

**UNIVERSIDADE FEDERAL DO AMAZONAS – UFAM**  
**INSTITUTO DE CIÊNCIAS SOCIAIS, EDUCAÇÃO E ZOOTECNIA – ICSEZ**  
**CURSO DE ADMINISTRAÇÃO**

**ARISTON TAVARES OLIVEIRA**

**MINERAÇÃO DE DADOS: UM ESTUDO DAS EMOÇÕES ACERCA DO  
FESTIVAL FOLCLÓRICO DE PARINTINS**

**PARINTINS – AM**

**2025**

**ARISTON TAVARES OLIVEIRA**

**MINERAÇÃO DE DADOS: UM ESTUDO DAS EMOÇÕES ACERCA DO  
FESTIVAL FOLCLÓRICO DE PARINTINS**

Trabalho de Conclusão de Curso apresentado a  
Universidade Federal do Amazonas – UFAM  
como requisito básico para a conclusão do  
curso de Administração.

Orientação: Prof. Dr. Pedro Marinho Amoedo.

**PARINTINS – AM**

**2025**

Ficha Catalográfica

Elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

---

O48m      Oliveira, Ariston Tavares  
            Mineração de dados: um estudo das emoções acerca do festival folclórico de parintins / Ariston Tavares Oliveira. - 2025.  
            30 f. : il., color. ; 31 cm.

            Orientador(a): Pedro Marinho Amocdo.  
            Trabalho de Conclusão de Curso (graduação) - Universidade Federal do Amazonas, Faculdade de Estudos Sociais, Curso de Administração - PROING, Parintins-AM, 2025.

            1. Análise de sentimentos. 2. Mineração de dados. 3. Festival folclórico. 4. Redes sociais. 5. Aprendizado de Máquina. I. Amocdo, Pedro Marinho. II. Universidade Federal do Amazonas. Faculdade de Estudos Sociais. Curso de Administração - PROING. III. Título

---

## SUMÁRIO

<b>RESUMO</b> .....	5
<b>INTRODUÇÃO</b> .....	6
<b>PROBLEMATIZAÇÃO</b> .....	6
<b>OBJETIVOS</b> .....	9
<b>Objetivo Geral</b> .....	9
<b>Objetivos específicos</b> .....	9
<b>REFERENCIAL TEÓRICO</b> .....	10
<b>O avanço das redes sociais</b> .....	10
<b>JUSTIFICATIVA</b> .....	20
<b>METODOLOGIA</b> .....	20
<b>OS DADOS DO ESTUDO</b> .....	20
<b>PRÉ-PROCESSAMENTO</b> .....	21
<b>CLASSIFICAÇÃO DOS SENTIMENTOS</b> .....	22
<b>RESULTADOS E DISCUSSÕES</b> .....	25
<b>ANÁLISE COM CLASSIFICADOR NAIVE BAYES</b> .....	29
<b>CONSIDERAÇÕES FINAIS</b> .....	30
<b>REFERÊNCIAS</b> .....	30

## RESUMO

Comunicações compartilhadas por usuários nas redes sociais carregam grande quantidade de informações que podem ser quantificadas e analisadas por procedimentos estatísticos apropriados. Modelos de aprendizagem de máquina, dicionários léxicos, redes neurais e entre outros, são exemplos de bons classificadores de informações. Estas técnicas, além da segurança dos seus resultados, apresentam como vantagens economia de tempo e de custos, pois os dados são capturados quase que em tempo real através de raspagens na internet, evitando que pesquisadores se desloquem a campo quando do uso de procedimentos tradicionais. Esta pesquisa tem por objetivo identificar e analisar os sentimentos e as emoções despertadas pelo festival folclórico de Parintins sobre os simpatizantes, brincantes, torcedores e turistas. Para tal, utilizou-se da técnica análise de sentimentos, em que se fez uso da análise léxica e do modelo de classificação Naive Bayes, ferramentas da aprendizagem de máquina. Para uso da classificação via Naive Bayes, o conjunto de dados foi rotulado por humanos em positivo e negativo e divididos em quatro bases consideradas (50% e 50%, 60% e 40%, 66% e 34% e 70% e 30% para treino e teste respectivamente). Os resultados via análise léxica mostram uma predominância de sentimentos positivos (67%) em relação aos negativos (33%), com substancial destaque para as emoções de confiança, alegria e surpresa. Vale observar que as emoções negativas como tristeza, medo e raiva aparecem bastante representativas para o momento festivo. O classificador Naive Bayes apresentou acurácia de 76%, um resultado mediano para um modelo classificador, sinalizando que dentre 10 textos analisados, classifica acertadamente 8.

**Palavras-chave:** Análise de Sentimentos, Mineração de Dados, Festival Folclórico, Redes Sociais, Aprendizado de Máquina, Classificação

Communications shared by users on social networks carry a large amount of information that can be quantified and analyzed using appropriate statistical procedures. Machine learning models, lexical dictionaries, neural networks, among others, are examples of effective information classifiers. These techniques, besides ensuring reliable results, offer advantages such as saving time and reducing costs, since data are captured almost in real time through web scraping, avoiding the need for researchers to go into the field when using traditional procedures.

This study aims to identify and analyze the sentiments and emotions triggered by the Parintins Folklore Festival among supporters, participants, fans, and tourists. To do so, the sentiment analysis technique was used, employing lexical analysis and the Naive Bayes classification model, both tools from machine learning.

For the Naive Bayes classification, the dataset was labeled by humans as positive or negative and divided into four considered splits (50% and 50%, 60% and 40%, 66% and 34%, and 70% and 30% for training and testing, respectively). The results from the lexical analysis show a predominance of positive sentiments (67%) in relation to negative ones (33%), with

substantial emphasis on the emotions of trust, joy, and surprise. It is worth noting that negative emotions such as sadness, fear, and anger appear quite representative for a festive moment. The Naive Bayes classifier achieved an accuracy of 76%, a median result for a classification model, indicating that out of 10 analyzed texts, it correctly classifies 8.

**Keywords:** Sentiment Analysis, Data Mining, Folk Festival, Social Networks, Machine Learning, Classification

## INTRODUÇÃO

Com o crescimento e o avanço da internet, indivíduos e organizações passaram a registrar e divulgar suas ideias, opiniões e atitudes nas redes sociais e em blogs por meio de diferentes formatos de comunicação, como textos, vídeos, imagens e emojis, entre outros. Essa popularização possibilitou o surgimento de uma nova categoria de comunicadores: indivíduos que, outrora, eram meros coadjuvantes passaram a assumir um papel protagonista, rompendo com a lógica de simples receptores de informação para participarem diretamente do processo comunicacional.

Dessa forma, não estão mais limitados às suas redes pessoais de contato ou às opiniões deste ou daquele especialista disponível publicamente (Becker; Tumitan, 2013; Pang; Lee, 2008).

Esse fluxo de comunicação na internet, no qual diferentes temas são abordados e debatidos diariamente — especialmente aqueles com maior destaque, popularmente conhecidos como temas “em alta” —, seja por profissionais da área ou por usuários comuns, abre espaço para diversas frentes de pesquisa, como política, economia, mudanças climáticas, assédio, lançamentos de produtos ou marcas, entre outros. As comunicações postadas, em formato de texto, imagens ou emojis, carregam grandes quantidades de informações que podem ser extraídas e analisadas de forma sistemática. Assim, torna-se possível compreender e, em certa medida, prever comportamentos humanos a partir do conteúdo presente nas informações veiculadas em publicações online (Pang; Lee, 2008).

Segundo Liddy (2001), análises obtidas a partir de dados extraídos da internet apresentam maior confiabilidade, pois os indivíduos, de forma consciente ou inconsciente, tendem a expressar suas emoções de maneira mais espontânea durante interações em ambientes digitais.

O estudo de dados provenientes da internet, por meio de rotinas e etapas computacionais, pode representar um diferencial relevante para novas pesquisas, sobretudo diante das

limitações enfrentadas por métodos tradicionais de coleta e análise de dados. Xavier et al. (2020) destacam que o aumento da produção textual em plataformas digitais possibilita a extração e a interpretação de informações úteis, gerando ganhos significativos em termos de agilidade, economia de tempo e redução de custos.

Nesse contexto, uma das metodologias amplamente aceitas no meio acadêmico para a análise de dados oriundos de redes sociais é a análise de sentimentos. Essa técnica constitui uma subárea do Processamento de Linguagem Natural (PLN) e utiliza métodos computacionais para identificar, extrair e classificar automaticamente os sentimentos expressos em comunicações digitais. Para Liu (2020), a análise de sentimentos busca identificar o sentimento presente em uma comunicação, bem como as emoções a ele associadas, por meio de ferramentas computacionais adequadas.

A análise de sentimentos pode ser empregada para classificar o conteúdo emocional expresso pelos usuários da internet em publicações, fornecendo uma visão clara de tendências, como a polaridade das opiniões — positiva, negativa ou neutra — e das emoções relacionadas a determinado evento ou tema em discussão. Essa técnica permite o levantamento e a análise de grandes volumes de dados de forma ágil, algo que, muitas vezes, não é alcançado por métodos tradicionais de análise (Yu; Duan; Cao, 2013).

Diante da complexidade inerente ao processo de análise, é comum a utilização de diferentes técnicas e algoritmos para tratar o problema em estudo. Conforme Pang e Lee (2008), as abordagens mais recorrentes para a extração de informações presentes em conteúdos digitais são os métodos de classificação supervisionados e não supervisionados.

Na abordagem supervisionada, podem ser utilizados dicionários léxicos e/ou modelos estatísticos de aprendizagem de máquina treinados a partir de conjuntos de dados rotulados por humanos. Esses modelos são empregados para prever os sentimentos expressos nas publicações. Tal abordagem tende a apresentar melhores resultados quando há uma quantidade significativa de dados rotulados, proporcionando maior precisão na classificação. A análise léxica, por sua vez, ocorre por meio de conjuntos de palavras previamente rotuladas e organizadas em dicionários léxicos (Turney; Littman, 2003).

Por outro lado, os métodos não supervisionados não dependem de dados rotulados, utilizando algoritmos de agrupamento para a identificação de padrões. De acordo com Melo (2019), técnicas como análise de clusters e análise de componentes principais são exemplos

de métodos utilizados nessa abordagem, em que os dados são agrupados conforme suas características de similaridade.

A escolha entre métodos supervisionados e não supervisionados depende de fatores como a disponibilidade de dados rotulados, o domínio do texto e os recursos computacionais disponíveis. Em alguns casos, pode-se adotar uma abordagem híbrida ou optar exclusivamente pela análise léxica, visto que esta permite identificar padrões de sentimentos sem a necessidade de rótulos prévios (Nascimento; Menandro, 2006).

Para o desenvolvimento deste estudo, foram coletados dados de postagens na rede social X relacionadas ao Festival Folclórico de Parintins, considerando os sentimentos e emoções manifestados por participantes e simpatizantes do evento. O Festival Folclórico de Parintins é realizado anualmente no município de Parintins, no estado do Amazonas, no último final de semana do mês de julho. Trata-se de uma celebração cultural marcada pela disputa entre os bois-bumbás Garantido e Caprichoso, cujos enredos abordam elementos da cultura amazônica, como povos indígenas, lendas regionais, crenças e costumes tradicionais.

Sua origem remonta ao ano de 1965, quando foi criado por membros da Juventude Alegre Católica (JAC) com o objetivo de arrecadar recursos para a construção da Catedral de Nossa Senhora do Carmo, padroeira da cidade (Castro, 2025).

Segundo a Associação Folclórica Boi-Bumbá Garantido (2024), o Boi-Bumbá Garantido foi criado em 1913 por Lindolfo Monteverde como cumprimento de uma promessa feita a São João Batista após sua recuperação de uma grave doença. A promessa consistia na criação de um boi que sairia anualmente pelas ruas no dia 24 de junho, data dedicada ao santo. A origem religiosa do Boi Garantido reforça a forte relação entre fé, cultura popular e identidade regional da população de Parintins (Garantido, 2023).

O Boi Caprichoso, por sua vez, foi fundado em 1913 por famílias nordestinas, como os Cid e Gonzaga, que trouxeram a tradição do boi-bumbá para a região (Cultura AM, 2023). A rivalidade entre os bois evoluiu para o festival como é conhecido atualmente, sendo oficializado em 1965 com o objetivo de promover a cultura local e fortalecer a identidade regional (Mundo Educação, 2024).

Além de seu valor simbólico e cultural, o Festival de Parintins exerce forte impacto econômico no município. Em 2024, a expectativa era de que o evento movimentasse cerca de R\$ 150 milhões, atraindo aproximadamente 120 mil visitantes, número superior à população local (Veja, 2024). Esse fluxo financeiro impulsiona diversos setores da economia local,

beneficiando artesãos, costureiras, vendedores autônomos, tricicleiros e outros profissionais ligados à cadeia turística e de serviços (CNA7, 2023).

Diante do exposto sobre a análise de dados de posts na internet e da contextualização do Festival Folclórico de Parintins, este trabalho tem como objetivo identificar os sentimentos e emoções expressos pelos usuários da rede social X em relação ao evento, por meio da técnica de análise de sentimentos. Essa abordagem mostra-se adequada, pois possibilita captar percepções presentes em grandes volumes de dados de forma automatizada.

## **PROBLEMATIZAÇÃO**

A popularização das redes sociais transformou os usuários em produtores ativos de conteúdo, fenômeno especialmente perceptível em eventos de grande relevância cultural, como o Festival Folclórico de Parintins. Apesar da intensa movimentação digital gerada pelo evento, ainda são escassas as análises sistemáticas voltadas à compreensão dos sentimentos e emoções expressos pelo público nas redes sociais. Diante dessa lacuna, surge a seguinte questão de pesquisa: quais sentimentos e emoções são manifestados nas redes sociais sobre o Festival Folclórico de Parintins, e de que forma essa compreensão pode contribuir para a interpretação da experiência do público e para a gestão do evento?

## **OBJETIVOS**

### **Objetivo Geral**

- Identificar os sentimentos e emoções expressos pelos usuários da rede social X sobre o Festival Folclórico de Parintins, utilizando técnicas de análise de sentimentos e algoritmos de aprendizagem de máquina.

### **Objetivos específicos**

- Realizar o pré-processamento dos tweets coletados, incluindo limpeza, padronização e formação do corpus textual.
- Classificar os sentimentos e emoções dos posts por meio da análise léxica com o dicionário NRC.
- Aplicar o classificador Naive Bayes para identificar a polaridade (positiva/negativa) dos sentimentos.

- Avaliar o desempenho do modelo de classificação utilizando métricas como acurácia, precisão e revocação.
- Identificar os principais sentimentos e emoções predominantes nas publicações relacionadas ao festival.

## **REFERENCIAL TEÓRICO**

### **O avanço das redes sociais**

As redes sociais digitais representam uma transformação significativa na forma como os indivíduos se comunicam na contemporaneidade. Seu desenvolvimento está diretamente associado à popularização da internet a partir das últimas décadas do século XX. No início dos anos 2000, essas plataformas passaram a se consolidar como espaços dinâmicos de interação, nos quais os próprios usuários produzem e compartilham conteúdos (Castro, 2025).

Inicialmente restritas a fóruns e blogs, as experiências digitais evoluíram para plataformas como Orkut, Facebook, Twitter, Instagram e TikTok. Mais do que simples produtos do avanço tecnológico, essas mídias reformularam as formas de sociabilidade, transcendendo fronteiras geográficas, culturais e afetivas para conectar pessoas e vivências em escala global. Segundo Santos e Santos (2014), as redes sociais digitais emergem como ferramentas de comunicação e informação que influenciam significativamente a sociedade, promovendo novas formas de interação e compartilhamento de conhecimento.

Atualmente, essas redes ocupam um lugar central na produção de sentidos e na mediação do cotidiano. Convertem-se em espaços de visibilidade pública e de construção identitária, funcionando simultaneamente como arenas políticas, culturais e afetivas. Sua influência vai além do entretenimento, moldando comportamentos, discursos, posicionamentos e afetos (Bucci, 2021).

Mais do que meras ferramentas de comunicação, as redes sociais constituem-se como dispositivos que estruturam relações sociais e econômicas em escala global, reconfigurando noções de privacidade, intimidade e coletividade. Estudos como o de Siqueira et al. (2024) evidenciam que o uso excessivo dessas plataformas pode impactar negativamente a saúde mental de adolescentes e jovens adultos, associando-se a sintomas de depressão, ansiedade e estresse, o que ressalta a necessidade de uma compreensão crítica sobre o papel das redes sociais na formação dos indivíduos.

As redes sociais, portanto, veiculam emoções e funcionam como catalisadoras das experiências dos usuários, alimentando dinâmicas de engajamento, mobilização e, por vezes, de polarização. De acordo com Guazzini et al. (2016), a expressão emocional nas redes sociais pode gerar um efeito de contágio emocional, no qual as emoções expressas em publicações influenciam as respostas emocionais de outros usuários, promovendo uma coerência emocional nas interações online.

Dessa forma, é possível afirmar que as redes sociais digitais transcendem sua função original de conexão entre pessoas, configurando-se como dispositivos complexos que interferem diretamente nos modos de viver, sentir e interagir no mundo contemporâneo. Analisar seus impactos sociais, psicológicos e culturais exige uma abordagem crítica e interdisciplinar, que considere tanto seus potenciais quanto seus riscos, bem como a dimensão afetiva inerente às interações digitais.

### **A utilização das redes sociais e análise de sentimentos**

Por constituírem um território privilegiado de manifestação de opiniões, afetos e sentimentos, as redes sociais tornam-se um campo fértil para a compreensão das percepções que se formam no universo digital. Uma ferramenta essencial nesse processo é a chamada análise de sentimentos (A.S). Segundo o estudo de Silva (2023), a A.S em redes sociais permite identificar padrões emocionais expressos em textos, comentários e interações, oferecendo ideias valiosas para diversas áreas, como a administração, o marketing, a política e a saúde pública, por exemplo.

Em uma pesquisa realizada por Medeiros *et al.* (2019), foram analisados tweets relacionados ao processo de impeachment no Brasil, utilizando léxicos específicos para identificar sentimentos predominantes. Os resultados indicaram uma predominância de sentimentos negativos durante os períodos de maior instabilidade política, evidenciando como as redes sociais refletem e amplificam as emoções coletivas em momentos críticos.

Além disso, estudos como o de Silva (2023), demonstram que a análise de sentimentos pode ser aplicada na avaliação da opinião pública sobre a atuação do governo, contribuindo para a melhoria dos serviços públicos. A pesquisa identificou que a A.S, baseada em textos extraídos de mídias sociais, pode ser um instrumento eficaz de mensuração da opinião pública, auxiliando na tomada de decisões mais alinhadas com as expectativas da sociedade.

Dentro do campo da Administração, a análise de sentimentos configura uma metodologia valiosa para administrar a comunicação da organização, seus vínculos com partes interessadas (tais como clientes e parceiros) e seu posicionamento no mercado. Empresas que analisam as interações em plataformas digitais conseguem extrair, instantaneamente, avaliações sobre produtos, serviços e políticas, permitindo correções de rumo alicerçadas em informações sólidas.

De acordo com Silva (2023), a análise de sentimentos em redes sociais permite compreender padrões emocionais expressos pelos usuários, sendo uma ferramenta poderosa na tomada de decisão gerencial. Seu uso na Administração revela-se essencial na antecipação de crises e no desenvolvimento de estratégias mais sensíveis ao comportamento do consumidor.

No setor de Recursos Humanos, essa abordagem também pode ser empregada para avaliar o clima organizacional. Como observado por Houana (2021), a aplicação de técnicas de ciência de dados em interações digitais de clientes e colaboradores permite identificar sentimentos predominantes e áreas críticas de insatisfação, orientando ações para melhorar o ambiente de trabalho e a comunicação interna.

Além disso, o monitoramento contínuo das redes sociais oferece vantagens significativas no contexto da experiência do cliente. Segundo Oliveira e Santos (2020), a análise de sentimentos permite identificar rapidamente sinais de descontentamento e pontos positivos nas interações dos consumidores com as marcas, favorecendo uma resposta ágil e assertiva por parte das organizações. Essa prática fortalece a fidelização e aprimora os processos de comunicação e atendimento.

No âmbito da gestão de crises e da governança corporativa, Ramos e Vieira (2021), apontam que a análise de sentimentos em tempo real pode orientar a comunicação institucional durante situações delicadas, como crises reputacionais, acidentes ou escândalos. A identificação do sentimento predominante entre o público permite que a organização module sua comunicação de forma mais eficaz, protegendo sua imagem e restaurando a confiança do consumidor.

Do ponto de vista da Administração Pública, acompanhar as interações dos cidadãos nas plataformas digitais tem se revelado uma ferramenta estratégica para desenvolver políticas públicas mais alinhadas às demandas sociais. Como afirma Silva (2023), a análise de

sentimentos pode ser aplicada como instrumento de avaliação da opinião pública, promovendo maior alinhamento entre governo e sociedade e fortalecendo os princípios da transparência e da gestão participativa.

Por fim, vale ressaltar que, embora a análise de sentimentos ofereça inúmeras possibilidades, seu uso deve ser ético e contextualizado. Malini (2017), chama atenção para os desafios metodológicos e políticos envolvidos na análise emocional das redes, especialmente em contextos polarizados, como no caso do impeachment de Dilma Rousseff. Portanto, a análise automatizada de emoções deve ser acompanhada de interpretação crítica, evitando leituras reducionistas ou enviesadas.

### **Análise de sentimentos**

A análise de sentimentos (A.S.), também conhecida como mineração de dados, é uma área da ciência que busca extrair, identificar e classificar opiniões, emoções e atitudes expressas em diferentes formatos de comunicação, como textos, imagens, emojis, entre outros, por meio de algoritmos computacionais. Segundo Liu (2012), a A.S. estuda as opiniões, sentimentos, atitudes e emoções das pessoas em relação a entidades como produtos, serviços, organizações, indivíduos, eventos e seus atributos. O objetivo principal é determinar a polaridade carregada no meio de comunicação, ou seja, se a comunicação expressa uma opinião positiva, negativa ou neutra e, em alguns casos, identificar a intensidade emocional associada.

De acordo com Xavier et al. (2020), a A.S. configura-se como um subcampo do Processamento de Linguagem Natural (PLN) que busca aproximar o processamento de linguagem realizado por máquinas daquele realizado por seres humanos. No contexto da sociedade contemporânea, marcada por interações digitais massivas, a análise de sentimentos assume um papel estratégico ao permitir a detecção de afetos e a compreensão das emoções expressas nos diferentes meios de comunicação.

Para Melhado e Rabot (2021), trata-se de uma ferramenta técnica que se configura como um mecanismo profundamente imbricado nos processos de psicopolítica e nos dispositivos de controle social, dado seu potencial de captar e mapear o campo emocional coletivo. A operacionalização da análise de sentimentos envolve a transformação de dados textuais, provenientes de interações em linguagem natural, em representações numéricas compreensíveis para os algoritmos computacionais.

Nessa perspectiva, Cambria et al. (2013) destacam que a análise de sentimentos não se limita à simples categorização de textos em positivos, negativos ou neutros. Ela demanda uma compreensão mais profunda do contexto, da subjetividade e das nuances linguísticas presentes nas expressões humanas.

Para exemplificar, apresenta-se a seguinte situação: se, em um post que fale sobre um filme, um usuário comentasse “Que filme incrível, mas não é para qualquer um...”, percebe-se que a mensagem contém um sentimento positivo e, além disso, uma nuance de exclusividade que só pode ser captada ao se compreender o contexto desse filme. Nesse sentido, os autores citados enfatizam o uso de técnicas avançadas de processamento de linguagem natural, aprendizado de máquina e inteligência artificial, capazes de analisar grandes volumes de dados textuais oriundos de diferentes fontes, como redes sociais, blogs, fóruns e avaliações de produtos.

### **Métodos utilizados na pesquisa de análise de sentimentos**

Segundo Wankhade, Rao e Kulkarni (2022), vários são os métodos para realizar a análise de sentimentos, destacando-se as abordagens principais, que são: os métodos supervisionados, os não supervisionados e os híbridos. Dentro dos métodos supervisionados, têm-se a análise léxica e os modelos de aprendizado de máquina. A abordagem léxica consiste em associar uma pontuação a cada palavra de acordo com sua polaridade — positiva, negativa ou neutra — e, a partir disso, somar essas pontuações para determinar o sentimento predominante no texto.

A abordagem baseada em aprendizado de máquina utiliza algoritmos para identificar, quantificar e categorizar sentimentos em textos, podendo ser aplicada nos métodos supervisionados ou não supervisionados. Na abordagem não supervisionada, o modelo infere a polaridade a partir de uma base de conhecimento, sem a necessidade de treinamento prévio, buscando reconhecer padrões sem conhecimento prévio. Já na abordagem supervisionada, o algoritmo aprende a partir de um conjunto de dados rotulados para, posteriormente, classificar a polaridade de novos textos de forma autônoma (Wankhade; Rao; Kulkarni, 2022).

Para aprofundar essa distinção, Sampaio (2021) explica que a aprendizagem de máquina supervisionada envolve quatro etapas principais: rotulação dos dados, identificação das características que distinguem esses dados, treinamento com um algoritmo específico e, por fim, aplicação do modelo treinado.

Esse processo exige a intervenção humana para rotular previamente os dados, fornecendo exemplos classificados que servem de base para o aprendizado do sistema. Por exemplo, em uma análise de sentimentos, comentários podem ser rotulados com emojis que indicam emoções positivas, negativas ou neutras, ajudando o algoritmo a reconhecer essas polaridades em novos textos.

Por sua vez, a abordagem híbrida combina análise léxica e aprendizado de máquina, aliando a simplicidade e a rapidez do dicionário léxico à capacidade de adaptação e aprendizado dos algoritmos supervisionados. Essa integração permite superar limitações isoladas de cada técnica, proporcionando uma análise de sentimentos mais precisa e robusta.

A análise de sentimentos pode ser realizada por meio de diferentes métodos, que variam em complexidade, capacidade de interpretação e necessidade de dados rotulados. A Tabela 1 apresenta uma síntese dos principais métodos utilizados, com descrições e exemplos desses modelos:

**Tabela 1:** Principais métodos de análise de sentimentos.

<b>Método</b>	<b>Descrição</b>	<b>Exemplos/Técnicas</b>
<b>Léxico</b>	Usa listas/dicionários de palavras com polaridades definidas (positivo/negativo/neutro).	SentiWordNet, AFINN, VADER.
<b>Aprendizado de Máquina</b>	Identifica padrões e classifica sentimentos com base em dados rotulados (supervisionado) ou sem rótulos (não supervisionado).	- <i>Naive Bayes</i> : assume independência dos atributos (Renault, 2020). - <i>SVM</i> : separa classes com um hiperplano ideal (Chen et al., 2015).

		<ul style="list-style-type: none"> <li>- <b>Random Forest:</b> combina várias árvores para maior robustez.</li> <li>- <b>MaxEnt (Maximum Entropy):</b> busca o modelo mais imparcial.</li> <li>- <b>MLP:</b> rede neural com múltiplas camadas (Renault, 2020).</li> <li>- <b>Regressão Logística:</b> adequada para problemas binários (Sainani, 2014).</li> <li>- <b>K-Means/LDA:</b> exemplos de métodos não supervisionados.</li> </ul>
<b>Redes Neurais Artificiais</b>	Modelam padrões complexos, inspiradas no cérebro humano; úteis para classificação e predição.	<i>Multilayer Perceptron</i> (MLP), Redes Recorrentes (RNNs), Redes Convolucionais (CNNs).
<b>Transformers</b>	Modelos avançados para PLN, com atenção para dependências de longo alcance.	BERT, GPT, RoBERTa (Wankhade; Rao; Kulkarni, 2022).
<b>Abordagem Híbrida</b>	Combina métodos léxicos e de aprendizado de máquina para superar limitações individuais.	Integra dicionários com algoritmos supervisionados.

Fonte: Castro, 2025.

Um número crescente de estudos tem utilizado as técnicas de análise de sentimentos como método para investigar as opiniões das pessoas sobre temas específicos, como educação, ensino e dados de mídias sociais. A ferramenta se destaca por sua capacidade de captar nuances emocionais, identificar tendências discursivas e acompanhar mudanças de percepção ao longo do tempo.

Tais atributos reforçam sua relevância como instrumento metodológico em pesquisas sociais e como apoio à tomada de decisões estratégicas em diferentes setores, como educação, saúde e políticas públicas. Diante desse potencial, torna-se relevante analisar os pontos fortes e as limitações técnicas de cada abordagem utilizada na análise de sentimentos. A Tabela 2, apresenta uma síntese comparativa das principais abordagens.

**Tabela 2:** Vantagens e desvantagens dos métodos de Análise de sentimentos.

<b>Método</b>	<b>Vantagens</b>	<b>Desvantagens</b>
<b>Abordagem Léxica</b>	<ul style="list-style-type: none"> <li>- Não requer dados anotados para treinamento-</li> <li>- Rápida e fácil de implementar-</li> <li>- Boa para análises exploratórias e contextos com poucos recursos computacionais.</li> </ul>	<ul style="list-style-type: none"> <li>- Dificuldade com ironias, sarcasmos e ambiguidade-</li> <li>- Não captura o sentido global do texto-</li> <li>- Depende de dicionários fixos, dificultando adaptação a gírias e neologismos.</li> </ul>
<b>Aprendizado de Máquina</b>	<ul style="list-style-type: none"> <li>- Aprende padrões complexos e contextuais-</li> <li>- Pode ser adaptada para diferentes domínios-</li> <li>- Evolui com novos dados.</li> </ul>	<ul style="list-style-type: none"> <li>- Requer grande volume de dados rotulados-</li> <li>- Risco de overfitting-</li> <li>- Interpretação do modelo pode ser limitada.</li> </ul>
<b>Abordagem Híbrida</b>	<ul style="list-style-type: none"> <li>- Combina simplicidade do léxico com precisão do aprendizado de máquina-</li> </ul>	<ul style="list-style-type: none"> <li>- Mais complexa de implementar e ajustar-</li> <li>- Demanda maior recurso</li> </ul>

	Reduz falhas isoladas- Análise mais robusta.	computacional- Desafios na integração dos resultados.
--	---	--

Fonte: Castro, 2025.

Conforme apresentado na tabela acima, cada abordagem possui implicações práticas que precisam ser consideradas conforme o objetivo da pesquisa e o contexto dos dados. Os métodos baseados em aprendizado de máquina, por exemplo, geralmente dependem de bases de dados rotuladas para treinar os classificadores, o que pode representar uma limitação significativa devido ao custo e o tempo envolvidos na obtenção desses dados. A rotulação manual é, inclusive, uma das principais restrições desses métodos (Cambria et al., 2013).

Já os métodos léxicos, por sua vez, utilizam dicionários associados a sentimentos específicos, os chamados léxicos de opinião, que permitem mapear a polaridade emocional do texto sem a necessidade de dados previamente rotulados (Camargo & Cabral, 2019; Pereira et al., 2020). Apesar dessa vantagem, sua eficiência depende diretamente da abrangência e da adaptação do vocabulário ao contexto analisado.

A generalização desses métodos é desafiada pela presença de figuras de linguagem, como sarcasmo e ironia, que podem distorcer a interpretação emocional do texto e reduzir a precisão da análise (Silva, 2016). Como alternativa à construção manual, que é trabalhosa e demorada, muitas pesquisas utilizam léxicos já validados, como o SentiLex-PT (Basile & Nunes, 2014), amplamente empregado em estudos em língua portuguesa.

### **A plataforma X como fonte de dados para a análise de sentimentos**

Após discutir os métodos léxicos e a importância de léxicos validados para a análise de sentimentos, é fundamental também considerar a escolha da fonte de dados, que influencia diretamente a qualidade e a relevância dos resultados. Nesse sentido, as redes sociais têm se destacado como ambientes ricos para a coleta de opiniões e emoções expressas em tempo real. Dentre elas, a rede social X apresenta características específicas, como mensagens curtas, uso intenso de hashtags e menções, além de um volume elevado de postagens, que o tornam uma plataforma particularmente valiosa para pesquisas de análise dessa natureza (Vaillant, 2022).

A dinâmica do X favorece a captura de reações imediatas a eventos sociais, políticos e culturais, permitindo análises temporais sobre o sentimento coletivo. O ritmo veloz de circulação das informações e a linguagem própria dos usuários, repleta de abreviações e gírias,

impõem, no entanto, desafios adicionais ao tratamento automático dos dados, exigindo técnicas específicas de pré-processamento e interpretação textual (Bontcheva; Rout, 2014).

Desse modo, a relevância do X como fonte de dados para pesquisas em análise de sentimentos pode ser entendida a partir de uma perspectiva histórica sobre as redes sociais digitais. O surgimento dessas plataformas no final dos anos 1990 mudou profundamente a forma como as pessoas interagem e compartilham informações.

Primeiramente, redes como Six Degrees (1997) e Friendster (2002) introduziram a lógica de criação de perfis e conexões em rede (Oliveira, 2011 apud Sampaio, 2021). Com o avanço tecnológico e o crescente desejo por conexão social, surgiram redes mais robustas, como o Orkut, Facebook e, em 2006, o Twitter, que atualmente é conhecido como X. Essa trajetória histórica ajuda a compreender por que o X se tornou uma fonte valiosa para estudos sobre sentimentos e comportamentos online.

Segundo Sampaio (2021), a rede social X foi idealizado como um serviço de comunicação por mensagens curtas, os “tweets”, que podem conter textos, links, imagens ou vídeos, sendo caracterizado pela agilidade na difusão de conteúdos e pela sua centralidade em eventos sociais, políticos e culturais. Seu diferencial sempre esteve no formato conciso e dinâmico das publicações, favorecendo a instantaneidade das reações e a viralização de conteúdo.

Embora o Facebook concentre o maior número de usuários globalmente, o X se consolidou como uma plataforma central em pesquisas de análise de sentimentos, justamente por permitir postagens curtas e espontâneas, características ideais para captar emoções e opiniões em tempo real. Além disso, seu sistema aberto de hashtags, menções e trending topics contribui para mapear padrões emocionais coletivos e oscilações na opinião pública diante de eventos de grande impacto (Sampaio, 2021).

Por esses motivos, o X acabou se tornando uma das fontes mais exploradas em estudos de mineração de sentimentos, servindo como campo fértil para experimentos com inteligência artificial, aprendizado de máquina e técnicas de processamento de linguagem natural.

Neste sentido, a presente pesquisa adota uma metodologia que combina abordagens léxicas com o classificador Naive Bayes, com o objetivo de mapear as percepções e sentimentos da população de Parintins durante o período do Festival Folclórico de Parintins. Ao articular rigor técnico e relevância social, busca-se produzir subsídios concretos para a formulação e o

aprimoramento de políticas públicas locais. Em síntese, a análise de sentimentos se configura como uma técnica eficaz de mineração de texto, cuja aplicabilidade depende dos objetivos da investigação e da complexidade do contexto analisado (Brito, 2017).

## **JUSTIFICATIVA**

O Festival Folclórico de Parintins é um dos maiores eventos culturais do Brasil, marcado por forte expressão identitária e por grande impacto econômico e turístico. Com a expansão das redes sociais, muitos participantes utilizam essas plataformas para relatar suas percepções, expectativas e experiências sobre o festival.

A compreensão dos sentimentos expressos online permite:

- Monitorar a aceitação do evento e identificar pontos de satisfação ou insatisfação.
- Ajudar gestores públicos e organizadores a compreenderem aspectos estruturais, como transporte, hospedagem, segurança e logística.
- Obter um retrato fiel do comportamento e da experiência do público, uma vez que, segundo Liddy (2001), indivíduos expressam emoções mais autênticas nas redes sociais.
- Utilizar metodologias modernas que garantem agilidade, economia de tempo e precisão frente a métodos tradicionais (Xavier et al., 2020).

Portanto, a pesquisa se justifica pela relevância cultural do festival, pela necessidade de estudos que associem mineração de dados ao comportamento do público e pelo potencial de geração de insights que contribuam para a gestão e melhoria contínua do evento.

## **METODOLOGIA**

### **OS DADOS DO ESTUDO**

Estes foram coletados da rede social X nos meses de abril, maio e junho de 2022, período em que acontecem os ensaios típicos dos bumbás e as apresentações no último final de semana de julho, no Bumbódromo. Tal período foi considerado em razão de as discussões nas redes crescerem com a proximidade do festival, visto que vários eventos ocorrem e os embates e opiniões tornam-se bem acirrados. Foram realizadas quatro coletas, obtendo-se 1.715 tweets não estruturados, por meio do programa R (R Core Team, 2022). Para a captura dos posts sobre a festa, utilizaram-se as seguintes expressões de busca: “#festa Parintins”, “#boi garantido”, “#boi caprichoso”, “#festival folclórico”, “#festa caprichoso”, “#festa garantido”, “#boi rua” e “#alvorada garantido”.

## PRÉ-PROCESSAMENTO

De posse dos dados não estruturados, procedeu-se à etapa do pré-processamento. Esta fase compreende todas as etapas relacionadas à organização e à adequação textual dos posts, como seleção, limpeza, integração, transformação, redução, entre outras, visto que o uso da linguagem em posts é carregado de textos abreviados, figuras, expressões sem sentido, símbolos, entre outros, ou seja, carregam uma quantidade de informações que não trazem contribuição para a análise (Teixeira, 2019; Soares, 2020).

Na etapa de limpeza, são removidos elementos textuais como URLs, menções a usuários (precedidas por “@”), hashtags irrelevantes, emojis, números, pontuações excessivas e caracteres especiais. Já a normalização consiste em converter expressões e palavras para padrões uniformes, como letras minúsculas e eliminação de acentuações, entre outros, com o objetivo de padronizar os termos e evitar variações desnecessárias que possam comprometer a análise (Morais; Ambrósio, 2007). Informações duplicadas ou irrelevantes são eliminadas para manter apenas os dados úteis.

Ainda, realiza-se a remoção das chamadas *stopwords*, que são palavras com pouco ou nenhum valor semântico para a tarefa de análise, como artigos, preposições e pronomes (por exemplo: “o”, “a”, “de”, “que”). Essa filtragem contribui para que os algoritmos se concentrem nos termos relevantes do conteúdo.

Após os dados estarem organizados em formato textual adequado, estes são estruturados no que se chama de *corpus*, que é um conjunto de textos previamente processados. Em um *corpus*, cada post representa uma unidade textual. Uma vez estando o *corpus* devidamente estruturado, realiza-se a tokenização, processo que divide o texto em unidades menores chamadas *tokens*, ou seja, a separação de um texto em palavras ou expressões. O processo de tokenização permite contar a frequência da ocorrência de cada termo de um post (texto).

Em posse da estrutura textual adequada, procedeu-se a uma análise exploratória por meio de gráficos de nuvens de palavras. Esta análise descritiva busca conhecer os termos mais frequentes, além de identificar possíveis expressões sem sentido que não foram eliminadas no pré-processamento. Nuvens de palavras são imagens compostas por palavras que demonstram, de maneira visual, a frequência de ocorrência das palavras em um dado texto, bem como a classificação das emoções e sentimentos carregados nas mensagens. Nesse tipo de gráfico, quanto maior for o número de ocorrências de uma palavra, maior ela será na nuvem de palavras (Ferreira; Correa, 2020; Kwartler, 2017).

## CLASSIFICAÇÃO DOS SENTIMENTOS

A classificação dos sentimentos e emoções foi realizada por meio da análise léxica e do classificador Naive Bayes. Na análise léxica, utilizou-se o dicionário NRC, contido no “pacote” *syuzhet* do programa R (R Core Team, 2022), o qual está disponível em vários idiomas. Sua escolha particular deu-se por apresentar um leque maior de opções para análise, permitir o uso do idioma em português e possuir rótulos para as semânticas de sentimento positivo e negativo, bem como rótulos para oito tipos de emoções: raiva, antecipação, desgosto, medo, alegria, tristeza, surpresa e confiança (Gonzaga, 2021).

Um dicionário léxico é uma lista composta por muitas palavras que, em vez de fazer a descrição particular da expressão, associa um significado quantitativo à expressão que carrega um determinado estado emocional. Esse valor pode ser numérico, como 0 ou 1, em que 0 representa o sentimento negativo e 1 o sentimento positivo, ou ainda um valor qualitativo para o rótulo (como positivo ou negativo). Em síntese, trata-se de um dicionário formado por termos, sendo que cada termo possui um rótulo associado ao sentimento (positivo ou negativo) e às emoções, que variam de acordo com o dicionário em uso (Kwartler, 2017; Castro et al., 2025).

Em posse dos posts em formato adequado, estes foram submetidos ao dicionário por meio de uma rotina computacional apropriada, realizando-se a classificação dos sentimentos (Benvenuto; Ribeiro; Araújo, 2017). Por exemplo, a palavra “amor”, no dicionário léxico, é rotulada com o valor 1, significando um sentimento positivo. Logo, na análise léxica, o algoritmo procura a ocorrência da palavra no texto e, uma vez encontrada, classifica-a como positiva ou como pertencente a emoções relacionadas ao sentimento positivo.

Na abordagem por meio do classificador Naive Bayes (NB), é fornecido ao algoritmo um conjunto de dados textuais previamente rotulados por humanos ou por máquina. Com isso, o algoritmo busca identificar padrões nos dados de entrada com base em suas respectivas categorias, caracterizando a etapa de treinamento. Após essa etapa, segue-se a fase de teste, na qual se avalia a qualidade do poder preditivo do método, ou seja, se o modelo é apropriado para realizar a classificação. Uma vez considerado adequado, procede-se à classificação dos sentimentos (Soumya; Pramod, 2020).

O classificador Naive Bayes utiliza o Teorema de Bayes para calcular a probabilidade de um texto ser positivo, negativo ou neutro (Aborisade; Anwar, 2018). A técnica consiste em empregar a probabilidade condicional a fim de classificar os dados

em categorias, como positivas, negativas ou neutras. Ou seja, a probabilidade do rótulo (ser positivo ou negativo), dado um texto, é igual à probabilidade do texto dado o rótulo, multiplicada pela probabilidade do rótulo, tudo dividido pela probabilidade do texto. Tal enunciado pode ser expresso conforme a Equação (1) (Gomes, 2019).

$$P(\text{rótulo}|\text{texto}) = \frac{P(\text{texto}|\text{rótulo}) * P(\text{rótulo})}{P(\text{texto})} \quad (1)$$

Ademais, como um texto é composto por várias palavras de (1) podemos escrever como em (2), resguardando que o rótulo pode ser positivo ou negativo.

$$p(\text{rótulo}|\text{palavra1,palavra2,,...,palavran}) = \frac{p(\text{palavra1,palavra2,...,palavran}|\text{rótulo})}{p(\text{palavra1,palavra2,...,palavran})} \quad (2)$$

Agora, assumindo que cada palavra em um texto é independente umas das outras, de (2) podemos escrever como em (3), que é a equação para o cálculo das probabilidades pelo classificador. Então,

$$p(\text{rótulo}|\text{palavra1,...,palavran}) = p(\text{palavra1}|\text{rótulo}) * p(\text{palavra2}|\text{rótulo}) * \dots * p(\text{palavran}|\text{rótulo}) * p(\text{rótulo}) \quad (3)$$

Por exemplo, “garantido meu amor”, se houver 4 ocorrências de "amor" em textos positivos e houver 16 palavras no total, então a  $P(\text{amor} | \text{positivo})$  seria calculado da seguinte maneira:

$$p(\text{amo}|\text{positivo}) = \frac{4}{16}$$

Uma vez que as probabilidades são calculadas para cada classe de sentimento pelo Naive Bayes, verifica-se qual das categorias de sentimento apresenta a maior probabilidade para, em seguida, com base nessa maior probabilidade, classificar o sentimento. Por exemplo, se um texto possui uma probabilidade mais alta de ser positivo do que negativo ou neutro, então o algoritmo irá classificá-lo como positivo (Carvalho, 2023).

Como critério para avaliar o desempenho das técnicas, utilizaram-se as métricas de acurácia, precisão e revocação, estatísticas que podem ser obtidas diretamente da matriz de

confusão. A matriz de confusão é uma tabela que fornece uma visão detalhada dos erros e acertos cometidos pelos classificadores (Jurafsky; Martin, 2017).

A leitura dos resultados em uma matriz de confusão ocorre da seguinte forma (Junior et al., 2022): nas entradas das linhas são apresentados os valores que de fato são (valores reais) e, nas entradas das colunas, os valores preditos pela técnica. Assim, um valor lido por linha nessa tabela corresponde ao valor verdadeiro, enquanto o valor lido por coluna corresponde ao valor predito. Por exemplo, a interseção entre a linha positiva e a coluna positiva representa o percentual de acertos. A diagonal principal contabiliza os acertos, enquanto os valores fora da diagonal correspondem aos erros. Em um problema de análise de sentimentos, as entradas por linhas e colunas representam as frequências de sentimentos, como positivo, negativo e neutro. Dessa forma, a matriz de confusão para três classes de sentimentos é apresentada na Tabela 1.

Assim, em uma matriz de confusão, a acurácia indica o percentual de sentenças que foram corretamente classificadas pelo algoritmo. Trata-se da soma de todos os acertos em todas as classes (elementos que compõem a diagonal principal) dividida pelo número total de sentenças. A precisão mede o percentual de sentenças corretamente classificadas para uma classe específica, sendo calculada pela divisão entre os acertos da classe e o total de elementos classificados como pertencentes a essa classe. Já a revocação indica a capacidade do modelo em identificar corretamente os elementos de uma determinada classe, sendo calculada como a razão entre o total de sentenças corretamente classificadas para uma classe e o total de sentenças existentes dessa classe.

Ademais, na Tabela 1, a precisão **P** de uma classe positiva **a** é calculada como:  $P(\text{Pos. a}) = a/(a+d+g)$ , a revocação **R** da classe positiva **a** é:  $R(\text{Pos. a}) = a/(a+b+c)$  e a acurácia **Ac** é a razão entre os elementos que compõem a diagonal principal da matriz e o total geral, assim,  $Ac = (a+e+i)/(a+b+c+d+e+f+g+h+i)$ . Uma outra métrica bastante utilizada é F1-Score, que nada mais é que a média harmônica entre precisão e revocação. Para o caso positivo tem-se o  $F1(\text{Pos}) = (2P(\text{Pos}) * R(\text{Pos})) / (P(\text{Pos}) + R(\text{Pos}))$ .

Tabela 1 – Matriz de confusão para os sentimentos positivo, neutro e negativo

		Predito		
		Positivo	Neutro	Negativo
Correto	Positivo	a	b	c
	Neutro	d	e	f
	Negativo	g	h	i

Fonte: Núcleo de Estudos e Pesquisas em Ambientes Amazônicos,

UFAM, 2022

## RESULTADOS E DISCUSSÕES

O Quadro 1 apresenta parte de uma retirada de posts em sua forma não estruturada, ou seja, sem nenhum tipo de limpeza, tratamento ou remoção de *stopwords*. Neste, pode-se observar posts com informações sobre o Festival Folclórico de Parintins e a sensação dos sentimentos transmitidos pelo escritor. Vale observar que os posts estão carregados de expressões que não apresentam significado prático, como <http://t.co/parintins,RT@> e outros.

Quadro 1 - Dados retirados da plataforma X, mês de maio a julho de 2022.

<p>[1] Parintins pro fui. Vocês tem noção disso? Já é mês que vem...<a href="https://t.co/parintins">https://t.co/parintins</a></p> <p>[2] @boicaprichoso: O Caprichoso é um boi construído a muitas mãos, erguido pelo dom abençoado das pessoas de Parintins, pelos braços dessa...<a href="http://t.co/Bq1Aw">http://t.co/Bq1Aw</a></p> <p>[3] RT@danielbrandaoph: Valentina Coimbra, Sinhazinha do Boi Garantido no Lançamento do Álbum 2023 "Garantido Por Toda Vida" em Parintins/AM..<a href="http://t.co/parintinsam">http://t.co/parintinsam</a></p> <p>[4] RT @boicaprichoso: O Caprichoso é um boi construído a muitas mãos, erguido pelo</p>
--

Fonte: Rede social X, junho 2022

O Quadro 2 apresenta exemplos de posts após a etapa de pré-processamento (transformação, integração, limpeza, eliminação de *stopwords* e remoção de demais expressões que não apresentavam significado). Observa-se, para os cinco casos exemplificados, uma boa qualidade do pré-processamento, visto que não se verificam mais palavras sem significado prático para o trabalho, bem como expressões que não fazem sentido, estando todas em letras minúsculas, enquanto outras foram removidas.



Fonte: Rede social X, abril à junho de 2022

O Gráfico 2 mostra as emoções captadas nos posts dos turistas e simpatizantes por meio da análise léxica. Neste, observam-se, para as emoções de confiança, surpresa e alegria, as maiores frequências na classe positiva. Tais resultados sinalizam a admiração, a grandiosidade, a beleza e o encanto pelo Festival Folclórico de Parintins. As postagens carregam aquilo que os participantes percebem e sentem; são as emoções produzidas nos indivíduos pelo festival folclórico e pelas correlações de acontecimentos em seu entorno. Em eventos festivos, como o Festival Folclórico de Parintins, é esperado que emoções como alegria e confiança sejam predominantes.

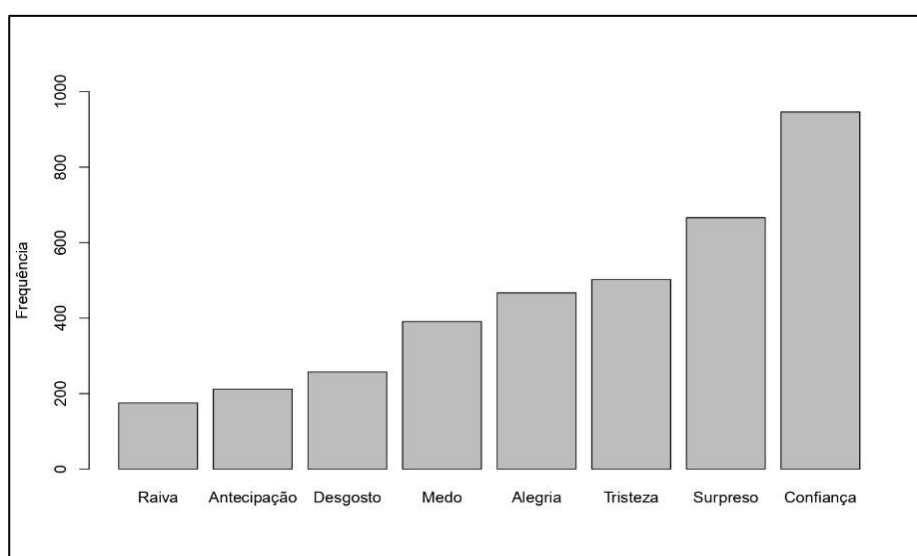
Para Izard (1991) e Ekman (1992), emoções de espectro positivo, como alegria e confiança, são frequentemente desencadeadas por experiências de celebração e momentos de confraternização e/ou união social. No contexto do Festival Folclórico de Parintins, estas são intensificadas pelo forte apego à identidade comunitária e pela grandiosidade do evento, que promove sensações de pertencimento e tradição. Também se destaca a contribuição da cidade, em parte, para essa experiência, com suas belezas naturais, como rios e paisagens exuberantes, passeios de triciclo pelas ruas da cidade e a culinária local, com seus pratos típicos, proporcionando aos visitantes um cenário imersivo e único.

Em relação ao Gráfico 2, observa-se, ainda, uma expressiva ocorrência de emoções negativas, como tristeza, medo e raiva, o que pode parecer contraditório em um evento festivo. Entretanto, tais emoções podem ocorrer e estão relacionadas diretamente a fatores pontuais vivenciados pelos participantes e às suas expectativas não correspondidas, como desconfortos, insegurança local, problemas com transporte e hospedagem, longas filas enfrentadas pelo público da arquibancada geral — público não pagante —, informações desencontradas, desgosto com o atendimento, inadequação dos espaços para alimentação, entre outros. Todavia, não se têm evidências de emoções negativas relacionadas diretamente às apresentações ligadas à festa dos bois.

Conforme explica Lazarus (1991), a tristeza pode surgir em contextos de comparação, competição ou perda, como a rivalidade entre os bois Caprichoso e Garantido no contexto do Festival de Parintins, em que os torcedores sentem-se frustrados quando suas expectativas não são atendidas, como, por exemplo, emoções fortes de xingamentos que acontecem ao boi de admiração ou ao opositor.

A identificação das emoções presentes nos posts, tanto positivas quanto negativas, relacionadas ao Festival de Parintins, traz informações valiosas para a administração pública e para a organização do evento. São evidências que indicam o rumo que o Festival Folclórico e seu entorno tomam, podendo servir de instrumento base para corrigir ou melhorar alguns aspectos negativos causadores de desconforto e insegurança encontrados na cidade e, assim, melhorar as vivências e expectativas do público participante do festival.

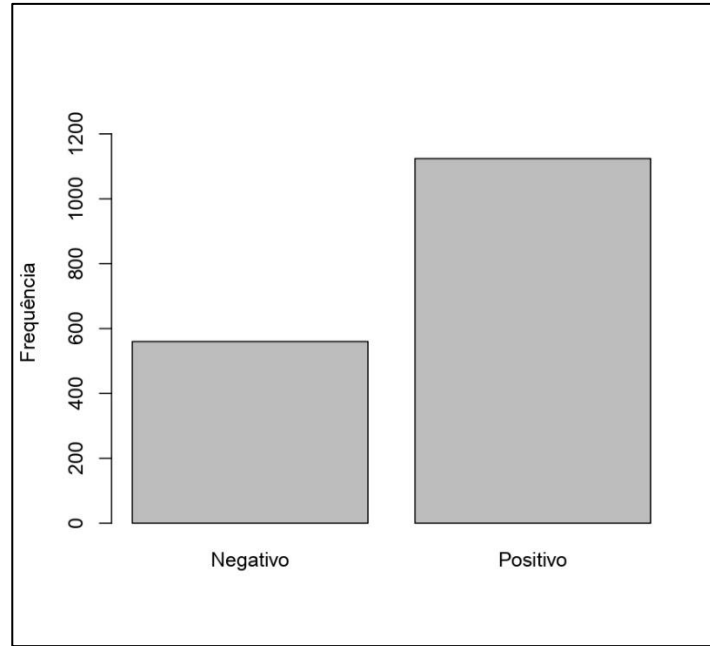
GRÁFICO 2 - Expressões de sentimentos de maior ocorrência de no mínimo duas vezes no mesmo post.



Fonte: Rede social X, abril à julho de 2022

O Gráfico 3 apresenta os resultados das polaridades positivas e negativas obtidas por meio do dicionário léxico. Neste, observa-se uma predominância de sentimentos positivos (67%), enquanto 33% refletem opiniões de sentimentos negativos. De acordo com Becker e Tumitan (2013), esses resultados fornecem informações valiosas para eventos culturais, pois sentimentos positivos indicam sucesso nas expectativas do público, enquanto os negativos apontam aspectos que podem ser aprimorados, como logística, acessibilidade e problemas pontuais ocorridos. Por outro lado, vale reafirmar que o festival segue na direção correta, com uma aceitação de quase 70% quanto à grandiosidade que é o Festival de Parintins.

GRÁFICO 3 - Gráfico das frequências das polaridades dos tweets



Fonte: Rede social X, junho de 2023

## ANÁLISE COM CLASSIFICADOR NAIVE BAYES

Para as análises utilizando o classificador Naive Bayes, foram rotulados por humanos 1.526 posts com rótulos positivos e negativos. Em seguida, vários ajustes foram realizados, nos quais se considerou a divisão do conjunto de dados em 50% para treino e 50% para teste, 60% para treino e 40% para teste, 70% para treino e 30% para teste e 66% para treino e 34% para teste. A Tabela 1 apresenta os resultados da matriz de confusão para o ajuste com 66% para treino e 34% para teste. A partir desta, verifica-se, para a classe positiva, uma precisão de 86,1%, com revocação de 80,8%. Já para a classe negativa, observam-se 52% de precisão e 62% de revocação, em que a acurácia é de aproximadamente 76%, um resultado mediano para um modelo classificador, significando que, se tomarmos 10 posts para classificar, oito serão classificados adequadamente e dois serão classificados de forma incorreta.

Tabela 1 -Matriz de confusão de Naive Bayes para as classes de sentimento negativo e positivo

Verdadeiro	Predito	
	Negativo	Positivo
Negativo	84	51
Positivo	75	316

Fonte: O autor, com base na execução do modelo Naive Bayes (2022)

## CONSIDERAÇÕES FINAIS

Conclui-se que o presente trabalho realiza uma análise dos sentimentos e emoções referentes ao Festival Folclórico de Parintins, em que a base de dados da pesquisa foi composta por tweets, postagens de pessoas que prestigiaram a festa no período de maio ao início de julho. Os resultados das classificações dos sentimentos, obtidos por meio do dicionário léxico NRC, mostraram uma polaridade positiva, ou seja, um nível de classificação dentro do aceitável. Portanto, aproximadamente 1.200 tweets foram classificados com sentimentos positivos, e uma margem de quase 600 tweets apresentou sentimentos negativos. Além disso, verificou-se que as emoções de confiança e surpresa foram predominantes nas postagens, enquanto as emoções de raiva e desgosto apresentaram as menores ocorrências. A análise realizada com o classificador Naive Bayes apresentou um valor de acurácia de 0,7795, sendo este um indicador obtido a partir de amostras divididas em 70% para treino e 30% para teste.

## REFERÊNCIAS

- ABORISADE, O.; ANWAR, M. Classification for authorship of tweets by comparing logistic regression and naive bayes classifiers. In: IEEE INTERNATIONAL CONFERENCE ON INFORMATION REUSE AND INTEGRATION (IRI), 2018. Anais [...]. [S. l.]: IEEE, 2018. p. 269–276.
- BASILE, V. S. A.; NUNES, M. G. V. S. SentiLex-PT: um léxico de opinião para análise de sentimentos em português. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION (LREC), 8., 2014. Proceedings [...]. 2014.
- BECKER, K.; TUMITAN, D. Introdução à mineração de opiniões: conceitos, aplicações e desafios. In: SIMPÓSIO BRASILEIRO DE BANCOS DE DADOS (SBBDD). Rio Grande do Sul, 2013.
- BENEVENUTO, F.; RIBEIRO, F.; ARAÚJO, M. Métodos para análise de sentimentos em mídias sociais. In: BRAZILIAN SYMPOSIUM ON MULTIMEDIA AND THE WEB (WEBMEDIA). 2015. p. 1–30.
- BONTCHEVA, K.; ROUT, D. Making sense of social media streams through semantics: a survey. *Semantic Web*, v. 5, n. 5, p. 373–403, 2014.
- BRITO, E. M. N. Mineração de textos: detecção automática de sentimentos em comentários nas mídias sociais. 2017. Dissertação (Mestrado em Sistemas de Informação) – Universidade FUMEC, Belo Horizonte, 2017.

- BUCCI, E. A superindústria do imaginário. São Paulo: Autêntica, 2021.
- CAMARGO, S. E. S.; CABRAL, L. A. L. Uma revisão sobre técnicas de análise de sentimentos para textos em língua portuguesa. *Revista de Informática Teórica e Aplicada*, v. 26, n. 1, p. 61–81, 2019.
- CASTRO, M. Festival de Parintins nasceu da fé católica e da campanha para construir a catedral da cidade. *G1 Amazonas*, 2025.
- CNA7. Mateus Assayag destaca importância econômica do Festival de Parintins para o povo da ilha. *CNA7*, 2023.
- CULTURA AM. Caprichoso e Garantido: rivalidade de gerações com identidade cultural. 2023.
- GARANTIDO. Garantido cumpre promessa de Lindolfo Monteverde e brinca nas ruas de Parintins. 2023.
- LIDDY, E. Natural Language Processing. In: *Encyclopedia of Library and Information Science*. New York: Marcel Dekker, 2001.
- LIU, B. *Sentiment analysis: mining opinions, sentiments, and emotions*. 2. ed. Cambridge: Cambridge University Press, 2020.
- MELO, P. F. An unsupervised approach based on self-learning for the combination of sentiment analysis methods. 2019. Dissertação (Mestrado em Ciência da Computação) – UFMG, Belo Horizonte, 2019.
- MUNDO EDUCAÇÃO. Festival de Parintins. 2024.
- NASCIMENTO, A. L. P.; MENANDRO, M. C. S. Análise lexical e análise de conteúdo. *Revista Psicologia: Teoria e Pesquisa*, v. 22, n. 3, p. 319–326, 2006.
- PANG, B.; LEE, L. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, v. 2, p. 1–135, 2008.
- VEJA. Caprichoso x Garantido: o que está por trás do Festival de Parintins. *Veja*, 2024.
- XAVIER, F. et al. Análise de redes sociais como estratégia de apoio à vigilância em saúde durante a Covid-19. *Estudos Avançados*, v. 34, n. 99, p. 261–281, 2020.
- YU, Y.; DUAN, W.; CAO, Q. The impact of social and conventional media on firm equity value: a sentiment analysis approach. *Decision Support Systems*, 2013.