



Universidade Presbiteriana Mackenzie



Universidade Presbiteriana Mackenzie

Pró-Reitoria de Pesquisa e Pós-graduação

Faculdade de Computação e Informática

Pós-Graduação Lato Sensu

**A percepção dos usuários de redes sociais sobre a tecnologia de
NFT e suas aplicações analisadas sob a ótica da análise de
sentimentos com processamento de linguagem natural**

Bruno Rocha Nascimento dos Santos

São Paulo

2022



Bruno Rocha Nascimento dos Santos

**A percepção dos usuários de redes sociais sobre a tecnologia de NFT
e suas aplicações analisadas sob a ótica da análise de sentimentos
com processamento de linguagem natural**

Trabalho de Aplicação de Conhecimento
apresentado ao Curso de Pós-Graduação em
Ciência de Dados – BigData Analytics da
Universidade Presbiteriana Mackenzie para a
obtenção do título de Especialista em Ciência
de Dados – Big Data Analytics.

Orientador: Prof. Ms. Vladimir Aparecido Esteves

São Paulo

2022



Resumo

xx

Palavras-chave: Machine Learning, Blockchain, Tokens Não Fungíveis, Python, Data Science, Ciências de Dados, Analise de Sentimentos, Twiter, Mineração de dados, Data Minning, Classificação, NFT.



SUMÁRIO

INTRODUÇÃO	XX
1 RELATÓRIO DA SITUAÇÃO	
1.1 A empresa.....	XX
1.2 Desafio.....	XX
1.3 Os sintomas	XX
1.4 Objetivo	XX
2 DIAGNÓSTICO	
2.1 As informações	XX
2.2 Análise e diagnóstico	XX
	XX
3 SOLUÇÃO	
3.1 Propostas de Solução	XX
3.2 Conexão Proposta com os Resultados Esperados	XX
4 PLANEJAMENTO	
4.1 Planos de ações	XX
5 CONSIDERAÇÕES FINAIS.....	XX
REFERÊNCIAS	XX
GLOSSÁRIO	XX
APÊNDICES	XX
ANEXOS	XX



LISTA DE QUADROS (se houver)

Quadro 1 (inserir nome do quadro) **XX**

Quadro 2 (inserir nome do quadro) **XX**



LISTA DE FIGURAS (se houver)

Figura 1 (inserir nome da figura) **xx**

Figura 2 (inserir nome da figura)..... **xx**



LISTA DE TABELAS (se houver)

Tabela 1 (inserir nome da tabela) **xx**

Tabela 2 (inserir nome da tabela) **xx**



LISTA DE GRÁFICOS (se houver)

Gráfico 1 (inserir nome do gráfico) **XX**

Gráfico 2 (inserir nome do gráfico) **XX**



INTRODUÇÃO

Em 2021 termo NFT (Nunfungible Token) ocupou uma posição de destaque nas manchetes dos jornais ao redor do mundo e nas pesquisas na internet, perfurando a bolha dos termos médicos e sanitários da Covid-19 e sendo escolhida como a palavra do ano em 2021 pelo dicionário Collins (Colins, 2021). No começo deste mesmo ano a venda da obra digital “The First 5000 Days” do artista Beeple por 69 milhões de dólares e a venda posterior da obra “The Merge” por 91 milhões trouxeram à tona o tema NFT, e as suas diferentes aplicações no meio digital se tornaram mais conhecidas. Trata-se de uma tecnologia baseada no blockchain que permite atribuir um certificado de propriedade digital a um token (item digital registrado) que o torna único, imutável, mas transferível, que pode ser trocado ou comercializado. Assim como as criptomoedas como o Bitcoin e o Ether são tecnologias também baseadas em Blockchain, diferindo-se na sua natureza não-fungível que as denomina, enquanto duas moedas de um mesmo tipo possuem exatamente o mesmo valor e características, cada NFT é único, o que o torna distinguível segundo suas qualidades, e o que permite que sejam comercializados por preços diferentes. Esta tecnologia permite aplicações em diferentes contextos, além das artes gráficas, pode ser aplicado sobre jogos, textos, itens colecionáveis, músicas vídeos e ser incorporada em metaversos e ambientes digitais em diferentes situações.

Neste contexto, e considerando-se a influência que essas novas tecnologias emergentes podem ter sobre nossas vidas, este estudo visa compreender a percepção dos usuários de redes sociais sobre esta tecnologia, explorar o tema com base nos conceitos de NLP (Natural Language Processing), e análise de sentimentos com ferramentas de Big Data a partir da coleta de um grande volume de dados textuais extraídos do Twitter e apresentar conclusões acerca da percepção dos usuários sobre o tema abordado. Dado o crescimento da presença digital das pessoas ao redor do mundo, e do crescimento acelerado do uso de plataformas de redes sociais para expressão, denúncia, divulgação de ideias, reclamações e recomendações, entende-se que pode haver comentários com discursos detratores e promotores à diferentes temas e produtos ocorrendo simultaneamente em volume bastante elevado, de forma que torna-se humanamente impossível o processamento e classificação destes registros manualmente, sendo assim, as ferramentas e técnicas de Big Data figuram entre as principais soluções para a compreensão de grandes volumes de dados, neste caso podem vir a facilitar o entendimento da voz que emana do consumidor/usuário pela internet, e elucidar o comportamento destes diante de novas tecnologias.



1 RELATÓRIO DA SITUAÇÃO

Os tópicos abaixo apresentam respectivamente: A empresa e o termo que são objetos da pesquisa, o desafio que esse trabalho visa a resolver, os sintomas que justificam a necessidade de uma aplicação de um sistema de monitoramento de redes sociais, e por fim, o objetivo geral e os objetivos específicos definidos com o intuito de resolver o problema de pesquisa explicado nos tópicos anteriores.

1.1 A empresa

A OpenSea é uma empresa norte americana sediada em Nova York focada na comercialização de Tokens não Fungíveis (NFTs). A empresa foi fundada por Devin Finzer e Alex Atallah em 2017. A plataforma da OpenSea permite a criação e comercialização de NFTs à preço fixo ou em formato de leilão, executando contratos inteligentes baseados na plataforma Ethereum e negociados na criptomoeda Ether. A OpenSea é a primeira e a maior empresa do segmento sendo avaliada em janeiro de 2022 em US\$ 13,3 bilhões. A empresa possui 70 funcionários, 600 mil usuários cadastrados na plataforma e 80 milhões de NFTs no portfólio (OpenSea, 2022). Os ativos tokenizados comercializados na plataforma podem ser obras de arte digital, vídeos, animações, músicas, propriedades virtuais, colecionáveis e qualquer conteúdo digital único que possa ser negociado online.

De acordo com Figueiredo (2020) um ativo tokenizado é a representação de um ativo real, que pode ser algo físico ou digital e que possui como vantagens a acessibilidade global, a facilidade das transações, a redução dos custos transacionais, a liquidez, e a transparência, garantida pela criação de regras codificadas e imunes à interferência que permitem a auditabilidade das operações. O dicionário Collins define NFT como *“um certificado digital único, registrado em uma blockchain, que é usado para registrar a propriedade de um ativo, como uma obra de arte ou um colecionável”* (Collins, 2021).

Uma das características fundamentais que diferencia uma NFT de uma Criptomoeda é a sua não-fungibilidade. *“Um item fungível para todos os propósitos, pode ser trocado por outra unidade desse mesmo item [...] Itens não fungíveis, por outro lado, não são trocáveis por outro, e tem propriedades únicas que podem fazer deles radicalmente diferentes um do outro, mesmo que pareçam similares”* (LAU, 2019, p. 6). Portanto, criptomoedas podem ser consideradas tokens fungíveis, que são transferíveis e de igual valor entre unidades iguais, enquanto tokens



não-fungíveis, igualmente transferíveis, mas possuem características que os distinguem entre si e possuem métodos para a verificação de propriedade, podendo assim ser negociados a diferentes valores, usados como colecionáveis, certificados de autenticidade entre várias outras utilizações possíveis. (FIGUEIREDO, 2020)

De acordo com LAU (2019, p.6) as NFTs possuem 5 características fundamentais que os diferenciam de outros ativos digitais, NFTs são únicos, rastreáveis, raros, indivisíveis e programáveis, ou seja, Cada token possui códigos que descrevem as propriedades que o tornam diferentes dos demais, cada token registra todas as transações de propriedade pelas quais ele passou desde a sua criação, a não-fungibilidade dos tokens e o caráter único torna-os escassos e atrativos para compradores e especuladores, as NFTs na maioria das vezes não podem ser transacionadas em frações, ou seja, não podem ser fracionadas em partes menores e como outros ativos digitais, NFTs podem ser incorporados a contratos de blockchain.

1.2 Desafio

O desafio consiste em uma pesquisa de mídias sociais aplicando ferramentas de Processamento de Linguagem Natural (NLP) com o intuito de compreender melhor o que os usuários de redes sociais associam aos produtos de NFT comercializados pela OpenSea e outras plataformas aos membros da comunidade para assim poder direcionar campanhas de esclarecimento e de atração a novos membros criadores e consumidores. Com esse intuito, espera-se que este estudo responda de forma exploratória ao seguinte problema de pesquisa: ***“Qual é o sentimento geral dos usuários de redes sociais sobre a tecnologia de NFT e quais associações são feitas a ela?”***

Neste trabalho será analisado especificamente o conteúdo das postagens da rede social Twitter em língua inglesa que contiveram o termo “NFT” em sua composição, a partir dos quais partiremos para a análise de sentimento com as ferramentas de machine learning do NLTK (Natural Language Toolkit) do Python para análise de sentimentos com modelos de processamento de linguagem natural não supervisionados. A amostra foi coletada por 30 dias consecutivos, considerando-se os 2000 últimos tweets únicos do dia na coleta, desconsiderando-se os repetidos, as repostagens e retweets (totalizando 60 mil registros). Cabe ressaltar que os dados coletados pela API do Twitter são dados públicos, e que não será atribuído o sentimento a pessoas ou instituições possam ter emitido suas opiniões ou tenham sido citadas em tais



mensagens e sim ao termo “NFT” referido pelo usuário, também se desconsiderará a localização geográfica do usuário, sua identificação e demais variáveis que possam ser passíveis de coleta por fugirem do escopo desta exploração, por tanto, a base de dados da pesquisa conteve apenas os campos “data”, com o dia da postagem, e “texto do tweet”, com o texto integral em língua inglesa postado pelo usuário, e é sobre este dataset que serão efetuados os devidos tratamentos e análises.

1.3 Os sintomas

A tecnologia relativamente nova de NFTs usada na empresa pode não ser bem compreendida pelo público geral, o conhecimento e as vendas desses ativos digitais tiveram destaque no ano de 2021, e muitas vezes associada negativamente a bolhas de especulação financeira de criptomoedas e ações fraudulentas. Este estudo será particularmente útil para a avaliação da percepção do público sobre a tecnologia e obras de NFT e o sentimento normalmente presente nas postagens de usuários de redes sociais sobre esse tema, e trará a oportunidade de entender quais são os pontos mais valorizados e apreciados pelos usuários e os que desagradam ou causam desconfiança entre eles e assim poder desenvolver materiais que promovam informações corretas que ajudem a popularizar o termo fora da comunidade e também ajudar a fomentar estratégias de marketing que atraiam novos criadores para a plataforma, que popularizem o comercio das obras e engaje potenciais compradores.

1.4 Objetivo

O objetivo geral do trabalho é analisar a percepção do público sobre o tema NFT, considerando o sentimento predominante, e em um panorama geral entender as palavras mais recorrentes associadas ao termo estudado. Diante do objetivo geral proposto podemos subdividir o objetivo geral nos seguintes objetivos específicos:

- a) Mensurar qual foi o sentimento predominante no período analisado
- b) Compreender quais foram as palavras correlacionadas ao tema NFT
- c) Compreender qual a percepção geral do público sobre o tema



2 DIAGNÓSTICO

Como o foco deste projeto é a obtenção de informações acerca da opinião dos usuários sobre o tema NFT em redes sociais, o Twitter foi escolhido por sua popularidade na expressão popular, pela liberdade e pelo caráter aberto dos dados, pois nela toda postagem dos usuários é pública a qualquer outro usuário, diferentemente de outras redes sociais em que o acesso às postagens pode ser restringido, tal como a política de compartilhamento de dados permite a realização de pesquisa e coleta de dados diversos da rede social via API, que possibilita a extração de grandes volumes de dados textuais para análise.

Nesta sessão será detalhada a metodologia da coleta dos dados, tal como a sua composição e o tratamento neles aplicado para a sua correta inserção em modelos de machine learnig que terão a finalidade de identificar e classificar as emoções nos textos coletados. Os dados serão também interpretados e explorados graficamente para a sua melhor compreensão. A análise empírica dos dados será pautada em 4 etapas que serão delimitadas e explanadas em maiores detalhes nos tópicos seguintes: coleta, processamento, visualização, e análise dos dados.

2.1 As informações

Coleta

Silva (2015) compilou as opções de unidades de conteúdo normalmente pelos pesquisadores como objetos de estudo em coletas de dados em pesquisas de mídias sociais e as organizou em 3 categorias, “Expressivo-descritivo”, que engloba as auto-descrições de blogs, biografias e perfis; “Interacional-Reativo” representado pelas interações de feedbacks simples expressos em curtidas e reações aos conteúdos; e o “Discursivo-Intencional”, referente à postagens, comentários, tweets e retweets em que as expressão se dá de forma textual em caráter opinativo e conversacional. A unidade de conteúdo adotada neste estudo é a discursivo-intencional pois por meio dela é possível analisar textualmente a opinião e sentimentos que o usuário exprime acerca dos temas abordados, e que é materializada nos textos nas postagens de tweets.



De acordo com a proposta de estudo deste trabalho, os textos analisados serão os que contêm o termo “NFT” em seu corpo, optou-se neste trabalho por analisar postagens em inglês, pois, a maioria dos usuários da internet se expressam em língua inglesa, assim como o inglês é a língua presente na maioria dos tweets que são postados diariamente e que possuem relevância global, sendo também as tecnologias de NLP desenvolvidas, em sua grande maioria, da avaliação de textos em inglês, o que em muitos casos, conforme observado na literatura de estudos correlatos em língua portuguesa, pode representar uma limitação que prejudica o desempenho da análise dos modelos de NLP (processamento de linguagem natural).

No que se refere às ferramentas adotadas, Silva (2016) organizou os programas voltados à coleta de dados de redes sociais em 3 categorias “plenas/comerciais” que são soluções comerciais que abrangem todas as etapas do processo; softwares acadêmicos/experimentais que oferecem soluções para pesquisas; e “scripts/APIs” que são soluções próprias desenvolvidas em linguagens de programação integradas com APIs das redes sociais, sendo esta última a categoria a adotada para a coleta neste trabalho.

A metodologia de coleta das postagens do Twitter foi realizada com o auxílio de scripts desenvolvidos na linguagem de programação Python, cuja biblioteca Tweepy fornece recursos para o acesso à API do Twitter e um pacote de funções que podem ser adotadas para facilitar a coleta segundo critérios pré-estabelecidos, neste estudo optou-se por analisar as postagens coletadas ao longo de 30 dias consecutivos, no período compreendido entre os dias 09 de fevereiro e 10 de Março de 2022. A cada dia do período descrito foram coletados 2000 tweets. De cada tweet fora armazenado apenas a informação de data e o texto publicado, sendo desconsideradas todas as demais informações passíveis de coleta pela API. O script em Python desenvolvido para a coleta desconsiderou o registro de tweets repetidos e retweets para evitar que uma grande quantidade de anúncios e spams fizesse parte da composição da amostra. Portanto, o conjunto total do arquivo XLSX que compõem o dataset usado nas análises conteve ao final da coleta 60.000 registros em duas colunas “data” e “texto/tweet”, já filtrados de mensagens repetidas e retweets,

Pré-processamento

Nesta etapa o texto foi padronizado no formato ideal para o processamento dos algoritmos de linguagem natural, para a correta aplicação da metodologia o texto foi padronizado em letras minúsculas, pois segundo Beyslow (2018) uma mesma palavra digitada



em maiúsculo, capitalizada ou em minúsculo será interpretada como palavras diferentes, portanto, é importante converter todo o texto para um mesmo padrão. Foram removidos caracteres especiais, e foram removidas as “Stop Words” que consistem em palavras que não carregam em si nenhuma informação relevante, para estas transformações e ajustes foram utilizados códigos em Python escritos com funções das bibliotecas Pandas (principalmente para a manipulação dos dados) e NLTK (Natural Language Toolkit) para o processamento de linguagem natural.

Tokenização

Ao ler um texto no formato de string um programa não consegue diferenciar as palavras dos espaços e pontuações e encara o texto inteiro como um objeto único, neste caso precisamos encontrar um modo de separar o corpo do texto em palavras de forma que permita ao computador avaliar cada palavra no formato de string como um objeto, individualmente. (BEYSOLOW, 2018). A tokenização dos dados consiste em separar cada palavra e caractere especial do texto transformando-as em tokens que serão encarados como objetos únicos, tornando um texto em um formato que seja mais facilmente processado pelas máquinas e que possa ser convertido no formato ideal para o input em modelos de processamento de linguagem natural. (BEYSOLOW, 2018)

Limpeza

Em seguida os tokens são padronizados, filtrados e são removidos os caracteres especiais, além das stopwords, que são palavras comuns e repetitivas que conectam as partes de um texto e que não carregam nenhum significado em si, e por isso não agregam às máquinas maior capacidade de compreensão de um texto. Alguns exemplos de stop words são as palavras em inglês: I, me, myself, it, its, that, this, am, are, was, they, entre muitas outras (BEYSOLOW, 2018).

Normalmente listas de stopwords estão inclusas em pacotes de software que lidam com NLP, na biblioteca NLTK (Natural Language Toolkit) do Python há uma lista de 184 Stop Words em língua inglesa e tais palavras foram exclusas do texto no script de processamento dos dados. O termo NFT e as suas possíveis variações (NFT, NFTs e Non Fungible Tokens) foram inclusos na lista de stopwords, tal como caracteres especiais e stopwords que não foram barrados na filtragem inicial.

Lemantização

Na sequência os tokens passam pelo processo de lematização, que consiste na redução das formas possíveis que um verbo pode se apresentar, suas diferentes variações no presente, passado e futuro são transformados em uma única forma no presente e cada aparição em qualquer possível variação será representada neste único formato. O texto lematizado possui o mesmo número de palavras, mas os verbos aparecem sempre no tempo verbal do presente. A figura abaixo exemplifica o processo de lematização e a transformação a que o texto é submetido.

Figura 1

A transformação dos dados



Fonte: Desenvolvido pelo autor

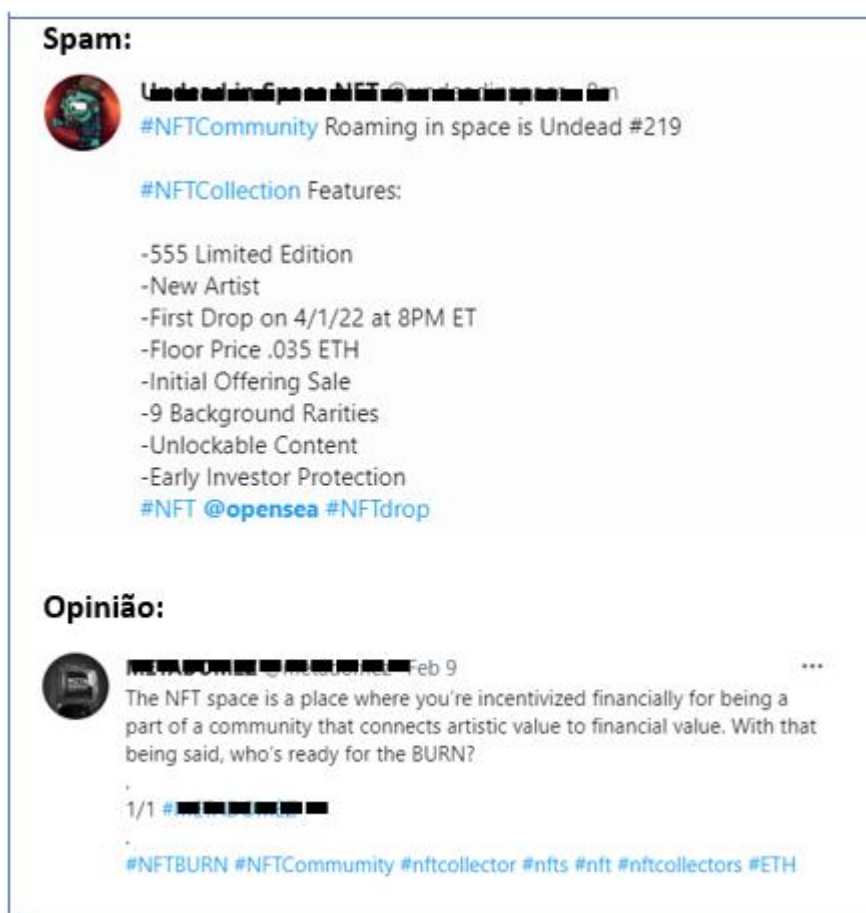
Remoção de Spam

A observação dos dados nos permite perceber que apesar de não constar registros repetidos que inicialmente foram filtrados na coleta, uma grande parte dos registros refere-se à divulgação de ofertas de NFTs por seus criadores e plataformas. Por se tratar de um dado que não carrega em si a percepção dos usuários sobre o tema abordado optou-se por remover os dados referentes à Spam (mensagem eletrônica indesejada, normalmente anúncios publicitários).

A remoção dos dados de Spam foi realizada com a utilização de um modelo supervisionado de classificação do tipo Naive Bayes, treinado com um dataset com 1500 registros tokenizados e rotulados manualmente como “Spam” ou “Opinião”, divididos em um conjunto de treino com 70% dos registros, e de teste com os demais 30%. A avaliação do modelo apresentou uma acurácia de 70%. Por fim, o modelo treinado e avaliado foi aplicado sobre os 60 mil registros e classificou 23.378 textos (39%) como “opinião” e os demais 36.622 (61%) como “Spam”, os modelos classificados como spam foram desconsiderados na aplicação do modelo de análise de sentimentos, na fase posterior. Abaixo apresenta-se um exemplo de postagem que contém um spam e uma opinião:

Figura 2

Exemplos de Tweets



Fonte: Elaborado pelo autor

Por fim, após a coleta e o devido tratamento, o dataset completo com a classificação das opiniões apresenta em sua composição a data e hora da coleta, o texto original, o texto



tokenizado e lematizado e a classificação de spam ou opinião. Conforme a amostra apresentada abaixo:

Tabela 1

Os 9 primeiros registros do dataset

indice	data	texto	token	classe
0	2022-02-09 23:59:59	@kezynfts @sadboyz_nft what an interesting project	['kezynfts', 'sadboyz_nft', 'interesting']	opinion
1	2022-02-09 23:59:59	look how cute this nft is 😊 check em out https://t.co/ouzc	['look', 'cute', '😊', 'check', 'em', 'ouzc']	opinion
2	2022-02-09 23:59:58	@ai_mai_nft done	['ai_mai_nft', 'do']	spam
3	2022-02-09 23:59:58	#nft this is the new collection of the washi artist limited to	['new', 'collection', 'washi', 'artist', 'li']	spam
4	2022-02-09 23:59:58	@legaleagle @bmc_nft offers massive utility, super commu	['legaleagle', 'bmc_nft', 'offer', 'mass']	spam
5	2022-02-09 23:59:58	i do keep working on my ai friendship, meanwhile dropped	['keep', 'work', 'ai', 'friendship', 'mea']	spam
6	2022-02-09 23:59:58	@shilldoebaggins @bobbybananasnft @cryptoland could s	['shilldoebaggins', 'bobbybananasnft']	opinion
7	2022-02-09 23:59:58	@murdockmurr hello everyone !)) 💖👉 check out my crye	['murdockmurr', 'hello', 'everyone', '!']	spam
8	2022-02-09 23:59:57	@themetatribes i'm so exited for this drop nftalpha3#7331	['themetatribes', 'exit', 'drop', 'nftalp']	opinion

Pode-se concluir ao final deste tópico que ao final da fase de processamento os dados encontram-se em um formato adequado para a aplicação em modelos supervisionados e não supervisionados e também em formato adaptado para a geração de gráficos e análises de frequências que serão necessárias para avaliações mais detalhadas nos capítulos seguintes, a análise dos dados brutos permitiu a constatação da presença de um grande volume de postagens comerciais de divulgação e spam que ao final do processo foi eliminado e assim permitindo o uso somente dos dados textuais relevantes para o propósito da pesquisa. No tópico a seguir os dados brutos e processados serão explorados e visualizados para a compreensão.

2.2 Análise e diagnóstico

Nesta sessão será apresentada uma análise exploratória em que são expostos graficamente informações sobre os dados que foram coletados na etapa anterior, esta etapa se mostra relevante pois nos permite avaliar qualitativamente a amostra e assim ter um panorama sobre a visão dos usuários sobre o tema pesquisado, o que permitiria a tomada de decisões mais assertivas em ações de marketing. São apresentadas as suas características e distribuição com a finalidade de aumentar a compreensão sobre as opiniões emitidas de forma qualitativa. Apesar do dataset original ter sido reduzido com o intuito de manter apenas as opiniões sem os anúncios, a avaliação será realizada em duas etapas: Para o conjunto completo dos dados (com os registros de anúncios e opinião) e isoladamente para os registros classificados como opinião, pois o conjunto completo possui os registros de criadores e consumidores dos conteúdos, o que neste caso pode ajudar na compreensão do tema e do fluxo de informações sobre o tema nas

redes sociais.

Analise do Dataset completo

O conjunto dos 60 mil tweets são compostos por 757.876 palavras, das quais 131.977 são termos únicos (com verbos já lematizados), destas palavras 75% se repetem menos de 2 vezes, 99% se repetem até 70 vezes e 99,9% se repetem menos de 700 vezes. Portanto podemos perceber que somente uma quantidade mínima de palavras se repetem um número grande de vezes e essas são as palavras de interesse para esta observação. As 50 palavras que mais apareceram nos textos representam 0,04% do total de palavras e cada uma mais de 1435 vezes, de todas as palavras consideradas apenas 31 apareceram mais de 2000 vezes, sendo o termo mais utilizado pelos usuários a palavras “OpenSea” (6972), “Project” (5919) e NFTMunity (5700).

Figura 3

Nuvem de Palavras



Fonte: Desenvolvido pelo autor

No Word Cloud acima é possível verificar graficamente as palavras que mais se repetem. Nuvens de Palavras normalmente são expressas visualmente de forma que o seu tamanho no quadro indica a sua repetição nos textos. No caso acima podemos observar o verbo “Get”, o substantivo “Project”, e as palavras “Collection”, “Opensea”, “Thank”, “Love” e “Check” em



destaque. Apesar de ser uma apresentação visual interessante que sintetiza a repetição dos termos não permite a identificação da sua frequência no texto, para essa finalidade apresenta-se abaixo a tabela de frequência das 50 palavras mais repetidas.

Tabela 2

As 30 palavras mais frequentes entre as 60 mil postagens sobre NFTs

Palavra	Frequência
Opensea	6972
Project	5919
Nftmunity	5700
Check	4791
Get	4789
Eth	4329
New	3424
Like	3285
Nftart	3209
Collection	3100
Go	3066
Art	2996
Crypto	2969
Mint	2834
Good	2758
Buy	2656
Make	2655
Via	2633
Nftcollection	2581
Love	2494
Join	2419
Look	2415
Nftdrop	2413
Nftcollector	2411
First	2307



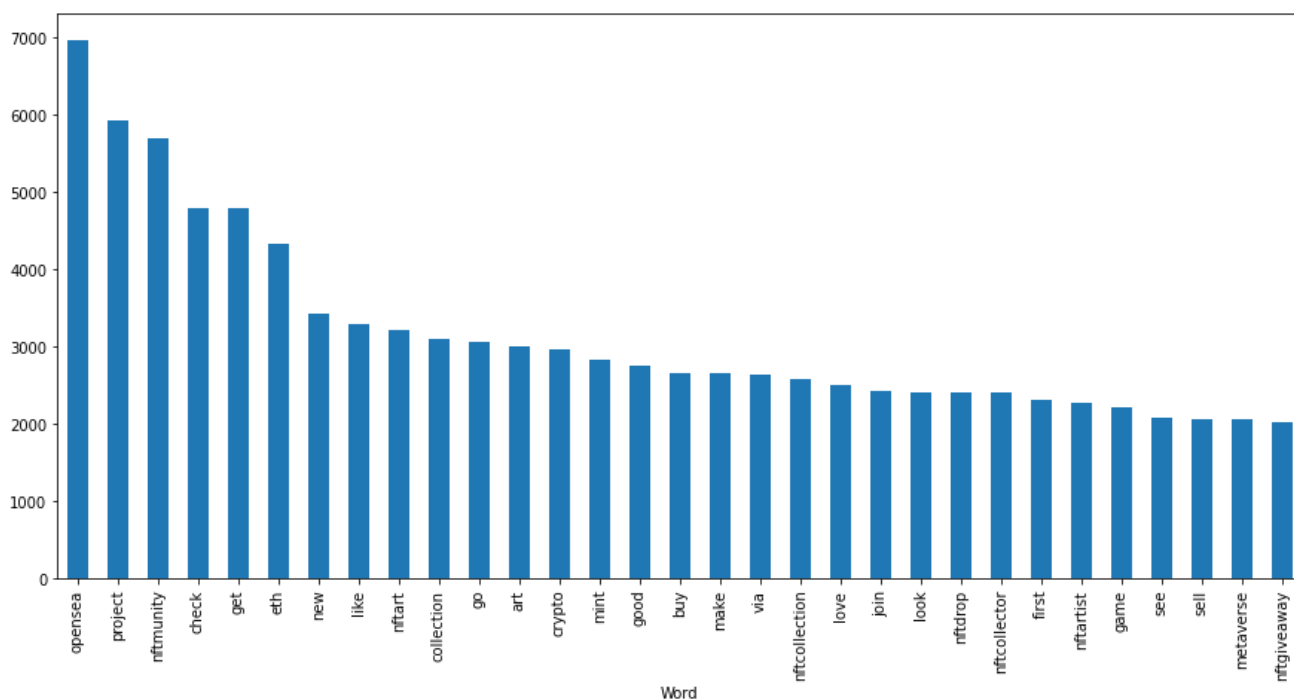
Nftartist	2262
Game	2205
See	2088
Sell	2063
Metaverse	2061
Nftgiveaway	2019

Fonte: Desenvolvido pelo autor

O gráfico abaixo nos permite comparar a ocorrência das palavras e compreender quais as que mais se repetem com o intuito de entender os motivos de sua ocorrência.

Gráfico 1

A distribuição da ocorrência das 31 palavras repetidas mais de 2000 vezes



Fonte: Desenvolvido pelo autor a partir de coletas do Twitter

Pode-se observar que há 31 palavras que se repetem mais de 2000 vezes nos textos. Graficamente observamos que as 6 primeiras palavras (Opensea, Project, NFTMunity, Check, Get e ETH) ultrapassam 4 mil repetições, enquanto as demais ficam entre 2000 e 3500 repetições.

Entre as palavras mais presentes podemos observar termos diretamente relacionadas ao termo



como plataformas e a moeda corrente OpenSea e Ethereum (plataformas de criação, registro e comércio de NFTs), NFTMunity e Munity (Comunidade de interessados por NFTs), ETH (Sigla da criptomoeda Ether, usada para as negociações na plataforma da Ethereum), Mint (emissão de NFT), AMP (moeda digital).

Por fim, é possível concluir que nesse conjunto de dados as palavras que são relacionadas a aplicações de NFT nas artes digitais, neste caso as palavras NFTart, Art, NFTartist e Crypto (que pode ser usada na combinação com “Art” formando o termo “Crypto Art”) são associadas ao uso recorrente da tecnologia para a certificação de artes digitais, também os termos collection, nftcollection e nftcollector que podem se referir tanto a artes digitais como a outras coleções de NFT. Há também a citação a games, ao metaverso e à plataforma de vendas Opensea.

Entre os verbos mais comuns nas postagens sobre NFTs, percebemos que isoladamente podem ser interpretadas como sendo verbos imperativos (check, get, go, make, buy, look, see, sell, join love, like) que convidam o leitor a tomar uma atitude com relação as NFTs, obter, vender, ir, ver, conferir, criar, juntar-se que podem indicar que boa parte das postagens tratam-se da divulgação de criadores ou de plataformas ou notícias sobre NFTs criadas, a presença da palavra new entre as mais recorrentes corrobora com essa conclusão. Há também o verbo Love e Like e as palavras good e Great e Thanks, que remetem a um sentimento positivo expresso pelos usuários com relação ao tema.

Análise dos Tweets filtrados

O conjunto dos tweets classificados como opinião é composto por 23.378 registros, com um total de 269.205 palavras, das quais 56.504 são termos únicos lematizados, destas palavras 75% se repetem menos de 2 vezes, 99% se repetem até 67 vezes e 99,9% se repetem até 477 vezes. As 50 palavras que mais apareceram nos textos representam 0,0008% do total de palavras e cada uma mais de 512 vezes, As 30 palavras que mais se repetiram apareceram mais de 675 vezes, sendo o termo mais utilizado pelos usuários a palavras “Project” (3420), “Good” (1899) e “Get” (1827), “Like” (1596), “Go” (1580) e “Love” (1536). Abaixo o wordcloud com as palavras dos textos classificados como opinião.

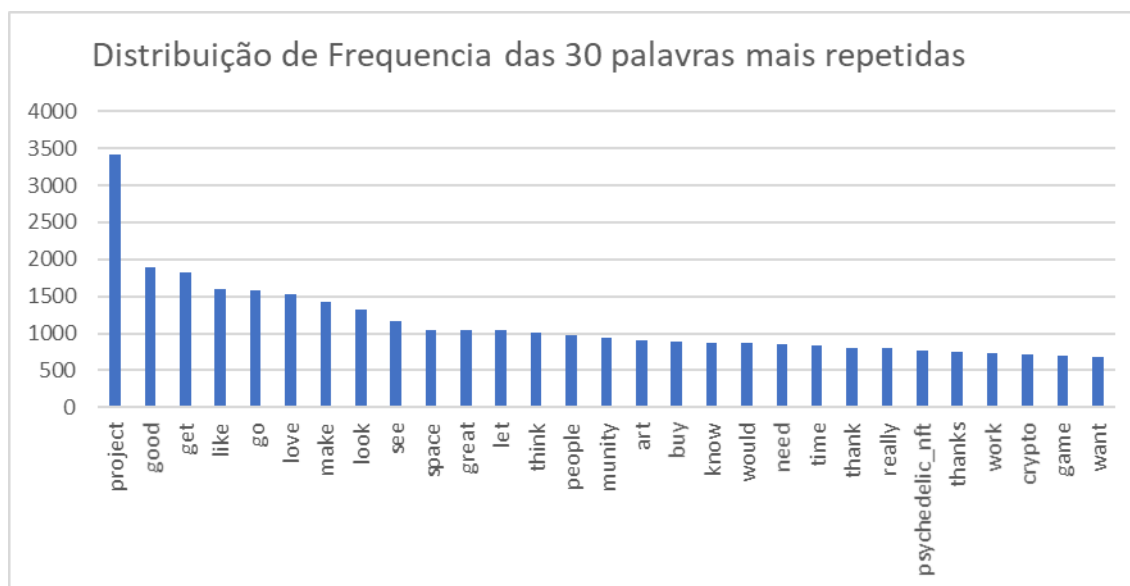
Figura 3

Nuvem de Palavras



Fonte: Desenvolvido pelo autor

Gráfico 2



Fonte: Desenvolvido pelo autor

Pode-se observar que das 30 palavras mais repetidas, 9 ultrapassam 1000 repetições e apenas a palavra Project ultrapassa 2000 repetições.

Tabela 3

As 30 palavras mais frequentes entre as 60 mil postagens sobre NFTs



Word	Frequency
Project	3420
Good	1899
Get	1827
Like	1596
Go	1580
Love	1536
Make	1420
Look	1328
See	1159
Space	1052
Great	1042
Let	1038
Think	1012
People	968
Munity	945
Art	905
Buy	893
Know	875
Would	868
Need	850
Time	832
Thank	796
Really	795
psychedelic_nft	760
Thanks	756
Work	737
Crypto	715
Game	695
Want	675

Fonte: Desenvolvido pelo autor



O conjunto das informações acima nos permite concluir que as palavras que são relacionadas a aplicações de NFT em games, artes e metaverso, mas percebe-se uma recorrência menor de substantivos próprios e termos específicos relacionados ao tema e uma predominância de verbos e de palavras que carregam sentimentos positivos.

Entre as palavras mais comuns nas postagens sobre NFTs, estão as palavras Good, Love, Like, Great, Nice, Amazing, Best, Thanks e Hope que carregam uma conotação positiva, e também verbos imperativos como Get, Go, See, Take e Mint.

Neste capítulo foram apresentados os dados que serão necessários para atingir os objetivos definidos e apresentados nos capítulos anteriores, foi explicada a metodologia adotada para a coleta dos dados, os critérios de amostragem, a composição do dataset com os dados da pesquisa, o formato bruto dos dados no input do pré-processamento, a explicação sobre os métodos adotados na transformação e filtragem dos dos tweets, e o output após o tratamento. As informações foram exploradas visualmente em gráficos, wordclouds e tabelas para a sua melhor compreensão que ajudará na correta aplicação da solução que será apresentada no capítulo seguinte.

3 SOLUÇÃO

Conforme explicado nos capítulos anteriores, busca-se uma solução que auxilie na análise dos dados coletados através do Twitter que permita uma avaliação sobre a opinião dos usuários sobre NFTs, produto principal da OpenSea. Neste capítulo é apresentada a proposta de solução que será adotada na resolução do problema de pesquisa, serão discutidos os impactos que espera-se que a solução cause, são explicitados os pontos favoráveis e desfavoráveis da proposta, e por fim é apresentado o embasamento teórico deste projeto, com o detalhamento da ferramenta que será utilizada e com a revisão bibliográfica de outros trabalhos que adotaram soluções parecidas em outros problemas em diferentes contextos, e que justificam o uso da solução.

3.1 Propostas de solução

Dentre as possíveis soluções que podem ser adotadas para a finalidade da pesquisa, apresenta-se neste como solução para o desafio proposto a criação de um script em Python para a aplicação de um modelo de análise de sentimentos, optou-se por utilizar da biblioteca livre



VADER (Valence Aware Dictionary for Sentiment Reasoning) especializada em mineração de opiniões em postagens de redes sociais.

➤ **Proposta : Aplicação do dicionário léxico VADER (Valence Aware Dictionary for Sentiment Reasoning) em um modelo de análise de sentimentos/emoções para a avaliação de tweets**

- **Descrição da proposta:** Dentre as possíveis soluções que podem ser adotadas para a finalidade da pesquisa, apresenta-se como solução para o desafio proposto a criação de um script em Python para a aplicação de um modelo de análise de sentimentos, optou-se por utilizar da biblioteca livre VADER (Valence Aware Dictionary for Sentiment Reasoning) por ser uma biblioteca completa para a análise e classificação de textos e especializada em mineração de opiniões em postagens de redes sociais. Será desenvolvido um script em Jupyter Notebook na linguagem de programação Python que fará a adequação necessária aos dados e aplicará sobre eles a função de classificação do Vader, que rotulará todos os tweets do dataset como positivo, negativo ou neutro, de acordo com o conteúdo do texto.
- **Possíveis impactos:** O uso de algoritmos para a coleta de dados de redes sociais aliados a ferramentas computacionais que auxiliem na correta interpretação destas informações ajudará a empresa na compreensão sobre o que os usuários de internet e possíveis negociadores em potencial entendem sobre o produto comercializado e sobre a própria imagem da empresa, e assim poderá focar as suas ações de marketing de forma mais assertiva e direcionada, também a partir da análise de sentimento automática poderá filtrar as mensagens negativas e assim entender quais os pontos de insatisfação dos consumidores e da comunidade relacionada às atividades da empresa ao mesmo tempo em que avaliam os pontos que são mais valorizados nas mensagens classificadas como positivas para reforça-las.
- **Prós:** A solução adotada não envolve a contratação de Softwares especializados tornando a implementação menos onerosa, por se tratar do desenvolvimento de um script próprio há a possibilidade de flexibilizar a finalidade das pesquisas e de fazer adaptações que podem não ser possíveis em softwares comerciais. A linguagem de programação Python adotada no projeto é uma linguagem de propósito geral com aplicações



especialmente relevantes em análise de Big Data e possui bibliotecas completas voltadas para esta finalidade, tem grande popularidade entre os programadores e cientistas de dados e uma comunidade ativa, o que pode facilitar a aquisição e o treinamento da mão de obra necessária para manter e modificar os scripts.

- **Contras:** O desenvolvimento dos scripts para a coleta, classificação e análise dos dados apesar de mais barata demandaria um esforço maior de aplicação pois envolve a elaboração dos códigos, das métricas e da avaliação manual dos pontos relevantes a serem considerados na arquitetura do código. Para uma possível finalidade comercial de adoção em larga escala de um modelo de monitoramento de redes sociais em tempo real (principalmente no caso de monitoramento da marca) pode ser necessário a utilização de softwares de Cloud Computing para o armazenamento e processamento em nuvem dos dados, e assim aumentar a produtividade, o que pode demandar custos mais elevados.
- **Recursos:** Por se tratar de um projeto desenvolvido em Python com IDE Jupyter com a implantação dos modelos de classificação e de análise de sentimento através de bibliotecas livres do Python e atuando dentro das limitações de amostragem da API o projeto tal como foi idealizado para a finalidade específica de estudar um termo de pesquisa por amostragem tende a ter um baixo custo de implementação, sendo restrito o custo ao acesso à API do Twitter que em nível empresarial envolve uma despesa mensal de US\$ 2900,00 que pode escalar conforme a utilização em caso de uma necessidade recorrente ou para a busca de dados em tempo real (streaming), além de custo das horas de trabalho de um profissional (ou equipe) da área de dados com conhecimentos em programação em Python.

- **Teoria de suporte e autor:**

Segundo Hutto & Gilbert (2014) modelos de análise de sentimentos podem ser úteis para um grande número de problemas que envolvem a interação entre humanos e máquinas para finalidades comerciais e de pesquisa, as aplicações podem ocorrer nos campos de sociologia, marketing, publicidade, economia e ciência política. No campo de estudo da Análise de Conteúdo em textos, Silva (2018) apresenta a análise de sentimentos como um método para a



realização de uma Análise Relacional de conteúdo, em que o foco da pesquisa é extrair sentido não somente da presença dos conceitos, mas também da relação entre esses conceitos no texto, e define a análise de sentimento/emoções como “*Um método que busca compreender a emoção dos conceitos presentes no texto, geralmente através de análises temáticas e de elementos relacionados a essas emoções*” (SILVA, 2018,p.15)

O modelo adotado para a classificação neste estudo segue a aplicação de um dicionário de sentimento léxico, que é uma lista de características lexicais em que as palavras são rotuladas de acordo com a sua orientação semântica como positiva ou negativa. O VADER (Valence Aware Dictionary for Sentiment Reasoning) é um dicionário léxico criado com uma combinação de métodos quantitativos e qualitativos desenvolvido e testado em diferentes contextos, mas com desempenho especialmente melhor em textos no formato de micro-blog, tal como postagens de redes sociais como o Twitter (Hutto & Gilbert, 2014). No caso do VADER este modelo de dicionário lexical foi criado com a utilização de recursos lexicais e com cinco regras gerais que incorporam convenções gramaticais e semânticas para determinar a intensidade do sentimento. No caso as 5 regras são compostas de: Pontuação (palavras que antecedem pontos de exclamação); Capitalização, palavras capitalizadas, especialmente a presença de palavras em capslock junto de outras palavras não capitalizadas podem indicar palavras com sentimentos relevantes; Advérbios de intensidade, consistem na análise das palavras que impactam no acréscimo ou decréscimo da intensidade do sentido de adjetivos e de outros advérbios no contexto da frase; a análise do trigramma lexical que precede uma palavra que carrega um sentimento e Conjunções adversativas, palavras tais como a palavra “mas” que indicam a mudança da polaridade do sentimento e apresenta uma contraposição entre duas partes de um texto. De acordo com o autor, a avaliação indica que o modelo é capaz de performar tão bem quanto um grupo de 20 avaliadores humanos. (Hutto & Gilbert, 2014)

A seguir apresentam-se diferentes estudos em que técnicas de mineração de opinião e modelos de análise de sentimentos foram adotados como soluções em pesquisas de conteúdo em redes sociais em diferentes contextos e para finalidades distintas. No contexto de mineração de opinião sobre produtos, Xavier (2019) avaliou a percepção de consumidores sobre o sistema operacional Android 10 em seu mês de lançamento, a autora coletou tweets com a hashtag #Android10 e com a função “Sentiment”, da biblioteca de mesmo nome classificou com um modelo os tweets como positivos ou negativos e obteve uma acurácia de 85% e com os resultados concluiu que as opiniões sobre o Android10 naquele momento foram



predominantemente positivas. Também no cenário de avaliação da satisfação do consumidor, Saura et al (2018) comparou as opiniões de consumidores sobre ofertas de Black Friday com a biblioteca do Python “Monkey Learn” aplicando o modelo a ofertas de diferentes empresas e comparando as opiniões e sentimentos sobre as promoções divulgadas.

Em avaliação de eventos e atrações turísticas temos o trabalho de Crescêncio et al (2020) que utilizou de modelos supervisionados de machine learning para avaliar qualitativamente o evento “Oktoberfest Blumenau”, o modelo probabilístico Naive Bayes conseguiu uma acurácia de 90% e a rede neural LSTM, 92%, e a metodologia permitiu aos autores a identificação dos pontos que os usuários mais valorizaram, como a música, a variedade de bebidas e também os fatores negativos, como a superlotação, a dificuldade de acesso aos banheiros e as filas. Feliciano & Farah (2018) utilizou algoritmos de análise de sentimentos para avaliar tweets sobre a Copa do Mundo de futebol de 2018, neste trabalho comparou-se o desempenho de diferentes algoritmos na classificação (OpinionFinder, Umigon, SentiStrength, Vader, OpinionLexicon e Affin) aplicados sobre 4000 tweets rotulados e a análise da matriz de confusão permitiu concluir que o OpinionFinder obteve o melhor desempenho entre os modelos avaliados, com acurácia de 95%, contra 85% do Umigon, 80% do SentiStrength, 67% do Vader, 65% do OpinionLexicon e 42% do Affin.

Medeiros et al (2016) realizou a mineração de opiniões para avaliar no Twitter dois momentos no processo de impeachment da presidenta Dilma Rousseff em 2016. A metodologia adotada por Medeiros neste trabalho consistiu em aplicar a base de dados do dicionário Lexico Emolex (base de dados com palavras em inglês associadas ao sentimento que carregam) traduzida para o português e aplicada a tweets em dois momentos do processo de impeachment e a partir dos sentimentos predominantes entender como o ânimo social se apresentou em diferentes momentos políticos. Também no sentido de entender a opinião pública para a adoção de medidas sociais e políticas, Oliveira et al (2017) testou se a análise de emoções pode refletir a opinião pública e contribuir para a gestão social, neste estudo os autores compararam as opiniões sobre 4 políticas públicas do governo brasileiro (Minha Casa Minha Vida, Bolsa Família, Mais Médicos e Pronatec) com o software de análise textual DiscoverText com uma técnica de classificação supervisionada e obteve precisão acima de 70% nas classificações. O estudo permitiu aos autores identificar quais programas sociais eram associados a sentimentos majoritariamente positivos e quais eram associados a sentimentos negativos.



No caso da avaliação de ativos digitais, o estudo de Franceschet (2021) analisou o sentimento presente nos comentários da plataforma da SuperRare (empresa que comercializa NFTs) sobre obras digitais de NFT anunciados na plataforma, com o dicionário léxico da NRC Emoticon Lexicon que avalia 8 emoções básicas (raiva, medo, antecipação, confiança, surpresa, tristeza, alegria e nojo) e a polarização positivo/negativo presentes nas avaliações, e o aplicou sobre os comentários de artistas e usuários separadamente. Há trabalhos que relacionam as opiniões no Twitter sobre o Bitcoin a oscilações da criptomoeda em tentativas de previsão de tendências para a cotação da moeda com base nestas opiniões mineradas. Mohapatra et al (2020) utilizou a biblioteca VADER em Python para propor um modelo de predição da cotação da criptomoeda a partir da análise de opiniões do Twitter, ressaltando que existem muitos outros fatores além de textos de redes sociais que podem inferir na cotação. Com esta mesma finalidade Wong (2021) utilizou do modelo não supervisionado Vader e de um modelo Supervisionado Naive Bayes para a previsão, e apesar de ter obtido uma acurácia baixa, de aproximadamente 50%, o modelo VADER apresentou um resultado superior ao Naive Bayes, o autor ressalta a necessidade de uma atenção especial na fase de pré-processamento dos dados para a obtenção de resultados melhores. Sattarov et al (2021) utilizou também o modelo VADER em um estudo que pôde comprovar a existência de uma correlação positiva entre o sentimento em postagens no Twitter sobre Bitcoins e o preço corrente da moeda digital. Os estudos apresentados acima demonstram a recorrência do uso de modelos de machine learning em trabalhos acadêmicos de finalidades práticas na análise textual e também a aceitação e aplicabilidade do dicionário VADER neste contexto.

3.2 Conexão da proposta com os resultados esperados

Analisando os pontos explicados nesta sessão podemos concluir que o uso de modelos de análise de sentimento supervisionados e não-supervisionados é recorrente na literatura, e uma solução eficiente para a mineração de emoções em textos de redes sociais, com aplicações práticas em diferentes contextos, desde usos para a avaliação de novos produtos, monitoramento de marcas, estudo sobre eventos e atrações turísticas em avaliações de programas sociais ou medidas políticas, podem ser adotados sobre dados textuais de diferentes plataformas, sejam textos de microblogs, notícias jornalísticas, e avaliações em plataformas de compras, entre outros. Sobre a ferramenta adotada (VADER) concluímos que ela se apresenta como uma



solução adequada ao propósito deste trabalho, fora validada por seus autores na comparação com avaliadores humanos e adotada amplamente por outros autores em diferentes pesquisas por seu caráter livre, gratuito, independente de supervisão humana (que demandam mais tempo para a rotulação manual dos conjuntos de dados de treinamento e teste para o modelo) e adequada ao propósito de classificar textos de redes sociais, sobre os quais costuma apresentar acurácia elevada e satisfatória conforme o referencial apresentado. A citação dos estudos relacionados e os seus resultados nos permite comprovar a eficácia desta metodologia em trabalhos acadêmicos e ilustra outras aplicações práticas. Espera-se que a sua aplicação no contexto deste estudo sobre a OpenSea e ativos digitais ajude a classificar com acurácia elevada o conjunto dos tweets previamente coletado e tratado e a partir da separação dos datasets de tweets positivos, neutros e negativos analisar separadamente cada um deles e assim conseguir alcançar o objetivo principal da pesquisa. Para a análise serão adotados recursos estatísticos de contagem de frequência de palavras, gráficos, wordclouds além de matriz de confusão para avaliar a acurácia, sensibilidade e especificidade do modelo na classificação dos dados. O detalhamento da implementação da proposta de solução, a descrição das etapas e métricas será o foco do capítulo seguinte.

4 PLANEJAMENTO

XX

4.1 Planos de ações

XX

Quadro 1 – Proposta de solução A: **XXXXXXXXXXXXXXXXXXXXXXXXXXXX**

Objetivo: (descreva-o, iniciando sempre com um verbo de ação. Exemplos: revisar, estruturar, promover, criar, inibir, etc.)		
Ação Detalhada	Prazo para Finalização/Implantação	Responsável (área/função)



Quadro 2 – Proposta de solução B: XXXXXXXXXXXXXXXXXXXXXXXXXXXX

Objetivo: (descreva-o, iniciando sempre com um verbo de ação. Exemplos: revisar, estruturar, promover, criar, inibir, etc.)		
Ação Detalhada	Prazo para Finalização/Implantação	Responsável (área/função)

5 CONSIDERAÇÕES FINAIS

XX



FIM

REFERÊNCIAS

COLLINS. **Word of the Year 2021**. 2021. Disponível em:

<<https://blog.collinsdictionary.com/language-lovers/get-your-crypto-at-the-ready-nfts-are-big-in-2021/>>. Acesso em: 17 fev. 2022

FIGUEIREDO, Daniel D. Fundamentos em Blockchain. **Instituto de Gestão e Tecnologia da Informação**. 2020. Disponível em: <https://d1wqtxts1xzle7.cloudfront.net/63983776/Apostila_-_FBL20200721-57366-p1hvct-with-cover-page-v2.pdf?Expires=1646178419&Signature=O5BraLL8qf1ou09tQ9qsAHWZiF8z3QX~rCmYlCXlB7yVDfWOHGXokHG5ZVV1ChEQoo2v-S3iBUFiPZ8glxq0QPFzCQ15-YZevDZxZlwYNQRW~gamMZiuz1HyLzAA6yc46N9JKBqWx8RDHMnWKDKJt7HVC~8seXUVK1uHnWzDN6iBzl42G8V-ZV~ztohSYRgi8VYev2MpXP~G7Ay8Lzq2MePsH2uYPjz9CoiJYqsVAV08IJfPk0EWjEHY3M792cLUvLfx6TmMjwl21Oyz2aUebPNFcvV5-IirJZT6aFjKp~ZC1zlia99-



[d1ApzaPiGVD63yGa2wosgsDL7hXyyX84sA_&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA>](#). Acesso em 01 Mar de 2022.

LAU, Kendrick. Non-Fungible Tokens. A Brief Introduction and History. Crypto.com. 2019. Disponível em: https://assets.ctfassets.net/hfgyig42jimx/6A8K5H6VrTydTDuEFHXQ5P/3cca896ad77bd967859a7a1256a5a91f/Crypto.com_Macro_Report_-_Non-Fungible_Tokens.pdf. Acesso em: 21 fev. 2022.

SILVA, Tarcízio. **Pesquisa baseada em Dados Sociais Digitais: mapeamento de ferramentas e táticas de coleta de dados no Intercom**. Razón y Palabra, n.90. Monterrey : Instituto Tecnológico de Monterrey (Departamento de Comunicação), 2015. Disponível em: <http://www.razonypalabra.org.mx/N/N90/Varia/28_Silva_V90.pdf>. Acesso em 05 abr 2022.

SILVA, Tarcízio; STABILE, Max (Orgs.). **Monitoramento e pesquisa em mídias sociais: metodologias, aplicações e inovações**. São Paulo: Uva Limão, 2016. Disponível em < [\(PDF\) Monitoramento e Pesquisa e Mídias Sociais: metodologias, aplicações e inovações \(researchgate.net\)](#)>. Acesso em 05 abr 2022.

BEYSOLOW T. II, **Applied Natural Language Processing with Python**, San Francisco, California, USA . Apress 2018. Disponível em: https://doi.org/10.1007/978-1-4842-3733-5_1. Acesso em 14 mar 2022

--

HUTTO, C., & GILBERT, E. **VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text**. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), 216-225. 2014. Disponível em:<<https://ojs.aaai.org/index.php/ICWSM/article/view/14550>>. Acesso em 05 abr 2022.

SILVA Tarcizio. BUCKSTEGGE Jaqueline. ROGEDO Pedro. **Estudando Cultura e Comunicação com Mídias Sociais**. Instituto Brasileiro de Pesquisa e Análise de Dados – IBPAD, 2018. Disponível em: <[\(PDF\) Estudando Cultura e Comunicação com Mídias Sociais \(researchgate.net\)](#)> . Acesso em 05 abr 2022.



XAVIER Juliana. **O uso da análise de sentimentos no Twitter para avaliar a opinião do público consumidor a respeito do sistema operacional mobile Android 10.** Faculdade Antonio Meneghetti AMF. Restinga Seca, Rio Grande do Sul, 2019. Disponível em:<

SAURA Jose Ramon; REYES-MENENDEZ Ana; PALOS-SANCHEZ Pedro. **Un Análisis de Sentimiento en Twitter con Machine Learning: Identificando el sentimiento sobre las ofertas de #BlackFriday.** Revista Espacios vol 39 nº 42. Pag 16. 2018. Disponível em:< <http://www.revistaespacios.com/a18v39n42/a18v39n42p16.pdf>>. Acesso em 03 mar 2022.

CRESCENCIO et al. **Um processo de classificação de texto: Análise de Sentimento das Opiniões no TripAdvisor sobre a atração Oktoberfest Blumenau.** Congresso Internacional de conocimiento e innovación. 2020. Disponível em: <<https://proceeding.ciki.ufsc.br/index.php/ciki/article/view/867>>. Acesso em 05 e Abril de 2022.

FELICIANO Patrícia; FARAH Paulo R. **Avaliação de Algoritmos de Análise de Sentimentos em Tweets no Domínio da Copa do Mundo FIFA 2018.** Anais da 2a Escola Regional de Engenharia de Software (ERES 2018) em Dois Vizinhos, PR, Brasil, 22 a 24 de outubro de 2018. Disponível em: <<https://sol.sbc.org.br/index.php/eres/article/view/10059>>. Acesso em: 25 fev 2022.

MEDEIROS et all. **Introdução à análise de sentimentos em redes sociais: O léxico construído em dois momentos do processo de Impeachment.** IX Simpósio Nacional ABCiber- PUC São Paulo – 8, 9, 10 Dezembro de 2016. <https://abciber.org.br/anaisletronicos/wp-content/uploads/2016/trabalhos/introducao_a_analise_de_sentimentos_em_redes_sociais_o_lexico_construido_em_dois_momentos_do_processo_de_impeachment_jean_maicon_rickes_medeiros.pdf>. Acesso em 05 abr 2022

OLIVEIRA et al. **A aplicação da técnica de análise de sentimento em mídias sociais como**



instrumento para as práticas da gestão social em nível governamental. RAP- Revista de Administração Pública- Rio de Janeiro 53(1):235-251, jan. - fev. 2019. Disponível em: <<https://www.scielo.br/j/rap/a/GD3F8HdkQKGSHy8zzV8w9Ys/?lang=pt>>. Acesso em 05 de Abril de 2022

FRANCESCHET Massimo. **The Sentiment of Crypto Art.** Department of Mathematics, Computer Science, and Physics – University of Udine – Italy. 2021. Disponível em; http://ceur-ws.org/Vol-2989/long_paper10.pdf. Acesso em 05 de Abril de 2022

MOHAPATRA et al. **KryptoOracle: A Real-Time Cryptocurrency Price Prediction Platform Using Twitter Sentiments.** 2020. Disponível em: <<https://doi.org/10.48550/arXiv.2003.04967>>. Acesso em 05 de Abril de 2022

WONG Eugene L.X. **Prediction of Bitcoin prices using Twitter Data and Natural Language Processing.** 2021. Pratt School of Engineering, Duke University. Disponível em: <<https://dukespace.lib.duke.edu/dspace/handle/10161/24081>>. Acesso em 04 de abril de 2022

SATTAROV et al. **Forecasting Bitcoin Price Fluctuation by Twitter Sentiment Analysis.** International Conference on Information Science and Communications Technologies. 2020. Disponível em < [\(PDF\) Forecasting Bitcoin Price Fluctuation by Twitter Sentiment Analysis \(researchgate.net\)](#)>. Acesso em 04 de Abril 2022



GLOSSÁRIO (se houver)



APÊNDICES (se houver)



ANEXOS (se houver)