



UNIVERSIDADE FEDERAL DO MARANHÃO  
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

# **Análise de sentimentos em *tweets* utilizando modelos tradicionais, *deep learning* e RoBERTa**

KARLA CAROLLYNE PEREIRA MENDES

São Luís - MA, Brasil

Julho, 2025

KARLA CAROLLYNE PEREIRA MENDES

**Análise de sentimentos em *tweets* utilizando modelos tradicionais, *deep learning* e RoBERTa**

Dissertação de Mestrado submetida à Coordenação do Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Maranhão (UFMA) como parte dos requisitos para obtenção do título de Mestre em Engenharia Elétrica na área de concentração de Ciência da Computação.

Orientador: Prof. Dr. Vicente Leonardo Paucar Casas.

São Luís - MA, Brasil

Julho, 2025

KARLA CAROLLYNE PEREIRA MENDES

**Análise de sentimentos em *tweets* utilizando modelos tradicionais, *deep learning* e *RoBERTa***

Dissertação de mestrado aprovada em 25/07/2025

---

Prof. Dr. Vicente Leonardo Paucar Casas, UFMA  
(Orientador)

---

Prof. Dr. Mauro Sérgio Silva Pinto, UEMA  
(Membro da Banca Examinadora)

---

Prof. Dr. Francisco José da Silva e Silva, UFMA  
(Membro da Banca Examinadora)

São Luís - MA, Brasil  
Julho, 2025

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).  
Diretoria Integrada de Bibliotecas/UFMA

Mendes, Karla Carollyne Pereira.

Análise de sentimentos em tweets utilizando modelos tradicionais, deep learning e RoBERTa / Karla Carollyne Pereira Mendes. - 2025.

57 p.

Orientador(a): Vicente Leonardo Paucar Casas.

Dissertação (Mestrado) - Programa de Pós-graduação em Engenharia Elétrica/ccet, Universidade Federal do Maranhão, São Luís, 2025.

1. Análise de Sentimentos. 2. Processamento de Linguagem Natural. 3. Modelos Tradicionais. I. Casas, Vicente Leonardo Paucar. II. Título.

*Dedicatória:* A Deus e aos meus pais por  
me darem forças de chegar até aqui.

# Agradecimentos

A Deus, por me conceder vida, saúde e propósito. Por ter me sustentado em cada etapa desta jornada, me fortalecendo diante das dificuldades e me guiando com sabedoria.

Aos meus pais, pelo exemplo de integridade, pelos valores ensinados e pelo apoio incondicional mesmo quando os caminhos eram difíceis de entender. Por serem minha verdadeira fonte de luz e motivação, que me inspira todos os dias a ser melhor. Cada ensinamento foi essencial para que eu me tornasse quem sou hoje.

À minha família, que mesmo nos momentos mais desafiadores nunca deixou de acreditar no meu potencial.

Aos amigos e colegas que fiz ao longo da pós-graduação, João Pedro, Igor, Bruno, obrigada, pela parceria, incentivo e aprendizado compartilhado. Obrigada por tornarem esse percurso mais leve, produtivo e inesquecível.

Ao colega Higo, aluno do professor Paucar, obrigada pelo suporte na dissertação e apoio mútuo.

Aos laboratórios LSDI e PIB, na pessoa dos professores Francisco e Allan Kardec, obrigada pelos suportes todas as vezes que foram necessários.

Ao meu orientador Professor Dr. Vicente Leonardo Paucar, por sua liderança inspiradora, por cada orientação, desafio e palavra de encorajamento. Seu exemplo deixará marcas profundas na minha vida profissional e pessoal.

A todos os professores do Programa de Pós-Graduação em Engenharia de Elétrica da UFMA, minha eterna gratidão pelo comprometimento, conhecimento transmitido e pela dedicação incansável à formação de pesquisadores críticos e preparados para o futuro.

Ao setor administrativo e técnico da UFMA, pelo suporte prestativo durante toda a trajetória acadêmica — com destaque especial ao Alcides, por sua atenção e prontidão em ajudar sempre que necessário.

À CAPES, pelo apoio financeiro que foi fundamental para viabilizar a continuidade deste estudo e o desenvolvimento desta pesquisa.

Por fim, agradeço a cada pessoa que, direta ou indiretamente, contribuiu para a realização deste trabalho. Este mestrado é resultado de uma construção coletiva e afetiva, que levarei comigo com gratidão e humildade.

*"Aprendizagem nunca esgota a mente."*

(Leonardo da Vinci)

# Resumo

Nesta pesquisa, desenvolve-se uma análise de sentimentos por meio da aplicação de diferentes modelos de classificação a textos curtos e informais, como aqueles encontrados em redes sociais, especificamente no Twitter, atualmente X. O estudo contempla algoritmos tradicionais de aprendizado supervisionado, como LSTM (Long Short-Term Memory), LinearSVC, Regressão Logística, BernoulliNB e Árvore de Decisão, além de modelos mais avançados baseados na arquitetura Transformer, com destaque para o modelo RoBERTa (Robustly Optimized BERT Pretraining Approach). Esses modelos são avaliados quanto à sua capacidade de identificar, com precisão, sentimentos positivos, negativos ou neutros, considerando diferentes técnicas de pré-processamento, vetorização e representação textual. A proposta busca comparar esses modelos em diferentes contextos e estruturas linguísticas, evidenciando suas vantagens e limitações no domínio do Processamento de Linguagem Natural (PLN). Adicionalmente, são empregadas técnicas de otimização para ajuste de hiperparâmetros, com o objetivo de aprimorar o desempenho dos modelos em um conjunto de dados com um milhão de tweets. Os modelos RoBERTa e LSTM foram os que tiveram melhores resultados de acurácia com 89,84% e 79,13%, que foram melhores e competitivos com os resultados da literatura recente.

**Palavras-chave:** Análise de sentimentos, X, Processamento de Linguagem Natural, LSTM, RoBERTa, Machine Learning, Deep Learning.

# Abstract

This research aims to develop a sentiment analysis framework using different classification models applied to short and informal texts, such as those commonly found on social media platforms like Twitter. The study covers traditional supervised learning algorithms, including LSTM (Long Short-Term Memory), LinearSVC, Logistic Regression, BernoulliNB, and Decision Tree, as well as more advanced models based on the Transformer architecture, with particular emphasis on RoBERTa. These models are evaluated based on their ability to accurately identify positive, negative, or neutral sentiments, considering different preprocessing, vectorization, and textual representation techniques. The proposal seeks to compare these models across various linguistic contexts and structures, highlighting their advantages and limitations within the domain of Natural Language Processing (NLP). Additionally, optimization techniques are employed to fine-tune hyperparameters, aiming to improve model performance on a dataset consisting of one million tweets. The RoBERTa and LSTM models achieved the best accuracy results, with 89.84% and 79.13% respectively, outperforming and remaining competitive with recent results in the literature.

**Keywords:** Sentiment analysis, X, Natural Language Processing, LSTM, RoBERTa, Machine Learning, Deep Learning.

# Lista de Figuras

Figura 2.1 – Revisões de análise de sentimentos - tipos de modelos . . . . .	8
Figura 2.2 – Revisões de análise de sentimentos e aplicações de 2002 a 2021 . . . . .	9
Figura 3.1 – Metodologia Proposta . . . . .	18
Figura 3.2 – Arquitetura LSTM . . . . .	19
Figura 3.3 – Arquitetura RoBERTa . . . . .	19
Figura 3.4 – LSTM: Distribuição de dados . . . . .	21
Figura 4.1 – LSTM: Treinamento e validação da acurácia . . . . .	28
Figura 4.2 – LSTM: Perda de treinamento . . . . .	28
Figura 4.3 – LSTM: Matriz de confusão . . . . .	30
Figura 4.4 – RoBERTa: acurácia na validação . . . . .	32
Figura 4.5 – RoBERTa: Loss no treino e validação . . . . .	33
Figura 4.6 – RoBERTa: Matriz de confusão . . . . .	34

# Lista de Tabelas

Tabela 2.1 – Comparativo de Modelos de Análise de Sentimentos com o dataset Sentiment140 . . . . .	12
Tabela 2.2 – Resumo dos principais modelos utilizados na análise de sentimentos . .	14
Tabela 3.1 – Descrição do conjunto de dados. . . . .	20
Tabela 3.2 – Conjunto de dados de tweets. . . . .	20
Tabela 4.1 – Modelo da arquitetura da LSTM . . . . .	27
Tabela 4.2 – Valores da Métricas de Desempenho dos Modelos . . . . .	29
Tabela 4.3 – Valores da Métricas de Desempenho dos Modelos . . . . .	35

# Lista de abreviaturas e siglas

BERT	<i>Bidirectional Encoder Representations from Transformers</i>
BNB	<i>Bernoulli Naive Bayes</i>
GD	<i>Geração Distribuída</i>
LinearSVC	<i>Linear Support Vector Classifier</i>
LLM	<i>Large Language Model</i>
LSTM	<i>Long Short-Term Memory</i>
NLP	<i>Natural Language Processing</i>
NB	<i>Naive Bayes</i>
RL	<i>Regressão Logística</i>
RNN	<i>Recurrent Neural Network</i>
RoBERTa	<i>Robustly Optimized BERT Pretraining Approach</i>
SVM	<i>Support Vector Machines</i>

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>1</b>
1.1	Generalidades	1
1.2	Formulação do problema	2
1.3	Objetivos	3
1.3.1	Objetivos Geral	3
1.3.2	Objetivos Específicos	3
1.4	Justificativa	3
1.5	Metodologia	5
1.6	Estrutura do Trabalho	5
<b>2</b>	<b>ANÁLISE DE SENTIMENTOS E TRABALHOS RELACIONADOS</b>	<b>7</b>
2.1	Análise de Sentimentos	7
2.2	Trabalhos relacionados	10
2.3	Mercado de energia elétrica	13
2.4	Modelos selecionados para comparação	14
2.4.1	LSTM	14
2.4.2	LinearSVC	15
2.4.3	Regressão Logística (RL)	15
2.4.4	Bernoulli Naive Bayes (BNB)	15
2.4.5	Árvores de Decisão	16
2.4.6	RoBERTa	16
<b>3</b>	<b>ANÁLISE DE SENTIMENTOS DE TWEETS COM MÉTODOS TRADICIONAIS, DEEP LEARNING E TRANSFORMER ROBERTA</b>	<b>18</b>
3.1	Dados	18
3.2	Arquitetura dos Modelos	20
3.2.1	Pré-processamento	20
3.2.1.1	Word2Vec	22
3.2.1.2	Encoder	22
3.2.1.3	Embedding	22
3.2.2	Classificador de sentimentos	22
3.2.2.1	LSTM	22
3.2.2.2	LinearSVC	23
3.2.2.3	Regressão Logística (RL)	23
3.2.2.4	BernoulliNB	23
3.2.2.5	Árvore de Decisão	23

3.3	Arquitetura do Modelo Transformer RoBERTa . . . . .	23
3.3.1	Pré-processamento . . . . .	23
3.3.1.1	Leitura e Limpeza dos Dados . . . . .	24
3.3.1.2	Tokenização . . . . .	24
3.3.1.3	Divisão dos dados . . . . .	25
3.3.2	Treinamento do modelo RoBERTa . . . . .	25
3.3.3	Evolução dos Parâmetros . . . . .	25
3.3.4	Teste do Modelo RoBERTa . . . . .	26
<b>4</b>	<b>SIMULAÇÕES E RESULTADOS . . . . .</b>	<b>27</b>
4.1	Modelos Utilizados . . . . .	27
4.1.1	LSTM . . . . .	27
4.2	Modelo RoBERTa . . . . .	31
<b>5</b>	<b>CONCLUSÃO . . . . .</b>	<b>36</b>
5.1	Conclusões . . . . .	36
5.2	Limitações técnicas . . . . .	37
5.3	Considerações sobre o modelo RoBERTa . . . . .	37
5.4	Trabalhos futuros . . . . .	38
5.5	Publicação decorrente da dissertação . . . . .	38
	<b>REFERÊNCIAS . . . . .</b>	<b>39</b>

# 1 Introdução

## 1.1 Generalidades

Análise de sentimentos, também conhecida como *opinion mining*, constitui um subcampo do Processamento de Linguagem Natural (PLN) que envolve abordagens de *Deep Learning* para identificar e classificar emoções expressas em texto [1]. Em outras palavras, a análise de sentimentos avalia a polaridade e o sentimento dos textos escritos para entender se é positivo, negativo ou neutro, podendo ser aplicado em diversos contextos, seja para compreender o sentimento público em vários domínios, como negócios e política, como também para tomadas de decisão estratégicas [2].

Com os avanços no uso da Internet e o envolvimento contínuo dos usuários nas redes sociais resultaram na produção diária de uma quantidade substancial de dados. O aumento da acessibilidade proporcionou novas oportunidades e desafios, incentivando os usuários a compartilharem seus pontos de vista, retratando não apenas o conteúdo sobre os acontecimentos, mas também seus sentimentos e emoções [3–5].

Consequentemente, o twitter é uma das redes sociais mais utilizadas pelo mundo, contando com cerca de 335 milhões de usuários ativos por mês [6]. A rede permite compartilhar diversas formas de conteúdo, seja por meio dos tweets, que são publicações de mensagens textuais, ou dos retweets, onde um usuário repercute a publicação de outros usuários. [7]. Assim, o uso de dados do Twitter para análise de sentimentos tornou-se uma tendência popular. O crescente interesse na análise de mídias sociais trouxe mais atenção ao Processamento de Linguagem Natural relacionadas à análise de texto. Grande parte da literatura disponível concentra-se em textos em inglês, mas há um interesse crescente na análise multilíngue [8–10]. A análise de sentimentos pode ser feita extraíndo comentários subjetivos sobre um determinado tópico usando diferentes sentimentos, como Positivo, Negativo ou Neutro [8].

O processo de análise de sentimento abrange várias etapas essenciais, incluindo pré-processamento, extração de recursos e classificação. Na fase de pré-processamento, os dados do texto bruto passam por limpeza para remover informações irrelevantes, como palavras irrelevantes, caracteres especiais e números. Esta etapa inclui a transformação do texto em representações numéricas por meio de técnicas de incorporação de palavras, frequentemente baseadas em redes neurais. No estágio de extração de recursos, o texto processado é então classificado em sentimentos usando modelos de classificação.

O conjunto de dados sentiment140 é utilizado neste estudo, que contém 1,6 milhão de tweets, entre os quais 800,000 tweets são tweets negativos e 800.000 tweets são

anotados como positivos. Os tweets neste conjunto de dados foram originalmente anotados considerando os emoticons, por exemplo, os tweets com emoticons felizes foram considerados positivos e os tweets contendo emoticons tristes foram considerados negativos [11]. Este estudo propõe analisar termos de precisão, além disso, da comparação de desempenho entre diferentes modelos de classificação.

Diante disso, as principais contribuições nesse trabalho são:

- **Avaliação e Comparação de Diferentes Modelos de Classificação:** O estudo compara e avalia a eficácia de diversos modelos de classificação LLMs e Transformer, oferecendo uma visão detalhada de seus desempenhos de cada um. No modelos LLMs foram: LSTM, LinearSVC, Regressão Logística, BernoulliNB e Árvore de Decisão e no modelo transformer RoBERTa. O destaque especial será dado à LSTM e RoBERTa, cujos resultados obtidos foram 79% de acurácia para LSTM e 89,84% para RoBERTa.
- **Compreensão de Emoções em Redes Sociais:** Exploração detalhada da capacidade de cada modelo em identificar e categorizar nuances emocionais presentes em mensagens curtas do Twitter, contribuindo para uma melhor compreensão das emoções expressas nas redes sociais.

O restante do trabalho está organizado da seguinte forma: O Capítulo 2 explora trabalhos relacionados à análise de sentimento, que fornece uma breve descrição de estudos anteriores. O Capítulo 3 Fundamentação Teórica relacionada aos conceitos explorados para o desenvolvimento do estudo, o Capítulo 4 explorando a metodologia adotada, conjunto de dados junto com as técnicas de pré-processamento utilizadas para criar dados limpos. Também explica as técnicas e algoritmos utilizados nesta pesquisa para conduzir experimentos, o Capítulo 5 apresenta uma discussão detalhada e análise dos resultados. Por fim, o Capítulo 6 é composta pela conclusão e trabalhos futuros.

## 1.2 Formulação do problema

A proposta desta dissertação consiste em investigar técnicas para entender como avaliar a performance de modelos híbridos na base de dados de tweets ou base de dados distintos. Nesse sentido, foi elaborada a seguinte Pergunta de Pesquisa (PP):

- PP1: Quais são as melhores tipos de modelos para classificação de textos adaptáveis na base de um milhão de dados de tweets?

Este trabalho busca investigar e responder a essa questão, avaliando a eficácia de diferentes modelos híbridos aplicados à análise de sentimentos em um grande conjunto de

dados de tweets, comparando-os com modelos tradicionais e identificando suas forças e limitações.

Com o objetivo de responder a pergunta de pesquisa levantada, neste trabalho são relatados os experimentos para a avaliação de ferramentas de classificação de texto e análise de sentimentos aplicados ao tema.

## 1.3 Objetivos

### 1.3.1 Objetivos Geral

Esta dissertação tem como objetivo geral avaliar modelos LLMs e Transformer RoBERTa para análise de sentimentos em tweets, comparando seu desempenho com resultados disponíveis em modelos da literatura.

### 1.3.2 Objetivos Específicos

Especificamente, este trabalho busca os seguintes objetivos específicos:

- Revisar a literatura sobre análise de sentimentos de modelos LLMs e transformer RoBERTa aplicados a Tweets;
- Implementar diferentes modelos de análise de sentimentos, incluindo LLMs como LSTM, LinearSVC, Regressão Logística, BernoulliNB, Árvore de Decisão e modelo Transformer RoBERTa;
- Avaliar o desempenho dos modelos de análise de sentimentos com base na métrica de acurácia, utilizando o conjunto de dados Sentiment140;
- Comparar os resultados na aplicação dos modelos implementados com resultados da literatura, identificando suas vantagens e limitações de cada modelo.

## 1.4 Justificativa

A análise de sentimentos em redes sociais se tornou uma ferramenta estratégica essencial para empresas e instituições que desejam compreender o comportamento, opiniões e emoções do público em relação a produtos, serviços, eventos ou tendências. Com o crescimento das mídias sociais, especialmente o Twitter, que conta com milhões de usuários ativos e é utilizado amplamente para manifestação de opiniões, a tarefa de extrair valor a partir desses dados textuais passou a exigir técnicas mais robustas de Processamento de Linguagem Natural.

Nesse cenário, técnicas de Processamento de Linguagem Natural (PLN), como a LSTM e o RoBERTa, têm se destacado pela capacidade de lidar com textos curtos e informais, como os presentes em redes sociais. Contudo, observa-se uma lacuna na literatura: embora existam estudos que utilizam modelos tradicionais (como Regressão Logística, BernoulliNB ou mesmo LSTM) e outros que exploram modelos baseados em Transformers, são raros os trabalhos que realizam uma análise comparativa sistemática entre essas abordagens em um mesmo conjunto de dados, especialmente em domínios caracterizados por informalidade linguística, como o Twitter. Essa ausência dificulta a compreensão mais aprofundada sobre as reais vantagens e limitações técnicas de cada modelo nesse tipo de aplicação.

A proposta desta pesquisa é motivada pela constatação de que, embora existam diversos estudos que utilizam modelos tradicionais de aprendizado de máquina (como Regressão Logística, BernoulliNB e LSTM) e outros que aplicam modelos baseados em Transformers (como o RoBERTa) em tarefas de análise de sentimentos, a literatura carece de análises comparativas diretas que coloquem essas abordagens lado a lado em um mesmo contexto experimental. Essa ausência dificulta a avaliação precisa das vantagens e limitações de cada técnica, especialmente em cenários de linguagem curta, informal e contextualizada, como os textos de redes sociais.

Diante disso, este estudo busca aprimorar os métodos existentes ao realizar uma análise comparativa sistemática entre modelos tradicionais e modelos recentes, destacando seus desempenhos, pontos fortes e limitações práticas. Dessa forma, além de aplicar técnicas já consolidadas, a pesquisa contribui para o avanço do conhecimento científico ao fornecer evidências mais robustas e atualizadas sobre a adequação de cada abordagem ao domínio específico das redes sociais.

Além disso, esta dissertação aborda uma problemática ainda não totalmente resolvida: como garantir acurácia e capacidade de generalização dos modelos em um ambiente textual tão dinâmico, informal e ambíguo como o Twitter. Ao utilizar variados modelos tradicionais, bem como modelo RoBERTa, a proposta amplia o escopo da pesquisa e oferece evidências concretas sobre a importância da especialização dos modelos conforme o domínio textual.

Dessa forma, o trabalho se justifica por sua contribuição técnica, científica e prática, ao propor uma solução que melhora os modelos existentes, cobre uma lacuna atual na literatura e oferece subsídios para aplicações reais em diversas áreas que dependem da análise de sentimentos em tempo real.

## 1.5 Metodologia

A metodologia deste trabalho envolve um conjunto de métodos e técnicas adotados para alcançar os objetivos propostos, organizados em etapas sequenciais que permitam tanto o embasamento teórico quanto a validação prática da proposta. A abordagem segue os princípios da pesquisa científica, com foco na análise empírica e na comparação de desempenho entre diferentes modelos de classificação de sentimentos aplicados a tweets.

O primeiro método adotado foi a pesquisa bibliográfica, com o objetivo de realizar um levantamento do estado da arte sobre análise de sentimentos, abordagens baseadas em Processamento de Linguagem Natural, modelos de aprendizado profundo e técnicas específicas como LSTM e Transformers. Esta etapa fundamentou o embasamento teórico do trabalho, possibilitando a identificação das lacunas existentes na literatura e a definição dos modelos mais promissores para a investigação.

A metodologia científica adotada neste estudo se caracteriza como pesquisa aplicada, de abordagem quantitativa e comparativa, com o uso de dados reais da plataforma Twitter. O conjunto de dados utilizado foi o Sentiment140, composto por 1,6 milhão de tweets rotulados automaticamente como positivos ou negativos, o que possibilita a realização de testes com alto volume de dados e maior representatividade.

Para a parte experimental, foram implementados os seguintes modelos: LSTM, LinearSVC, Regressão Logística, BernoulliNB, Árvore de Decisão e RoBERTa. O processo metodológico envolveu ainda o pré-processamento dos dados (remoção de ruídos, tokenização, normalização e vetorização com técnicas como Word2Vec e Embedding), o treinamento dos modelos, validação e análise comparativa dos resultados com base em métricas como acurácia e matriz de confusão.

A combinação de pesquisa teórica com validação prática em dados reais garante à proposta deste trabalho rigor metodológico e aplicabilidade concreta no campo da análise de sentimentos em redes sociais.

## 1.6 Estrutura do Trabalho

Este trabalho está estruturado da seguinte forma:

- O Capítulo 2 descreve o conceito da análise de sentimentos e trabalhos relacionados, nisso engloba as discussões dos conceitos fundamentais de análise de sentimentos e comparação de modelos utilizados.
- O Capítulo 3 apresenta as etapas adotadas que compõem a metodologia proposta para este trabalho: qual será o conjunto de dados utilizado, pré-processamento,

os modelos implementados, arquitetura desses modelos e os critérios de avaliação adotados.

- O Capítulo 4 trata sobre os resultados obtidos e discussões em relação aos experimentos realizados, análise comparativa dos modelos e a discussão dos estudos da arte.
- O Capítulo 5 apresenta as considerações finais sobre os resultados, principais contribuições do estudo e suas limitações, artigo científico desenvolvido, bem como trabalhos futuros.

## 2 Análise de sentimentos e trabalhos relacionados

### 2.1 Análise de Sentimentos

A análise de sentimentos é uma subárea do Processamento de Linguagem Natural, que visa extrair, categorizar e analisar sentimentos, além de opiniões de textos, classificando em positivo, negativo e neutro ou até mesmo em emoções mais específicas, tais como: surpresa, confiança, raiva, medo, tristeza, nojo, alegria, entre outras. Além disso, a análise de sentimentos pode ser aplicada à análise política para entender a opinião pública em relação a partidos políticos, candidatos, bem como também pode ser utilizada no setor financeiro para analisar notícias e postagens em mídias sociais, a fim de prever preços de ações e identificar potenciais oportunidades de investimento. [12–15].

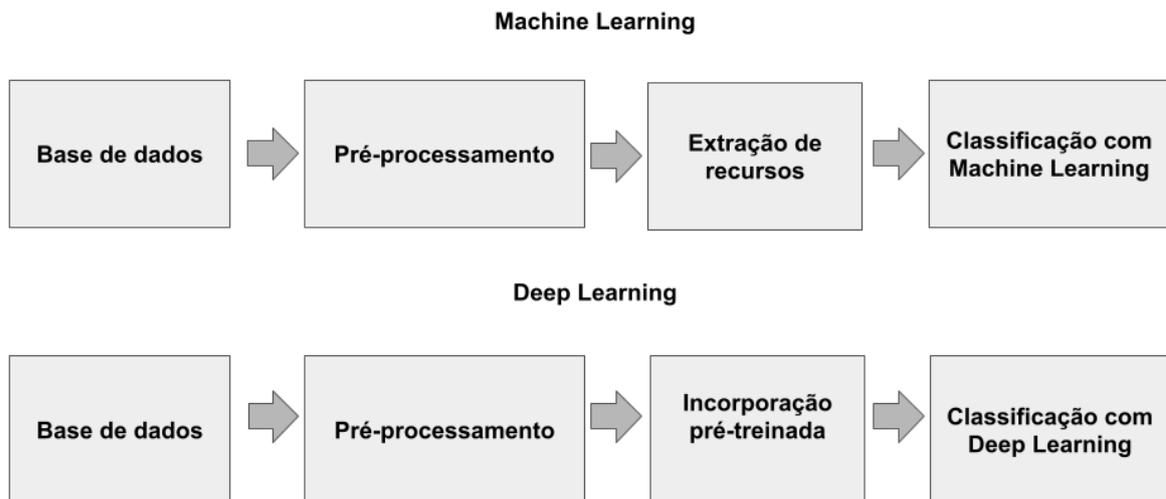
No trabalho de [16], os pesquisadores utilizaram um aplicativo de sensoriamento pessoal instalado nos *smartphones* dos participantes para coletar dados de mensagens de texto. A partir desses dados, eles calcularam sentimento de linguagem anonimizados. Essa análise foi gerada a partir das mensagens de texto enviadas pelos participantes, utilizando categorias de linguagem pré-definidas, como graus de depressão, linguagem emocional e pronomes pessoais. Essas categorias foram correlacionadas com a gravidade da depressão autorrelatada pelos participantes e o nível de sentimento presente.

Para realizar essa tarefa, os dados de texto devem passar por várias etapas, incluindo pré-processamento de dados, extração de características e classificação. O pré-processamento de dados é uma etapa importante no processo de análise de sentimentos, pois ajuda a padronizar os dados de texto e remover quaisquer elementos irrelevantes ou com ruído. Essa etapa pode incluir técnicas como lematização e remoção de stop words e caracteres especiais. Os dados de texto limpos são então transformados em recursos ou *embeddings*, que são inseridos no classificador para previsão de sentimentos. Os classificadores utilizados na análise de sentimentos podem ser em categorizados em: *Machine Learning* e *Deep Learning*.

Classificadores de *Machine Learning*, como regressão logística, *Bernoulli Naive Bayes* e máquina de vetores de suporte (SVM), utilizam modelos matemáticos para prever sentimentos. Classificadores de *Deep Learning*, como redes neurais recorrentes e modelos de memória de longo prazo (LSTM), utilizam redes neurais artificiais para fazer previsões de sentimentos. A Figura 2.1 fornece uma representação visual do processo de análise de sentimentos utilizado.

Diante disso, a escolha do classificador depende dos requisitos específicos e do caso

de uso da tarefa de análise de sentimentos, ou seja, depende do volume de dados disponível, do nível de complexidade linguística, da necessidade de interpretabilidade e dos recursos computacionais. Modelos tradicionais tendem a apresentar bom desempenho em cenários com dados limitados, baixo custo computacional e necessidade de rápida implementação. Por outro lado, modelos baseados em *Deep Learning*, embora mais exigentes em termos de processamento e tempo de treinamento, são mais indicados para conjuntos de dados extensos, linguagem contextualizada e tarefas que demandam maior robustez na detecção semânticas.

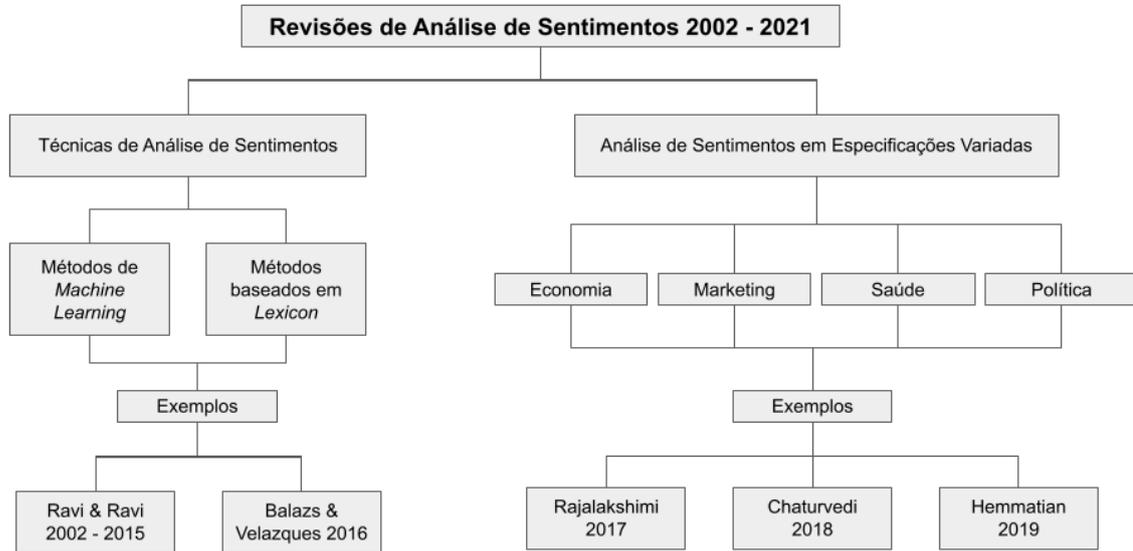


**Figura 2.1** – Revisões de análise de sentimentos - tipos de modelos

Com o crescimento do volume de dados textuais disponíveis, especialmente nas redes sociais, a análise de sentimentos tornou-se uma ferramenta essencial para que empresas e organizações compreendam a percepção do público em relação a seus produtos, serviços ou marcas [17]. Um estudo de revisões sistemáticas sobre análise de sentimentos, com foco especial em redes sociais, mostrou que, até 2020, a literatura publicada abordava principalmente duas questões: (i) as técnicas utilizadas — em especial *Machine Learning* e métodos baseados em léxico (*Lexicon*); e (ii) os domínios específicos de aplicação, como emergências, inteligência de negócios, marketing, previsão de resultados eleitorais, entre outros.

O método baseado em léxico consiste no uso de dicionários ou listas de palavras previamente associadas a polaridades (positiva, negativa ou neutra). Assim, a análise é realizada a partir da frequência e da intensidade com que essas palavras aparecem no texto, sem a necessidade de treinamento prévio de modelos de aprendizado de máquina. Essa abordagem é simples e interpretável, mas pode ser limitada diante de contextos

complexos, ironias ou gírias, especialmente em ambientes dinâmicos como as redes sociais. Um esquema de classificação dessa estrutura de artigos é apresentado na Figura 2.2.



**Figura 2.2** – Revisões de análise de sentimentos e aplicações de 2002 a 2021

Um exemplo claro para o caso de análise de sentimentos é a publicação de [18], que realiza um estudo resumindo mais de cem artigos publicados entre 2002 e 2015, com foco nas aplicações da análise de sentimentos, nas diferentes abordagens e nas questões em aberto na área. Em 2016, [19] destacaram o valor relevante da mineração de opiniões e da fusão de informações que podem ser encontradas na análise de sentimentos.

De acordo com [20], foi realizado um levantamento abrangente sobre os principais métodos empregados na análise de sentimentos, destacando seus domínios de aplicação e os desafios associados à tarefa. Complementarmente, Chaturvedi et al. (2018) [21] abordaram a questão da detecção de subjetividade, que consiste em identificar se um texto expressa uma opinião, emoção ou julgamento pessoal (*subjetivo*).

[22] revisa as técnicas de classificação aplicadas à mineração de opinião, entendida como o processo de identificar, extrair e analisar automaticamente opiniões, sentimentos e emoções expressos em textos, bem como à análise de sentimentos e à extração de características. Diferentemente desses trabalhos, [23], em 2019, obteve-se como objetivo revisar de maneira abrangente as abordagens mais significativas e difundidas que vinham sendo introduzidas no campo da análise de sentimentos por meio do aprendizado profundo. Esse estudo reuniu mais de 130 artigos da área, oferecendo um panorama detalhado das principais técnicas de *Deep Learning* utilizadas até aquele momento.

[20] se concentra em métodos de análise de sentimentos, domínios de aplicação e desafios. Além disso, a detecção de subjetividade foi analisada aprendizado profundo

trabalho extenso incluindo mais de 130 artigos de pesquisa, que fornece uma pesquisa detalhada das técnicas de aprendizado profundo mais populares na época.

Modelos Tradicionais como Máquinas de Vetores de Suporte (SVM) e Naive Bayes, foram utilizados para a classificação de sentimentos. Contudo, esses modelos apresentam limitações ao tentar capturar dados mais complexas, como ironia, sarcasmo e ambiguidade. Essa dificuldade motivou o desenvolvimento e a adoção de modelos mais avançados de *deep learning*, como as Redes Neurais Recorrentes (RNN) e, mais recentemente, os modelos baseados em *transformers*, como o BERT e o RoBERTa.

## 2.2 Trabalhos relacionados

As abordagens de *Machine Learning* e *Deep Learning* na análise de sentimentos começam com a padronização dos dados de texto por meio do pré-processamento e limpeza de informações irrelevantes. Em seguida, técnicas de extração de características são aplicadas para representar o texto como características numéricas que podem ser inseridas em classificadores de *Machine Learning*. Alguns dos classificadores comumente utilizados em análise de sentimentos incluem máquina de vetores de suporte (SVM), *Naive Bayes*, regressão logística, floresta aleatória e árvores de decisão. Tais modelos, adotados na literatura, demonstram eficácia em tarefas de categorização de sentimentos, mas enfrentam limitações na interpretação de contexto e ironia.

Este subcapítulo apresenta uma visão detalhada dos principais métodos e modelos utilizados na análise de sentimentos, com foco nas abordagens de *machine learning* e *deep learning*. Estudos mostram que a análise de sentimentos tem se mostrado uma ferramenta essencial no processamento de linguagem natural (PLN), especialmente com o avanço dos modelos de transformers, como o RoBERTa, no qual será destaque nesta pesquisa.

Nesse contexto, [24] propuseram uma investigação aprofundada sobre o desempenho de modelos *LLMs* — *Large Language Models* aplicados à tarefa de análise de sentimentos em textos extraídos do Sentiment140, conjunto de dados com 1,6 milhão de tweets rotulados automaticamente. A motivação do estudo se baseia na crescente demanda por modelos que consigam captar semânticas em textos curtos, informais e ambíguos, típicos das redes sociais.

O estudo adotou uma abordagem baseada em técnicas de classificação com *Deep Learning*, com modelos da arquitetura *Transformer*, uma das mais eficientes no campo do Processamento de Linguagem Natural (PLN). O objetivo foi medir o desempenho desses modelos na tarefa de classificar sentimentos entre positivo e negativo, utilizando a acurácia como métrica de comparação.

Os modelos testados por [24] e seus resultados foram: LLaMA\_7b\_v2, com acurácia

de 50%; GPT-3.5-turbo com acurácia de 62%; GPT-4 com acurácia de 64%; GPT-4o com acurácia de 72%; Doubao com acurácia de 75%; Doubao-pro com acurácia de 77%. Todos os modelos foram classificados como LLMs e aplicaram metodologias de *Deep Learning* para processar os dados. A adoção da arquitetura Transformer permite que esses modelos capturem relações contextuais complexas e dependências semânticas entre palavras, sendo especialmente eficazes em tarefas onde o entendimento do tom e da intenção do autor é essencial.

Os resultados apresentados acima indicaram que modelos mais recentes e otimizados, como o Doubao-pro, superam versões anteriores como o LLaMA\_7b\_v2, demonstrando o impacto direto de melhorias na arquitetura, no número de parâmetros e no treinamento específico para o domínio da linguagem informal. Esses achados reforçam a tendência de substituição de modelos clássicos de *Machine Learning* por abordagens baseadas em LLMs, especialmente em aplicações mais sofisticadas de análise de sentimentos.

No trabalho de [25], os autores propuseram uma abordagem clássica de análise de sentimentos utilizando algoritmos de *Machine Learning* na mesma base de dados Sentiment140. O foco do estudo foi comparar o desempenho de diferentes modelos de classificação, além de propor um modelo de votação e um ensemble próprio para aumentar a acurácia preditiva da classificação de sentimentos.

Foram testados os seguintes algoritmos: Naive Bayes, com acurácia de 75,19%; Random Forest, com acurácia de 71,76%; Support Vector Machine (SVM), com acurácia de 75,61%; Logistic Regression, com acurácia de 74,15%; Proposed Ensemble, um modelo de combinação otimizada proposto pelos autores, que alcançou a maior acurácia do estudo: 75,81%.

Todos os classificadores utilizaram representações vetoriais do texto, geradas a partir do pré-processamento e da extração de características como unigramas e bigramas. O estudo reforça a relevância dos métodos de ensemble para o aumento da performance preditiva em tarefas de análise de sentimentos, mostrando que a combinação de diferentes modelos pode superar abordagens individuais, inclusive em domínios com ruído e linguagem informal como o Twitter.

A pesquisa de [25] destaca, portanto, o potencial das técnicas de *Machine Learning* tradicionais em tarefas de classificação de sentimentos, mesmo quando comparadas a modelos mais recentes. Ainda assim, seus resultados se aproximam da performance de LLMs intermediários como o GPT-3.5-turbo, indicando que, embora os modelos clássicos ofereçam boa acurácia, os LLMs otimizados superam essas abordagens com maior robustez e sensibilidade contextual.

Além disso, [26] compararam métodos de *machine learning* and *deep learning* para análise de sentimentos usando dois conjuntos de dados: Sentiment140 e Twitter US Airline

Sentiment. O conjunto de dados foi pré-processado usando tokenização e limpeza de caracteres como URL, hashtag e pontuação. Os métodos comparados foram *multinomial naive Bayes*, regressão logística, máquina de vetores de suporte *support vector machine - SVM*, LSTM e um conjunto de *Naive Bayes* e regressão logística. No conjunto de dados Sentiment140, o LSTM alcançou 66%.

Por fim, [27] conduziram um estudo comparativo para avaliar o desempenho de quatro modelos de aprendizado de máquina para análise de sentimento: máquina de vetores de suporte, multinomial naive Bayes, LSTM e BERT. O conjunto de dados sentiment140, que consiste em 1,6 milhão de Tweets classificados como positivos ou negativos, foi utilizado no estudo. Após pré-processar os dados usando tokenização, lematização, lematização e remoção de stop-words e pontuação, os resultados mostraram que o modelo BERT teve o melhor desempenho, alcançando 85,4% de precisão e superando os demais modelos.

Também foi feita uma análise comparativa com outros estudos que utilizaram o conjunto de dados de referência sentiment140, resumido na Tabela 2.1.

**Tabela 2.1** – Comparativo de Modelos de Análise de Sentimentos com o dataset Sentiment140

Literatura	Classificador	Tipo de Modelo	Acurácia (%)
[24]	LLaMA_7b_v2	LLM	50.00
	GPT-3.5-turbo	LLM	62.00
	GPT-4	LLM	64.00
	GPT-4o	LLM	72.00
	Doubao	LLM	75.00
	Doubao-pro	LLM	77.00
[25]	Naive Bayes	Machine Learning	75.19
	Random Forest	Machine Learning	71.76
	SVM	Machine Learning	75.61
	Logistic Regression	Machine Learning	74.15
	Proposed Ensemble	Ensemble	75.81
[26]	LSTM	Deep Learning	66.00
[27]	BERT	LLM	85.40

Assim, observa-se que, neste capítulo, foi apresentada uma análise dos principais métodos e modelos aplicados a análise de sentimentos, com ênfase nas abordagens de *machine learning* e *deep learning*. Estudos mostraram que análise de sentimentos tem se consolidado como uma ferramenta essencial no contexto do Processamento de Linguagem Natural (PLN), principalmente diante do crescimento de dados gerados em redes sociais. Com o avanço dos modelos baseados em arquiteturas do tipo *Transformer*, como o BERT (Bidirectional Encoder Representations from Transformers) e suas variantes mais recentes — incluindo o RoBERTa —, tornou-se possível capturar maiores performances. Neste cenário, o modelo RoBERTa será aprofundado nesta pesquisa por sua capacidade de

interpretar textos curtos e informais, características predominantes em corpora como o Sentiment140.

## 2.3 Mercado de energia elétrica

Na área do mercado de energia elétrica, a análise de sentimentos surge como um campo para sua aplicação. No Brasil, a *Geração Distribuída* (GD), sobretudo a solar fotovoltaica, vem transformando o cenário energético desde a Resolução Normativa nº 482 da ANEEL, de 2012, que instituiu o sistema de compensação de energia elétrica. Esse marco regulatório possibilitou que consumidores se tornassem também produtores, os chamados “prosumidores”, alterando as relações de consumo e produção [28].

Grande parte da literatura existente sobre GD foca em aspectos técnicos, regulatórios e econômicos, como descentralização da matriz elétrica, sustentabilidade e eficiência do sistema [29, 30]. Contudo, a percepção do consumidor comum ainda é pouco explorada, mesmo sendo fator determinante na adoção dessas tecnologias. O consumidor avalia sua satisfação principalmente em termos práticos, como redução da conta de energia, retorno do investimento e qualidade no atendimento das distribuidoras [31].

As redes sociais desempenham papel essencial nesse processo, uma vez que não se configuram mais como um espaço restrito a especialistas, mas como canais abertos de expressão da experiência cotidiana de milhões de usuários. Nessas plataformas, consumidores compartilham sentimentos de entusiasmo diante da economia proporcionada pela energia solar, mas também frustrações relacionadas a tarifas, regulamentações e suporte técnico. Esse conjunto de percepções representa um insumo valioso para compreender a aceitação social da GD.

Estudos internacionais indicam que a adoção de sistemas fotovoltaicos em residências é mais influenciada por fatores subjetivos, como percepção de benefícios, inovação e preocupação ambiental, do que por características sociodemográficas clássicas [32]. Dessa forma, a aplicação da análise de sentimentos em dados de redes sociais pode contribuir para identificar tendências de comportamento, avaliar níveis de satisfação e fornecer insumos para políticas públicas e estratégias empresariais.

Segundo [33], a análise de sentimentos vai além da simples classificação binária entre positivo e negativo, sendo capaz de identificar emocionais e relações contextuais. No contexto do setor elétrico, essa abordagem permite ampliar a visão técnica com uma leitura mais humana, fundamentada nas experiências e percepções do consumidor comum.

Assim, integrar a análise de sentimentos ao estudo da geração distribuída no Brasil oferece um caminho para alinhar inovação, formulação de políticas públicas e expectativas sociais, tendo assim um setor energético mais inclusivo e orientado ao usuário final.

## 2.4 Modelos selecionados para comparação

Este capítulo apresenta uma análise comparativa dos principais métodos utilizados na literatura para de análise de sentimentos. Foram considerados desde os modelos tradicionais de *Machine Learning*, amplamente aplicados em contextos supervisionados, até as abordagens mais recentes baseadas em *Deep Learning* e modelos de linguagem de larga escala (*LLMs*) com arquiteturas do tipo *Transformer*. Destacam-se, nesse cenário, modelos como o LSTM e, especialmente, o RoBERTa e suas variantes, que têm demonstrado desempenho superior na classificação de sentimentos em textos curtos, ruidosos e informais, como os encontrados em plataformas como o Twitter. A capacidade de captura contextual desses modelos justificam sua escolha como objeto de aprofundamento nesta pesquisa. Abaixo, na Tabela 2.2 verifica-se compreender melhor os principais modelos utilizados na pesquisa.

**Tabela 2.2** – Resumo dos principais modelos utilizados na análise de sentimentos

Modelo	Tipo de Modelo	Arquitetura Principal	Características Notáveis
LSTM	Deep Learning	Rede Neural Recorrente (LSTM)	Captura dependências temporais e sequenciais de longo prazo. [34]
LinearSVC	Machine Learning	Máquina de Vetores de Suporte (Linear)	Boa performance em dados vetorizados com margens lineares. [35] e [25]
Regressão Logística	Machine Learning	Modelo Estatístico de Classificação	Eficiente em bases balanceadas e com variáveis independentes. [35]
BernoulliNB	Machine Learning	Classificador Probabilístico (Bernoulli)	Simples e rápido, bom desempenho em textos curtos binários. [27,36]
Árvore de Decisão	Machine Learning	Árvore de Decisão Binária	Interpretável, porém limitado para dados ruidosos. [37]
RoBERTa	LLM	Transformer com Masked Language Modeling	Modelagem contextual profunda com embeddings dinâmicos. [38]

### 2.4.1 LSTM

As redes neurais do tipo *Long Short-Term Memory (LSTM)*, introduzidas por [34], destacam-se por sua capacidade de capturar dependências de longo prazo em seqüências textuais, superando limitações de redes recorrentes tradicionais (RNNs), como o desaparecimento do gradiente. Essa característica as torna particularmente adequadas para tarefas de Processamento de Linguagem Natural (PLN), como a análise de sentimentos.

Em [39], os autores exploram uma abordagem híbrida, combinando dicionários léxicos com camadas empilhadas de LSTM, o que resultou em um desempenho superior aos métodos clássicos de classificação. O estudo evidencia que o LSTM é capaz de aprender automaticamente representações relevantes dos dados de entrada, reduzindo a necessidade de engenharia manual de atributos e aumentando a eficácia na detecção de sentimentos em textos curtos e informais.

### 2.4.2 LinearSVC

O *Linear Support Vector Classifier* (LinearSVC) é uma variante linear da Máquina de Vetores de Suporte (SVM), amplamente utilizada em tarefas de classificação binária no contexto de *Machine Learning*. Sua principal característica é a capacidade de encontrar o hiperplano ótimo que separa classes distintas em um espaço vetorial de alta dimensão, maximizando a margem entre os dados rotulados.

Na análise de sentimentos, o LinearSVC tem se mostrado eficiente, sobretudo quando combinado com técnicas de vetorização textual, como unigramas, bigramas e representações TF-IDF. Devido à sua natureza linear, o modelo é computacionalmente mais leve do que versões kernelizadas da SVM, tornando-o adequado para aplicações em larga escala ou com restrições de processamento.

Segundo [35] e [25], o LinearSVC obteve resultados expressivos na tarefa de classificação de sentimentos em tweets, superando algoritmos mais simples como o Naive Bayes e a Regressão Logística. Os estudos indicam que o desempenho do LinearSVC está fortemente ligado à qualidade da extração de características, sendo especialmente sensível à escolha da representação vetorial dos textos.

Embora apresente boa generalização em contextos lineares, o modelo pode ter limitações em tarefas que envolvam ambiguidade semântica ou linguagem altamente subjetiva, como sarcasmo, para as quais modelos contextuais como os baseados em *Transformers* tendem a ser mais adequados.

### 2.4.3 Regressão Logística (RL)

É um dos modelos estatísticos mais empregados na análise de sentimentos, principalmente por sua capacidade de modelar relações lineares entre variáveis explicativas e variáveis-alvo binárias. No estudo de [35], a RL foi comparada com outros algoritmos de classificação e demonstrou desempenho competitivo, especialmente em bases de dados balanceadas, utilizando unigramas e bigramas como atributos.

### 2.4.4 Bernoulli Naive Bayes (BNB)

É um classificador probabilístico simples, mas eficaz, sobretudo quando aplicado a textos curtos e com vocabulário bem delimitado. [36] aplicaram a versão binária do Naive Bayes em um ambiente de streaming de dados, obtendo bons resultados com seleção de atributos. Já [27] observaram que, embora o desempenho do BNB seja inferior a modelos mais modernos como BERT e LSTM, ele se destaca pelo baixo custo computacional e simplicidade de implementação.

### 2.4.5 Árvores de Decisão

Também foram avaliadas por [37], que empregaram a técnica em conjunto com a representação TF-IDF para classificar sentimentos em múltiplos conjuntos de dados. Os autores destacam a alta interpretabilidade das árvores, mas também apontam suas limitações em generalizar para dados ruidosos ou não estruturados, como os encontrados em tweets.

### 2.4.6 RoBERTa

Os modelos baseados na arquitetura *Transformer*, como o RoBERTa, têm ganhado destaque nos últimos anos por sua robustez e alto desempenho em tarefas de Processamento de Linguagem Natural (PLN). Proposto por [38], o RoBERTa é uma versão otimizada do BERT [40], que aprimora aspectos fundamentais do pré-treinamento para aumentar sua eficácia. As principais melhorias incluem o uso de um volume significativamente maior de dados, um número mais elevado de épocas e a remoção da tarefa de *Next Sentence Prediction* (NSP), focando exclusivamente na tarefa de *Masked Language Modeling* (MLM).

A arquitetura *Transformer*, base estrutural do RoBERTa, é caracterizada pelo mecanismo de atenção (self-attention), que permite ao modelo capturar relações contextuais complexas entre palavras, independentemente de sua posição no texto. Essa capacidade garante aos modelos *Transformer* uma vantagem significativa em relação às redes neurais recorrentes, como o LSTM, sobretudo na interpretação de textos mais longos ou com estruturas semânticas sofisticadas [41].

Diferente dos modelos tradicionais de representação vetorial de palavras, como o Word2Vec, que geram embeddings estáticos, o RoBERTa produz representações dinâmicas ajustadas ao contexto da sentença. Essa característica permite uma compreensão mais refinada da linguagem, o que o torna especialmente eficaz em domínios informais, como a análise de sentimentos em tweets.

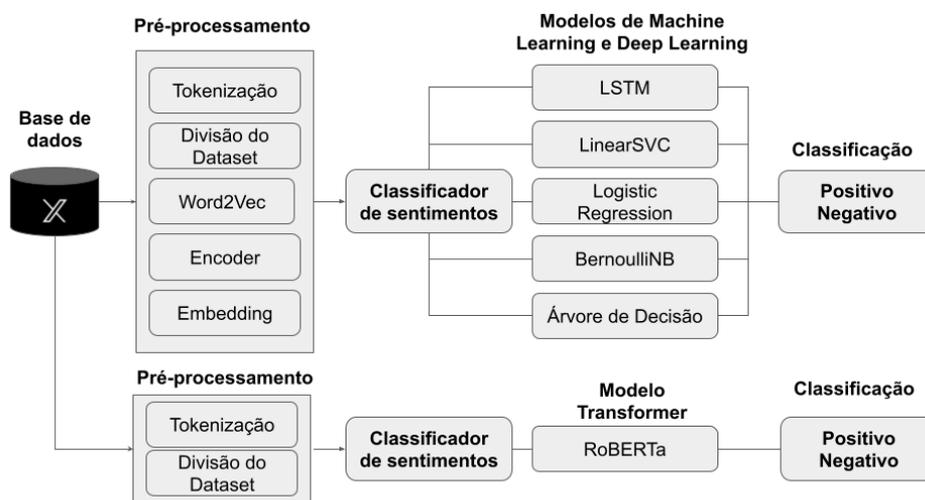
Em um estudo comparativo conduzido por [27], o RoBERTa demonstrou desempenho superior em relação a modelos tradicionais e redes neurais simples, como o Naive Bayes e o LSTM. Sua capacidade de compreender nuances semânticas e lidar com linguagem informal o posiciona como uma das abordagens mais promissoras para tarefas de classificação de sentimentos.

Assim, considerando seu desempenho e sua arquitetura, o modelo RoBERTa se destaca como objeto de aprofundamento nesta pesquisa, sobretudo diante dos desafios característicos da análise de sentimentos em textos curtos, como aqueles encontrados no Twitter. Esses desafios incluem características presentes no dataset, como volumetria dos dados, presença de linguagem informal, uso de gírias e abreviações, limitação de caracteres, além de ambiguidade semântica, ironia e sarcasmo, aspectos que impactam no treinamento.

A utilização do modelo RoBERTa nesta pesquisa buscou enfrentar parte dessas limitações, aproveitando sua capacidade de capturar representações contextuais por meio do mecanismo de atenção (*self-attention*), o que lhe permite lidar melhor com variações linguísticas e identificar significados em textos curtos. Ao realizar uma comparação direta entre diferentes modelos, esta pesquisa contribui para evidenciar de como o modelo RoBERTa supera as restrições observadas em abordagens anteriores.

### 3 Análise de sentimentos de *tweets* com métodos tradicionais, *deep learning* e transformer RoBERTa

Este estudo tem como objetivo classificar os sentimentos dos tweets através de diferentes modelos de classificação e analisar a precisão de cada um. A metodologia proposta envolve o uso de várias técnicas que são brevemente descritas na seção seguinte. A Figura 3.1 representa como será feito.



**Figura 3.1** – Metodologia Proposta

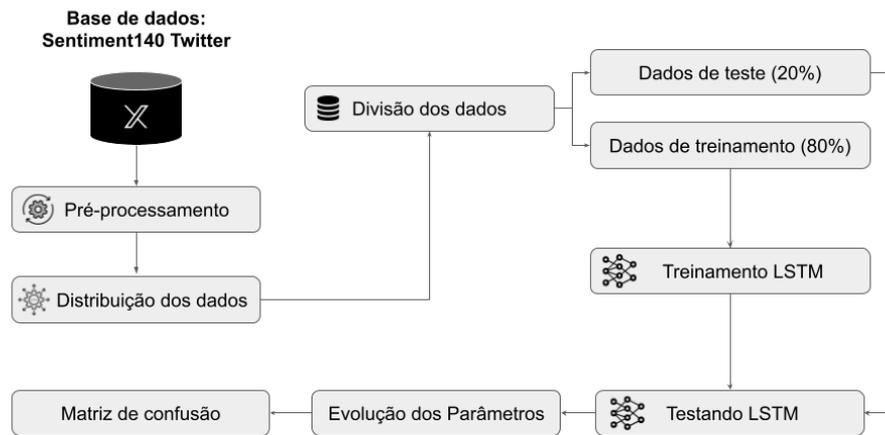
De forma precisa, observa-se como procedeu a arquitetura do modelo LSTM proposto, no qual o modelo LSTM será destaque aos demais modelos citados acima. A Figura 3.2 representa abaixo:

Por fim, também pode-se verificar na Figura 3.3 como se deu a arquitetura do modelo RoBERTa proposto, no qual será destaque no estudo:

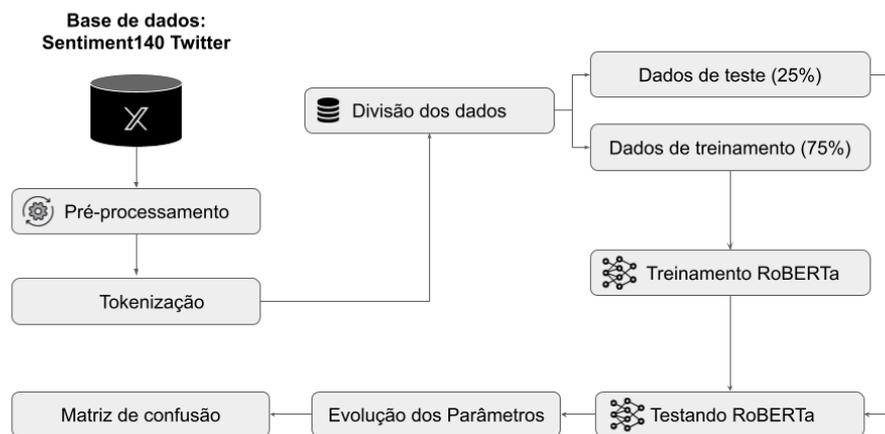
A comparação de desempenho entre os modelos LSTM e RoBERTa será discutida em detalhes na seção a seguir, apresenta-se os dados utilizados para treinamento e a preparação dos mesmos para análise de sentimentos.

#### 3.1 Dados

O conjunto de dados escolhido é o sentiment140 [42, 43], composto por 1,6 milhões de tweets obtidos do Twitter, cada um rotulado como positivo, negativo ou neutro. Possui



**Figura 3.2** – Arquitetura LSTM



**Figura 3.3** – Arquitetura RoBERTa

800.000 tweets positivos e 800.000 tweets negativos, cada um rotulado como 0 para sentimento negativo, 2 para neutro e 4 para sentimento positivo.

O que diferencia o Sentiment14 [42, 43] é sua abordagem para rotulagem: ele usa emoticons nos tweets (*como :-) ou :-(*) como indicadores de sentimento, que são então removidos para criar uma representação mais autêntica do conteúdo de rede social. Este conjunto de dados é particularmente valioso para lidar com os desafios de texto ruidoso, incluindo abreviações, erros de ortografia e linguagem informal típica do Twitter. O Sentiment140 [42, 43] é utilizado para treinar e avaliar modelos de análise de sentimentos, especialmente aqueles projetados para analisar dados de rede social.

Criado em 2009 por alunos da Universidade Stanford, essa rotulação possibilita a detecção automatizada de sentimentos, tornando o conjunto de dados útil e valioso para pesquisas em análise de sentimentos em redes sociais. No método empregado, a equipe

assumi que tweets contendo emoticons positivos, como :), eram considerados positivos, enquanto tweets com emoticons negativos, como :(, eram considerados negativos.

A abordagem dos autores proporciona uma perspectiva única e automatizada para a criação de dados de treinamento, destacando a eficácia do uso de emoticons como indicadores de sentimento. É importante destacar que, anotar manualmente os tweets teria sido uma tarefa trabalhosa e demorada devido a quantidade. Os autores do conjunto de dados anotaram os tweets considerando o ruído dos emoticons para a previsão do tweet como positivo ou negativo. [44]

O conjunto de dados consiste em seis recursos que são listados e descritos na Tabela 3.1. Algumas amostras de tweets do conjunto de dados são mostradas na Tabela 3.2.

**Tabela 3.1** – Descrição do conjunto de dados.

Características	Descrição
<b>target</b>	Polaridade do tweet (0 = negativo, 2 = neutro, 4 = positivo)
<b>id</b>	ID do tweet
<b>date</b>	Data de publicação do tweet
<b>flag</b>	A consulta (lyx). Se não houver consulta, o valor será SEM CONSULTA.
<b>user</b>	Nome de usuário que postou o tweet
<b>text</b>	Texto escrito no tweet

**Tabela 3.2** – Conjunto de dados de tweets.

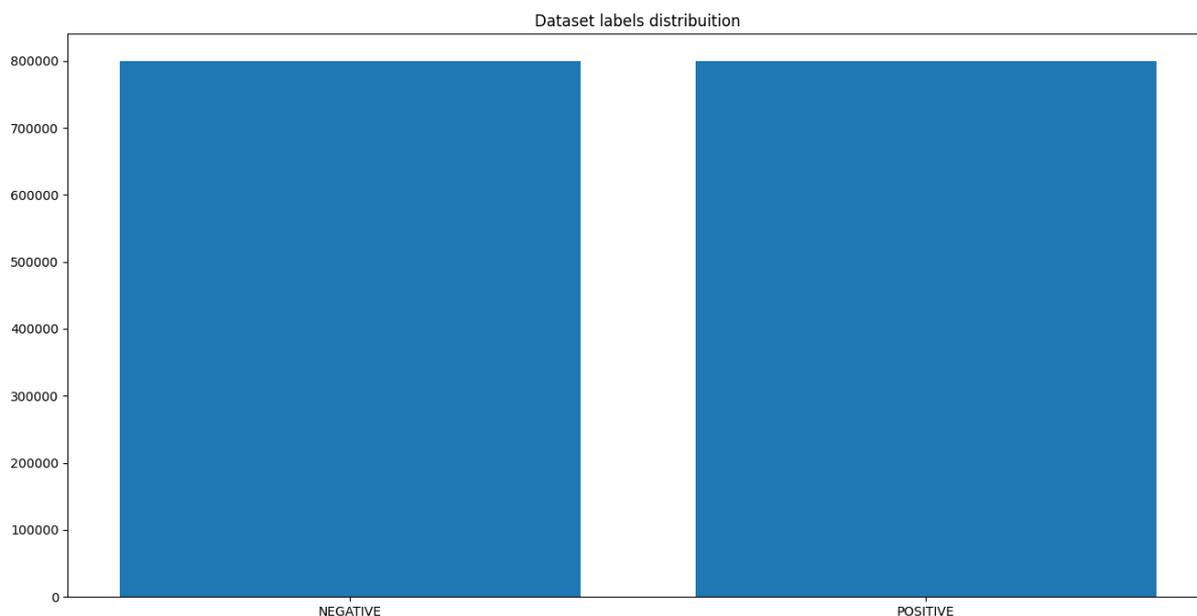
No.	Target	Text
1	0	@Kenichan I dived many times for the ball.
2	0	@octolinz16 It counts, Idk why I did either. you never talk to me anymore
3	4	@ElleCTF I would like to call you at night after dinner

A seguir, verifica-se o pré-processamento dos dados para o treinamento dos modelos LLMs e Transformers. Esta etapa é essencial para obtermos boa performance durante o treinamento dos modelos e qualidade dos resultados. Destaca-se quais foram as técnicas adotadas, com o objetivo compararmos quais modelos performaram melhor na mesma base de dados treinada.

## 3.2 Arquitetura dos Modelos

### 3.2.1 Pré-processamento

Após a leitura dos dados, se obtém a distribuição dos tweets para verificar a equivalência entre as classes de sentimentos. O objetivo era garantir que o número de tweets positivos e negativos fosse balanceado. A Figura 3.4 ilustra a distribuição dos dados, evidenciando que ambos os sentimentos estão igualmente representados, com 800.000 tweets em cada categoria. Isso assegura que o conjunto de dados não apresente distorções, facilitando a tarefa de modelagem.



**Figura 3.4** – LSTM: Distribuição de dados

Verifica-se na Figura 3.4, o conjunto de dados está equilibrado, com uma quantidade igual de tweets positivos e negativos, o que é um fator importante para garantir a imparcialidade no treinamento do modelo.

Para garantir uma análise eficaz dos dados, realizou-se um processo de preparação dos dados, com o objetivo de remover inconsistências e facilitar o treinamento dos modelos. A ferramenta *Natural Language Toolkit* (NLTK) foi utilizada para o pré-processamento, seguindo as etapas descritas a seguir:

- (1) Transformação de texto para minúsculo;
- (2) Remoção de *Stop Words*<sup>1</sup>;
- (3) Remoção de caracteres numéricos;
- (4) Remoção de pontuações ou caracteres especiais

Além disso, as três classes de sentimentos originais do banco de dados foram simplificadas para duas categorias: positiva e negativa, para facilitar o treinamento e a análise.

Em seguida, os dados pré-processados foram divididos em dois conjuntos: *Training Data* (dados de treinamento) e *Test Data* (dados de teste). A proporção dos conjuntos foi definida com 80% dos dados para treinamento, ou seja, 1.280.000 exemplos, e 20% para teste, com 320.000 exemplos. Para melhorar a representação das palavras e suas relações

<sup>1</sup> Stopwords: palavras muito frequentes em textos (como artigos, pronomes e preposições) que não acrescentam significado relevante ao sentimento expresso. No presente trabalho, foram removidas no pré-processamento para reduzir ruídos e melhorar o desempenho dos modelos de classificação.

semânticas, utilizou-se as técnicas word2vec, encoder e embedding, conforme detalhado nas subseções a seguir.

### 3.2.1.1 Word2Vec

Utilizando a biblioteca Gensim, treinou-se o modelo Word2Vec nos textos dos tweets, ajustando parâmetros como o tamanho do vetor, a janela de contexto e a contagem mínima de ocorrências. O vocabulário resultante contém 30.369 palavras. A partir dessa representação vetorial, foi possível explorar a similaridade semântica entre palavras, evidenciando suas relações. Também foi realizada a tokenização dos textos e a padronização do comprimento dos tweets, convertendo-os em sequências numéricas fixas. Esse processo preparou os dados para a etapa seguinte de classificação de sentimentos, demonstrando a aplicabilidade do Word2Vec na análise de sentimentos em dados de redes sociais.

### 3.2.1.2 Encoder

O Encoder transforma dados textuais em formato numérico. Neste estudo, utilizamos o Encoder foi utilizado para lidar com as categorias de sentimentos do conjunto de dados. Após identificar as categorias ('POSITIVE', 'NEGATIVE', 'NEUTRAL'), empregamos a classe LabelEncoder do pacote scikit-learn para criar um encoder. O encoder associa valores numéricos únicos a cada categoria de sentimento. Os resultados são apresentados em arrays, onde as dimensões refletem o número de exemplos no conjunto de dados. Por exemplo, um array de dimensão (1280000, 1) indica a presença de 1.280.000 rótulos de classes no conjunto de treinamento.

### 3.2.1.3 Embedding

A técnica de embedding foi utilizada para representar palavras ou tokens em um espaço vetorial contínuo, de menor dimensão. Essa técnica é essencial para capturar as relações semânticas e contextuais entre as palavras, permitindo que o modelo entenda melhor os significados implícitos no texto.

## 3.2.2 Classificador de sentimentos

Para a classificação dos sentimentos dos tweets, foram utilizados vários modelos de Deep Learning, sendo eles: LSTM, LinearSVC, Regressão Logística, BernoulliNB e Árvore de Decisão.

### 3.2.2.1 LSTM

O modelo foi estruturado da seguinte forma: a) uma camada LSTM com 100 neurônios, incluindo uma taxa de dropout de 0.2 e dropout recorrente de 0.2 para

regularização; b) última camada com um neurônio e função de ativação sigmoide para gerar a saída.

#### 3.2.2.2 LinearSVC

Utilizando a técnica de Support Vector Machines (SVM) com kernel linear, o LinearSVC foi treinado com base nos dados de treinamento disponíveis. Ele determina separadores lineares no espaço vetorial e facilita a separação de diferentes categorias de dados vetoriais de entrada. Após a obtenção do hiperplano, os recursos do texto extraídos podem ser colocados no classificador para prever os resultados. Além disso, a ideia central é encontrar uma linha mais próxima dos vetores de suporte.

#### 3.2.2.3 Regressão Logística (RL)

Classificador baseado em estatística usado principalmente para a análise de dados binários nos quais uma ou mais variáveis são usadas para encontrar os resultados. Também é usado para avaliar a probabilidade de associação de classe [45]. Durante o estudo, os parâmetros ajustados para o modelo incluem um parâmetro de regularização C com valor 2, que controla a força da regularização, bem como um número máximo de iterações configurado para 1000.

#### 3.2.2.4 BernoulliNB

No estudo, este modelo foi aplicado como classificador binário com um parâmetro de suavização de 2 ( $\alpha=2$ ). Após o treinamento com os dados correspondentes, sua performance foi avaliada utilizando o método `model_Evaluate(BNBmodel)`.

#### 3.2.2.5 Árvore de Decisão

Árvore de Decisão foi empregada utilizando a classe `DecisionTreeClassifier` do pacote `scikit-learn`. O modelo foi treinado com o conjunto de treinamento, aprendendo padrões nas características dos dados relacionados à análise de sentimentos.

### 3.3 Arquitetura do Modelo Transformer RoBERTa

#### 3.3.1 Pré-processamento

O pré-processamento de dados é uma etapa essencial no treinamento de modelos, especialmente em tarefas de Processamento de Linguagem Natural (NLP), onde a qualidade dos dados de entrada tem um impacto na performance do modelo. Será apresentado como foi realizado o processo de pré-processamento para o treinamento do modelo RoBERTa, desde a preparação da base de dados até a divisão dos dados para treinamento e teste.

Vale destacar que, conforme apresentado no fluxo ilustrado nas Figuras 2.1 e 2.2, o pré-processamento foi estruturado de maneira a garantir que os dados estivessem prontos para as fases subsequentes de tokenização, treinamento e avaliação do modelo RoBERTa.

Embora o pré-processamento do modelo RoBERTa siga uma abordagem similar ao que foi apresentado nos modelos LLMs, é importante destacar que o RoBERTa é projetado para lidar de maneira mais eficiente com entradas complexas, aproveitando as características avançadas de sua arquitetura. A seguir, serão apresentados os detalhes do tratamento e ajustes aplicados aos dados, visando termos resultados melhores no desempenho do modelo RoBERTa durante o treinamento.

### 3.3.1.1 Leitura e Limpeza dos Dados

Inicialmente, é importante destacar que a base de dados utilizada é a mesma, Sentiment140 [42, 43] Twitter, que contém tweets rotulados com sentimentos positivos e negativos. A partir dessa base, os dados foram carregados utilizando a biblioteca pandas, que oferece flexibilidade na manipulação de dados.

Os rótulos da coluna target, que inicialmente continham os valores 0 (negativo) e 4 (positivo), foram convertidos para um formato binário, cada tweet é classificado com um rótulo: 0 para sentimentos negativos e 4 para sentimentos positivos.

### 3.3.1.2 Tokenização

Após a limpeza e mapeamento dos rótulos, a próxima etapa do pré-processamento foi a tokenização dos textos. O modelo RoBERTa é baseado na arquitetura de transformers e foi treinado para compreender e processar grandes volumes de texto.

Assim, visto que o modelo RoBERTa requer que o texto seja dividido em unidades menores, chamados tokens, que podem ser entendidas e processadas pelo modelo. Foi utilizado o AutoTokenizer da biblioteca transformers, que é específico para o modelo RoBERTa. Este tokenizer divide os textos em tokens, que são as unidades de processamento para o modelo. A tokenização foi configurada para garantir que as sequências de tokens geradas tivessem um tamanho máximo de 128 tokens, com padding (preenchimento) adicionado aos textos menores e truncamento nos textos mais longos, para que todos os textos fossem uniformemente processados.

O uso do tokenizer de RoBERTa também garantiu que as palavras fossem corretamente transformadas em suas representações numéricas, com base no vocabulário aprendido pelo modelo durante o pré-treinamento. Esse processo de tokenização é uma das etapas mais importantes, pois a qualidade da tokenização afeta diretamente a qualidade das representações geradas pelo modelo.

### 3.3.1.3 Divisão dos dados

Após o pré-processamento e a tokenização, os dados são divididos em dois conjuntos: dados de treinamento e dados de teste. A divisão ocorre da seguinte forma:

Dados de treinamento (75%): São utilizados para treinar o modelo a realizar a tarefa de análise de sentimentos. O modelo aprende a partir desses dados, ajustando seus parâmetros para prever corretamente os sentimentos nos tweets.

Dados de teste (25%): São reservados para avaliar o desempenho do modelo após o treinamento. O modelo é avaliado com base nesses dados para verificar sua capacidade de generalizar e classificar corretamente os sentimentos de tweets não vistos durante o treinamento.

Essa divisão é importante para garantir que o modelo não aprenda apenas os dados de treinamento, mas também seja capaz de generalizar para novos dados, evitando overfitting.

### 3.3.2 Treinamento do modelo RoBERTa

Com os dados de treinamento preparados, o próximo passo é o treinamento do modelo RoBERTa. O modelo é treinado para aprender a relação entre os textos dos tweets e seus respectivos sentimentos (positivo ou negativo). Durante o treinamento, o modelo ajusta seus parâmetros internos com base na minimização da função de perda (loss function), que mede a discrepância entre as previsões do modelo e os rótulos reais dos dados de treinamento.

O treinamento é realizado utilizando a arquitetura do modelo RoBERTa, que é um modelo de transformer. Ele usa o mecanismo de atenção para aprender as relações contextuais entre palavras em uma sequência de texto, o que permite ao modelo compreender o significado completo de um tweet, mesmo que ele contenha palavras fora de contexto.

### 3.3.3 Evolução dos Parâmetros

Durante o treinamento, o modelo passa por várias épocas, em que os parâmetros internos são ajustados. Ao longo dessas épocas, o modelo melhora suas previsões e se ajusta cada vez mais aos dados de treinamento. A evolução dos parâmetros refere-se ao processo de atualização dos pesos do modelo com base nas perdas calculadas em cada iteração de treinamento.

Além disso, o código implementa um monitoramento das métricas do modelo durante o treinamento, incluindo acurácia, precisão, recall e F1-score, para acompanhar como o modelo está melhorando a cada época.

### 3.3.4 Teste do Modelo RoBERTa

Após o treinamento, o modelo é testado utilizando o conjunto de dados de teste. O objetivo dessa etapa é avaliar a capacidade do modelo de generalizar o aprendizado para dados que ele não viu durante o treinamento. Durante o teste, as métricas de desempenho são calculadas para verificar a acurácia do modelo na classificação dos sentimentos dos tweets.

## 4 Simulações e Resultados

Foram realizados experimentos para atestar a qualidade dos modelos de classificação de sentimentos. Os resultados foram categorizados para análise de modelos LLMs, Transformers e comparação entre os mesmos. A seguir, foram detalhados os experimentos realizados, com foco nos modelos LSTM e RoBERTa, além de uma comparação com os resultados de modelos anteriores e com outros autores.

### 4.1 Modelos Utilizados

O código fonte, o treinamento e o dataset utilizado na experimentação se encontram disponíveis no repositório deste projeto no Github <sup>1</sup>. Uma gama diversificada de experimentos foi conduzida envolvendo vários modelos, LSTM, LinearSVC, Regressão Logística, BernoulliNB, Árvore de Decisão e RoBERTa. O propósito de realizar uma variedade de experimentos é adquirir a mais alta precisão relativa à classificação de sentimento do conjunto de dados sentiment140. Nesta seção, foi realizada uma comparação entre a abordagem proposta e estudos anteriores conduzidos com o conjunto de dados Sentiment140.

#### 4.1.1 LSTM

A primeira etapa dos experimentos foi realizada com o modelo LSTM (*Long Short-Term Memory*). Na Tabela 4.1 apresenta a arquitetura, que se inicia com uma camada de Embedding, responsável por transformar os dados de entrada em representações vetoriais. Em seguida, há uma camada de Dropout, que serve para prevenir overfitting, seguida pela camada LSTM composta por 100 unidades. Por fim, uma camada dense com ativação sigmoide é utilizada para a saída binária.

**Tabela 4.1** – Modelo da arquitetura da LSTM

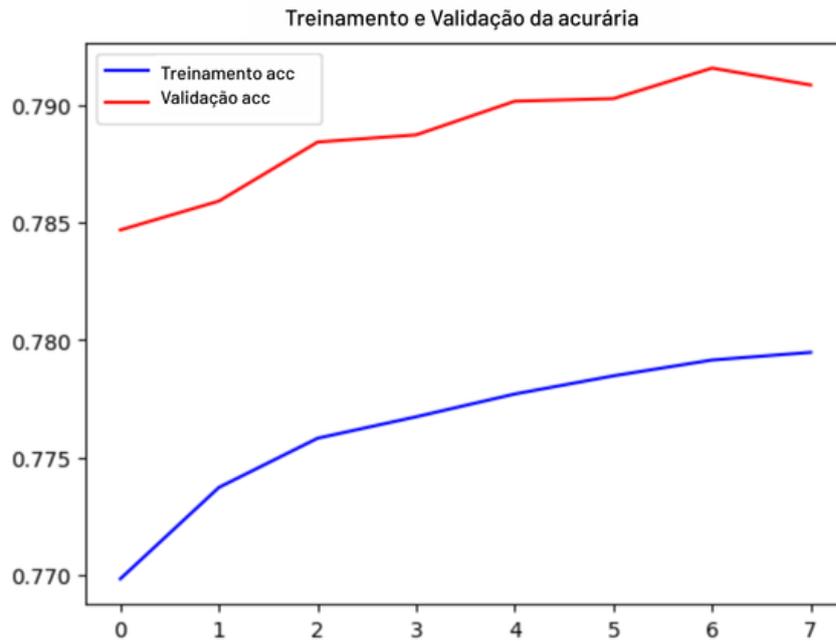
Tipo de camada	Formato de saída	Parâmetro #
embedding (Embedding)	(None, 300, 300)	87,125,700
dropout (Dropout)	(None, 300, 300)	0
lstm (LSTM)	(None, 100)	160,400
dense (Dense)	(None, 1)	101
Parâmetros Totais: 87,286,201		
Parâmetros Treináveis: 160,501		
Parâmetros não treináveis: 87,125,700		

utilizando o ambiente computacional do Google Colab, que disponibiliza recursos de GPU para acelerar o processamento. Durante esse período, a perda (loss) foi reduzida

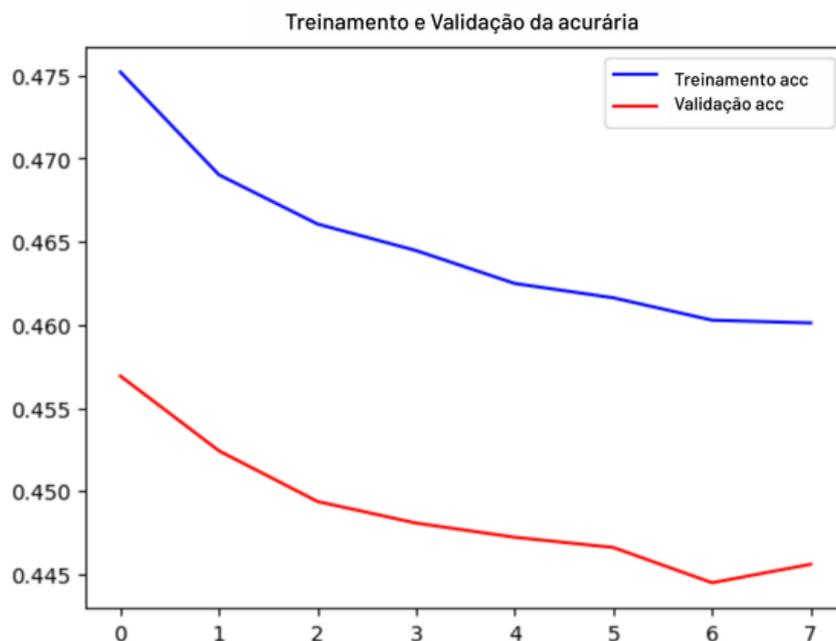
<sup>1</sup> <https://github.com/karlamendes0/Analise-de-sentimentos-de-tweet.git>

de 0,4752 para 0,4601. Os resultados da avaliação do modelo apresentaram uma acurácia de 79,13% no conjunto de teste, com uma perda (loss) de 0,444.

Os gráficos de precisão e perda durante o treinamento, apresentados nas Figuras 4.1 e 4.2, mostram um crescimento inicial na precisão, seguido de uma leve estagnação, indicando possível overfitting. Esse comportamento sugere que ajustes nos hiperparâmetros e a introdução de regularização poderiam melhorar a capacidade de generalização do modelo.



**Figura 4.1** – LSTM: Treinamento e validação da acurária



**Figura 4.2** – LSTM: Perda de treinamento

Nas Figuras 4.1 e 4.2 observa-se a evolução das métricas durante o treinamento e a perda de treinamento. Ela exibe a trajetória do modelo ao longo das épocas durante o treinamento. O eixo das abscissas representa o número de épocas, enquanto o eixo das ordenadas indica os valores das métricas.

Observa-se, na Figura 4.1 um crescimento inicial na precisão, tanto no conjunto de treinamento quanto no de validação, indicando aprendizado e generalização. Contudo, a precisão de validação tende a estagnar ou declinar após um pico, o que sugere a ocorrência de *overfitting*, conforme ilustrado na Figura 4.2. Em relação à perda, verifica-se uma redução inicial em ambos os conjuntos, mas a perda de validação se estabiliza ou aumenta, revelando uma possível falta de generalização do modelo. A discrepância entre os conjuntos de treinamento e validação aponta para uma superestimação nos dados de treinamento, o que sugere necessidades de ajustes, como regularização ou mudanças nos hiperparâmetros, visando melhorar a capacidade de generalização do modelo.

Enquanto a Figura 4.3 proporciona uma representação visual através da Matriz de Confusão da capacidade do modelo em distinguir entre classes, revelando as previsões corretas e as possíveis fontes de erros.

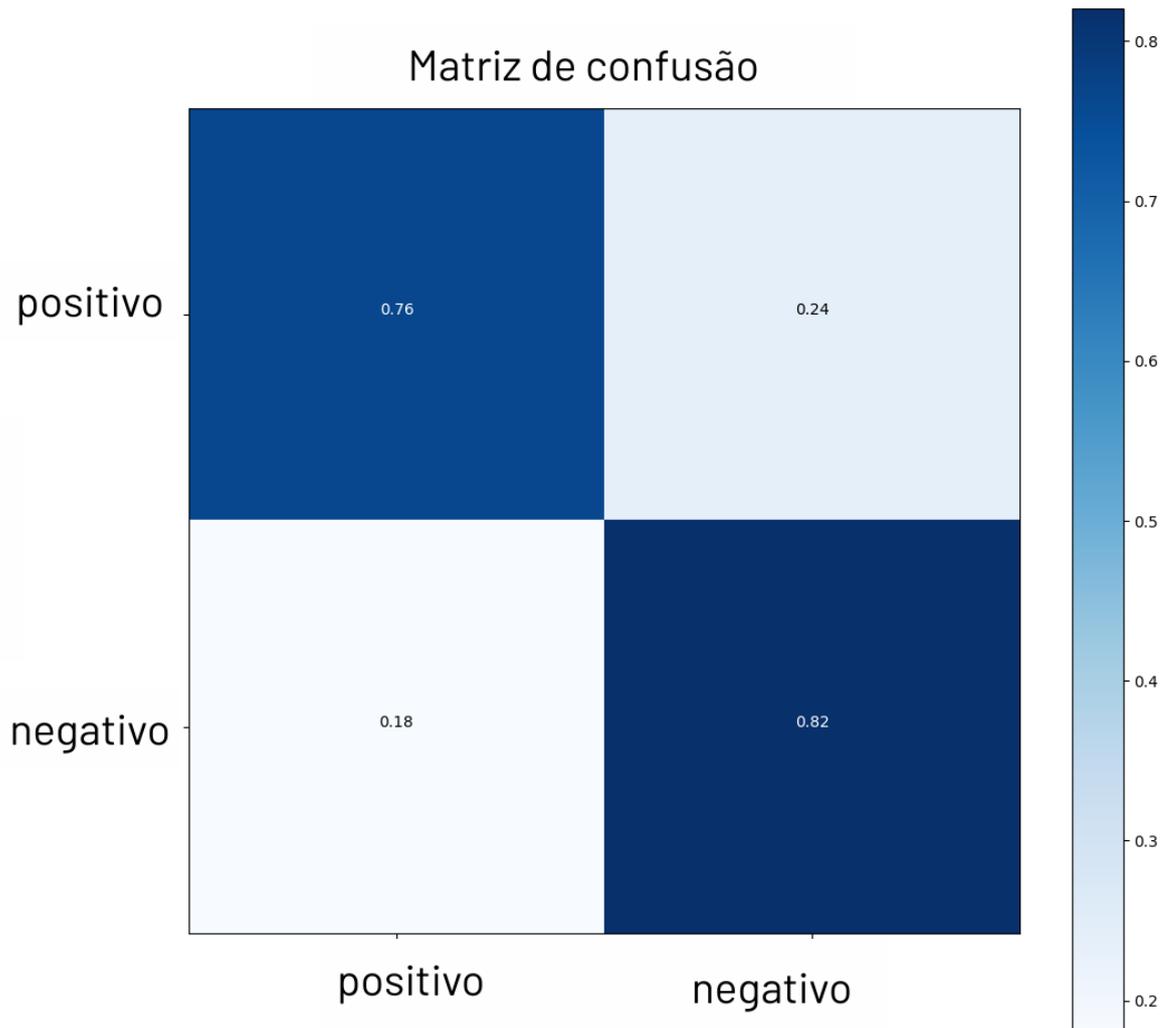
Após o treinamento do modelo LSTM, foram realizados experimentos com os demais modelos, incluindo LinearSVC, Regressão Logística, BernoulliNB e Árvore de Decisão. Na Tabela 4.2 apresenta o relatório de classificação com as métricas de desempenho de cada modelo.

**Tabela 4.2** – Valores da Métricas de Desempenho dos Modelos

Modelo	Acurácia (%)
LSTM	79,13
LinearSVC	50,36
Regressão Logística	50,17
BernoulliNB	51,38
Árvore de Decisão	59,96

Baseado nisto, observa-se que, entre os modelos LLMs treinados, o modelo LSTM exibiu resultados notáveis, com uma acurácia de 79,13% enquanto os demais modelos tiveram resultados variados. Essa variação pode ser atribuída a diversos fatores relacionados tanto à natureza dos modelos quanto às características específicas do conjunto de dados. As razões pelas quais o LSTM obteve um desempenho superior são as seguintes:

LinearSVC, Regressão Logística, BernoulliNB e Árvore de Decisão são modelos mais simples, que, por serem lineares ou baseados em probabilidades simples, têm limitações para capturar padrões complexos em dados de texto. Esses modelos não conseguem lidar com dependências temporais e contextuais, que são cruciais para a análise de sentimentos, o que se reflete no desempenho inferior em comparação ao LSTM.



**Figura 4.3** – LSTM: Matriz de confusão

Arquitetura do Modelo: A arquitetura do modelo LSTM é mais completa que as dos demais, podendo capturar padrões nos dados de treinamento e generalizar bem para os dados de teste, permitindo que possa aprender e entender as relações temporais e contextuais entre as palavras em uma frase, o que é essencial para tarefas de processamento de linguagem natural (PLN). Esse nível de complexidade é necessário para que o modelo entenda o contexto completo da sentença e, por conseguinte, faça classificações mais precisas.

Tamanho e Qualidade dos Dados: A quantidade e a qualidade dos dados de treinamento têm um impacto no desempenho de qualquer modelo. O LSTM foi treinado em um conjunto de dados representativo, com uma boa distribuição de classes de sentimento (positivo e negativo). Modelos como LinearSVC e Regressão Logística podem sofrer mais com dados desbalanceados ou com conjuntos de dados de menor qualidade, enquanto o

LSTM consegue aprender melhor os padrões mesmo em dados mais complexos.

**Hiperparâmetros Ajustados:** O desempenho do LSTM também foi influenciado pela escolha cuidadosa dos hiperparâmetros, como o tamanho da rede neural, a taxa de aprendizado, o número de épocas de treinamento e o uso de regularização (como Dropout). A escolha apropriada desses hiperparâmetros ajudou o modelo a se ajustar melhor aos dados, evitando overfitting e permitindo que ele generalizasse bem para novos dados. Esses ajustes são cruciais, especialmente em modelos mais complexos como o LSTM, que necessitam de um equilíbrio cuidadoso para atingir um bom desempenho.

**Pré-processamento dos dados:** A forma como o texto é pré-processado (tokenização, remoção de stopwords, normalização) e como as palavras são representadas (embedding, como Word2Vec ou GloVe) tem grande impacto na capacidade do modelo de capturar informações contextuais importantes. O LSTM, por ser um modelo sequencial, consegue melhor processar as dependências entre palavras e entender o contexto completo das frases. Outros modelos como LinearSVC e Regressão Logística podem falhar em capturar as relações mais sutis entre palavras em frases longas ou complexas.

Portanto, o LSTM se destacou por sua capacidade de capturar relações temporais e contextuais nos dados de texto, algo essencial em tarefas de análise de sentimentos. Já os outros modelos tiveram desempenho inferior por conta de suas arquiteturas e tratamento dos dados.

## 4.2 Modelo RoBERTa

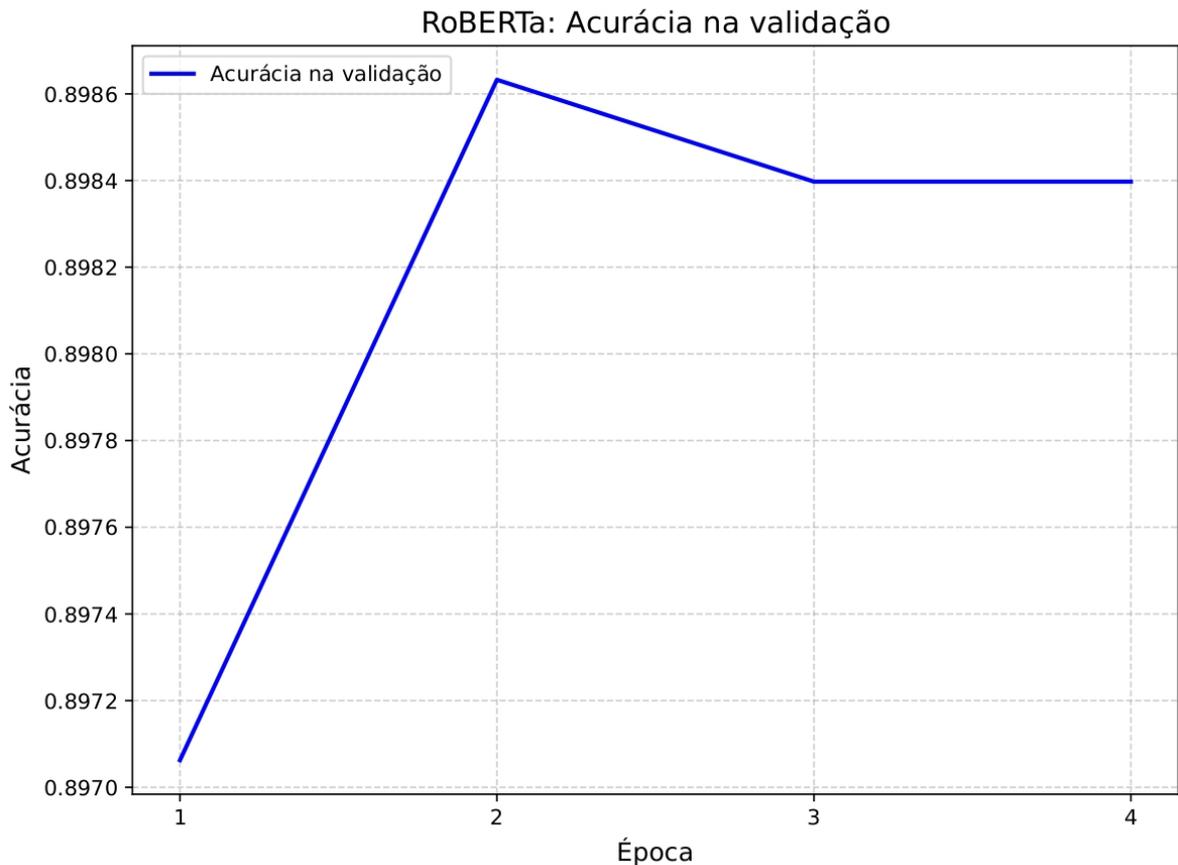
O modelo RoBERTa, após ser treinado e avaliado, gerou resultados satisfatórios, como demonstrado pelas métricas de avaliação e pelos gráficos gerados durante o processo. A seguir, descrevem-se os principais resultados e as métricas associadas, com base nas imagens fornecidas e no código implementado <sup>2</sup>.

Após a avaliação do modelo LSTM, foi realizado o treinamento com o modelo RoBERTa, que, como mencionado, é uma versão robusta do BERT. O modelo foi treinado por 4 épocas, seguindo as melhores práticas do artigo original sobre o BERT [40]. Essa escolha de 4 épocas foi fundamentada na literatura, que sugere que essa quantidade de épocas proporciona uma boa convergência do modelo, garantindo melhores resultados, podendo ser observado a seguir nas Figuras 4.4 e 4.5.

Na Figura 4.4, o gráfico de acurácia apresenta que o modelo RoBERTa obteve aumento na acurácia após a primeira época, atingindo acurácia de 89,84% ao final do treinamento. Esse aumento rápido indica que o modelo teve ótimo desempenho, o que é característico de modelos Transformers como o RoBERTa. A estabilidade da acurácia ao

---

<sup>2</sup> <https://github.com/karlamendes0/Analise-de-sentimentos-com-RoBERTa>



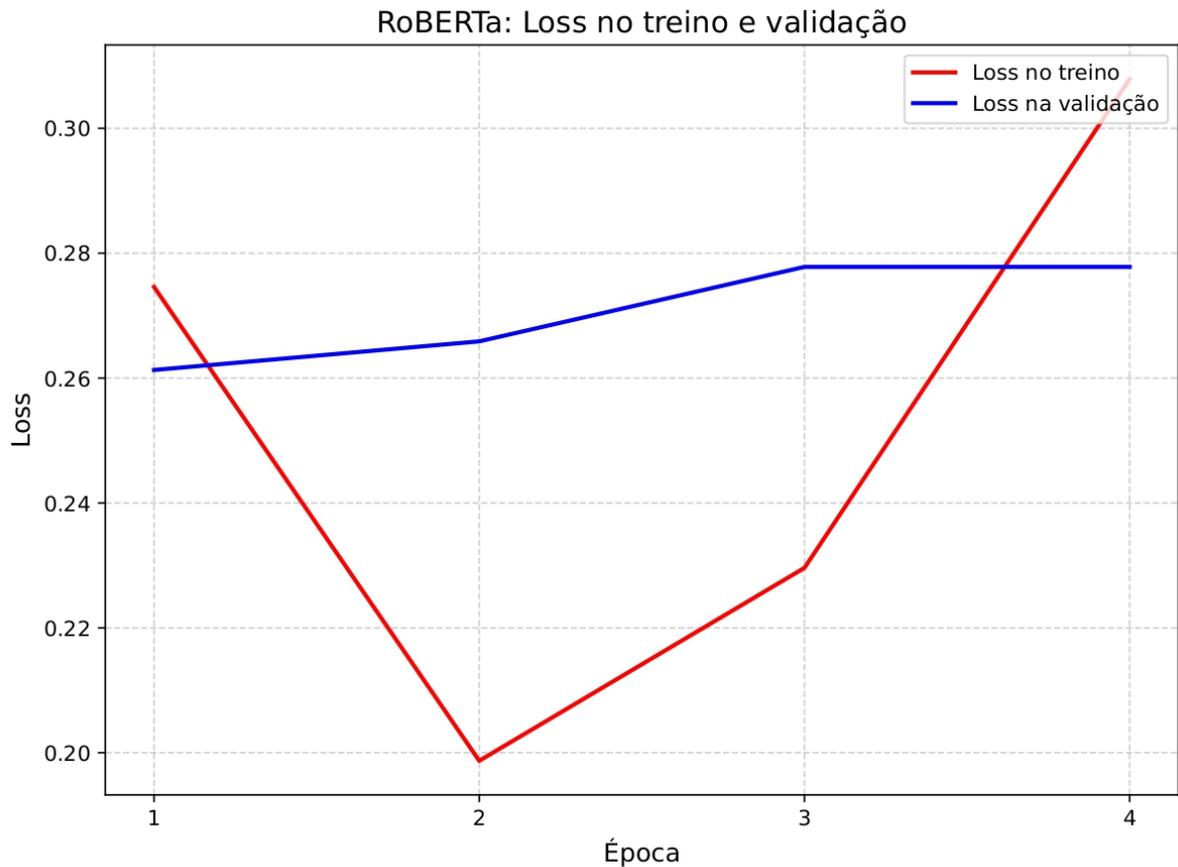
**Figura 4.4** – RoBERTa: acurácia na validação

longo das 4 épocas sugere que o modelo não está sofrendo de overfitting, o que é um bom indicativo de que o treinamento foi bem-sucedido em termos de generalização.

Além disso, a análise da performance da acurácia nas épocas pode ser feita considerando os seguintes pontos: crescimento acelerado e estabilização. No crescimento inicial acelerado, a primeira época, a acurácia começa em torno de 0.8970 e experimenta um aumento considerável, alcançando 0.8986 ao final da segunda época. Esse crescimento sugere que o modelo está aprendendo rapidamente as características do conjunto de dados, ajustando seus pesos de forma eficaz nas primeiras iterações. Enquanto a partir da terceira época, a curva de acurácia se estabiliza, mantendo-se em torno de 0.8986, sem mais aumentos significativos. Isso indica que o modelo atingiu um ponto de convergência no aprendizado, onde as melhorias na acurácia são mínimas, mesmo após mais épocas de treinamento. Esse comportamento é comum em modelos de aprendizado profundo, indicando que o modelo pode ter aprendido todos os padrões relevantes dos dados disponíveis ou que há limitações no conjunto de dados ou hiperparâmetros que precisam ser ajustados para melhorar a performance.

Em termos práticos, a acurácia está boa e pode ser considerada como um ponto de equilíbrio entre treinamento eficiente e convergência de desempenho. Para melhorar o

desempenho, no qual será citado em trabalhos futuros, será considerado ajustes finos nos hiperparâmetros.



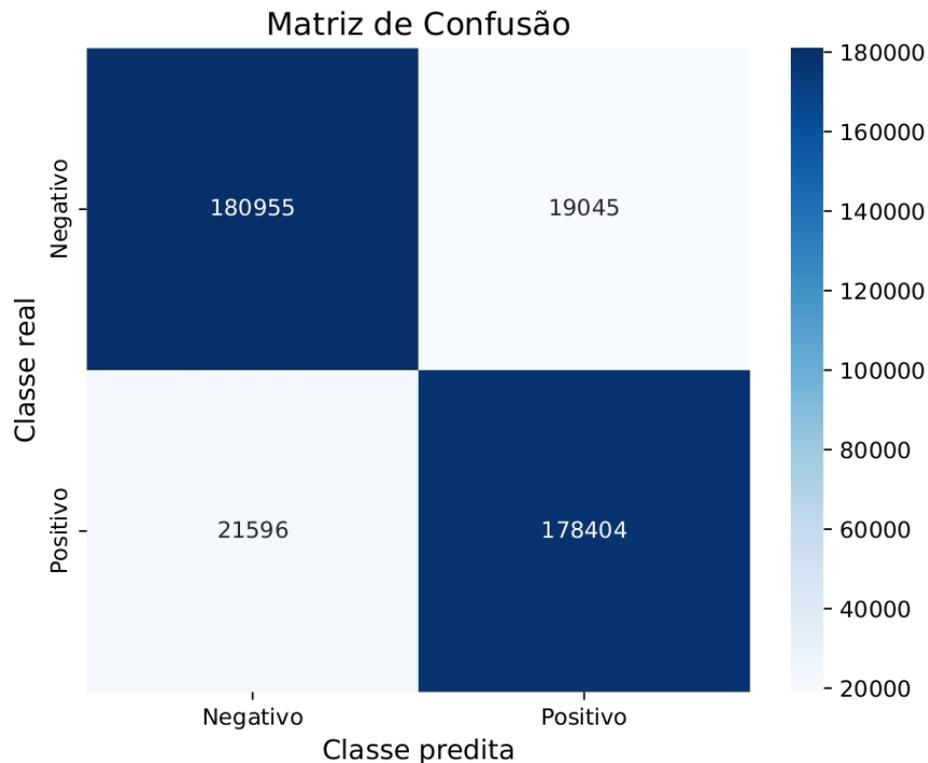
**Figura 4.5** – RoBERTa: Loss no treino e validação

Na Figura 4.5 mostra o gráfico de loss durante o treinamento e a validação. O loss de treinamento (em vermelho) diminui nas primeiras épocas, indicando que o modelo está aprendendo de maneira eficiente e ajustando seus parâmetros para reduzir o erro nas previsões durante o treinamento. Contudo, a perda de validação (em azul) apresenta um leve aumento após a segunda época, o que sugere overfitting gradual. Esse comportamento é esperado, e a leve discrepância entre o loss de treinamento e validação pode ser corrigida com ajustes nos hiperparâmetros, para encontrar o ponto ótimo de treinamento, no qual será também destaque nos trabalhos futuros.

Ainda é possível observar ponderações em relação ao *loss* (ou função de perda) no treinamento e na validação. Para a linha vermelha, representada pela função de perda no treinamento, tem-se uma queda até a segunda época, indicando que o modelo estava aprendendo rapidamente. O loss inicialmente começa em torno de 0.26, caindo para 0.20 no final da segunda época. Após a segunda época, o loss no treinamento começa a aumentar novamente, chegando a valores acima de 0.28 nas épocas seguintes. Esse aumento pode sugerir que o modelo está começando a overfit, ou seja, ele está se ajustando muito aos dados de treinamento e não generalizando tão bem para os dados de validação.

Enquanto na linha azul, que representa o loss na validação, segue uma trajetória diferente. Inicialmente, o loss na validação começa em um valor relativamente alto e permanece estável até a segunda época. A partir da terceira época, o loss de validação começa a subir drasticamente, o que é um indicativo claro de overfitting. Ou seja, o modelo parece estar perdendo a capacidade de generalizar para dados que não foram vistos durante o treinamento, com o loss de validação ultrapassando o loss de treinamento.

Por fim, é possível destacar ainda a análise de desempenho do modelo RoBERTa através da matriz de confusão, visto na Figura 4.6, no qual demonstra um equilíbrio satisfatório entre os Verdadeiros Positivos e os Verdadeiros Negativos, o que indica que o modelo é eficaz na análise entre sentimentos positivos e negativos na maioria dos casos. A alta taxa de acerto em ambos os casos, positivos e negativos, demonstra que o modelo tem uma boa capacidade de performance. A análise de matriz se dá por: verdadeiro positivos, falso positivos, falsos negativos, verdadeiros negativos. Verifica-se com mais detalhes a seguir:



**Figura 4.6** – RoBERTa: Matriz de confusão

Verdadeiros Positivos (TP): O modelo previu corretamente 178.404 instâncias como positivas, que realmente eram positivas. Falsos Positivos (FP): O modelo classificou 19.045 instâncias como positivas, que na realidade eram negativas. Esse erro ocorre quando o modelo sobrestima a quantidade de sentimentos positivos em relação ao que realmente está presente. Falsos Negativos (FN): O modelo classificou 21.596 instâncias como negativas, que na realidade eram positivas. Esse tipo de erro representa uma falha do modelo em

identificar corretamente as instâncias com sentimento positivo. Verdadeiros Negativos (TN): O modelo corretamente identificou 180.955 instâncias como negativas, que de fato eram negativas.

Logo, os valores de Falsos Positivos e Falsos Negativos fornecem áreas importantes para melhorias. Os Falsos Positivos (19.045) indicam que o modelo tende a classificar algumas instâncias negativas como positivas, sugerindo a necessidade de ajustes no threshold de decisão ou no balanceamento das classes durante o treinamento, para reduzir esse erro. Por outro lado, os Falsos Negativos (21.596) mostram que o modelo falha ao identificar sentimentos positivos, o que pode ser corrigido com ajustes no processo de treinamento ou com uma reavaliação das amostras de dados.

Assim sendo, observa-se na Tabela 4.3 que o modelo RoBERT alcança resultado de acurácia de 89,84%, sendo superior ao modelo LSTM que obteve 79,13%. Os erros apresentados na matriz de Falsos Positivos e Falsos Negativos indicam áreas onde o modelo poderá ser melhorado, como por exemplo, ajustes nos hiperparâmetros e análise detalhada dos dados de treinamento, que será visto em posteriormente em trabalhos futuros.

**Tabela 4.3** – Valores da Métricas de Desempenho dos Modelos

<b>Modelo</b>	<b>Acurácia (%)</b>
Regressão Logística	50,17
LinearSVC	50,36
BernoulliNB	51,38
Árvore de Decisão	59,96
LSTM	79,13
RoBERTa	89,84

# 5 Conclusão

## 5.1 Conclusões

Este estudo propôs uma abordagem abrangente para a classificação de sentimentos em mensagens do Twitter, utilizando uma variedade de modelos de Modelos LLMs, incluindo LSTM, LinearSVC, Regressão Logística, BernoulliNB, Árvore de Decisão, e o modelo Transformer, com destaque para o RoBERTa.

O processo envolveu a coleta e pré-processamento dos dados do Twitter, seguido pela implementação e treinamento dos modelos mencionados. Os resultados obtidos evidenciam o desempenho de cada modelo, destacando o LSTM como uma escolha notável devido aos seus resultados na acurácia.

Ao comparar os resultados com estudos relacionados, foi observado que, dentre os modelos Tradicionais de *Machine Learning* e *Deep Learning*, a arquitetura LSTM obteve o melhor desempenho, alcançando uma acurácia de aproximadamente 79,13%. Os outros modelos (LinearSVC, Regressão Logística, BernoulliNB e Árvore de Decisão) apresentaram desempenhos variados, porém, em geral, com acurácias inferiores ao modelo LSTM. Esses resultados indicam que a LSTM teve um desempenho superior aos outros modelos considerados, com métricas de desempenho mais consistentes. Isso se deve à arquitetura da rede neural na modelagem de sequências temporais, capturando dependências de longo alcance entre palavras em textos, sendo particularmente eficiente em tarefas de análise de sentimentos, onde a compreensão contextual é essencial para aprender padrões complexos e relações semânticas entre palavras. Diferentemente dos modelos lineares, as LSTMs são mais aptas a lidar com a complexidade textual, processando sequências e retendo informações de longo prazo para uma representação mais rica. Além disso, o treinamento em embeddings, como os Word2Vec pré-treinados, pode ter fortalecido a capacidade da LSTM de capturar representações semânticas robustas.

Por outro lado, o modelo Transformer RoBERTa obteve uma acurácia de 89,84%, sendo superior aos resultados de outros estudos relacionados. Esse desempenho superior pode ser atribuído à sua arquitetura robusta, que é mais avançada e eficiente em comparação com modelos LLMs mais tradicionais. A melhoria do modelo RoBERTa já havia sido prevista no artigo de [38], que destacou sua superioridade em relação ao BERT, apresentando resultados significativamente melhores em diversas métricas de avaliação, especialmente em tarefas de processamento de linguagem natural mais complexas, como a análise de sentimentos.

## 5.2 Limitações técnicas

Durante o desenvolvimento desta pesquisa, algumas limitações de cunho técnico-computacional foram observadas e impactaram o processo de experimentação. A principal dificuldade esteve relacionada à necessidade de elevado poder computacional para lidar com o grande volume de dados do conjunto Sentiment140 [42, 43] e com a complexidade dos modelos implementados.

O tempo de processamento mostrou-se um fator crítico. O treinamento do modelo LSTM demandou aproximadamente 8 horas, enquanto o treinamento do modelo RoBERTa exigiu cerca de 12 horas para ser concluído. Esse tempo elevado se deve tanto ao tamanho do dataset (1,6 milhão de tweets) quanto à natureza intensiva das operações envolvidas em arquiteturas de aprendizado profundo.

Além disso, durante a execução no ambiente Google Colab, enfrentaram-se limitações técnicas relacionadas à capacidade computacional disponível. Em diversas ocasiões, o treinamento dos modelos não foi concluído devido a restrições de uso da plataforma, como tempo limite de execução e memória insuficiente para processar o dataset completo.

Outra limitação encontrada esteve associada às atualizações das bibliotecas utilizadas nos experimentos, em especial durante a fase inicial de treinamento com o modelo LSTM. Mudanças nas versões das bibliotecas de suporte (como *TensorFlow* e *Keras*) ocasionaram incompatibilidades e erros de execução, exigindo revisões de código e ajustes no ambiente de desenvolvimento. Esse processo resultou em atrasos adicionais e aumento do tempo total de experimentação, além de evidenciar a dependência de infraestrutura estável e bem documentada para pesquisas em aprendizado profundo.

Adicionalmente, para viabilizar a execução do segundo experimento, referente ao treinamento do modelo RoBERTa, contou-se com o apoio do aluno de doutorado Higo, orientado pelo Professor Vicente Leonardo Paucar. Ele disponibilizou acesso a uma VPN com infraestrutura de Lisboa do laboratório *INESC-ID's Human Language Technology* que permitiu rodar o código em ambiente mais estável. Essa colaboração foi essencial, uma vez que no ambiente do Google Colab o treinamento não se concluíria devido a limitações técnicas. Ainda assim, mesmo com esse suporte, o tempo de execução permaneceu elevado, reforçando a necessidade de infraestrutura computacional especializada para lidar com modelos da arquitetura Transformer.

## 5.3 Considerações sobre o modelo RoBERTa

A utilização do modelo RoBERTa nesta pesquisa se justificou pela necessidade de adotar uma arquitetura mais robusta para a análise de sentimentos em textos curtos, informais e ruidosos, como os tweets. Diferentemente de modelos tradicionais de aprendizado

de máquina e de redes neurais recorrentes, o RoBERTa se baseia na arquitetura Transformer, caracterizada pelo mecanismo de atenção, que permite capturar relações contextuais profundas entre palavras de uma sentença.

Essa característica o torna mais adequado para lidar com ambiguidades semânticas, gírias, abreviações e estruturas linguísticas típicas das redes sociais. Enquanto o modelo LSTM obteve um desempenho competitivo, com 79,13% de acurácia, o RoBERTa alcançou 89,84%, evidenciando sua maior capacidade de generalização e seu poder de modelagem contextual.

O ganho de desempenho se deve principalmente ao uso de representações dinâmicas de palavras, em contraste com embeddings estáticos, permitindo que cada termo seja interpretado conforme o contexto em que está inserido. Isso reforça o RoBERTa como uma ferramenta mais adequada para a análise de sentimentos em textos informais.

## 5.4 Trabalhos futuros

Em trabalhos futuros, pretende-se melhorar o algoritmo para que se possa obter melhores resultados, como por exemplo: RoBERTa-Tweet (para esse modelo já foi iniciado o treinamento e testes, onde acurácia alcançou 90,07%, no qual pretende-se aprofundar a pesquisa futuramente), além de aplicação com outros modelos híbridos: RoBERTa-GPU, GPT 4, entre outros, aplicando-os tanto na base de dados Sentiment140 quanto em outras bases de dados.

Esses modelos, ao serem treinados em diferentes conjuntos de dados e aplicados a tarefas de análise de sentimentos, têm o potencial de proporcionar avaliações mais precisas e permitir a validação de hipóteses sobre o impacto das arquiteturas transformers em domínios específicos, como sentimentos expressos em textos curtos e linguagem informal, característicos do Twitter.

Além disso, novos ajustes de hiperparâmetros e de treinamento serão explorados para otimizar ainda mais a performance dos modelos, aproveitando as potencialidades dos transformers em tarefas de processamento de linguagem natural.

## 5.5 Publicação decorrente da dissertação

Durante o desenvolvimento deste mestrado, um artigo científico foi aceito e publicado em conferência:

- [46] K. C. P. Mendes, V. L. Paucar, R. N. D. Costa Filho, “Análise de sentimentos de tweets utilizando diferentes técnicas de deep learning.” XXV Congresso Brasileiro de Automática (CBA 2024), 2024, Rio de Janeiro.

## Referências

- [1] K. L. Tan, C. P. Lee, K. M. Lim, and K. S. M. Anbananthen, “Sentiment analysis with ensemble hybrid deep learning model,” *IEEE Access*, vol. 10, pp. 103 694–103 704, 2022. Citado na página 1.
- [2] K. L. Tan, C. P. Lee, and K. M. Lim, “A survey of sentiment analysis: Approaches, datasets, and future research,” *Applied Sciences*, vol. 13, no. 7, 2023. [Online]. Available: <https://www.mdpi.com/2076-3417/13/7/4550> Citado na página 1.
- [3] F. Persia and D. D’Auria, “A survey of online social networks: Challenges and opportunities,” in *2017 IEEE International Conference on Information Reuse and Integration (IRI)*, 2017, pp. 614–620. Citado na página 1.
- [4] A. M. Khattak, R. Batool, F. A. Satti, J. Hussain, W. A. Khan, A. M. Khan, and B. Hayat, “Tweets Classification and Sentiment Analysis for Personalized Tweets Recommendation,” *Complexity*, vol. 2020, pp. 1–11, Dec. 2020. [Online]. Available: <https://www.hindawi.com/journals/complexity/2020/8892552/> Citado na página 1.
- [5] S. Yoo, J. Song, and O. Jeong, “Social media contents based sentiment analysis and prediction system,” *Expert Systems with Applications*, vol. 105, pp. 102–111, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417418302124> Citado na página 1.
- [6] M. Osman, “Estatísticas e fatos do twitter sobre a nossa rede favorita,” outubro 2023. Citado na página 1.
- [7] M. A. S. Pinto, A. F. L. J. Junior, A. J. G. Busson, and S. Colcher, “Relacionando modelagem de tópicos e classificação de sentimentos para análise de mensagens do twitter durante a pandemia da covid-19,” in *Anais Estendidos do XXVI Simpósio Brasileiro de Sistemas Multimídia e Web*. Porto Alegre, RS, Brasil: SBC, 2020, pp. 61–64. [Online]. Available: [https://sol.sbc.org.br/index.php/webmedia\\_estendido/article/view/13064](https://sol.sbc.org.br/index.php/webmedia_estendido/article/view/13064) Citado na página 1.
- [8] K. Arun and A. Srinagesh, “Multilingual twitter sentiment analysis using machine learning,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 6, p. 5992, Dec. 2020. [Online]. Available: <http://ijece.iaescore.com/index.php/IJECE/article/view/20770> Citado na página 1.
- [9] K. Dashtipour, S. Poria, A. Hussain, E. Cambria, A. Y. A. Hawalah, A. Gelbukh, and Q. Zhou, “Multilingual Sentiment Analysis: State of the Art and Independent

- Comparison of Techniques,” *Cognitive Computation*, vol. 8, no. 4, pp. 757–771, Aug. 2016. [Online]. Available: <https://doi.org/10.1007/s12559-016-9415-7> Citado na página 1.
- [10] O. Y. Adwan, M. Al-Tawil, A. Huneiti, R. Shahin, A. Abu Zayed, and R. Al-Dibsi, “Twitter Sentiment Analysis Approaches: A Survey,” *International Journal of Emerging Technologies in Learning (iJET)*, vol. 15, no. 15, p. 79, Aug. 2020. [Online]. Available: <https://online-journals.org/index.php/i-jet/article/view/14467> Citado na página 1.
- [11] A. Go, R. Bhayani, and L. Huang, “Twitter sentiment classification using distant supervision,” *CS224N project report, Stanford*, vol. 1, no. 12, p. 2009, 2009. Citado na página 2.
- [12] Y. Mao, Q. Liu, and Y. Zhang, “Sentiment analysis methods, applications, and challenges: A systematic literature review,” *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 4, p. 102048, 2024. Citado na página 7.
- [13] M. Rodríguez-Ibáñez, A. Casánez-Ventura, F. Castejón-Mateos, and P.-M. Cuenca-Jiménez, “A review on sentiment analysis from social media platforms,” *Expert Systems with Applications*, vol. 223, p. 119862, 2023. Citado na página 7.
- [14] U. Kamath, J. Liu, and J. Whitaker, *Deep learning for NLP and speech recognition*. Springer, 2019, vol. 84. Citado na página 7.
- [15] K. L. Tan, C. P. Lee, and K. M. Lim, “A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research,” *Applied Sciences*, vol. 13, no. 7, p. 4550, Apr. 2023. [Online]. Available: <https://www.mdpi.com/2076-3417/13/7/4550> Citado na página 7.
- [16] T. Liu, J. Meyerhoff, J. C. Eichstaedt, C. J. Karr, S. M. Kaiser, K. P. Kording, D. C. Mohr, and L. H. Ungar, “The relationship between text message sentiment and self-reported depression,” *Journal of Affective Disorders*, vol. 302, pp. 7–14, 2022. Citado na página 7.
- [17] M. Rodríguez-Ibáñez, A. Casánez-Ventura, F. Castejón-Mateos, and P.-M. Cuenca-Jiménez, “A review on sentiment analysis from social media platforms,” *Expert Systems with Applications*, vol. 223, p. 119862, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417423003639> Citado na página 8.
- [18] K. Ravi and V. Ravi, “A survey on opinion mining and sentiment analysis: Tasks, approaches and applications,” *Knowledge-Based Systems*, vol. 89, pp.

- 14–46, 2015. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0950705115002336> Citado na página 9.
- [19] J. A. Balazs and J. D. Velásquez, “Opinion mining and information fusion: a survey,” *Information Fusion*, vol. 27, pp. 95–110, 2016. Citado na página 9.
- [20] S. Rajalakshmi, S. Asha, and N. Pazhaniraja, “A comprehensive survey on sentiment analysis,” in *2017 fourth international conference on signal processing, communication and networking (ICSCN)*. IEEE, 2017, pp. 1–5. Citado na página 9.
- [21] I. Chaturvedi, E. Cambria, R. E. Welsch, and F. Herrera, “Distinguishing between facts and opinions for sentiment analysis: Survey and challenges,” *Information Fusion*, vol. 44, pp. 65–77, 2018. Citado na página 9.
- [22] F. Hemmatian and M. K. Sohrabi, “A survey on classification techniques for opinion mining and sentiment analysis,” *Artificial intelligence review*, vol. 52, no. 3, pp. 1495–1545, 2019. Citado na página 9.
- [23] A. Yadav and D. K. Vishwakarma, “Sentiment analysis using deep learning architectures: a review,” *Artificial Intelligence Review*, vol. 53, no. 6, pp. 4335–4385, 2020. Citado na página 9.
- [24] Y. Liu, X. Zhu, Z. Shen, Y. Liu, M. Li, Y. Chen, B. John, Z. Ma, T. Hu, Z. Li, Z. Xu, W. Luo, and J. Wang, “Do large language models possess sensitive to sentiment?” 2025. [Online]. Available: <https://arxiv.org/abs/2409.02370> Citado 2 vezes nas páginas 10 e 12.
- [25] N. Saleena *et al.*, “An ensemble classification system for twitter sentiment analysis,” *Procedia computer science*, vol. 132, pp. 937–946, 2018. Citado 4 vezes nas páginas 11, 12, 14 e 15.
- [26] P. Harjule, A. Gurjar, H. Seth, and P. Thakur, “Text classification on twitter data,” in *2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE)*. IEEE, 2020, pp. 160–164. Citado 2 vezes nas páginas 11 e 12.
- [27] K. Dhola and M. Saradva, “A comparative evaluation of traditional machine learning and deep learning classification techniques for sentiment analysis,” in *2021 11th international conference on cloud computing, data science & engineering (Confluence)*. IEEE, 2021, pp. 932–936. Citado 4 vezes nas páginas 12, 14, 15 e 16.
- [28] C. A. G. Garcez, “Distributed electricity generation in brazil: An analysis of policy context, design and impact,” pp. 353–362, 2017. Citado na página 13.

- [29] A. C. Silva, “Geração distribuída no brasil: impactos regulatórios e desafios,” pp. 780–799, 2019. Citado na página 13.
- [30] R. J. Heffron and D. McCauley, “The concept of energy justice across the disciplines,” pp. 658–667, 2014. Citado na página 13.
- [31] S. Soares, “Justiça energética e transição sustentável no brasil,” pp. 1–22, 2019. Citado na página 13.
- [32] E. Schulte, C. Montag, and M. Reuter, “Determinants of residential adoption of solar pv: A meta-analysis,” p. 102187, 2021. Citado na página 13.
- [33] B. Liu, “Sentiment analysis: Mining opinions, sentiments, and emotions,” 2020. Citado na página 13.
- [34] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, pp. 1735–1780, 11 1997. Citado na página 14.
- [35] N. Iqbal, A. M. Chowdhury, and T. Ahsan, “Enhancing the performance of sentiment analysis by using different feature combinations,” in *2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)*. IEEE, 2018, pp. 1–4. Citado 2 vezes nas páginas 14 e 15.
- [36] Y. G. Jung, K. T. Kim, B. Lee, and H. Y. Youn, “Enhanced naive bayes classifier for real-time sentiment analysis with sparkr,” in *2016 international conference on information and communication technology convergence (ICTC)*. IEEE, 2016, pp. 141–146. Citado 2 vezes nas páginas 14 e 15.
- [37] M. Rathi, A. Malik, D. Varshney, R. Sharma, and S. Mendiratta, “Sentiment analysis of tweets using machine learning approach,” in *2018 Eleventh international conference on contemporary computing (IC3)*. IEEE, 2018, pp. 1–3. Citado 2 vezes nas páginas 14 e 16.
- [38] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, “Roberta: A robustly optimized bert pretraining approach,” 2019. Citado 3 vezes nas páginas 14, 16 e 36.
- [39] M. Gaye, C. Zhang, and A. Wulamu, “Twitter sentiment analysis combining lexical dictionary, machine learning, and deep learning models,” *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 2, 2021. Citado na página 14.
- [40] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” 2019. [Online]. Available: <https://arxiv.org/abs/1810.04805> Citado 2 vezes nas páginas 16 e 31.

- [41] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Kaiser, and I. Polosukhin, “Attention is all you need,” in *Advances in neural information processing systems*, vol. 30, 2017. Citado na página 16.
- [42] A. Go, R. Bhayani, and L. Huang, “Twitter sentiment classification using distant supervision,” *CS224N project report, Stanford*, vol. 1, no. 12, p. 2009, 2009. Citado 4 vezes nas páginas 18, 19, 24 e 37.
- [43] M. Michailidis, “Sentiment140 dataset with 1.6 million tweets,” <https://www.kaggle.com/datasets/kazanov/sentiment140>, 2017, (Accessed on 01/09/2024). Citado 4 vezes nas páginas 18, 19, 24 e 37.
- [44] B. Gaye, D. Zhang, and A. Wulamu, “A tweet sentiment classification approach using a hybrid stacked ensemble technique,” *Information*, vol. 12, no. 9, 2021. [Online]. Available: <https://www.mdpi.com/2078-2489/12/9/374> Citado na página 20.
- [45] D. G. Kleinbaum, M. Klein, and E. R. Pryor, *Logistic regression: a self-learning text*. Springer, 2002, vol. 94. Citado na página 23.
- [46] K. C. P. Mendes, V. L. Paucar, and R. N. D. C. Filho, “Análise de sentimentos de tweets utilizando diferentes técnicas de deep learning,” in *Proceedings of the XXV Congresso Brasileiro de Automática (CBA 2024)*, 2024. Citado na página 38.