

SOFTWARE WEBIAUS PARA ANÁLISE DE SENTIMENTOS EM MÍDIAS SOCIAIS: ATRIBUIÇÃO DE POLARIDADE EM NÍVEL DE SENTENÇA

Autores

Juan Carlos de Jesus Santos¹

Iuri Israel Pereira dos Reis Salla²

José Walmir Gonçalves Duque³

Resumo

No século XXI existem diversas formas das pessoas se comunicarem, isso sem a necessidade de um telefone fixo ou telegrama. Muitas dessas pessoas optam por postagens em redes sociais, sobre o seu dia a dia, o que compraram no mercado, atividades que pretendem realizar e até mesmo sua opinião sobre um assunto ou produto específico. O objetivo do software é implementar um algoritmo de Inteligência Artificial capaz de extrair informações de postagens publicados por usuários do Twitter utilizando uma técnica de elementos de textos com opiniões, podendo defini-las em um nível de sentença, onde o determinado trabalho se dedica integralmente, classificando-os como positivo ou negativo. É esperado que as informações obtidas pela Inteligência Artificial possam auxiliar em eventuais tomadas de decisão. Para esse software, será utilizada uma tecnologia baseada na linguagem de programação Python, uma linguagem interpretada estruturada de alto nível, fazendo também uso do banco de dados SQLite, um software de banco de dados simples e eficaz para aplicações ágeis, de código aberto. Os resultados em prol do projeto evidenciaram a necessidade e usabilidade deste software no dia a dia para análises rápidas de enormes datasets relacionados a comentários positivos e negativos. Com o uso do software, percebeu-se a possibilidade de analisar vários comentários de uma rede social por meio de uma técnica de Inteligência Artificial que define os dados positivos e negativos, por fim, o software retorna os resultados das análises em forma de gráfico.

Palavras-chave: Inteligência Artificial; Análise de Sentimento; *Machine Learning*.

Abstract

In the 21st century there are many ways for people to communicate, without the need for a landline or telegram. Many of these people choose to post on social networks, about their daily lives, what they bought in the market, activities they intend to carry out and even their opinion on a specific subject or product. The purpose of the software is to implement an Artificial Intelligence algorithm capable of extracting information from posts published by Twitter users using a technique of text elements with opinions, which can be defined at a sentence level, where the given work is fully dedicated, classifying them as positive or negative. It is expected that the information obtained by Artificial Intelligence can assist in eventual decision making. For this software, a technology based on the Python programming language will be used, a high-level structured interpreted language, also making use of the SQLite database, a simple and effective database software for agile, open source applications. The results in favor of the

¹ Graduação em Ciência da Computação pelo Centro Universitário Salesiano São Paulo – Unisal – Lorena. E-mail: curso.computacao@lo.unisal.br

² Graduação em Ciência da Computação pelo Centro Universitário Salesiano São Paulo – Unisal – Lorena. E-mail: curso.computacao@lo.unisal.br

³ Mestrado em Engenharia Eletrônica e Computação pelo Instituto Tecnológico de Aeronáutica – ITA, docente pelo Centro Universitário Salesiano São Paulo – Unisal – Lorena e docente pela Faculdade de Tecnologia do Estado de São Paulo – FATEC – E-mail: walmirduque@gmail.com

project evidenced the necessity and usability of this software on a daily basis for quick analysis of huge datasets related to positive and negative comments. With the use of the software, it was noticed the possibility of analyzing several comments from a social network through an Artificial Intelligence technique that defines the positive and negative data, finally, the software returns the results of the analysis in graphic form.

Keywords: Python; SQLite; Artificial Intelligence.

1. INTRODUÇÃO

No século XXI a sociedade encontra-se imersa na tecnologia e diversas são as formas das pessoas se comunicarem, não sendo necessário um meio físico para que essa comunicação ocorra. Muitas dessas pessoas optam por postagens em redes sociais, divulgando atividades que pretendem realizar e até mesmo sua opinião sobre um assunto ou produto específico (ARAÚJO, 2012).

Em virtude do aumento do número de expressões sentimentais em ambientes virtuais, as empresas encontraram na análise desses dados uma forma muito eficaz de identificar novas oportunidades, compreendendo a necessidade de seus clientes, usando essas informações para o crescimento de sua empresa no mercado (GOMES, 2019).

Essas empresas estão cada vez mais interessadas em automatizar o processo de análise dos dados extraídos das mídias sociais, tendo em vista a importância que tal análise tem para uma tomada de decisão estratégica das empresas, podendo dessa forma acrescentar em um ganho de maiores vendas e aumentar o seu público presente em determinado *site*, potencializando os benefícios financeiros, bem como reduzindo o tempo despendido para análise de público-alvo (IGNOATTO; WEBBER, 2019; SOUZA, 2021).

Análise de sentimento, segundo Liu (2010), é o estudo de sentimentos e emoções expressas em forma de texto. Existem diferentes frentes de pesquisas e diferentes níveis de abstração por meio dos quais é possível categorizar informações de texto como fatos ou opiniões. Este trabalho se dedica à análise de polarização (positiva ou negativa) em sentenças.

Diante dessa necessidade, foi desenvolvido o *software Webiaus*, totalmente gratuito, em interface *web*, em linguagem *Django*, com *layout* simples, contendo tela de *login*, cadastro e análise de resultados, cuja finalidade é realizar uma extração e coleta de dados embasadas em um *dataset* do *Twitter*, voltado para realizar um treinamento supervisionado de uma

Inteligência Artificial (IA), com base em comentários positivos e negativos pré-classificados, além de agregar um sistema de tolerância a falhas.

Essa aplicação realizará um treinamento que retornará um resultado para quem estiver utilizando, no qual essa análise será embasada em sentimentos com aspecto e atribuição de polaridade em nível de sentença, cujas atribuições de peso são baseadas em cada palavra presente em uma frase comentada em uma postagem no *Twitter*.

O objetivo da aplicação é fornecer às empresas uma IA capaz de extrair informações de postagens publicadas por usuários do *Twitter* utilizando uma técnica de elementos de textos com opiniões, podendo defini-las em uma análise baseada em nível de sentença, categorizando as informações de acordo com o peso de cada palavra presente em uma frase, servindo de base para o desenvolvimento de ferramentas ainda mais eficientes e especializadas para mineração/classificação de dados em redes sociais.

2. REFERENCIAL TEÓRICO

2.1. Inteligência Artificial

A tecnologia encontra-se difundida em todas as áreas do conhecimento, isso gera um grande fluxo de informações, tornando o processo de análise inviável para ser feito por seres humanos. A IA, segundo Russel e Norvig (2010), é um ramal da computação com objetivo de desenvolver algoritmos que sejam capazes de tomar decisões como um ser humano, fornece a capacidade de analisar diversos tipos de dados centralizados. Dessa forma, a IA visa o entendimento da inteligência humana para efetuar sua reprodução na computação objetivando um retorno de análise mais rápida e eficiente (ENDERLE; BRONZINO, 2012).

Feigenbaum (1981, apud FERNANDES, 2003) conceitua IA como uma parte da ciência da computação que visa à criação de aplicações que possam simular a inteligência humana que objetive entendimento de linguagem, aprendizado, raciocínio, resolução de problemas, entre outras problemática.

Russell e Norvig (2010) afirma que os sistemas integram duas formas de interpretação, “pensando” e “agindo”. Essas duas frentes são primordiais para definir os objetivos da IA. Com base na tabela 1, esse conceito é complementado, assim do são apresentadas 4 categorias para definir o que compõe as diversas áreas da IA, dentre elas:

Tabela 1: As 4 Categorias de IA

Sistemas que pensam iguais seres humanos	Sistemas que pensam de forma racional
Sistemas que agem iguais seres humanos	Sistemas que agem de forma racional

Fonte: (RUSSEL, NORVIG, 2010 [modificado em 2021 Set 17]).

2.2. Objetividade, Subjetividade e Emoção

De acordo com Feldman (2013), a objetividade procura apresentar sempre um acontecimento, bem como uma informação concreta ou fatos. A subjetividade apresenta algo mais fora de padrões concretos, bem como opiniões e sentimentos pessoais. Algumas técnicas utilizam a análise da objetividade para deduzir se compensa a realização da análise de sentimentos.

Conforme fora citado anteriormente, uma frase subjetiva expressa sentimentos pessoais ou opiniões, assim, a subjetividade lhe oferece um engajamento às emoções envolvidas nas frases analisadas e emite s específicos presentes, tais como: amor, alegria, surpresa ou raiva (ORTONY *et al.*, 1987).

Com base em estudos anteriores, a subjetividade com a emoção traz a análise de sentimento em mídias sociais com a polaridade em nível de sentença com mais eficácia e consistência no treinamento da IA.

Tendo como base que a posição emocional representa um dado de extrema importância para o desenvolvimento, bem como para empresas analisarem os dados de satisfação de suas vendas por meio dos dados apresentados pelos usuários. Um exemplo para analisar o estado emocional diferenciado em redes sociais, tal como o *Facebook*, é muito importante para prever ações de mães deprimidas que apresentam menos afeto pelo filho no pós-parto, se comparado com mães não deprimidas (CHOUDHURY *et al.* 2014).

2.3. Classificação e Análise de Polaridade

A polaridade tem uma representatividade muito forte e presente em qualquer análise de sentimento com uma análise baseada em textos e frases, e se beneficia quando diz respeito à análise de sentimentos.

De acordo com Camila (2016, P. 1):

A polaridade de uma molécula está diretamente relacionada à forma na qual os elétrons são distribuídos ao redor dos átomos. Se houver uma distribuição simétrica, a molécula será apolar, porém, se a distribuição for assimétrica, e uma das partes da molécula possuir uma grande densidade eletrônica, então será uma molécula polar.

De forma análoga, a análise de polaridade em uma sentença expressa como o ‘peso’ de cada expressão ou palavra analisada pelo analisador de sentimentos está distribuído no corpo do texto, podendo assim classificar aquele grupo de frases ou palavras em um nível subjetivo (positivo ou negativo).

De acordo com Silva (2015, p. 17), “Ao estudar o problema de análise de polaridade queremos classificar um documento em relação à sua polaridade: positiva, negativa e neutra.”. Ainda com base no mesmo autor, é ressaltado com uma ênfase cabal de que esses documentos não expressam claramente suas opiniões, bem como quando expressam opiniões, essas são classificadas como neutras e, contudo, classificar uma opinião com polaridade neutra é uma requisição extremamente difícil e complicada.

2.4. Análise de Sentimentos em Nível de Documento, Palavra e Aspecto

De acordo com Ortony *et al.* (1987), a ascendência do sentimento trará consistência na decisão tomada pela IA. A análise de sentimentos para comparação define perfeitamente o nível de sentença para cada análise, assim, o objetivo é classificar se um documento, em geral, expressa um sentimento positivo ou negativo. Esta importância é adequada quando o documento trata de uma única frase que disponibiliza uma opinião sobre um dado produto. (BECKER e TUMITAN, 2013).

Nesse nível de granularidade, uma frase pode ser julgada por várias entidades e pode conter vários sentidos relacionados a ela. A frase “Eu não tive um relacionamento muito agradável, porém os dias pareciam ótimos ao lado dela”, o sujeito do relacionamento “não” era muito carinhoso e “não” era muito bom, o sujeito envolvido tem duas polaridades diferentes relacionadas a “relacionamento” e “dias”. Embora os “dias” sejam considerados positivos, o “relacionamento” pode ser analisado de forma negativa.

Apenas a opinião sobre os “dias” pode ser considerada positiva, sendo que a opinião sobre o “relacionamento” é considerada negativa (BENEVENUTO; RIBEIRO; ARAÚJO, 2021). Com base em frases como essa, a IA tem maior dificuldade de análise ou requer maior grau de complexidade e processamento, podendo assim exigir um *hardware* mais dedicado em núcleos de processadores para a análise, pelo fato de analisar palavras com aspectos diferentes;

é nesse conceito em que a sentença terá uma aplicabilidade importante e a polaridade das palavras também irá definir um valor para as palavras e os aspectos.

Existem casos em que os usuários não identificam a sua opinião com tanta objetividade, adentrando na subjetividade e posicionamento emocional das palavras e frases, por exemplo, um “não” pode definir o peso em uma análise de sentimento e ser primordial para a sentença final em uma análise de sentimentos.

Em casos como esse, a IA deverá se preocupar em medir os pesos cabíveis em cada palavra presente em uma frase e, dessa forma, poderem ser comparadas; ela utilizará de advérbios que são palavras que modificam um verbo, um adjetivo ou outro advérbio (FERNANDES, 2021). São eles que irão extrair a entidade referida daquela opinião e definirão se aquela análise é positiva ou negativa (FELDMAN, 2013).

2.5. Nível de Sentença

Neste nível de análise em que o trabalho se integra absolutamente, um documento com as frases pode ter diversas opiniões diferentes umas das outras (THET; NA; KHOO, 2010). A análise em nível de sentença irá ressaltar claramente o peso para cada palavra e, para cada frase, o software preocupa-se em focalizar no geral, identificando postagens e comentários em mídias sociais que seguem um padrão de sentenças não muito extenso.

A análise de sentimento visa identificar e extrair automaticamente as opiniões, sentimentos e emoções expressos no texto (NARAYANAN; LIU; CHOUDHARY, 2009). O software de análise de sentimento percorre o texto e usa um algoritmo para gerar uma estimativa de seu conteúdo emocional, podendo gerar diferentes formas de sentença, sejam elas: fortemente positiva, positiva, fortemente negativa e negativa. O software geralmente usa algoritmos de análise de sentimento com abordagem à aprendizagem de máquina ou lexical. (THELWALL, 2013).

2.6. Aprendizagem de Máquina (Machine Learning)

Gomes (2019) aborda a aprendizagem de máquina como sendo voltada para encontrar padrões e para teoria de aprendizado voltado para a área de IA. Uma das primeiras pessoas a dar definições sobre aprendizado das máquinas foi Arthur Samuel, que o conceituou “campo de estudo que dá aos computadores a habilidade de aprender sem serem explicitamente programados” (SIMON, 2019).

De acordo com Thelwall (2013, p. 2, tradução nossa),

Uma abordagem de aprendizado de máquina pode classificar "não estou feliz" como negativo porque o bigrama "não estou feliz" ocorre quase sempre em textos no conjunto de treinamento codificados como negativos por humanos, enquanto a abordagem lexical pode optar pela escolha negativa porque "feliz" é uma palavra positiva conhecida e "não" é uma negação conhecida palavra que ocorre imediatamente antes dela.

A aprendizagem de máquina deve ser capaz de analisar essas pequenas frases, definir uma polaridade e, com isso, ser capaz de definir uma sentença para essas frases apontadas. Esses apontamentos serão capazes de dizer se aquele conjunto de palavras são positivas ou negativas, isso tudo com base no aprendizado de máquinas, mais conhecido como *Machine Learning* (THELWALL, 2013).

3. MATERIAIS E MÉTODOS

Nesta seção pretende-se apontar a metodologia utilizada no intuito de avaliar e analisar a usabilidade do *software* em redes sociais, sendo mais específico, a rede social *Twitter*, voltada para todos os tipos de pessoas.

A metodologia também fez uso de pesquisas teóricas a fim de reunir apontamentos e opiniões sobre a utilização do *software* em redes sociais. A obtenção dos resultados das pesquisas sobre analisador de sentimentos em redes sociais trouxe um grande *feedback* do trabalho, auxiliando a chegar a novas conclusões e desenvolver o *software* com mais clareza e segurança. Para visualização dos resultados da análise de sentimento, será apresentado um gráfico de barras (quantidade x polaridade), ainda em ambiente *web*.

3.1. Linguagem de Programação

Para o desenvolvimento de um *software website*, é necessário que a linguagem de programação possua recursos para atender todas as necessidades e o entendimento entre as linguagens envolvidas.

O *Python* é uma linguagem que já está há um bom tempo no mercado e vem sendo usada com muito mais frequência em projetos mais atuais, isso acontece pelo fato de ser uma linguagem de alto nível, deixando as origens de linguagens anteriores e adotando medidas de atualização com o decorrer do mercado e a sua necessidade.

De acordo com Coutinho (2019), *Python* é uma linguagem de programação de alto nível voltada para facilitar a aplicabilidade em *Django*, para uma programação ágil, voltada para

Zope Object Data Base (ZODB), aplicada para gerenciamento de banco de dados e direcionada também para programação de jogos com o *Pygame*.

Para o desenvolvimento do projeto, a linguagem de programação com mais recursos é o *Python* voltado para *Django*, podendo atender todas as expectativas e necessidades ágeis do *software*.

3.2. Django

De acordo com a documentação do *Django*, ele foi desenvolvido com o intuito de tornar-se um ambiente *web* com um desenvolvimento mais ágil e fácil. O *site* da documentação já utiliza bastante referência para tornar simples a sua programação com tutoriais disponíveis e o *Django* é uma estrutura *web Python* de alto nível, contudo, é importante ressaltar que ele é um *software* livre e livre de custos monetários.

Para o desenvolvimento do projeto será utilizado *Django* para toda a aplicação *web*, envolvendo *Python Web*, podendo utilizar de *templates* pré-definidos existentes nas ferramentas do *Django*.

3.3. HTML & CSS

Segundo Guedes (2021), o *HyperText Markup Language (HTML)* e *Cascading Style Sheets (CSS)* são as linguagens base utilizadas para o desenvolvimento de páginas *web* e elas atuam de forma a aprontar todo o desenvolvimento. Ela é baseada em marcações, que, de acordo com Guedes (2021), são responsáveis por definir elementos dentro da página *web*. Todas as especificações sobre HTML e CSS são controladas e administradas pelo W3C (*World Wide Web Consortium*).

O CSS é uma linguagem de folha de estilo, por assim dizer é complementar do HTML. É utilizado para auxiliar todo o HTML e tornar o código mais limpo e legível por qualquer programador e desenvolvedor *web*. Por meio do CSS, é possível organizar um padrão de classes do HTML para facilitar o desenvolvimento de cada página *web*, agilizando e padronizando todos os *templates* desenvolvidos.

3.4. SQLite

Segundo a própria documentação do *SQLite* (2020), o *SQLite* é uma biblioteca de linguagens C, que tem por objetivo uma automação, implementação e agregação rápida e de alta confiabilidade de um mecanismo de banco de dados *Structured Query Language (SQL)*.

No software foi utilizado o *SQLite* como o banco de dados padrão para o desenvolvimento da aplicação em *Django*, isso pelo fato de ser uma ferramenta, ou bem dizendo, um banco de dados implementado na estrutura do *Python* com *Django*, podendo, assim, facilitar e agilizar todo o desenvolvimento do software.

3.5. Docker

Delgado (2020) assevera que o *Docker* é uma ferramenta *open source* e desenvolvida na linguagem de programação *Go*, que é uma linguagem de programação de alta performance desenvolvida pela *Google*. O objetivo do *Docker* é a criação e administração de ambientes únicos e isolados.

Esses ambientes isolados são nomeados como Containers, que são ambientes embasados em uma imagem de uma aplicação. Ainda, Delgado (2020, p.1) afirma: “imagine que o container *Docker* é como se fosse um container real em um navio (servidor), todos os contêineres estão lado a lado, porém seu conteúdo (ecossistema) não tem interferência de outros containers.”

Ainda com base no mesmo autor, a utilização de containers é muito útil para tolerância a falhas em grandes e pequenas empresas e aplicações reais. Se um container é excluído, é possível criar outro com as mesmas configurações e utilizações, podendo agilizar todo o desenvolvimento.

3.6. Mineração de Dados

Para a obtenção de dados, foi desenvolvido um algoritmo utilizando a linguagem *Python* com auxílio na biblioteca NLTK (*Natural Language Toolkit*) com base na *Application Programming Interface* (API) do *Twitter*. Esse algoritmo é responsável por criar um *dataset* embasado nas postagens feitas no *Twitter*, cuja busca é elaborada por meio de uma chamada de método e são obtidos dados estruturados em um arquivo CSV.

De acordo com Coutinho (2019), os dados estruturados podem ser definidos por três etapas, cada caso utiliza-se de uma etapa diferente. Existem os dados estruturados, no qual é obtido um dado muito rígido e organizado, esse tipo de estrutura é localizada em arquivos como: planilhas de *excel*; um banco de dados relacional; um arquivo CSV, entre outros. Ainda, conforme Coutinho (2019), existem outros dois tipos de estrutura, bem conhecidos como: dados semiestruturados e dados não estruturados.

Nessa aplicação foi utilizada a estrutura de dados estruturados, onde será manipulado um arquivo CSV com os dados totalmente organizados para a possível tomada de decisão e classificação dos dados.

3.7. Naive Bayes

Após o apanhamento dos dados, foi desenvolvido um algoritmo com base em *Naive Bayes* para classificá-los, que faz a chamada do método de *Naive Bayes*. O algoritmo pega a base de dados e classifica-os de forma probabilística e precisa.

Segundo Becker (2019), o algoritmo de *Naive Bayes* constitui-se em encontrar uma probabilidade a posteriori em um dado positivo, mas ele ressalta também de que não se pode esquecer da negação, portanto, deve-se registrar o posteriori da negação.

Para a realização dessa classificação, é utilizado o cálculo desenvolvido pela teoria de *Bayes*, como é representado na imagem 1:

Imagem 1: Fórmula de Naive Bayes

$$P(y | x_1, \dots, x_n) = \frac{P(y)P(x_1, \dots, x_n | y)}{P(x_1, \dots, x_n)}$$

Fonte: Página da fórmula de *Naive Bayes* no *Scikit-learn*⁴

Após a análise da imagem 1, a fórmula de *Naive Bayes* é dividida em algumas partes. Imagina-se que se está trabalhando no crescimento de uma nova empresa e foram realizadas 200 entrevistas com pessoas distintas.

Após coletar os dados das entrevistas, foi identificado que 80 pessoas possuíam o currículo pouco viável (40%) e 120 pessoas estavam com o currículo aplicável para a empresa (60%), sendo que das pessoas que possuíam o currículo pouco viável, 90% receberam aprovado na entrevista e 30% das pessoas que estavam com o currículo aplicável também foram aprovados na entrevista. Analisando esses dados mais claramente, pode-se observar da seguinte forma:

- 200 pessoas realizaram a entrevista.
- 40% das pessoas que realizaram a entrevista possuem um currículo pouco viável.
- 90% das pessoas que realizaram a entrevista foram aprovadas.

⁴ Disponível em: <https://scikit-learn.org/stable/modules/naive_bayes.html#>. Acesso em: 21 out. 2020.

→ 30% das pessoas que estavam com o currículo aplicável também foram aprovadas na entrevista.

Com os dados coletados, o algoritmo de *Naive Bayes* constitui-se em definir a probabilidade do que é positivo e é aprovado para entrevista.

É imprescindível salientar que é preciso aquilatar a probabilidade a posteriori da negação ou, bem dizer, por negatividade (as pessoas que não possuem o currículo viável, levando em conta de que passou pela entrevista e foi aprovado inicialmente, definindo como positivo). Assim:

$$\rightarrow P(\text{currículo pouco viável} \mid \text{positivo}) = 40\% * 90\%$$

$$\rightarrow P(\text{currículo pouco viável} \mid \text{positivo}) = 0,4 * 0,9$$

$$\rightarrow P(\text{currículo pouco viável} \mid \text{positivo}) = 0,36$$

$$\rightarrow P(\text{currículo aplicável} \mid \text{positivo}) = 60\% * 30\%$$

$$\rightarrow P(\text{currículo aplicável} \mid \text{positivo}) = 0,6 * 0,3$$

$$\rightarrow P(\text{currículo aplicável} \mid \text{positivo}) = 0,18$$

Em seguida, é preciso acertar os dados para que a adição das duas probabilidades resulte em 1 (100%), para isso, dividiu-se o resultado pela soma das duas probabilidades.

$$\rightarrow P(\text{currículo pouco viável} \mid \text{positivo}) = 0,36 / (0,36 + 0,18) = 0,6666$$

$$\rightarrow P(\text{currículo aplicável} \mid \text{positivo}) = 0,18 / (0,36 + 0,18) = 0,3333$$

$$\rightarrow 0,6666 + 0,3333 = 0,9999 \text{ ou aproximadamente } 1.$$

Após o entendimento do algoritmo de *Naive Bayes*, para o desenvolvimento do software foi utilizado uma API chamada de *MultinomialNB*, visto que o objetivo é de classificar os dados do *dataset* em *Python*, método voltado especificamente para classificação de textos (SCIKIT-LEARN, 2020).

3.8. API

Para desenvolvimento desse *software*, está sendo utilizada uma tecnologia de *Application Programming Interface* (API) do *Twitter*, que pode ser moldada de acordo com o decorrer da aplicação, facilitando a comunicação de dados supervisionados, possibilitando a entrega de conteúdos adaptados a cada grau de necessidade do algoritmo de uma IA aplicada.

Com base nessa API, é investigado esse ramo tecnológico permitindo a utilização de dados distintos da rede social *Twitter*, que são gerados por postagens de usuários da rede social.

Será com esses dados que o trabalho irá criar formas da sua IA voltada para análise de sentimentos.

Em sua competência mais relevante, a API é uma tecnologia que permite uma “conversação” entre um produtor ou um serviço, isso possibilita o fluxo de informações entre sistemas. Tal tecnologia possibilita o acesso aos dados e funcionalidades a outros desenvolvedores, outras empresas ou até mesmo entre departamentos e localizações. Existem duas formas para se fazer essas operações: API interna e API externa.

A API externa permite a troca de dados de forma aberta com o público, sem a necessidade de permissões ou restrições, geralmente são consumidas por aplicações criadas por desenvolvedores que não trabalham na empresa.

APIs internas são projetadas para serem consumidas por código escrito em casa, por desenvolvedores que trabalham internamente em uma empresa. Esses atributos tecnológicos aumentam as formas como as empresas trocam dados, serviços e recursos complexos com os seus parceiros e os abrem ao público, tanto de maneira interna quanto externa (LANE, 2013).

Para a utilização da API do *Twitter*, foi utilizado o método interno, na qual foi necessária a solicitação de uma chave criptografada entre a aplicação que está utilizando a API e os motivos pertinentes àquela utilização.

A API é uma estrutura de dados fornecida por bibliotecas tecnológicas para apoiar o desenvolvimento e o uso de aplicações através de uma interface legível. APIs abertas são aquelas cujo acesso é permitido a qualquer pessoa disposta a construir um aplicativo que necessita acessar a sua biblioteca, embora que por vezes sejam necessários um registro e uma chave de autorização. É um aplicativo ou aplicação Web acessado via navegador na Web e se baseia numa arquitetura cliente-servidor com o objetivo de gerar o resultado desejado pelos usuários (PIETOSO, 2009, p. 92, tradução nossa).

O intuito da API é a viabilidade de várias empresas e pessoas poderem agilizar o seu desenvolvimento e, por conseguinte, facilitar a criação de novos projetos e startups que as utilizam.

4. ANÁLISE E DISCUSSÃO DOS RESULTADOS

Com base no fluxograma (imagem 2), ao ser iniciado o software *Webiaus*, o processo é iniciado desde a íntegra da aplicação até o usuário final, assim pode-se apresentar toda sua temática e interface envolvida.

Assim que a página *web* do software *Webiaus* é apresentada (imagem 3), estão disponíveis duas opções, cadastro de usuário ou login; caso o usuário escolha efetuar o login

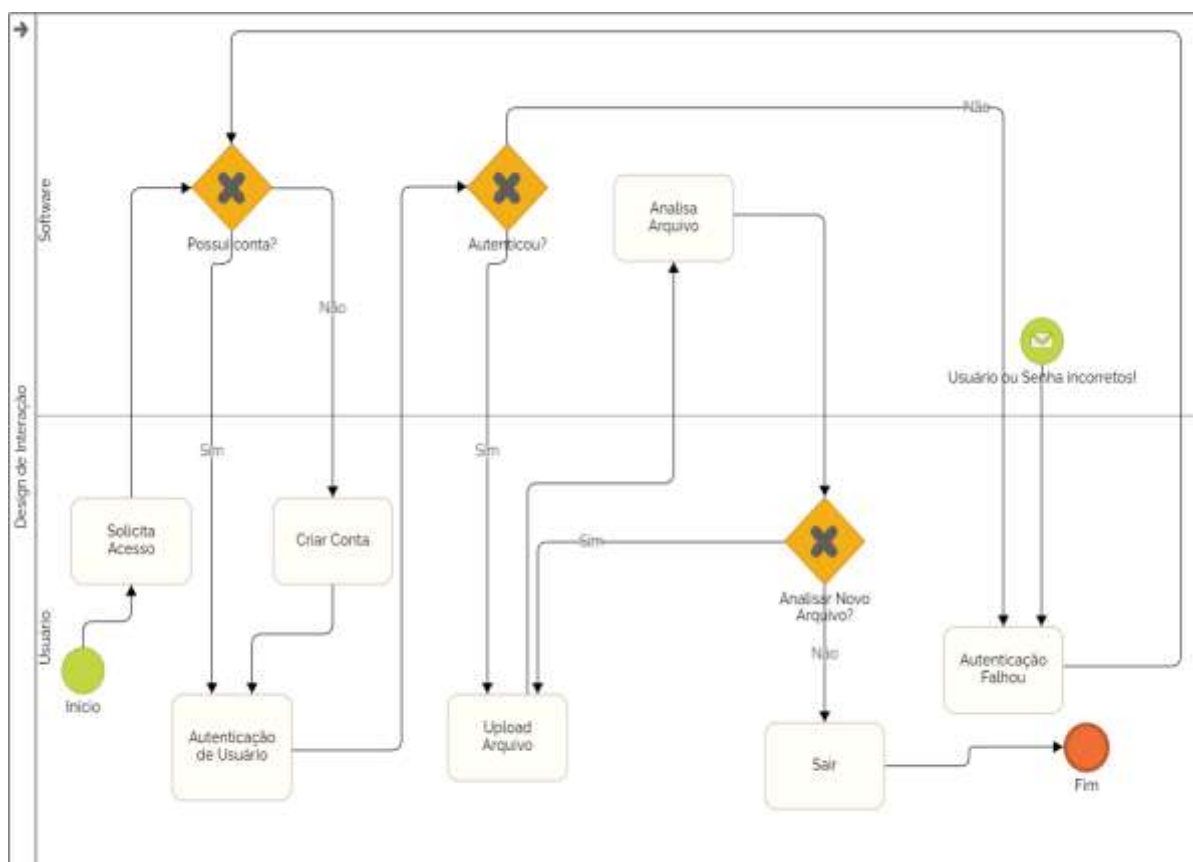
(imagem 5), é verificado se o usuário possui conta, caso afirmativo, é solicitada uma autenticação de usuário, que é direcionado para uma tela de *upload* de arquivo (imagem 6); após ser feito o *upload* deste arquivo, o usuário solicita a análise de sentimentos, logo é retornado o resultado positivo ou negativo com base na frase contida dentro do arquivo (imagem 9); em seguida, será perguntado se o usuário deseja analisar outro arquivo.

Caso o usuário não possua uma conta, ele é direcionado para uma tela para criação de conta (imagem 4), onde é solicitado seu endereço de *e-mail*, usuário e senha. Logo em seguida, o usuário é redirecionado para a tela de login e efetuará a autenticação de acesso (imagem 5).

Se a autenticação falhar, o usuário será informado da falha no acesso e poderá optar por redefinir sua senha (imagem 7) ou para efetuar um novo cadastro (imagem 4).

A partir do desenvolvimento do *software Webiaus*, foi realizado teste na plataforma do *Twitter*, retornando resultados que foram processados pelas entidades dessa aplicação, com dados positivos e negativos com base no campo de análise.

Imagem 2: Fluxograma de Interação do Software



Fonte: Autores (2020).

Imagem 3: Tela Inicial do Analisador de Sentimentos.



Fonte: Autores (2020).

A imagem 4 apresentada é a representação do formulário de Cadastro, assim que o usuário fizer a requisição no botão da aba apresentada na imagem 2.

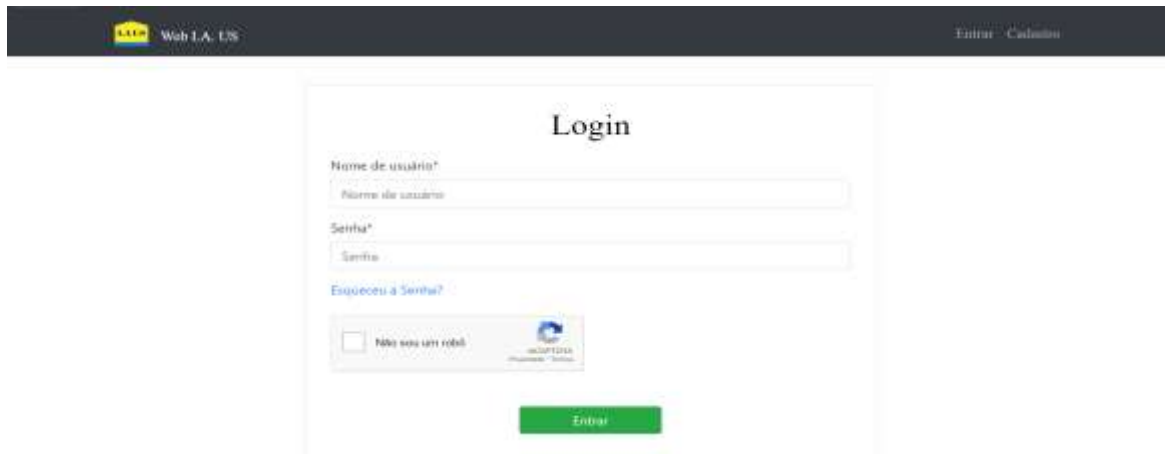
Imagem 4: Tela de Cadastro do Analisador de Sentimentos.

Fonte: Autores (2020).

A imagem 5 apresentada é a representação do formulário de login de autenticação do usuário. Esta autenticação é primordial para a funcionalidade principal do Analisador de

Sentimentos, pois sem a autenticação correta do usuário, é impossível utilizar o restante do software *Webiaus*.

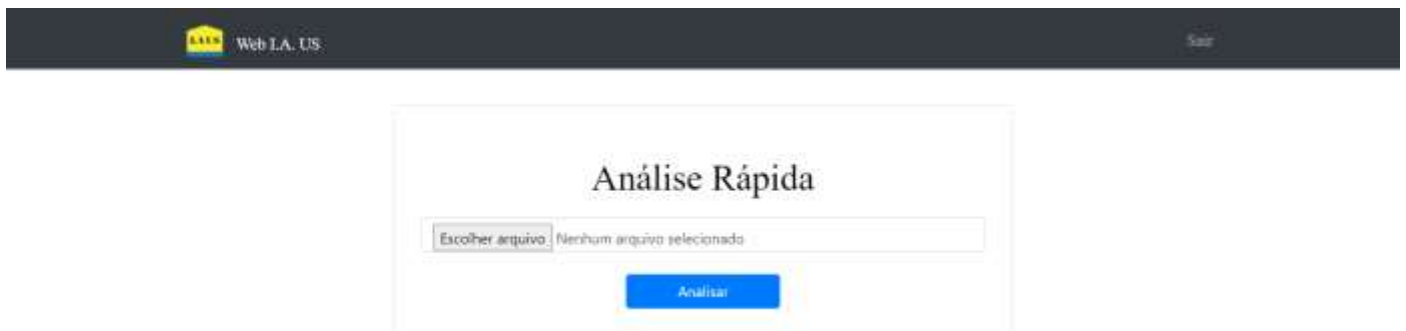
Imagem 5: Tela de Login do Analisador de Sentimentos.



Fonte: Autores (2020).

A imagem 6 apresentada é a representação da tela de Upload de Arquivos após a autenticação de usuário ser aprovada pelo formulário de Login. Como ressaltado na descrição da imagem 5, a autenticação é primordial para a funcionalidade principal do Analisador de Sentimentos.

Imagem 6: Tela Inicial do Analisador de Sentimentos após usuário autenticado.



Fonte: Autores (2020).

A imagem 7 apresentada é a representação da tela de quando o usuário esquecer seu usuário ou senha de autenticação.

Imagem 7: Tela de Redefinição de Senha do Analisador de Sentimentos.



Fonte: Autores (2020).

A imagem 8 apresentada é a representação da tela responsável por sair do Analisador de Sentimentos; após o usuário sair do *software*, é retornado para a tela inicial. Caso o usuário opte por não sair do *software*, pode clicar no *link* referente à parte *Web I.A US*.

Imagem 8: Tela de Sair do Analisador de Sentimentos.



Fonte: Autores (2020).

A imagem 9 apresentada é a representação da aplicabilidade da IA no *software Webiaus*, na qual faz jus ao nome de Analisador de Sentimentos. Nela são extraídos os valores do *upload* do arquivo que é apresentado na imagem 6, os quais retornam dados positivos e negativos de acordo com a tomada de decisão da IA.

Imagem 9: Tela de Análise Gráfica do Analisador de Sentimentos.



Fonte: Autores (2020).

A imagem 10 apresentada é um dos testes feitos com a coleta do arquivo com o comentário inserido. Inicialmente foi feito um teste para uma análise simples, cujo texto analisado foi o seguinte: “A comida está ótima”. A IA tem de retornar o resultado como positivo pelo fato de ser uma análise mais simples.

Imagem 10: Análise do Resultado Positivo da Análise de Sentimentos.

```

classificador2 = NaiveBayes("naive_bayes_1_4_pt2.pickle", vetorizador2, recomendacoes)
print(classificador2.marcaador())

for i in range(3):
    texto = input()
    analisar_sentimento(classificador2, texto)
    print("-----")
    
```

```

Windows PowerShell
Copyright (C) Microsoft Corporation. Todos os direitos reservados.

Experimente a nova plataforma cruzada PowerShell https://aka.ms/pscore6

PS C:\Users\juaens\OneDrive\Documents\Análise-de-Sentimento> python base.py
Tentando carregar revisões processadas...
As revisões processadas foram carregadas.

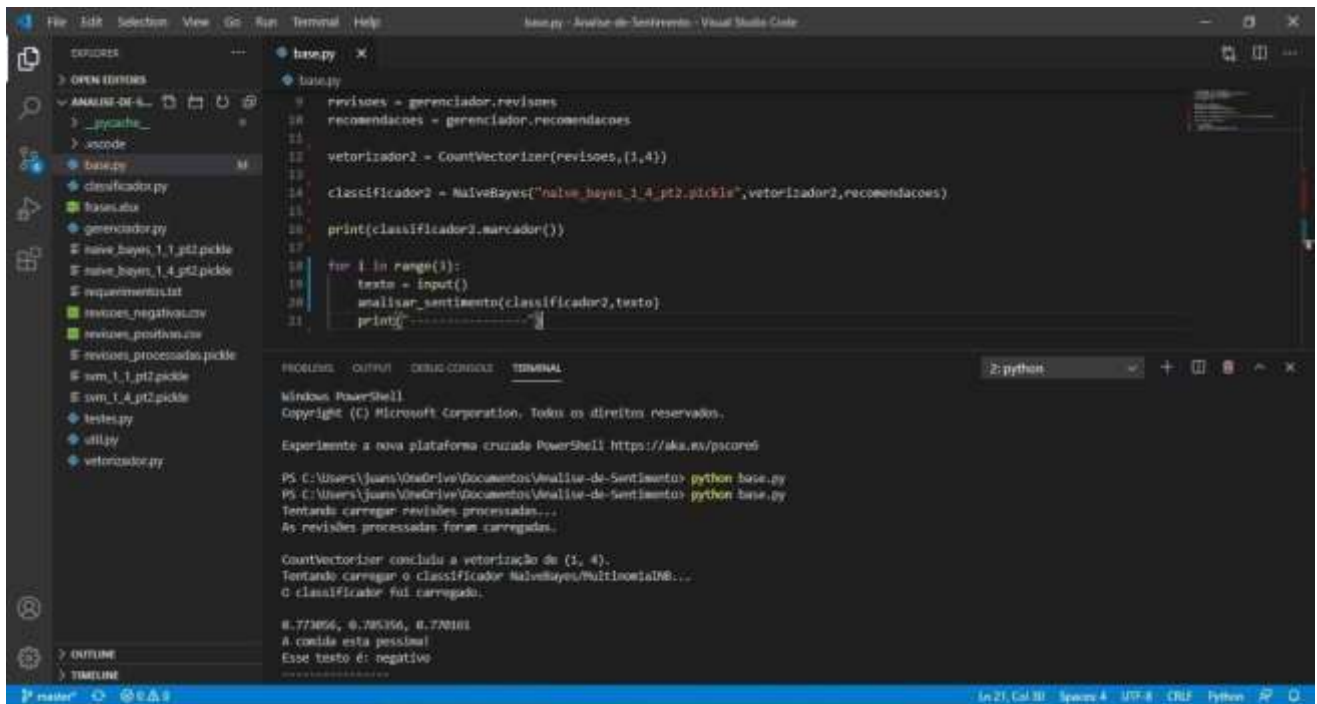
CountVectorizer concluiu a vetorização de (1, 4).
Tentando carregar o classificador NaiveBayes/multinomialNB...
O classificador foi carregado.

0.773056, 0.785356, 0.770103
A comida está ótima!
Esse texto é: positivo
    
```

Fonte: Autores (2020).

A imagem 11 apresentada é um dos testes feitos com a coleta do arquivo com o comentário inserido; inicialmente foi feito um teste para uma análise simples, em que o texto analisado foi o seguinte: “A comida está péssima”. A IA tem de retornar o resultado como negativo pelo fato de ser uma análise mais simples e negativa.

Imagem 11: Análise do Resultado Negativo da Análise de Sentimentos.



```
base.py
base.py
revisoes = gerenciador.revisoes
recomendacoes = gerenciador.recomendacoes
vetorizador2 = CountVectorizer(revisoes,(1,4))
classificador2 = NaiveBayes("naive_bayes_1_4_pt2.pkl",vetorizador2,recomendacoes)
print(classificador2.marcar())

for i in range(1):
    texto = input()
    analisar_sentimento(classificador2,texto)
    print(".....")
```

```
Windows PowerShell
Copyright (C) Microsoft Corporation. Todos os direitos reservados.

Experimente a nova plataforma cruzada PowerShell https://aka.ms/powershell

PS C:\Users\João\OneDrive\Documentos\Análise-de-Sentiment> python base.py
PS C:\Users\João\OneDrive\Documentos\Análise-de-Sentiment> python base.py
Tentando carregar revisoes_processadas...
As revisoes processadas foram carregadas.

CountVectorizer concluiu a vetorização de (1, 4).
Tentando carregar o classificador NaiveBayes/MultinomialNB...
O classificador foi carregado.

0.773856, 0.785356, 0.720181
A comida está péssima!
Esse texto é: negativo
.....
```

Fonte: Autores (2020).

A imagem 12 apresentada é um dos testes feitos com a coleta do arquivo com o comentário inserido. Nessa etapa foi feito um teste para uma análise com grau de dificuldade mais avançada, em que foram analisados três textos diferentes: “Incrível como conseguiram deixar um filme bom tão péssimo”; “Incrível como conseguiram deixar um filme bom tão chato” e, “Incrível como conseguiram deixar um filme bom tão ruim”. A IA tem dificuldade em analisar comentários positivos no início e conclusivos como negativos, mas ainda assim ela retorna o resultado como negativo, o resultado correto.

Imagem 12: Análise do Resultado da Análise de Sentimentos.

```

9  revisoes = gerenciador.revisoes
10 recomendacoes = gerenciador.recomendacoes
11
12 vetorizador2 = CountVectorizer(revisoes,(1,4))
13
14 classificador2 = NaiveBayes("naive_bayes_1_4_pt2.pickle",vetorizador2,recomendacoes)
15
16 print(classificador2.marcaador())
17
18 for i in range(3):
19     texto = input()
20     analisar_sentimento(classificador2,texto)
21     print("-----")

```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

```

Tentando carregar revisões processadas...
As revisões processadas foram carregadas.

CountVectorizer concluiu a vetorização de (1, 4).
Tentando carregar o classificador NaiveBayes/MultinomialNB...
O classificador foi carregado.

0.773056, 0.785356, 0.770161
Incrível como conseguiram deixar um filme bom tão pessimo
Esse texto é: negativo
-----
Incrível como conseguiram deixar um filme bom tão chato
Esse texto é: negativo
-----
Incrível como conseguiram deixar um filme legal tão ruim
Esse texto é: negativo
-----
PS C:\Users\Juaan\OneDrive\Documents\Análise-de-Sentimento>

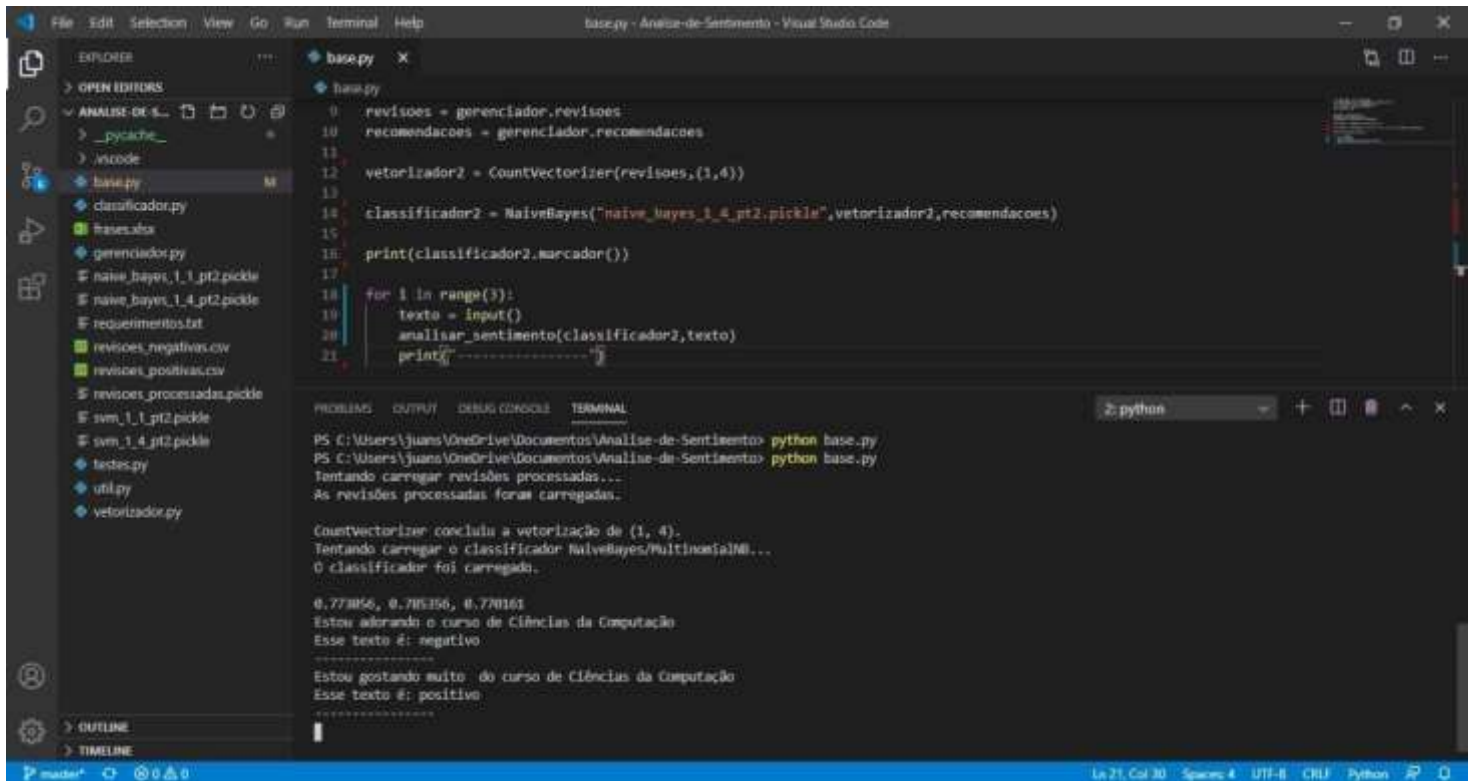
```

Fonte: Autores (2020).

De acordo com a imagem 13 apresentada nesta etapa, foi feito um teste para uma análise com grau de dificuldade avançada, em que foram analisados dois textos diferentes: “Estou adorando o curso de Ciências da Computação” e “Estou gostando muito do curso de Ciências da Computação”. A IA tem dificuldade em analisar comentários nos quais a polaridade das palavras tem um peso bastante relevante entre elas.

É possível observar isso nas duas frases apresentadas acima, em que “estar gostando muito” define uma positividade muito forte para a IA. Logo, a outra frase não tem essa representatividade muito convincente, dessa forma, a IA definiu como negativo.

Imagem 13: Análise do Resultado da Análise de Sentimentos.



```

0  revisoes = gerenciador.revisoes
10  recomendacoes = gerenciador.recomendacoes
11
12  vetorizador2 = CountVectorizer(revisoes,(1,4))
13
14  classificador2 = NaiveBayes("naive_bayes_1_4_pt2.pickle",vetorizador2,recomendacoes)
15
16  print(classificador2.marcaador())
17
18  for i in range(3):
19      texto = input()
20      analisar_sentimento(classificador2,texto)
21      print("-----")

```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

```

PS C:\Users\juans\OneDrive\Documents\Análise-de-Sentimento> python base.py
PS C:\Users\juans\OneDrive\Documents\Análise-de-Sentimento> python base.py
Tentando carregar revisões processadas...
As revisões processadas foram carregadas.

CountVectorizer concluiu a vetorização de (1, 4).
Tentando carregar o classificador NaiveBayes/MultinomialNB...
O classificador foi carregado.

0.778156, 0.785356, 0.778151
Estou adorando o curso de Ciências da Computação
Esse texto é: positivo
-----
Estou gostando muito do curso de Ciências da Computação
Esse texto é: positivo
-----

```

Fonte: Autores (2020).

5. CONSIDERAÇÕES FINAIS

Esse software apresentou uma proposta para implementação com o objetivo geral de implementar um algoritmo de IA capaz de extrair informações de postagens publicadas por usuários do *Twitter* utilizando uma técnica de elementos de textos com opiniões, podendo defini-las em um nível de sentença em que o determinado trabalho se dedica integralmente, classificando-os como positivo ou negativo. É esperado que as informações obtidas pela IA possam servir de auxílio para que os usuários possam utilizá-las em eventuais tomadas de decisão.

O desenvolvimento desse software mostrou a otimização de tempo, possibilitando a redução do trabalho humano para análise e interpretação de dados publicados na rede social *Twitter*. A partir desses resultados, infere-se que o projeto atendeu às necessidades dos objetivos apontados e a utilização de um software capaz de definir o grau de cada frase, sendo eles positivos ou negativos.

É notório por toda a fundamentação teórica que um analisador de sentimentos traz muito mais do que apenas uma simples análise; foi notável que uma análise de sentimentos

bem aplicada é capaz de prevenir diversos tipos de problemas ou doenças apresentadas pelas pessoas em suas redes sociais.

A experiência obtida por meio do desenvolvimento e dos testes revelou a empregabilidade do software no mercado. Foi possível visualizar a necessidade de um aplicativo capaz de identificar e definir um grau de escala entre positivo e negativo entre comentários postados em diversas redes sociais.

Uma possível atualização é a implantação dessa função em redes sociais na parte dos comentários, podendo filtrar os comentários positivos e negativos diretamente da página web, servindo diretamente como uma API.

Desta forma, propõe-se para trabalhos futuros facilitar e agilizar a aplicabilidade desse analisador de sentimentos como uma ferramenta em listas de comentários em redes sociais, já que essa ferramenta está disponível de forma gratuita e sua implementação é simples e intuitiva, podendo trazer uma melhoria na saúde mental de diversos usuários.

REFERÊNCIAS

ARAÚJO, L. L. P. **A Popularização das Redes Sociais e o Fenômeno da Orkutização**. In: **Intercom – Sociedade Brasileira de Estudos Interdisciplinares da Comunicação XIV Congresso de Ciências da Comunicação na Região Nordeste**, Recife, 2012. Disponível em: <<http://www.intercom.org.br/papers/regionais/nordeste2012/resumos/R32-0590-1.pdf>>. Acesso em: 2 set. 2020.

BECKER, L. **Algoritmo de classificação Naive Bayes**. 2019. Disponível em: <<https://www.organicadigital.com/blog/algoritmo-de-classificacao-naive-bayes/#:~:text=O%20algoritmo%20de%20Naive%20Bayes,dado%20que%20tem%20a%20doen%C3%A7a%E2%80%9D>>. Acesso em: 21 out. 2020.

BENEVENUTO, F.; RIBEIRO, F.; ARAÚJO, M. **Métodos para Análise de Sentimentos em mídias sociais**. 2021. Disponível em: <<https://homepages.dcc.ufmg.br/~fabricio/download/webmedia-short-course.pdf>>. Acesso em: 18 set. 2020.

CAMILA, P. **Como definir o conceito de polaridade?**. 2016. Disponível em: <<https://descomplica.com.br/artigo/como-definir-o-conceito-de-polaridade/49V/>>. Acesso em: 17 set. 2020.

COUTINHO, T. **O que é Python e para que serve**. 2019. Disponível em: <<https://www.voitto.com.br/blog/artigo/python>>. Acesso em: 14 out. 2020.

CHOUDHURY, D. M.; COUNTS, S.; HORVITZ, E. J.; Hoff, A. **Characterizing and predicting postpartum depression from shared facebook data**. 2014. Disponível em: <<http://www.dataluna.com/moca-papers/625.pdf>>. Acesso em: 17 set. 2020.

DELGADO, C. **Docker 101 - Primeiros passos e muito mais!**. 2020. Disponível em: <<https://caiodelgado.dev/docker-101/>>. Acesso em: 15 out. 2020.

ENDERLE, J. D.; BRONZINO, J. D. **Introduction to biomedical engineering**. 3. ed. Burlington: Academic press. 2012.

FELDMAN, R. (2013). Techniques and applications for sentiment analysis. **Communications of the ACM**, 56(4):82–89.

FERNANDES, A. **Inteligência artificial: noções gerais**. Florianópolis: Visual Books, 2003.

FERNANDES, M. **Classificação dos Advérbios**. 2012. Disponível em: <<https://www.todamateria.com.br/adverbio/>>. Acesso em: 16 set. 2020.

GOMES, P. C. T. **Análise de Sentimentos com Machine Learning**. 2019. Data Geeks (site). Disponível em: <<https://www.datageeks.com.br/analise-de-sentimentos/>>. Acesso em: 2 set. 2020.

GOMES, P. **Conheça as etapas do Pré-processamento de dados**. 2019 Disponível em: <<https://www.datageeks.com.br/pre-processamento-de-dados/#:~:text=Neste%20artigo%2C%20apresentarei%20algumas%20t%C3%A9cnicas,em%20qualquer%20conjunto%20de%20dados.>>. Acesso em: 21 out. 2020.

GUEDES, M. **O que é e como começar com HTML e CSS?**. [s.d.] 2021 Disponível em: <<https://www.treinaweb.com.br/blog/o-que-e-e-como-comecar-com-html-e-css/>>. Acesso em: 2 out. 2020.

IGNOATTO, M. L.; WEBBER, C. G. Inteligência Competitiva nas Mídias Sociais: Um Estudo de Caso na Moda. **SCIENTIA CUM INDUSTRIA**, V. 7, N. 2, P. 156 — 164, 2019. Disponível em: <<http://ucs.br/etc/revistas/index.php/scientiacumindustria/article/view/7749/pdf#>>. Acesso em: 17 set. 2021.

LANE, K. **API 101**. 2013. Disponível em: <<https://s3.amazonaws.com/kinlane-productions/whitepapers/API+Evangelist+-+API+101.pdf>>. Acesso em: 20 out. 2020.

LIU, B. **Sentiment Analysis and Subjectivity**. In. Handbook of Natural Language Processing. Segunda Edição. 2010.

NARAYANAN, R.; LIU, B.; CHOUDHARY, A. **Sentiment analysis of conditional sentences**. In Proc. of the Conference on Empirical Methods in Natural Language Processing: Volume 1, pages 180–189. Association for Computational Linguistics, 2009.

ORTONY, A.; CLORE, G. L.; FOSS, M. A. **Os fundamentos psicológicos do léxico afetivo**. **Revista de Personalidade e Psicologia Social**, vol. 53, P. 751–766. 1987.

PANG, B.; LEE, L.; VAITHYANATHAN, S. **Thumbs up? Sentiment Classification using Machine Learning Techniques**. Proceedings of the ACL-02 conference on Empirical

methods in natural language processing-Volume 10. Association for Computational Linguistics, PA - USA 2002.

PIETOSO, C. R. **Newspapers as Platforms: how Open APIs can impact journalism.** 2009. 108 f. Dissertação (Electronic Publishing) - City University, London, 2009.

RUSSELL, S. J.; NORVIG, P. **Artificial intelligence: a modern approach.** Upper Saddle River: Pearson Education Limited. Terceira Edição. 2010.

SAYAD, S. **Naive Bayesian.** [s.d.] 2021. Disponível em: <https://www.saedsayad.com/naive_bayesian.htm>. Acesso em: 21 out. 2020.

SCIKIT-LEARN. **Naive Bayes.** [s.d.] 2020. Disponível em: <https://scikit-learn.org/stable/modules/naive_bayes.html#multinomial-naive-bayes>. Acesso em: 21 out. 2020.

SILVA, J. R. **DETECÇÃO DE OPINIÕES E ANÁLISE DE POLARIDADE EM DOCUMENTOS FINANCEIROS COM MÚLTIPLAS ENTIDADES.** 2015. Dissertação (Mestrado em Informática) – Curso de Pós-Graduação em Informática – Universidade Federal do Amazonas, Manaus, 2015.

SIMON, P. **Too Big to Ignore: The Business Case for Big Data.** Wiley and SAS Business Series. Wiley, 2013.

SOUZA, G. C. **DESENVOLVIMENTO E VALIDAÇÃO DE UMA PLATAFORMA DE AUXÍLIO, BASEADA EM TWEETS, À TOMADA DE DECISÃO DO INVESTIDOR BRASILEIRO NO MERCADO FINANCEIRO.** 2021. Monografia – Curso Bacharel em Ciências da Computação – Universidade Federal de Ouro Preto, Minas Gerais, 2021.

SQLITE. **O que é SQLite?.** 2020 Disponível em: <<https://www.sqlite.org/index.html>>. Acesso em: 15 out 2020.

STRAPPARAVA, C., MIHALCEA, R. **Learning to Identify Emotions in Text.** ACM Symposium on Applied Computing. 2008, P. 1556-1560.

THELWALL, M. 2013. **Heart and soul: Sentiment strength detection in the social web with sentistrength.** Disponível em: <<http://sentistrength.wlv.ac.uk/documentation/SentiStrengthChapter.pdf>>. Acesso em: 18 set. 2020.

THET, T. T.; NA, J. C.; KHOO, C. S. G. Aspect-based sentiment analysis of movie reviews on discussion boards. **Journal of Information Science**, vol. 36 (6), P. 823–848, 2010.