

PERSPECTIVAS DA BLACK FRIDAY NO BRASIL EM 2022 ATRAVÉS DE PUBLICAÇÕES NO TWITTER

Guilherme Velinças Chalhoub e Silva¹
Murilo Henrique Tank Fortunato²

RESUMO

Seja em formato de vídeo, fotografias ou até mesmo escrito não dá para negar o crescimento e a importância que as redes sociais vêm obtendo no decorrer dos anos. Cada vez mais as pessoas utilizam desse meio para se comunicarem, compartilhem experiências e apresentarem opiniões sobre os mais variados temas. Nesse contexto, as empresas passaram a usufruí-las a fim de estreitar ainda mais as relações com os clientes e, ao mesmo tempo, coletar informações valiosas para o negócio, como por exemplo o que melhorar em um produto ou serviço e ter dados em tempo real sobre acontecimentos ao redor do mundo. O Twitter é uma dessas redes sociais onde os usuários expressam opiniões de maneira sucinta. Uma maneira de transformar os textos em informações relevantes para as organizações é através da análise dos sentimentos, onde avalia-se se o que estão escrevendo sobre um determinado assunto é positivo ou negativo. Nesse sentido, o presente trabalho analisou o sentimento dos tweets a respeito da Black Friday no Brasil em 2022, cujo objetivo era verificar a polaridade dos comentários na rede social. Para isso, foi necessário coletar diversos tweets sobre o tema, aplicar as técnicas de Processamento de Linguagem Natural, treinar e validar o modelo de machine learning até que, no fim, fosse possível realizar as análises. Conclui-se que, embora os resultados obtidos não sejam verdades absolutas devido a alguns fatores, os consumidores escreveram, em sua maioria, comentários de cunho negativo a respeito da Black Friday no Brasil em 2022.

Palavras-chave: análise de sentimentos; consumidores; linguagem; tweets.

ABSTRACT

Whether in vídeo, photographs or even in text format there is no way to deny the growth and importance that social networks have obtaining over the years. More and more people are using this environment to communicate, share experiences and present opinions on the most varied topics. In this context, companies began to take advantage of them in order to further strengthen relationships with customers and, at the same time, collect valuable information for the business, such as what to improve in a product or service and collect data about happenings around the world in real time. Twitter is one of those social networks where users express opinions succinctly. One way to transform texts into relevant information for organizations is through sentiment analysis, where it is evaluated whether what they are writing about a particular subject is positive or negative. In this sense, the present work analyzed the sentiment of tweets regarding Black Friday in Brazil in 2022, whose objective was to verify the polarity of comments on the social network. For this, it was necessary collecting several tweets about the subject, apply Natural Language Processing techniques, train and validate the machine learning model until, in the end, it was possible to perform the analyzes. It was concluded that, although the results obtained are not absolute truths due to some factors, consumers mostly wrote negative comments about Black Friday in Brazil in 2022.

Keywords: consumers; language; sentiment analysis; tweets.

¹ Engenheiro Mecânico. Especialista em Data Science e Analytics. Rua Desembargador Motta, 3428 – Mercês; 80430-232 Curitiba, Paraná, Brasil. <https://orcid.org/0009-0004-3987-4239>

² Pecege. Doutor em Agricultura Sustentável. Rua Cezira Giovanoni Moretti, 580 – Santa Rosa; 13414-157 Piracicaba, São Paulo, Brasil. <https://orcid.org/0000-0002-1695-8120>. guilherme.chalhoub@usp.br

Introdução

Para se comunicarem e se expressarem entre si, seja de forma escrita ou falada, os humanos usufruem da denominada linguagem natural. Por estar em constante evolução, conter certas ambiguidades e possuir diversas variações, o uso de computadores com a finalidade de assimilar e produzir linguagem é bem desafiador, entretanto torna-se cada vez mais necessário fazê-lo devido a imensas quantidades de informações no formato de linguagem natural sendo produzidas a cada minuto, a partir do relacionamento dos seres humanos com as máquinas. Nesse contexto, Chowdhury (2003) define que o Processamento de Linguagem Natural (PLN) permite o entendimento e a manipulação por meio de computadores da linguagem natural, em forma de texto ou fala, para uma determinada utilidade. Turban et al. (2019) destaca que o PLN tem a finalidade de fazer a conversão de textos em dados, os quais são mais facilmente manipulados em programas computacionais. Para Miranda (2009) é possível usufruir todo o conteúdo de um texto utilizando o PLN, além de ser possível extrair entidades, detectar sinônimos, fazer a correção de palavras escritas de maneira incorreta e evitar a ambiguidade. Existem diversas aplicabilidades, como por exemplo: tradução de idiomas, conversação computacional, acesso a informações e análise de sentimentos. (AYTEKIN; SUTCU, 2019).

Wilson *et al.* (2014) define que a Análise de Sentimentos tem o intuito de julgar a polaridade de uma expressão como sendo positiva, negativa ou, caso não seja possível detectar sentimentos, neutra. Além disso, é considerada por Bhadane *et al.* (2015) como uma parte do Processamento de Linguagem Natural cujo objetivo é detectar elementos subjetivos no texto. Ding et al. (2009) afirma que a análise de sentimentos está sendo foco em diversos estudos e análises no âmbito acadêmico, mas também vêm sendo bastante utilizadas pela indústria. (ROSA, 2015). Para Liu (2012) a análise de sentimentos está diretamente relacionada com as mídias sociais.

Com o surgimento datado em meados da década de 1990, as redes sociais virtuais vêm sendo cada vez mais utilizadas com o decorrer do tempo. Esse crescimento casa com um maior número de pessoas tendo mais facilidade ao acesso à internet devido a políticas de inclusão digital ou por meio do avanço tecnológico. Atualmente existem diversos ambientes de interação social online onde as pessoas conseguem compartilhar notícias, sentimentos, opiniões, interesses profissionais, produtos, serviços, fotos, vídeos e mais uma infinidade de funções (CORRÊA, 2017).

O Twitter é uma rede social que, de acordo com Aslam (2022), surgiu em 2006 com o intuito de permitir que os usuários pudessem compartilhar textos curtos - inicialmente limitados a 140 caracteres e posteriormente aumentado para 280 – que são comumente chamados de tweets. Acabou se tornando popular por se tratar de uma ideia simples, onde qualquer pessoa pode compartilhar facilmente e de maneira bem rápida qualquer notícia, ideia ou opinião sobre os mais variados assuntos. Além disso, segundo Aslam (2022), a rede conta com mais de 200 milhões de usuários ativos, os quais, por sua vez, acabam gerando uma quantidade enorme de informações que podem ser utilizadas para diversas finalidades, tais como: recomendação de produtos através dos dados de navegação dos usuários, avaliação da satisfação dos clientes de uma determinada marca ou entendimento de opiniões de cunho político. Para Ghisi *et al.* (2016) a Black Friday é considerada um dos dias mais importantes para o varejo mundial. Uma consequência disso é que diversos tweets são gerados antes, durante e, até mesmo, depois desta data.

Com origem nos Estados Unidos, a Black Friday é a sexta-feira que sucede o Dia de Ação de Graças no país e é considerada o marco inicial das vendas de Natal. Durante este dia os consumidores contam com diversos descontos e promoções realizadas pelas lojas para comprar os produtos desejados de maneira mais acessível. Tal data se tornou referência nos Estados Unidos, pois é considerada o dia com o maior volume de compras por lá, o que acabou chamando a atenção de diversos lojistas espalhados pelo mundo. No Brasil, a Black Friday teve sua estreia em 2010 e, a partir de então, acabou se consolidando como uma data importante para o comércio no país, alcançando um faturamento referente a R\$1,9 bilhão em 2016 (GHISI *et al.*, 2016).

Em sintonia com o avanço tecnológico, as empresas vêm cada vez mais utilizando as redes sociais para entender o comportamento dos consumidores, assim como suas sugestões, críticas e sentimentos a respeito de um determinado produto ou serviço. Neste contexto, este trabalho visa, a partir de publicações na rede social Twitter, mensurar quais são as perspectivas dos consumidores em relação à Black Friday no Brasil em 2022, utilizando-se das técnicas de aprendizado de máquinas como a análise de sentimentos.

Material e Métodos

Para a realização deste trabalho, algumas etapas foram cumpridas antes da obtenção do resultado. Entre elas, destacam-se a geração do dataset através de web scrapping, limpeza e preparação dos dados, criação do modelo de aprendizado de máquina voltado a análise de

sentimentos e, por fim, sua validação. A Figura 1 demonstra os processos que foram realizados.

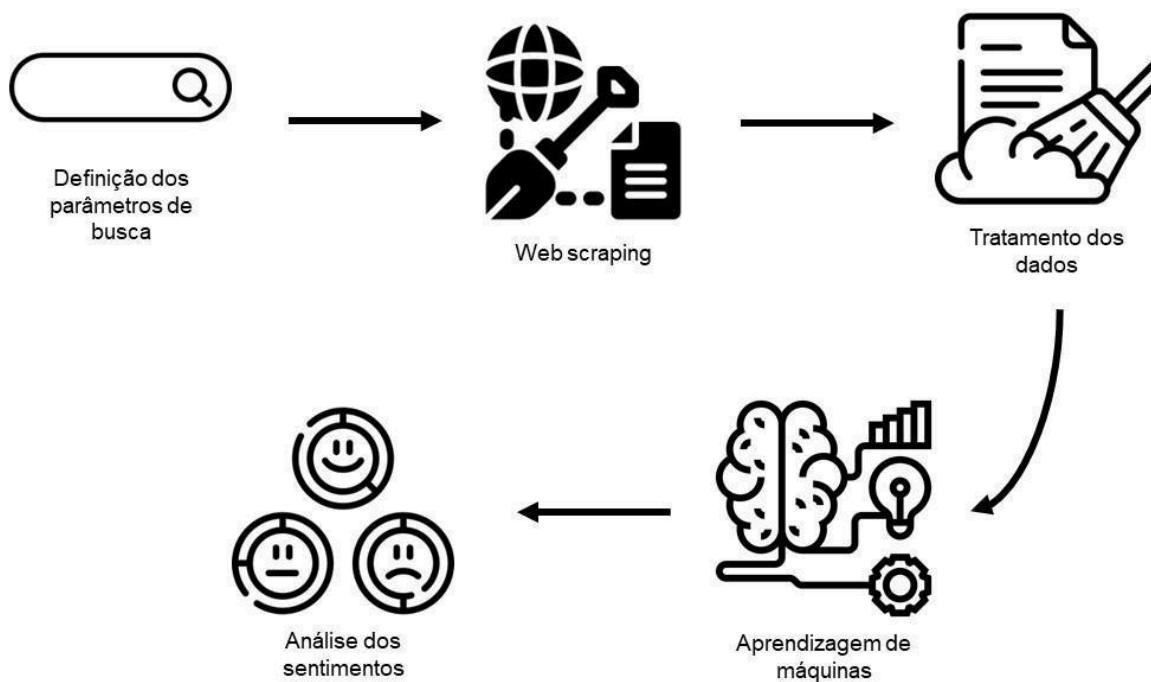


Figura 1. Fluxograma do trabalho
Fonte: Dados originais da pesquisa

Antes de gerar o dataset, foi preciso estabelecer alguns parâmetros que seriam necessários mais para frente quando fosse realizado a busca pelos tweets. O primeiro deles foi o termo de busca, ou seja, qual deveria ser o texto procurado na barra de pesquisa do Twitter de modo que o resultado da busca fosse pertinente com o tema do trabalho. Então, definiu-se que o “black friday” seria o termo de busca, fazendo com que o Twitter mostrasse apenas tweets que continham essas palavras em sequência no corpo do texto. Em seguida, optou-se por selecionar apenas os tweets redigidos em português. O último parâmetro a ser selecionado foi o período em que esses tweets foram publicados. Neste caso, ficou definido que deveriam ser aqueles cuja data de publicação foi entre os dias 18 nov. 2022 e 02 dez. 2022, pois desta maneira conseguia-se ter também uma avaliação da semana anterior e posterior da data da Black Friday (25 nov. 2022).

Com os parâmetros já estabelecidos foi preciso encontrar uma maneira de coletar esses tweets, o que poderia ser algo extremamente desafiador e trabalhoso. Neste contexto, web scraping e crawlers surgiram como opções para realizar essa tarefa de forma automatizada. De modo genérico, Neil (2016) afirma que um software capaz de retirar informações de sites e que seja capaz de imitar o comportamento humano de navegar na web - seja por protocolo

HTTP ou por ações automatizadas - pode ser definido como um sistema de web scraping. Paralelamente, Olston e Najork (2010) pontuam que os crawlers (conhecidos também como robôs ou spiders) podem ser usados para diversas finalidades, sendo sua principal função permitir que os usuários realizem buscas no índice das páginas da internet, de modo que sejam encontrados elementos específicos correspondentes à consulta. Para aplicar as técnicas de web scraping no contexto do trabalho, fez-se o uso da linguagem de programação Python.

Segundo McKinney (2017), em 1991 foi criada a linguagem de programação denominada Python, e desde então vem conquistando cada vez mais espaço, tornando-se uma das mais populares hoje em dia e, conseqüentemente, sendo bastante utilizada em âmbitos industriais e acadêmicos. Existem diversas bibliotecas disponíveis para serem utilizadas no Python. Optou-se por utilizar a biblioteca Selenium para automatizar o processo de entrar no navegador da web, ir ao site do Twitter, realizar a busca dos tweets utilizando os parâmetros previamente definidos e coletar os tweets através de crawlers. A partir destas ferramentas, foi possível gerar uma base com os dados brutos.

Os dados brutos precisavam, na sequência, passar por uma série de tratamentos, com a finalidade de estarem prontos para a realização da análise de sentimentos. Para tratá-los, foi novamente utilizada a linguagem de programação Python, só que desta vez fez-se o uso de uma biblioteca chamada de NLTK. O Natural Language Toolkit (NLTK) permite que os usuários trabalhem com os dados de linguagem humana, fornecendo interfaces de simples utilização e diversos recursos lexicais, além de possibilitar o processamento de textos para classificação, tokenização, lematização, marcação análise e raciocínio semântico. Através desta biblioteca foi possível padronizar os textos dos tweets.

Primeiramente foram excluídos da base os tweets de empresas divulgando promoções de produtos. Em seguida, os links de sites ou qualquer outro tipo de hiperlink foram removidos dos textos, assim como as datas, valores monetários, números e as palavras “black” e “friday”, as quais foram utilizadas no termo de busca dos tweets. Na sequência, foi necessário realizar um tratamento em caracteres especiais e nos excessos de espaços entre as palavras, eliminando-os dos textos em ambos os casos. Removeu-se, também, as stopwords, que são palavras que podem ser suprimidas, ocultadas ou omitidas de frases sem alterar completamente o sentido do que quer ser escrito. Geralmente as stopwords são compostas por artigos, conjunções e preposições. Alguns exemplos de stopwords são palavras como “as”, “os”, “uns”, “de”, “para”, “com” e “por”. Por fim, os verbos presentes nos tweets foram ajustados para o infinitivo.

Com isso, foi possível obter um dataset contendo todos os tweets padronizados e prontos para serem aplicados no modelo de análise de sentimentos. Alguns exemplos dos tweets em seus formatos originais e após os tratamentos podem ser verificados na Tabela 1.

Tabela 1. Tweets originais e pós-tratamento

Tweet original	Data	Tweet após tratamento
Essa Black friday é mó farsa, tô acompanhando um fone faz mais de mês e agora eles estão anunciando como promoção, mas não mudou o valor de nadaaaa	18/11/2022	farsa acompanhar fone Mes estao anunciar promocao nao mudar nadaaaa
Que diabo de black friday é essa que não baixa o preço das coisas	19/11/2022	diabo nao baixo preco coisa
A Black Friday no Brasil é o golpe!	20/11/2022	Brasil golpe
Se tem uma coisa que eu não acredito no Brasil, é black friday	22/11/2022	nao acreditar Brasil
E olha que não comprei nada na Black Friday	23/11/2022	olhar nao comprar
maldita black friday acabando com a minha felicidade	24/11/2022	maldita acabar felicidade
Ninguém falou de blackfriday hoje ne	25/11/2022	ninguem falar hoje
Gente que black friday mais mixuruca é essa dessa ano? Não vi nada de promoção	25/11/2022	gente mixuruca ano nao promocao
queria ter sido mais responsável durante o ano pra ter aproveitado melhor essa black friday	27/11/2022	querer ser responsavel durante ano pra aproveitar bom
Nessa black friday comprei mais produtos, porém de preços menores, então foi bom.	29/11/2022	comprar produto por preco pequeno entao bom
minha maior decepção é não ter comprado um livro nessa black friday	30/11/2022	decepcao nao comprar livro
E cá entre nós, 10, 20% de desconto? Isso pra mim não é nada ...fora que o que eu queria comprar não entrou pra Black friday kkkk	01/12/2022	desconto pra eu nao querer comprar nao entrar pra kkkk
Perfume que comprei na Black Friday ainda não chegou :/	02/12/2022	perfume comprar nao chegar

Fonte: Dados originais da pesquisa

Além disso, foi necessário gerar uma segunda base de dados, a qual seria usada para treinar e testar o modelo de análise de sentimentos gerado em seguida. Buscou-se uma base pública e disponível na internet, que continha textos em português relacionados ao tema da Black Friday e que já estivessem avaliados ou fosse possível inferir o sentimento positivo ou negativo. Portanto foi escolhida uma base em que diversos clientes de uma empresa de varejos digitais realizaram revisões de produtos e serviços ofertados em 2018, escrevendo textos (inclusive sobre a Black Friday) e dando uma nota de satisfação, a qual varia de 1 a 5, sendo 1 a pior nota e 5 a melhor. Para os fins deste trabalho, foram considerados apenas os textos cuja avaliação possuía as notas 1, 2 e 5, definindo-se as notas 1 e 2 como sentimento negativo e nota 5 como sentimento positivo.

Para normalizar a base, foi novamente utilizada a linguagem Python para selecionar aleatoriamente quarenta mil textos, onde metade era formada por textos de cunho positivo e a outra metade de textos considerados negativos. Por fim, essa base normalizada também passou pelas etapas anteriores de tratamento. A Tabela 2 demonstra alguns exemplos dos textos e notas de avaliação dos clientes, assim como o sentimento inferido e o texto após o tratamento dos dados.

Tabela 2. Base de dados para testar o modelo antes e após o tratamento

Texto	Nota	Sentimento Inferido	Texto após tratamento
Melhor relação custo-benefício que encontrei em janeiro/2018.	5	Positivo	bom relacao custo beneficio encontrei janeiro
Fiz a compra na Black Friday de 2017 e até hoje (08/01/2018) ainda não recebi o produto.	1	Negativo	fiz compra ate hoje nao receber produto
Produto atendeu as expectativas, a entrega foi realizada antes do prazo, requisito que também conta muito para a minha avaliação, por isso merece 5 estrelas.	5	Positivo	produto atender expectativa entrega realizar prazo requisito tambem contar avaliacao merecer estrela
Não vale o que se paga, shampoo resseca o cabelo e o	1	Negativo	nao vale pagar shampoo resseca cabelo

condicionador deixa super embaraçado		condicionador deixar super embaracar
a tv é muito boa. A qualidade de som e de imagem é excelente.	5 Positivo	qualidade som imagem excelente
Demora resolver um probleminha oito dias, descaso com cliente.	2 Negativo	demorar resolver probleminho dia descaso cliente

Fonte: Dados originais da pesquisa

Com ambas as bases geradas e devidamente tratadas, foram realizadas mais algumas etapas aos textos presentes nelas, com o intuito de transformar os textos em informações numéricas, tornando-os aptos a serem aplicados no modelo de aprendizado de máquina. Para isso, foram aplicados os conceitos tanto do modelo Bag of Words quanto o Term Frequency-Inverse Document Frequency (TF-IDF). O modelo Bag of Words, segundo Brownlee (2017), é bastante utilizado em PLN e tem o objetivo de mensurar a frequência que cada palavra foi utilizada em um texto, sem levar em conta a ordem e a estrutura dele. Já o TF-IDF, para Manning et al. (2008), ao contrário do Bag of Words, impõe um peso às palavras de acordo com a importância que elas têm para o documento. Para obter uma matriz com o peso de cada palavra em cada um dos tweets foi necessário utilizar uma biblioteca do Python chamada de Scikit-learn, a qual possui uma gama diversa de algoritmos de aprendizado de máquina. A Figura 2 detalha o funcionamento do conceito Bag of Words.

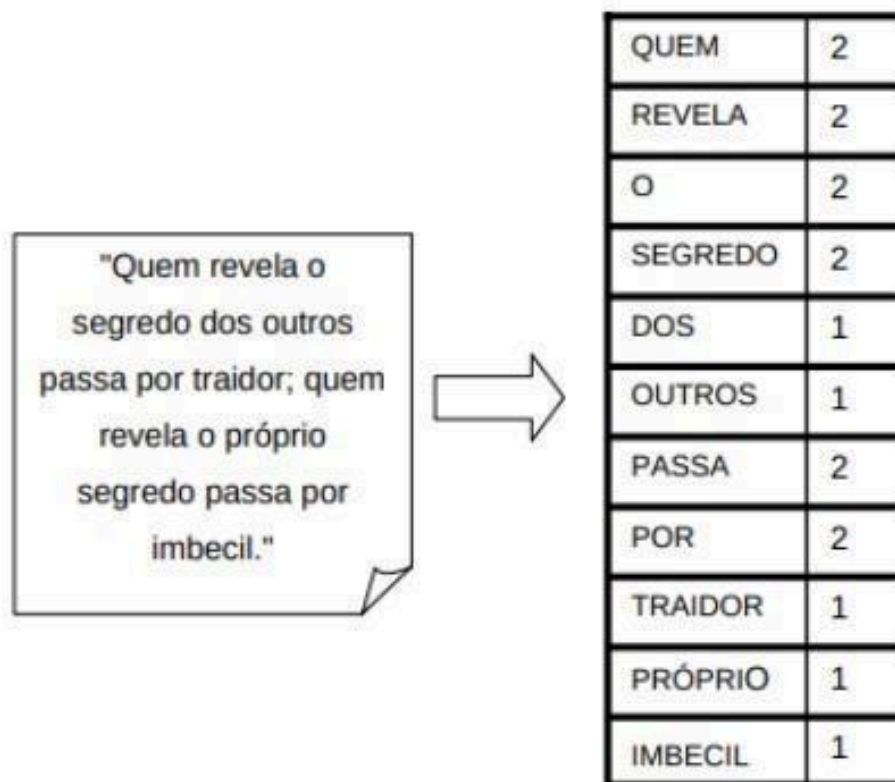


Figura 2. Exemplo do conceito Bag of Words
Fonte: Miranda (2009)

Em seguida foi elaborado o modelo de machine learning. Silva (2020) afirma que aprendizado de máquina é um ramo da inteligência artificial onde os sistemas podem aprender com dados, identificar padrões, classificar e tomar decisões com uma mínima intervenção humana. Neste trabalho, foi utilizado um aprendizado de máquina supervisionado, onde Carvalho e Padilha (2017) assentam que um algoritmo de classificação irá generalizar as informações contidas no conjunto de treinamento classificando, posteriormente, outros objetos cujas características sejam desconhecidas. Nesse sentido, foi utilizada a base com os textos já classificados como positivo ou negativo para treinar o modelo com a finalidade que o próprio consiga, na sequência, classificar os tweets coletados a respeito da Black Friday no Brasil em 2022.

Novamente foi utilizada a biblioteca Scikit-learn do Python para a criação do modelo. Para a realização deste trabalho selecionou-se o algoritmo de classificação de Naive Bayes, o qual, de acordo com Ribeiro (2015), é um classificador que possui uma abordagem probabilística baseada no Teorema de Bayes, além de assumir que as variáveis não são dependentes entre si. Conforme corrobora Amaral (2016), durante a etapa de treinamento do modelo, o algoritmo estabelecerá uma tabela de valores, assim como atribuir pesos para cada

atributo contido em uma classe de classificação, de modo que quando uma nova instância for submetida, o modelo realizará a soma dos pesos de cada atributo em cada uma das classes, sendo que a classe que obtiver a maior soma dos pesos será considerada a classe classificadora do item. Corrêa (2017) afirma que tal algoritmo tornou-se popular entre aqueles que são utilizados para categorizar textos com base nas frequências das palavras devido ao desempenho computacional ser relativamente melhor que os demais algoritmos.

Então o dataset com os textos previamente classificados, após ter sido devidamente tratado e os textos convertidos em matrizes numéricas através do método TF-IDF, foi dividido aleatoriamente onde 80% dele serviu de treinamento para o modelo enquanto os 20% restantes foi usado para testá-lo. Por fim, com o modelo já treinado e validado conseguiu-se analisar os sentimentos dos tweets coletados, classificando-os como positivo ou negativo.

Resultados e Discussão

Os resultados obtidos e demonstrados nessa seção irão passar pela seguinte ordem: primeiro serão demonstrados os resultados da análise exploratória dos dados, seguido da matriz de confusão obtida através do treinamento e validação do modelo de machine learning e por fim serão abordadas as análises dos tweets já classificados pelo modelo.

O crawler levou cerca de quatro horas e coletou um total de 29891 tweets. Constatou-se que havia 6977 tweets que continham propagandas, informações de preço e divulgações de cupons de desconto presentes no corpo do texto. Como o intuito do trabalho é avaliar o sentimento dos consumidores durante a Black Friday, optou-se por remover esses tweets da base, a fim de evitar que eles interfiram no resultado. Após a exclusão destes, a base passou a possuir um total de 22914 tweets. Em média, os tweets possuem 104 caracteres, sendo que alguns atingem o limite máximo de caracteres estipulados pela plataforma, ou seja, 280. Além disso, também foi possível identificar a quantidade de publicações feitas por dia, conforme mostra a Figura 3.

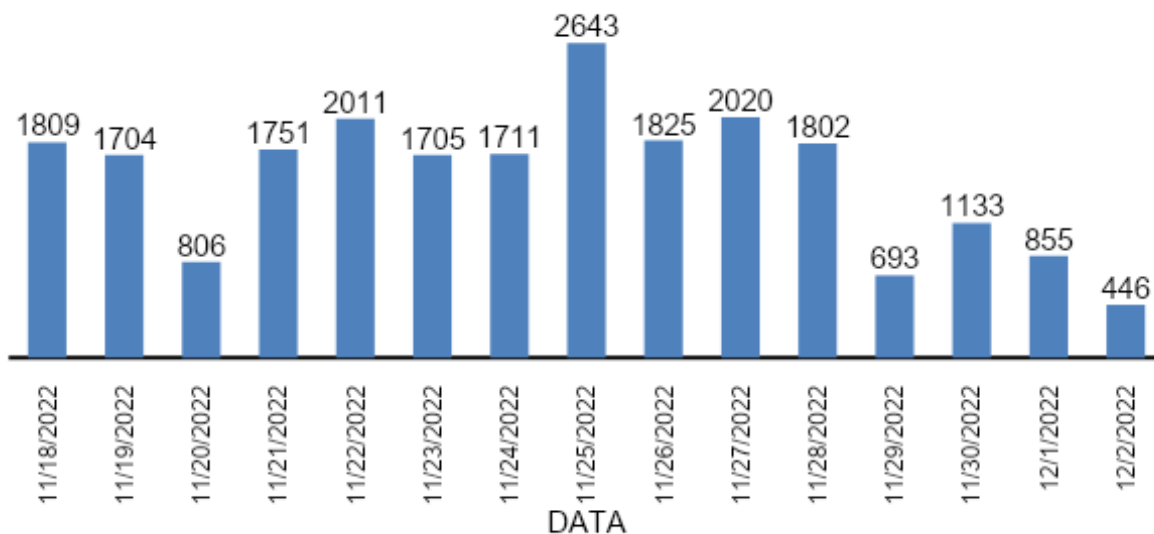


Figura 3. Quantidade de tweets coletados por data de publicação
 Fonte: Resultados originais da pesquisa

A partir da Figura 3 percebe-se que o dia com mais publicação na rede social Twitter, a respeito da Black Friday, foi justamente o dia do evento. Também é possível notar que a quantidade de textos nos sete dias anteriores é superior a quantidade de textos nos sete dias seguintes à data do evento, conforme evidencia a Figura 4.

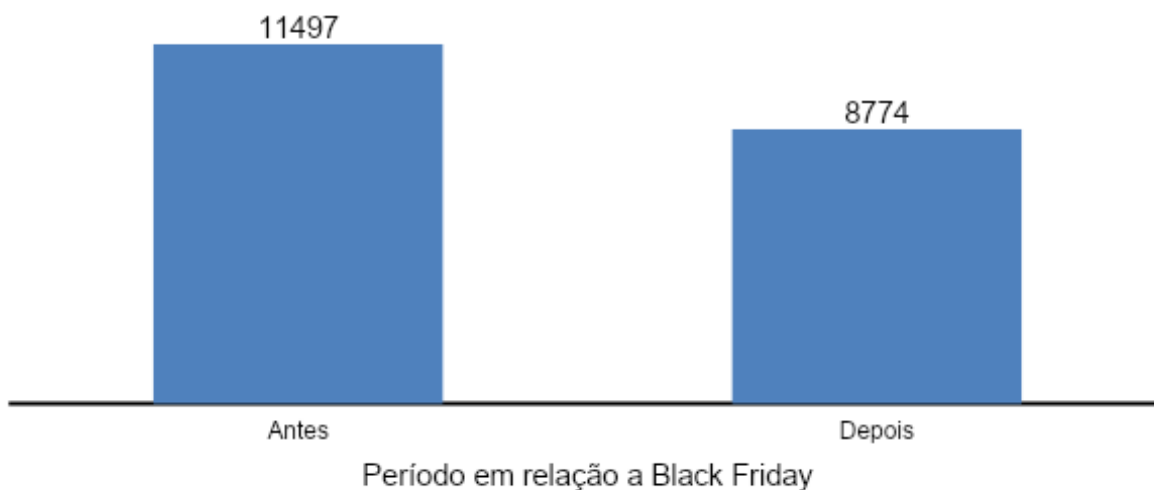


Figura 4. Quantidade de tweets antes e depois da Black Friday
 Fonte: Resultados originais da pesquisa

É possível que isso esteja relacionado com o fato de que, antes da Black Friday, os consumidores ficaram ansiosos com as prováveis promoções que estariam por vir ou criaram expectativas e teorias para a data. No entanto, após o evento, por já saberem de tudo que

aconteceu no dia, esse tópico entrou em declínio nessa rede social. Isso fica mais evidente na Tabela 3, que demonstra as palavras mais redigidas nos tweets. Nela, observa-se que houve uma redução na frequência de utilização das palavras “desconto” e “promoção” quando se compara os períodos anterior e posterior a data da Black Friday.

Tabela 3. Contagem de palavras nos tweets

Palavra	Contagem Total	Contagem Antes	Contagem Depois
nao	5147	2395	2101
comprar	4173	1873	1787
pra	4118	2191	1460
desconto	1787	1050	577
promocao	1703	912	604
aproveitar	1582	783	650
ano	1494	689	633
ir	1482	905	455
dia	1475	720	603
ate	1280	697	474
chegar	1220	533	616
livro	1120	553	457
voce	1084	626	387
real	1044	489	418
querer	1027	531	354

Fonte: Resultados originais da pesquisa

Constata-se, também, que mesmo após realizar a remoção das stopwords a palavra “pra” continuou aparecendo com frequência. Isso indica que mesmo com os pacotes da linguagem de programação Python estarem em constante evolução, alguns erros simples ainda estão susceptíveis de ocorrerem, além de evidenciar ainda mais como nossa língua possui diversas variações para a mesma finalidade, dificultando ainda mais essa melhoria na tecnologia.

Após separar a base já classificada em treinamento e aplicá-la ao modelo de aprendizado de máquina e em seguida testá-lo, utilizando os 20% restantes da mesma, obteve-se a matriz de confusão mostrada na Figura 5.

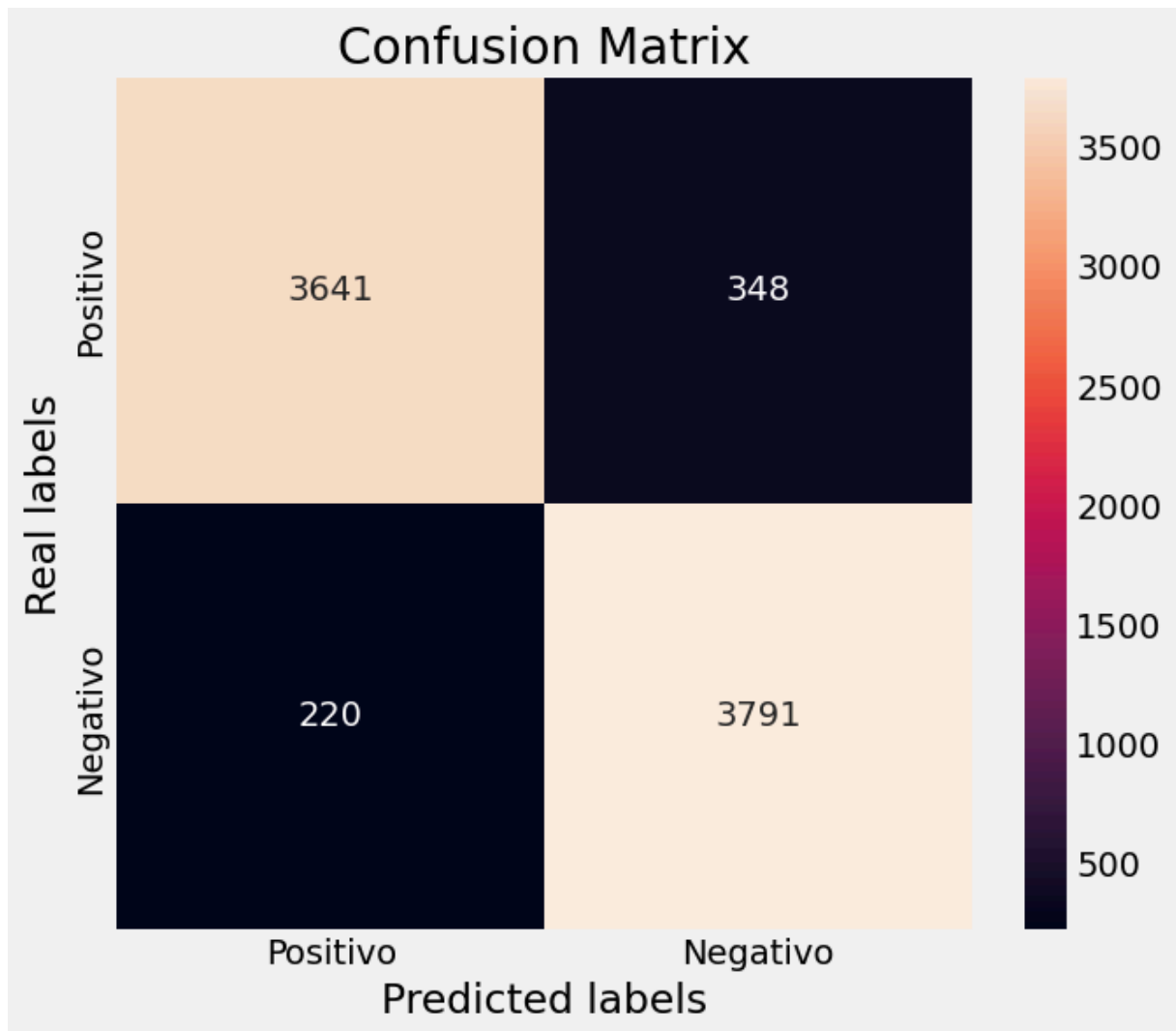


Figura 5. Matriz de confusão do modelo
 Fonte: Resultados originais da pesquisa

A partir da matriz de confusão é possível observar que o modelo possui uma acurácia de 92,9%. Dos 8000 textos que foram selecionados aleatoriamente para testar o algoritmo, 3641 eram classificados como positivo e o modelo conseguiu prever corretamente, enquanto 348 deveriam ser classificados como positivo, mas o resultado obtido foi negativo. Da mesma forma, 3791 comentários possuíam um cunho negativo e o modelo os classificou precisamente, enquanto apenas 220 foram classificados como positivos quando o correto deveria ser o contrário. Além disso, fica evidente que o algoritmo teve mais acertos para frases com sentimento negativo e mais erros ao avaliar frases positivas, levando-o a possuir uma especificidade de 94% enquanto a sensibilidade do modelo foi de 92%.

Por fim, foi passada a base com os tweets sobre a Black Friday para que o modelo já treinado pudesse avaliar tweet por tweet e classificá-los como positivo ou negativo. Alguns exemplos dos sentimentos avaliados podem ser visualizados na Tabela 4.

Tabela 4. Exemplos de tweets classificados pelo modelo

Tweet	Data	Sentimento
eu to ha 1 mes acompanhando preco de tenis e AUMENTOU MUITO nesses 3 ultimos dias pq entramos na semana da Black Friday vcs me enganaram	18/11/2022	Negativo
Esse mouse e incrivel, so compro em epoca de black friday, meu ultimo foi em 2019 por 99, maravilhoso custo beneficio	19/11/2022	Positivo
To esperando uma promocao top de black friday	20/11/2022	Positivo
eu odeio black friday pq nunca tenho dinheiro p comprar coisa	21/11/2022	Negativo
gente essa semana black friday ta devagar quase parando cade os cupons	22/11/2022	Negativo
uma das minhas unicas comprinhas da black friday chegou to chocada ainda o quanto valeu a pena	23/11/2022	Positivo
essa black friday ta fraca	24/11/2022	Negativo
A black Friday mais decepcionante que teve	25/11/2022	Negativo
muito feliz com minha aquisicao da black friday	26/11/2022	Positivo
Essa foi a primeira Black Friday que aproveitei e que valeu a pena	27/11/2022	Positivo
Cei nao hem propaganda enganosa da Black Friday	28/11/2022	Negativo
meudeus eu AMO BLACK FRIDAY	29/11/2022	Positivo
Foi a pior blackfriday q eu ja vi	30/11/2022	Negativo
nossa black fraude engana trouxa cada ano que passa pior! #BlackFriday de araque	01/12/2022	Negativo
Eu te amo (um pouco), Black Friday	02/12/2022	Positivo

Fonte: Resultados originais da pesquisa

Realizado a contagem de tweets positivos e negativos em toda a base, obteve-se o resultado mostrado na Figura 6.

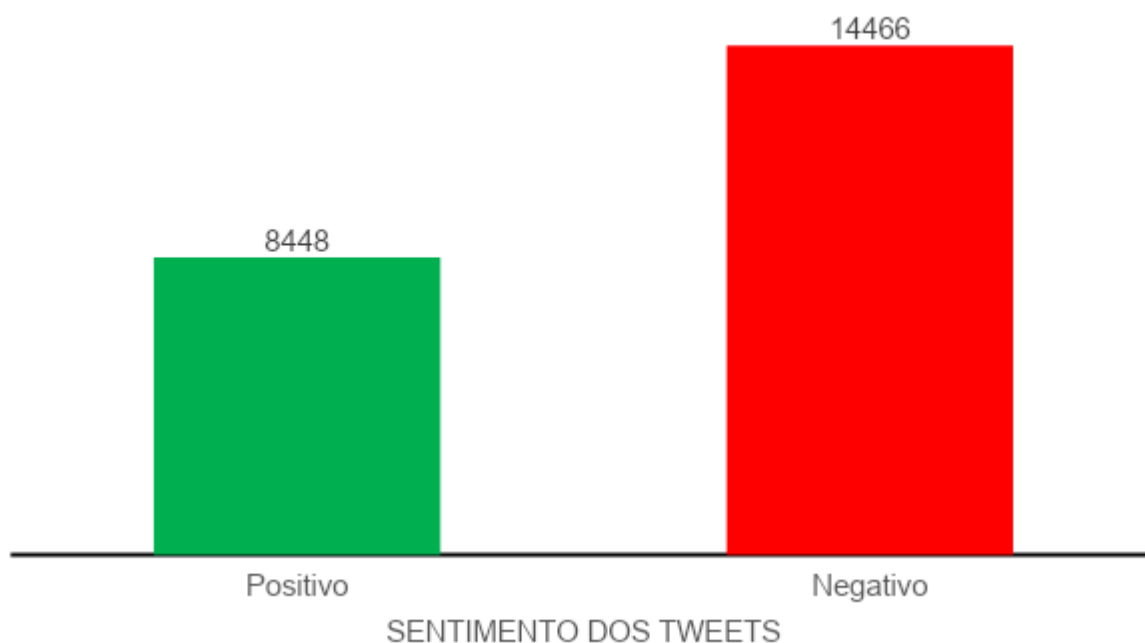


Figura 6. Sentimento dos tweets
Fonte: Resultados originais da pesquisa

Percebe-se que a quantidade de tweets classificados como positivos foi menor que aqueles classificados como negativos. Ou seja, apenas 36% dos tweets foram classificados como positivo de acordo com as palavras presentes no texto e sua frequência de utilização. Mesmo realizando uma separação por dia da publicação do tweet, constatou-se a predominância de tweets classificados como negativo, conforme evidencia a Figura 7.

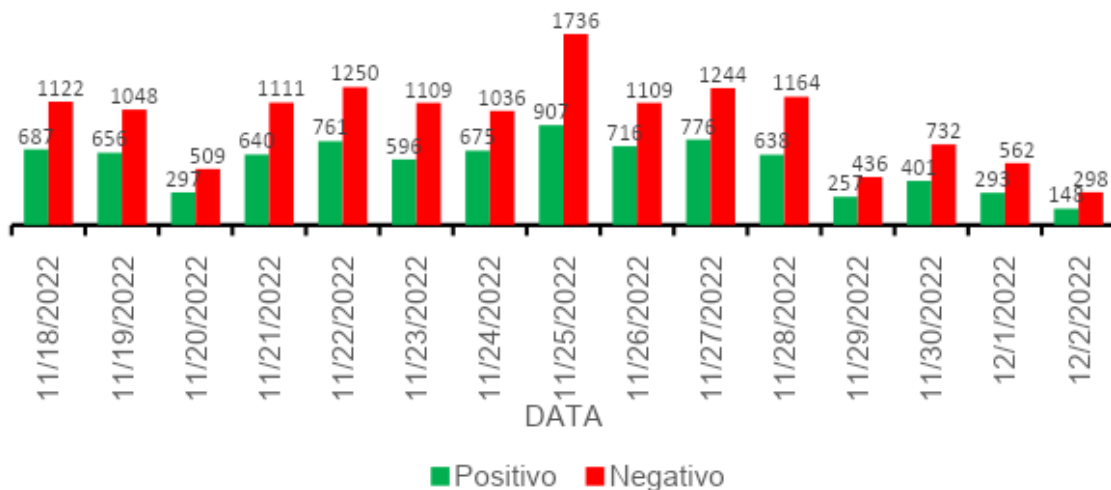


Figura 7. Sentimento dos tweets por data de publicação
 Fonte: Resultados originais da pesquisa

Novamente, o pico de avaliações positivas e negativas foi no dia do evento. Porém fica claro que a quantidade de tweets com sentimento negativo nessa data foi quase o dobro. Isso pode demonstrar uma possível insatisfação dos consumidores com a Black Friday. Além disso, constata-se que a quantidade de comentários classificados como positivo diminuiu quando comparamos o período anterior com o período posterior ao do evento em questão. De acordo com a Figura 8, nota-se que houve uma diminuição de 25% desses tweets. Essa queda pode ser explicada pela diminuição na quantidade de tweets coletados nesses períodos, conforme mostrado anteriormente.

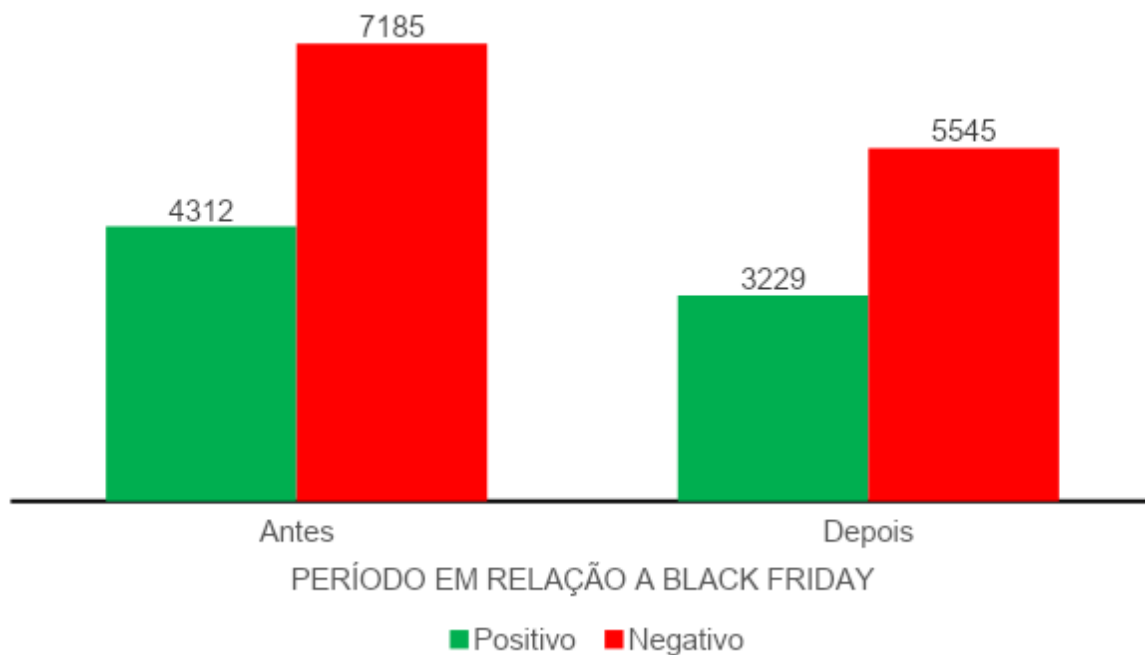


Figura 8. Tweets classificados de acordo com o período em relação a Black Friday
 Fonte: Resultados originais da pesquisa

Por fim, ressalta-se que foram observadas diversas classificações errôneas do modelo devido a ironias e sarcasmos presentes nos textos. O algoritmo utilizado neste trabalho não foi treinado para lidar com ironias, muitas vezes presentes nas redes sociais, sendo capaz apenas de impor peso nas palavras presentes no texto, de acordo com a metodologia TF-IDF, classificando-o em positivo ou negativo. Outro elemento que pode afetar a análise final é que neste trabalho os tweets foram classificados apenas como positivo ou negativo, entretanto nem todos os tweets realmente expressam esses sentimentos. Em alguns casos há apenas algum comentário de polaridade neutra ou ambígua que não cabem em nenhuma das duas opções, mas como o modelo foi treinado apenas para classificar os tweets em positivo ou negativo, esses casos também o foram. A Tabela 5 destaca algumas ocorrências que foram classificadas de maneira errada.

Tabela 5. Classificações precipitadas

Tweet	Data	Sentimento
O triste dessa Black Friday e que nao sou rico	18/11/2022	Negativo

quase que eu cometi uma loucura financeira por conta da black friday	18/11/2022	Negativo
Acabei de cair na pegadinha de dar uma passada no shopping em Black Friday pra dar so uma olhada e literalmente estourei meu cartao	22/11/2022	Negativo
Nossa bem q o dinheiro podia cair na minha conta a tempo da black friday ne	23/11/2022	Negativo
Expectativa: juntar dinheiro o ano todo pra comprar um monte de coisa na Black Friday Realidade: nao juntar dinheiro e se endividar na Black Friday	25/11/2022	Negativo
Unica coisa que faco esses ultimos 3 dias e gastar comprando BLackFriday hauahaua	27/11/2022	Negativo
GARANTI A MINHA NA BLACK FRIDAY. NUNCA FUI TRISTE	02/12/2022	Negativo

Fonte: Resultados originais da pesquisa

Como esperado, há imperfeições no modelo e devido a isso os resultados obtidos devem ser interpretados apenas com cunho indicativo e não como verdade absoluta. Outrossim, há limitações também no corpus de treinamento, pois pode haver palavras ou expressões que não foram identificadas na etapa de classificação dos textos, além de erros de digitação que podem não ter sido reconhecidos pelo modelo de machine learning.

Conclusões

Este trabalho teve como objetivo analisar as perspectivas da Black Friday no Brasil em 2022 através de publicações na rede social Twitter. Nesse contexto foi gerada uma base contendo tweets a respeito da Black Friday para terem os sentimentos analisados e classificados como positivo ou negativo. É notório que diversos tweets foram classificados de maneira incorreta. Isso pode ter ocorrido devido a base selecionada para treinamento do modelo ser sobre avaliação de produtos de maneira geral, ao invés de ser composta por comentários exclusivos sobre a Black Friday ou possuir avaliações restritas a esse período. Outro ponto que deve ter causado tais equívocos é a própria linguagem escrita nas redes sociais, que contém muita ironia, sarcasmos, ambiguidade e erros de digitação, os quais o modelo de análise de sentimentos desenvolvido neste trabalho não é capaz de reconhecer. Não obstante, foi possível ter uma ideia de como se comportam os usuários do Twitter no Brasil durante o período da Black Friday.

Através dos resultados obtidos, constatou-se que a maioria dos comentários redigidos na rede social em questão são de cunho negativo. Além disso, é importante ressaltar que o período que antecede a Black Friday é acompanhado tanto de excitação pela chegada da data quanto pela frustração que ocorre quando as empresas divulgam as ofertas desse período. Isso fica evidente quando se observa que as palavras “promoção” e “desconto” são bastante utilizadas nesses dias que antecedem o evento. Ademais, o abandono de engajamento do tópico torna-se evidente à medida que os dias vão passando após a realização da Black Friday, ou seja, o número de comentários a respeito da Black Friday foi diminuindo gradativamente devido ao término das promoções e início das campanhas de Natal por parte das empresas.

Como sugestão para um trabalho futuro é necessário que seja elaborado um algoritmo capaz de interpretar emojis – muito utilizados nas redes sociais hoje em dia para expressar sentimentos, todas as variações linguísticas presentes na língua portuguesa, além de incluir a polaridade neutra na análise dos sentimentos. Também é válido testar outros modelos classificadores além do Naive Bayes que foi implementado neste trabalho, a fim de verificar se a acurácia deles é melhor ou pior. Pode-se realizar, por exemplo, a mesma análise utilizando Regressão Logística, Support Vector Machine (SVM) ou um modelo baseado em Floresta Aleatória.

Referências

Amaral, F. (2016). *Aprenda Mineração de Dados: Teoria e Prática.* 1ed. Alta Books, Rio de Janeiro, RJ, Brasil.

Aslam, S. (2022). Twitter by the Numbers: Stats, Demographics & Fun Facts. Disponível em: <<https://www.omnicoreagency.com/twitter-statistics/>>. Acesso em: 07 ago. 2022.

Aytekin, C.; Sutcu, C. S. (2019). An Example of Pragmatic Analysis in Natural Language Processing: Sentimental Analysis of Movie Reviews. In: Communication and Technology Congress, 2019, Istambul, Turquia.

Bhadane, C.; Dalal, H.; Doshi, H. (2015). Sentiment analysis: Measuring opinions. p. 808-814 In: Vasudevan, H.; Joshi, A. R.; Shekokar, N. M. *Procedia Computer Science*, v.45. Elsevier, 2015, Amsterdã, Holanda.

Brownlee, J. (2017). A Gentle Introduction to the Bag-of-Words Model. Disponível em: <<https://machinelearningmastery.com/gentle-introduction-bag-words-model/>>. Acesso em: 11 jan. 2023.

Chowdhury, G. G. (2003). Natural language processing. p. 51-89. In: *Annual Review of Information Science and Technology*, v.37. Wiley, Nova Iorque, Estados Unidos da América.

Corrêa, I.T. (2017). Análise dos sentimentos expressos na rede social Twitter em relação aos filmes indicados ao Oscar 2017. *Monografia de Bacharelado em Sistemas de Informação.* Universidade Federal de Uberlândia, Uberlândia, MG, Brasil.

Delen, D.; Sharda, R.; Turban, E. (2019). *Business Intelligence e Análise de Dados para Gestão do Negócio.* 4ed. Bookman, Porto Alegre, RS, Brasil.

Ding, X.; Liu, B.; Zhang, L. (2009). Entity discovery and assignment for opinion mining applications. In: *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2009, Nova Iorque, NW, Estados Unidos da América. Anais... p. 1125-1134.

Ghisi, M. A.; Guerra, A. A. C.; Nielsen, F. A. G. (2017). Black Friday no Brasil: Aspectos positivos e negativos desta ação promocional para o varejo brasileiro. In: *XX SemeAd*, 2017, São Paulo, SP, Brasil. Anais... p. 3-4.

Hoffmann, P.; Wiebe, J.; Wilson, T. (2005). Recognizing contextual polarity in phrase level sentiment analysis. p. 347-354. In: Mooney, R.; Brew, C.; Chien, L. F.; Kirchoff, K. *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing.* Association for Computational Linguistics, 2005, Vancouver, British Columbia, Canadá.

Liu, B. (2012). *Sentiment analysis and opinion mining.* p. 1-167. In: Hirst, G. *Synthesis Lectures on Human Language Technologies.* Morgan & Claypool Publishers, 2012, Toronto, Canadá.

Manning, C. D.; Raghavan, P.; Schütze, H. (2008). *Introduction to Information Retrieval.* 1ed. Cambridge University Press, Cambridge, Reino Unido.

McKinney, W. (2017). *Python for Data Analysis: Data Wrangling with Pandas, Numpy, and IPython.* 2ed. O'Reilly Media, Sebastopol, Califórnia, Estados Unidos da América.

Miranda, A. R. O. (2009). Descoberta de conhecimento em texto aplicada a um sistema de atendimento aos usuários de um plano de assistência à saúde. *Dissertação de Mestrado em Engenharia Civil.* Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brasil.

Najork, M.; Olston, C. (2010). Web Crawling. p. 175-246. In: Kelly, D.; Castells, P.; Liu, Y. *Foundations and Trends in Information Retrieval*, v. 4. Now, 2023, Hanover, Massachusetts, Estados Unidos da América.

Neil, Y. (2016). Web Scrapping the Easy Way. *Tese para obtenção de créditos em Sistemas de Informação.* Georgia Southern University, Geórgia, Estados Unidos da América.

Padilha, V. A.; Carvalho, A. C. P. L. F. (2017). *Mineração de dados em Python.* 1ed. Instituto de Ciências Matemáticas e de Computação – Universidade de São Paulo, São Paulo, SP, Brasil.

Ribeiro, L. B. (2015). Análise de sentimentos em comentários sobre aplicativos para dispositivos móveis: estudo do impacto do pré-processamento. *Monografia de Bacharelado em Ciências da Computação.* Universidade de Brasília, Brasília, DF, Brasil.

Rosa, R. L. (2015). Análise de sentimentos e afetividade de textos extraídos das redes sociais. *Tese de Doutorado em Engenharia de Computação.* Universidade de São Paulo, São Paulo, SP, Brasil.

Silva, A. R. (2020). Uma visão geral sobre Machine Learning - Classificação. Disponível em: <<https://statplace.com.br/blog/uma-visao-geral-sobre-machine-learning/>>. Acesso em: 12 jan. 2023.