

**ANÁLISE DE DADOS DO TWITTER:
UMA PERSPECTIVA PSICOLÓGICA DO BAIXO RENDIMENTO DOS
ESTUDANTES BRASILEIROS DURANTE E PÓS PANDEMIA**

ALEX SANTOS DA SILVA¹
PAULO HENRIQUE MIRANDA LOPES²
JOSÉ ROBERTO FONTOURA³

UNIVERSIDADE DO ESTADO DA BAHIA – UNEB
2023

RESUMO

Durante a pandemia, ampliou-se o uso das redes sociais como meio de comunicação, expressão e interação entre pessoas ao redor do mundo. Para exploração disso, foi desenvolvida uma aplicação para coletar e analisar dados de postagens através de uma máquina de aprendizado usando processamento de linguagem natural (PLN) para identificação e análise de sentimentos em relação à universidade. Para a construção do embasamento teórico foram usados autores como: MEDHAT; HASSAN; KORASHY. Essa pesquisa se configura como exploratória e descritiva, tendo como objetivo cruzar dados retirados de publicações do Twitter visando analisar assuntos mais relevantes durante os últimos quatro anos para identificação de palavras, frases e expressões para determinado sentimento nos textos coletados. Os resultados revelaram uma predominância de sentimentos negativos nas postagens dos alunos, indicando o confinamento decorrente da pandemia como um impacto negativo tanto no bem-estar emocional quanto no desempenho acadêmico.

Palavras-chave: Twitter. Análise de sentimentos. Processamento de linguagem natural.

ABSTRACT

During the pandemic, the use of social networks as a means of communication, expression and interaction between people around the world has expanded. To explore this, an application was developed to collect and analyze data from posts through a learning machine using natural language processing (NLP) to identify and analyze feelings towards the university. For the construction of the theoretical basis, authors such as MEDHAT; HASSAN; KORASHY were used. This research is configured as exploratory and descriptive, aiming to cross data taken from Twitter publications in order to analyze more relevant subjects during the last four years to identify words, phrases and expressions for a certain feeling in the collected texts. The results revealed a predominance of negative feelings in the students' posts, indicating the confinement resulting from the pandemic as a negative impact on both emotional well-being and academic performance.

Keywords: Twitter. Analyze feelings. Natural language processing.

¹ Graduando em Sistemas de Informação - Alagoinhas/BA - alesansilva@outlook.com.br

² Graduando em Sistemas de Informação - Alagoinhas/BA - paulolopes599@gmail.com

³ Doutorado em Difusão do Conhecimento - Salvador/BA - Professor da UNEB - jfontoura@uneb.br

1 INTRODUÇÃO

Durante a pandemia, ampliou-se o uso das redes sociais como meio de comunicação, expressão e interação entre pessoas ao redor do mundo.

As universidades que atuavam de forma presencial tiveram que se adaptar ao ensino remoto, para suprir a necessidade de continuidade das aulas, que por causa do período de confinamento foram interdidas.

Em vista disso, diversos estudantes começaram a sentir dificuldades perante as grandes demandas. Muitos deles recorreram às redes sociais como meio de expressar a insatisfação da sua vida acadêmica, o baixo rendimento que tiveram por causa desse cenário.

Gundim et al. (2021) investigaram reações emocionais negativas, como estresse, ansiedade, luto, raiva e pânico, em relação ao desempenho acadêmico, com base em sete artigos publicados em 2020.

Os resultados destacam que esse impacto não recebe a devida atenção nas universidades, enquanto as redes sociais se tornam a principal plataforma utilizada pelos jovens para expressar seus sentimentos e opiniões em relação à vida acadêmica. (Gundim et al., 2021).

De acordo com Araújo et al. (2020), a pandemia de COVID-19 teve um impacto psicológico importante, resultando em ansiedade e depressão ampliadas pela incerteza e pela sobrecarga de informações.

Nesse sentido, é necessário que as instituições implementem medidas preventivas em larga escala para promover a saúde física e psicológica de estudantes e professores, além de desenvolver planos para o retorno às aulas e o manejo de infecções nos campi.

Segundo a pesquisa realizada por Maia, Dias (2020), foi observado um aumento significativo de perturbação psicológica, como ansiedade, depressão e estresse, entre os estudantes durante a pandemia, em comparação com períodos considerados normais.

No entanto, os autores não abordaram explicitamente ações e estratégias de proteção e promoção da saúde mental nesse contexto. De acordo com Schmidt et al. (2020), pesquisas indicam que o medo de contrair um vírus altamente contagioso e potencialmente fatal, cujas características e trajetórias ainda são pouco compreendidas, tem impacto negativo no bem-estar psicológico de muitas pessoas.

Os estudantes enfrentaram incerteza em relação à progressão acadêmica, o que afetou negativamente sua saúde mental. Fatores como ser do sexo feminino, estar estudando e apresentar sintomas físicos específicos foram associados a um maior impacto psicológico, incluindo níveis mais altos de estresse, ansiedade e depressão.

Medidas de apoio mencionadas incluem intervenções psicológicas precoces para grupos de alto risco, recursos de suporte psicológico, redes de suporte online e informações precisas baseadas em evidências em formatos acessíveis. Teixeira, Dahl (2020) relataram que os estudantes experimentaram uma variedade de sentimentos, como medo, estresse, ansiedade, luto, culpa, raiva e desamparo, devido à incerteza causada pela emergência.

Foram identificadas ações para reduzir esses impactos, incluindo a criação de uma rotina de estudos baseada nos interesses individuais, encontros virtuais para práticas integrativas e complementares, orientações acadêmicas remotas, apoio psicossocial por meio de grupos de apoio e acesso a profissionais por meios

virtuais. Yuan S. et al. (2020) descobriram que 65,92% da amostra do estudo eram estudantes.

A maioria dos participantes relatou que seu estado emocional permaneceu estável nas duas semanas iniciais da pandemia. Os autores enfatizam a necessidade de implementar ações para minimizar o impacto na saúde mental daqueles que experimentaram alterações emocionais. Cao W et al. (2020) constataram que a saúde mental dos estudantes foi impactada de maneiras diversas durante o surto de COVID-19.

A pesquisa revelou que 24,9% dos estudantes apresentaram níveis de ansiedade elevados, atribuídos a fatores como estabilidade financeira, pressão psicológica e econômica, preocupações acadêmicas e impacto da epidemia na vida cotidiana, além de receios em relação aos estudos e ao futuro profissional, falta de comunicação interpessoal e disseminação de notícias falsas. O aumento no número de pacientes e casos suspeitos, assim como a expansão do surto para diferentes regiões, gerou preocupação pública, intensificando a ansiedade.

Diante do exposto anteriormente, a fim de suprir essa necessidade a partir de uma análise mais profunda dos dados que demonstram o maior índice desses distúrbios mentais que causam/causaram esse baixo rendimento, essa pesquisa busca responder a seguinte questão-problema: "Como identificar padrões de sentimentos expressos através das postagens do Twitter?"

Para isso foi necessário desenvolver uma *machine learning* (máquina de aprendizado) com capacidade de identificar padrões de textos através do processamento de linguagem natural, buscando similaridades e padrões de sentimentos perceptíveis nas postagens. Com a linguagem de programação *Python* para o desenvolvimento da *machine learning*, foram usadas diversas bibliotecas para simplificar e auxiliar o desenvolvimento do código, como a *pandas* e *matplotlib*, *requests*, *re*, *random*, *wordcloud*, *nltk*, *sklearn*, *pickle*.

Tendo como objetivos específicos:

a) cruzar dados retirados de publicações do Twitter visando analisar *tweets* de assuntos mais relevantes durante os últimos quatro anos para comparar uma base de dados de sentimentos pré-classificados, para identificação de palavras, frases e expressões chaves para cada determinado sentimento ou emoção expresso nos textos coletados do Twitter; b) Apontar os impactos visíveis aparentes nos dados recolhidos com o intuito de analisar os sentimentos predominantes nas publicações referentes a cada um dos últimos quatro anos e sua influência na vida acadêmica; c) Validar, a partir do estudo levantado, a predominância de sentimentos negativos que permeiam as dificuldades psicológicas enfrentadas durante a pandemia e a repercussão pós pandemia, gerando insatisfação acadêmica.

O interesse pelo tema se fez a partir das inquietações pessoais fundamentadas pela universidade em si, especificamente através do conhecimento crítico adquirido, no qual compreende-se a intencionalidade e as razões por trás do grande aumento em índices de depressão e ansiedade no período pandêmico e pós pandêmico, o que causa, de maneira significativa, consequências negativas.

A análise de dados possibilita uma facilidade em filtrar informações de forma direta. Através disso, foi criado um algoritmo de reconhecimento de padrões com o processamento de linguagem natural. A solução trouxe relatórios sobre a expressão dos sentimentos dos estudantes no Twitter, que por sua vez, têm a maior predominância de sentimentos negativos. Isso possibilitou, a partir dos autores supracitados, a dedução de que os alunos tiveram baixo rendimento causado pelo período pandêmico.

Este trabalho é composto por: introdução apresenta uma visão geral do artigo, destacando os principais temas abordados. Em seguida, é apresentado o referencial teórico, no qual são utilizadas fontes de autores externos e confiáveis para embasar o trabalho.

A metodologia é descrita detalhadamente, destacando o tipo de pesquisa realizada e os procedimentos adotados. A análise de resultados discute as informações coletadas e apresenta os principais achados da pesquisa. Por fim, as considerações finais resumem a análise realizada e apresentam as conclusões alcançadas a partir da pesquisa.

2 REFERENCIAL TEÓRICO

Neste tópico serão apresentados conceitos sobre redes sociais, análise de sentimentos e processamento de linguagem natural. Também serão apresentadas algumas das principais tecnologias que serão utilizadas durante o desenvolvimento. Para a construção do embasamento foram usados trabalhos de autores como: Medhat; Hassan; Korashy 2014.

2.1 REDES SOCIAIS

As redes sociais são aplicações que geram espaços virtuais para um grupo de pessoas se relacionarem por meio do envio de mensagens. De acordo com Marteleto (2001, p.72) as redes sociais, representam “[...] um conjunto de participantes autônomos, unindo ideias e recursos em torno de valores e interesses compartilhados”. Sendo ela parte de um grupo social que atua no cotidiano. Redes sociais na Internet possuem elementos característicos, que servem de base para que a rede seja percebida e as informações a respeito dela sejam compreendidas. (Recuero 2009).

Assim como exposto, o Twitter, mídia social de fama e uso mundial, é uma rede que permite a flexibilidade para os seus usuários para expressar sua livre opinião. Criado em março de 2006, é uma rede social gratuita que permite o compartilhamento de opiniões, informações e variados tipos de conteúdos.

O Twitter também é utilizado para compartilhar notícias, conhecimentos pessoais e fatos em geral, opiniões, eventos (Naaman; Boase, 2010). Com as hashtags, curtidas, tweets e retweets é possível acompanhar a opinião pública sobre cada determinado assunto. O mesmo é extremamente relevante para dois grupos de usuários: aqueles que estão em busca de informações e aqueles que desejam compartilhar informações de maneira ampla.

Uma das principais ferramentas do Twitter, é os *Trending Topics*, onde disponibiliza aos usuários os assuntos mais falados do mundo ou do país no momento. Também é possível fazer uso de hashtags, que servem para destacar palavras-chave e colocar em alta no *Trending Topics* o conteúdo do texto publicado de forma resumida (Santos, 2023).

O Twitter é uma plataforma importante para o engajamento e mobilização social. Através dos recursos de compartilhamento rápido de informações e de amplificação de conteúdo, os assuntos ganham visibilidade e podem alcançar um público vasto em tempo real. Movimentos como o ativismo online, campanhas de conscientização e petições têm encontrado no Twitter um espaço propício para difundir suas mensagens e promover mudanças na sociedade. A capacidade de conectar pessoas de diferentes partes do mundo, permitindo o acesso a

perspectivas diversas e a oportunidade de participar de discussões relevantes, faz do Twitter uma ferramenta poderosa no cenário atual de interconectividade global.

No próximo tópico serão expostos detalhes sobre análise de sentimentos, que é essencial para compreender as emoções e opiniões expressas nas redes sociais.

2.2 ANÁLISE DE SENTIMENTOS

A análise de sentimentos desempenha um papel fundamental na compreensão e interpretação das informações compartilhadas nas plataformas de mídia social. Com o crescimento exponencial do volume de dados gerados nessas redes, torna-se cada vez mais desafiador extrair insights valiosos dessas informações. A análise de sentimentos surge como uma abordagem promissora para lidar com esse desafio, permitindo a categorização automática dos sentimentos expressos nos textos, bem como a identificação de sua polaridade.

Segundo Liu (2012), a análise de sentimentos é um campo de estudo que se dedica a examinar as emoções, avaliações, atitudes e sentimentos associados a diversos elementos, como organizações, serviços, produtos, indivíduos, eventos, questões, tópicos e seus atributos. Considerando a influência significativa que as opiniões exercem no comportamento das pessoas, a análise de sentimentos se mostra uma ferramenta extremamente valiosa. Ao proporcionar uma análise aprimorada e uma alocação mais eficiente dos recursos relacionados a essas entidades, essa avaliação das opiniões produz resultados significativos.

O principal propósito da análise de sentimentos é categorizar automaticamente os sentimentos expressos em um texto, utilizando um filtro de palavras. Segundo Medhat, Hassen e Korashy (2014), a finalidade desse processo é extrair opiniões, identificar os sentimentos expressos e classificar sua polaridade.

A utilização da análise de sentimentos proporciona uma forma automatizada de averiguar as redes sociais identificando nas publicações os sentimentos imersos nelas. A utilização desses dados pode auxiliar nas tomadas de decisões, fazendo com que essas sejam mais eficientes e detenham maiores chances de serem bem-sucedidas.

A percepção sobre a análise de sentimento mencionada por Hussein (2018) destaca a importância dessa prática para identificar e extrair informações subjetivas presentes no texto. Ao aplicar técnicas de processamento de linguagem natural e análise de texto, é possível compreender as emoções, opiniões e atitudes expressas pelos indivíduos, fornecendo uma visão mais abrangente do conteúdo textual. Essa abordagem permite uma análise mais precisa das reações emocionais dos estudantes, contribuindo para a compreensão dos impactos psicológicos em contextos acadêmicos e aprimorando estratégias de apoio e intervenção.

Para compreender as emoções expressas em textos, é necessário realizar uma análise profunda de sentimentos. Essa abordagem busca entender os sentimentos presentes na comunicação humana, como textos e avaliações.

A análise profunda de sentimentos requer o uso do processamento de linguagem natural (PLN). O PLN capacita as máquinas a compreender e interpretar a linguagem humana, sendo essencial na identificação e classificação das expressões emocionais presentes nos textos analisados. Essa combinação de análise de sentimentos e PLN proporciona insights valiosos sobre as opiniões e emoções das pessoas. Com o avanço contínuo do PLN, podemos esperar melhorias na análise de sentimentos, resultando em análises mais precisas e

detalhadas. A união entre a análise profunda de sentimentos e o PLN traz benefícios para diversas áreas da sociedade.

2.3 PROCESSAMENTO DE LINGUAGEM NATURAL

O Processamento de Linguagem Natural (PLN) é um campo de pesquisa em inteligência artificial que busca formas de utilizar computadores para manipular e compreender a linguagem natural. De acordo com Allen (2003), um dos desafios fundamentais do PLN é lidar com a ambiguidade presente na linguagem humana. Brascher (2002) destaca e apresenta exemplos das diferentes formas de ambiguidade na linguagem natural:

Ambiguidade lexical: ocorre quando não é possível interpretar o significado de uma palavra isolada. Por exemplo, a palavra "manga" pode se referir à fruta manga ou a manga da camiseta.

Ambiguidade semântica: ocorre quando uma frase pode ter várias interpretações possíveis devido ao relacionamento dos termos. Por exemplo, a frase "Ela não chora mais porque ele partiu". (ou ela chorava porque ele havia partido, ou ela parou de chorar uma vez que ele já foi embora).

Ambiguidade pragmática: está relacionada ao contexto e à situação do falante no momento da comunicação. Por exemplo, na frase "Paulo vai à escola", não fica claro se Paulo é um estudante ou se está apenas indo à escola naquele momento.

Ambiguidade predicativa: ocorre quando é necessário interpretar as relações entre o predicado, os participantes e os argumentos. Por exemplo, na frase "Eu a deixei feliz. (feliz = ou atributo do sujeito ou atributo do objeto).

Ambiguidade sintática: surge na estruturação das frases em constituintes hierárquicos. Por exemplo, na frase "o policial viu o ladrão com a câmera.", não fica claro se o policial estava com uma câmera ou se o ladrão estava com uma câmera quando foi visto.

Além disso, a resolução da ambiguidade também contribui para a extração de informações relevantes e precisas de grandes volumes de texto, auxiliando em tarefas como análise de sentimentos, análise de opiniões e extração de conhecimento. A contínua pesquisa e avanços no campo do Processamento de Linguagem Natural visam aprimorar cada vez mais a capacidade dos sistemas de compreender e resolver a ambiguidade presente na linguagem humana, tornando a interação com a tecnologia mais intuitiva e eficiente.

Para a execução do processamento de linguagem natural com a análise de sentimentos de forma automatizada é utilizado a abordagem de aprendizado de máquina.

2.4 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 "Esse semestre está muito exaustivo! Os professores só estão preocupados em fechar a unidade no prazo." o sentimento associado ao semestre é decepção.

Os dados que serão utilizados para a produção deste modelo são chamados de dados de treinamento. Essa base de dados será uma que já foi pré-classificada, isto é, que já reconhece esses padrões descritos 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 pronta 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 Máquinas de vetores de suporte 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).

No evento *Computer on the Beach* (2018) é citado o método que utiliza palavras-chaves, dicionários ou estatística para a classificação, para assim reconhecer e identificar dados chaves que se enquadram com os dados em análise. Isso torna necessário o tratamento e filtragem desses dados. Esse é o método não supervisionado, que não necessita de uma base de dados de treinamento. Este método tem como principal vantagem a independência do domínio, isto é, este modelo não tem como base um conjunto limitado de treinamento (Sonj; Patel, 2014). O método permite que não se limite apenas a base de dados que já possui, mas também aprendendo conforme novos dados são adicionados.

Portanto, a análise de sentimentos por meio da abordagem de aprendizagem de máquina permite a identificação e categorização automatizada de sentimentos em textos, contribuindo para uma compreensão mais abrangente das opiniões e atitudes expressas pelos indivíduos. Essa análise é realizada tanto por meio de métodos supervisionados, que dependem de dados de treinamento previamente classificados, quanto por métodos não supervisionados, que utilizam técnicas de classificação baseadas em palavras-chave, dicionários ou estatísticas, permitindo um aprendizado contínuo e independente do domínio.

Para o desenvolvimento do trabalho foi escolhida a linguagem de programação *Python*, que será detalhada no próximo tópico.

2.5 PYTHON

Python é uma linguagem de altíssimo nível orientada a objeto, de tipagem dinâmica e forte, interpretada e interativa (Borges, 2010). De acordo com Borges (2014), possui uma sintaxe clara e concisa, favorecendo a legibilidade do código fonte e com isso, torna-se possível aumentar a produtividade no desenvolvimento utilizando esta linguagem.

Atualmente, ela é uma das linguagens de programação mais referenciadas em análise de dados e ferramentas de programação, devido à sua facilidade de aprendizagem e ampla aplicabilidade. O *Python* foi desenvolvido por Guido van Rossum em 1991, seu nome veio de uma série humorística britânica *Monty Python's Flying Circus* (Santana; Galesi, 2010).

De acordo com o Índice de Comunidade de Programação da TIOBE - *The Software Quality Company* (2023), que mede a popularidade das linguagens de programação, o *Python* ocupa a primeira posição em maio de 2023. Além de ser fácil de aprender também é multiplataforma, ou seja, funciona em diversos sistemas operacionais.

“Nos últimos dez anos, *Python* passou de uma linguagem de computação científica inovadora, ou para ser usada "por sua própria conta e risco", para uma das linguagens mais importantes em ciência de dados, aprendizado de máquina (*machine learning*) e desenvolvimento de *softwares* em geral, no ambiente acadêmico e no mercado [...]” (Mckinney, 2018, p.20). Sua escolha para esse projeto foi pela facilidade em trabalhar nessa área de dados.

3 METODOLOGIA

A metodologia tem como objetivo apresentar os meios de pesquisa a serem utilizados para corroborar com a proposta do estudo referido, visando demonstrar de maneira clara o objetivo da pesquisa.

O tipo de pesquisa é exploratória e descritiva, adequada para investigar a relação entre o baixo rendimento acadêmico e os sentimentos dos estudantes. Essa abordagem permitirá uma análise inicial e descritiva dos dados coletados no Twitter, buscando compreender melhor essa relação.

No que diz respeito à abordagem, se caracteriza como quantitativa, que envolve o uso de técnicas de análise de dados do Twitter, como mineração de texto e processamento de linguagem natural (PLN), para identificar sentimentos em relação à universidade nas postagens dos estudantes.

Essa pesquisa se qualifica como bibliográfica que, para Fonseca (2002) configura-se:

[...] A partir do levantamento de referências teóricas já analisadas, e publicadas por meios escritos e eletrônicos, como livros, artigos científicos, páginas de web sites. Qualquer trabalho científico inicia-se com uma pesquisa bibliográfica, que permite ao pesquisador conhecer o que já se estudou sobre o assunto. Existem porém pesquisas científicas que se baseiam unicamente na pesquisa bibliográfica, procurando referências teóricas publicadas com o objetivo de recolher informações ou conhecimentos prévios sobre o problema a respeito do qual se procura a resposta (Fonseca, 2002, p. 32).

Quanto aos procedimentos, a utilização de uma máquina de aprendizado para identificar e classificar os sentimentos negativos expressos nas postagens, é utilizado o modelo de aprendizado de máquina supervisionado, não supervisionado e processamento de linguagem natural.

Durante a primeira parte do projeto foi necessário a identificação das técnicas que foram utilizadas para análise de sentimento. Existem dois tipos de técnicas que foram utilizados: A abordagem baseada em análise léxica usada a partir do processamento de linguagem natural; e a abordagem baseada em aprendizado de máquina, através da qual foi criado um programa para o treinamento com uma base de dados que contém textos que já foram validados na identificação de sentimentos.

A segunda parte do projeto foi o desenvolvimento da aplicação para a coleta e armazenamento dos dados do Twitter, que pode ser encontrado nesse link (<https://project-tcc.vercel.app/>), fazendo a filtragem de dados desnecessários e buscando somente os dados que tiverem relevância durante os quatro últimos anos, de acordo com que foi proposto pelo projeto para ser tratado e validado os sentimentos e emoções negativas referente às universidades.

A terceira parte foi feito o tratamento dos dados removendo tudo que não seja relevante como links, hashtags e qualquer outra coisa escrita que não seja compreendido pelo processamento da linguagem natural. Para isso foi desenvolvido um script utilizando a linguagem de programação *Python* para simplificar o trabalho de limpeza dos dados.

A quarta parte foi feito o tratamento dos dados da base de dados de treinamento para separar os sentimentos nos quais foram feitas as análises. Para isso foi necessário identificar o que foi fornecido pela base e quais são as “descrições” dos sentimentos classificados como positivos, negativos e neutros.

A quinta parte foi desenvolver a *machine learning* que utilizou a abordagem de aprendizado de máquina não supervisionada e a abordagem de processamento de linguagem natural, para que a máquina fizesse interpretação do sentimento descrito. Para isso, foi criado uma “mochila de palavras”, que consiste em vetorizar as palavras para contabilizar a repetição de palavras. Também foi utilizado o TF-IDF (*Term Frequency-Inverse Document Frequency*, em português Frequência Termo-Frequência Inversa do Documento), para assim conseguir medir a frequência de repetição, a qual leva em conta a importância relativa de uma palavra em relação a um conjunto de documentos. A *machine learning* foi treinada com uma base de dados de treino com os sentimentos pré-classificados, para assim, identificar os padrões necessários para reconhecer os sentimentos encontrados na base de dados do Twitter.

A partir dos dados coletados, foi possível realizar uma conclusão embasada, a qual foi apresentada por meio da análise de sentimentos e da visualização de nuvens de palavras e gráficos para identificar os principais sentimentos expressos nos textos. Com base nessas informações, os resultados da pesquisa foram utilizados para elaborar gráficos que evidenciam o impacto dos sentimentos negativos no desempenho dos estudantes brasileiros, considerando uma perspectiva psicológica.

4 ANÁLISE DE RESULTADOS

O trabalho tem como proposta demonstrar a partir da análise e classificação dos sentimentos expressos no Twitter por estudantes, fatores que impactaram no rendimento acadêmico. Para desenvolvê-lo foram feitos procedimentos metodológicos, por exemplo, o desenvolvimento de uma aplicação para fazer a coleta de dados e de uma *machine learning* capaz de identificar os sentimentos expressos nos textos das postagens.

Sentimentos são as reações subjetivas e conscientes que surgem em resposta a eventos, experiências e interações, refletindo a experiência pessoal e individual da emoção. Eles envolvem uma combinação complexa de processos cognitivos, fisiológicos e sociais, dando forma à experiência afetiva única de cada indivíduo. (Davidson; Scherer; Goldsmith, 2003). Podemos categorizar os sentimentos em três amplas categorias: positivos, negativos e neutros

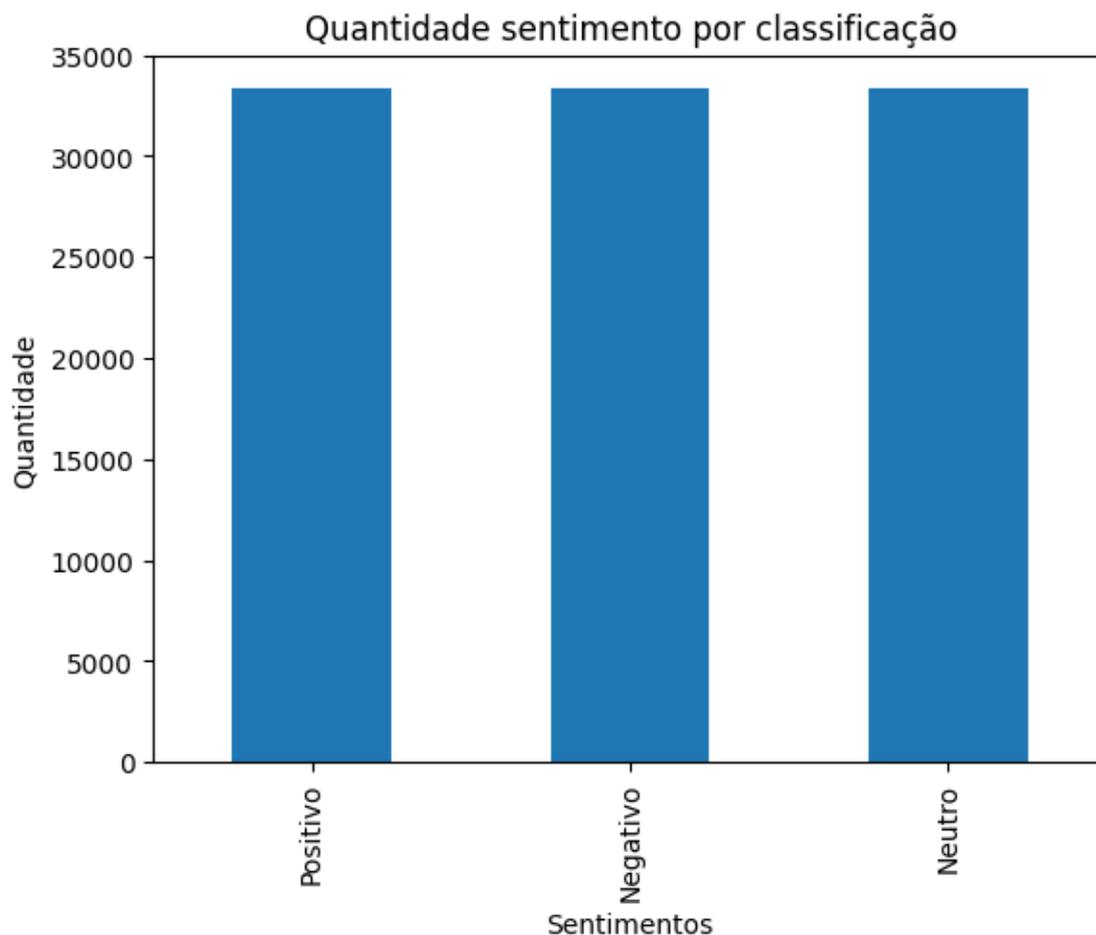
Os sentimentos positivos são geralmente experimentados como alegres e satisfatórios. Eles trazem uma boa sensação. Alguns exemplos comuns de sentimentos positivos incluem felicidade, alegria, gratidão, esperança, amor, entusiasmo, confiança e serenidade.

Os sentimentos negativos são geralmente experimentados como desagradáveis ou perturbadores. Eles envolvem emoções como tristeza, raiva, medo, culpa, frustração, inveja e ansiedade.

Os sentimentos neutros são aqueles que não são especificamente positivos nem negativos. Eles são geralmente experimentados como uma sensação de calma, indiferença ou neutralidade emocional.

Para classificação foi utilizado uma base de dados textuais do *kaggle* que foi coletada do twitter para fins de criação de uma base de treinamento da *machine learning*, nela contém 100.000 tweets que já foram classificados em 3 classes de sentimentos: positivo, negativo ou neutro, distribuídos de forma equilibrada, exibidos abaixo no gráfico 1 e demonstrado uma prévia na figura 1.

Gráfico 1: Distribuição de base de treino por classificação de sentimento



Fonte: De autoria própria (2023)

Figura 1: Exemplos de dados da base de treinamento

75463	Tartaruga ganha cadeira de rodas feita de Lego...	Neutro	g1
53367	@katrisilva1 @1IsabelleLima Vcs q me esquecera...	Negativo	:(
58041	@jeonttuntold te amo mais :(Negativo	:(
54909	Visao q eu amaria ter :(https://t.co/XMnGbNy9FE	Negativo	:(
62354	Ela deve ter ficado sem créditos :(https://t.co/...	Negativo	:(
95515	#Trabalho: Governo aumenta para R\$ 1.006 previ...	Neutro	#trabalho
21108	Não espere das pessoas aquilo que vc faria por...	Positivo	:)
52024	Fico vendo esse vídeo pra animar um pouco minh...	Negativo	:(
62398	@ihatesnakeau nem vejo mas wofnkwndke acompanh...	Negativo	:(
1627	#AMANHÃ é o grande dia :D https://t.co/SrVnW79fBx	Positivo	:)
93410	A pessoa que você gosta passa a noite pensando...	Neutro	#fato
56562	@nunes_maiza muita saudade :(❤️ podia voltar	Negativo	:(
57599	Agora vou parar de postar .. infelizmente tenh...	Negativo	:(
63961	@brszzz nunca mãe me chamou :((*	Negativo	:(
44242	@gwbrielaf parabéns nordeste :(Negativo	:(

Fonte: De autoria própria (2023)

Os dados para o treino da *machine learning*, foram previamente tratados antes do uso para o treinamento, para isso, foram utilizadas algumas bibliotecas *Natural Language Toolkit (NLTK)*, utilizada para o processamento de texto para classificação, tokenização, lematização, marcação, análise e raciocínio semântico. Em resumo, fornece recursos para a máquina conseguir compreender textos humanos.

Outra biblioteca que foi fundamental para o desenvolvimento é a *scikit-learn (sklearn)*, ela fornece recursos necessários para o aprendizado de máquina, como exemplos, para criação de uma máquina de aprendizado utilizando o método *Naive Bayes Multinomial* que implementa o algoritmo para dados com a distribuição multinomial e é uma das duas variantes clássicas de *Naive Bayes* usadas na classificação de texto (onde os dados são normalmente representados como contagens de vetores de palavras).

Uma adaptação de uma biblioteca utilizado, foi o LelA (Léxico para Inferência Adaptada) que é um *fork* do léxico e ferramenta para análise de sentimentos *VADER (Valence Aware Dictionary and Sentiment Reasoner)* adaptado para textos em português, com suporte para emojis e foco na análise de sentimentos de textos expressos em mídias sociais.

Também foram utilizadas, algumas bibliotecas para tratamento de texto, fazer requisição à aplicação que foi desenvolvida para a coleta dos dados, organização de *dataframe* e algumas bibliotecas para montagem de gráficos demonstrativos, estatísticos e matemáticos.

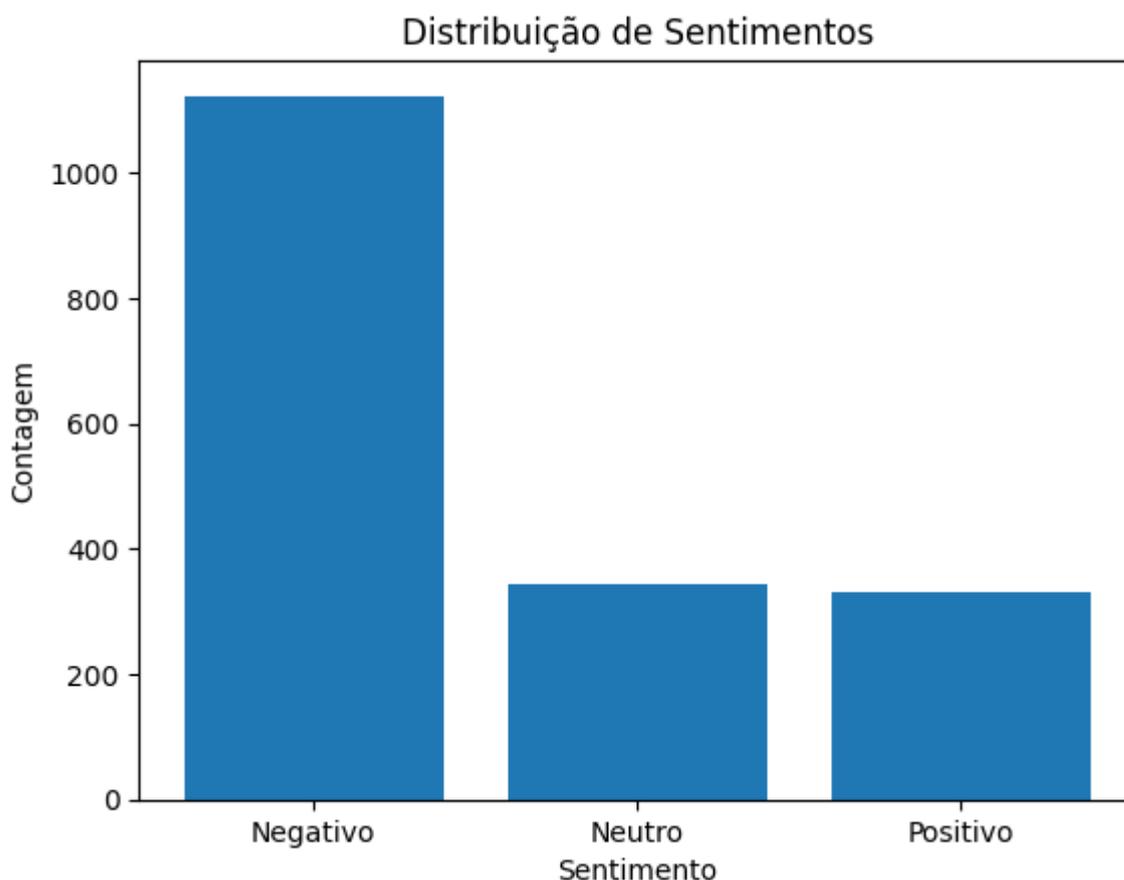
Após treinar a máquina, foi feita a análise pela *machine learning* dos dados que foram coletados de postagens dos estudantes no período de maio de 2020 até

maio de 2023, utilizando no filtro de busca do twitter a palavra chave “aula online”, totalizando 1800 postagens extraídas.

Com o resultado da análise de dados, foi atingido um nível de acurácia de 81%, é importante mencionar que algoritmos de *machine learning* podem ter dificuldade em interpretar termos específicos, como gírias ou palavras escritas incorretamente, o que pode afetar levemente a precisão da análise.

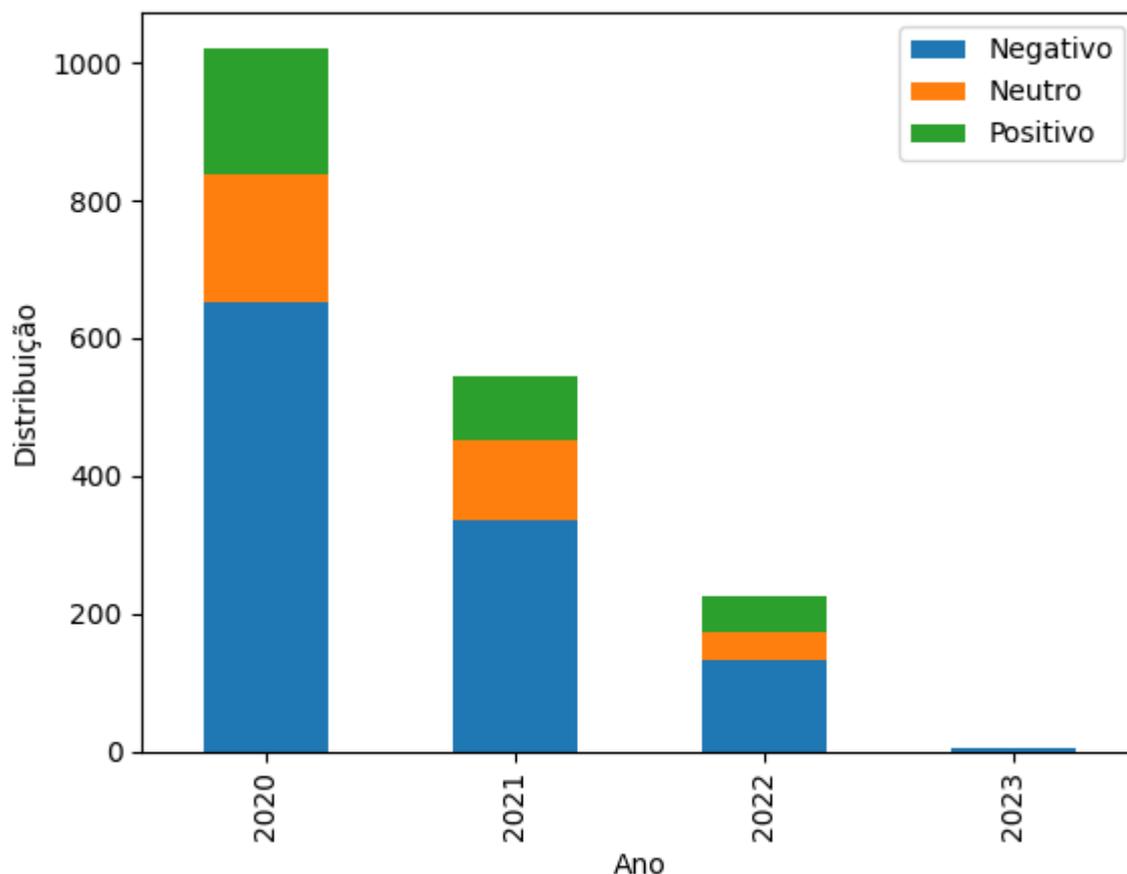
Com isso foi possível identificar uma maior existência de sentimentos negativos nas postagens dos estudantes, exibido no gráfico 2 logo abaixo, sendo 1123 negativos, 331 positivos e 343 neutros.

Gráfico 2: Classificação dos sentimentos analisados nas postagens



Fonte: De autoria própria (2023)

Pode-se observar na figura 2 uma prévia das postagens que foram classificadas.

Gráfico 3: Distribuição dos dados por ano

Fonte: De autoria própria (2023)

A distribuição por ano mostra que a maioria dos dados são de 2020, ano de início da pandemia, foi neste período que mais postagens foram encontradas. Nos anos seguintes, as postagens com a palavra “aula online” ficaram menos recorrentes.

Durante a pandemia foi demonstrado através da análise dos dados que o confinamento causou maior impacto negativo nos estudantes, o isolamento prolongado pode levar ao aumento do estresse, ansiedade, depressão e sentimentos de solidão, afetando diretamente o desempenho acadêmico.

5 CONSIDERAÇÕES FINAIS

Neste trabalho, foi proposto uma abordagem para analisar e classificar os sentimentos expressos por estudantes no Twitter, com o objetivo de identificar os fatores que impactaram seu rendimento acadêmico. Ao longo do estudo, desenvolvemos uma aplicação para coleta de dados e utilizamos técnicas de aprendizado de máquina para realizar a classificação dos sentimentos presentes nas postagens feitas por estudantes no twitter.

Os principais resultados da pesquisa revelaram uma predominância de sentimentos negativos nas postagens dos alunos, indicando que o confinamento decorrente da pandemia teve um impacto negativo significativo em seu bem-estar emocional e, conseqüentemente, em seu desempenho acadêmico. Como foi relatado pelos autores Gudin et al(2021), Araújo et al(2020), Maia et al(2020),

Schmidt et al(2020) e demais, o isolamento prolongado gerou um aumento do estresse, ansiedade, depressão e sentimentos de solidão.

Dessa forma, foi possível responder ao problema levantado neste trabalho, que, a partir da análise de dados feita pela *machine learning* foi possível identificar a predominância de sentimentos negativos dos estudantes nas postagens. A análise dos sentimentos expressos no Twitter permitiu compreender a importância do contexto socioemocional na educação durante a pandemia, destacando a necessidade de atenção e suporte aos alunos em períodos desafiadores.

Ao comparar/conferir os resultados com a hipótese apresentada, manteve-se que o impacto negativo causado pelo confinamento e isolamento social estava atendendo com a expectativa inicial. A predominância de sentimentos negativos nas postagens validou a ideia de que as condições adversas impostas pela pandemia afetaram a saúde emocional dos estudantes e, por consequência, seu desempenho acadêmico.

Com esses resultados, pode-se classificar o projeto como funcional e inovador, trazendo um método objetivo de coleta e processamento de dados, podendo ser adaptado para outras abordagens, caso necessário.

As contribuições deste trabalho são diversas. Primeiramente, uma aplicação desenvolvida para a coleta de dados possibilita o monitoramento dos sentimentos dos estudantes. Além disso, uma análise de sentimentos por meio de técnicas de aprendizado de máquina contribui para uma compreensão mais ampla dos fatores emocionais envolvidos no processo de aprendizagem durante situações adversas.

É importante ressaltar que o uso de algoritmos de *machine learning* possui limitações, como a dificuldade de interpretar termos específicos e gírias, o que pode afetar levemente a precisão das análises. Portanto, é fundamental considerar essas limitações ao interpretar os resultados.

No contexto educacional, as descobertas deste estudo destacam a importância de investimentos em suporte socioemocional para os estudantes, especialmente em períodos desafiadores como o vivenciado durante uma pandemia. Estratégias de apoio psicossocial, como programas de aconselhamento e acompanhamento emocional, podem contribuir significativamente para a melhoria do bem-estar dos estudantes e, conseqüentemente, para o seu rendimento.

Em suma, este trabalho demonstra a influência dos sentimentos expressos no Twitter pelos estudantes em relação ao seu rendimento acadêmico durante a pandemia. Os resultados obtidos ressaltaram que em períodos desafiadores a falta de auxílio psicológico pode influenciar diretamente e negativamente no rendimento dos estudantes.

REFERÊNCIAS

- ALLEN, J. F. **Natural language processing**. John Wiley and Sons Ltd., 2003.
- ARAÚJO, F. J. de, O., de Lima, LSA, Cidade, PIM, Nobre, CB, and Neto. MLR, 2020. **Aprendendo com o Imprevisível: Saúde Mental dos Universitários e Educação Médica na Pandemia de Covid-19**
- BORGES, Luis E. **Python para desenvolvedores**. 2a. ed. Rio de Janeiro: Edição do Autor, 2010.
- BRASCHER, M. **A ambigüidade na recuperação da informação**. IASI, 2002.
- CAO, Wenjun et al. **The psychological impact of the COVID-19 epidemic on college students in China**. *Psychiatry research*, v. 287, p. 112934, 2020.
- 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: 12 dez. 2022.
- DAVIDSON, R.J., SCHERER, K.R., GOLDSMITH, H.H. (Eds.). **Handbook of Affective Sciences**. Oxford University Press, 2003.
- FONSECA, J. J. S. **Metodologia da pesquisa científica**. Fortaleza: UEC, 2002. Apostila.
- Gundim, V. A. Encarnação, J. P. da, Santos, F. C., Santos, J. E. dos, Vasconcellos, E. A., & Souza, R. C. de. (2020). **Saúde mental de estudantes universitários durante a pandemia de covid-19**. *Revista Baiana De Enfermagem*, 35. <https://doi.org/10.18471/rbe.v35.37293>
- 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.
- MAIA, Berta Rodrigues; DIAS, Paulo César. **Ansiedade, depressão e estresse em estudantes universitários: o impacto da COVID-19**. *Estudos de psicologia (Campinas)*, v. 37, 2020.
- 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
- MCKINNEY, Wes. **Python para Análise de Dados**. 2. ed. São Paulo: Novatec, 2018.
- 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.

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: 07 dez. 2022.

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

SANTOS, Daiana Kathrin Santana. **Análise de sentimentos dos usuários do Twitter sobre a eleição presidencial de 2022 no Brasil**. 2023

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: 17 dez. 2022.

TEIXEIRA, Melissa Ribeiro; DAHL, Catarina Magalhães. **Recriando cotidianos possíveis: construção de estratégias de apoio entre docentes e estudantes de graduação em Terapia Ocupacional em tempos de pandemia**. Revista Interinstitucional Brasileira de Terapia Ocupacional-REVISBRATO, v. 4, n. 3, p. 509-518, 2020

TIOBE - THE SOFTWARE QUALITY COMPANY. TIOBE Index for May 2021. [S. l.], 2023. Disponível em: <https://www.tiobe.com/tiobe-index/>. Acesso em: 13 jun. 2023.

SCHMIDT, Beatriz et al. **Saúde mental e intervenções psicológicas diante da pandemia do novo coronavírus (COVID-19)**. Estudos de Psicologia (campinas), v. 37, 2020.

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: 13 dez. 2022.

YUAN, Shuai et al. **Comparison of the indicators of psychological stress in the population of Hubei province and non-endemic provinces in China during two weeks during the coronavirus disease 2019 (COVID-19) outbreak in February 2020**. Medical science monitor: international medical journal of experimental and clinical research, v. 26, p. e923767-1, 2020

ZHANG, Yin; JIN, Rong; ZHOU, Zhi-Hua. Understanding Bag-of-Words Model: A Statistical Framework. **International Journal of Machine Learning and Cybernetics**, [s. l.], 32 2010. Disponível em: https://www.researchgate.net/publication/226525014_Understanding_bag-ofwordsmodel_A_statistical_framework. Acesso em: 16 dez. 2022.

GLOSSÁRIO

Kaggle: é uma plataforma de competição de ciência de dados e comunidade on-line de cientistas de dados e praticantes de aprendizado de máquina sob o Google LLC.

Lematização: é o ato de representar as palavras através do infinitivo dos verbos e masculino singular dos substantivos e adjetivos.

Machine learning: é um subcampo da Engenharia e da ciência da computação que evoluiu do estudo de reconhecimento de padrões e da teoria do aprendizado computacional em inteligência artificial.

Naive Bayes: é um algoritmo que gera uma tabela de probabilidades a partir de uma técnica de classificação de dados.

Natural Language Toolkit (NLTK): é algo como uma caixa de ferramentas para linguagem natural.

Processamento de linguagem natural (PLN): é uma subárea da ciência da computação, inteligência artificial e da linguística que estuda os problemas da geração e compreensão automática de línguas humanas naturais.

Raciocínio semântico: é o estudo da semântica ou interpretações (idealizações) de linguagens naturais e formais, geralmente tentando captar a noção pré-teórica de implicação (ou consequência lógica).

Scikit-learn: é uma biblioteca de aprendizado de máquina de código aberto para a linguagem de programação Python.

Tokenização: é transformar um ativo qualquer em uma representação digital, ou seja, que só existe no meio eletrônico.

Trending Topics: Traduzido do inglês - No Twitter, uma palavra, frase ou tópico mencionado em uma taxa maior do que outros é considerado um "trending topic" ou simplesmente uma "tendência".

APÊNDICES

Apêndice A: Interface da aplicação de coleta dos dados

Coleta de dados

Insira as informações solicitadas

Username

Data

Post

Posts (1800)

@rafawweber 05/08/2022
n consigo mais prestar atenção em aula online

@daniel_sturiao 08/08/2022
eu odeio aula online

@bianca_csousa 09/08/2022
pra mim essa aula online é tortura psicológica

@milennacordeiro 09/08/2022
eu sei que ninguém aguentava aula online mais, mas vocês não acham que a realidade de hoje também não comporta mais aula 100% presencial?

Apêndice B: Partes principais do código da *machine learning*

▾ Importações das bibliotecas

```
[ ] import os
import requests
import pandas as pd
import numpy as np
import random
import string
from collections import Counter
import re
import nltk
from nltk.tokenize import word_tokenize
from nltk.tokenize import wordpunct_tokenize
from nltk.stem import SnowballStemmer
from nltk.probability import FreqDist
from nltk.metrics import ConfusionMatrix
from wordcloud import WordCloud
import matplotlib.pyplot as plt
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfTransformer
from sklearn.naive_bayes import MultinomialNB, BernoulliNB
from sklearn.linear_model import SGDClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.svm import SVC
from sklearn.ensemble import RandomForestClassifier
from sklearn.pipeline import Pipeline
from sklearn import metrics
from sklearn.model_selection import GridSearchCV, train_test_split
from sklearn.metrics import accuracy_score
import importlib.util
from dateutil.parser import parse
```

1. substituição de emojis e emoticons

```
[ ] def replace_emojis(data, emoji_list):
    ls = []

    for line in data:
        for emoji in emoji_list:
            line = line.replace(emoji, ' ' + emoji_list[emoji] + ' ')

        ls.append(line)

    return ls
```

```
[ ] emoji_list = {
    '😊': 'positive_emoji', '😄': 'positive_emoji', '😁': 'positive_emoji', '😆': 'positive_emoji', '😂': 'positive_emoji', '😃': 'positive_emoji',
    '😅': 'positive_emoji', '😇': 'positive_emoji', '😈': 'positive_emoji', '😍': 'positive_emoji', '😘': 'positive_emoji', '😙': 'positive_emoji',
    '😚': 'positive_emoji', '😛': 'positive_emoji', '😜': 'positive_emoji', '😝': 'positive_emoji', '😞': 'positive_emoji', '😟': 'positive_emoji',
    '😠': 'positive_emoji', '😡': 'positive_emoji', '😢': 'positive_emoji', '😣': 'positive_emoji', '😤': 'positive_emoji', '😥': 'positive_emoji',
    '😦': 'negative_emoji', '😧': 'negative_emoji', '😨': 'negative_emoji', '😩': 'negative_emoji', '😪': 'negative_emoji', '😫': 'negative_emoji',
    '😬': 'negative_emoji', '😭': 'negative_emoji', '😮': 'negative_emoji', '😯': 'negative_emoji', '😰': 'negative_emoji', '😱': 'negative_emoji',
    '😲': 'negative_emoji', '😳': 'negative_emoji', '😴': 'negative_emoji', '😵': 'negative_emoji', '😶': 'negative_emoji', '😷': 'negative_emoji',
    '😸': 'neutral_emoji', '😹': 'neutral_emoji', '🙄': 'negative_emoji',
    '❤️': 'positive_emoji', '💕': 'positive_emoji', '💖': 'positive_emoji', '💗': 'positive_emoji', '💘': 'positive_emoji', '💙': 'positive_emoji', '💜': 'positive_emoji',
    '💚': 'positive_emoji', '💛': 'positive_emoji', '💜': 'positive_emoji', '💚': 'positive_emoji', '💛': 'positive_emoji', '💜': 'positive_emoji',
    '💖': 'positive_emoji', '💗': 'positive_emoji', '💘': 'positive_emoji', '💙': 'positive_emoji', '💜': 'positive_emoji',
    '👉': 'positive_emoji', '👈': 'positive_emoji', '👊': 'positive_emoji', '👋': 'positive_emoji'
}
```

```
[ ] X_values = replace_emojis(X_values, emoji_list)
for i in range(0, 5):
    print(X_values[i])
```

2. Mochila de palavras (BOW)

```
[ ] vectorizer = CountVectorizer()
X_vect = vectorizer.fit_transform(X_values)
```

```
[ ] print(X_vect.shape)
```

```
(100000, 39520)
```

```
[ ] print(vectorizer.vocabulary_.get(u'positive_emoji'))
```

```
29319
```

```
[ ] tfidf_transformer = TfidfTransformer()
X_tfidf = tfidf_transformer.fit_transform(X_vect)
```

```
[ ] print(X_tfidf.shape)
```

```
(100000, 39520)
```

```
[ ] X_train, X_test, y_train, y_test = train_test_split(X_tfidf, y_classification, test_size=0.2, random_state=9)
```

↳ Construindo um pipeline MultinomialNB

```
[ ] model_MNB = Pipeline([
    ('vect', CountVectorizer()),
    ('tfidf', TfidfTransformer()),
    ('clf', MultinomialNB()),
])
```

```
[ ] model_MNB.fit(X_train, y_train)
    predictions_MNB = model_MNB.predict(X_test)
```

```
[ ] acc = accuracy_score(predictions_MNB, y_test)
    print(acc)
```

0.81238

2. Requisição dos dados

```
▶ res = requests.get('https://project-tcc.vercel.app/api/posts')
    data_analyse = pd.read_json(res.text)
    print(data_analyse['comment'].values)
    data_analyse.to_csv('twitter_data.csv')
```

```
☹ ['achei que tava abalando faltando aula hoje e o professor manda mensagem colocando aula online'
   'Eu me sinto a pessoa mais burra do mundo. Eu achei que hoje era quarta (eu não sei o motivo) e que tinha aula'
   'Odeio aula ead, eu faço uma pergunta no chat e o professor demora pra responder AAAAAAAAAAAAAAAAA'
   ... 'pra mim essa aula online é tortura psicológica'
   'eu odeio aula online' 'n consigo mais prestar atenção em aula online']
```

```
[ ] analyse_corpus = data_analyse['comment'].values
    X_analyse, X_analyse_vect, X_analyse_tfidf=tratament(analyse_corpus)
```

```
[ ] predictions_MNB = model_MNB.predict(X_analyse)
    new_list=[]
    for doc, prediction in zip(analyse_corpus,predictions_MNB):
        new_list.append([doc,prediction])

    df = pd.DataFrame(new_list, columns=['text', 'sentiment'])
```

```
▶ sentiment_counts = df['sentiment'].value_counts()

    plt.bar(sentiment_counts.index, sentiment_counts.values)

    plt.xlabel('Sentimento')
    plt.ylabel('Contagem')
    plt.title('Distribuição de Sentimentos')

    plt.show()
```