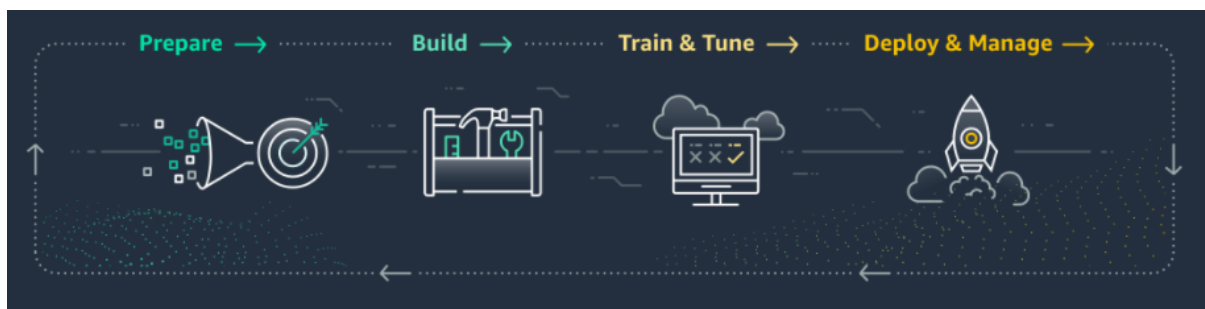
Aprendizado de Máquina é uma área da ciência da computação que se concentra em criar algoritmos e modelos estatísticos para fazer com que os computadores executem tarefas sem precisar de instruções diretas. Eles fazem isso procurando padrões e fazendo suposições a partir dos dados. Os computadores usam esses algoritmos para analisar grandes quantidades de informações do passado e encontrar padrões nos dados. Isso ajuda a fazer previsões mais precisas com base em novos dados.

Segundo Matos (2021, p. 21): “Os algoritmos de *Machine Learning* encontram padrões naturais nos dados que geram conhecimento e ajudam a calcular previsões e tomar melhores decisões”.

Por exemplo, pode-se treinar um programa de computador para diagnosticar câncer através imagens de raios-x. Para fazer isso, os cientistas de dados utilizam milhões de imagens digitalizadas e os diagnósticos correspondentes para ensinar o computador a reconhecer sinais de câncer. A compreensão do processo de aprendizagem da máquina pode ser observada na figura 1.

O AM é uma ferramenta que ajuda as empresas a crescerem, encontrar novas maneiras de ganhar dinheiro e resolver problemas difíceis. Para fazer isso, ele usa dados, que são como o combustível para a tomada de decisões corporativas. No passado, costumavam reunir dados de muitas fontes, como o retorno/ponto de vista dos clientes, informações de funcionários e dados financeiros. Esta ferramenta automatiza a busca e melhora significativamente esse processo. Isso significa que as empresas podem usar um tipo de *software* especial que examina uma grande quantidade de dados muito rapidamente, permitindo que elas obtenham resultados mais rápidos e tomem decisões de negócios mais informadas e assertivas (AWS, 2023).

Figura 1. Processo de aprendizagem do Aprendizado de Máquina



Fonte: AWS Amazon, 2023. Disponível em: <https://aws.amazon.com/pt/what-is/machine-learning/>.

Acesso em: 09 out. 2023.

No contexto da ciência de dados, o AM desempenha um papel fundamental, treinando algoritmos para classificar e prever com base em métodos estatísticos. Existem dois principais tipos de aprendizado por máquinas: a Aprendizagem Profunda e AM clássico.

Aprendizagem Profunda (em inglês *Deep Learning*) é mais automatizado, aprendendo com dados não estruturados e identificando automaticamente características distintivas. Por outro lado, o AM clássico depende mais da intervenção humana para definir essas características. O AM opera com três componentes principais: um processo de decisão, uma função de erro e um processo de otimização. Os modelos de AM se enquadram em três categorias: aprendizado supervisionado, não supervisionado e semissupervisionado (IBM, 2023).

No supervisionado utiliza conjuntos de dados rotulados para prever com precisão – no caso – o VADER, enquanto o não supervisionado analisa conjuntos de dados não rotulados para encontrar padrões ocultos e, por fim, o semi supervisionado combina elementos dos dois para lidar com conjuntos de dados com poucos rótulos.

Dados os temas abordados, em seguida serão discutidas as especificações do *software* Bafafá.

## 2. ESPECIFICAÇÃO DO SOFTWARE

Com o objetivo de apresentar as características e funcionalidades do *software*, detalhamos as especificações que definem a excelência e a versatilidade do mesmo são elas:

- **Coleta de Dados:** o Bafafá realizará a coleta de dados via API do termo em inglês *Application Programming Interface* (em português Interface de Programação de Aplicação) e Raspagem de Dados (em inglês *Web Scraping - WS*). Será capaz de coletar comentários e postagens de várias redes sociais como X (antigo Twitter), Instagram, dentre outros;
- **Pré-processamento de Texto:** a ferramenta realizará a limpeza e o pré-processamento dos dados coletados, incluindo a remoção de caracteres especiais, *stopwords* e normalização de texto;
- **Análise de Sentimento:** o *software* é capaz de determinar o sentimento geral (positivo, negativo ou neutro) de cada comentário;
- **Geração de Relatórios:** o sistema pode criar relatórios detalhados e visualizações de dados para apresentar os resultados da análise;
- **Alertas em Tempo Real:** o Bafafá pode ser configurado para enviar alertas em tempo real quando for detectada uma mudança significativa no sentimento em relação a uma marca ou produto;
- **Segurança e Privacidade:** a ferramenta foi construída em conformidade com a Lei Geral de Proteção de Dados (LGPD) e garante segurança dos dados do cliente além de cumprir com as regulamentações de privacidade restrita apenas no que é possível extrair com as APIs;
- **Linguagem de Programação:** o Bafafá foi desenvolvido na linguagem de programação Python devido sua ampla adoção em análise de texto e PLN;
- **Plataformas Suportadas:** foi projetado para ser compatível com várias plataformas, incluindo Windows, macOS e Linux. Isso visa garantir maior flexibilidade na implementação.

No capítulo subsequente, será detalhado o escopo do *software*.

## 2.1. Escopo

O escopo serve para definir as funcionalidades, objetivos, atividades, requisitos e os recursos que o sistema incluirá. Basicamente, este processo visa delimitar quais resultados esperam-se alcançar com a ferramenta.

Os recursos do *software* serão detalhados abaixo.

## 2.2. Entrada

As entradas de um *software* são os dados, informações ou comando fornecidos ao programa durante a sua execução.

O processo do Bafafá se inicia com uma base de dados para se trabalhar. Neste contexto, o programa primeiro realiza a busca dos dados que serão utilizados como entrada.

No próximo assunto, será apresentado o tema coleta de dados.

## 2.3. Coleta de Dados

No *software* desenvolvido pelos autores o processo de coleta de dados é realizado de duas formas, através de técnicas de busca orgânica (em inglês *Web Screaming*) e por meio de solicitações via API.

As técnicas de WS podem ser também chamadas de “raspagem de dados” feita diretamente da web (Barbosa; Cavalcanti, 2020). Tais informações não poderiam ser coletadas manualmente por conta de seu alto volume, desta maneira, automatizamos a coleta com base em palavras chaves definidas previamente pelo usuário e que contextualize a empresa, a marca ou o produto que se pretende avaliar.

No Bafafá, a coleta de dados utilizando técnicas de RD, foi realizada por meio da biblioteca *snsrape*<sup>7</sup> disponível em Python. A biblioteca se autointitula um *scraper* para redes sociais capaz de coletar dados como perfis de usuários, hashtags,

---

Mais informações consultar:

<sup>7</sup> SNSCRAPE. Disponível em: <https://pypi.org/project/snsrape/>. Acesso em: 29 set. 2023.

comentários, posts relevantes (PyPi, 2023) e no Bafafá foi utilizada para raspar dados do Instagram e X.

A RD do X também foi feita por meio de requisições via API. Essa abordagem foi necessária pois esta rede social bloqueia as requisições feitas pelo *snsrape* como forma de proteção de dados e a este tipo de técnica. Todavia, o X permite a coleta de dados da rede social via requisições à API oficial, para tal foi necessária a criação de uma conta no ambiente para desenvolvedores, onde nos foi disponibilizado tokens e chaves para acesso à API via programação *back-end*. Os *scripts* Python criados para requerer estes dados, foram construídos em um ambiente de nuvem denominado *DataBricks*.

É importante pontuar que o X limita o número de informações que podem ser requeridas à API no plano básico disponibilizado pela empresa. Para conseguir acessar mais dados é necessário aderir a algum dos planos pagos disponibilizados.

A seguir, detalharemos o armazenamento de dados.

## 2.4. Armazenamento dos Dados

Os *scripts* Python utilizando a biblioteca *snsrape* e os de coleta via API, nos permite armazenar os dados coletados em variáveis no ambiente de desenvolvimento *Visual Studio Code* (VSC) por palavra-chave consultada.

Por exemplo, digamos que o *software* seja utilizado para acompanhar o desempenho de uma determinada marca de biscoitos nas redes sociais. A palavra-chave utilizada para as buscas foi “VINHO”, então o *software* armazenará todas os comentários coletados em uma variável chamada (VINHO), que nada mais é do que um *data frame* criado via Pandas<sup>8</sup> para aquela palavra-chave.

Várias palavras-chave podem ser definidas pelo usuário. Ao fim de todas as coletas o sistema concatena todos os *data frames* criados em uma nova variável única. Para isso é necessário que todos os *data frames* criados sigam a mesma

---

Mais informações consultar:

<sup>8</sup> PANDAS. Disponível em: <https://pandas.pydata.org>. Acesso em: 27 mai. 2024.

padronização no nome e na quantidade de colunas, conforme exemplo apresentado na tabela 1:

Tabela 1. Modelo da estruturação com nome/total de colunas dos *data frames* por palavra-chave

SLNO	TEXT	SENTIMENT	VARIETY	COUNTRY
99999	XXXXX	XXXXXX	XXXXXX	XXXXX

Fonte: elaborado pelos autores (2023).

Após a concatenação dos *data frames* a variável é convertida para um formato onde possa ser acessada pela extensão Pandas que viabiliza o uso de arquivos em Excel no código.

No próximo item, descreveremos sobre o pré-processamento da base de dados final.

## 2.5. Pré-processamento da Base de Dados Final

O pré-processamento da base final é uma etapa crucial e anterior a análise de sentimentos. Esta etapa se torna necessária pois ao se coletar os dados das redes sociais, coletamos também muitas informações irrelevantes, isto é, textos que citam a palavra-chave pesquisada, mas que não necessariamente tem relação com o que se deseja analisar. Os pré-processamentos realizados na base final do *software* obtido pelos autores, foram:

- **Remoção de informações irrelevantes:** para a sentimentalização dos comentários as colunas de texto que não são necessárias e podem ser descartadas;
- **Remoção de linhas em branco:** o código faz análises buscando e excluindo linhas em branco, e substituindo letras maiúsculas por minúsculas de todo o arquivo que será sentimentalizado.

O processamento dos dados será explicado no próximo tópico.

## **2.6. Processamento**

A etapa de processamento dos dados segue a importância das etapas anteriores. Nesta operação foram criadas funções em Python que atuam na base de dados do sistema realizando os mais diversos tipos de transformações nos textos coletados.

As funções aplicadas à base final são apresentadas a seguir.

## **2.7. Remoção de Pontuação e Acentuação**

A presença de caracteres de pontuação como ponto, ponto e vírgula, vírgula, e a acentuação de palavras no geral podem dificultar o processo de sentimentalização do texto. Para garantir a uniformização do texto, simplificação das análises e redução de ruídos aplicamos esta função de processamento.

Em seguida, será abordado sobre o tema lematização.

## **2.8. Lematização**

A lematização converte palavras para a sua forma raiz. Por exemplo, ao encontrar a palavra sou no contexto de alguma frase, a aplicação irá automaticamente converter a palavra para o seu radical, neste caso, ser.

Será retratado a seguir a padronização do texto.

## **2.9. Padronização do Texto**

A aplicação padroniza os comentários salvos na base de dados, convertendo o texto dos comentários para caixa baixa (letras minúsculas).

A fio será exemplificado a sentimentalização dos comentários.



## 2.10. Sentimentalização dos Comentários

A análise de sentimentos é o processo em que o Bafafá dá sentimento a um comentário da base de dados. Para tal, a aplicação utiliza o *Valence Aware Dictionary and Sentiment Reasoner* (VADER) uma biblioteca popular para análise de sentimento em linguagem natural.

O VADER se baseia em técnicas de AM e em dicionários que atribuem pontuações de sentimento a textos com base nas palavras e suas polaridades associadas. Esta biblioteca além de atribuir polaridades (positiva, negativa, neutra), também considera a intensidade do sentimento e valência, o que permite uma análise mais granular.

A biblioteca se sai bem na análise de comentários de redes sociais, onde gírias e sarcasmo são comuns, pois entende bem as nuances do texto.

O Bafafá aplica a biblioteca VADER na base de dados processada para atribuir sentimentos positivos, negativos ou neutros para cada frase.

De acordo com Silva e Serrano (2023, p. 7):

A vantagem de se usar inteligência artificial para detectar sentimentos é a capacidade de analisar os dados com mais velocidade e de forma automatizada, tornando o processo mais simples e tendo um maior aproveitamento do tempo.

Continuamente será abordado o tema dicionário léxico.

## 2.11. Dicionário Léxico

Modelos de AM e classificadores por dicionários não são perfeitos e podem errar. Como forma de garantir uma classificação de sentimentos mais assertiva, o Bafafá conta com um dicionário léxico em português que dá suporte as bibliotecas e ferramentas utilizadas para a sentimentalização.

A aplicação utiliza este dicionário para contextualizar o VADER com a realidade específica dos dados que estão sendo imputados. Por exemplo, se o *software* estiver sendo utilizado por uma empresa que atue na área de alimentos e bebidas, o dicionário léxico será criado com base em palavras que façam sentido para este contexto empresarial.

O próximo capítulo tangerá a ressentimentalização dos comentários.

## 2.12. Ressentimentalização dos Comentários

Dado que os modelos classificadores podem errar e que o dicionário léxico é criado pela ferramenta para melhor contextualização da realidade dos dados, o Bafafá propõe automaticamente uma ressentimentalização dos comentários com base em divergências que possam ser encontradas.

A lógica programada funciona da seguinte maneira, suponha que a biblioteca VADER classifique um comentário como *negativo*, mas o dicionário léxico entende que no contexto dos dados que estou analisando este mesmo comentário seria classificado como *neutro*, então uma reclassificação é proposta para aquele comentário em específico. Se não há entendimentos diferentes entre a classificação realizada pelo VADER e o contexto dos dados através do dicionário léxico, nenhuma reclassificação é proposta.

No capítulo a seguir, serão abordadas as saídas do *software*.

## 3. SAÍDAS

### 3.1. Base de Dados e Saída de dados

Após todos os passos anteriormente mencionados o que obtemos como saída de todo o processo é uma base de dados final processada, sem pontuações, sem acentos, lematizada, padronizada, tokenizada e sentimentalizada.

Nossa saída será um arquivo em .CSV contendo todos os comentários extraídos das redes sociais, sentimentalizados e prontos para serem adicionados a uma ferramenta de *Business Intelligence* nesse caso o Power BI.

Esta tabela é posteriormente utilizada pela aplicação como fonte de dados para a construção de uma visão *front-end* onde gestores e líderes tem acesso de forma visual (gráficos dinâmicos, *big numbers*, *banners* informativos) aos dados analisados. A tabela utilizada para a visão é periodicamente atualizada, de acordo com o que está

acontecendo no momento nas redes sociais. A periodicidade de atualização destes dados é de suma importância, pois pode significar maior agilidade na tomada de decisão.

Na próxima seção, será explicado sobre a visão *front-end* do *software*.

### 3.2. Visão Front-end

O *software* Bafafá desempenha uma função pivotal ao conceber um *dashboard*, que se configura como a interface principal da aplicação. Este *dashboard* é desenvolvido utilizando a plataforma Power BI, onde indivíduos responsáveis pela tomada de decisões podem analisar visualmente dados por meio de gráficos e monitorar os comentários que circulam nas redes sociais relacionadas à marca, produto e outros temas pertinentes.

Os mencionados *dashboards* desempenham a tarefa de elaborar relatórios de percepções com base em eventos ocorridos em países e marcas. Caso surjam tendências ou comentários significativos relacionados ao tópico em discussão nas redes sociais, o sistema automaticamente emite alertas para os gestores responsáveis, proporcionando uma resposta ágil às dinâmicas do ambiente digital.

No tópico seguinte será referido os clientes do *software*.

### 3.3. Clientes do Software

O *software* possibilita atender aos clientes dos mais diversos ramos do mercado, onde cada empresa consegue adaptar o *software* de acordo com as suas necessidades. Temos como ilustração os seguintes comércios:

- I. Alimentos e Bebidas;
- II. Cosméticos;
- III. Finanças;
- IV. Têxtil;
- V. Transportes;
- VI. Turismo (Companhias Aéreas, Ônibus, Hotelaria);

- VII. Artigos de Papelaria;
- VIII. Universidades;
- IX. Negócios Digitais.

Nosso foco principal está em atender às necessidades de pequenas e médias empresas que tenham como cultura a mentalidade de olhar voltado para o futuro. Essa ferramenta pode ser uma valiosa aliada, capacitando as empresas a ajustarem-se às exigências dos clientes, elevando a satisfação e otimizando suas operações e estratégias de vendas.

Se houver um aumento nas avaliações positivas, a empresa consegue prever um aumento nas vendas e ajustar sua produção e estoque de acordo. Por outro lado, se houver algum retorno negativo, a empresa pode tomar medidas proativas para mitigar possíveis perdas, essa capacidade é viabilizada pelo Bafafá, que opera em tempo real para atender às demandas e solucionar problemas antes que se agravem, a monitorização do sentimento do cliente desempenha um papel crucial para proteger a reputação da marca.

Segundo Palkovsky (2023): “Afiml, por trás de cada informação, existe uma narrativa relacionada a comportamentos, preferências e experiências”.

Em seguida será explanado os requisitos de sistema do Bafafá.

## 4. REQUISITOS DO SISTEMA

O dicionário *Oxford Languages* define a palavra *requisito*<sup>9</sup> como a condição para se alcançar determinado fim. Um requisito equivale a um atributo essencial de um *software*, sendo este, crucial para a resolução de um problema do usuário com o intuito de atingir um objetivo específico (Thayer; Dorfmann, 2000). Para Siddiqi e Shekaran (1996) os requisitos são uma declaração de tudo o que um *software* se propõe a fazer, entretanto, sem abordar os detalhes de como será feito.

---

Mais informações consultar:

<sup>9</sup> REQUISITO. *In*: Oxford Languages. Google, c2023. Disponível em:

<https://www.google.com/search?client=opera-gx&q=requisito&sourceid=opera&ie=UTF-8&oe=UTF-8>.

Acesso em: 26 out. 2023.

Já no ponto de vista de Goguen e Linde (1993) os requisitos são especificidades que uma ferramenta precisa dispor para operar com sucesso no contexto em que esta for utilizada.

Para autores como Engholm Junior (2010), a elicitación de requisitos é a fase inicial de qualquer projeto, sendo de extrema importância para o seu sucesso. O levantamento dos requisitos do sistema não serve apenas para o planejamento correto, análise de viabilidade e estimativa de custos, mas também para que o cliente receba um sistema a altura de suas expectativas.

Nos que diz respeito aos requisitos de sistema de *software*, estes são classificadores como requisitos funcionais, não funcionais ou de domínio, e estes serão abordados a fio.

#### 4.1. Requisitos Funcionais

Os requisitos funcionais descrevem explicitamente as funcionalidades e serviços do sistema além de documentar como o sistema deve se comportar em casos de entradas específicas e o que o sistema não deve fazer (Sommerville, 2011). Em outras palavras, os requisitos funcionais podem ser definidos como as condições necessárias para a obtenção de certo objetivo.

Alguns dos requisitos funcionais do Bafafá são:

- **Coleta de Dados de Redes Sociais:** o Bafafá é capaz de extrair automaticamente dados relevantes de redes sociais específicas, tais quais postagens, comentários e menções;
- **Armazenamento dos Dados:** o *software* armazena em arquivos de Excel, para serem utilizados em ferramentas de BI, ou adição em um banco de dados.
- **Pré-processamento da Base de Dados:** consiste na remoção de informações como linhas em branco;
- **Processamento da Base de Dados:** remoção de *stopwords*, e padronizar palavras para minúsculo.
- **Análise de Sentimentos:** o Bafafá utiliza algoritmos de PLN, AM e metodologias léxicas para classificar o sentimento das mensagens (postagens,

comentários) em três categorias de sentimentos: positivos, negativos ou neutros;

- **Dicionário Léxico:** o Bafafá utiliza dicionários léxicos específicos do meio empresarial para contextualizar os algoritmos de AM sobre os dados que estão sendo imputados;
- **Produto:** base de dados única contendo todos os comentários pré-processados, processados e sentimentalizados.
- **Visão *front-end*:** *dashboard* desenvolvido em Power BI como interface principal de nosso aplicativo. Este *dashboard* é o ponto principal para percepções/visões e tomada de decisão dentro da aplicação;
- **Automação Programada:** o sistema é capaz de, automaticamente, coletar dados de redes sociais, processá-los, sentimentalizá-los e disponibilizá-los em sua totalidade ou qualquer atualização incremental, na visão *front-end*. As atualizações ocorrem automaticamente a medida com o que se observa em redes sociais, quase que em tempo real.

Subsequente serão abordados os requisitos não funcionais.

## 4.2. Requisitos Não Funcionais

Os Requisitos não funcionais vem sendo citados em processos de desenvolvimento de *software*, dentre outras formas como restrições e condições de contorno. Estão diretamente relacionados com as funções/serviços do *software*/sistema (Sommerville, 2011). Os requisitos não funcionais desempenham papel de criticidade durante o desenvolvimento de um *software*, e erros devido a não elicitação ou a elicitação incorreta destes estão entre os mais caros e difíceis de se corrigir, principalmente após a implantação do sistema (Brooks Junior, 1987).

Uma boa elicitação de requisitos não funcionais tentará responder três principais perguntas:

- I. Quais as necessidades dos usuários do sistema? O que eles esperam que o sistema faça?
- II. Quais são os objetivos da empresa com a implementação do *software*?

III. Quais são os limites do *software*? Quais os recursos e capacidades disponíveis?

Abaixo apresenta-se os requisitos não funcionais do Bafafá:

- **Desempenho:** o Bafafá é capaz de coletar, pré-processar, processar, analisar, sentimentalizar, produtizar as informações e por fim disponibilizá-las em um *dashboard*, em tempo quase que real, grandes volumes de dados e informações de redes sociais a respeito do desempenho de marcas e produtos. O tempo de resposta do *software* depende da velocidade da internet em uso, do tempo de leitura e gravação no banco de dados em nuvem, e nos acontecimentos e movimentação que ocorrem nas redes. Em condições normais de uso o *software* é automaticamente atualizado a cada uma hora, sendo possível a personalização desta *feature*;
- **Segurança:** o sistema atende a todos os requisitos de segurança da empresa antes de entrar em funcionamento. Os dados são coletados a partir do que a API das redes sociais proporciona, então só será captado informações públicas ali inseridas nas redes de acordo com suas permissões.
- **Disponibilidade:** o *software* está disponível 24 horas por dia, 7 dias por semana;
- **Usabilidade:** o Bafafá foi projetado para fácil uso, garantindo que os usuários consigam aproveitar ao máximo os recursos disponíveis, além de ser acessível à diferentes públicos;
- **Conformidade:** o Bafafá tem alto comprometimento com a segurança das informações coletadas, estando em conformidade com a LGPD de setembro de 2020. Os dados coletados no âmbito do *software* são beneficiados e exclusivamente utilizados para geração de percepções e entendimentos de mercado da empresa detentora da licença de uso;
- **Integração:** o Bafafá se integra à API das redes sociais para coletar as informações relevantes para cada marca e/ou produto.
- **Escalabilidade:** o sistema é capaz de escalar para atender às necessidades da empresa no futuro;

- **Manutenção:** o sistema é foi projetado e desenvolvido de maneira modular, com documentação clara para facilitar atualizações e manutenções futuras. As atualizações no Bafafá ocorrem de maneira periódica, visando sempre a melhoria contínua do sistema. São realizadas de maneira a minimizar o impacto nas operações;
- **Insights:** o Bafafá pode gerar percepções automáticas utilizando a IA disponível no Microsoft Power BI. Todavia, as principais visões serão geradas pela alta cúpula, entendedora e tomadora de decisões para o negócio;
- **Backup e Recuperação:** o processo de *backup* e recuperação de dados no Bafafá é simples e acontece de maneira integrada e periódica, conjuntamente ao sistema de *backup* da empresa ou descartado caso seja uma informação temporária.

Seguidamente serão expostos os requisitos de domínio do sistema.

### 4.3. Requisitos de Domínio

Os requisitos de domínio podem ser baseados em requisitos funcionais ou não funcionais e são derivados do domínio de aplicação e seus processos (Sommerville, 2011). Em termos mais simples, os requisitos de domínio podem ser compreendidos como requisitos do processo, do departamento e/ou da empresa cliente.

O levantamento dos requisitos de domínio para o Bafafá é uma etapa muito importante e muitas vezes muito difícil pelo fato dos requisitos de domínio serem expressos em uma linguagem específica do domínio da aplicação. Ou seja, por exemplo, em um sistema de Departamento de Recursos Humanos (RH), são utilizados termos e regras específicas desta área, em que muitas vezes o analista de requisitos não conhece, por não trabalhar nesta área e não ter a experiência que seu cliente tem neste segmento (Maretti, 2014).

Assim, os requisitos abaixo elencados representam os requisitos mínimos para implantação do Bafafá na empresa cliente. Requisitos de domínio mais complexos são adicionados de acordo com a especificidade do negócio onde o *software* será implantado.



- **Conhecimento de Redes Sociais:** os membros da equipe que utilizarão o Bafafá devem ter compreensão das principais plataformas de redes sociais utilizadas no *software* e de como elas funcionam (tipo de conteúdo, público e outras coisas mais). É importante que técnico consiga compreender minimamente como os usuários interagem nas redes, como as postagens são criadas e os tipos de dados que o *software* consegue extrair e processar;
- **Técnicas para Análise de Sentimentos:** é de grande importância que os membros utilizadores do *software* detenham conhecimento sobre o que é o PLN e como funciona, bem como a AL e como funciona. Os técnicos serão responsáveis pela calibragem do sistema, para tal imputarão palavras-chave específicas do negócio e seus respectivos sentimentos a um dicionário léxico;
- **Percepções de Negócio:** o Bafafá oferece visões de negócio gerados automaticamente com base em IA em relação ao que ele enxerga nas redes sociais definidas no escopo. Dito isto, o valor da ferramenta se dá com a inteligência de negócios dos tomadores de decisão, que serão os principais responsáveis por conectar as informações do SAD à outras informações da empresa, para chegar a melhor decisão para os negócios dela. Como um SAD, é altamente recomendado que o *software* não seja a única ferramenta utilizada no processo de tomada de decisão;
- **Ética e Privacidade:** ter uma compreensão sólida dos princípios da empresa, interpretar corretamente os resultados fornecidos pelo Bafafá, e não se apoiar unicamente neles para tomar decisões de negócio.

No próximo capítulo discorreremos sobre a definição do projeto.

## 5. DEFINIÇÃO DO PROJETO

Com base na avaliação de requisitos do projeto e na capacidade técnica e conhecimento dos desenvolvedores, definimos que o Bafafá será implantado em ambiente de nuvem, visando também a disponibilidade e a facilidade de acesso à ferramenta.

A seguir, será abordado sobre a linguagem de programação.

## 5.1. Linguagem de Programação

O Bafafá utiliza o Python como linguagem de programação principal. Python é uma linguagem de programação de propósito geral, podendo ser utilizada para análise de dados, construção de *dashboards*, automações até o desenvolvimento de *softwares*. Por ter esta característica, o uso da linguagem tem crescido entre as empresas (Lopes *et al.*, 2019).

Esta linguagem foi escolhida devido ao conhecimento prévio que a equipe de desenvolvimento possuía na mesma. Esta medida visou rapidez e agilidade no desenvolvimento.

O próximo conteúdo discorrerá sobre o banco de dados.

## 5.2. Banco de Dados

Um BD é uma coleção organizada de informações, ou dados, estruturados. Normalmente, esses dados são armazenados em servidores físicos ou na nuvem, de acordo com a necessidade e as políticas da empresa. Posto isso, a utilização de um BD, permite manter um grande conjunto de dados organizados por tabelas de acordo com sua classificação, facilitando assim, o acesso, gerenciamento e análise dessas informações.

No caso do *software* colocamos em uso o modelo *SQL Server Management Studio* (SSMS) já que este é um ambiente versátil e integrado para gerenciar qualquer infraestrutura SQL.

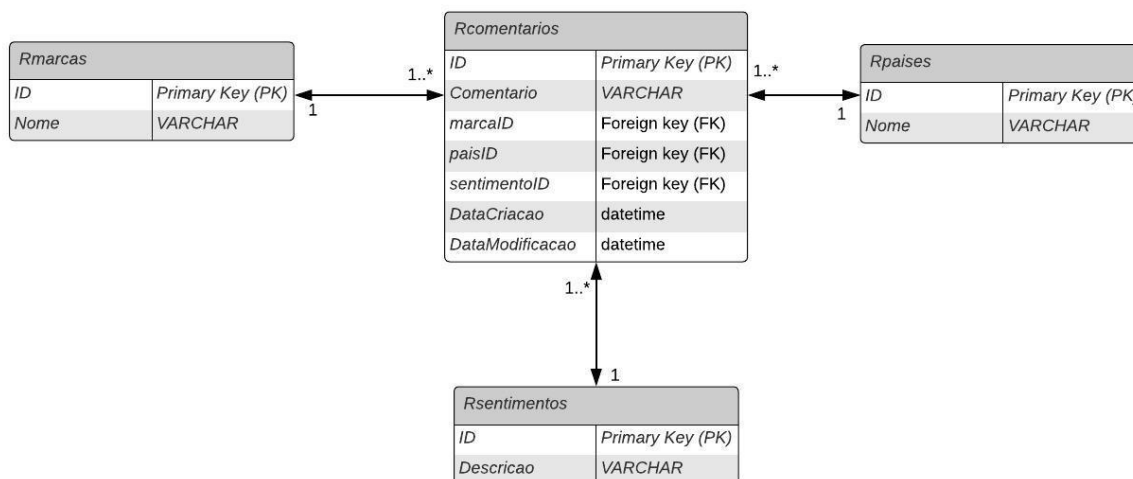
Segundo a Microsoft (2023, tradução própria):

Use o SSMS para acessar, configurar, gerenciar, administrar e desenvolver todos os componentes do *SQL Server*, do BD SQL do Azure [...]. O SSMS fornece um utilitário único e abrangente que combina um amplo grupo de ferramentas gráficas com muitos editores de *script* avançados para fornecer acesso ao *SQL Server* para desenvolvedores e administradores de BD de todos os níveis de habilidade.

A ferramenta Bafafá após fazer todo o levantamento e processamento das informações, disponibiliza um arquivo em .CSV que é um arquivo padrão para upload em qualquer BD disponível hoje em mercado.

O Bafafá tem a opção de utilizar direto o arquivo fonte gerado em .CSV ou um BD, já que o arquivo em formato Excel dispõe da funcionalidade de importação e exportação de dados. Em suma, o *software* tem compatibilidade para inclusão de um BD, caso o usuário opte por esta opção, como exemplificado na figura 2.

Figura 2. Diagrama do Banco de Dados



Fonte: elaborado pelos autores (2024).

O diagrama de BD foi separado o resultado do .CSV em quatro tabelas para que possa ser alimentado o corretamente sendo elas as tabelas: **Rmarcas** onde estará explicito as marcas ou nome de um produto ou o que foi pesquisado no caso deste trabalho marcas de vinho, em seguida temos a tabela **Rpaíses** que registrará todos ou apenas os países pesquisados dependendo da demanda, na sua continuação com a tabela **Rsentimentos** onde se tem apenas três resultados positivo, neutro e negativo. Por fim, a tabela **Rcomentarios** que trará as informações dos comentários, data e hora da alimentação feita no BD, data de modificação dos dados do BD e todas as chaves estrangeiras (em inglês *foreign Keys - FK*) que fazem ligação com as outras tabelas listadas, obtendo a informação completa desejada.

As ferramentas de auxílio serão apresentadas a seguir.

### 5.3. Ferramentas de Auxílio

O Bafafá utiliza as ferramentas de auxílio: VSC (ambiente, no qual, será estruturado os códigos do *software*), utilizando a linguagem Python 3.11 e as extensões; utiliza-se também as ferramentas da Microsoft como Excel para a exportação de dados via raspagem e Power BI para a visualização das informações coletadas.

A captura dos comentários gerados pelos usuários é realizada através do: X e o Instagram, que são os principais meios de captação de dados utilizados.

Na seção a seguir, será exemplificado o digrama de fluxo de dados.

### 5.4. Diagramas de Fluxo de Dados

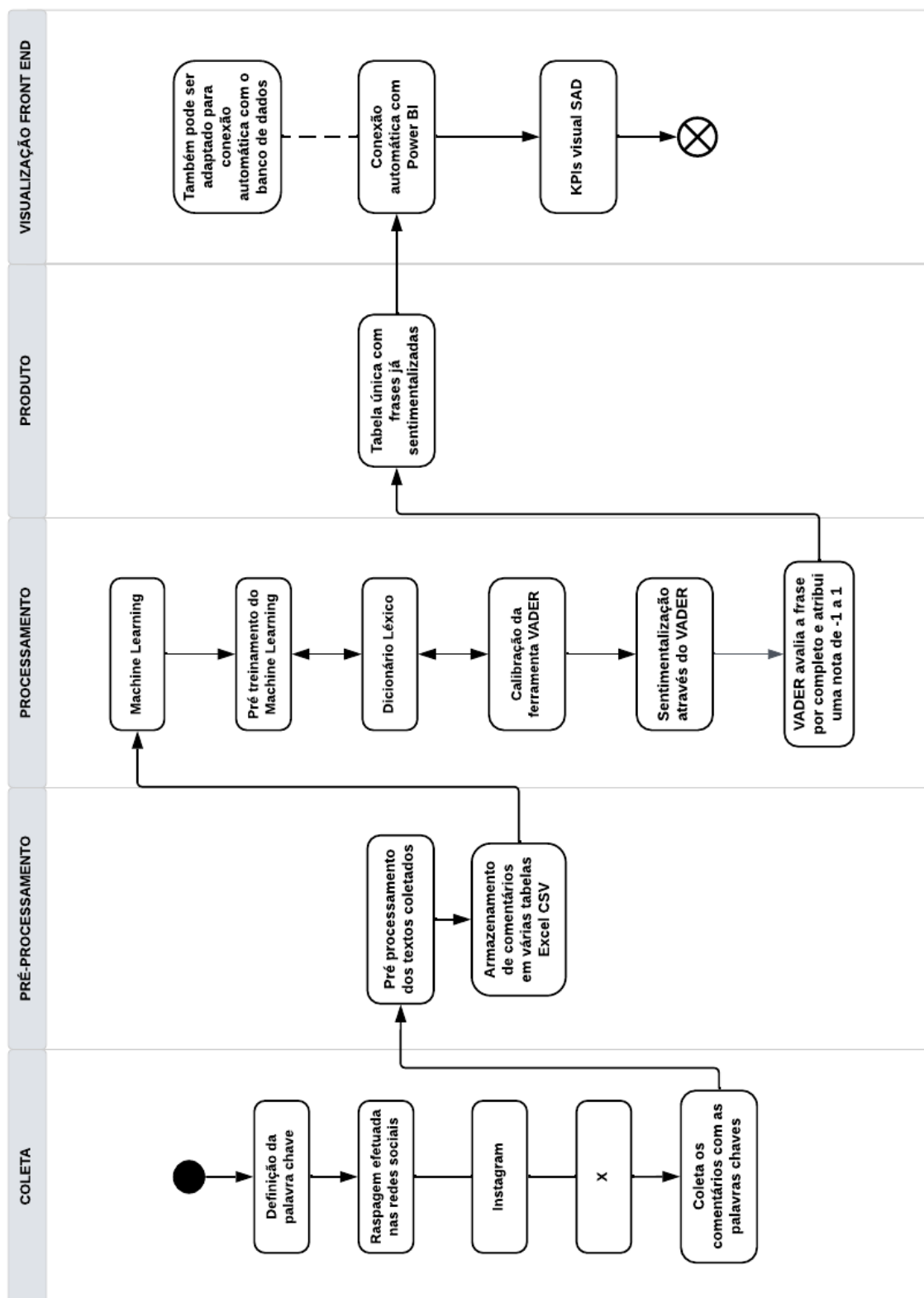
O Diagrama de Fluxo de Dados (DFD) é uma representação gráfica de como as informações se movem em um determinado sistema ou processo.

De acordo com Jóia e Werner (2021):

O Diagrama de Fluxo de Dados é uma ferramenta de modelagem gráfica de Análise Estruturada que descreve a transformação de entradas e saídas, ou seja, a técnica de análise estruturada: utilizada para demonstrar o fluxo de informações de uma organização, permitindo, assim, à atividade de organização sistema & métodos utilização racional dos sistemas mecanizados.

Pode ser observado a representação do DFD do sistema na figura 3.

Figura 3. Diagrama de fluxo de dados do Bafafá



Fonte: elaborado pelos autores (2023).

O processo se inicia na **Coleta** com o usuário inserindo a palavra-chave para pesquisa, quando definida a palavra-chave inicia-se o processo de raspagem nas redes sociais escolhidas – X ou Instagram, coletando assim comentários com a palavra pré-definida anteriormente. No **Pré-processamento** é realizada depuração dos dados obtidos, e o armazenamento dos comentários em tabelas .CSV. Já no **Processamento** é o processo responsável pela execução do AM, o usuário realiza o pré-treinamento do AM, a alimentação das palavras – positivas e negativas no dicionário léxico, também realiza a calibração da ferramenta VADER, este, iniciando o processo de sentimentalização dos comentários atribuindo notas de -1 a 1. O **Produto** é a tabela final com a exibição das postagens já sentimentalizadas, feito isso o usuário escolherá se usará direto o arquivo .CSV para visualização final ou se será integrado a um banco de dados, por fim, a **Visualização front-end** a versão atual do *software* possui conexão direta com a ferramenta Power BI – mas também pode ser adaptada/aprimorada para realizar a conexão direta com BD, a ação final é o Bafafá através de KPIs e *dashboards* para realizar a exibição dos indicadores visuais – SAD.

No próximo capítulo, será explanado o assunto dicionário de dados.

## 5.5. Dicionário de Dados

De acordo com Assunção (2024):

O Dicionário de Dados, fundamental em qualquer projeto de gestão de dados, atua como uma ponte entre o conhecimento técnico e o acesso à informação. Ele é, essencialmente, um catálogo organizado de todos os elementos de dados relevantes para um sistema, projeto ou banco de dados. Inclui descrições detalhadas de cada dado, como seu significado, relações, origem, uso e formato.

Em outras palavras, o Dicionário de Dados (DD) fornece uma compreensão estruturada dos dados e suas conexões, registrando detalhes importantes, como tipos de atributos, valores permitidos e relações com demais elementos, permitindo a simplificação da interpretação dos dados pelos usuários. Alguns dos elementos essenciais são: nome do campo, tipo de dado, descrição e restrições. O DD também funciona como um repositório centralizado para informações sobre os dados do banco de dados, como relacionamentos, significados e origens (Date, 2003).

Podem ser observados alguns exemplos de dicionário de dados nas tabelas 2 e 3 a seguir:

Tabela 2. Entidade: dados de redes sociais

ENTIDADE: DADOS DE REDES SOCIAIS						
DESCRIÇÃO:		Essa entidade armazena as informações relacionadas às postagens em redes sociais sobre o produto/serviço.				
VARIÁVEL	NOME VARIÁVEL	TIPO DE DADO	DESCRIÇÃO	VALORES PERMITIDOS	VALORES NULOS	TIPO DE RESTRIÇÃO
ID da Postagem	ID_Postagem	INT	Identificador único da postagem.	n – MAX	não	PK/not null
ID do usuário	ID_Usuário	VARCHAR	Identificador único do usuário.	30	não	Unique/not null
Texto da Postagem	Texto_Postagem	VARCHAR	O texto da postagem feita nas redes sociais.	1 – MAX	não	Unique/not null
Pais	Pais_Postagem	STRING	País em que foi feita a postagem	1 – 30	não	Unique/not null
Plataforma	Plataforma	VARCHAR	A plataforma de redes sociais em que a postagem foi feita.	30	não	Unique/not null

Fonte: elaborado pelos autores (2024).

A tabela 2 representa: alguns dados que podem ser coletados via API e sua construção em arquivo .CSV, exportado.



Pode ser observada a representação do arquivo resultante após a aplicação da ferramenta VADER na tabela 04 atribuída ao trabalho de graduação (TG), a qual, fornece as informações necessárias para um estudo.

Tabela 3. Tabela final após aplicação do VADER

ENTIDADE: DADOS DE REDES SOCIAIS						
DESCRIÇÃO:	Essa entidade armazena as informações relacionadas às postagens em redes sociais sobre o produto/serviço, país e marca.					
VARIÁVEL	NOME VARIÁVEL	TIPO DE DADO	DESCRIÇÃO	VALORES PERMITIDOS	VALORES NULOS	TIPO DE RESTRIÇÃO
ID - slno	Sub-Linha Numerado Ordenadamente	INT	Identificador único da postagem.	0 a N	Tratado	Tratado
text	Comentário do usuário	VARCHAR	O texto da postagem feita nas redes sociais.	1 - MAX	Tratado	Unique/not null
Sentiment	Sentimento adicionado pela ferramenta	VARCHAR	O texto da ferramenta: Positivo, Neutro, Negativo.	Definido	Tratado	Not Null
Variety	Marca, ou tipo uma informação adicional do conteúdo buscado	VARCHAR	Marca do Vinho no caso	1 – 30	Tratado	Not null
Country	País	VARCHAR	País na qual a informação foi extraída	1 - 30	Tratado	Not null

Fonte: elaborado pelos autores (2024).

A tabela acima demonstra o padrão que o arquivo final será retornado, mostrando o padrão de colunas, tipo de dado, tamanho. Para que a ferramenta de análise Power BI possa buscar os dados corretamente e alimentar os *dashboards*.

Será retratado a seguir a descrição lógica dos processos.

## 5.6. Descrição Lógica dos Processos

Nessa seção exemplificamos a descrição lógica dos processos oferecendo uma visão detalhada de como o sistema funciona, desde a coleta de dados até a correlações dos dados coletados:

- **Coleta de dados de redes sociais:** Este processo é responsável pela raspagem/coleta de comentários nas redes sociais. O processo ocorre quando o usuário informa as entradas, como palavra-chave que será utilizada na pesquisa, data limite da API e quantidade limite de postagens a serem pesquisadas. Assim, o Bafafá faz conexão com as APIs das redes sociais e realiza a raspagem das postagens, comentários e/ou menções relacionados que contenham os parâmetros pré-definidos anteriormente, após isso, *software* exibe as informações obtidas em arquivo .CSV.
- **Pré-processamento dos dados:** Processo encarregado de realizar a depuração dos dados obtidos da base final para que não haja redundância de dados. Retirando assim, todas informações irrelevantes para o processo de sentimentalização, como colunas de ID, nome de usuário e palavra-chave.
- **Processamento:** Já o processamento dos dados é responsável pela lapidação destas informações.
- **Análise de sentimentos:** Esta parte do processo é responsável pela sentimentalização das postagens, comentários e/ou menções coletadas em sentimentos positivos, neutros e negativos com o auxílio do dicionário léxico e aprendizado de AM utilizando o VADER. Nessa parte do processo também são efetuados testes com as postagens com e sem processamento para analisar qual informação está mais fiel a uma análise plenamente humana.
- **Análise de correlações:** Nesta etapa o combina os dados obtidos pela sentimentalização e dados definidos pelo usuário como, por exemplo, dados de variações nas vendas por mês – uma vez que esteja integrado ao sistema de vendas da empresa, para que possa obter dados relevantes para a análise. Posto isso, o Bafafá calcula as correlações entre os sentimentos (positivo, neutro e negativo) e suas variações de venda, esses resultados serão armazenados de acordo com o processo definido pela empresa.

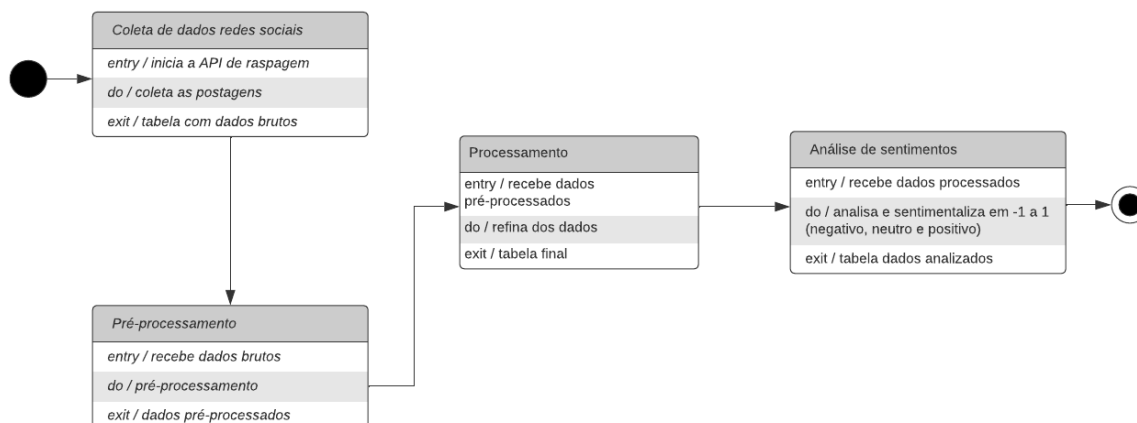
O próximo assunto a ser abordado é diagramas de transição de estado.

## **5.7. Diagramas de Transição de Estado**

Diagramas de Transição de Estado, também conhecidos como Diagrama de Máquinas de Estado, são representações visuais utilizadas para modelar o comportamento de um sistema, exemplificando como ele pode se modificar de um estado para outro conforme ocorrem eventos internos e externos. Estes podem ser expandidos e/ou ajustados com mais especificidades, eventos singulares e/ou estados intermediários conforme a necessidade de representação do processo em questão.

Os diagramas de estado para dados de redes sociais e de previsão de vendas do Bafafá podem ser observados nas figuras 4, 5 e logo abaixo, a figura 6 exemplificará o diagrama de estado para análise de sentimentos:

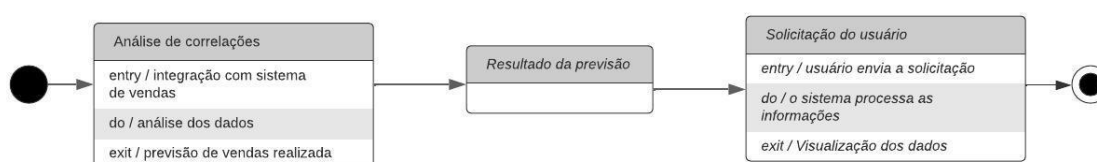
Figura 4. Diagrama de estado para dados de redes sociais



Fonte: elaborado pelos autores (2024).

A figura 4 apresenta o processo de diagrama de estado dos dados de redes sociais. O primeiro componente **Coleta de Dados das Redes Sociais** é responsável pela raspagem, no qual, a entrada (*entry*) é a iniciação da API de raspagem, a ação (*do*) é a coleta dos comentários e a saída (*exit*) é a tabela com os dados brutos. Posteriormente, se inicia o **Pré-processamento** dos comentários, sendo a entrada o recebimento dos dados brutos, ação o pré-processamento destes e a saída os dados pré-processados. Logo após o *software* irá executar o **Processamento dos Dados**, recebendo como entrada os dados pré-processados, a ação será refinar a limpeza já executada e como saída exibirá a tabela com os dados finais. Na **Análise de Sentimentos** a entrada recebe os dados que foram processados anteriormente, executa a ação de analisar e sentimentalizar entre -1 a 1 e sua saída é exibir a tabela com os comentários sentimentalizados.

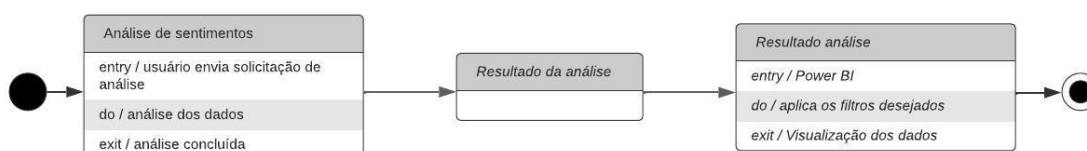
Figura 5. Diagrama de estado para previsão de vendas



Fonte: elaborado pelos autores (2024).

Já na figura 5 podemos observar o Diagrama de Estado para Previsões de Venda. O componente **Análise de Correlações** tem como entrada a integração com o sistema de vendas da empresa, sua ação é a análise dos dados compartilhados e saída é a previsão de vendas realizada pelo Bafafá, gerado o **Resultado da Previsão**, o componente subsequente é a **Solicitação do Usuário**, tendo como entrada a solicitação enviada pelo usuário, o software realiza a ação de processar as informações e como saída exibe a visualização dos dados.

Figura 6. Diagrama de estado para análise de sentimentos



Fonte: elaborado pelos autores (2024).

Como é possível notar na figura 6, será descrito o processo de Análise de Sentimentos. A entrada do primeiro componente **Análise de Sentimentos** é o usuário enviado uma solicitação de análise, a ação a ser executada é a análise dos dados solicitados e a saída são os dados analisados. Com o **Resultado da Análise** a inicia-se a entrada de dados dos dados no Power BI, a aplicação de filtros será a ação a ser realizada e sua saída será a visualização dos dados através de *dashboards*.

No capítulo seguinte se desenvolverá o tema diagramas estruturais.

## 5.8. Diagramas Estruturais

Os diagramas estruturais funcionam como uma espécie de mapa que auxilia no entendimento do funcionamento do sistema, mostrando as partes envolvidas e suas funções e como cada uma se relaciona entre si.

Os três tipos de diagramas são: o Diagrama de Entidade-Relacionamento (DER), o Diagrama de Componentes e o Diagrama de Camadas, auxiliando na compreensão da estrutura do sistema e suas interações, mostrando como deve ser alcançado os objetivos do projeto.

No próximo tópico, será explicado sobre o diagrama entidade-relacionamento.

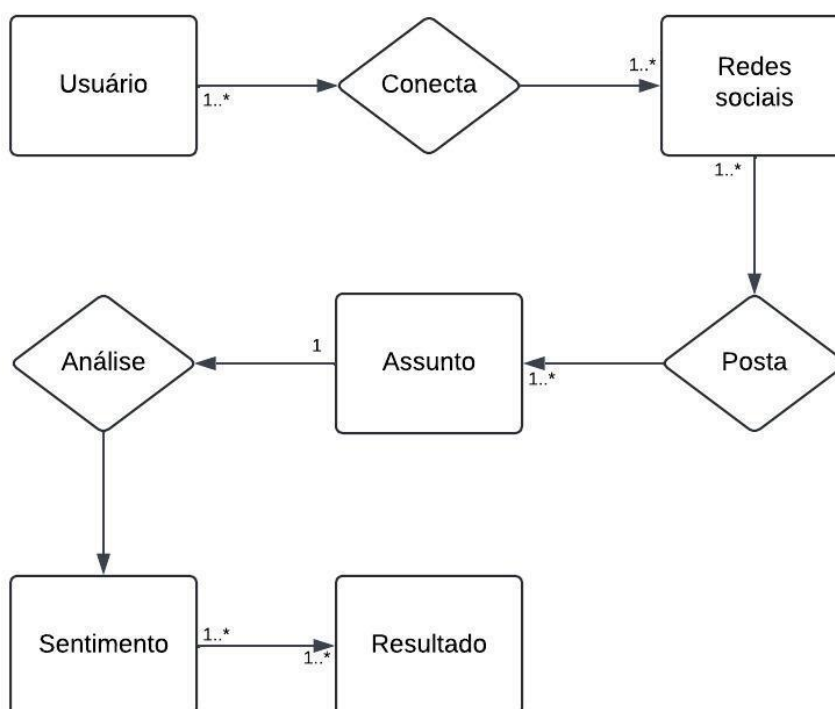
## 5.9. Diagrama Entidade-Relacionamento

O DER é uma ferramenta gráfica utilizada para representar as interconexões entre as entidades em um sistema, auxiliando no processo de entendimento em como as postagens em redes sociais afetam diretamente nas vendas, onde é analisado as postagens dos usuários nas redes sociais e através da análise de sentimentos é possível verificar seu impacto na compra de produtos e serviços.

De acordo com Nogueira (1988, *apud* Franck; Pereira; Dantas Filho, 2021, p. 2): "Um diagrama entidade relacionamento (ER) é um tipo de fluxograma que ilustra como "entidades", pessoas, objetos ou conceitos, se relacionam entre si dentro de um sistema".

A figura 7 é a representação do DER do *software* proposto neste trabalho de graduação:

Figura 7. Diagrama entidade de relacionamento



Fonte: elaborado pelos autores (2024).

A figura 7 descreve os processos do diagrama entidade de relacionamento: um usuário se conecta a uma ou mais redes sociais, e nesta pode efetuar uma ou mais postagens sobre um ou diversos assuntos. O Bafafá coletará uma dessas postagens para ser analisada e este comentário receberá um sentimento, gerando um ou mais resultados, conforme o número de postagens coletadas.

Continuamente será exposto caso de uso.

### **5.10. Caso de uso:**

O caso de uso selecionado envolve uma busca complexa pelas APIs que busca a palavra-chave “VINHO”, verifica-se que este termo se ramifica em diversos setores como: marca, país, comentários que podem variar dependendo da região de consumo. A escolha da palavra-chave foi devido à complexidade no processo de separação de dados, seguindo os passos do *software* se obtém comentários de diversos países, então usa-se um recurso do Excel para traduzir os comentários e sedimentar da maneira correta em relação à estrutura para que o *software* possa processar todos os dados dos arquivos do X e do Instagram.

Após o VADER realizar o processamento, tratativa dos dados e efetuar a sentimentalização das frases, é gerado um documento em formato .CSV contendo o arquivo de saída que será processado pelo *software* de visualização Power BI que mostrará as informações sobre o termo utilizado. A partir dessas informações é possível chegar a diversas tomadas de decisões, dados valiosos como por exemplo: quais os vinhos possuem as melhores avaliações, quais as melhores marcas, os vinhos mais consumidos, onde são consumidos diversos estopins para uma estratégia de vendas.

O *software* realiza a RD através das postagens dos usuários nas redes sociais, um comentário sobre determinado assunto, produto ou serviço e retorna informações valiosas para os cientistas de dados que auxilia nas estratégias de *marketing*, vendas, além de apresentar outras variáveis vantagens para uma empresa.

Tabela 4. Caso de uso

<b>Caso de uso:</b> utilização da ferramenta	
<b>Descrição Geral:</b>	o caso de uso iniciasse quando o usuário deseja fazer uma pesquisa sobre algum termo nas redes sociais.
<b>Atores:</b>	usuário e/ou analista de dados
<b>Pré-condições:</b>	o usuário e/ou analista de dados inicia a ferramenta, palavra-chave, e palavras do dicionário léxico inseridas.
<b>Garantia de sucesso:</b>	raspagem de dados bem-sucedida.
<b>Fluxo básico de funcionamento:</b>	<ol style="list-style-type: none"> <li>1. A ferramenta faz a leitura da amostra da raspagem das redes sociais;</li> <li>2. Aplicado as primeiras tratativas dos comentários, como acentuação, letras maiúsculas e minúsculas;</li> <li>3. Aplicação de notas (-1 a 1) sobre as palavras treinadas no AM junto ao dicionário léxico;</li> <li>4. Aplicação do sentimento (positivo, neutro e negativo) nos comentários;</li> <li>5. Define como será o arquivo final, padrões do nome das colunas;</li> <li>6. Exporta o arquivo já sentimentalizado;</li> <li>7. Power BI faz a leitura do arquivo e mostra o resultado em tela.</li> </ol>
<b>Fluxo alternativo:</b>	<p>os processos do 1 ao 6 são iguais, com exceção do 7, tendo sua alternativa exemplificada abaixo:</p> <ol style="list-style-type: none"> <li>7. Arquivo exportado passa por adição a um BD para criação de histórico de determinado assunto.</li> </ol>

Fonte: elaborado pelos autores (2024).

A tabela 4 aborda uma explicação geral, de como o usuário/ferramenta trabalham em conjunto para obter o resultado de KPIs e *dashboards* dos resultados, e seus fluxos principais e alternativos.

O próximo assunto abordará sobre os diagramas de classe.



## 5.11. Diagramas de Classe

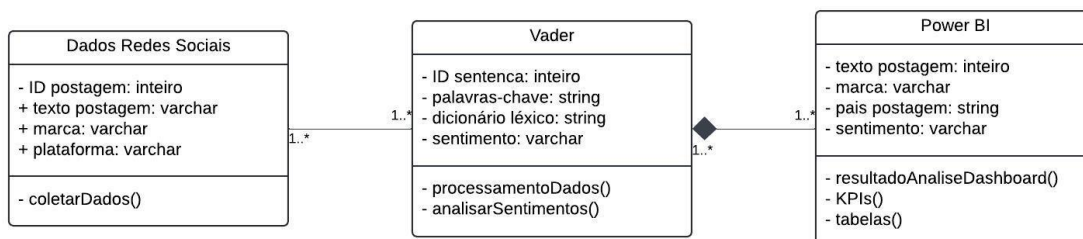
O diagrama de classe é uma ferramenta que descreve os elementos essenciais de um sistema, sendo fundamental para a representação de classes, atributos e métodos.

De acordo com Sadowska (2020, *apud* Souza Filho, 2023, p. 28):

O diagrama de Classes permite representar a estrutura lógica do sistema, baseando-se no paradigma da orientação a objetos. No entanto, é necessário um grande esforço para conduzir o processo de diagramação de Classes com precisão, para evitar falhas do sistema no futuro.

Na figura 8 é possível verificar essa representação:

Figura 8. Diagrama de classe do Bafafá



Fonte: elaborado pelos autores (2024).

Na parte superior de cada componente contém o nome da classe, por exemplo **“Dados Redes Sociais”**. Já a parte do meio contém os atributos de cada classe, como **“-ID postagem: inteiro”** e cada um destes devem possuir uma linha exclusiva. É importante destacar que todas as classes possuem diferentes níveis de acesso, e este dependerá do modificador de acesso atribuído, o qual trará a visibilidade para a informação desejada: público receberá o sinal de (+) e privado (-). Posteriormente aos “:” será informada a variável desta informação como: inteiro, varchar, dateline entre outros. E na parte inferior será incluída as operações da classe (métodos), no qual, também possui uma única linha para cada operação **“coletarDados()”**. Exemplificando o primeiro componente, a primeira célula informações da tabela **“Dados Redes Sociais”**, onde serão armazenadas as informações, o bloco do meio trará informações dos dados que serão utilizados no processo e, por fim, a última

célula exibe qual ação será executada/exibida pelo sistema, a mesma lógica deve ser aplicada para os demais componentes.

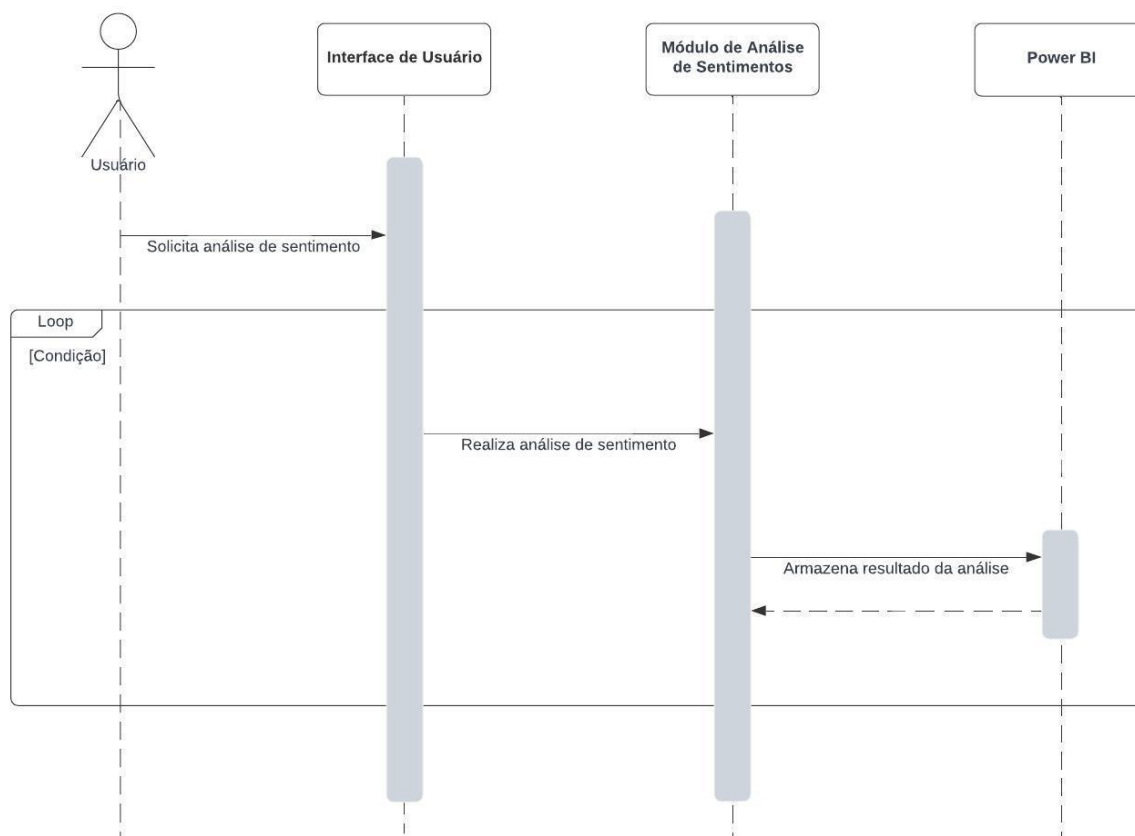
A seguir, será explicado sobre os diagramas de sequência.

## 5.12. Diagramas de Sequência

Através de uma representação gráfica, um Diagrama de Sequência (DS) permite aos desenvolvedores visualizarem de forma detalhada e ordenada a troca de mensagens que ocorrem entre os objetos para realizar determinado processo em um sistema, este, pode fazer parte de uma função de *software* ou de um processo de negócio. O DS possui uma perspectiva temporal, abrangendo duas dimensões: na dimensão horizontal será representada a interação dos objetos, e na dimensão vertical será indicada a sequência temporal dos processos (Pereira, 2011).

O DS de análise de sentimentos é representado pela figura 9 na qual, descreve a sequência de ações para que seja realizada a análise de sentimentos dos comentários coletados das redes sociais:

Figura 9. Diagrama de sequência de análise de sentimentos



Fonte: elaborado pelos autores (2024).

O diagrama exemplificado na figura 9 traz informações sobre o seguinte processo de laço de repetição (em inglês *loop*): o usuário irá solicitar a análise de sentimento, o *software* Bafafá irá processar a ação solicitada e irá realizar a análise conforme os dados fornecidos para tal. Após efetuado todo o processo de coleta, pré-processamento, processamento e análise dos comentários, o Bafafá irá alimentar o Power BI e retornará as informações da análise realizada para o usuário em forma de dashboards inteligentes.

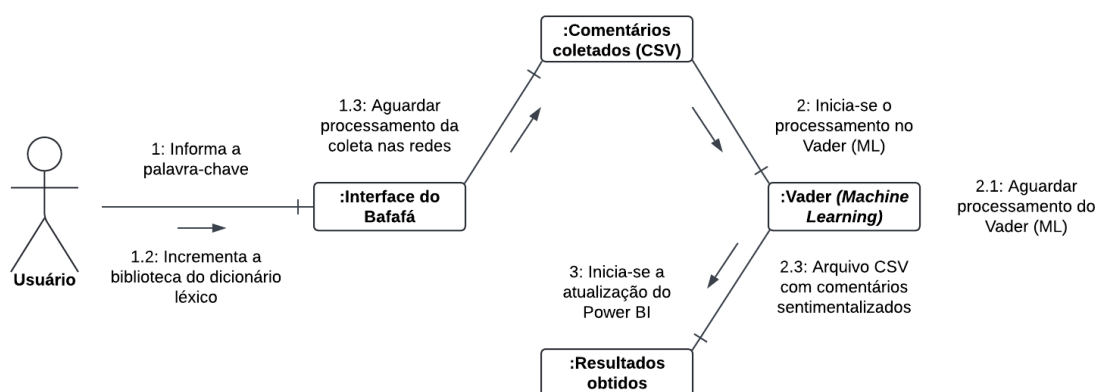
Será abordado no tópico seguinte sobre os diagramas de comunicação.

### 5.13. Diagramas de Comunicação

O diagrama de comunicação é uma representação gráfica da interação entre objetos ou funções de linhas de vida e mensagens, auxiliando no processo de entendimento da relação entre objetos. De acordo com a IBM (2021): “Diagramas de comunicação são um tipo de diagrama de interação que você pode usar para explorar o comportamento dinâmico de um sistema ou aplicativo de software”.

Temos a representação do diagrama de comunicação do Bafafá na figura 10 a seguir.

Figura 10. Diagrama de comunicação do Bafafá



Fonte: elaborado pelos autores (2024).

Na figura 10 pode ser observado o processo de comunicação do usuário com o Bafafá. O usuário informará a palavra-chave que deseja pesquisar, subsequentemente, irá incrementar a biblioteca do dicionário léxico com palavras positivas, negativas e neutras, e aguardará o processamento da coleta dos comentários nas redes sociais. Após a coleta destes, inicia-se o processamento do AM no VADER - este pode levar alguns minutos, dependendo da amostra que será processada pelo *software*. O resultado deste, trará um arquivo .CSV (Excel) com os comentários já sentimentalizados. Em seguida, inicia-se a atualização do Power BI para exibição dos resultados obtidos.

No próximo capítulo, será discorrida a arquitetura da solução.

## 6. ARQUITETURA DA SOLUÇÃO

Este TG foi elaborado de uma forma híbrida dispondo de algumas ferramentas para chegar ao resultado obtido:

- Foi utilizado o programa de códigos VSC, um programa gratuito da Microsoft para assim desenvolver e executar os códigos necessários para o projeto;
- A linguagem de programação aplicada foi Python 3.11, por obter as mais atualizadas bibliotecas de dados do mercado, tais como Pandas (trabalhar utilizando Excel), VADER.sentiment (Extensão do VADER; sentimentalizador escolhido no projeto), tweepy e Instapy (extensão que permite fazer a RD das respectivas redes sociais);
- Utiliza-se o Excel, programa da Microsoft como arquivo para o trânsito de dados em .xlsx, e .CSV;
- Foi empregado também o programa Power BI da Microsoft, para apresentação de dados, e o uso da linguagem DAX, do programa para chegar ao resultado desejado de *dashboards*.

A seguir, apresentaremos as considerações finais deste trabalho.

## 7. CONSIDERAÇÕES FINAIS

Podemos concluir que o objetivo da introdução foi apresentar a necessidade de inovação e a importância das redes sociais para a tomada de decisões estratégicas em um mercado cada vez mais digitalizado e competitivo. A ferramenta Bafafá foi desenvolvida com o objetivo de fornecer percepções valiosas sobre o sentimento do público em relação a um assunto específico, o que pode ser útil para empresas e as partes interessadas em diversos campos. A contribuição do presente deste TG é a obtenção e viabilização de resultados como uma espécie de nota em um local (redes sociais), através dos próprios comentários dos usuários que utilizam ou falam sobre determinado assunto e/ou produto.

Investigando o impacto nas redes sociais sobre um determinado assunto a ser estudado, por exemplo, ao escrever o termo “VINHO” (termo genérico), ocorre uma busca geral nas redes sociais (X e Instagram), com a finalidade de fazer uma sentimentalização nos comentários das pessoas que fazem publicações. Além disso é possível verificar se representa algo positivo ou negativo em nosso exemplo a ser apresentado o termo “VINHO” um termo genérico que vai fazer uma busca geral nas redes sociais (X e Instagram), a fim de fazer uma sentimentalização nos comentários das pessoas que fazem suas publicações e verificar se é positiva ou negativa sobre o assunto pesquisado.

Na ferramenta a absorção de dados é viabilizada através de APIs, que realizam toda a raspagem de dados de um determinado assunto, e exportará para arquivos de Excel, em que será convertido para uma visualização de dados mais sucinta, através de ferramentas de visualização, como no caso o Power BI.

Nossa contribuição para o Trabalho de Conclusão de Curso (TCC) Bafafá se dá na obtenção e viabilização de resultados como uma espécie de nota em um local (redes sociais), através do próprio comentário dos que utilizam ou falam sobre determinado assunto. Local este no qual, as pessoas realmente expressam suas emoções reais sobre algo, a fim de receber informações valiosas para serem utilizadas em um estudo.

Uma sugestão para pesquisas futuras é a aplicação em outros locais, como outras redes sociais, resumos, textos e filmes. Visto que, a aplicação em diversos locais, onde se é possível adicionar comentários, que o local não tenha uma tecnologia de análise de sentimentos (as famosas estrelas que normalmente vão de 1 a 5).

A reflexão se dá na gigantesca quantidade de informação preciosa presente nesses locais, pois colocando em uma situação real, onde alguém faz a compra de algo e não faz a avaliação no local da compra (seja ela em shoppings, feiras, comércios de rua), a pessoa vai fazer a avaliação do produto ou serviço adquirido nas redes sociais, tornando o Bafafá uma poderosa ferramenta de absorção de dados de avaliação sobre qualquer assunto que o usuário desejar. A ferramenta Bafafá fica aberta para modificações sobre informações a mais, que é possível captar de acordo com o poder da API das redes, como datas, sexo, horários, temporalidade, reações, uma diversidade de outras informações que possam somar para projetos futuros.

## REFERÊNCIAS

- ANTUNES, M. *et al.* Monitoramento de informação em mídias sociais: o e-Monitor Dengue. **TransInformação**, Campinas, v. 26, n. 1, p. 09-18, jan./abr. 2014. DOI 10.1590/S0103-37862014000100002 Disponível em: [https://www.researchgate.net/publication/273942273\\_Monitoramento\\_de\\_informacao\\_em\\_midias\\_sociais\\_o\\_e-Monitor\\_Dengue](https://www.researchgate.net/publication/273942273_Monitoramento_de_informacao_em_midias_sociais_o_e-Monitor_Dengue). Acesso em: 08 set. 2023.
- ASSUNÇÃO, D. Dicionário de Dados: Guia Completo para Iniciantes. *In*: Blog Faspec. **FASPEC Evoluir importa**. Goiânia, 31 jan. 2024. Disponível em: <https://blog.faspec.edu.br/dicionario-de-dados/>. Acesso em: 17 mar. 2024.
- AWS. O que é machine learning? *In*: Amazon Web Services, Inc. **AWS**. [S.l.], c2023. Disponível em: <https://aws.amazon.com/pt/what-is/machine-learning/>. Acesso em: 09 out. 2023.
- BACCIANELLA, S.; ESULI, A.; SEBASTIANI, F. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. **INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION (LREC'10)**, Malta, v. 10, p. 2200-2204, mai./2010. Disponível em: [http://www.lrec-conf.org/proceedings/lrec2010/pdf/769\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2010/pdf/769_Paper.pdf). Acesso em: 29 set. 2023.
- BARBOSA, A. B. G.; CAVALCANTI, A. B. Web scraping e análise de dados. *In*: Congresso Nacional de Pesquisa e Ensino em Ciências V CONAPESC, 5., 2020. **Anais [...]**. Campina Grande: Realize Eventos Científicos e Editora, 2020, p 1-4, 2020. Disponível em: <https://www.editorarealize.com.br/index.php/artigo/visualizar/73189>. Acesso em: 26 set. 2023.
- BENEVENUTO, F.; RIBEIRO, F.; ARAÚJO, M. Métodos para análise de sentimentos em mídias sociais. *In*: XXI Simpósio Brasileiro de Sistemas Multimídia e Web. Minicursos, 21., 2015, Manaus. **Anais [...]**. Manaus: Sociedade Brasileira de Computação, 2015, cap. 2, p. 31-59, out./2015. Disponível em: <https://books-sol.sbc.org.br/index.php/sbc/catalog/book/19>. Acesso em: 11 set. 2023.
- BOROVKOVA, S. Como dados de sentimento de notícias embasam estratégias de negociação e investimento. *In*: Refinitiv Eikon. **Refinitiv**. Nova Iorque, [S.d.], c2020. Disponível em: <https://www.refinitiv.com/pt/blog/future-of-investing-trading/como-dados-de-sentimento-de-noticias-embasam-estrategias-de-negociacao-e-investimento/>. Acesso em: 29 set. 2023.
- BROOKS JUNIOR; F. P. No Silver Bullet: Essence and Accidents of Software Engineering. **IEEE Computer**, Carolina do Norte, v. 20, n. 4, p. 10-19, abr./1987. DOI 10.1109/MC.1987.1663532. Disponível em: [https://www.researchgate.net/publication/220477127\\_No\\_Silver\\_Bullet\\_Essence\\_and\\_Accidents\\_of\\_Software\\_Engineering](https://www.researchgate.net/publication/220477127_No_Silver_Bullet_Essence_and_Accidents_of_Software_Engineering). Acesso em: 29 set. 2023.

CARVALHO, P.; SILVA, M. J. SentiLex-PT: Principais características e potencialidades. **Oslo Studies in Language**, Universidade de Oslo, v. 7, n. 1, p. 425-438, mar./2015. Disponível em: <https://journals.uio.no/osla/article/download/1444/1341/5475>. Acesso em: 29 set. 2023.

DATE, C. J. **Introdução a Sistemas Bancos de Dados**. Tradução: Daniel Vieira. 9. ed. Rio de Janeiro: Elsevier, 2003. 896 p. Disponível em: <https://www.google.com.br/books/edition/Introdução+a+sistemas+de+bancos+de+da/d/xBeO9LSIK7UC?hl=pt-BR&gbpv=1&pg=PP1&printsec=frontcover>. Acesso em: 05 abr. 2024.

DURAN, M. S.; ALUÍSIO, S. M. Propbank-Br: a Brazilian Treebank annotated with semantic role labels. **Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)**, Turquia, v. 12, p. 1862–1867, mai./2012. Disponível em: [http://www.lrec-conf.org/proceedings/lrec2012/pdf/272\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2012/pdf/272_Paper.pdf). Acesso em: 29 set. 2023.

ENGHOLM JUNIOR, H. **Engenharia de software na prática**. 1. ed. São Paulo: Novatec Editora, 2010. 440 p.

FRANCK, K. M.; PEREIRA, R. F.; DANTAS FILHO, J. V. Ratio-Entity Diagram: a tool for conceptual data modeling in Software Engineering. **Research, Society and Development**, [S.l.], v. 10, n. 8, p. 1-12, 2021. DOI: 10.33448/rsd-v10i8.17776. Disponível em: <https://rsdjournal.org/index.php/rsd/article/view/17776>. Acesso em: 4 jun. 2024.

GOGUEN, J. A.; LINDE, C. Techniques for Requirements Elicitation\*. **IEEE Computer**, Califórnia, v. 6, n. 6, p. 152-164, jan./1993. Disponível em: [https://www.researchgate.net/publication/3556118\\_Techniques\\_for\\_Requirements\\_Elicitation](https://www.researchgate.net/publication/3556118_Techniques_for_Requirements_Elicitation). Acesso em: 26 out. 2023.

HEMPEL, M. **The use of social media in environmental health research and communication: an evidence review**. 2014. 20 p. Relatório (Mestrado em Saúde Pública) - Universidade de Guelph, Canadá, 2014. Disponível em: <https://ncceh.ca/content/ncceh-student-project-award-2014-m-hempel>. Acesso em: 08 set. 2023.

HUTTO, C.; GILBERT, E. VADER: A Parsimonious Rule-Based Model for Sentiment aAnalysis of Social Media Text. **ICWSM**, Michigan, v. 8, n. 1, p. 216-225, 2014. DOI 10.1609/icwsml.v8i1.14550. Disponível em: <https://ojs.aaai.org/index.php/ICWSM/article/view/14550>. Acesso em: 29 set. 2023.

IBM. Diagramas de comunicação. *In*: Copyright IBM Corporation. **IBM**. [S.l.], 16 mar. 2021. Disponível em: [https://www.ibm.com/docs/pt-br/radfws/9.6?topic=SSRTLW\\_9.6.0/com.ibm.xtools.sequence.doc/topics/ccommndia\\_g.html](https://www.ibm.com/docs/pt-br/radfws/9.6?topic=SSRTLW_9.6.0/com.ibm.xtools.sequence.doc/topics/ccommndia_g.html). Acesso em: 25 abr. 2024.



IBM. O que é aprendizado de máquina (ML)? *In: IBM. IBM. [S.l.]*, c2023. Disponível em: <https://www.ibm.com/br-pt/topics/machine-learning>. Acesso em: 21 fev. 2024.

JÓIA, A.; WERNER, I. Diagrama de fluxos de dados. **Revista Terra & Cultura: Cadernos De Ensino E Pesquisa**, [S.l.], v. 15, n. 30, p. 46-69, jul./dez.2021. Disponível em: <http://periodicos.unifil.br/index.php/Revistateste/article/view/2302>. Acesso em: 4 jun. 2024.

KHAN, R.; DAS, A. **Build Better Chatbots: A Complete Guide to Getting Started with Chatbots**. 1. ed. v. 2. Berkeley: Apress, 2018. E-book. 109 p. DOI [10.1007/978-1-4842-3111-1](https://doi.org/10.1007/978-1-4842-3111-1). Disponível em: <https://link.springer.com/book/10.1007/978-1-4842-3111-1>. Acesso em: 26 out. 2023.

KLEIN, G. H.; GUIDI NETO, P.; TEZZA, R. Big Data e Mídias Sociais: monitoramento das redes como ferramenta de gestão. **Saúde Soc**. São Paulo, v. 26, n. 1, 2007. p. 208-217. DOI 10.1590/S0104-12902017164943. Disponível em: <https://www.scielo.br/j/sausoc/a/hMsMwVXk8G6bqkzNkXJ9hJJ/abstract/?lang=pt>. Acesso em: 08 set. 2023.

LOPES, G. R. *et al.* Introdução à análise exploratória de dados com python. **ERCEMAPI**, Ceará, v. 7, p. 87-94, set./2019. Disponível em: [https://www.researchgate.net/publication/336778766\\_Introducao\\_a\\_Analise\\_Exploratoria\\_de\\_Dados\\_com\\_Python](https://www.researchgate.net/publication/336778766_Introducao_a_Analise_Exploratoria_de_Dados_com_Python). Acesso em: 10 set. 2023.

MARETTI, A. Requisitos funcionais, não funcionais e de domínio. *In: Grupo de Estudos | Prof. André Maretti. Grupo de Estudos | Prof. André Maretti. [S.l.]*, 09 mar. 2014. Disponível em: <https://andremaretti.wordpress.com/2014/03/09/requisitos-funcionais-nao-funcionais-e-de-dominio/>. Acesso em: 21 nov. 2023.

MATOS, G. R. D. **Machine learning aplicado à gestão de activos físicos industriais**. 2021. 71 p. Dissertação (Mestrado em Engenharia Mecânica) - Instituto Superior de Engenharia de Lisboa, Lisboa, 2021. DOI [10400.21/13524](https://doi.org/10.10400.21/13524) Disponível em: <https://repositorio.ipl.pt/handle/10400.21/13524>. Acesso em: 8 set. 2023.

MICROSOFT. WHAT is SQL Server Management Studio (SSMS)? *In: Microsoft. Learn Microsoft. [S.l.]*, 31 mar. 2023. Disponível em: <https://learn.microsoft.com/en-us/sql/ssms/sql-server-management-studio-ssms?view=sql-server-ver16>. Acesso em: 21 maio 2024.

OLIVEIRA, D. J. S. *et al.* A aplicação da técnica de análise de sentimento em mídias sociais como instrumento para as práticas da gestão social em nível governamental. **Revista de Administração Pública**, Rio de Janeiro, v. 53, n. 1, p. 235-251, jan./fev. 2019. DOI 10.1590/0034-7612174204. Disponível em: [http://scielo.br/scielo.php?script=sci\\_arttext&pid=S0034-76122019000100235](http://scielo.br/scielo.php?script=sci_arttext&pid=S0034-76122019000100235). Acesso em: 22 mar. 2024.

PALKOVSKY, B. Dados: quantos geramos e como eles se transformam em insights.

*In*: Linkages. **Linkages**. Campinas, 29 mar. 2023. Disponível em: <https://linkages.com.br/2023/03/29/dados-quantos-geramos-e-como-isso-impacta-nossa-vida/#:~:text=Diariamente%2C%20o%20mundo%20gera%20cerca,2%2C5%20quinti lh%C3%B5es%20de%20dados>. Acesso em: 20 set. 2023.

PASSARI, A. F. L. **Exploração de dados atomizados para previsão de vendas no varejo utilizando redes neurais**. 2003. P. 143. Dissertação (Mestrado em Administração) - Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo, 2003. DOI 10.11606/D.12.2003.tde-30082004-105146. Disponível em: <https://www.teses.usp.br/teses/disponiveis/12/12139/tde-30082004-105146/pt-br.php>. Acesso em: 13 set. 2023.

PEREIRA, L. A. M. **Análise e Modelagem de Sistemas com a UML**. 1. ed. Rio de Janeiro: Luiz Antônio, 2011. 208 p.

PYPI - Python Package Index. **snsrape 0.7.0.20230622**. [S.l.], Python Software Foundation, c2024. Disponível em: <https://pypi.org/project/snsrape/>. Acesso em: 29 set. 2023.

REBOUÇAS, L. A era dos dados: como a conectividade está mudando o mundo dos negócios (para melhor). *In*: Telefônica Brasil S.A. **Vivo Meu Negócio**. [S.l.], 14 jul. 2021. Disponível em: <https://vivomeunegocio.com.br/especialistas/era-dos-dados/>. Acesso em: 08 set. 2023.

ROSA, R. L. **Análise de sentimentos e afetividade de textos extraídos das redes sociais**. 2015. 99 p. Tese (Doutorado em Sistemas Digitais) - Escola Politécnica, Universidade de São Paulo, São Paulo, 2015. DOI 10.11606/T.3.2016.tde-19072016-115713. Disponível em: <https://www.teses.usp.br/teses/disponiveis/3/3141/tde-19072016-115713/publico/RenataLopesRosa2015.pdf>. Acesso em: 08 set. 2023.

SIDDIQI, J.; SHEKARAN, M. C. Requirements Engineering: the emerging wisdom. **IEEE Computer**, California, v. 13, n. 2, p. 15-19, mar./1996. DOI 10.1109/MS.1996.506458. Disponível em: <https://www.computer.org/csdl/magazine/so/1996/02/s2015/13rRUy0qnEb>. Acesso em: 26 out. 2023.

SILVA, M. E. R.; SERRANO, P. H. S. M. Análise de sentimentos em textos de redes sociais: uma comparação entre o ChatGPT e métodos tradicionais. **Cadernos De Comunicação**, Rio Grande do Sul, v. 27, n. 3, 2023. Disponível em: <https://periodicos.ufsm.br/ccomunicacao/article/view/84828>. Acesso em: 22 mar. 2024.

SOMMERVILLE, I. **Engenharia de Software**. Tradução: Ivan Bosni; Kalinka Oliveira. 9. ed. São Paulo: Pearson Prentice Hall, 2011. 529 p. Disponível em: <https://www.facom.ufu.br/~william/Disciplinas%202018-2/BSI-GSI030->

[EngenhariaSoftware/Livro/engenhariaSoftwareSommerville.pdf](#). Acesso em: 26 out. 2023.

SOUZA FILHO, I. P. **Uma Proposta de Incorporação do Diagrama de Classes da UML à Linguagem MASRML Adaptado ao Contexto de Sistema Multiagentes**. 2023. 117p. Dissertação (Mestrado Profissional em Engenharia de Software) – Universidade Federal do Pampa, Campus Alegrete, Alegrete, 2023. Disponível em: <https://repositorio.unipampa.edu.br/jspui/handle/riu/8558>. Acesso em: 4 jun. 2024.

TAUSCZIK, Y.; PENNEBAKER, J. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. **Journal of Language and Social Psychology**, [S.l.], v. 29, n. 1, p. 24-54, fev./2010. DOI 10.1177/0261927X09351676. Disponível em: <https://journals.sagepub.com/doi/10.1177/0261927X09351676>. Acesso em: 14 set. 2023.

THAYER, R; DORFMAN, M. **System and Software Requirements Engineering**. 2. ed. Los Alamitos: IEEE Computer Society Press Tutorial, 2000. 549 p.

## APÊNDICE A – MANUAL DO USUÁRIO

### 1. UTILIZAÇÃO E MANUSEIO DO SOFTWARE

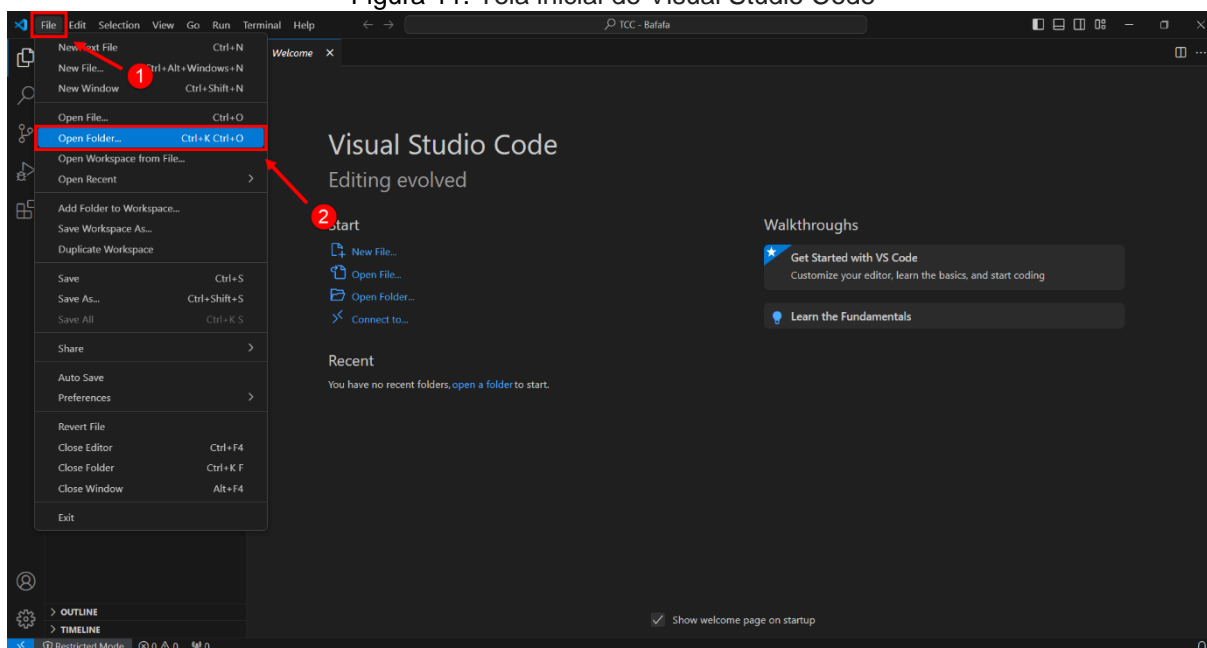
Neste capítulo é orientado ao usuário como utilizar e alimentar o código do *software* Bafafá de acordo com o tema requerido.

#### 1.1. Iniciando o software Bafafá no Visual Studio Code

O usuário deve importar a pasta “TCC – Bafafá” no Visual Studio Code (VSC) seguindo o processo abaixo:

Clicar em **File** (em português “Arquivo”) localizado no carto superior esquerdo, irá abrir o menu, procure pela opção **Open Folder** (“Abrir Pasta”), conforme exibido na figura 11.

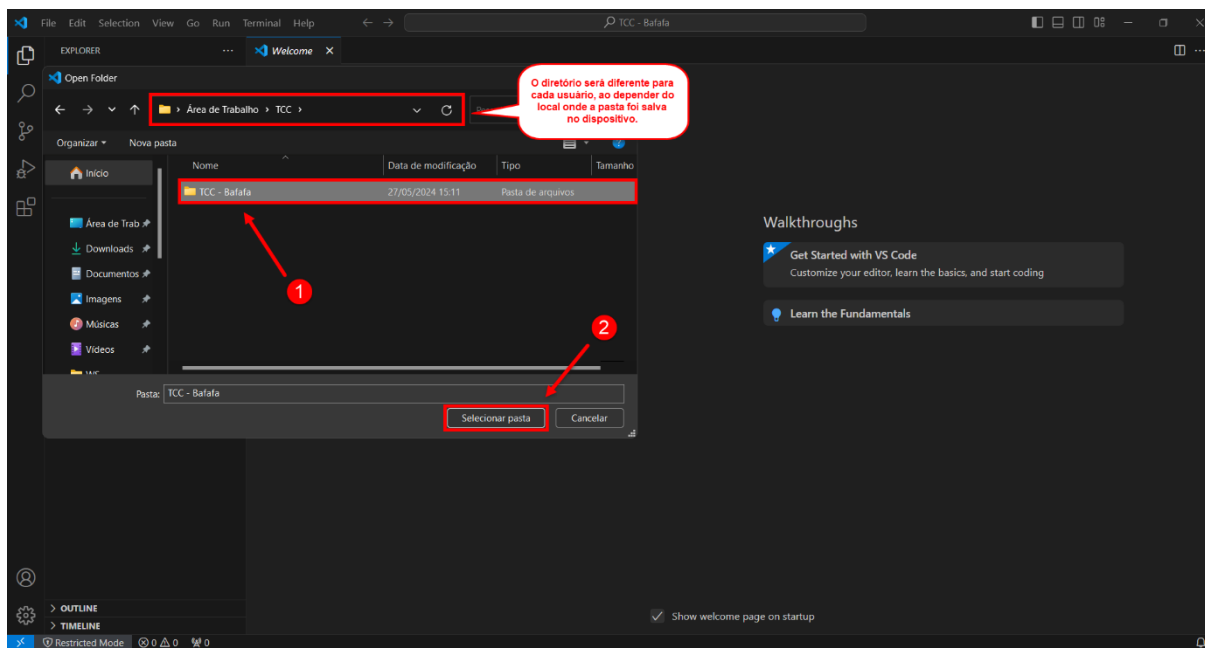
Figura 11. Tela inicial do Visual Studio Code



Fonte: elaborado pelos autores (2024).

A ferramenta irá abrir uma janela para que possa selecionar o local de onde pasta deverá ser importada para o VSC, ao localizá-la e selecione a pasta, clique no botão Selecionar Pasta, conforme pode ser observado na figura 12.

Figura 12. Tela de importação do Visual Studio Code



Fonte: elaborado pelos autores (2024).

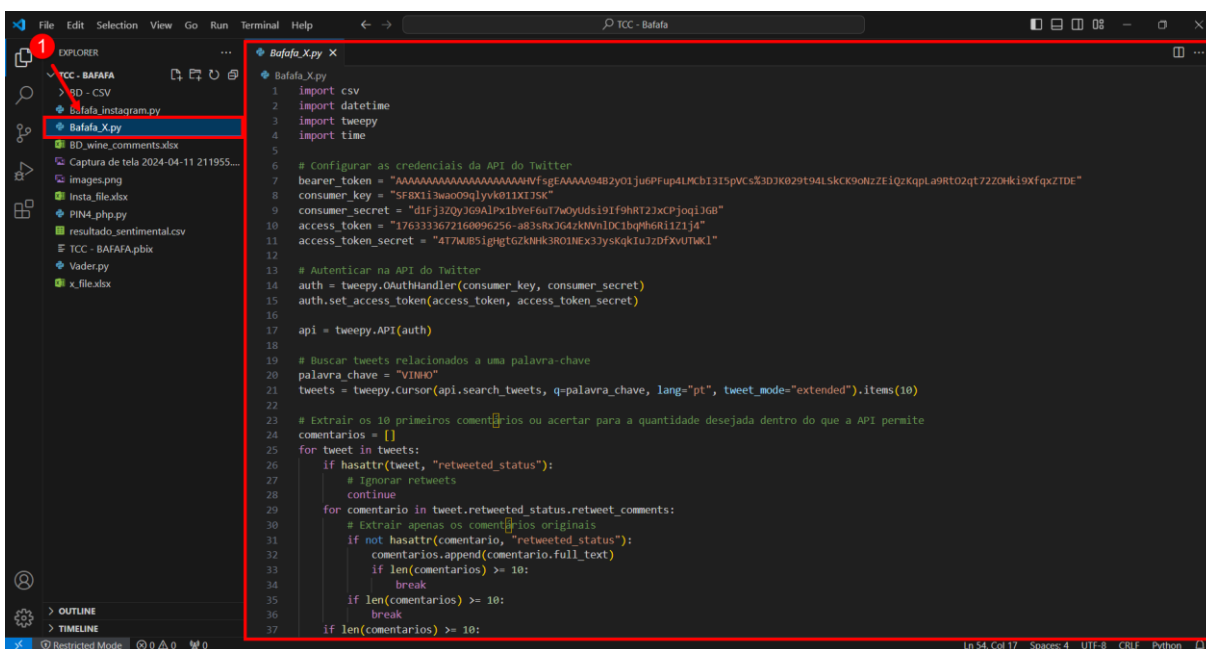
Após isso, será carregado no lado esquerdo do software todo o conteúdo da pasta “TCC – Bafafá”, no próximo tópico será abordado como informar a palavra-chave para pesquisa.

## 1.2. Alterando a palavra-chave a ser pesquisada:

A palavra-chave é o termo a ser buscado e auxilia no processo de mineração de dados do *software* para retornar o resultado esperado.

Na coluna de opções do lado esquerdo, deverá clicar na opção nomeada como **Bafafa\_X.py** para poder visualizar o código fonte de raspagem do X conforme a figura 13.

Figura 13. Tela código fonte do X do Bafafá



```

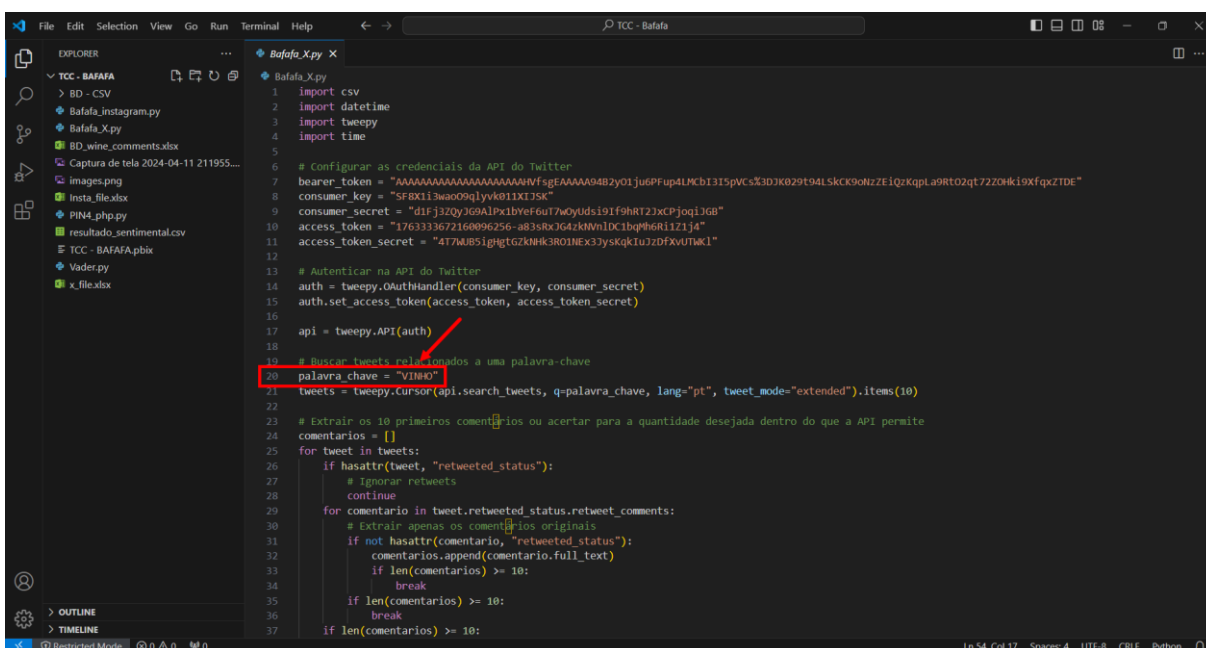
1 import csv
2 import datetime
3 import tweepy
4 import time
5
6 # Configurar as credenciais da API do Twitter
7 bearer_token = "AAAAAAAAAAAAAAAAAAAAHVsfEAAAAM94B2yO1ju6PFUp4LMCbI315pvcS3DJK029t94LSkck90NZZE1QzKqplA9RT02qT72Z0hk19XfqzZTDE"
8 consumer_key = "SF8X13wao09qlyvk011XI75K"
9 consumer_secret = "d1Fj3ZQyJG9AlPx1bveF6u7w0yUdsI9IF9hRT2JxCPjoq1JGB"
10 access_token = "1763333672160096256-a838rxJGdzkNwn1DC1bqMh6R11Z1j4"
11 access_token_secret = "4T7MUB5jghgtGzKkHk3RO1NE33JysKqk1uJzDFXvUTwK1"
12
13 # Autenticar na API do Twitter
14 auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
15 auth.set_access_token(access_token, access_token_secret)
16
17 api = tweepy.API(auth)
18
19 # Buscar tweets relacionados a uma palavra-chave
20 palavra_chave = "VINHO"
21 tweets = tweepy.Cursor(api.search_tweets, q=palavra_chave, lang="pt", tweet_mode="extended").items(10)
22
23 # Extrair os 10 primeiros comentários ou acertar para a quantidade desejada dentro do que a API permite
24 comentarios = []
25 for tweet in tweets:
26     if hasattr(tweet, "retweeted_status"):
27         # Ignorar retweets
28         continue
29     for comentario in tweet.retweeted_status.retweet_comments:
30         # Extrair apenas os comentários originais
31         if not hasattr(comentario, "retweeted_status"):
32             comentarios.append(comentario.full_text)
33             if len(comentarios) >= 10:
34                 break
35     if len(comentarios) >= 10:
36         break
37 if len(comentarios) >= 10:

```

Fonte: elaborado pelos autores (2024).

Na linha de código **número 20**, o usuário poderá alterar a palavra-chave que está informada dentro das aspas, no exemplo, “**VINHO**” para qualquer outro termo que deseja procurar – lembrando de certificar-se de que **não removeu acidentalmente as aspas do código**, exemplificado na figura 14.

Figura 14. Como informar a palavra-chave para pesquisa



```

1 import csv
2 import datetime
3 import tweepy
4 import time
5
6 # Configurar as credenciais da API do Twitter
7 bearer_token = "AAAAAAAAAAAAAAAAAAAAHVsfEAAAAM94B2yO1ju6PFUp4LMCbI315pvcS3DJK029t94LSkck90NZZE1QzKqplA9RT02qT72Z0hk19XfqzZTDE"
8 consumer_key = "SF8X13wao09qlyvk011XI75K"
9 consumer_secret = "d1Fj3ZQyJG9AlPx1bveF6u7w0yUdsI9IF9hRT2JxCPjoq1JGB"
10 access_token = "1763333672160096256-a838rxJGdzkNwn1DC1bqMh6R11Z1j4"
11 access_token_secret = "4T7MUB5jghgtGzKkHk3RO1NE33JysKqk1uJzDFXvUTwK1"
12
13 # Autenticar na API do Twitter
14 auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
15 auth.set_access_token(access_token, access_token_secret)
16
17 api = tweepy.API(auth)
18
19 # Buscar tweets relacionados a uma palavra-chave
20 palavra_chave = "VINHO"
21 tweets = tweepy.Cursor(api.search_tweets, q=palavra_chave, lang="pt", tweet_mode="extended").items(10)
22
23 # Extrair os 10 primeiros comentários ou acertar para a quantidade desejada dentro do que a API permite
24 comentarios = []
25 for tweet in tweets:
26     if hasattr(tweet, "retweeted_status"):
27         # Ignorar retweets
28         continue
29     for comentario in tweet.retweeted_status.retweet_comments:
30         # Extrair apenas os comentários originais
31         if not hasattr(comentario, "retweeted_status"):
32             comentarios.append(comentario.full_text)
33             if len(comentarios) >= 10:
34                 break
35     if len(comentarios) >= 10:
36         break
37 if len(comentarios) >= 10:

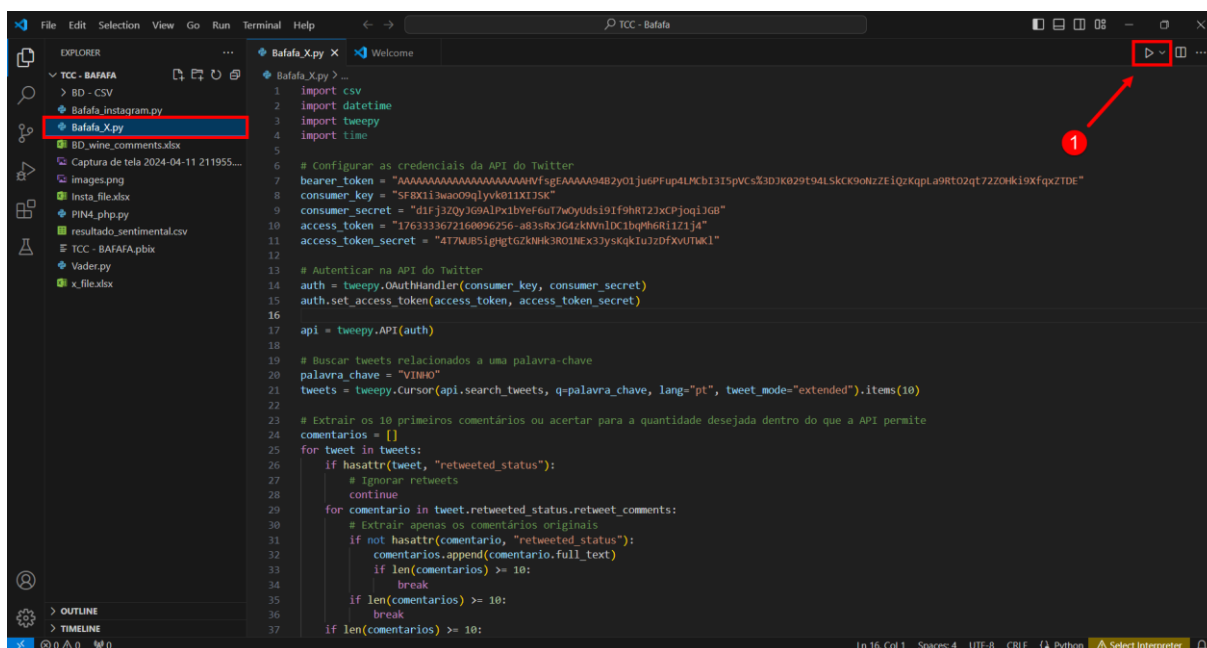
```

Fonte: elaborado pelos autores (2024).

### 1.3. Executando a pesquisa da palavra-chave nas redes sociais:

Para executar a pesquisa do programa, deve-se clicar no ícone de **Play** ▷ (“Iniciar”) localizado no canto superior direito da tela, conforme demonstrado na figura 15.

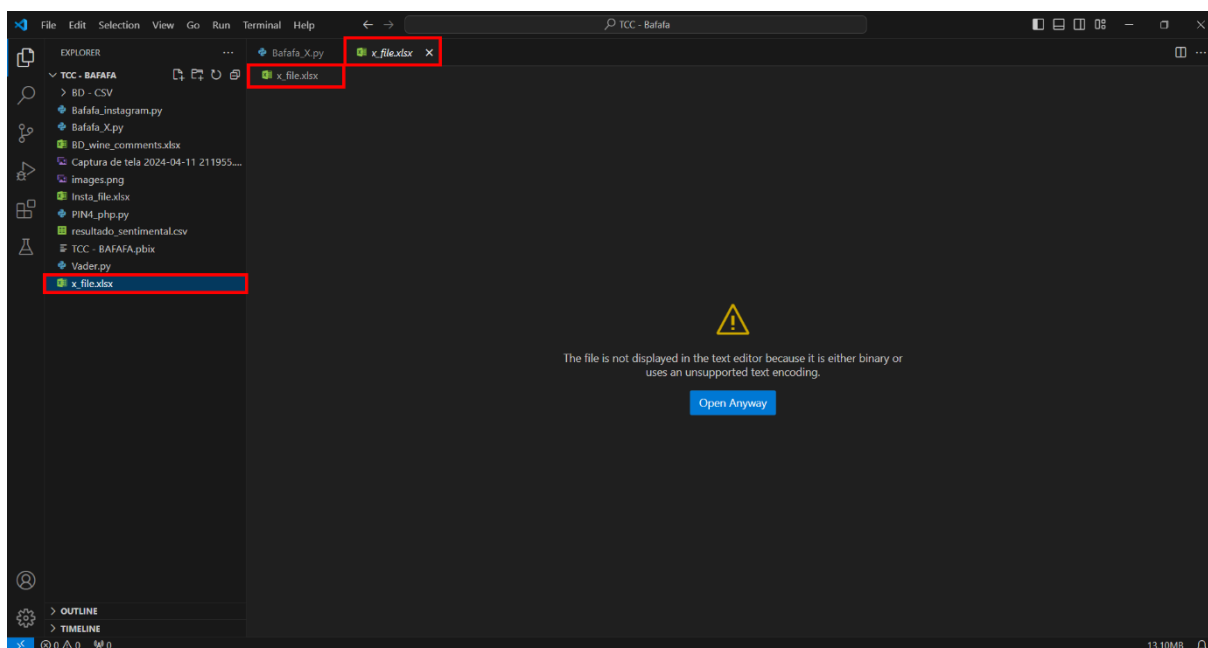
Figura 15. Como executar a pesquisa da palavra-chave na API do X



Fonte: elaborado pelos autores (2024).

Logo após o programa executar as raspagens das informações no X, será gerado um arquivo Excel (extensão .xlsx) nomeado como **x\_file.xlsx**, conforme a figura 16.

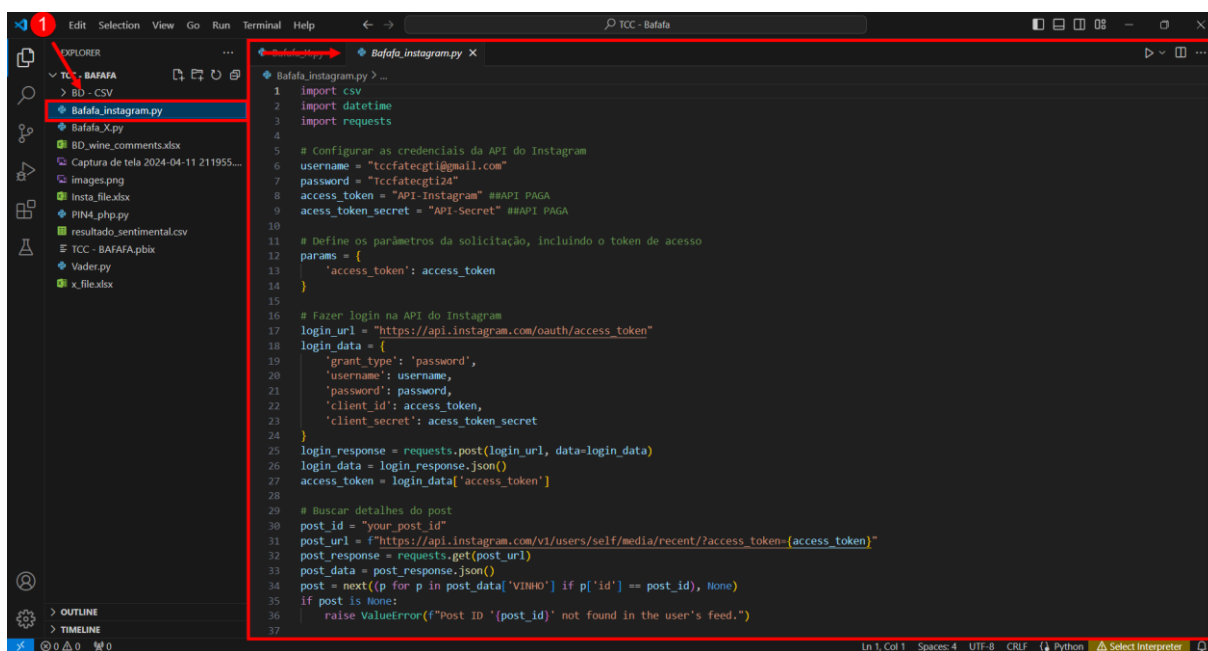
Figura 16. Arquivo .xlsx dos comentários coletados no X



Fonte: elaborado pelos autores (2024).

Depois de gerado o arquivo anterior, o usuário deverá prosseguir para o próximo passo, no menu lateral esquerdo deverá clicar na opção nomeada como **Bafafa\_instagram.py** para poder visualizar o código fonte da rede social Instagram conforme ilustrado na figura 17.

Figura 17. Tela código fonte do Instagram do Bafafá

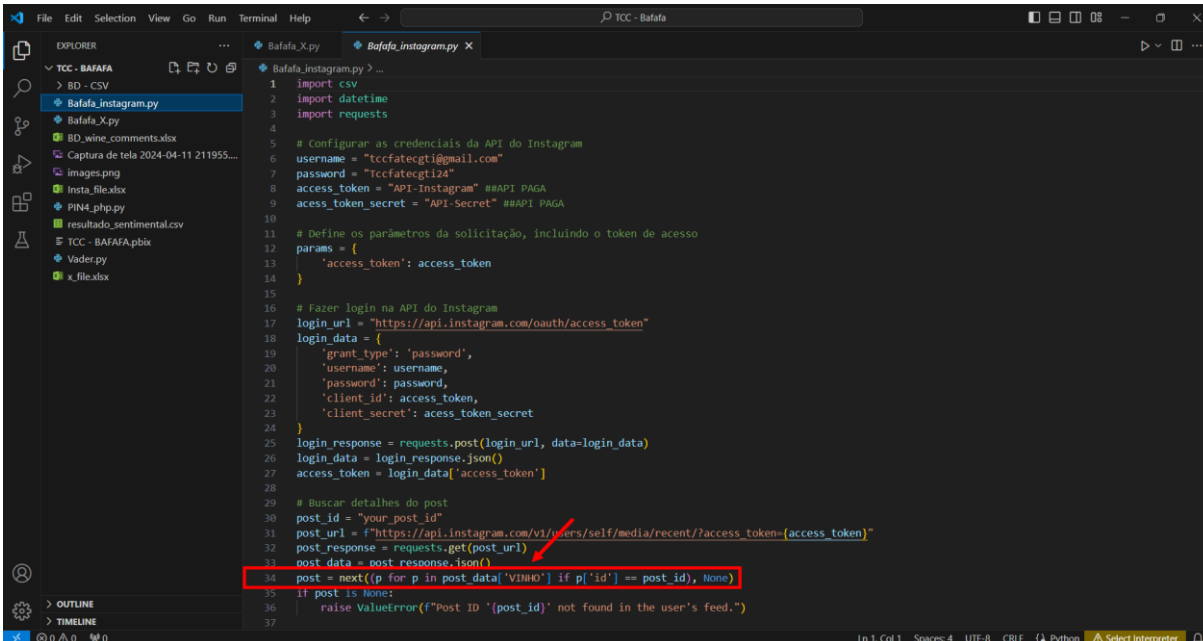


Fonte: elaborado pelos autores (2024).



Na linha de código **número 34** da variável **post**, o usuário poderá alterar a palavra-chave que está informada dentro dos colchetes e aspas simples, tal como: **['VINHO']** para qualquer outro termo que deseja procurar – certificando-se de que **não alterou acidentalmente a estrutura código**, como mostra a figura 18.

Figura 18. Como informar a palavra-chave para pesquisa



```

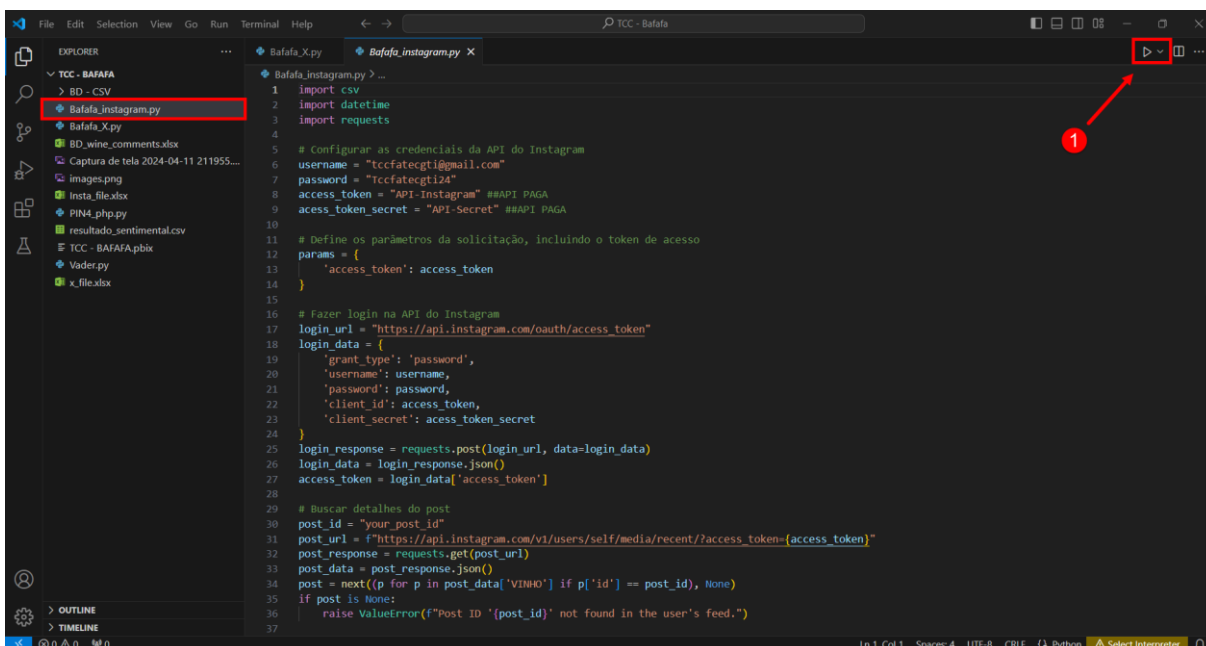
1 import csv
2 import datetime
3 import requests
4
5 # Configurar as credenciais da API do Instagram
6 username = "tccfatecti@gmail.com"
7 password = "tccfatecti24"
8 access_token = "API-Instagram" ##API PAGA
9 access_token_secret = "API-Secret" ##API PAGA
10
11 # Define os parâmetros da solicitação, incluindo o token de acesso
12 params = {
13     'access_token': access_token
14 }
15
16 # Fazer login na API do Instagram
17 login_url = "https://api.instagram.com/oauth/access_token"
18 login_data = {
19     'grant_type': 'password',
20     'username': username,
21     'password': password,
22     'client_id': access_token,
23     'client_secret': access_token_secret
24 }
25 login_response = requests.post(login_url, data=login_data)
26 login_data = login_response.json()
27 access_token = login_data['access_token']
28
29 # Buscar detalhes do post
30 post_id = "your_post_id"
31 post_url = f"https://api.instagram.com/v1/users/self/media/recent/?access_token={access_token}"
32 post_response = requests.get(post_url)
33 post_data = post_response.json()
34 post = next((p for p in post_data['posts'] if p['id'] == post_id), None)
35 if post is None:
36     raise ValueError(f"Post ID '{post_id}' not found in the user's feed.")
37

```

Fonte: elaborado pelos autores (2024).

Para poder executar a pesquisa dos comentários na rede social Instagram, novamente o usuário deverá clicar no ícone de **Play** ▷ (“Iniciar”) localizado no canto superior direito da tela segundo a figura 19.

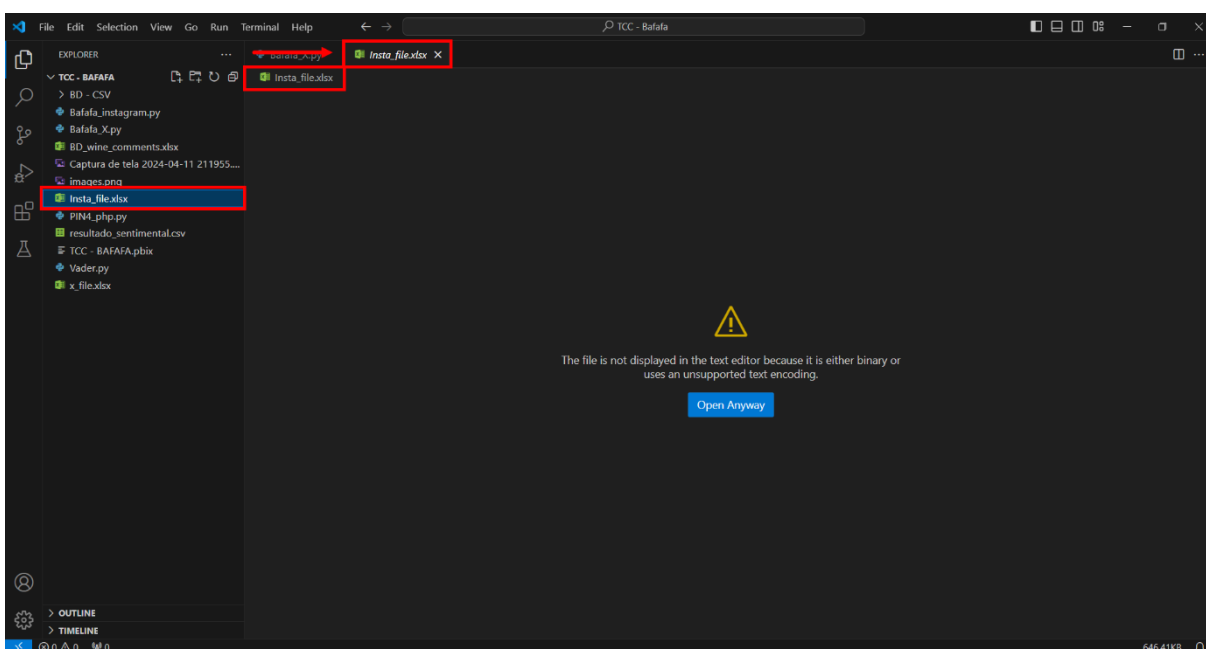
Figura 19. Como executar a pesquisa da palavra-chave na API do Instagram



Fonte: elaborado pelos autores (2024).

Após o executada as raspagens das postagens no Instagram, será gerado um arquivo Excel (extensão .xlsx) nomeado como **Insta\_file.xlsx**, de acordo com a figura 20.

Figura 20. Arquivo .xlsx dos comentários coletados no Instagram

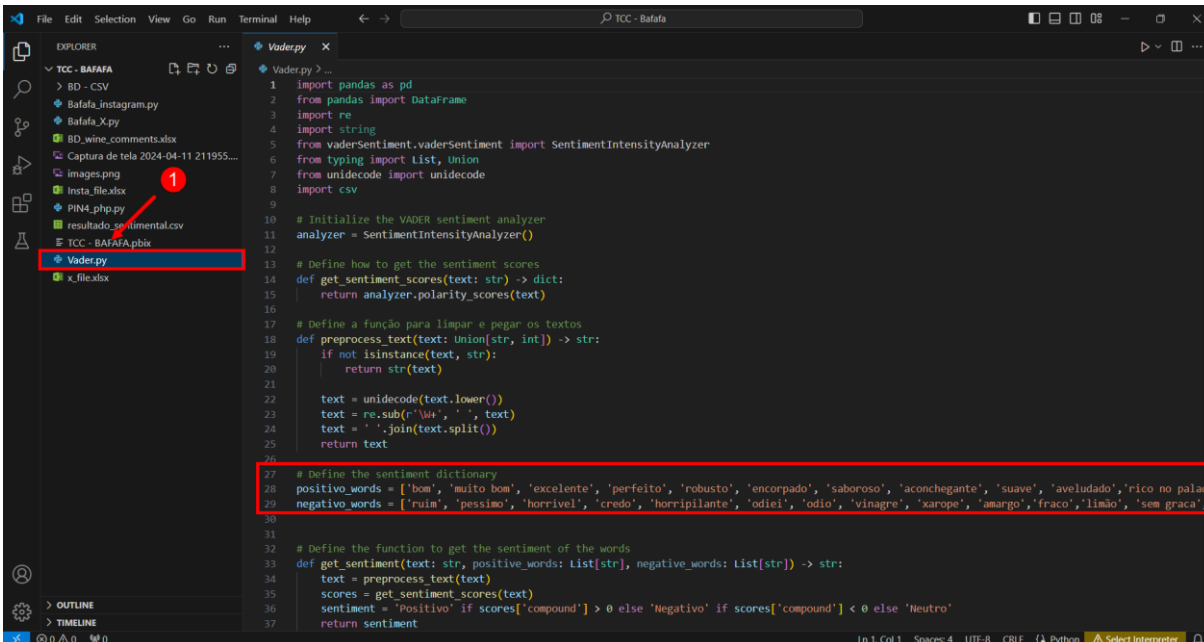


Fonte: elaborado pelos autores (2024).

## 2. AGRUPANDO OS ARQUIVOS EXPORTADOS

Logo após o usuário executar as alterações das palavras-chaves e realizar as exportações dos arquivos **.xlsx (Excel)**, deverá clicar na opção **VADER.py** e procurar no código fonte as **linhas 28 e 29**, nas quais, está disposto o Dicionário Léxico do Bafafá. E nestas linhas, dentro de colchetes, aspas simples e separadas por vírgulas, como por exemplo: “positivo\_words = [‘bom’, ‘muito bom’, ‘excelente’]” deverão inseridas as palavras Positivas “**positivo\_words**” e Negativas “**negativo\_words**” respectivamente, esses vocábulos irão auxiliar o programa a efetuar a sentimentalização dos comentários e informar se estes serão positivos, negativos ou neutros. Conforme exibido na figura 21.

Figura 21. Tela código fonte do VADER do Bafafá



```

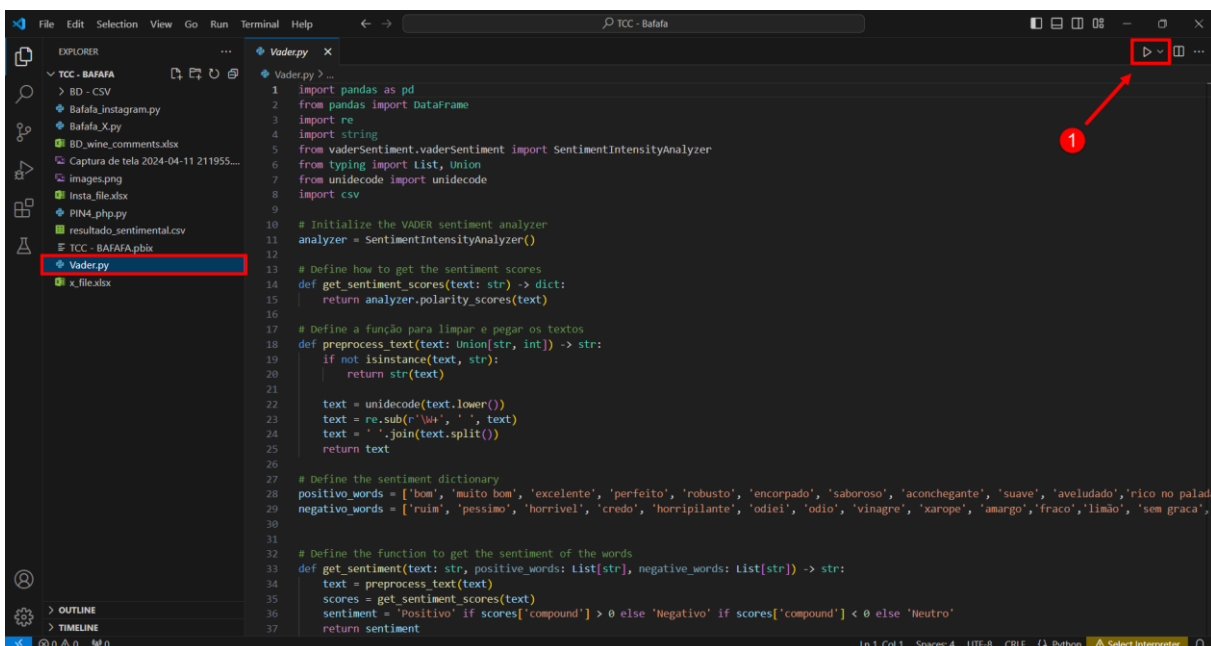
1 import pandas as pd
2 from pandas import DataFrame
3 import re
4 import string
5 from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
6 from typing import List, Union
7 from unicode import unicode
8 import csv
9
10 # Initialize the VADER sentiment analyzer
11 analyzer = SentimentIntensityAnalyzer()
12
13 # Define how to get the sentiment scores
14 def get_sentiment_scores(text: str) -> dict:
15     return analyzer.polarity_scores(text)
16
17 # Define a função para limpar e pegar os textos
18 def preprocess_text(text: Union[str, int]) -> str:
19     if not isinstance(text, str):
20         return str(text)
21
22     text = unicode(text.lower())
23     text = re.sub(r'\W+', ' ', text)
24     text = ' '.join(text.split())
25     return text
26
27 # Define the sentiment dictionary
28 positivo_words = ['bom', 'muito bom', 'excelente', 'perfeito', 'robusto', 'encorpado', 'saboroso', 'aconchegante', 'suave', 'aveludado', 'rico no palad
29 negativo_words = ['ruim', 'pessimo', 'horriovel', 'credo', 'horripilante', 'odiel', 'odio', 'vinagre', 'xarope', 'amargo', 'fraco', 'limao', 'sem graca',
30
31
32 # Define the function to get the sentiment of the words
33 def get_sentiment(text: str, positive_words: List[str], negative_words: List[str]) -> str:
34     text = preprocess_text(text)
35     scores = get_sentiment_scores(text)
36     sentiment = "Positivo" if scores['compound'] > 0 else "Negativo" if scores['compound'] < 0 else "Neutro"
37     return sentiment

```

Fonte: elaborado pelos autores (2024).

O usuário deverá clicar no ícone de **Play** ▷ (“Iniciar”) para executar o programa. Feito isso, o *software* agrupará os dois arquivos de Excel gerados e irá realizar o processo de sentimentalização destes. No qual, pode levar de 1 a 5 minutos dependendo do volume de informações contidas na amostra de dados coletados das redes sociais, ilustrado na figura 22.

Figura 22. Como executar a sentimentalização dos comentários



```

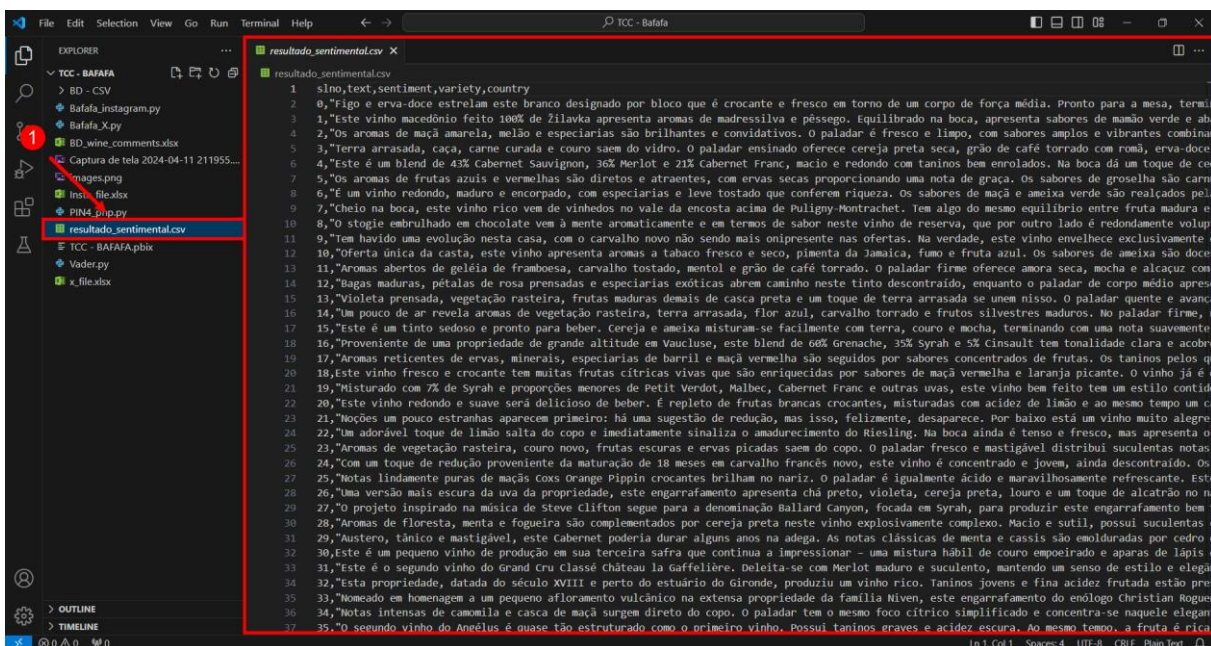
1 import pandas as pd
2 from pandas import DataFrame
3 import re
4 import string
5 from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
6 from typing import List, Union
7 from unicode import unicode
8 import csv
9
10 # Initialize the VADER sentiment analyzer
11 analyzer = SentimentIntensityAnalyzer()
12
13 # Define how to get the sentiment scores
14 def get_sentiment_scores(text: str) -> dict:
15     return analyzer.polarity_scores(text)
16
17 # Define a função para limpar e pegar os textos
18 def preprocess_text(text: Union[str, int]) -> str:
19     if not isinstance(text, str):
20         return str(text)
21
22     text = unicode(text.lower())
23     text = re.sub(r'\W+', '', text)
24     text = ' '.join(text.split())
25     return text
26
27 # Define the sentiment dictionary
28 positivo_words = ['bom', 'muito bom', 'excelente', 'perfeito', 'robusto', 'encorpado', 'saboroso', 'aconchegante', 'suave', 'aveludado', 'rico no palada
29 negativo_words = ['ruim', 'pessimo', 'horrible', 'credo', 'horripilante', 'odiei', 'odio', 'vinagre', 'xarope', 'amargo', 'fraco', 'limao', 'sem graca',
30
31
32 # Define the function to get the sentiment of the words
33 def get_sentiment(text: str, positivo_words: List[str], negativo_words: List[str]) -> str:
34     text = preprocess_text(text)
35     scores = get_sentiment_scores(text)
36     sentiment = 'Positivo' if scores['compound'] > 0 else 'Negativo' if scores['compound'] < 0 else 'Neutro'
37     return sentiment

```

Fonte: elaborado pelos autores (2024).

Em seguida, ao término da operação, será gerado um novo arquivo em Excel (extensão .CSV) nomeado como, denominado como **resultado\_sentimental.csv**, esse será o resultado do VADER já com os comentários tratados e finalizados. Explicitada figura 23 abaixo.

Figura 23. Arquivo .csv com os comentários tratados



```

1 sino, text, sentiment, variety, country
2 0, "Figo e erva-doce estrelam este branco designado por bloco que é crocante e fresco em torno de um corpo de força média. Pronto para a mesa, termi
3 1, "Este vinho macedônio feito 100% de Zilavka apresenta aromas de madressilva e pássego. Equilibrado na boca, apresenta sabores de mamão verde e abo
4 2, "Os aromas de maçã amarela, melão e especiarias são brilhantes e convidativos. O paladar é fresco e limpo, com sabores amplos e vibrantes combin
5 3, "Terra arrasada, caça, carne curada e couro saem do vidro. O paladar ensinado oferece cereja preta seca, grão de café torrado com romã, erva-doce
6 4, "Este é um blend de 43% Cabernet Sauvignon, 36% Merlot e 21% Cabernet Franc, macio e redondo com taninos bem enrolados. Na boca dá um toque de ces
7 5, "Os aromas de frutas azuis e vermelhas são diretos e atraentes, com ervas secas proporcionando uma nota de graça. Os sabores de groselha são carni
8 6, "É um vinho redondo, maduro e encorpado, com especiarias e leve tostado que conferem riqueza. Os sabores de maçã e ameixa verde são realçados pela
9 7, "Cheio na boca, este vinho rico vem de vinhedos no vale da encosta acima de Puligny-Montrachet. Tem algo do mesmo equilíbrio entre fruta madura e
10 8, "O stogie embrulhado em chocolate vem à mente aromáticamente e em termos de sabor neste vinho de reserva, que por outro lado é redondamente volup
11 9, "Tem havido uma evolução nesta casa, com o carvalho novo não sendo mais onipresente nas ofertas. Na verdade, este vinho envelhece exclusivamente e
12 10, "Oferta única da casta, este vinho apresenta aromas a tabaco fresco e seco, pimenta da Jamaica, fumo e fruta azul. Os sabores de ameixa são doces
13 11, "Aromas abertos de geleia de framboesa, carvalho tostado, mentol e grão de café torrado. O paladar firme oferece amora seca, mocha e alcaçuz com
14 12, "Bagas maduras, pétalas de rosa prensadas e especiarias exóticas abrem caminho neste tinto descontraído, enquanto o paladar de corpo médio apres
15 13, "Violeta prensada, vegetação rasteira, frutas maduras demais de casca preta e um toque de terra arrasada se unem nisso. O paladar quente e avança
16 14, "Um pouco de ar revela aromas de vegetação rasteira, terra arrasada, flor azul, carvalho torrado e frutos silvestres maduros. No paladar firme, r
17 15, "Este é um tinto sedoso e pronto para beber. Cereja e ameixa misturam-se facilmente com terra, couro e mocho, terminando com uma nota suavemente
18 16, "Proveniente de uma propriedade de grande altitude em Voucluse, este blend de 60% Grenache, 35% Syrah e 5% Cinsault tem tonalidade clara e acobre
19 17, "Aromas reticentes de ervas, minerais, especiarias de barril e maçã vermelha são seguidos por sabores concentrados de frutas. Os taninos pelos qu
20 18, "Este vinho fresco e crocante tem muitas frutas cítricas vivas que são enriquecidas por sabores do corpo vermelho e laranja picante. O vinho já é c
21 19, "Misturado com 7% de Syrah e proporções menores de Petit Verdot, Malbec, Cabernet Franc e outras uvas, este vinho bem feito tem um estilo contid
22 20, "Este vinho redondo e suave será delicioso de beber. É repleto de frutas brancas crocantes, misturadas com acidez de limão e ao mesmo tempo um ca
23 21, "Noções um pouco estranhas aparecem primeiro: há uma sugestão de redução, mas isso, felizmente, desaparece. Por baixo está um vinho muito alegre
24 22, "Um adorável toque de limão salta do copo e imediatamente sinaliza o amadurecimento do Riesling. Na boca ainda é tenso e fresco, mas apresenta o
25 23, "Aromas de vegetação rasteira, couro novo, frutas escuras e ervas picadas saem do copo. O paladar fresco e mastigável distribui suculentas notas
26 24, "Com um toque de redução proveniente da maturação de 18 meses em carvalho francês novo, este vinho é concentrado e jovem, ainda descontraído. Os
27 25, "Notas lindamente puras de maçãs Cox's Orange Pippin crocantes brilham no nariz. O paladar é igualmente ácido e maravilhosamente refrescante. Este
28 26, "Uma versão mais escura da uva da propriedade, este engarrafamento apresenta chá preto, violeta, cereja preta, louro e um toque de alcaçuz no na
29 27, "O projeto inspirado na música de Steve Clifton segue para a denominação Ballard Canyon, focada em Syrah, para produzir este engarrafamento bem t
30 28, "Aromas de floresta, menta e fogueira são complementados por cereja preta neste vinho explosivamente complexo. Macio e sutil, possui suculentas c
31 29, "Austero, tânico e mastigável, este Cabernet poderia durar alguns anos na adega. As notas clássicas de menta e cassis são emoluradas por cedro e r
32 30, "Este é um pequeno vinho de produção em sua terceira safra que continua a impressionar — uma mistura hábil de couro empoado e aparas de lápis e
33 31, "Este é o segundo vinho do Grand Cru Classé Château la Gaffelière. Deleita-se com Merlot maduro e suculento, mantendo um senso de estilo e elegar
34 32, "Esta propriedade, datada do século XVIII e perto do estuário do Gironde, produziu um vinho rico. Taninos jovens e fina acidez frutada estão prus
35 33, "Nomeado em homenagem a um pequeno afloramento vulcânico na extensa propriedade da família Niven, este engarrafamento do enólogo Christian Roguer
36 34, "Notas intensas de canela e casca de maçã surgem direto do copo. O paladar tem o mesmo foco cítrico simplificado e concentra-se naquele elegant
37 35, "O segundo vinho do Anéplus é quase tão estruturado como o primeiro vinho. Possui taninos graves e acidez escura. Ao mesmo tempo, a fruta é rica

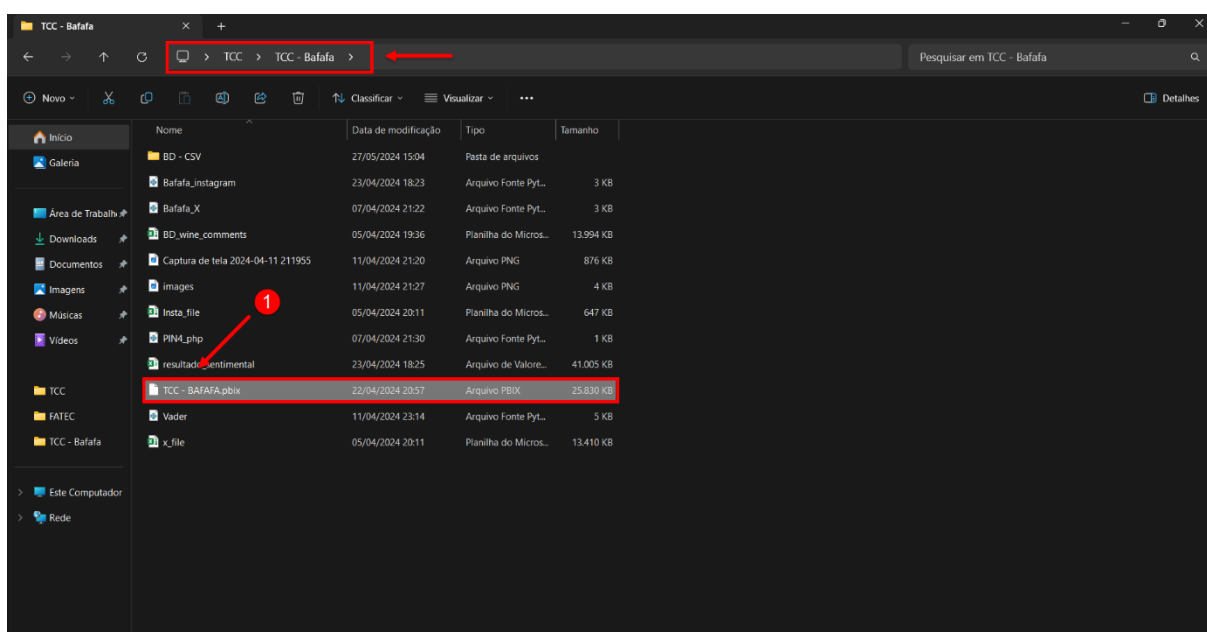
```

Fonte: elaborado pelos autores (2024).

### 3. VISUALIZAÇÃO – POWER BI

Posteriormente a exportação do arquivo final, o usuário deverá abrir o arquivo de visualização em Power BI (extensão .pbix) que estará localizado no mesmo diretório da pasta “**TCC - Bafafá**” importada inicialmente no VSC, este arquivo estará nomeado como **TCC - BAFafa.pbix**, como ilustrado na figura 24.

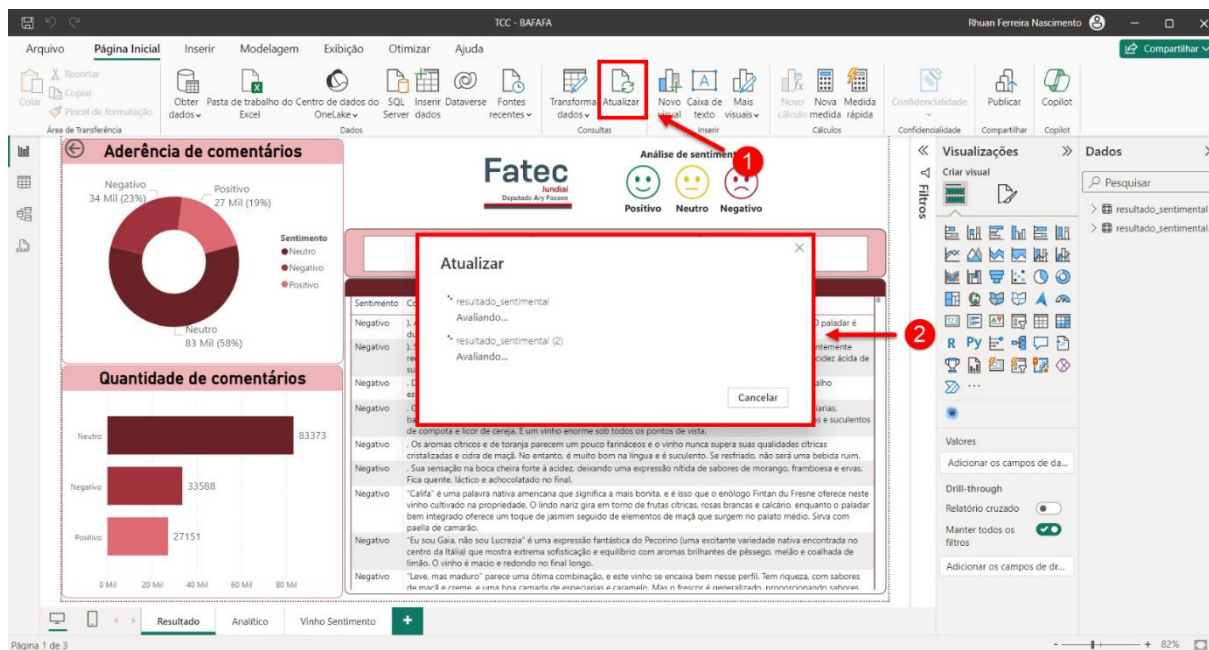
Figura 24. Arquivo Power BI do Bafafá



Fonte: elaborado pelos autores (2024).

Logo após abrir o arquivo no *software* Power BI, o usuário deverá clicar em atualizar e aguardar o processo finalizar, conforme a figura 25.

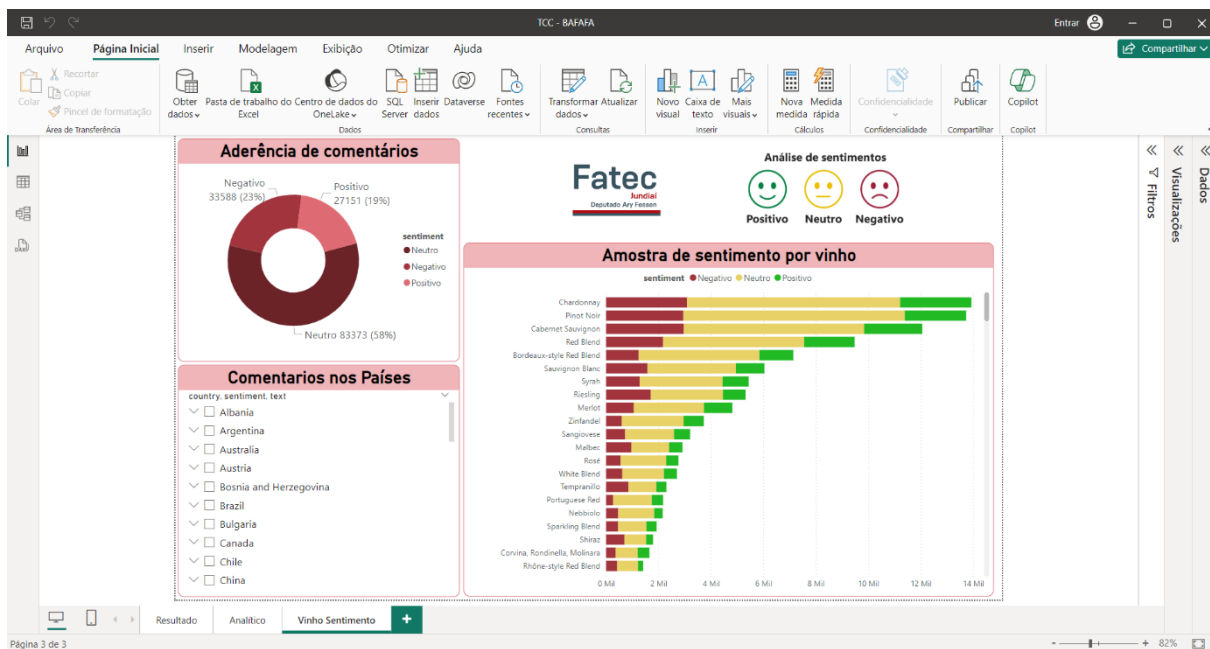
Figura 25. Como atualizar os dados do Power BI do Bafafá



Fonte: elaborado pelos autores (2024).

Depois da atualização, conforme consta na figura 26, o usuário irá visualizar as informações dos sentimentos processados de acordo com cada comentário sobre o termo e tema apresentados, sendo possível visualizar a porcentagem de cada sentimento sendo eles: positivo, negativo e neutro, os países onde foram realizadas as postagens e a barra de análise da amostra de sentimentos de acordo com as marcas analisadas.

Figura 26. Resultado visual dos comentários sentimentalizados



Fonte: elaborado pelos autores (2024).

## APÊNDICE B – CÓDIGO FONTE DO INSTAGRAM

```
import csv
import datetime
import requests
import instapy

# Configurar as credenciais da API do Instagram
username = "tccfatecgti@gmail.com"
password = "Tccf*****"
access_token = "API-Instagram" ##API PAGA
access_token_secret = "API-Secret" ##API PAGA

# Define os parâmetros da solicitação, incluindo o token de acesso
params = {
    'access_token': access_token
}

# Fazer login na API do Instagram
login_url = "https://api.instagram.com/oauth/access_token"
login_data = {
    'grant_type': 'password',
    'username': username,
    'password': password,
    'client_id': access_token,
    'client_secret': access_token_secret
}
login_response = requests.post(login_url, data=login_data)
login_data = login_response.json()
access_token = login_data['access_token']

# Buscar detalhes do post
```



```

post_id = "your_post_id"
post_url = f"https://api.instagram.com/v1/users/self/media/recent/?access_token={access_token}"
post_response = requests.get(post_url)
post_data = post_response.json()
post = next((p for p in post_data['VINHO'] if p['id'] == post_id), None)
if post is None:
    raise ValueError(f"Post ID '{post_id}' not found in the user's feed.")

# Extrair os comentários
comments_url = f"https://api.instagram.com/v1/media/{post_id}/comments?access_token={access_token}"
comments_response = requests.get(comments_url)
comments_data = comments_response.json()
comments = comments_data['data']

# Imprimir os comentários
for i, comment in enumerate(comments, start=1):
    print(f"Comentário {i}: {comment['text']}")

# Exportar os comentários para um arquivo CSV
now = datetime.datetime.now()
filename = f"Raspagem_{now.year}_{now.month}_{now.day}_{now.hour}_{now.minute}.csv"
with open(filename, "w", newline="", encoding="utf-8") as csvfile:
    fieldnames = ["Comentário"]
    writer = csv.DictWriter(csvfile, fieldnames=fieldnames)
    writer.writeheader()
    for i, comment in enumerate(comments, start=1):
        writer.writerow({"Comentário": f"Comentário {i}: {comment['text']}"})

```

#<https://developers.facebook.com/docs/instagram/>

## APÊNDICE C – CÓDIGO FONTE DO X

```

import csv
import datetime
import tweepy
import time

# Configurar as credenciais da API do Twitter
bearer_token = "AAAAAAAAAAAAAAAA***** ..."
consumer_key = "SF8X1i3waoO9q***** ..."
consumer_secret = "d1Fj3ZQyJG9AIPx1bYeF6u***** ..."
access_token = "1763333672160096256***** ..."
access_token_secret = "4T7WUB5igHgt***** ..."

# Autenticar na API do Twitter
auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)

api = tweepy.API(auth)

# Buscar tweets relacionados a uma palavra-chave
palavra_chave = "VINHO"
tweets = tweepy.Cursor(api.search_tweets, q=palavra_chave, lang="pt",
tweet_mode="extended").items(10)

# Extrair os 10 primeiros comentários ou acertar para a quantidade desejada dentro
do que a API permite
comentarios = []
for tweet in tweets:
    if hasattr(tweet, "retweeted_status"):
        # Ignorar retweets
        continue

```

```

for comentario in tweet.retweeted_status.retweet_comments:
    # Extrair apenas os comentários originais
    if not hasattr(comentario, "retweeted_status"):
        comentarios.append(comentario.full_text)
        if len(comentarios) >= 10:
            break
    if len(comentarios) >= 10:
        break
if len(comentarios) >= 10:
    break

# Imprimir os comentários
for i, comentario in enumerate(comentarios, start=1):
    print(f"Comentário {i}: {comentario}")

# Exportar os comentários para um arquivo CSV
now = datetime.datetime.now()
filename =
f"Raspagem_{now.year}{now.month}{now.day}{now.hour}{now.minute}.csv"
with open(filename, "w", newline="", encoding="utf-8") as csvfile:
    fieldnames = ["Comentário"]
    writer = csv.DictWriter(csvfile, fieldnames=fieldnames)
    writer.writeheader()
    for i, comentario in enumerate(comentarios, start=1):
        writer.writerow({"Comentário": f"Comentário {i}: {comentario}"})

#https://developer.twitter.com/en/docs/twitter-api/getting-started/about-twitter-api

```

## APÊNDICE D – CÓDIGO FONTE DO VADER

```
import pandas as pd
from pandas import DataFrame
import re
import string
from VADERSentiment.VADERSentiment import SentimentIntensityAnalyzer
from typing import List, Union
from unidecode import unidecode
import csv

# Iniciar VADER sentimentalizador
analyzer = SentimentIntensityAnalyzer()

# Define como serão analisados os "scores" dos textos
def get_sentiment_scores(text: str) -> dict:
    return analyzer.polarity_scores(text)

# Define a função para limpar e pegar os textos
def preprocess_text(text: Union[str, int]) -> str:
    if not isinstance(text, str):
        return str(text)

    text = unidecode(text.lower())
    text = re.sub(r'\W+', ' ', text)
    text = ' '.join(text.split())
    return text

# Define o dicionário léxico
positivo_words = ['bom', 'muito bom', 'excelente', 'perfeito', 'robusto', 'encorpado',
'saboroso', 'aconchegante', 'suave', 'aveludado', 'rico no paladar', 'frutado',
```

```
'harmonioso', 'refrescante', 'delicioso', 'sedutor', 'complexo', 'elegante', 'aromático',
'persistente', 'suculento','harmonizado','maduro']
negativo_words = ['ruim', 'pessimo', 'horriovel', 'credo', 'horripilante', 'odiei', 'odio',
'vinagre', 'xarope', 'amargo','fraco','limão', 'sem graca', 'sem gosto', 'sofrido', 'angustia',
'angustiante', 'azedo', 'ácido', 'amadeirado em excesso', 'agressivo', 'adstringente',
'desbalanceado', 'plano', 'áspero', 'diluído', 'insípido', 'desinteressante', 'monótono',
'desagradável']
```

```
# Define como serão as notas dos comentarios
```

```
def get_sentiment(text: str, positive_words: List[str], negative_words: List[str]) -> str:
    text = preprocess_text(text)
    scores = get_sentiment_scores(text)
    sentiment = 'Positivo' if scores['compound'] > 0 else 'Negativo' if scores['compound']
< 0 else 'Neutro'
    return sentiment
```

```
# Define como será aplicado o sentimento
```

```
def analyze_sentiment(df: DataFrame, column: str) -> DataFrame:
    df[column] = df[column].astype(str) # convert the column to string type
    df['sentiment'] = df[column].apply(get_sentiment, args=(positivo_words,
negativo_words))
    return df
```

```
# Define como será pego as informações do excel
```

```
def analyze_excel_file(file_path: str, sheet_name: str, column: str, variety_column: str,
country_column: str) -> DataFrame:
    df = pd.read_excel(file_path, sheet_name=sheet_name)
    df = analyze_sentiment(df, column)
    df['variety'] = df[variety_column]
    df['country'] = df[country_column]
    return df
```

```
# Define a análise do que está no excel
def analyze_sentiment_of_excel_files(file_path1: str, sheet_name1: str, column1: str,
variety_column1: str, country_column1: str, file_path2: str, sheet_name2: str, column2:
str, variety_column2: str, country_column2: str) -> tuple[DataFrame, DataFrame]:
    df1 = analyze_excel_file(file_path1, sheet_name1, column1, variety_column1,
country_column1)
    df2 = analyze_excel_file(file_path2, sheet_name2, column2, variety_column2,
country_column2)
    return df1, df2

# Define os caminhos dos arquivos Excel
file_path1 = 'x_file.xlsx'
file_path2 = 'insta_file.xlsx'

# Define os nomes das planilhas
sheet_name1 = 'Planilha1'
sheet_name2 = 'Planilha1'

# Define a coluna a ser analisada em cada arquivo
column1 = 'Tradução'
column2 = 'Tradução'

# Define a coluna variety em cada arquivo
variety_column1 = 'variety'
variety_column2 = 'variety'

# Define a coluna country em cada arquivo
country_column1 = 'country'
country_column2 = 'country'

# Chama a função de análise
```

```
df1, df2 = analyze_sentiment_of_excel_files(file_path1, sheet_name1, column1,
variety_column1, country_column1, file_path2, sheet_name2, column2,
variety_column2, country_column2)
```

```
# Define os nomes das colunas no arquivo exportado
```

```
fieldnames = ['sln0', 'text', 'sentiment', 'variety', 'country']
```

```
# Exporta o CSV
```

```
with open('resultado_sentimental.csv', mode='w', newline="", encoding='utf-8') as f:
```

```
    writer = csv.DictWriter(f, fieldnames=fieldnames)
```

```
    writer.writeheader()
```

```
    for i, row in df1.iterrows():
```

```
        writer.writerow({'sln0': i, 'text': row['Tradução'], 'sentiment': row['sentiment'],
'variety': row['variety'], 'country': row['country']})
```

```
    for i, row in df2.iterrows():
```

```
        writer.writerow({'sln0': i, 'text': row['Tradução'], 'sentiment': row['sentiment'],
'variety': row['variety'], 'country': row['country']})
```