



UFOP

Universidade Federal
de Ouro Preto

**Universidade Federal de Ouro Preto
Instituto de Ciências Exatas e Aplicadas
Departamento de Computação e Sistemas**

Cibele Francisca de Carvalho

**Manifestações de Inteligência
Compartilhada em Mídias Sociais**

**João Monlevade
2022**

Cibele Francisca de Carvalho

Manifestações de Inteligência Compartilhada em Mídias Sociais

Monografia apresentada ao curso de Engenharia de Computação do Instituto de Ciências Exatas e Aplicadas da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Engenharia de Computação.

Orientador: Diego Zuquim Guimarães Garcia.

João Monlevade

2022

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

C331m Carvalho, Cibele Francisca de.
Manifestações de inteligência compartilhada em mídias sociais.
[manuscrito] / Cibele Francisca de Carvalho. - 2022.
42 f.

Orientador: Prof. Dr. Diego Zuquim Guimarães Garcia.
Monografia (Bacharelado). Universidade Federal de Ouro Preto.
Instituto de Ciências Exatas e Aplicadas. Graduação em Engenharia de
Computação .

1. Engenharia de software. 2. Mídia social. 3. Twitter (Rede social on-
line). 4. Mineração de dados (Computação). 5. Mineração de dados
textuais. I. Garcia, Diego Zuquim Guimarães. II. Universidade Federal de
Ouro Preto. III. Título.

CDU 004.62:007

Bibliotecário(a) Responsável: Sione Galvão Rodrigues - CRB6 / 2526



FOLHA DE APROVAÇÃO

Cibele Francisca de Carvalho

Manifestações de inteligência compartilhada em mídias sociais

Monografia apresentada ao Curso de Engenharia de Computação da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Engenharia de Computação

Aprovada em 22 de junho de 2022

Membros da banca

Doutor - Diego Zuquim Guimarães Garcia - Orientador - Universidade Federal de Ouro Preto
Mestre - Euler Horta Marinho - Universidade Federal de Ouro Preto
Bacharel - Silvano Sergio Martins Oliveira

Diego Zuquim Guimarães Garcia, orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 17/07/2022



Documento assinado eletronicamente por **Diego Zuquim Guimaraes Garcia, PROFESSOR DE MAGISTERIO SUPERIOR**, em 17/07/2022, às 23:04, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0363533** e o código CRC **DAD50B3B**.

TERMO DE RESPONSABILIDADE

Eu, **Cibele Francisca de Carvalho** declaro que o texto do trabalho de conclusão de curso intitulado “*Manifestações de Inteligência Compartilhada em Mídias Sociais*” é de minha inteira responsabilidade e que não há utilização de texto, material fotográfico, código fonte de programa ou qualquer outro material pertencente a terceiros sem as devidas referências ou consentimento dos respectivos autores.

João Monlevade, 22 de Junho de 2022

Cibele Francisca de Carvalho

Dedico este trabalho as pessoas que direta ou indiretamente me ajudaram para que eu pudesse concluir a minha graduação, me apoiando e incentivando mesmo que nos momentos mais difíceis, não puderam estar ao meu lado.

Agradecimentos

Agradeço aos meus familiares e amigos que estiveram ao meu lado durante toda a minha trajetória acadêmica seria injusto citar somente alguns pois foram muitos que se dispuseram a me apoiar nos meus piores momentos, mas que também puderam conhecer a minha melhor versão.

Agradeço também a todos professores e funcionários da UFOP que durante todo esse tempo de vida universitária me fizeram uma pessoa melhor e mais humana para saber lidar com as adversidades da vida, e MUITO OBRIGADO UFOP por me dar a oportunidade de concluir a minha graduação em uma universidade de excelência e me apoiar durante todo esse tempo, acreditando em mim. Meu muito obrigado e que essa amizade e carinho possa durar por muitos anos.

Agradeço também a IBM que me deu a oportunidade de poder estagiar em uma grande multinacional e aos meus colegas de trabalho que sempre estão dispostos a me apoiar no meu crescimento profissional, são vocês que diariamente têm me moldado para que eu me torne uma profissional de excelência.

“Muda, que quando a gente muda o mundo muda com a gente. A gente muda o mundo na mudança da mente. E quando a mente muda a gente anda pra frente. E quando a gente manda ninguém manda na gente! Na mudança de atitude não há mal que não se mude nem doença sem cura. Na mudança de postura a gente fica mais seguro. Na mudança do presente a gente molda o futuro!”

— Gabriel, o Pensador,
em: Seja Você Mesmo (mas não Seja sempre o Mesmo).

Resumo

Este trabalho trás uma abordagem dos conteúdos que são compartilhados em mídias sociais e sua influência na tomada de decisão no cotidiano das pessoas. As novas formas instantâneas nas trocas de informações e o considerável aumento no número de usuários que possuem acesso à internet fizeram com que houvesse um crescimento exponencial dos dados compartilhados, aumentando o número de estudos comportamentais desses conteúdos compartilhados em mídias sociais, juntamente a isso viu-se a necessidade na automatização de processos para respostas rápidas, devido ao enorme volume e a velocidade em que os dados eram gerados. Neste trabalho foi utilizado a Mineração de Textos na determinação de padrões interessantes e as coletas se deram a partir do código desenvolvido na linguagem de programação *Python*, para o desenvolvimento deste trabalho foi feito uma classificação e avaliação dos dados. Alguns softwares e mídias sociais foram considerados para fazer a coleta dos dados no desenvolvimento do trabalho. A escolha do *Twitter* se deu por ser uma mídia social dinâmica e permitir o compartilhamento de *tweets* sem que haja uma conexão entre os usuários e as informações serem postadas constantemente.

Palavras-chaves: Mídias Sociais. Twitter. Mineração de texto

Abstract

This work brings an approach to the content that is shared on social media and its influence on decision-making in people's daily lives. The instant forms of studies of children or increase in exchanges do not have access to the internet with new users that have grown exponentially from a number of shared data, increasing the number of behavioral studies of these users in social ways, creating the need for automation of processes for quick responses, due to the sheer volume and speed at which the data is generated. In this work we used Text Mining in the determination of interesting patterns and the mining was developed with code in the *Python* programming language, for the development of this study a classification and evaluation of the data was made. Some software and social media were considered to collect data in the development of the work. The choice of *Twitter* was due to the site being a dynamic social media and allow sharing of *tweets* without there being a connection between users and the information constantly posted.

Keywords: Social Media. Twitter. Text Mining

Lista de ilustrações

Figura 1 – <i>Erro ao retornar mais de 2500 tweets</i>	27
Figura 2 – <i>Tweets compartilhados</i>	28
Figura 3 – <i>Análise de Sentimentos #educação no período da manhã</i>	30
Figura 4 – <i>Palavras mais citadas no período da manhã</i>	30
Figura 5 – <i>Análise de Sentimentos #educação no período da tarde</i>	31
Figura 6 – <i>Palavras mais citadas no período da tarde</i>	32
Figura 7 – <i>Análise de Sentimentos #educação no período da noite</i>	33
Figura 8 – <i>Palavras mais citadas no período da noite</i>	34
Figura 9 – <i>Trigrams - Manhã</i>	35
Figura 10 – <i>Trigrams - Tarde</i>	36
Figura 11 – <i>Trigrams - Noite</i>	36
Figura 12 – <i>Cluster - Manhã</i>	38
Figura 13 – <i>Cluster - Tarde</i>	38
Figura 14 – <i>Cluster - Noite</i>	39

Lista de tabelas

Tabela 1 – Número de <i>tweets</i> e polaridades	29
--	----

Lista de abreviaturas e siglas

API Application Programming Interface

NLP Natural Language Processing

PEC Proposta de Emenda Constitucional

VSCoDe Visual Studio Code

Sumário

1	INTRODUÇÃO	14
1.1	Problema	15
1.2	Justificativa	16
1.3	Objetivos	16
1.3.1	Objetivo geral	16
1.3.2	Objetivos específicos	17
1.4	Estrutura do trabalho	18
2	REVISÃO BIBLIOGRÁFICA	19
2.1	Referenciais Teóricos	19
2.2	Trabalhos Relacionados	20
3	METODOLOGIA	22
3.1	Levantamento dos dados que serão utilizados	23
3.2	Coleta dos dados	23
3.3	Avaliação dos dados	24
4	DESENVOLVIMENTO	26
4.1	Visão geral	26
4.2	Testes	26
5	APRESENTAÇÃO E ANÁLISE DOS RESULTADOS	28
5.1	Coleta e análise dos <i>tweets</i>	28
5.2	Período das coletas	29
5.2.1	Coletas realizadas no período da manhã	29
5.2.2	Coletas realizadas no período da tarde	31
5.2.3	Coletas realizadas no período da noite	32
5.2.4	Comparação dos resultados	34
5.2.5	Clusterização	37
6	CONCLUSÃO	40
6.1	Trabalhos Futuros	41
	REFERÊNCIAS	42

1 Introdução

Com o advento da internet e o pluralismo do seu ambiente surgiram práticas de interação social que transformaram o modo como a sociedade se comunica e se organiza de acordo com os grupos de seu interesse. Juntamente com essa transformação e reorganização nas comunicações, houve também uma evolução das ferramentas da internet que partiram – "desde a Web 1.0, caracterizada pelos sites estáticos e pela interatividade limitada, nos anos 90, até as plataformas da Web 2.0, cuja ênfase está na interação social, no compartilhamento de informações e no acesso às ferramentas de produção" (ACKLAND, 2013; BAYM, 2010). Na Web 3.0 ou Web semântica todas as informações são categorizadas e armazenadas de forma que os computadores compreendam tanto quanto um ser humano, isso muitas vezes é visto como uma combinação de Inteligência Artificial e Web Semântica, no qual a web semântica ensinará o computador o que os dados significam e a inteligência artificial poderá utilizar essa informação ¹.

Entende-se por inteligência compartilhada, a ação que menos exige dos intelectuais coletivos, permitindo a divulgação e publicação de conteúdos compartilhados por indivíduos isolados nas redes sociais, gerando um impacto na coletividade. Esses compartilhamentos podem estar voltados para algo relevante que esteja acontecendo na internet e ser caracterizado como agregação de sujeitos, para um melhor uso e disseminação do conhecimento das pessoas para um fim em comum.

As redes sociais por serem de baixo custo, descentralizada e simultânea, são um ambiente propício para trocas no dia a dia, com uma comunicação interpessoal rápida, que podem ser privadas ou em grupos. Esses grupos ou comunidades são formados por interesses em comum, compartilham um grande volume de conteúdos com assuntos de seus interesses (BAYM, 2010; BURGESS; GREEN, 2009; JENKINS; FORD; GREEN, 2013), além de permitem o armazenamento e a recuperação de uma enorme quantidade de dados eficientemente, o foco está em desenvolver metodologias para extração de informações e criação de conhecimento (BENEVENUTO; RIBEIRO; ARAÚJO, 2015).

As empresas usam cada vez mais a internet para divulgar seus produtos através dos influenciadores digitais para que possam representar a marca. Eles são contratados e através de suas plataformas digitais divulgam os produtos para seus seguidores e os influenciam a consumir o produto. No entanto, as plataformas digitais também são palco para interações acerca de determinados assuntos que está gerando discussões no momento, havendo uma alta interação entre os usuários, que muitas vezes podem estar interligados ou não. Podemos exemplificar as *hashtags* que são um meio de interação entre usuários de

¹ <https://www.lifewire.com/what-is-web-3-0>

diferentes localizações geográficas que interagem entre si, por meio destas.

Segundo (BAYM, 2010) são enumerados sete conceitos-chave para refletir acerca de como o uso de tecnologias digitais (especificamente internet e celulares) modificam as interações e, potencialmente, os relacionamentos mediados. São eles: a) a interatividade – social (conexões entre pessoas e grupos) ou técnica (interação com a tecnologia e suas plataformas); b) a estrutura temporal; c) as "pistas" sociais – informações sobre contexto, conteúdos, significados e identidades; d) o alcance; e) a replicabilidade; f) o arquivamento; e g) a mobilidade (Baym, 2010, p. 6-12). Como há diferentes ambientes para interações entre os usuários, essas características devem ser observadas a priori como condições das interações medianas.

Assim com a proliferação da internet houve uma aceleração nos meios de comunicação gerando uma maior interatividade entre os povos, uma aceleração na globalização e fez com que as pessoas se tornassem dependentes dos serviços disponibilizados na web. Segundo (MACHADO; TIJIBOY, 2005) as redes sociais visam impulsionar as relações humanas através da tecnologia. Tornando a internet um grande gerador de dados para as empresas e também para estudos comportamentais das pessoas nas mídias sociais através das análises dados. Segundo (GOMES, 2013) desta forma o *Text Mining*, em particular a Análise de Sentimentos gerou grande interesse, uma vez que permite de uma forma automatizada tratar e analisar grandes volumes de dados não estruturados e gerar conhecimento.

1.1 Problema

Ao se trabalhar com a coleta de dados em mídias sociais, algumas questões devem ser levadas em consideração, como os problemas éticos e as limitações dos dados que serão visualizados e analisados. De acordo com (MARKHAM; BUCHANAN, 2012) alguns elementos abordados devem ser levados em consideração como: vulnerabilidade da população estudada, possível prejuízo dos sujeitos pesquisados, balanço entre direitos e benefícios da pesquisa, questões éticas que podem surgir durante a pesquisa e um maior número de fontes para decisão.

Logo após a análise de que o trabalho está em conformidade com as questões éticas, será feito através da inteligência coletiva e da análise dos conteúdos, a verificação da forma que os dados são compartilhados e como as pessoas cooperam para uma produção colaborativa, e através da ação coletiva, como gera as responsabilidades compartilhadas, além de verificar os sentimentos e o pertencimento que são definidos devido a determinados grupos que são formados com as opiniões debatidas.

1.2 Justificativa

Com o advento da internet cada vez mais surgem redes sociais que possibilitam as pessoas a se comunicarem umas com as outras de forma rápida, precisa e instantânea, diante disso mais dados são gerados e utilizados para análises. Respeitando-se a ética e a privacidade dos indivíduos que utilizam das mídias sociais para compartilhar as suas opiniões e se fazer uma análise comportamental dessas opiniões nas redes sociais.

Muitas vezes as opiniões compartilhadas refletem os sentimentos que as pessoas tem para se sentirem pertencentes a determinados grupos que compartilham da mesma opinião. E isso gera um grande debate devido as divergências entre determinados grupos e suas opiniões contrárias sobre assuntos que estão sendo debatidos nas mídias sociais, tornando assim a internet um lugar para discussão de ideias e ajudam as pessoas que possuem um sentimento neutro a criarem um pensamento lógico para tomada de decisões quanto há assuntos que estão sendo abordados.

E isso torna as mídias sociais um ambiente propício para análises de como a inteligência coletiva influencia em tomadas de decisões. Segundo (MAIA et al., 2015) a literatura focada em conversação mediada por computador e redes sociais, cujo interesse são a construção de identidades e comportamento em ambientes virtuais, destaca que aspectos como a identificação dos participantes, a liberdade discursiva e a moderação são elementos que influenciam o engajamento discursivo e a interação social mediada.

1.3 Objetivos

1.3.1 Objetivo geral

A partir da utilização da mídia social escolhida houve uma verificação desses dados compartilhados. A forma que esses dados impactam os usuários nas suas tomadas de decisões a partir do compartilhamento de opiniões e como ele pode cooperar para agregar sobre o que ele consome na internet e através dos dados coletados fazer uma análise dos sentimentos que as pessoas tem ao compartilhar suas opiniões na internet. Segundo (GOMES, 2013) descobrir o que as pessoas pensam sobre determinado assunto, tem sido um importante motivo para armazenamento de informações.

Com a grande quantidade de informações que tem sido compartilhadas na internet, foi necessário a automatização de tarefas para se fazer uma rápida análise dos dados e obter assim *insights* das opiniões que estavam sendo compartilhadas. Para a análise de mídias sociais os dados que foram coletados eram abertos e a mídia social que foi utilizada para fazer as coletas foi o *Twitter*. O foco principal abordado foi as *hashtags* (marcadores que agrupam *tweets*) que fazem menção a palavra educação.

1.3.2 Objetivos específicos

Para a realização das análises os dados foram coletados do *Twitter* e os gráficos foram gerados na execução do *script* em *Python* que foi implementado no *VSCode*. O código de coletas que foi utilizado está disponível na internet e é aplicável à situação, a partir de então foi analisado métricas para cada um dos pontos críticos levantados utilizando ferramentas específicas que foram desenvolvidas no código, os teste foram realizados da seguinte forma:

- Levantamento dos dados da mídia social utilizada.
- Realização de uma comparação e avaliação desses dados através de métricas do software utilizado.
- Após a comparação e avaliação, foi feito uma análise dos dados coletados pelo software.

1.4 Estrutura do trabalho

Este trabalho está organizado de acordo os seguintes capítulos:

O capítulo 2 a revisão dos trabalhos correlatados à técnica utilizadas.

No capítulo 3 apresenta a metodologia utilizada e o conjunto de dados.

O capítulo 4 apresenta o desenvolvimento do trabalho e testes que foram utilizados.

O capítulo 5 apresenta os resultados obtidos no trabalho.

Por fim, o capítulo 6 apresenta as conclusões obtidas.

2 Revisão bibliográfica

2.1 Referenciais Teóricos

Algumas das características importantes acerca da inteligência coletiva tais como, colaboração, busca pelo enriquecimento mútuo e formação de redes de cooperação, estão presentes na construção de ambientes colaborativos (FRANCO et al., 2018). Ou seja, a inteligência coletiva tem como característica a atividade compartilhada entre diversas pessoas. A forma como elas participam dessas atividades geram os níveis de dificuldade, aqueles que possuem um maior comprometimento coletivo nas ações executadas possuem um maior grau de dificuldade para com seu grupo pertencente. Indo "do menor nível de dificuldade para o maior tem-se: o compartilhamento, a cooperação e a ação coletiva "(SHIRKY, 2011).

Nas mídias sociais, muitas vezes o que é produzido está mais voltado para o compartilhamento e reuso de informação do que necessariamente um conhecimento coletivo que é o foco da inteligência coletiva (FRANCO et al., 2018). No novo molde como os sujeitos se comunicam, há também uma alteração em como ele "desenvolveu novas formas de ser, agir, construir, usar e disseminar informação, sendo esses processos viabilizados por uma estrutura digital "(FRANCO et al., 2018). E conforme o acesso aos dados de redes e sua análise crescem, uma visão mais conectada do mundo está emergindo (RECUERO; BASTOS; ZAGO, 2015).

Houve um considerável aumento no desenvolvimento dos métodos para análises de redes, devido a quantidade de dados a serem analisados terem crescido exponencialmente nos últimos anos com o aumento do comércio eletrônico e de dispositivos móveis, as ferramentas para análise tiveram que aumentar o seu poder computacional e de complexidade.

As diversas formas de redes tanto móveis, sociais e dispositivos sensoriais, estão remodelando as formas da sociedade com a tecnologia para adequar-se as novas estruturas existente e com respostas rápidas que as redes proporcionam para gerir as opiniões emitidas pelos usuários e como a interação entre eles pode produzir dados para uma análise comportamental e de sentimentos.

As novas formas de conexão possuem um alto potencial para o estudo de como ocorre os processos sociais de interesse crítico, no qual "transformar dados em conhecimento e conhecimento em inteligência ajudará as empresas a conquistar novos clientes e engajá-los."(BOUFIM; BARKA, 2015). Saindo do tradicional método usado nas empresas, para "os mapas de rede, que são úteis para capturar, simplificar e exemplificar sistemas complexos grandes"(RECUERO; BASTOS; ZAGO, 2015).

Com a evolução da web os usuários não são mais meros receptores com restrição de serviços, mas sim receptores e produtores de informações, no qual podem produzir seu próprio conteúdo através de plataformas gratuitas possibilitando assim a interação entre seus usuários e com uma otimização de recursos (COUTINHO; JUNIOR, 2007). E fazem uso dessas plataformas digitais para abordarem assuntos que muitas vezes estão em alta devido a sua repercussão na rede.

Muitos são os influenciadores que fazem o uso da sua plataforma para tratar de assuntos que são de interesse e expor a sua opinião acerca de tal assunto, podendo haver uma interação simultânea com os usuários que estão online no momento. No entanto devido a grande polarização política que estamos vivendo, muitos influenciadores preferem não opinar sobre determinados assuntos que estão em alta para não haver uma "perda" de seguidores que não compartilham dos mesmos ideais. E com a proliferação dos equipamentos digitais e a sua utilização como meios de comunicação rápida e contínua houve um aumento na procura de sistemas e algoritmos capazes de extrair conhecimentos através de dados (GOMES, 2013).

2.2 Trabalhos Relacionados

A forma de análise e coleta dos dados foi feita a partir do aspecto bibliográfico. Dessa forma, as fontes que poderão ser utilizadas para coleta de dados são: primárias (livros, teses, dissertações, relatórios entre outros), secundárias (artigos, manuais, dicionários e enciclopédias), e terciárias (bibliografias, catálogos, índices e outros) (FRANCO et al., 2018). A construção do modelo com as características de inteligência coletiva apresentadas nas manifestações do compartilhamento, da cooperação e da ação coletiva se dará a partir do arcabouço teórico (FRANCO et al., 2018).

Inicialmente o trabalho tinha como proposta fazer a coleta de dados pelo *Twitter* e gerar os grafos a partir da plataforma *Gephi*, mas depois de algumas análises foi mais vantajoso prosseguir por meio de uma coleta e análise de dados a partir de Mineração de Textos, usando a linguagem de programação *Python*, e com esses dados coletados foi observado as manifestações da inteligência coletiva em mídias sociais e os sentimentos das pessoas ao compartilhar as suas opiniões nas redes sociais, com o propósito de obter o conhecimento estruturado podendo ser utilizado por um sistema de apoio ou para tomadas de decisão (BENEVENUTO; RIBEIRO; ARAÚJO, 2015).

Alguns trabalhos relacionados fizeram a análises de sentimentos utilizando o *Facebook* como rede social para se fazer as coletas, nele (OLIVEIRA, 2021) coleta os dados do *Facebook* para traçar o sentimento das pessoas a respeito das eleições de 2018. Já (FILHO, 2014) faz uma análise de sentimentos a partir do *Twitter* em que os dados foram minerados para mapear as opiniões durante a Copa de 2014, e através desses *tweets* extrair

conhecimentos relacionados ao evento.

3 Metodologia

Para o desenvolvimento do trabalho, foi utilizado um notebook Lenovo S145 com 8GB de ram, HD de 1TB e processador intel inside core i7-8565U.

O objetivo deste trabalho relaciona-se com a análise de sentimentos e de qual forma essas análises podem cooperar com o que as pessoas manifestam nas redes sociais e a forma que as suas opiniões expressas continuamente podem impactar a tomada de decisão através dos *tweets* que foram propagados, porém determinar esse sentimento expresso pode ser trabalhoso e estar sujeito a erros.

As coletas foram feitas pelo *Twitter* no qual é descrito como um microblog ou site de rede social e tem como atividade central postar mensagens curtas de atualização de status (*tweets*) via web ou telefone celular (THELWALL; BUCKLEY; PALTOGLOU, 2011). O Brasil possui a quarta maior base de usuários no *Twitter*, com 19.05 milhões de usuários cadastrados em janeiro de 2022 ¹. Sua estrutura dinâmica permite aos usuários terem acesso a informações no momento em que elas são postadas, é uma rede que não restringe a conexão entre os usuários sendo possível dar *retweet* em *tweets* que foram publicados por outros usuários.

O *Twitter* foi escolhido por ter um alto nível de popularidade, conter um grande volume e velocidade de dados sobre assuntos relacionados aos interesses em comum ou até mesmo àqueles que estão em "alta" no momento. Muitas vezes essas manifestações são expressas com palavras que podem caracterizar algum sentimento, sendo essas possíveis de serem analisadas através da *NLP* (*Natural Language Processing*). Segundo (FILHO, 2014) a *NLP* é um conjunto de técnicas teórico-computacionais que visam representar dados textuais e processar a linguagem natural humana para diversas tarefas.

Para as análises que foram feitas a menção em questão que foi abordada para o desenvolvimento desse trabalho foi a #educação. A #educação foi escolhida por estar em constante debate na sociedade, é um tema que envolve todas as fases da vida de uma pessoa, desde a pré-escola ao universitário, seus debates são essenciais na discussão de melhorias para uma educação pública e privada de qualidade na inserção das pessoas na sociedade.

As coletas foram feitas com dados abertos públicos, respeitando-se a privacidade dos usuários que compartilharam suas opiniões sobre a menção escolhida para fazer as análises. As etapas deste trabalho foram assim definidas:

¹ <https://valorinveste.globo.com/mercados/internacional-e-commodities/noticia/2022/04/25/brasil-tem-a-quarta-maior-base-de-usuarios-do-twitter-no-mundo.ghtml>

1. Revisão de literatura, bem como trabalhos correlatados na análise de sentimentos.
2. Levantamento e coleta dos dados que foram utilizados.
3. Avaliação dos dados.
4. Análise e discussão dos resultados obtidos e identificação de possíveis melhorias.

3.1 Levantamento dos dados que serão utilizados

Para o levantamento dos dados utilizados foram usados os seguintes métodos para a classificação e o agrupamento dos dados. O objetivo era conhecer as estruturas das conexões e como essas conexões atuam em grupos, o conhecimento dos indivíduos do coletivo não se fez necessário.

- Nós (*nodes*): são representados pelo indivíduos, tais como perfis e *nicknames*, textos, posts publicados.
- Arestas (*edges*): coleção de conexões de objetos com um objetivo individual.
- Grupo: pode ser compreendido como um *cluster* de nós.
- Polaridade: grau de positividade, negatividade ou neutralidade de um texto.
- Sentimento: reconhecimento de sentimentos específico presente em um texto.
- Subjetividade: expressam sentimentos pessoais ou opiniões.

3.2 Coleta dos dados

A coleta dos dados se deu através da Mineração de Textos, conhecida como Descoberta de Conhecimentos em Textos, fornece um conjunto de técnicas na automatização do processo e estruturação das informações, a Mineração de Textos pode ser vista como uma extensão da Mineração de Dados que é baseada em dados estruturados, já a Mineração de Textos aplica-se sobre dados não estruturados (FILHO, 2014).

Neste trabalho foi utilizado um *script* na linguagem de programação *Python* e a biblioteca *Tweepy*, que é uma biblioteca em *Python* de fácil acesso à *API* (*Application Programming Interface*) do *Twitter* e código foi implementado usando a IDE *VSCode*. O *Twitter* disponibiliza chaves de acesso para que se possa fazer as análises, chaves estas que são privadas, individuais e podem ser atualizadas.

Durante o desenvolvimento houve alguns contra tempos com as chaves de acesso para fazer a coleta, pois o *Twitter* não permite fazer a coleta de dados antigos, essa chave

só é permitida para pesquisadores, sendo necessário alguns dados para que se possa obter o acesso, foram seguidos todos os passos pedidos para tentar obter a chave de pesquisador, mas não obtivemos sucesso, essa chave possui um número máximo de *tweets* que podem ser coletados, sendo necessária a chave com permissão de acesso para fazer a coleta de dados antigos. Foram utilizadas duas chaves de acesso nas coletas, uma com uma conta do *Twitter* mais antiga e uma conta mais recente, para esse trabalho foi realizada a coleta com a chave de acesso da conta mais recente.

Para fazer as coletas foi utilizado os códigos de (YENER, 2020) e (CARVALHO, 2021), que está disponível na internet e se mostrou efetivo para o objetivo desse trabalho. O algoritmo classificava os *tweets* em negativos, positivos e neutros fazendo essas classificações utilizando as bibliotecas *Tweety* para obter os textos e o *Textblob* para classificar, algumas alterações foram feitas para fazer a coleta dos *tweets* em português.

A coleta ocorreu no dia 8 de Junho de 2022 algumas semanas após a votação que estava ocorrendo sobre a Proposta de Emenda Constitucional (PEC) 206/2019. E quanto mais distante da data que ocorreu o evento "menos subjetivo e menos relacionado aos fatos eram os texto das mensagens"(MACHADO; TIJIBOY, 2005).

Segundo (HASSAN; ABBASI; ZENG, 2013) a demanda por dados do *Twitter* é motivada por cada vez mais interessados em entender o que as pessoas estão postando. Nos *tweets* coletados, segundo (MACHADO; TIJIBOY, 2005) busca-se através das redes sociais espaços de construção e produção de discursos, manifestações das múltiplas "verdades" sociais e suas representações. Através dos dados coletados foi possível indentificar variáveis para fazer uma análise de como as pessoas estavam reagindo e quais as palavras estavam sendo mais utilizadas, visto que é um assunto que impacta diretamente milhares de alunos que estudam em universidades públicas.

3.3 Avaliação dos dados

Após o levantamento dos dados e as coletas, foi feito a avaliação dos dados. Essa avaliação consistiu em utilizar métodos e filtros para remoção de informações desnecessárias para a aplicação. Algumas palavras são utilizadas para classificar quanto aos sentimentos que são expressos ao fazer o compartilhamento de opiniões em mídias mídias sociais.

A seguinte frase "eu gosto de estudar" pode ser associada a um sentimento positivo, já a frase "eu odeio o estudar" pode ser associada a um sentimento negativo, no entanto a frase "eu gosto de ir para a faculdade, mas odeio estudar" pode ser caracterizada como um sentimento neutro, pois ela expressa dois sentimentos em uma mesma frase com o mesmo contexto.

Esses sentimentos serão classificados como positivos, negativos e neutros. No entanto

algumas frases podem expressar mais de um sentimento quando são frases compostas e complexas, sendo assim a avaliação dos dados tem por objetivo extrair aspectos dos sentimentos e produzir um resumo dos sentimentos. Possibilitando segundo (MACHADO; TIJIBOY, 2005) as redes sociais ser um lugar onde pode-se contribuir para a mobilização de saberes, reconhecimento de diferentes identidades e articulações dos pensamentos que compõem a coletividade.

4 Desenvolvimento

4.1 Visão geral

Nesse trabalho foi utilizado a Mineração de Textos para fazer a análise de sentimentos, que é o processo de associar um texto a um sentimento diferenciando-os em positivo, negativo e neutro, e entender como essas análises podem colaborar com o processo de tomada de decisões das pessoas em determinados assuntos.

Segundo (HEMMATIAN; SOHRABI, 2017) a análise de sentimentos requer métricas de similaridades de forma que os sentimentos tenham alto nível de similaridade e vive-versa. Durante o processo deve-se considerar o uso de ironia ou sarcasmo para reconhecer quando as pessoas usam palavras positivas para expressar um sentimento negativo e vice-versa, podendo gerar uma ambiguidade no sentimento.

Nas análises de textos pode-se "conter uma única opinião e pode ser considerada neutra"(LIU, 2012) descrevendo diferentes aspectos. Nas análises de polaridades negativas, positivas e neutras, ao treinar ferramentas de aprendizado de máquina com exemplo de emoções em texto o uso de palavras semelhantes com frequência expressam seus sentimentos, com isso máquinas aprendem automaticamente a detectar sentimentos sem intervenção humana, afim de tornar as análises mais qualitativas. No entanto segundo (BENEVENUTO; RIBEIRO; ARAÚJO, 2015) elas não são capazes de fazerem as mesmas interpretações que os humanos, necessitando assim de algoritmos precisos e não ambíguos para a realizar as tarefas.

Com os softwares descritos na seção 3, dado que todos foram instalados e configurados corretamente, o trabalho estava pronto para ser executado e realizar as análises que foram propostas, respeitando todas as normas de privacidade nas coletas, sendo apenas possível ver o texto que foi compartilhado por cada usuário, ficando assim o usuário oculto de qualquer comentário que possa expor a sua opinião quanto ao assunto abordado na realização desse trabalho.

4.2 Testes

Com os softwares devidamente configurados e as chaves de desenvolvedor aprovadas, foi iniciado as coletas para o desenvolvimento. As coletas foram feitas através do *VSCode* que possui um limite máximo de 2500 *tweets* de amostragem, limitando assim a fazer análises com um grande volume de dados. Para esse trabalho o limite máximo que foi usado para executar é de 2500 *tweets*, número máximo de *tweets* que o *VSCode* amostra.

Para fazer as coletas e gerar os resultados que eram esperados, o código foi executado algumas vezes, para que amostrasse os *tweets* do assunto no momento da execução. Muitas vezes o assunto do momento não era de fato Educação, mas sempre estava envolvida em debates que acontecia no momento, visto que o *Twitter* é uma mídia social em que as pessoas compartilham as suas opiniões expressando assim seus sentimentos em tempo real. E como mencionado acima o *Twitter* não permite fazer coleta de dados mais antigos, devido as chaves de acesso, dificultando a coleta de dados antigos do assunto no dia em que ele estava em debate ou por um período de tempo maior.

Na Figura 1 que se segue abaixo é possível ver o erro gerado no *VSCode* quando queremos ter como a entrada do parâmetro um número maior que 2500 tweets, neste caso tentou-se executar com 2501 *tweets* e ele ficou executando por um tempo indeterminado até que houvesse uma pausa na execução, quando isso ocorre é necessário que todas as células anteriores sejam executadas novamente.



Figura 1 – Erro ao retornar mais de 2500 tweets

Fonte: Autora

Durante o desenvolvimento do trabalho houve uma dificuldade em fazer as coletas através do *Pycharm*, IDE que tinha sido escolhida para a execução do código, o *Pycharm* coletava um grande volume de dados e ao executar esses dados no *Gephi*, plataforma escolhida para geração dos grafos, o mesmo não gerava os grafos esperados. As coletas eram feitas no *Pycharm* com a extensão *.txt* e era necessário convertê-la em *Json* e depois em *.csv*, extensão que era aceita pelo *Gephi*.

Devido a essas dificuldades foi necessário a execução do código no *VSCode* e com um número de dados menor, além de ter sido possível executar o código passo a passo para um melhor entendimento do processo, os gráficos também foram gerados no *VSCode*.

5 Apresentação e análise dos resultados

5.1 Coleta e análise dos *tweets*

As coletas dos dados do *Twitter* foi a primeira parte da execução deste trabalho, elas foram realizadas no dia 8 do mês de Junho de 2022 e foi definido diferentes períodos do dia para a realização das coletas, iniciando assim a coleta dos *tweets* que faziam menção ao termo a *#educação*. Para as coletas o *script* em Python recebe como parâmetro *hashtags* e retorna os *tweets* marcados com essas *hashtags*.

Na Figura 2 abaixo pode-se verificar os 5 primeiros e os 5 últimos *tweets* que foram compartilhados com menção a *#educação*.

Index	Tweet Text
0	Acabando com os bloqueios no aprendizado de id...
1	Curso de Espanhol. Veja em detalhes no site ht...
2	Curso de Japones. Veja em detalhes no site htt...
3	Guia de Abertura e Legalizacao de Empresas. Ve...
4	Luzes do painel de seu carro. Veja em detalhes...
...	...
935	Curso de Paisagismo para Leigos e Experts. Vej...
936	Bebedouro e Aviso sedia a 2ª rodada da 1ª divi...
937	Curso de Tubulacoes Industriais. Veja em detal...
938	Biblia e Terrorismo. Veja em detalhes no site ...
939	Papas que eram realmente pessoas terriveis. Ve...

940 rows × 1 columns

Figura 2 – *Tweets compartilhados*

Fonte: *Autora*

A coleta ocorreu após alguns dias da votação da PEC 206/2019 que estava em discussão e estava gerando grandes discussões nas mídias sociais, por partes de pessoas que eram a favor e outras que eram contra, podendo assim ainda ter menções sobre o resultado da votação ou não, uma vez que o termo "Educação" está em constante discussão no *Twitter*.

Outro ponto a ressaltarmos é que estamos em ano eleitoral para presidente da república, evento este que repercute na população aumentando a força do sentimento sobre temas tão relevantes como educação, sendo assim os sentimentos podem ser tanto positivo, negativo ou neutro, visto que é um tema que está sofrendo constantes ataques em diversas formas pelo governo e até mesmo pela sociedade.

5.2 Período das coletas

Para fazer a inicialização das coletas foi definido que elas seriam realizadas em três intervalos de tempo diferentes durante o dia da coleta, assim sendo realizadas no período da manhã às 10:30, no período da tarde às 14:30 e no período da noite às 19:30. Foi escolhido esses horários com diferenças iguais de 4 horas pois ao executar os código para testes verificou-se que as polaridades mudavam de acordo com o período do dia podendo assim entender melhor como as pessoas reagem sobre o assunto.

Na Tabela 1 contém o número total de *tweets* coletados e o número de *tweets* positivos, negativos e neutros.

Dia/Hora	Número Total	Números Positivos	Números Negativos	Números Neutros
08/06 às 10:30	940	21	643	276
08/06 às 14:30	1086	25	756	306
08/06 às 19:30	1307	26	928	353

Tabela 1 – Número de *tweets* e polaridades
Fonte: Autora

5.2.1 Coletas realizadas no período da manhã

Na inicialização da execução entrou-se com a #educação que era o parâmetro de entrada definido e 2500 *tweets*, como mencionado acima na Tabela 1, no momento da execução do código havia 940 menções com a educação contendo um número menor que o previsto.

Na Figura 3 abaixo observou-se que no período da manhã, foi verificado que o número de *tweets* com sentimentos negativos teve uma porcentagem maior em relação aos outros, sendo de 25,7 por cento, seguido pelos sentimentos neutros representando 11,0 por cento e por último os sentimentos positivos com 0,8 por cento.

Análise de Sentimento para #educação

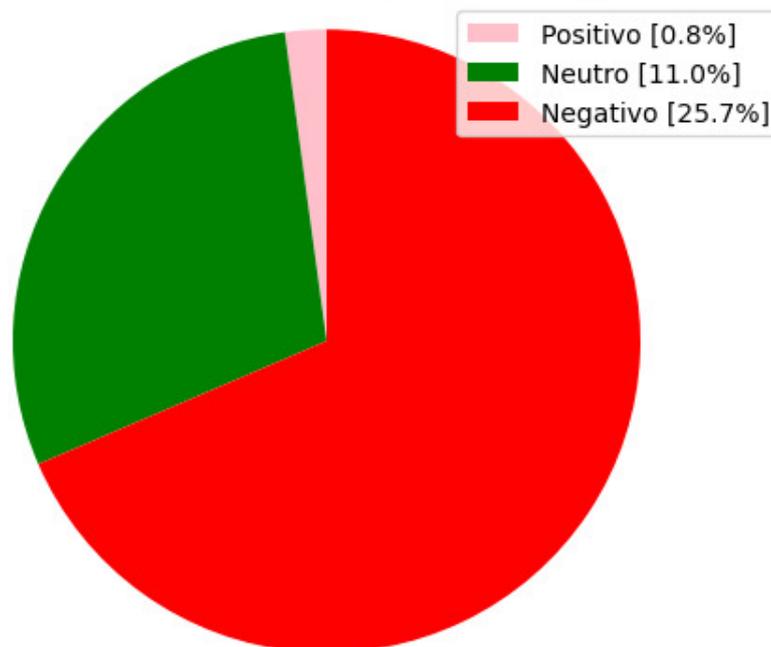


Figura 3 – *Análise de Sentimentos #educação no período da manhã*
 Fonte: Autora

Na Figura 4 que se segue abaixo observa-se as palavras que foram mais usadas nas coletas no período da manhã em ordem decrescente e quantas vezes eles foram citadas.

	0
site	603
veja	600
detalh	592
via	576
cultura	418
curso	234
conhecimento	187
educação	116
mpsnet	110
aula	39

Figura 4 – *Palavras mais citadas no período da manhã*
 Fonte: Autora

5.2.2 Coletas realizadas no período da tarde

Para as coletas realizadas no período da tarde foram utilizadas os mesmos parâmetros de entradas que no período da manhã, e como mencionado na Tabela 1, o número total de *tweets* com a menção pesquisada foi de 1086, sendo menor que o número que era esperado de 2500 *tweets*, mas nota-se um aumento no número de *tweets* que faziam menção a #educação.

Na Figura 5 abaixo, observou-se que os número de *tweets* com sentimento negativo era de 30.2 por cento, seguido de sentimentos neutros correspondendo a 12.2 por cento e por último os sentimentos positivos com 1.0 por cento. Não havendo uma diferença significativa dos sentimentos que foram coletados no período da manhã, mesmo havendo um número maior de *tweets* coletados.

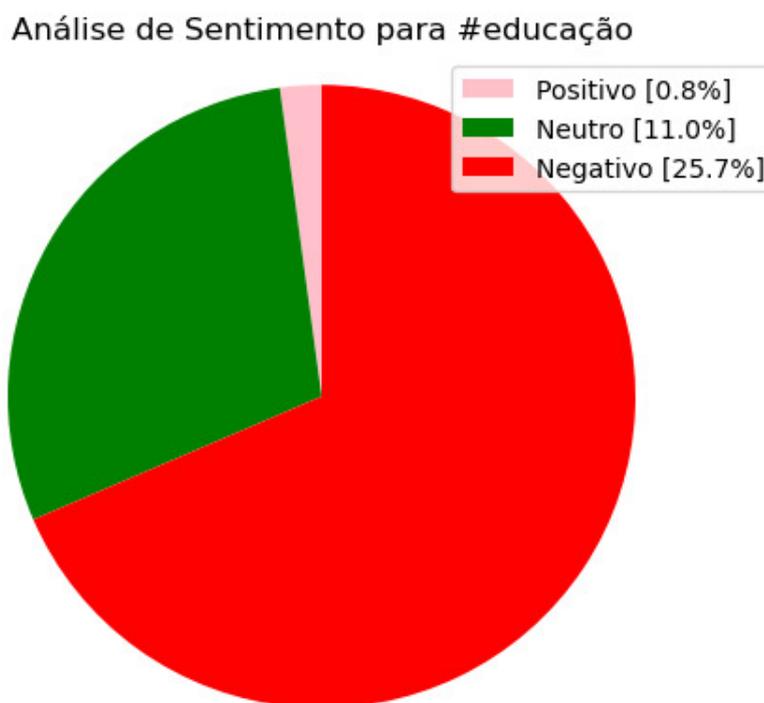
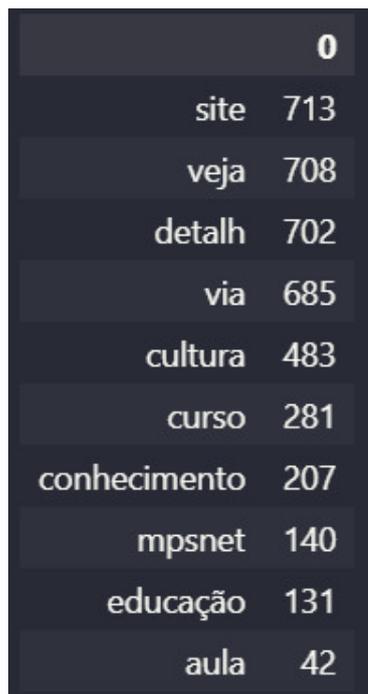


Figura 5 – *Análise de Sentimentos #educação no período da tarde*
Fonte: *Autora*

Na Figura 6 abaixo observa-se as palavras que foram mais usadas no período da tarde em ordem decrescente e quantas vezes eles foram citadas. Em relação ao período da manhã, verifica-se que houve um aumento na quantidade de vezes que a palavra foi citada, mas isso se deve também ao fato de que no período da tarde o número de *tweets* coletados ser maior que no período da manhã e as palavras mais citadas continuam na mesma posição, somente a palavra educação e mpsnet que modificaram a sua posição.



	0
site	713
veja	708
detalh	702
via	685
cultura	483
curso	281
conhecimento	207
mpsnet	140
educação	131
aula	42

Figura 6 – Palavras mais citadas no período da tarde

Fonte: Autora

5.2.3 Coletas realizadas no período da noite

No período da noite foram utilizadas os mesmos parâmetros de entradas que no período da manhã e da tarde, e como mencionado na Tabela 1, o número total de *tweets* com a menção pesquisada foi de 1307, também como nos períodos anteriores houve um número menor de *tweets* coletados do que era esperado, mas maior que nos períodos anteriores. Esse aumento no número de *tweets* pode ter ocorrido devido as pessoas estarem mais ativas nas mídias sociais no período da noite.

Na Figura 7 abaixo, observou-se que os número de *tweets* com sentimento negativo era de 37,1 por cento, seguido de sentimentos neutros correspondendo a 14,1 por cento e por último os sentimentos positivos com 1,0 por cento. No período da noite já observa-se um aumento nas porcentagens de sentimentos negativos e neutros em relação aos períodos anteriores.

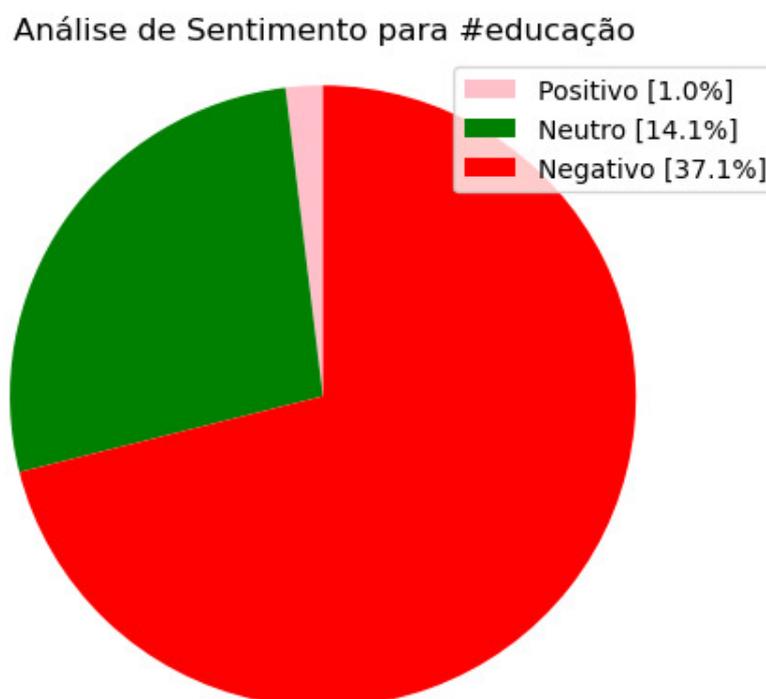
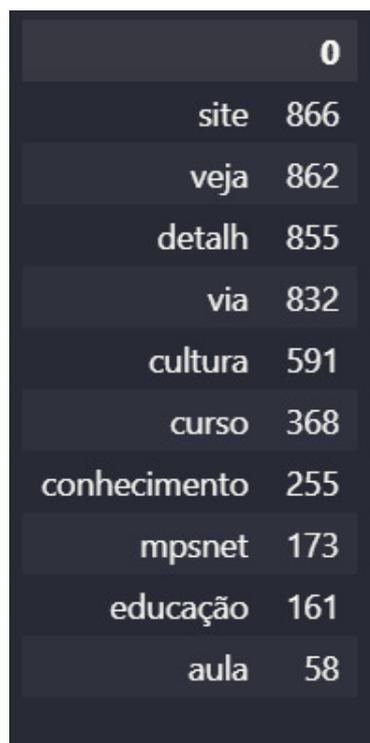


Figura 7 – *Análise de Sentimentos #educação no período da noite*
Fonte: *Autora*

Na Figura 8 abaixo observa-se as palavras que foram mais usadas no período da noite em ordem decrescente e quantas vezes elas foram citadas. Em relação ao período da manhã e da tarde, houve um aumento na quantidade de vezes que as palavras foram citadas, mas isso se deve ao fato do período da noite o número de *tweets* coletados ser maior que no período da manhã e da tarde, e não houve uma alteração significativa em comparação ao período da tarde quanto as palavras mais citadas.



	0
site	866
veja	862
detalh	855
via	832
cultura	591
curso	368
conhecimento	255
mpsnet	173
educação	161
aula	58

Figura 8 – Palavras mais citadas no período da noite

Fonte: Autora

Através dos gráficos nota-se que nesse dia o sentimento negativo em relação a #educação não modificou e no decorrer do dia com o aumento no número de *tweets* coletados o sentimento negativo foi mantendo-se constante podendo-se compreender com isso que as postagens negativas poderiam estar sendo utilizadas com o objetivo de criticar. E isso foi ocorrendo durante o dia, não havendo assim uma alteração significativa do sentimento, ou seja, nota-se que as opiniões não se alteraram e as pessoas continuaram a seguir padrões de sentimentos negativos supostamente com tons de crítica nas suas postagens.

5.2.4 Comparação dos resultados

O compartilhamento de opiniões semelhantes nas redes sociais deve-se muito ao fato das pessoas se identificarem com as questões que estão sendo compartilhadas no momento, seguindo o coletivo. As redes sociais podem segundo (MACHADO; TIJIBOY, 2005) contribuir para a mobilização dos saberes, reconhecimento das diferentes identidades e a articulação de pensamentos que compõem a coletividade. E isso foi notado com os *tweets* que foram sendo coletados durante o dia, os sentimentos não se modificaram muito e a frequência de determinadas palavras utilizadas continuaram constantes sendo possível perceber que as pessoas até mesmo ao compartilhar suas opiniões seguiam um padrão de escrita.

Foi possível observar uma constância nas palavras usadas, o *tri-grams* segundo (NASCIMENTO; OSIEK; XEXÉO, 2015) calcula a probabilidade de cada texto ser positivo ou negativo analisando a sequência de três palavras. Como mostrado abaixo na Figura 9 dos *tri-grams* nos três períodos analisados observa-se uma constância nas combinações das palavras que ocorreram juntas e com uma maior frequência ao longo do dia, podendo haver uma alteração no decorrer do dia nas sequências que possuem menores combinações.

Em dias que não ocorrem eventos no qual a menção está em alta, a quantidade de *tweets* coletados foi inferior a dias em que a menção era alvo de discussão e isso se deve também ao fato do *Twitter* ser uma mídia social dinâmica e ocorre diversos eventos diferentes no decorrer do dia, podendo uma menção estar em "alta" em determinado período do dia e em outro já não estar sendo mais discutida, nas imagens abaixo observa-se os *tri-grams* que foram mais utilizados durante os períodos da manhã, tarde e noite.

```
[('veja em detalhes', 589),  
 ('em detalhes site', 589),  
 ('detalhes site cultura', 412),  
 ('site cultura conhecimento', 182),  
 ('detalhes site mpsnet', 83),  
 ('cultura conhecimento cursos', 57),  
 ('detalhes site mpsnet2', 26),  
 ('aulas veja em', 14),  
 ('video aulas veja', 13),  
 ('acesse siga nossas', 13),  
 ('acesse saiba mais', 12),  
 ('saiba mais notícias', 12),  
 ('mais notícias política', 12),  
 ('notícias política salvador', 12),  
 ('política salvador bahia', 12),  
 ('salvador bahia prefeituradesalvador', 12),  
 ('completo veja em', 11),  
 ('3000 esquemas circuitos', 11),  
 ('esquemas circuitos eletronicos', 11),  
 ('circuitos eletronicos vol', 11)]
```

Figura 9 – *Trigrams - Manhã*

Fonte: *Autora*

```
[('veja em detalhes', 697),
 ('em detalhes site', 697),
 ('detalhes site cultura', 477),
 ('site cultura conhecimento', 202),
 ('detalhes site mpsnet', 104),
 ('cultura conhecimento cursos', 63),
 ('detalhes site mpsnet2', 35),
 ('aulas veja em', 15),
 ('video aulas veja', 14),
 ('acesse siga nossas', 13),
 ('3000 esquemas circuitos', 12),
 ('esquemas circuitos eletronicos', 12),
 ('circuitos eletronicos vol', 12),
 ('acesse saiba mais', 12),
 ('saiba mais notícias', 12),
 ('mais notícias política', 12),
 ('notícias política salvador', 12),
 ('política salvador bahia', 12),
 ('salvador bahia prefeituradesalvador', 12),
 ('completo veja em', 11)]
```

Figura 10 – *Trigrams - Tarde*Fonte: *Autora*

```
[('veja em detalhes', 835),
 ('em detalhes site', 835),
 ('detalhes site cultura', 584),
 ('site cultura conhecimento', 250),
 ('detalhes site mpsnet', 122),
 ('cultura conhecimento cursos', 84),
 ('detalhes site mpsnet2', 50),
 ('aulas veja em', 25),
 ('video aulas veja', 23),
 ('em video aulas', 18),
 ('veja detalhes site', 17),
 ('completo veja em', 15),
 ('videos veja em', 13),
 ('curso completo veja', 13),
 ('3000 esquemas circuitos', 13),
 ('esquemas circuitos eletronicos', 13),
 ('circuitos eletronicos vol', 13),
 ('acesse siga nossas', 13),
 ('conhecimento cursos educação', 12),
 ('acesse saiba mais', 12)]
```

Figura 11 – *Trigrams - Noite*Fonte: *Autora*

O *Twitter* é um importante instrumento para pesquisas comportamentais nas redes

sociais, devido a sua característica em ser uma mídia social dinâmica e sem custo. E segundo (MACHADO; TIJIBOY, 2005) por transcender a distância, apresentar natureza assíncrona e a rápida propagação de informação. Esse dinamismo pôde ser visto nos sentimentos que as pessoas expressaram nos *tweets* que foram coletados durante o desenvolvimento desse trabalho.

5.2.5 Clusterização

A clusterização dos dados consiste em um método não supervisionado que visa agrupar os dados em grupos semelhantes. Segundo (CASSIANO, 2014) a ideia básica é elementos que compõem um mesmo cluster devem apresentar uma alta similaridade, e tem por objetivo maximizar a homogeneidade dentro de cada cluster e maximizar a heterogeneidade entre cluster.

As opiniões contidas no conjunto de dados de *tweets* foram gerados através dos gráficos da clusterização e pode ser observado com os *tweets* que estão agrupados em cada cluster, havendo um entendimento maior do conjunto de dados original, e assim descobrir correlações entre o conjunto de opiniões que foram coletadas, pois ao se analisar os sentimentos eles não podem pertencer a mais de um cluster. Ainda segundo (CASSIANO, 2014) na literatura, a análise de clusters pode ser chamada também de Clusterização, *Clustering*, *Q-analysis*, *Typology*, *Classification Analysis* ou *Numerical Taxonomy*, sendo a clusterização uma técnica mais primitiva.

Nas Figuras 10, 11 e 12 abaixo, os gráficos de clusterização que foram gerados para fins de comparação e validação na identificação de relacionamentos não possuem uma centralização de agrupamentos para compreensão, devido a quantidade de *tweets* coletados e também a baixa polaridade, não possuindo assim uma boa visualização dos clusters vizinhos dificultando a percepção.

Para uma melhor visualização dos clusters eles foram definidos da seguinte maneira para uma análise de sentimentos:

- Cluster 1 - foram representados como sentimento negativo
- Cluster 2 - foram representados como sentimento positivo
- Cluster 3 - foram representados como sentimento neutro

Análise de Sentimentos

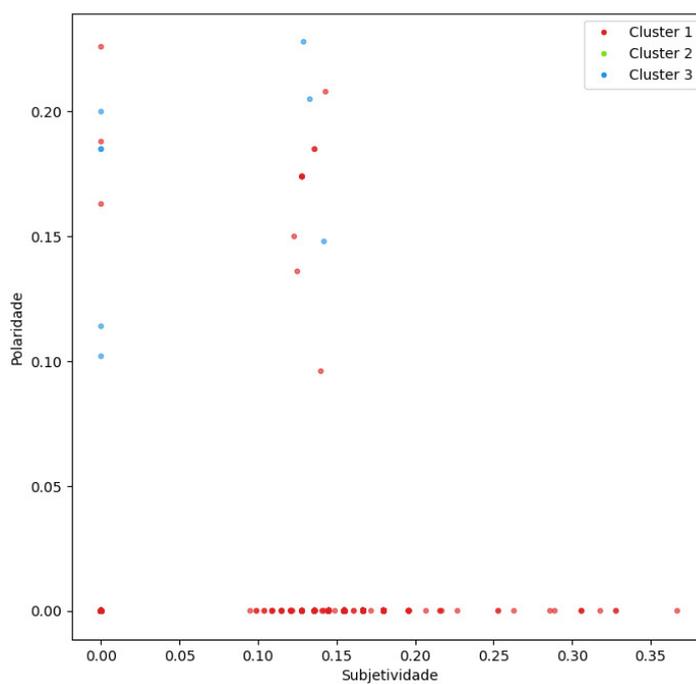


Figura 12 – *Cluster - Manhã*
 Fonte: *Autora*

Análise de Sentimentos

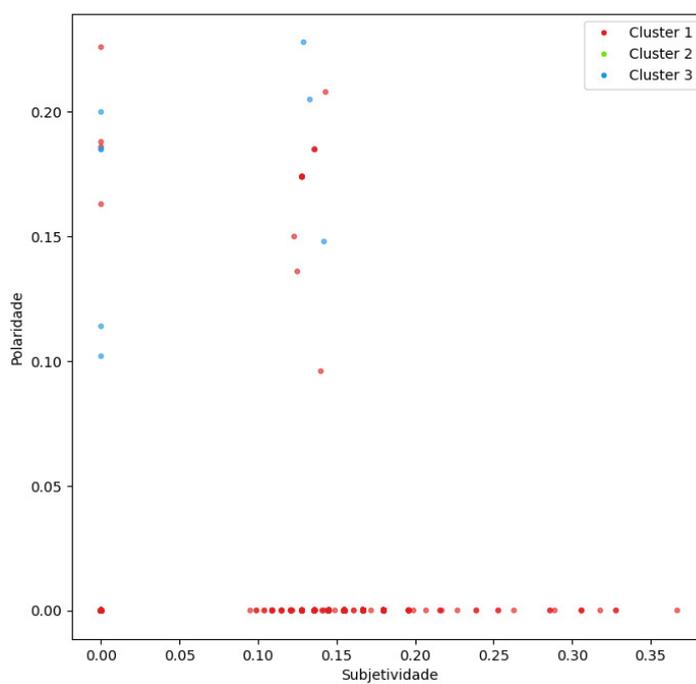
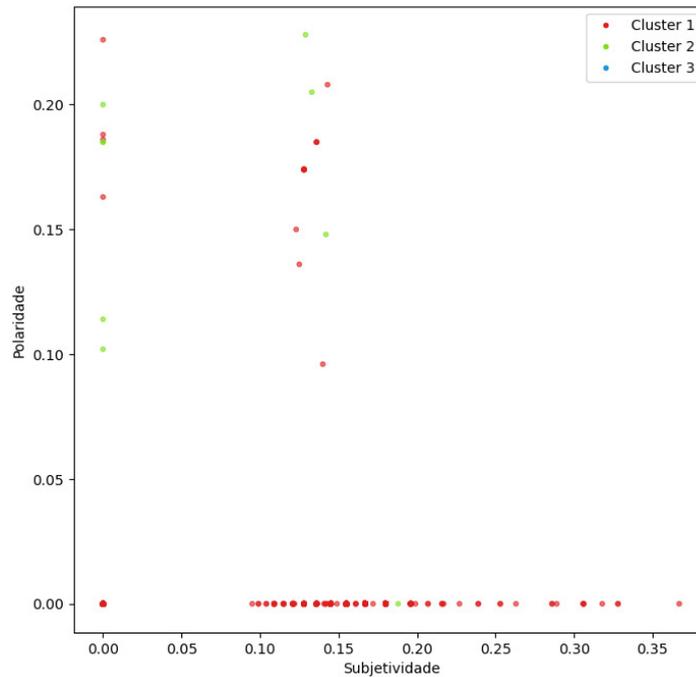


Figura 13 – *Cluster - Tarde*
 Fonte: *Autora*

Análise de Sentimentos

Figura 14 – *Cluster - Noite*Fonte: *Autora*

Nos gráficos de clusterização percebe-se uma linearidade e dispersão dos dados, dificultando com isso obter uma compressão dos dados, sendo perceptível apenas os cluster com sentimentos negativos que através das coletas houve um maior número de *tweets* coletados, assim nota-se que os sentimentos negativos possuem um conjunto de informações que os demais sentimentos não possuem.

Observa-se também que os gráficos no período da manhã e tarde não diferem muito entre si, notando-se que os sentimentos não modificaram de forma significativa de um período para o outro, sendo os *tweets* com sentimentos negativos e neutros com uma melhor visualização. Já o gráfico de cluster do período da noite nota-se presença de *tweets* positivos.

6 Conclusão

Este trabalho teve como objetivo usar a Mineração de Textos para analisar a forma com que as pessoas compartilham as suas opiniões nas mídias sociais e como elas podem colaborar para uma formação de opinião e na tomada de decisões. Para a realização do trabalho foi realizada a coleta de 2500 *tweets* que era a quantidade máxima que o *VSCode* suportava para fazer análises, a menção que usamos foi a *#educação* e a coleta foi realizada no dia 8 de Junho de 2022.

As análises de mídias sociais têm sido de extrema importância devido ao seu dinamismo, que pode ser observado nas postagens para analisar o comportamento e as opiniões das pessoas nas redes sociais. Estas mídias sociais contêm diversos tipos de informações e tornam-se ambientes propícios para verificar padrões nos textos compartilhados. E com isso podemos fazer análises de gostos, assuntos mais procurados e debates em torno de determinados assuntos que são de interesses da sociedade.

Com isso as pessoas tendem a tomar opiniões ou até mesmo mudar pontos de vista através de discussões que são geradas com as postagens. Esse trabalho mostrou que as mídias sociais diariamente se torna palco de discussões de ideias que geram grandes debates e podem fazer com a que a sociedade procure se manter mais atualizada sobre assuntos que impactam diretamente a vida. As mídias sociais se tornaram um importante meio de propagação de informações e compartilhamento de ideias no cotidiano da sociedade.

Durante o dia que ocorreu as coletas, o sentimento negativo predominou nos *tweets* coletadas em todos os períodos que foram analisados, não havendo uma alteração significativa das menções durante o decorrer do dia e isso pode ser observado pelas imagens dos períodos, em que as porcentagens de sentimentos negativos aumentava com o decorrer do dia. Em um mesmo *tweet* pode ocorrer de haver mais de uma *hashtag* podendo então a *#educação* estar relacionada a outro assunto e durante a coleta ser classificada como sentimento negativo ou com os demais sentimentos positivos e neutros.

Apesar do *Twitter* ser uma boa mídia social para a coleta de dados, as pessoas tendem a seguir um padrão de sentimentos no decorrer do dia, podendo haver uma modificação caso algum evento esteja ocorrendo, sendo isso possível para qualquer tipo de sentimento. O *Twitter* por ser uma mídia social dinâmica pode haver diferentes movimentos de opiniões com o decorrer do tempo, permitindo assim analisar ao longo do dia como as pessoas se comportavam nas mídias sociais e o que compartilhavam. Notou-se que uma quantidade de dados maior da que foi utilizada poderia ser possível tirar conclusões mais precisas nas análises.

6.1 Trabalhos Futuros

Como trabalhos futuros diante das limitações que o *Twitter* coloca sobre as chaves de acesso, pode-se considerar obter a chave de acesso para pesquisador para uma análise mais ampla com quantidades mais expressivas e fazer uma análise para a aplicação de mais métricas na realização de análises de sentimentos e métricas de validação estatística. Pode-se experimentar o processo em outras mídias sociais no qual o processo de obtenção de chaves de acesso sejam menos limitados que o do *Twitter*.

Uma possível melhoria seria a utilização de várias *hashtags* com a menção educação para uma coleta mais ampla do termo. Um aumento no conjunto de menções pode ser mais eficaz que o conjunto de dados coletados com apenas a *#educação*. E a partir disso construir modelos com as informações geradas no desenvolvimento desse trabalho.

Observou-se que na coleta dos 2500 dados que foram utilizados para análise as formas de escrita são semelhantes, assim como as palavras que foram utilizadas nessas escritas. Mas devido ao grande vocabulário português palavras diferentes podem ter sido utilizadas para uma mesma definição.

Foi realizada uma análise de sentimentos utilizando Mineração de Textos, em português, tendo como observar se os *tweets* compartilhados pelas pessoas tinham características com sentimentos positivos, negativos ou neutros. Além de poder ver quais palavras eram mais usadas para expressarem os sentimentos e quais foram mais citadas nos *tweets*, vale ressaltar que uma das principais dificuldades da automação é identificar quais *tweets* são de fato relacionados a menção utilizada que foi *#educação*.

Referências

- ACKLAND, R. *Web social science: Concepts, data and tools for social scientists in the digital age*. [S.l.]: Sage, 2013. Citado na página 14.
- BAYM, N. K. Personal connections in the digital age. In: . [S.l.]: Polity, 2010. p. 6–20. Citado 2 vezes nas páginas 14 e 15.
- BENEVENUTO, F.; RIBEIRO, F.; ARAÚJO, M. Métodos para análise de sentimentos em mídias sociais. *Sociedade Brasileira de Computação*, 2015. Citado 3 vezes nas páginas 14, 20 e 26.
- BOUFIM, M.; BARKA, H. Building holistic social media strategy referring to social intelligence and digital maturity. In: IEEE. *2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)*. [S.l.], 2015. p. 1–8. Citado na página 19.
- BURGESS, J. E.; GREEN, J. B. *YouTube: online video and participatory culture*. [S.l.]: Polity Press, 2009. Citado na página 14.
- CARVALHO, T. *Visualizing Clusters with Python's Matplotlib*. 2021. Url <https://towardsdatascience.com/visualizing-clusters-with-pythons-matplotlib-35ae03d87489>. Citado na página 24.
- CASSIANO, K. M. Análise de séries temporais usando análise espectral singular (ssa) e clusterização de suas componentes baseada em densidade. *Pontifícia Universidade Católica do Rio de Janeiro*, 2014. Citado na página 37.
- COUTINHO, C. P.; JUNIOR, J. B. B. Blog e wiki: os futuros professores e as ferramentas da web 2.0. 2007. Citado na página 20.
- FILHO, J. A. C. Mineração de textos: análise de sentimentos utilizando tweets referentes à copa do mundo 2014. 2014. Citado 3 vezes nas páginas 20, 22 e 23.
- FRANCO, A. H. C. et al. Inteligência coletiva: manifestações nos ambientes digitais. Universidade Estadual Paulista (UNESP), 2018. Citado 2 vezes nas páginas 19 e 20.
- GOMES, H. J. C. *Text Mining: análise de sentimentos na classificação de notícias [em linha]*. Lisboa: Universidade Nova de Lisboa. Tese (Doutorado) — Tese de Mestrado.[Consult. 12 maio. 2014]. Disponível em WWW:< URL: http . . . , 2013. Citado 3 vezes nas páginas 15, 16 e 20.
- HASSAN, A.; ABBASI, A.; ZENG, D. Twitter sentiment analysis: A bootstrap ensemble framework. In: IEEE. *2013 international conference on social computing*. [S.l.], 2013. p. 357–364. Citado na página 24.
- HEMMATIAN, F.; SOHRABI, M. K. A survey on classification techniques for opinion mining and sentiment analysis. *Artificial Intelligence Review*, p. 1–51, 2017. Citado na página 26.

- JENKINS, H.; FORD, S.; GREEN, J. *Spreadable Media: Creating Value and Meaning in a Networked Culture*. [S.l.]: NYU Press, 2013. Citado na página 14.
- LIU, B. Sentiment analysis and opinion mining. In: *Synthesis Lectures on Human Language Technologies*. [S.l.: s.n.], 2012. Citado na página 26.
- MACHADO, J. R.; TIJIBOY, A. V. Redes sociais virtuais: um espaço para efetivação da aprendizagem cooperativa. *RENOTE*, v. 3, n. 1, jun. 2005. Disponível em: <<https://seer.ufrgs.br/index.php/renote/article/view/13798>>. Citado 5 vezes nas páginas 15, 24, 25, 34 e 37.
- MAIA, R. et al. Sobre a importância de examinar diferentes ambientes online em estudos de deliberação. *Opinião pública*, Universidade Estadual de Campinas, v. 21, n. 2, 2015. Citado na página 16.
- MARKHAM, A.; BUCHANAN, E. Ethical decision-making and internet research: Version 2.0. recommendations from the aoir ethics working committee. *Available online: aoir.org/reports/ethics2.pdf*, 2012. Citado na página 15.
- NASCIMENTO, P.; OSIEK, B.; XEXÉO, G. Análise de sentimento de tweets com foco em notícias. *Revista Eletrônica de Sistemas de Informação*, v. 14, n. 2, 2015. Citado na página 35.
- OLIVEIRA, R. F. d. Análise de sentimentos das postagens e comentários dos principais candidatos à presidência durante a corrida eleitoral de 2018. 2021. Citado na página 20.
- RECUERO, R.; BASTOS, M.; ZAGO, G. *Análise de redes para mídia social*. [S.l.]: Editora Sulina, 2015. Citado na página 19.
- SHIRKY, C. *A cultura da participação: criatividade e generosidade no mundo conectado*. [S.l.]: Zahar, 2011. Citado na página 19.
- THELWALL, M.; BUCKLEY, K.; PALTOGLOU, G. Sentiment in twitter events. *Journal of the American Society for Information Science and Technology*, Wiley Online Library, v. 62, n. 2, p. 406–418, 2011. Citado na página 22.
- YENER, Y. *Step by Step: Twitter Sentiment Analysis in Python*. 2020. [Urlhttps://towardsdatascience.com/step-by-step-twitter-sentiment-analysis-in-python-d6f650ade58d](https://towardsdatascience.com/step-by-step-twitter-sentiment-analysis-in-python-d6f650ade58d). Citado na página 24.