

CENTRO UNIVERSITÁRIO DE ANÁPOLIS – UniEVANGÉLICA
BACHARELADO EM ENGENHARIA DE SOFTWARE

DESENVOLVIMENTO DE UMA APLICAÇÃO QUE ANALISARÁ DADOS DO
TWITTER.

THYERRE DOS SANTOS ESBALTAR

Anápolis
2021

THYERRE DOS SANTOS ESBALTAR

**DESENVOLVIMENTO DE UMA APLICAÇÃO QUE ANALISARÁ DADOS DO
*TWITTER.***

Trabalho de Conclusão de Curso I apresentado como requisito parcial para a conclusão da disciplina de Trabalho de Conclusão de Curso I do curso de Bacharelado em Engenharia de Software do Centro Universitário de Anápolis – UniEVANGÉLICA.

Orientador(a): Prof.^a. Ms. Aline Dayany Lemos.

Anápolis
2021

THYERRE DOS SANTOS ESBALTAR

**DESENVOLVIMENTO DE UMA APLICAÇÃO QUE ANALISARÁ DADOS DO
*TWITTER.***

Trabalho de Conclusão de Curso I apresentado como requisito parcial para a obtenção de grau do curso de Bacharelado em Engenharia de Software do Centro Universitário de Anápolis – UniEVANGÉLICA.

Aprovado(a) pela banca examinadora em [dia] de [mês] de 2021, composta por:

Prof. [nome do professor]
Orientador

Prof. [nome do professor]

Resumo

É da natureza humana valorizar as opiniões e percebe-se que com o avanço da *internet* o compartilhamento dela tornou-se mais abrangente sendo capaz de alcançar pessoas de todo o mundo. Essas opiniões que são compartilhadas na *internet* por meio de aplicações de redes sociais, demonstram-se bastante importantes, haja vista que as opiniões têm um amplo alcance e a possibilidade de impactar a marca da *e-commerce* de forma positiva ou negativa. A fim de auxiliar no entendimento de como está a situação da marca do *e-commerce*, faz-se necessário a utilização da análise de emoções, onde seria realizada uma investigação sobre opinião compartilhada e retornaria à classificação da mesma. Para a produção da aplicação foi analisada e escolhida a linguagem de programação *Python* nela fazemos uso de diversas bibliotecas que tem como objetivo auxiliar no desenvolvimento do código, dentre essas bibliotecas podemos citar a *tweepy*, *preprocessor*, *NLTK*, *pandas* e *matplotlib*. Este estudo busca apresentar a possibilidade de aplicação da análise de emoções utilizando a abordagem de aprendizagem de máquina empregando o classificador *Naive Bayes* sobre frases coletadas na rede social *Twitter* que foram destinadas a somente um *e-commerce*.

Palavras-chave: *Python*. *Twitter*. Análise de emoções.

Lista de Ilustrações

Figura 1 – Primeiros passos para a coleta.....	23
Figura 2 – Inserindo os dados nas variáveis.....	24
Figura 3 - Código para salvar no formato .txt.....	25
Figura 4 - Código para limpar as frases.....	25

Lista de abreviaturas e siglas

API	<i>Application Programming Interface</i>
NLTK	<i>Natural Language Toolkit</i>
PLN	Processamento de Linguagem Natural
SVM	<i>Support Vector Machines</i>
URL	<i>Uniform Resource Locator</i>

SUMÁRIO

1.	PROBLEMA.....	8
2.	OBJETIVOS.....	9
2.1	Objetivo Geral.....	9
2.2	Objetivos Específicos.....	9
3.	JUSTIFICATIVA.....	10
4.	FUNDAMENTAÇÃO TEÓRICA.....	11
4.1	Redes Sociais.....	11
4.1.1	<i>Twitter</i>	12
4.2	Análise de sentimentos.....	12
4.3	Processamento de Linguagem Natural.....	13
4.4	Classificadores de sentimentos.....	14
4.4.1	Abordagem de aprendizado de máquina.....	14
4.4.1.1	Classificador <i>Naive Bayer</i>	15
4.4.1.2	Entropia Máxima.....	16
4.4.1.3	Rede Bayesiana.....	16
4.4.1.4	Rede Neural.....	16
4.4.1.5	SVM.....	17
4.4.1.6	Modelo baseado em regras.....	17
4.4.1.7	Árvore de decisão.....	18
4.5	<i>Python</i>	18
4.6	NLTK.....	19
5.	METODOLOGIA.....	20
6.	CRONOGRAMA.....	22
7.	RESULTADOS ALCANÇADOS.....	23
8.	RESULTADOS ESPERADOS.....	27
	REFERÊNCIA BIBLIOGRÁFICAS.....	28

1. PROBLEMA

Com a *Internet*, surgiu um meio no qual as pessoas são capazes de compartilhar opiniões, onde elas podem ser vistas em qualquer lugar do mundo, com isto cada vez mais pessoas começam a utilizar as redes sociais *on-line*, segundo dados de 2012 aproximadamente 96% da população *on-line* utilizava redes sociais (MARQUES; PINTO; ALVAREZ, 2016).

O *Twitter* tornou-se uma das maiores redes sociais do mundo, em junho de 2012 cerca de 41.200.000 pessoas produziram cerca de 2,3 bilhões de *tweets* expondo suas opiniões, considerando somente as cidades de Rio de Janeiro e São Paulo (SEMIOCAST, 2012). Essa rede social é o local onde os usuários podem divulgar todos os tipos de informações em tempo real por meio de mensagens, inicialmente essas mensagens tinham limite de 140 caracteres, mas atualmente o limite é de 280 caracteres (NASCIMENTO; OSIEK; XEXÉO, 2013).

As opiniões são de extrema importância para os humanos, pois muitas das vezes com o auxílio delas decidem se irão comer em um determinado restaurante, se irão comprar determinado carro ou em qual *e-commerce* farão as compras *on-line*. De acordo com Liu (2012), as opiniões são extremamente importantes para quase todas as atividades humanas tendo em vista que são os grandes influenciadores de nosso comportamento. Segundo Becker (2014), empresas baseiam suas estratégias de negócios e *marketing* na opinião de seus clientes em relação aos seus produtos ou serviços.

Levando em consideração a importância das opiniões, a grande utilização das redes sociais e o que isto proporciona para as *e-commerce* em possibilidades de coleta de opinião sobre a sua marca, podemos levantar a seguinte questão: É possível descobrir as opiniões de clientes através de mensagens escritas em redes sociais?

2. OBJETIVOS

2.1 Objetivo Geral

Desenvolver uma aplicação que fará uma análise de emoções nos dados coletados da rede social *Twitter* com o objetivo de demonstrar o ponto de vista dos clientes em relação ao *e-commerce*.

2.2 Objetivos Específicos

- Identificar métodos e técnicas de mineração de dados direcionado para análise de emoções;
- Desenvolver uma aplicação utilizando a linguagem de programação *Python*;
- Realizar a análise de emoções direcionado as emoções em tweets acerca de clientes e *e-commerces*;

3. JUSTIFICATIVA

As opiniões dos clientes são de grande importância para as empresas, visto que podem apresentar o ponto de vista do cliente acerca da empresa, ser capaz de analisar essas opiniões provê a possibilidade da empresa de descobrir qual a receptividade para com clientes. De acordo com Antunes (2014), as redes sociais vêm ganhando destaque no setor privado, porque permitem uma melhor comunicação entre as empresas e os clientes possibilitando um melhor levantamento de informações referente a receptividade dos consumidores em relação aos produtos e serviços.

Opiniões negativas são capazes de prejudicar a reputação de qualquer indivíduo. Para as empresas esse tipo de opinião demonstra ser igualmente importante ou quiçá mais, haja vista que, tem potencial de influenciar negativamente os consumidores, podendo ocasionar a perda de um possível futuro cliente. Segundo Evangelista e Padilha (2014), as empresas já têm a noção de que um comentário desfavorável pode atingir uma imensa quantidade de clientes de forma rápida e dessa maneira afetar a imagem da empresa no mercado.

De acordo com Gomes (2012), algumas empresas também já sabem como tirar vantagens dessas opiniões compartilhadas em redes sociais, por meio da análise das opiniões que foram compartilhadas em diversos canais de comunicação. O resultado dessa análise de opinião pode possibilitar à empresa um investimento de recursos de maneira eficiente e assertiva.

4. FUNDAMENTAÇÃO TEÓRICA

Neste tópico serão apresentados conceitos sobre redes sociais, análise de sentimentos, aprendizagem de máquina, processamento de linguagem natural. Também será apresentado algumas das principais tecnologias que serão utilizadas durante o desenvolvimento.

4.1 Redes sociais

As redes sociais são aplicações que fazem-se amplamente utilizadas por pessoas com acesso à *internet*. Conforme dito por Marteleto (2001) as redes sociais, representam “[...] um conjunto de participantes autônomos, unindo ideias e recursos em torno de valores e interesses compartilhados”. Para Souza e Quandt (2008), as redes sociais são estruturas complexas e dinâmicas constituídas por pessoas com princípios ou com um propósito em comum.

Segundo Recuero (2009), a rede social pode ser definida como um grupo de dois elementos, sendo atores que consistem em pessoas, grupos ou instituições e suas conexões que seriam os laços sociais ou interações.

De acordo com Kietzmann et al. (2011) as redes sociais têm as seguintes características:

- **Identidade:** representa o local onde os usuários revelam suas identidades no ambiente das redes sociais.
- **Conversas:** representam a extensão da comunicação entre os usuários na rede social.
- **Compartilhamento:** representa a possibilidade dos usuários de distribuírem e receberem conteúdo.
- **Presença:** trata-se dos outros usuários notarem que alguém se encontra *on-line* na rede social.
- **Relações:** representa até onde os usuários relacionam entre si.
- **Reputação:** seria a oportunidade de um usuário checar a sua popularidade e a de outros usuários nas redes sociais, por meio de quantidade de seguidores, *likes*, postagem de fotos e vídeos.
- **Grupos:** representa a possibilidade da formação de comunidades e grupos nas redes sociais.

Essas características das redes sociais proporcionaram-na sua expansão, tornando-se um local onde possibilita o compartilhamento de diversos tipos de opiniões e conteúdos, a formação de diversas comunidades e grupos com vários tipos de temas e objetivos.

As redes sociais provaram ser um local que produz enormes volumes de dados, isto ocorre devido a sua grande quantidade de usuários e ao fato de ser um local no qual o compartilhamento de conteúdo é uma de suas principais funções. Conforme foi dito no Simpósio Brasileiro de Banco de Dados (2014) “A popularidade dessas plataformas pode ser evidenciada através da capacidade que possuem de produzir enormes volumes de conteúdo”.

4.1.1 Twitter

O *Twitter* é uma rede social que comporta um mundo de diferentes tipos de pessoas, ideias, pontos de vista e informações. O *Twitter* foi criado em março de 2006, sendo uma rede social gratuita que permite o compartilhamento de opiniões, informações e outros tipos de conteúdos.

Barbosa et al. (2012) afirma que o padrão de interação do *Twitter* faz com que o usuário sinta a necessidade de compartilhar e expressar de maneira contínua sua opinião e sentimento. Por ser em tempo real, fácil e rápido publicar no *Twitter*, os usuários interagem mais com a rede social (ZHAO; ROSSON, 2009).

O *Twitter* é utilizado por seu usuário não apenas como uma ferramenta que tem o único propósito de divulgar informações pessoais, mas também o empregam para compartilhar opiniões e conhecimentos sobre eventos e fatos em geral (NAAMAN; BOASE, 2010). Conforme Java et al., (2007), o *Twitter* é uma ferramenta de grande importância para dois tipos de usuário: os que procuram informação e os que querem compartilhar amplamente informações.

4.2 Análise de sentimentos

A utilização da análise de sentimentos, também conhecida como mineração de opinião, visa a captação da emoção predominante do seu público-alvo. A análise de sentimentos é uma ferramenta que se tornou valiosa para empresas, políticos, influenciadores e entre outros. “Empresas, eventos (e.g. Olimpíadas), personalidades, estão interessadas na compreensão de como são percebidas pelo público em geral em tempo real, e nas mais variadas mídias” (BECKER, 2013).

De acordo com o autor Liu (2012) a análise de sentimentos é uma área de estudo que tem como objetivo a análise das emoções das pessoas, avaliações, atitudes, emoções e sentimentos relacionados a entidades como organizações, serviços, produtos, indivíduos, eventos, questões, tópicos e seus atributos.

Visando o grande fator de importância que as opiniões têm para o comportamento do ser humano, a análise de sentimentos é uma ferramenta de grande valia, pois ambicionando uma melhor análise e um investimento dos recursos de forma mais correta da entidade, essa análise da opinião ganha grande relevância.

A análise de sentimentos tem como fundamental objetivo classificar, por intermédio da apuração das palavras e de maneira automática, os sentimentos que foram expressados no texto. De acordo com Medhat, Hassen e Korashy (2014), o propósito da análise de sentimentos é obter opiniões, identificar os sentimentos expressados nela e classificar sua polaridade.

A utilização da análise de sentimentos implementada para analisar frases que são postadas em redes sociais proporcionam uma forma automatizada de averiguar o ponto de vista do cliente para com a empresa. A utilização desses dados pode amparar nas tomadas de decisões, fazendo com que essas sejam mais eficientes e detenham maiores chances de serem bem-sucedidas.

4.3 Processamento de Linguagem Natural

O Processamento de Linguagem Natural (PLN) trata-se de um ramo da pesquisa da inteligência artificial que busca maneiras de como computadores poderiam ser utilizados para manipular e entender a linguagem natural. Conforme Allen (2003), o desafio básico enfrentado em PLN refere-se a ambiguidade que está presente na linguagem humana. Brascher (2002) salienta e apresenta motivos para tais ambiguidades na linguagem natural:

- Ambiguidade lexical: ocorre quando não se é possível interpretar o significado da unidade lexical. Exemplo da palavra “banco”, a palavra pode se referir ao banco de uma praça ou a instituição bancária.
- Ambiguidade semântica: ocorre quando se tem várias interpretações possíveis para o relacionamento dos termos na frase. Exemplo: "Um rio corre através de cada país europeu." A frase significa que apenas um rio corre através de cada país europeu, ou vários rios correm por entre diferentes países europeus.

- Ambiguidade pragmática: se relaciona ao que está ligado à situação do locutor no momento da enunciação. Exemplo na frase “Paulo vai à escola.” o sujeito seria um estudante ou está apenas indo à escola naquele momento?
- Ambiguidade predicativa: se dá por meio da interpretação das relações temáticas que encadeiam o predicado, participantes e argumentos. Exemplo na frase “A crítica deste autor.” o autor seria o objeto da crítica, ou seria o agente da crítica.
- Ambiguidade sintática: se dá no momento da estruturação das frases em constituintes hierarquizados. Exemplo a frase “A professora de dança espanhola.” a professora é espanhola ou da dança que é espanhola.

4.4 Classificadores de sentimentos

Há diversas formas de classificar sentimentos em textos. Segundo Maynard e Funk (2011), as técnicas de detecção de sentimentos seriam divididas basicamente em abordagem baseada em léxico, abordagem de aprendizagem de máquina e abordagem híbrida. Nos subtópicos abaixo discutiremos sobre a abordagem de aprendizagem de máquina e sobre o classificador *Naive Bayes*.

4.4.1 Abordagem de aprendizado de máquina

Na abordagem de aprendizagem de máquina é desenvolvido um modelo de classificação onde uma mensagem está associada a uma polaridade ou sentimento, na seguinte frase “Comprei uma TV e chegou voando! Tenho muita simpatia por essa loja” o sentimento associado ao *e-commerce* é de satisfação.

Os dados que utilizamos para a produção deste modelo são chamados de dados de treinamento e normalmente são formados por uma mensagem e cada uma está vinculada a uma classe, podendo ser um sentimento ou polaridade (*COMPUTER ON THE BEACH*, 2018).

Podemos dividir os métodos de classificação que utilizam a abordagem de aprendizagem de máquina em métodos de aprendizagem supervisionados e não supervisionados Medhat, Hassan e Korashy (2014).

A abordagem de aprendizagem supervisionada necessita de uma coleção prévia de treinamento, ou seja, depende de uma numerosa quantidade de documentos classificados. As técnicas que podem ser utilizadas em aprendizagem supervisionada com modelos probabilísticos são *Naive Bayes*, Máxima Entropia e Redes Bayesianas, ou com modelos

lineares Rede Neural e SVM e ainda modelos baseados em regras e técnica como árvore de decisão (WILSON; WIEBE; HOFFMANN 2008; MEDHAT; HASSAN; KORASHY 2014; TAN, STEINBACH; KUMAR 2009; ZHANG; ZHOU 2010).

Conforme citado no evento *Computer on the Beach* (2018), o método não supervisionado não necessita da coleção prévia de treinamento, o método utiliza uma coleção de palavras, dicionários ou estatística para a classificação. Este método tem como principal vantagem a independência do domínio, isto é, este modelo não tem como base um conjunto limitado de treinamento (SONI; PATEL, 2014).

4.4.1.1 Classificador *Naive Bayer*

Como descrito no item 4.4.1 o Naive Bayes é um dos classificadores do método de aprendizagem supervisionada. Segundo Costa e Lopes (20-?), o motivo deste classificador ser bastante utilizado é em função do mesmo assumir que o valor do atributo é independente, ou seja, o valor de um atributo não sofre influência pelo valor de outro atributo.

O *Naive Bayes* é um classificador probabilístico que foi construído tendo como base o teorema de *Bayes*. O teorema de *Bayes* utiliza o teorema para calcular a probabilidade de que uma frase pertença a uma classe específica (MEDHAT; HASSAN; KORASHY, 2014).

$$P(C|F) = (P(F|C) * P(C)) / P(F)$$

Onde: $P(C)$ é a probabilidade da ocorrência do C, $P(F)$ é a probabilidade de F ter ocorrido, $P(F|C)$ é a probabilidade do F ocorrer tendo o C já ocorrido e $P(C|F)$ será o resultado da probabilidade de ocorrer o C dado que já ocorreu F.

O *Naive Bayes* visa rotular uma frase, tendo como base a probabilidade do objeto ser de determinada classe ou não. De acordo com Gomes (2013) o *Naive Bayes* é considerado bastante eficaz nas situações em que necessite do processamento e precisão para classificar novas amostras. O *Naive Bayes* demonstra bom resultado na categorização de textos (ARANHA; VELLASCO, 2007).

4.4.1.2 Entropia Máxima

O classificador de entropia máxima transforma conjuntos de recursos classificados em vetores por meio da codificação. O vetor codificado é utilizado para calcular os pesos de cada recurso, através da combinação dos pesos que ocorre a classificação das frases (MEDHAT; HASSEN; KORASHY ,2014).

O classificador de entropia máxima foi utilizado por Kaufmann (2012) que tinha como objetivo utilizar uma pequena quantidade de dados de treinamento onde fosse possível detectar sentenças paralelas de qualquer par de idiomas.

4.4.1.3 Rede Bayesiana

A rede Bayesiana é um modelo bastante utilizado nos campos que fazem-se necessário realizar diagnósticos (Moreno, 2015). A rede bayesiana difere-se do classificador *Naive Bayes* na suposição de que todos os recursos são dependentes, como já foi dito o classificador *Naive Bayes* supõe que todos os recursos são independentes (MEDHAT; HASSEN; KORASHY ,2014).

A autora Moreno (2015) dá o exemplo da utilização da rede bayesiana para realizar um diagnóstico médico, neste exemplo um paciente deveria informar os sintomas, e tendo-os como base, o software buscaria na base de dados, que foi alimentado com diversos dados estatísticos adquiridos ao longo do tempo ou pré-definidos e retornaria quais as possíveis doenças que estão relacionadas aos sintomas que foram apresentados.

Para a análise de emoções a utilização da rede bayesiana torna-se muito cara devido à complexidade de computação, por este motivo, não é utilizada com frequência (AGGARWAL; ZHAI, 2012).

4.4.1.4 Rede Neural

A rede neural trata-se de algoritmos sofisticados utilizados com o intuito de modelar relações complexas, normalmente não lineares, entre a entrada e a saída de dados (SANTOS, 2017). A rede neural constitui-se de vários neurônios no qual o neurônio é sua unidade básica (MEDHAT; HASSEN; KORASHY, 2014).

TAFNER et al (1996) define a rede neural como “uma rede massivamente paralela de elementos interconectados e suas organizações hierárquicas que estão preparadas para iterar com objetos do mundo real do mesmo modo que um sistema nervoso biológico faz”.

As redes neurais buscam se aproximar o máximo possível das redes neurais humanas, pois, as redes neurais humanas apresentam uma desenvoltura no desempenho de suas funções Paula (2002).

4.4.1.5 SVM

Adotando a Teoria Estatística da Aprendizagem, os primeiros trabalhos sobre as SVM surgiram no ano de 1979 com o trabalho de Vapnik (VAPNIK, 2006). O SVM é um classificador que busca definir bons separadores lineares que terão como objetivo separar da melhor maneira as diferentes classes (MEDHAT; HASSEN; KORASHY, 2014).

Os dados textuais apresentam ser idealmente adequado a classificação SVM pela natureza espalhada do texto, onde poucos recursos podem ser considerados irrelevantes, contudo as palavras tendem a correlacionar entre si e normalmente organizadas em tipos linearmente separáveis (JOACHIMS, 1997).

As aplicações que utilizam esse modelo demonstram um resultado comparável e às vezes superior ao conseguido por outros algoritmos de aprendizado (BRAGA et al., 2000).

4.4.1.6 Modelo baseado em regras

Um modelo baseado em regras é um classificador que utiliza para realizar a classificação um conjunto de regras (SILVA; LOH, 2016). Há diversos critérios para a criação das regras, a fase de treinamento que construiu todas as regras seguindo os critérios (MEDHAT; HASSEN; KORASHY, 2014).

O modelo baseado em regras depende de um treinamento antecipado e faz uso de uma base de conhecimento com a finalidade de aplicar as regras (TAN; STEINBACH; KUMAR, 2009).

4.4.1.7 Árvore de decisão

Conforme Souza et al. (1998) a árvore de decisão é uma maneira simples de classificar amostras em uma quantidade finita de classes. As árvores de decisão é um meio eficiente de produzir classificadores que tem como objetivo prever as classes baseando-se em valores de atributos de um determinado conjunto (GARCIA, 2000).

A árvore de decisão consiste em determinar regras para a divisão de um conjunto de dados tornando-os grupos, a caracterização da amostra inserida ocorrerá tendo como base o comportamento do grupo no qual está inserida, as regras para a divisão mudam de acordo com o algoritmo que foi escolhido (SANTOS, 2014).

A principal diferença entre o modelo baseado em regras e a árvores de decisão é que os classificadores baseado em regras permitem haver a sobreposição no espaço de decisão enquanto que na árvore de decisão o particionamento é hierárquico rigoroso (MEDHAT; HASSEN; KORASHY, 2014).

4.5 Python

A linguagem de programação *Python* é uma linguagem de altíssimo nível, de tipagem dinâmica e forte, iterativa e interpretada e orientada a objeto (BORGES, 2010). Segundo Borges (2014), possui uma sintaxe clara e concisa, favorecendo a legibilidade do código fonte e com isso faz com que o desenvolvimento utilizando esta linguagem seja mais produtiva.

O *Python* foi criado por Guido van Rossum em 1991, tendo como origem de seu nome a série humorística britânica *Monty Python's Flying Circus* (SANTANA; GALESI, 2010). Atualmente tem conquistado bastante espaço entre as outras ferramentas de programação, por ter uma interface "amigável" de fácil aprendizagem e pela sua grande aplicabilidade.

De acordo com o TIOBE - *The Software Quality Company* (2021) o *Python* é a segunda linguagem mais popular dentre os desenvolvedores no mês de maio de 2021, além de ser uma linguagem popular é utilizado por grandes empresas de tecnologia, como a *Google*, *Microsoft*, *Instagram*, *Spotify* e várias outras. *Python* é deveras popular na comunidade científica e bastante utilizado na análise de dados (CAELUM, 20-?). Bird, Klein e Loper (2009) afirmou que *Python* é uma linguagem que dispõe de excelentes funcionalidades destinadas ao processo de dados linguísticos.

4.6 NLTK

O pacote *Natural Language Toolkit* (NLTK) tem como função realizar o processamento de linguagem natural em *Python*, foi criado em 2001 como parte do curso de Linguística Computacional no Departamento de Ciência da Computação e informação da Universidade da Pensilvânia e desde então foi utilizado em cursos de diversas universidades (BIRD; KLEIN; LOPER, 2009).

De acordo com Garg (2016) o NLTK dispõe de diversas funções que são utilizadas no pré-processamento dos dados a fim de que os dados disponibilizados no *Twitter* se tornem próprios para a extração e a mineração de recursos.

Conforme Bird, Klein e Loper (2009), o NLTK tinha como foco atingir quatro objetivos principais quando foi projetado:

- Simplicidade: este objetivo busca fornecer uma estrutura que seja intuitiva, fornecendo ao usuário, de forma que não torne-se entediante, conhecimento prático de PLN.
- Consistência: procura fornecer uma ferramenta que seja uniforme com interfaces e estruturas de dados consistentes e com métodos que tenham nomenclatura clara, ou seja, facial de adivinha a funcionalidade do método.
- Extensibilidade: busca fornecer uma estrutura em que novos módulos do software possam ser bem acomodados, onde haja a possibilidade de incluir em uma tarefa novas implementações e abordagens concorrentes.
- Modularidade: este objetivo procura fornecer componentes que são capazes de serem utilizados de forma independente, excluindo assim a necessidade de entender todo o *kit* de ferramentas.

O NLTK é uma biblioteca que executa um papel muito significativo na conversão de textos que estão em linguagem natural para um sentimento (GARG, 2016). Esta biblioteca terá um papel importante na conversão dos dados de textos que coletamos do *Twitter* em um formato que seja possível a extração de sentimentos.

5. METODOLOGIA

Tendo como origem a pesquisa bibliográfica para a composição deste trabalho, tentou-se levantar, descobrir e analisar conhecimentos e informações antecipadas de extração de conhecimento sobre documentos textuais, buscando uma resposta ou uma contribuição na resolução do problema (VILAÇA, 2010).

Para a primeira etapa do projeto faz-se necessário a identificação das técnicas que serão utilizadas para a análise dos sentimentos. Como já foi dito, existem basicamente três tipos de técnicas para a detecção de emoções, abordagem baseada em léxico, abordagem de aprendizagem de máquina e abordagem híbrida. A abordagem que deverá ser utilizada será a de aprendizagem de máquina.

A segunda etapa é produzir uma aplicação que deverá coletar dados e analisá-los, é necessário cumprir alguns passos antes de produzir a aplicação. O primeiro passo que tem que ser feito é buscar uma maneira de fazer a coleta de frases no *Twitter*, esse passo pode ser cumprido simplesmente copiando e colando uma frase relacionada ao *e-commerce* que buscamos analisar porém esse tipo de método não é interessante, para tornar o trabalho de coleta mais simples e mais veloz deverá ser feito um *script* que terá como propósito a coleta de frases que estejam associadas ao *e-commerce* que desejamos analisar.

O segundo passo busca produzir a base de treinamento e a base de teste, neste passo iremos analisar e classificar manualmente cada frase que foi coletada. Haja visto que é indispensável que as frases sejam salvas em um tipo de arquivo que seja permitido a edição. Para isso criaremos um *script* de código que deverá gerar um arquivo que seja editável e inserirá os dados que foram coletados no passo anterior nesse arquivo. Após os dados estarem salvos no arquivo iniciaremos a classificação, nesta parte analisaremos cada frase e colocaremos a classe que ela possui, podendo ser “neutro”, “insatisfeito” e “satisfeito”.

O terceiro passo objetiva tornar a frase limpa, ou seja, retirar as menções, *hashtags*, *emoji*, *URLs*. Para fazer este processo faremos um *script* que irá limpar as frases, para auxiliar na codificação do *script* para a limpeza utilizaremos o método *clean* que está presente na biblioteca *preprocessor*, é este método que deverá limpar as frases. Após a limpeza das frases há possibilidade de frase terem ficados vazias, isto ocorre quando uma frase é composta por apenas itens que serão retirados pelo *script* de limpeza, para solucionar este problema devemos codificar uma estrutura de condição que terá como objetivo analisar se a frase está vazia, caso esteja vazia esta frase será apagada se não a frase deverá ser salva.

O quarto passo busca criar um *script* que analisará as frases e as classificará, para isso precisaremos acessar os arquivos externos no qual salvamos a nossa base de treinamento e a base de teste. A seguir deverá ser utilizado a biblioteca NLTK para auxiliar na produção do código que analisará as frases.

O quinto passo tem como objetivo levantar uma maneira de apresentar os dados que foram adquiridos. Para isto será escrito mais um *script*, para produzir este código faremos uso da biblioteca matplotlib que terá a função de produzir os gráficos que utilizaremos no projeto e por meio da biblioteca pandas faremos uma análise dos dados obtidos, como a quantidade de emoções de cada classe.

O sexto passo visa fazer a refatoração de todos os *scripts* que foram produzidos até então. A refatoração será feita com objetivo de criar uma aplicação que ao ser executada inicializará os *scripts* já feitos. As modificações que serão realizadas na refatoração serão mínimas, por exemplo, no *script* de coleta que foi produzido no primeiro passo será modificado os dados que serão coletados, até então o *script* coletava apenas o *tweet* agora passará a coletar o *tweet*, a linguagem que foi publicado, a quantidade de *likes*, a quantidade de *retweets*, a data da postagem, por qual plataforma foi postado, a localização do usuário que postou, caso esteja disponível, a data que ocorreu a postagem e a quantidade de caracteres que há no *tweet*, além disso o arquivo externo em que estão sendo salvo os dados coletados será modificado. Já no *script* escrito no quinto passo a partir do momento que for feita a refatoração este *script* irá analisar não só a quantidade de emoção de cada classe, mais também a quantidade de *tweet* que foram coletados, a quantidade total de *likes*, as regiões que mais postaram sobre o *e-commerce*, a quantidade de postagens feitas em cada tipo de plataforma e a quantidade média caracteres.

Na terceira e última etapa iremos realizar a análise de emoções, nesta etapa será executado a aplicação que foi desenvolvida na fase anterior onde deverá ser apresentado o resultado da análise, ou seja, será exibido os dados da quantidade de frases classificadas como “satisfeito”, “insatisfeito” e “neutro”, os três *tweets* com mais *likes*, e os três com mais *retweets*, o gráfico de qual a região que mais realiza postagens sobre o *e-commerce*, qual plataforma que é mais utilizada para realizar as postagem que estão relacionadas ao *e-commerce* e a quantidade média de caracteres utilizados para transmitir a opinião sobre o *e-commerce*.

6. RESULTADOS ALCANÇADOS

Este trabalho tem como objetivo explicitar uma maneira de descobrir as opiniões dos clientes as quais foram expressadas em uma rede social. Para atingir este objetivo foram elaboradas e executadas várias etapas.

Como já foi abordado, para se utilizar o método de aprendizagem supervisionada se faz necessário uma coleção prévia de treinamento, a qual é denominada base de treinamento. Para o preenchimento desta base de dados foram utilizadas frases extraídas do *Twitter*, as quais foram coletadas utilizando a biblioteca *tweepy* e com o auxílio dela fizemos acesso a API do *Twitter*.

Antes de poder coletar os dados fazendo uso do *tweepy*, faz-se necessário estar cadastrado no *Twitter* e a conta está vinculada com o *Twitter Developer*. Esta vinculação é necessária pois carecemos de chaves, que são fornecidas pelo *Twitter Developer*, para que as requisições do acesso a API do *Twitter* sejam bem sucedidas.

Para podermos utilizar o *tweepy* precisamos instalá-lo, para isso executaremos a seguinte linha de comando em um terminal: `pip install tweepy`. Após a instalação do *tweepy* ser concluído o utilizaremos para a coleta de dados. Na figura 1 demonstramos o código para realizar a coleta.

Figura 1 - Primeiros passos para a coleta

```
1 import tweepy as tw
2
3 with open('Key_twitter.txt', 'r') as tfile:
4     consumer_key = tfile.readline().strip('\n')
5     consumer_secret = tfile.readline().strip('\n')
6     access_token = tfile.readline().strip('\n')
7     access_token_secret = tfile.readline().strip('\n')
8
9 auth = tw.OAuthHandler(consumer_key, consumer_secret)
10 auth.set_access_token([access_token, access_token_secret])
11
12 api = tw.API(auth)
13
14 query_search = "@e-commerce" + "-filter:retweets"
15
16 cursor_tweet = tw.Cursor(api.search, q=query_search).items(1000)
```

Fonte: Autor

O bloco de código da linha 3 até a 7 da figura 1 são referentes às chaves de acesso que adquirimos com o *Twitter Developer*, essas chaves foram salvas em um arquivo externo e

fazemos acesso ao arquivo utilizando a palavra reservada *with*. As variáveis `consumer_Key`, `consumer_secret`, `access_token` e `access_token_secret` guardarão os valores que foram salvos no arquivo externo.

Na linha de código 9 e 10 da figura 1 criamos um objeto chamado `auth`, obtemos este objeto por intermédio do método `OAuthHandler`, este método recebeu como argumentos as variáveis `consumer_Key` e `consumer_secret`, a seguir informamos a plataforma da API os nossos `access_token` e `access_token_secret` por meio do método `set_access_token`. O próximo passo é a criação de um objeto que seja do tipo `API`, é necessário este objeto pois é por meio dele que será gerado a conexão. Para a concepção deste objeto utilizaremos o método `API`, enviamos como argumento o `auth`.

Na linha 14 da figura 1 a variável `query_search` receberá as *strings* “@e-commerce” e “-filter:retweets”, o primeiro valor, “@e-commerce”, é a palavra chave que desejamos buscar, o segundo valor, “-filter:retweets”, informa que se deve retirar os *retweets*.

Na linha 16 da figura 1 a variável `cursor_tweet` receberá o resultado do método `Cursor`, método `item` é referente a quantidade máxima de itens que devem ser coletados. A seguir, armazenaremos os dados coletados em uma coleção, variável que permite o armazenamento de múltiplos itens.

Figura 2 - Inserindo os dados nas variáveis

```
18 tweet_data_list = []
19 print(type(tweet_data_list))
20 for tweet in cursor_tweet:
21     status = api.get_status(tweet.id, tweet_mode="extended")
22
23     tweet_data = {
24         'Frase': status.full_text,
25         'Emocao': "NULL"
26     }
27
28     tweet_data_list.append(tweet_data)
29
```

Fonte: Autor

Na linha 18 da figura 2 inicializamos a variável `tweet_data_list`, esta variável neste momento será transformada em uma variável do tipo *list*. A seguir no bloco de código que se inicia na linha 20 e finaliza na linha 27 iremos lidar com os dados que coletamos.

Na linha 20 criamos uma estrutura de repetição, esta estrutura tem como objetivo percorrer todo o objeto `cursor_tweet`. O método `get_status` irá retornar o único *status* que

especificamos no ID, *tweet.id*, o segundo argumento que enviamos para este método é o “*extended*”, para o parâmetro *tweet_mode*, esse argumento terá como função fazer com que o *tweet* que coletamos venha com o texto completo.

A variável *tweet_data* que foi inicializada na linha 23 receberá como valor um dicionário, dentro deste dicionário estará presente o texto do *tweet* que coletamos e a emoção.

Antes de finalizarmos esse bloco de código será inserido dentro da *tweet_data_list* a variável *tweet_data*.

Os dados que são coletados foram salvos em um arquivo externo no formato .txt. Podemos observar o código utilizado para salvar os dados na figura 3.

Figura 3 - Código para salvar no formato .txt

```
31 with open("Frases_Twitter.txt", "a", encoding='UTF-8') as arquivo:
32     arquivo.writelines(str(tweet_data_list))
33
```

Fonte: Autor

Após a coleta destas frases ocorreu uma análise e identificação das emoções transmitidas de maneira manual.

Antes de podermos torná-las nossa base de treinamento primeiro temos que limpá-las, ou seja, iremos retirar das frases elementos que iriam atrapalhar na análise de sentimentos.

Para fazer a limpeza das frases utilizaremos a biblioteca *preprocessor*, para fazer uso desta biblioteca é necessário instalá-la. O *preprocessor* irá retirar as menções, *hashtags*, *URLs*, *emojis*, palavras reservadas entre outras coisas. Podemos observar o código que fará a limpeza das frases na figura 4.

Figura 4 - Código para limpar as frases

```
def limpar_frase(base):
    dados_atualizados = []
    for dado in base:
        dados_atualizados.append((p.clean(dado['Frase']), dado['Emocao']))
    return dados_atualizados
```

Fonte: Autor

A base de treinamento encontra-se com 600 frases analisadas sendo 200 com emoção classificada como “neutro”, 200 classificadas como “insatisfeito” e 200 classificadas como “satisfeito”.

8. RESULTADOS ESPERADOS

Para a conclusão plena do objetivo deste projeto, o qual é descobrir as opiniões dos clientes que foram publicadas em redes sociais, existem etapas que ainda não foram finalizadas, as quais são: a implementação das técnicas e métodos levantados sobre análise de sentimentos, a utilização da biblioteca NLTK e a criação de *script* em *Python* que fará a análise de sentimento, produzir um código que terá a função de checar se a frase após ser limpa esteja vazia, a criação de *scripts* para a análise dos resultados obtidos na análise de sentimento e para a apresentação dos resultados desta análise, realizar a refatoração dos códigos produzidos transformando-os em uma aplicação que coletará dados no *Twitter*, os salvará, os limpará, executará a análise de sentimentos e apresentará dados sobre a quantidade de frases classificadas para cada tipo de classe, os três *tweets* com mais *likes* e os três mais retuitados, as regiões que mais realizaram postagens e quantidade média de caracteres presentes nos *tweet*.

REFERÊNCIAS BIBLIOGRÁFICAS

- AGGARWAL, Charu C.; ZHAI, Chengxiang. *Mining Text Data*. [S. l.: s. n.], 2012. Disponível em: <https://www.springer.com/gp/book/9781461432227#>. Acesso em: 20 abr. 2021.
- ANTUNES, M. N.; SILVA, C. H.; GUIMARÃES, M. C. S.; RABAÇO, M. H. L. Monitoramento de informação em mídias sociais: o e-Monitor Dengue. **TransInformação**, v. 26, n. 1, p. 9-18, 2014.
- ALLEN, J. F. *Natural language processing*. John Wiley and Sons Ltd., 2003.
- ARANHA, C. N; VELLASCO, M. Uma abordagem de pré-processamento automático para mineração de textos em português: sob o enfoque da inteligência computacional. **Pontifícia Universidade Católica do Rio de Janeiro**, 2007.
- BARBOSA, Alexandre N. **Descoberta de Conhecimento Aplicado à Base de Dados Textual de Saúde**. São Leopoldo, 2012.
- BECKER, Karin; TUMITAN, Diego. Introdução à Mineração de Opiniões: Conceitos, Aplicações e Desafios. [S.l.: s.n.], [s. l.], 20-?.
- BIRD, Stven; KLEIN, Ewan; LOPER, Edward. *Natural Language Processing with Python - Analyzing Text with the Natural Language Toolkit*. 1ª. ed. 2009. 479 p.
- BORGES, Luis E. **Python para desenvolvedores**. 2a. ed. Rio de Janeiro: Edição do Autor, 2010.
- BORGES, L. E. **Python para desenvolvedores**. Rio de Janeiro: Novatec, 2014, p. 27 - 29.
- BRAGA, Antônio De Pádua; CARVALHO, André Ponce De Leon E De; LUDERMIR, Teresa Bernarda. **Redes Neurais Artificiais: Teoria e Aplicações**. Editora LTC, 2000.
- BRASCHER, M. **A ambigüidade na recuperação da informação**. IASI, 2002.
- CAELUM. **PYTHON E ORIENTAÇÃO A OBJETOS**. [S. l.], 20-?. Disponível em: https://www.academia.edu/257818/Metodologia_De_Análise_De_Redes_Sociais. Acesso em: 9 maio 2021.
- COMPUTER ON THE BEACH, 2018, Florianópolis. **Análise de Sentimento em Redes Sociais no Idioma Português com Base em Mensagens do Twitter [...]**. [S. l.: s. n.], 2018. Disponível em: <https://www.aclweb.org/anthology/J09-3003/>. Acesso em: 18 fev. 2021.
- COSTA, Ana Carolina Bras; LOPES, Êmile Cunha. Análise de Sentimentos em Notícias Utilizando Dicionário Léxico e Aprendizado de Máquina. [S.l.: s.n.], 20-?. Disponível em: https://nca.ufma.br/~geraldo/vc/20152/trab/sentimentos_artigo.pdf. Acesso em: 11 maio 2021.
- EVANGELISTA, T. ; PADILHA, T. P. P. . **Monitoramento de Posts Sobre Empresas de E-Commerce em Redes Sociais Utilizando Análise de Sentimentos**. **III Brazilian Workshop on Social Network Analysis and Mining (BraSNAM 2014)**, 2014, Brasília.

GARCIA, S.C. **O uso de árvores de decisão na descoberta de conhecimento na área da saúde.** SEMANA ACADÊMICA, 2000. Rio Grande do Sul: Universidade Federal do Rio Grande do Sul, 2000.

GARG, Prateek. *Sentimental Analysis of Twitter Data using Python NLTK.* 2016. *Thesis (Master of Technology in Computer Science and Applications)* - Thapar University, [S. l.], 2016. Disponível em: <http://tudr.thapar.edu:8080/jspui/bitstream/10266/4273/4/4273.pdf>. Acesso em: 10 maio 2021.

GOMES, Helder Joaquim Carvalheira. **TEXT MINING: ANÁLISE DE SENTIMENTOS NA CLASSIFICAÇÃO DE NOTÍCIAS.** 2012. Trabalho de conclusão de curso (Mestre em Estatística e Gestão de informação, Especialização em Gestão do Conhecimento e Business Intelligence) - Instituto Superior de Estatística e Gestão de Informação Universidade Nova de Lisboa. Disponível em: <https://run.unl.pt/bitstream/10362/9182/1/TEGI0325.pdf>. Acesso em: 30 mar. 2021.

JAVA, Akshay et al. *Why We Twitter: Understanding Microblogging Usage and Communities.* **Joint 9th WEBKDD and 1st SNA-KDD Workshop**, [s. l.], 2007. Disponível em: https://www.researchgate.net/publication/234786863_Why_we_Twitter_Understanding_microblogging_usage_and_communities. Acesso em: 21 abr. 2021.

JOACHIMS, Thorsten et al. *A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization.* **Presented at the ICML conference**, [s. l.], 1997.

Kaufmann JM. **JMaxAlign: A Maximum Entropy Parallel Sentence Alignment Tool.** Disponível em: <<https://www.aclweb.org/anthology/C12-3035/>>. Acesso em : 10 de maio de 2021.

KIETZMANN, et. Al. **Social media? Get serious! Understanding the functional building blocks of social media.** Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0007681311000061> >. Acesso em 9 maio. 2021.

KUMAR, Vipin; STEINBACH, Michael; TAN, Pang-Ning. *Introdução ao Data Mining: Mineração de Dados.* [S. l.: s. n.], 2009.

LIU, B. *Sentiment analysis and opinion mining.* **Synthesis Lectures on Human Language Technologies**, Morgan & Claypool Publishers, v. 5, n. 1, p. 1–167, 2012.

MARQUES, Teresa P.; PINTO, Alexandra Marques; ALVEZ, Maria João. Estudo psicométrico da Escala de Avaliação dos Riscos e Oportunidades dos jovens utilizadores do Facebook. **Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica.**, [s. l.], 2016. Disponível em: <https://www.aidep.org/sites/default/files/articles/R41/Art12.pdf>. Acesso em: 28 mar. 2021.

MARTELETO, Regina Maria. Análise de redes sociais: aplicação nos estudos de transferência da informação. **Ciência da informação**, v. 30, n. 1, p. 71-81, 2001.

MAYNARD, D.; FUNK, A. *Automatic detection of political opinions in tweets*. In: SPRINGER. The semantic web: ESWC 2011 workshops. [S.l.], 2011. p. 88–99.

MEDHAT, W.; HASSAN, A.; KORASHY, H. *Sentiment analysis algorithms and applications: A survey*. *Ain Shams Engineering Journal*, Elsevier, v. 5, n. 4, p. 1093–1113, 2014.

MORENO, Águeda Cabral. **Análise de Sentimentos na Classificação de Comentários Online aplicando Técnicas de Text Mining**. 2015. Dissertação (Mestre em Sistemas Integrados de Apoio à Decisão) - Instituto Universitário de Lisboa, [S. l.], 2015. Disponível em: <https://repositorio.iscte-iul.pt/handle/10071/11504>. Acesso em: 5 maio 2021.

NAAMAN, Mor; BOASE, Jeffrey; LAI, Chih-hui. *Is it Really About Me? Message Content in Social Awareness Streams*. [S. l.: s. n.], 2010. Disponível em: https://www.academia.edu/5446826/Is_it_really_about_me_message_content_in_social_awareness_streams. Acesso em: 7 abr. 2021.

NASCIMENTO, Paula; OSIEK, Bruno Adam; XEXÉO, Geraldo. Análise de sentimento de tweets com foco em notícias. **Revista Eletrônica de Sistemas de Informação**, [S. l.], 2015. Disponível em: <http://www.periodicosibepes.org.br/index.php/reinfo/article/view/1525/pdf>. Acesso em: 7 abr. 2021.

PAULA, Maurício Braga de. **Indução automática de árvores de decisão**. 2002. Dissertação (Mestre em Ciência da Computação) - Universidade Federal de Santa Catarina, [S. l.], 2002. Disponível em: [https://repositorio.ufsc.br/xmlui/bitstream/handle/123456789/82465/189971.pdf?sequence=1&isAllowed=y#:~:text=Segundo%20QUINLAN%20\(1979\)%2C%20a,eficiente%20de%20aprendizado%20por%20exemplos.&text=O%20objetivo%20do%20aprendizado%20é,instâncias%20em%20suas%20respectivas%20classes](https://repositorio.ufsc.br/xmlui/bitstream/handle/123456789/82465/189971.pdf?sequence=1&isAllowed=y#:~:text=Segundo%20QUINLAN%20(1979)%2C%20a,eficiente%20de%20aprendizado%20por%20exemplos.&text=O%20objetivo%20do%20aprendizado%20é,instâncias%20em%20suas%20respectivas%20classes). Acesso em: 12 maio 2021.

RECUERO, Raquel. **Redes sociais na internet**. Porto Alegre: Sulina, 2009.

SANTOS, Amanda Alves dos. **Paisagem do Parque Nacional da Serra da Canastra e de sua zona de amortecimento - MG: análise de padrões espaciais a partir de árvore de decisão e métricas de paisagem**. 2014. Dissertação (Mestre em Análise e Modelagem de Sistemas Ambientais) - Universidade Federal de Minas Gerais, [S. l.], 2014. Disponível em: <https://repositorio.ufmg.br/handle/1843/IGCM-9UXRPK>. Acesso em: 11 maio 2021.

SANTOS, Igor Pedro Pinto dos. **Análise de Sentimento Usando Redes Neurais de Convolução**. 2017. Dissertação (Mestre) - Universidade do Estado do Rio de Janeiro, [S. l.], 2017. Disponível em: <https://docplayer.com.br/87455104-Igor-pedro-pinto-dos-santos-analise-de-sentimento-usando-redes-neurais-de-convolucao.html>. Acesso em: 13 maio 2021.

SEMIICAST. (2012). *Twitter reaches half a billion accounts. More than 140 millions in the U. S.* Disponível em: http://semicast.com/publications/2012_07_30_Twitter_reaches_half_a_billion_accounts_140m_in_the_US. Acesso em: 08 de março de 2012.

SIMPÓSIO BRASILEIRO DE BANCO DE DADOS, 2014, Curitiba. **Big Social Data: Princípios sobre Coleta, Tratamento e Análise de Dados Sociais [...]**. [S. l.: s. n.], 2014.

Disponível em: <https://www.inf.ufpr.br/sbbd-sbsc2014/sbbd/proceedings/artigos/pdfs/Topicos-em-Gerenciamento-de-Dados-e-Informacoes-2014.pdf>. Acesso em: 9 maio 2021.

SILVA, Rafael Silva da; LOH, Stanley. Sistema para identificação de sentimentos em textos na *web*. [S.l.: s.n.], 20 jun. 2016. Disponível em: <https://www.intext.com.br/tcc-rafael-silva.pdf>. Acesso em: 13 maio 2021.

SONI, Vibha *et al.* *Unsupervised Opinion Mining From Text Reviews Using SentiWordNet. International Journal of Computer Trends and Technology*, [s. l.], 2014. Disponível em: <https://www.ijcttjournal.org/archives/ijctt-v11p150>. Acesso em: 20 abr. 2021.

SOUZA, M. S. *et al.* *Data Mining: a database perspective. INTERNATIONAL CONFERENCE ON DATA MINING*, 1998. Rio de Janeiro: COPPE/UFRJ, p.413-432, 1998,.

SOUZA, Queila; QUANDT, Carlos. Metodologia de Análise de Redes Sociais. In: DUARTE, Fábio; QUANDT, Carlos; SOUZA, Queila. **O Tempo das Redes**. [S. l.: s. n.], 2008. Disponível em: https://www.academia.edu/257818/Metodologia_De_Análise_De_Red_Sociais. Acesso em: 25 mar. 2021.

TAFNER M. A. **Reconhecimento de palavras isoladas usando redes neurais artificiais**. Programa de Pós-Graduação em Engenharia de Produção. Universidade Federal de Santa Catarina. Florianópolis, Dissertação de Mestrado, 1996.

TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. **Introdução ao DATAMINING Mineração de Dados**. Rio de Janeiro: Editora Ciência Moderna Ltda., p-245-387, 2009.

TIOBE - *THE SOFTWARE QUALITY COMPANY*. TIOBE *Index for May 2021*. [S. l.], 2021. Disponível em: <https://www.tiobe.com/tiobe-index/>. Acesso em: 11 maio 2021.

VILAÇA, M.L.C. **Pesquisa e Ensino: Considerações e Reflexões**. *Revista Escrita*, v. 1, 2010.

VAPNIK, Vladimir. *Estimation of Dependences Based on Empirical Data*. [S.I.:s.n], 2006.

WILSON, Theresa; WIEBE, Janyce; HOFFMANN, Paul. *Recognizing Contextual Polarity: An Exploration of Features for Phrase-Level Sentiment Analysis. Computational Linguistics*, [s. l.], 16 abr. 2008. Disponível em: <https://www.aclweb.org/anthology/J09-3003/>. Acesso em: 18 fev. 2021.

ZHAO, Dejin *et al.* *How and Why People Twitter: The Role that Micro-blogging Plays in Informal Communication at Work*. [S. l.: s. n.], 2009. Disponível em: <https://www.semanticscholar.org/paper/How-and-why-people-Twitter%3A-the-role-that-plays-in-Zhao-Rosson/c473cebc334ba5bc3b11beeaff6f9d772d0e9b89>. Acesso em: 23 abr. 2021.

ZHANG, Yin; JIN, Rong; ZHOU, Zhi-Hua. *Understanding Bag-of-Words Model: A Statistical Framework. International Journal of Machine Learning and Cybernetics*, [s. l.],

2010. Disponível
https://www.researchgate.net/publication/226525014_Understanding_bag-of-words_model_A_statistical_framework. Acesso em: 16 mar. 2021.

em:

