

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

MATHEUS FELIPE MILESKI LOPES

**ANÁLISE DE SENTIMENTOS EM REDES SOCIAIS DURANTE A TEMPORADA
2022 DA FÓRMULA 1: UM ESTUDO UTILIZANDO PROCESSAMENTO DE
LINGUAGEM NATURAL NO TWITTER**

APUCARANA

2023

MATHEUS FELIPE MILESKI LOPES

**ANÁLISE DE SENTIMENTOS EM REDES SOCIAIS DURANTE A TEMPORADA
2022 DA FÓRMULA 1: UM ESTUDO UTILIZANDO PROCESSAMENTO DE
LINGUAGEM NATURAL NO TWITTER**

**Sentiment Analysis in Social Networks during the 2022 Formula 1 Season: A
Study Using Natural Language Processing on Twitter**

Trabalho de Conclusão de Curso de Graduação apresentado como requisito para obtenção do título de Bacharel em Engenharia de Computação do Bacharelado em Engenharia de Computação da Universidade Tecnológica Federal do Paraná.

Orientador: Prof. Dr. Rafael Gomes Mantovani

APUCARANA

2023



[4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)

Esta licença permite compartilhamento, remixe, adaptação e criação a partir do trabalho, mesmo para fins comerciais, desde que sejam atribuídos créditos ao(s) autor(es). Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

MATHEUS FELIPE MILESKI LOPES

**ANÁLISE DE SENTIMENTOS EM REDES SOCIAIS DURANTE A TEMPORADA
2022 DA FÓRMULA 1: UM ESTUDO UTILIZANDO PROCESSAMENTO DE
LINGUAGEM NATURAL NO TWITTER**

Trabalho de Conclusão de Curso de Graduação
apresentado como requisito para obtenção
do título de Bacharel em Engenharia de
Computação do Bacharelado em Engenharia
de Computação da Universidade Tecnológica
Federal do Paraná.

Data de aprovação: 21/novembro/2023

Daniel Prado de Campos
Doutorado
Universidade Tecnológica Federal do Paraná

Luiz Fernando Carvalho
Doutorado
Universidade Tecnológica Federal do Paraná

Rafael Gomes Mantovani
Doutorado
Universidade Tecnológica Federal do Paraná

APUCARANA

2023

Dedico este trabalho à minha família, pelo amor incondicional, e aos meus amigos, pelos momentos de riso e apoio que fizeram toda a diferença durante a graduação.

AGRADECIMENTOS

À medida que concluo esta etapa significativa da minha jornada acadêmica, sou tomado por um profundo sentimento de gratidão a todos aqueles que contribuíram para o meu crescimento e sucesso.

Primeiramente, devo meu sincero agradecimento ao Prof. Dr. Rafael Gomes Mantovani, cuja orientação e inspiração foram fundamentais nestes últimos três anos. Sob sua tutela, mergulhei profundamente no fascinante mundo da ciência de dados e inteligência artificial, participando de projetos de iniciação científica e extensão que moldaram minha trajetória acadêmica e profissional. Sua paixão pelo conhecimento e dedicação incansável foram uma fonte constante de motivação e aprendizado.

À minha família, não tenho palavras suficientes para expressar minha gratidão. Vocês foram a rocha sobre a qual construí minha resiliência e determinação. Nos momentos em que decidi mudar de curso para seguir meu sonho de trabalhar com tecnologia, vocês não apenas entenderam minha decisão, mas também me apoiaram incondicionalmente. Sem vocês, as dificuldades teriam sido insuperáveis e as vitórias, menos doces.

Aos meus amigos, agradeço por serem a válvula de escape das pressões da graduação. Nos momentos de estresse e cansaço, suas risadas e companheirismo foram o bálsamo que me manteve são. As memórias que criamos juntos serão tesouros que levarei comigo para sempre.

Por fim, gostaria de deixar uma mensagem de encorajamento a todos aqueles que trilham o caminho da educação e do conhecimento. A jornada acadêmica é repleta de desafios, mas cada obstáculo superado nos torna mais fortes e sábios. Persistam em seus sonhos, busquem sua paixão e lembrem-se de que, com dedicação e apoio, todos os objetivos são alcançáveis.

"Things get done only if the data we gather can inform and inspire those in a position to make difference."(SCHMOKER, 2000)

"As coisas só são realizadas se os dados que coletamos puderem informar e inspirar aqueles em posição de fazer a diferença."
— Mike Schmoker, tradução livre

RESUMO

LOPES, Matheus F. M.. **Análise de sentimentos em redes sociais durante a temporada 2022 da Fórmula 1:** Um estudo utilizando processamento de linguagem natural no Twitter. 2023. Trabalho de Conclusão de Curso de Graduação (Bacharelado em Engenharia de Computação), Universidade Tecnológica Federal do Paraná, Apucarana, 2023.

A importância cada vez maior dos dados para a tomada de decisões estratégicas e o reconhecimento das redes sociais como vastos repositórios de opinião pública impulsionaram a realização deste estudo, que teve como objetivo explorar a interação entre as emoções humanas e os eventos de automobilismo. A pesquisa centrou-se na aplicação do Processamento de Linguagem Natural (PLN) para extrair e analisar sentimentos manifestados em *tweets* sobre a Fórmula 1. Aplicou-se técnicas avançadas de aprendizado de máquina e aprendizado profundo para treinar vários modelos na tarefa de identificação de sentimentos. Dentre eles, o *Logistic Regression* e o modelo *Long Short-Term Memory* (LSTM) destacaram-se, alcançando acurácias de 78,21% e 78,08%, respectivamente. O modelo LSTM, em particular, foi implementado em um conjunto de dados público de *tweets* coletados durante as temporadas de 2021 e 2022 da Fórmula 1. O modelo foi utilizado para classificar os sentimentos expressos pelos fãs, permitindo uma análise exploratória dos dados correlacionados a eventos específicos das corridas. Os resultados obtidos revelaram padrões significativos de engajamento, com picos notáveis de reações emocionais coincidindo com momentos-chave das temporadas. As descobertas encontradas ilustram como eventos particulares podem influenciar profundamente as emoções e o comportamento dos fãs, bem como, a partir da análise detalhada dos sentimentos expressos, é possível obter dados valiosos que podem ser aproveitados para o desenvolvimento de estratégias de marketing e comunicação mais eficazes no esporte.

Palavras-chave: mineração de dados; aprendizado do computador; análise linguística; opinião pública.

ABSTRACT

LOPES, Matheus F. M.. **Sentiment Analysis in Social Networks during the 2022 Formula 1 Season:** A Study Using Natural Language Processing on Twitter. 2023. Undergraduate Thesis (Bachelor's Degree in Computer Engineering), Federal University of Technology - Paraná, Apucarana, 2023.

The growing importance of data in strategic decision-making and the recognition of social networks as vast repositories of public opinion have propelled this study, which aimed to explore the interaction between human emotions and motorsport events. The research focused on applying Natural Language Processing (NLP) to extract and analyze sentiments expressed in tweets about Formula 1. Advanced machine learning and deep learning techniques were employed to train various models in the sentiment classification task. Among these, the Logistic Regression and the Long Short-Term Memory (LSTM) model stood out, achieving accuracies of 78.21% and 78.08%, respectively. The LSTM model, in particular, was implemented on a public dataset of tweets collected during the 2021 and 2022 Formula 1 seasons. The model was used to classify the sentiments expressed by fans, allowing for an exploratory analysis of data correlated to specific events of the races. The findings revealed significant engagement patterns, with notable spikes in emotional reactions coinciding with key moments of the seasons. These discoveries illustrate how particular events can profoundly influence the emotions and behavior of fans, and from a detailed analysis of expressed sentiments, valuable data can be obtained that may be leveraged for developing more effective marketing and communication strategies in the sport.

Keywords: data mining; machine learning; linguistic analysis; public opinion.

LISTA DE FIGURAS

Figura 1 – Linha do tempo dos avanços do PLN.	19
Figura 2 – Disciplinas que constituem a Mineração de Dados.	34
Figura 3 – Categorias do aprendizado de máquina.	35
Figura 4 – Arquitetura básica do MLP.	40
Figura 5 – Arquitetura básica de uma CNN.	41
Figura 6 – Processo de aprendizado de uma RNN ao longo do tempo.	42
Figura 7 – Arquitetura dos portões de uma rede <i>Long Short-Term Memory</i> (LSTM).	43
Figura 8 – Arquitetura dos portões de uma rede <i>Gated Recurrent Unit</i> (GRU).	43
Figura 9 – Pipeline desenvolvida para a análise de sentimentos.	54
Figura 10 – Código QR para o repositório deste trabalho.	63
Figura 11 – Gráfico de violino da acurácia dos modelos em relação aos <i>datasets</i> de n-gramas.	67
Figura 12 – Histórico de treinamento da CNN.	70
Figura 13 – Histórico de treinamento do LSTM.	70
Figura 14 – Matriz de confusão do treinamento do LR.	72
Figura 15 – Matriz de confusão do treinamento do LSTM.	72
Figura 16 – Curva ROC para o treinamento do LR.	73
Figura 17 – Curva ROC para o treinamento do LSTM.	73
Figura 18 – Curva <i>Precision-Recall</i> para o treinamento do LR.	73
Figura 19 – Curva <i>Precision-Recall</i> para o treinamento do LSTM.	73
Figura 20 – Predições dos modelos de LR e LSTM.	73
Figura 21 – Histograma de <i>tweets</i> positivos e negativos classificados pelo LSTM.	76
Figura 22 – Quantidade de <i>tweets</i> durante o GP de Abu Dhabi 2021.	77
Figura 23 – Quantidade de <i>tweets</i> durante o GP do México 2021.	78
Figura 24 – Quantidade de <i>tweets</i> durante o GP da Arábia Saudita 2021.	78
Figura 25 – Quantidade de <i>tweets</i> durante o GP da Holanda 2021.	78
Figura 26 – Quantidade de <i>tweets</i> durante o GP da Grã-Bretanha 2022.	79
Figura 27 – Gráfico de Violino das métricas dos modelos em relação aos <i>datasets</i> de n-gramas.	95
Figura 28 – Matriz de Confusão do <i>Random Forest</i> (RF).	95

Figura 29 – Matriz de Confusão do <i>Multilayer Perceptron</i> (MLP).	95
Figura 30 – Matriz de Confusão do <i>Convolutional Neural Networks</i> (CNN).	96
Figura 31 – Curva ROC do RF.	96
Figura 32 – Curva ROC do MLP.	96
Figura 33 – Curva ROC do CNN.	96
Figura 34 – Curva <i>Precision-Recall</i> do RF.	97
Figura 35 – Curva <i>Precision-Recall</i> do MLP.	97
Figura 36 – Curva <i>Precision-Recall</i> do CNN.	97

LISTA DE TABELAS

Tabela 1 – Exemplo de uma Matriz de Confusão	46
Tabela 2 – Colunas presentes no <i>dataset</i> Sentiment140.	55
Tabela 3 – Colunas presentes no <i>dataset</i> de <i>tweets</i> sobre Fórmula 1.	56
Tabela 4 – Ferramentas e bibliotecas utilizadas no projeto	61
Tabela 5 – Resultados do experimento com o TF-IDF	66
Tabela 6 – Resultados do treinamento dos modelos de AM e AP	69
Tabela 7 – Comparação entre as previsões de sentimentos do LR e do LSTM.	74
Tabela 8 – Corridas presentes no dataset de F1	77

LISTA DE ABREVIATURAS E SIGLAS

Siglas

AM	Aprendizado de Máquina
AP	Aprendizado Profundo
API	<i>Application Programming Interface</i>
AUC	<i>Area Under the Curve</i>
BERT	<i>Bidirectional Encoder Representations from Transformers</i>
BPTT	<i>Backpropagation Through Time</i>
CD	Ciência de Dados
CNN	<i>Convolutional Neural Networks</i>
DT	<i>Decision Tree</i>
F1	Fórmula 1
FIA	<i>Fédération Internationale de l'Automobile</i> , em português Federação Internacional de Automobilismo
FN	Falso Negativo
FP	Falso Positivo
GPT	<i>Generative Pre-trained Transformers</i>
GRU	<i>Gated Recurrent Unit</i>
IA	Inteligência Artificial
KNN	<i>K-Nearest Neighbors</i>
LR	<i>Logistic Regression</i>
LSTM	<i>Long Short-Term Memory</i>
MD	Mineração de Dados
MLP	<i>Multilayer Perceptron</i>
NB	<i>Naïve Bayes</i>

NLTK	<i>Natural Language Toolkit</i>
PLM	Modelos de Linguagem Pré-treinado
PLN	Processamento de Linguagem Natural
QA	<i>Questions and Answers</i>
RF	<i>Random Forest</i>
RNN	<i>Recurrent Neural Networks</i>
ROC	<i>Receiver Operating Characteristic</i>
SVM	<i>Support Vector Machine</i>
TF-IDF	<i>Term Frequency-Inverse Document Frequency</i>
TM	Tradução por Máquina
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Considerações iniciais	14
1.2	Objetivos	15
1.2.1	Objetivo geral	15
1.2.2	Objetivos específicos	15
1.3	Justificativa	15
1.4	Estrutura do trabalho	16
2	REFERENCIAL TEÓRICO	18
2.1	Processamento de Linguagem Natural	18
2.1.1	Histórico e evolução do PLN	18
2.1.2	Principais aplicações de PLN	24
2.1.3	Conceitos e técnicas fundamentais do PLN	26
2.1.4	Desafios do PLN	28
2.1.5	Ferramentas e Recursos	29
2.1.5.1	Bibliotecas de PLN	29
2.1.5.2	Modelos de Linguagem Pré-Treinados (PLM)	30
2.1.5.3	<i>Corpus e Datasets</i>	30
2.2	Análise de Sentimentos	31
2.2.1	Definição e importância	31
2.2.2	Análise de sentimentos em redes sociais	32
2.2.3	Aplicação da análise de sentimentos no contexto da Fórmula 1 (F1)	32
2.3	Mineração e Processamento de Dados	34
2.4	Aprendizado de Máquina (AM)	35
2.4.1	Tipos de Algoritmos de AM	36
2.4.2	Algoritmos de AM	36
2.4.3	Aplicação de AM na Análise de Sentimentos	38
2.5	Aprendizado Profundo (AP)	38
2.5.1	Modelos de AP	40
2.5.2	Aplicação de AP na Análise de Sentimentos	44
2.6	Métricas de Desempenho para Modelos de Análise de Sentimentos	45

3	TRABALHOS RELACIONADOS	49
4	METODOLOGIA EXPERIMENTAL	54
4.1	Dataset	54
4.2	Pré-processamento	56
4.3	Extração de Características	57
4.4	Algoritmos de AM	58
4.5	Algoritmos de AP	59
4.6	Setup Experimental	60
5	RESULTADOS	64
5.1	Definição de n-gramas para TF-IDF	64
5.2	Resultado do treinamento dos modelos de AM e AP	68
5.2.1	Comparação dos resultados do LR e do LSTM	71
5.3	Análise dos sentimentos dos fãs de F1	75
5.3.1	Análise dos resultados de sentimento por corrida	76
6	CONCLUSÃO	81
6.1	Resultados Gerais	81
6.2	Limitações e Dificuldades	81
6.3	Trabalhos Futuros	82
6.4	Considerações Finais	83
	REFERÊNCIAS	84
	APÊNDICE A GRÁFICOS COMPLEMENTARES	95
	A.1 Resultados do Experimento de Definição de N-gramas para o TF-IDF	95
	A.2 Resultados do Treinamento dos Modelos de AM e AP	95

1 INTRODUÇÃO

Este capítulo traça o panorama do campo de estudo e situa o objeto de pesquisa. Em seguida, são apresentados os objetivos, tanto o geral quanto os específicos. A subseção da justificativa ressalta a importância e a relevância do estudo, enfatizando a contribuição que este trabalho oferece ao corpo acadêmico e à prática aplicada. Por fim, é apresentada a estrutura do trabalho, fornecendo um guia para o leitor sobre a organização do documento, esboçando os capítulos e as principais seções que compõem esta monografia.

1.1 Considerações iniciais

O trabalho aqui apresentado tem como cerne a análise dos sentimentos manifestados pelos usuários do Twitter/X¹ em relação à temporada de 2022 da Fórmula 1 (F1)², empregando métodos de Processamento de Linguagem Natural (PLN) (HIRSCHBERG; MANNING, 2015). A pesquisa se debruça sobre o desafio de capturar e interpretar os sentimentos e as perspectivas dos entusiastas do automobilismo, expressos por meio das interações digitais na mencionada plataforma de mídia social.

Através da criação de um *pipeline* automatizado³, o estudo superou os obstáculos trazidos pelas recentes alterações nas políticas de acesso à *Application Programming Interface* (API) do Twitter, recorrendo ao uso de dois *datasets* públicos substanciais. Essas coleções de dados forneceram a base para o processo de pré-processamento, treinamento de modelos preditivos de sentimentos e análise subsequente, possibilitando uma avaliação minuciosa dos padrões de engajamento e sentimentos.

O processo de análise envolveu etapas de preparação dos dados, aplicação de modelos de aprendizado de máquina e aprendizado profundo para a classificação de sentimentos, e a extração de percepções significativas. As estratégias de visualização de dados empregadas visam representar as tendências detectadas graficamente e desenvolver uma interface interativa que simplifique a interpretação dos resultados (WILBANKS; LANGFORD, 2014).

A pesquisa busca traçar um panorama entre os eventos que permeiam o cenário da F1 com o impacto nos sentimentos e engajamento dos seus fãs, bem como, procura oferecer um instrumento analítico de alto valor para estratégias de marketing e comunicação no contexto esportivo. Assim, o trabalho contribui para um domínio de pesquisa em expansão, que integra

¹ O Twitter é uma plataforma de mídia social que permite aos usuários compartilhar mensagens curtas chamadas "*tweets*". Durante o desenvolvimento deste trabalho, a plataforma passou a se chamar X, após aquisição pelo novo proprietário.

² A F1 é a principal categoria do automobilismo mundial, organizada pela *Fédération Internationale de l'Automobile*, em português Federação Internacional de Automobilismo (FIA).

³ *Pipelines* automatizados são utilizados no campo do aprendizado de máquina para integrar etapas de pré-processamento de dados, extração de características, modelagem e avaliação (DAS *et al.*, 2021).

análise de dados, entendimento do comportamento do consumidor e administração esportiva (WORATSCHEK; HORBEL; POPP, 2014).

Finalmente, a meta deste trabalho é revelar as dinâmicas sentimentais presentes na comunidade do Twitter em torno da F1, mapeando as emoções, tanto positivas quanto negativas, que envolvem os eventos da temporada. As descobertas são direcionadas a enriquecer a compreensão e apoiar a tomada de decisões dos envolvidos no universo do automobilismo.

1.2 Objetivos

Nesta seção, serão definidos o objetivo geral e os objetivos específicos que guiaram o desenvolvimento do trabalho.

1.2.1 Objetivo geral

Investigar e aplicar as técnicas de PLN para realizar uma análise de sentimentos dos usuários de redes sociais, especificamente no Twitter, durante a temporada de 2022 da F1, a fim de identificar padrões de engajamento, sentimentos e compreender as motivações por trás dessas emoções.

1.2.2 Objetivos específicos

- Desenvolver um *pipeline* automatizado para coleta e pré-processamento de dados;
- Realizar uma análise exploratória do engajamento e dos sentimentos;
- Identificar padrões individuais de sentimentos e engajamento nas corridas específicas, bem como padrões gerais ao longo de toda a temporada de 2022;
- Investigar possíveis correlações entre os sentimentos expressos e eventos importantes no cenário mundial, que possam influenciar a percepção do público;

1.3 Justificativa

A relevância da Ciência de Dados (CD) e do PLN tem se intensificado com a expansão do volume de dados gerados e com o avanço das técnicas de Inteligência Artificial (IA) (PHAM *et al.*, 2020), que incluem inovações como ChatGPT, DALL-E, MusicLM, entre outros (CAO *et al.*, 2023). Estas ferramentas têm remodelado a forma como informações são extraídas, interpretadas e transformadas em decisões estratégicas e em conteúdo autônomo. As técnicas de Mineração de Dados (MD) são fundamentais neste cenário para destilar informações valiosas,

possibilitando a compreensão de fenômenos complexos e aprimorando a geração de IA mais refinada (JOSEPH; THANAKUMAR, 2019).

A mineração de sentimentos e opiniões é uma das aplicações mais consolidadas da combinação de MD e PLN. Esta área de estudo foca na extração de inferências semânticas e afetivas a partir da linguagem natural, frequentemente sem necessidade de um entendimento profundo do texto, abrindo caminho para análises que lidam com o volume substancial de dados gerados em redes sociais (CAMBRIA *et al.*, 2013). As redes sociais, em particular, se tornaram fontes ricas de dados em tempo real, onde usuários compartilham opiniões, emoções e percepções sobre uma miríade de tópicos. A capacidade de analisar estes sentimentos proporciona uma compreensão valiosa das opiniões do público, que pode ser aplicada em diversas áreas, incluindo marketing, política, entretenimento e esporte (MAULUD *et al.*, 2021).

Neste contexto, o presente trabalho se propôs a aplicar estas técnicas avançadas para analisar os sentimentos do público da F1 durante a temporada de 2022, com o Twitter servindo como principal fonte de dados. A motivação por trás desta escolha é dupla: primeiramente, oferecer uma nova lente de análise sobre a percepção do público em relação a equipes, pilotos e corridas, identificando padrões e tendências que possam ser estratégicos para os envolvidos na competição; e, em segundo lugar, contribuir para o entendimento da relação entre as emoções do público, o desempenho das equipes e pilotos, e o impacto das decisões estratégicas dos organizadores.

A aplicação prática do PLN em um domínio específico como o esporte automobilístico não apenas demonstra a viabilidade de tais técnicas em campos especializados, mas também sublinha a importância de analisar o conteúdo gerado em redes sociais para extrair informações e compreender as interações entre eventos esportivos e a emoção do público.

Por fim, este trabalho destaca a aplicabilidade do conhecimento técnico adquirido e seu potencial para contribuir com a produção científica na área de PLN, bem como sua relevância prática para os interesses e estratégias dos profissionais envolvidos no cenário esportivo global.

1.4 Estrutura do trabalho

Este trabalho está organizado em capítulos que discorrem sobre os diversos aspectos da pesquisa realizada, abarcando desde a base conceitual até a exposição dos resultados e considerações finais.

- **Referencial teórico:** fornece a fundamentação teórica indispensável para a compreensão da pesquisa, este capítulo explora o domínio do PLN, os princípios e as práticas subjacentes à análise de sentimentos, as técnicas avançadas de mineração e processamento de dados, oferece uma perspectiva introdutória sobre aprendizado de máquina e aprendizado profundo e descreve os princípios por trás das métricas de resultados utilizados no trabalho.

- **Trabalhos Relacionados:** analisa estudos anteriores no campo da análise de sentimentos, delineando o posicionamento do trabalho desenvolvido no contexto acadêmico.
- **Metodologia Experimental:** detalha os procedimentos metodológicos adotados na pesquisa, englobando desde a seleção dos conjuntos de dados até as técnicas de pré-processamento e análise implementadas.
- **Resultados:** apresenta e discute os resultados alcançados pela pesquisa, fornecendo uma análise das métricas e dos modelos de aprendizado de máquina e aprendizado profundo empregados.
- **Conclusão:** sintetiza os principais achados do estudo e pondera sobre suas implicações práticas e acadêmicas, refletindo sobre as limitações enfrentadas e propondo direcionamentos para pesquisas futuras.

2 REFERENCIAL TEÓRICO

O presente capítulo introduz os principais conceitos para a compreensão do trabalho.

2.1 Processamento de Linguagem Natural

Segundo Liddy (2001) o Processamento de Linguagem Natural (PLN) é um campo interdisciplinar que combina conhecimentos da Linguística, Ciência da Computação e Inteligência Artificial, com o objetivo de criar interfaces que permitem a interação entre computadores e seres humanos por meio da linguagem natural. A intenção é que máquinas sejam capazes de entender e processar a linguagem humana de uma maneira que seja tanto natural quanto útil para diversas aplicações. A área de PLN abrange desde a compreensão de texto e voz até a geração de linguagem que se aproxima da comunicação humana.

O desenvolvimento de sistemas baseados em PLN envolve o entendimento e a manipulação de linguagem, requerendo que as máquinas realizem tarefas complexas como reconhecimento de fala, análise de sentimentos, tradução automática, sumarização de textos, entre outras. Este campo é particularmente desafiador devido à complexidade e variação inerentes às línguas humanas (JONES; GALLIERS, 1995).

No contexto do trabalho proposto, o PLN é utilizado como ferramenta fundamental para análise de sentimentos em redes sociais, uma aplicação que requer não apenas o reconhecimento de palavras e suas estruturas gramaticais, mas também a interpretação de nuances contextuais e emocionais expressas nas publicações dos usuários. A seguir, são apresentados os tópicos que compõem o referencial teórico da área de PLN, abordando a origem do PLN e seus marcos históricos, suas principais aplicações, conceitos básicos e técnicas, desafios enfrentados na área e uma visão geral das principais ferramentas, bibliotecas e recursos disponíveis.

2.1.1 Histórico e evolução do PLN

A Figura 1, apresenta uma linha do tempo que oferece uma visão geral das principais conquistas e marcos no campo do PLN. Esta linha do tempo ilustra a evolução dessa área desde seus primórdios no final dos anos 1940 até os avanços mais recentes, destacando os desenvolvimentos cruciais que moldaram o campo. Os detalhes e implicações de cada marco serão explorados e aprofundados nesta seção, proporcionando uma compreensão abrangente da trajetória histórica do PLN e seu impacto contínuo no mundo da tecnologia e comunicação.

Figura 1 – Linha do tempo dos avanços do PLN.



Fonte: Autoria própria (2023).

Embora o termo PLN não fosse conhecido no final dos anos 1940, os esforços pioneiros já estavam em andamento, particularmente no campo da Tradução por Máquina (TM). As pesquisas desta época tinham um escopo internacional, com o russo e o inglês sendo as línguas predominantes nos estudos de TM, conforme indicado por Otten e Pacak (1971). Contrariando as expectativas, a pesquisa em TM e PLN enfrentou um período de declínio após o relatório ALPAC (1966), que infamemente declarou que a TM não tinha um futuro promissor. No entanto, desafiando essas previsões, sistemas de produção de TM começaram a fornecer resultados práticos para seus clientes, como documentado por Hutchins (1986).

Paralelamente, o uso de computadores para estudos literários e linguísticos já havia começado. Trabalhos influentes impulsionados pela IA surgiram no início da década de 1960, com sistemas de *Questions and Answers* (QA) como o BASEBALL (GREEN *et al.*, 1961). Seguiram-se sistemas como o LUNAR¹ (WOODS, 1978) e o SHRDLU² (WINOGRAD, 1971), que representaram um avanço significativo em termos de capacidades linguísticas e de processamento de tarefas. Segundo Gardner *et al.* (1979), acreditava-se amplamente que progressos substanciais seriam possíveis em duas frentes: projetos de pesquisa de compreensão de fala e em projetos de desenvolvimento de sistemas com foco em interfaces de banco de dados, que se propunham a avançar além do LUNAR.

Na década de 1980, a Teoria da Gramática Computacional tornou-se uma área de pesquisa extremamente ativa, ligada às lógicas para significado e à capacidade de lidar com as crenças e intenções do usuário, além de funções como identificação de ênfases e temas abordados (BRISCOE *et al.*, 1987). Ao final dessa década, processadores de frases de propósito geral como o *Core Language Engine* da SRI International³ (ALSHAWI, 1992) e a Teoria de Representação do Discurso (KAMP; REYLE, 1993) ofereceram meios para abordar discursos mais extensos dentro de uma estrutura gramático-lógica. Este foi um período de crescimento comunitário, com a disponibilização de recursos práticos, gramáticas, ferramentas e analisadores, como o *Alvey Natural Language Tools*⁴ (BRISCOE *et al.*, 1987).

¹ O sistema LUNAR, apresentado em uma convenção de ciências lunares em 1971, foi um dos primeiros sistemas de QA projetados para interagir com um sistema especialista no domínio de amostras de rochas lunares.

² SHRDLU foi um programa pioneiro no entendimento de linguagem natural, desenvolvido por Terry Winograd no MIT entre 1968 e 1970, onde o usuário interage com o computador em um mundo virtual simplificado, por meio de blocos.

³ SRI International, anteriormente conhecido como Stanford Research Institute, é uma organização de pesquisa científica e inovação independente, que realiza trabalhos contratados para o governo, empresas privadas e fundações. Fundado em 1946, tem desempenhado um papel fundamental no desenvolvimento de tecnologias que vão desde a internet até assistentes virtuais de voz.

⁴ O programa Alvey financiado pelo governo do Reino Unido e desenvolvido pelas Universidades de Cambridge, Edimburgo e Lancaster, apresentou em 1987 a primeira versão do *Alvey Natural Language Tools*, contendo um analisador morfológico, uma gramática e um léxico (BRISCOE *et al.*, 1987).

Os anos 90 foram marcados pelas conferências de reconhecimento de fala e compreensão de mensagens (extração de informações) da DARPA⁵, notáveis não apenas pelas tarefas abordadas, mas pelo foco em avaliações rigorosas, iniciando uma tendência que se tornou proeminente na década (YOUNG; CHASE, 1998; SUNDHEIM; CHINCHOR, 1993). O avanço em temas como a desambiguação de sentidos de palavras (SMALL; COTTRELL; TANENHAUS, 2013) e as redes probabilísticas (JONES, 1994) marcaram o início de uma nova era no PLN: a do processamento linguístico estatístico. Segundo Charniak (1996), o processamento linguístico estatístico é uma abordagem dentro do PLN que utiliza modelos matemáticos para compreender e gerar linguagem natural. Baseia-se na ideia de que padrões linguísticos podem ser capturados e previstos por meio de probabilidades e estatísticas. Esta abordagem ganhou destaque no final dos anos 90, especialmente nas pesquisas de Manning e Schutze (1999), que ressaltaram a relevância do processamento estatístico da linguagem não somente para os analistas de dados, mas também para profissionais e acadêmicos de áreas diversas, como Linguística, Matemática, Psicologia e Filosofia. A perspectiva era de que palavras e frases transcendiam sua existência como mero dado, elas são entidades carregadas de significados, intenções, opiniões e valores culturais. O processamento estatístico da linguagem, portanto, não se limitava a uma análise quantitativa, ele se estendia para a modelagem de fenômenos linguísticos complexos e multidimensionais. Ainda no final do milênio, a extração de informações e o resumo automático também ganharam destaque através dos estudos de Mani e Maybury (1999).

Segundo Khurana *et al.* (2023), foi no início dos anos 2000 que a modelagem de linguagem neural, no qual a probabilidade de ocorrência da próxima palavra é determinada dado um número n de palavras anteriores, começou a ganhar atenção. Bengio, Ducharme e Vincent (2000) propuseram o conceito de rede neural *feedforward* e uma tabela de consulta para representar as n palavras anteriores em sequência. Collobert e Weston (2008) propuseram a aplicação de aprendizado multitarefa no campo do PLN, onde dois modelos convolucionais com *max pooling* foram usados para realizar tarefas de etiquetagem de partes da fala e reconhecimento de entidades nomeadas.

Mikolov *et al.* (2013) introduziram um processo de *word embedding*, onde a representação vetorial densa do texto foi endereçada, destacando também os desafios enfrentados pela tradicional representação esparsa, conhecida como saco-de-palavras. Com o avanço do *word embedding*, redes neurais foram introduzidas no campo do PLN, permitindo o processamento de entradas de comprimento variável. Sutskever, Vinyals e Le (2014) propuseram um *framework* geral para o mapeamento de sequência para sequência, onde redes codificadoras e decodificadoras são usadas para mapear de sequência para vetor e de vetor para sequência, respectivamente.

⁵ A DARPA, acrônimo do inglês *Defense Advanced Research Projects Agency* é uma agência de pesquisa e desenvolvimento do Departamento de Defesa dos Estados Unidos, responsável pelo desenvolvimento de tecnologias emergentes para uso militar.

A aplicação de CNN no campo do PLN abordou problemas como classificação de sentenças, análise de sentimentos, classificação de textos, sumarização de texto, tradução automática e relações de resposta, com trabalhos notáveis apresentados por autores como Widiastuti (2019) e Wang e Gang (2018). As *Recurrent Neural Networks* (RNN), ideais para dados sequenciais como texto, séries temporais, dados financeiros, fala, áudio e vídeo, também foram amplamente utilizadas, conforme discutido por Tarwani e Edem (2017).

Uma variante modificada das RNNs, as redes LSTMs, provou ser muito útil nos casos em que apenas informações importantes desejadas precisam ser retidas por mais tempo, descartando informações irrelevantes, como apontado por Hochreiter e Schmidhuber (1997). O desenvolvimento subsequente do LSTM levou a uma variante ligeiramente mais simples, a GRU, que mostrou resultados melhores que os LSTMs padrão em muitas tarefas de PLN (CHUNG *et al.*, 2014).

Mecanismos de auto-atenção, que permitem que as redes neurais concentrem-se em aspectos específicos dos dados de entrada, juntamente com o uso de PLN, permitiram o desenvolvimento dos Transformadores, como apresentado por Vaswani *et al.* (2017a). Divergindo das arquiteturas baseadas em RNN e CNN, os Transformadores são fundamentados no mecanismo de auto-atenção, que permite ao modelo processar sequências de dados de forma paralela e mais eficiente. Este mecanismo capacita os Transformadores a entenderem as relações entre diferentes partes de uma sequência de dados, como palavras em uma sentença, sem a necessidade de processamento sequencial.

Ainda no trabalho de Vaswani *et al.* (2017a), é apresentada a estrutura básica de um Transformador, baseado em blocos de codificação e decodificação, cada um contendo múltiplas camadas de auto-atenção e redes neurais de *feedforward*. A atenção multi-cabeça, uma característica central dos transformadores, possibilita que o modelo foque em diferentes partes da sequência simultaneamente, melhorando a capacidade de capturar nuances contextuais e semânticas. Essa abordagem melhora significativamente o desempenho em tarefas como tradução automática, geração de texto e compreensão de linguagem, ultrapassando os modelos baseados em RNN e LSTM em várias métricas de desempenho (VASWANI *et al.*, 2017a).

É importante notar que os Transformadores têm o potencial de aprender dependências de longo prazo, mas são limitados por um contexto de comprimento fixo na configuração da modelagem de linguagem. Por esse motivo, Dai *et al.* (2019) propuseram uma nova arquitetura neural, o *Transformer-XL*⁶, que permite aprender dependências além de um comprimento fixo de palavras. O trabalho de Rae *et al.* (2019) sobre o *Compressive Transformer*, expande a proposta por trás do *Transformer-XL* com um modelo de sequência e auto-atenção, que com-

⁶ A sigla XL refere-se a *extra-large* ou, em português, extra-longo.

prime memórias para aprendizado de sequências de longo alcance, dessa forma, evitando a necessidade de Retropropagação Através do Tempo⁷, que é computacionalmente custosa.

Em 2019, surge o primeiro Modelos de Linguagem Pré-treinado (PLM) baseado em Transformadores, o *Bidirectional Encoder Representations from Transformers* (BERT) (DEVLIN *et al.*, 2019). Com o BERT, pesquisadores começam a desenvolver novas ferramentas e a aprimorar diversos sistemas já então consolidados, como os analisador de sentimentos, os etiquetadores de partes da fala, e os reconhecedores de entidades nomeadas.

A introdução dos PLMs impulsionaram diversos avanços no PLN. Estes modelos emergiram como uma metodologia poderosa, caracterizada pelo pré-treino em larga escala seguido de um refinamento específico para uma determinada tarefa, como a de classificação de sentimentos (MIN *et al.*, 2023). A abordagem "*pre-train then fine-tune*" (ROGERS; KOVALEVA; RUMSHISKY, 2021) permite que os modelos se beneficiem de um vasto *corpus*⁸ não rotulado para aprender representações de linguagem gerais, que são posteriormente ajustadas para tarefas específicas com uma quantidade menor de dados de treino rotulados.

Outra técnica inovadora é a aprendizagem baseada em *prompts*, na qual um PLM é solicitado de tal forma que a solução de uma tarefa de PLN é reduzida a uma tarefa semelhante à tarefa de pré-treino do PLM, como a previsão de uma palavra ausente, ou uma tarefa mais simples, como a implicação textual. Esta técnica pode alavancar de maneira mais eficaz o conhecimento codificado nos PLMs, facilitando abordagens com poucos exemplos de dados textuais (YUAN; NEUBIG; LIU, 2021).

A reformulação de tarefas de PLN como geração de texto é outra metodologia que busca aproveitar ao máximo o conhecimento codificado em um modelo de linguagem generativo, como o GPT-2 (RADFORD *et al.*, 2018) e o T5 (RAFFEL *et al.*, 2020). Ao invés de limitar-se a classificações ou etiquetagem de sequências, esses modelos generativos são capazes de realizar tarefas complexas de geração de texto. A geração de dados é um método complementar que envolve o uso de PLMs para gerar automaticamente dados para tarefas de PLN. Os dados

⁷ A Retropropagação Através do Tempo (*Backpropagation Through Time* (BPTT)), é uma técnica de treinamento de redes neurais utilizada especificamente para dados sequenciais. Ela é uma extensão da retropropagação usada em redes neurais comuns e é aplicada em RNNs. BPTT funciona desdobrando a RNN ao longo do tempo e aplicando a retropropagação em cada passo de tempo, atualizando os pesos em relação às contribuições de cada etapa temporal no erro final. Essa técnica é essencial para que as RNNs aprendam dependências de longo prazo, mas pode ser computacionalmente intensiva (WERBOS, 1990).

⁸ Segundo Sinclair (1991), um *corpus* em PLN é definido como uma coleção sistemática e criteriosa de textos autênticos ou falas, que são coletados e organizados conforme critérios linguísticos explícitos para servir como uma amostra representativa do uso da linguagem.

gerados podem ser dados com rótulos prateados⁹, onde o PLM generativo é normalmente refinado para a tarefa, ou algum dado auxiliar, como contraexemplos, esclarecimentos, contextos ou outros. No primeiro caso, os dados com rótulos prateados podem ser adicionados aos dados rotulados existentes. No segundo caso, os dados auxiliares apoiam a tarefa alvo de alguma forma.

Em suma, os PLMs e os *Generative Pre-trained Transformers* (GPTs) expandiram o horizonte do PLN, facilitando não apenas a melhoria dos resultados nas tarefas-alvo tradicionais, mas também abrindo novos caminhos para abordar o PLN como uma questão de geração de texto e aprendizado interativo, explorando as capacidades inatas desses modelos complexos e profundamente treinados (MIN *et al.*, 2023).

2.1.2 Principais aplicações de PLN

A área de PLN engloba uma ampla gama de aplicações, muitas das quais se tornaram indispensáveis em vários domínios da atividade humana. Essas aplicações vão desde a interação básica com computadores até o processamento complexo de grandes volumes de dados textuais. A seguir, exploramos algumas das principais aplicações do PLN, evidenciando o seu papel fundamental na era digital.

- **Tradução de Máquina (TM):** a TM ou tradução automática é uma das aplicações mais antigas e bem estabelecidas do PLN, permitindo a conversão de texto de um idioma para outro. Sistemas de TM, como o *Google Translate*, têm evoluído significativamente com o avanço dos modelos baseados em IA, resultando em traduções cada vez mais precisas e fluentes (WU *et al.*, 2016).
- **Reconhecimento de Fala:** transforma a fala oral em texto escrito e é fundamental em assistentes virtuais como Siri e Alexa. Esta aplicação do PLN permite a interação por voz com dispositivos e sistemas, tornando a tecnologia mais acessível e conveniente (HINTON *et al.*, 2012).
- **Sumarização Automática:** refere-se à criação de um resumo conciso e relevante a partir de um texto extenso. Com o aumento exponencial de informações disponíveis, essa aplicação torna-se crucial para permitir que os usuários compreendam rapidamente o conteúdo essencial de documentos longos (NENKOVA; MCKEOWN *et al.*,

⁹ Um dado com rótulo prateado (*silver labeled data*) refere-se a informações que foram etiquetadas automaticamente usando uma combinação de diferentes classificadores, em contraste com os dados de rótulo dourado (*gold labeled data*), que são geralmente anotados manualmente por especialistas humanos. Enquanto os dados com rótulo dourado são considerados melhores em termos de precisão e confiabilidade para treinar modelos de aprendizado de máquina, os dados com rótulo prateado oferecem uma alternativa mais rápida e escalável, embora possam não alcançar o mesmo nível de precisão. Eles são particularmente úteis em cenários onde o custo ou a viabilidade de obter anotações humanas é proibitivo (BOLAND; KRÜGER, 2019).

2011). Existem diversos sistemas gratuitos de sumarização automática, entre eles podemos citar o RavBug AutoSummarizer¹⁰ e o QuillBot Summarize¹¹.

- **Análise de Sentimentos:** a análise de sentimento ou opinião é uma aplicação do PLN que identifica e classifica as opiniões expressas em um texto, auxiliando empresas a entenderem a percepção do público sobre produtos e serviços. Este campo tem uma relevância especial nas redes sociais, onde as opiniões são compartilhadas livremente (VINODHINI; CHANDRASEKARAN, 2012). Esta aplicação será aprofundada na próxima seção.
- **Extração de Informação:** envolve a identificação automática de informações estruturadas, como entidades nomeadas, relações e eventos, de textos não estruturados. Esta aplicação é essencial para a construção de bases de conhecimento e suporte à tomada de decisões (GRISHMAN, 2015).
- **Perguntas e Respostas:** sistemas de perguntas e respostas (QA) fornecem respostas automáticas a perguntas formuladas em linguagem natural. Essa aplicação de PLN está se tornando cada vez mais sofisticada, com a capacidade de fornecer respostas contextuais e precisas, beneficiando-se amplamente dos avanços dos PLMs e GPTs (CHEN *et al.*, 2017).
- **Aprendizado Assistido por Computador:** o aprendizado assistido por computador apoia o aprendizado de idiomas através de ferramentas que avaliam e auxiliam na correção e aprendizado da escrita e fala em línguas estrangeiras. Tais ferramentas são fundamentais em plataformas de educação online e aplicativos de aprendizado de idiomas (HEIFT; SCHULZE, 2007).
- **Reconhecimento de Entidades Nomeadas:** é uma sub-tarefa de Extração de Informação que envolve a localização e classificação de entidades em categorias como nomes de pessoas, organizações, locais e quantidades em textos. Essa aplicação é amplamente utilizada em sistemas de busca inteligente e análise de dados (NADEAU; SEKINE, 2007).
- **Desambiguação de Sentidos de Palavras:** a desambiguação de sentidos de palavras é crucial para a compreensão adequada do significado das palavras no contexto em que são usadas. Esta aplicação é importante em todas as outras áreas do PLN, pois lida diretamente com a polissemia inerente à linguagem humana (NAVIGLI, 2009).
- **Diálogo e Sistemas Conversacionais:** sistemas conversacionais e de diálogo, como *chatbots*, são desenvolvidos para simular conversas com usuários humanos. Esses

¹⁰ <https://www.ravbug.com/summarizer/>

¹¹ <https://quillbot.com/summarize>

sistemas são cada vez mais utilizados em atendimento ao cliente, *e-learning*, e terapia online, representando uma interação natural e intuitiva (MCTEAR; CALLEJAS; GRIOL, 2016).

Cada uma dessas aplicações demonstra a versatilidade e a importância do PLN no processamento e na análise de grandes volumes de dados linguísticos, evidenciando seu papel crítico na interseção entre a linguagem humana e a computação.

2.1.3 Conceitos e técnicas fundamentais do PLN

Para compreender os intrincados padrões da linguagem, o PLN recorre a um conjunto diversificado de conceitos e técnicas. Esta subseção aborda os fundamentos essenciais que sustentam o campo do PLN.

- **Linguagem Computacional:** a interseção entre linguagem e computação é a pedra angular do PLN. A linguagem é um sistema complexo de comunicação com nuances e variações que desafiam a análise computacional. O PLN aborda este desafio aplicando algoritmos que processam e analisam dados linguísticos (MANNING; SCHUTZE, 1999). A representação computacional da linguagem natural envolve a conversão de texto e fala em formatos que máquinas possam manipular, o que implica em modelagem linguística e computacional (PARSING, 2009).
- **Análise Lexical:** a análise lexical, ou tokenização, é o processo de segmentar um texto em suas unidades constituintes, chamadas *tokens*. Estes *tokens* geralmente correspondem a palavras, números e símbolos. A precisão na tokenização é vital, pois influencia diretamente todas as etapas subsequentes do PLN (CHOWDHURY, 2003).
- **Análise Sintática:** a análise sintática, ou *parsing*, é o processo que estrutura um texto conforme a gramática da língua. Ela determina como as palavras em uma sentença se relacionam entre si e formam uma estrutura hierárquica, identificando sujeitos, objetos, e outros elementos gramaticais (KLEIN; MANNING, 2003).
- **Análise Semântica:** a análise semântica procura entender o significado e a interpretação de palavras, frases e textos completos, considerando o contexto em que são usados. Este processo pode incluir a desambiguação de sentidos de palavras e a compreensão de referências e relações semânticas (NAVIGLI, 2009).
- **Recursos Lexicais:** recursos lexicais, como dicionários e ontologias, são fundamentais para o PLN. Eles fornecem estruturas de dados que permitem a associação de palavras com seus significados e relações semânticas. O WordNet é um exemplo amplamente utilizado de um recurso lexical que auxilia na análise semântica (MILLER, 1995).

- **Lematização:** a lematização é o processo de agrupar as diferentes formas flexionadas de uma palavra para que possam ser analisadas como um único item, identificado pelo lema da palavra, ou sua forma de dicionário. A lematização depende do correto reconhecimento da classe gramatical e do uso da palavra no texto, resultando em uma forma base que tem significado no dicionário. Esse processo é fundamental para tarefas que requerem compreensão semântica completa das palavras, como a tradução automática e a recuperação de informações (PLISSON *et al.*, 2004).
- **Stemmização:** a stemmização é uma técnica que corta as terminações morfológicas das palavras, reduzindo-as a uma forma base ou raiz, conhecida como *stem*. Ela não requer conhecimento da classe gramatical e geralmente se baseia em algoritmos heurísticos de corte. Apesar de menos precisa que a lematização, a stemmização é útil em aplicações onde a correspondência exata de palavras não é necessária, como a indexação de pesquisa e algumas tarefas de classificação de texto (PORTER, 2006).
- **Modelagem de Tópicos:** a modelagem de tópicos é uma técnica estatística que analisa as palavras de documentos para descobrir os temas que permeiam um grande volume de texto. *Latent Dirichlet Allocation* (LDA) é um dos métodos mais conhecidos para a modelagem de tópicos (BLEI; NG; JORDAN, 2003).
- **Embeddings de Palavras:** *embeddings* de palavras, como Word2Vec e GloVe, são representações vetoriais de palavras que capturam o contexto semântico e as relações entre elas. Essas representações são geradas usando redes neurais e são cruciais para muitas aplicações avançadas do PLN (MIKOLOV *et al.*, 2013; PENNINGTON; SOCHER; MANNING, 2014).
- **Processamento de Linguagem Sequencial:** O processamento de sequências lida com a natureza temporal e ordenada da linguagem. Técnicas como RNN e Transformadores são projetadas para capturar dependências de longo prazo em sequências de texto, essenciais para a geração de linguagem e tradução automática (VASWANI *et al.*, 2017b).
- **Term Frequency-Inverse Document Frequency (TF-IDF):** o TF-IDF é uma técnica estatística que avalia a importância de uma palavra em um documento em relação a uma coleção de documentos, o *corpus*. A frequência do termo (TF) mede a frequência com que uma palavra aparece em um documento, enquanto a frequência inversa do documento (IDF) avalia a importância da palavra ao longo do *corpus*. Palavras que aparecem frequentemente em um documento, mas não em todo o *corpus*, recebem uma pontuação mais alta, destacando sua relevância para o documento específico (CHOWDHURY, 2010).

- **Saco-de-Palavras:** é uma representação simplificada de um texto usada no PLN e na recuperação de informação. Neste modelo, um texto (como uma frase ou documento) é representado como o saco (multiconjunto) de suas palavras, desconsiderando a gramática e até a ordem das palavras, mas mantendo a multiplicidade (HARRIS, 1954).
- **Polaridade das Palavras:** é um conceito usado na análise de sentimentos para identificar a orientação emocional que as palavras expressam. A polaridade pode ser positiva, negativa ou neutra e é frequentemente determinada por meio de léxicos de sentimento ou de forma computacional usando técnicas de PLN. A análise de polaridade é essencial para interpretar opiniões e sentimentos em textos (TURNEY, 2002).

Esses conceitos são elementos chave no campo do PLN e desempenham um papel crítico na análise de sentimentos, particularmente no contexto das mídias sociais. A compreensão e aplicação dessas técnicas são vitais para a construção de modelos precisos e eficazes de classificação de sentimentos.

2.1.4 Desafios do PLN

O campo do PLN apresenta uma série de desafios intrínsecos à complexidade da linguagem humana e à sua representação computacional. Esses desafios abrangem desde a compreensão profunda das nuances linguísticas até a adequação de modelos em cenários com dados limitados ou de baixa qualidade. A seguir, é destacado os principais obstáculos enfrentados na área:

A ambiguidade é uma característica inerente à linguagem natural. Ela pode ocorrer em diferentes níveis, como lexical (uma palavra com múltiplos significados), sintático (mais de uma estrutura gramatical possível) e semântico (significados múltiplos dependendo do contexto). Resolver ambiguidades requer métodos sofisticados e conhecimento contextual, o que ainda é um desafio para sistemas de PLN atuais (COLEMAN, 2005).

Sarcasmo e ironia são formas de linguagem figurativa que costumam expressar o oposto do significado literal das palavras utilizadas. A detecção e interpretação correta dessas expressões exigem a compreensão de contextos sociais e culturais complexos, algo que modelos computacionais ainda lutam para realizar com precisão (MAYNARD; GREENWOOD, 2014).

O uso de expressões idiomáticas e figurativas, como metáforas e expressões coloquiais, constitui um desafio significativo para o PLN, uma vez que o significado dessas expressões muitas vezes não pode ser deduzido diretamente das palavras individuais que as compõem (KATZ; GIESBRECHT, 2006).

A variação contextual e a adaptação a diferentes domínios são também grandes desafios no PLN. Sistemas desenvolvidos para um domínio específico muitas vezes não performam bem quando aplicados a outros, devido à variação na terminologia e no estilo linguístico (III; MARCU, 2006).

Outro grande desafio é a busca por grandes quantidades de dados anotados para o treinamento de modelos de PLN. A anotação manual é cara e demorada, e a qualidade dos dados anotados pode variar significativamente, afetando o desempenho dos modelos treinados (HOVY *et al.*, 2006).

Além desses, outros desafios incluem a compreensão de novas gírias e jargões que surgem continuamente, a necessidade de modelos robustos a erros de digitação ou gramática, e a integração efetiva de conhecimento do mundo real no processamento linguístico. Abordar esses desafios é fundamental para avançar no desenvolvimento de sistemas de PLN que possam entender e processar linguagem humana de maneira eficaz e confiável.

2.1.5 Ferramentas e Recursos

O desenvolvimento e a implementação de soluções de PLN são fortemente apoiados por uma variedade de ferramentas, bibliotecas e recursos que facilitam desde a manipulação de linguagem natural até o treinamento de modelos complexos de aprendizado de máquina e aprendizado profundo. Abaixo, descrevemos algumas das ferramentas e recursos mais significativos nesta área.

2.1.5.1 Bibliotecas de PLN

Abaixo é apresentado as principais bibliotecas de PLN utilizadas no campo da Ciência da Computação para o desenvolvimento de aplicações que demandam compreensão, geração ou manipulação de linguagem natural. Essas bibliotecas fornecem uma gama de ferramentas e funcionalidades que permitem aos pesquisadores e desenvolvedores realizar desde tarefas básicas como tokenização, até operações mais complexas, como reconhecimento de entidades nomeadas e análise semântica profunda.

- **Natural Language Toolkit (NLTK):** uma das bibliotecas mais antigas e abrangentes para o processamento de linguagem natural em Python, o NLTK oferece funcionalidades para a tokenização, etiquetagem, análise sintática, semântica e outras tarefas de PLN em diversos idiomas, incluindo o português (JURAFSKY; MANNING, 2012).
- **spaCy:** focada na produção de *pipelines* de PLN para aplicações do mundo real, spaCy é uma biblioteca que fornece desempenho robusto e eficiente para tarefas como tokenização, lematização, reconhecimento de entidades nomeadas e outras (VASILIEV, 2020).
- **TextBlob:** uma biblioteca simplificada para tarefas básicas de PLN, TextBlob é popular por sua interface intuitiva, permitindo fácil realização de análise de sentimentos, classificação de texto e tradução (LORIA *et al.*, 2018).

Todas as bibliotecas apresentadas estão disponíveis para a linguagem de programação Python. Essa linguagem emergiu como a líder na área de PLN devido à sua sintaxe clara, bibliotecas ricas e comunidade ativa, tornando-se o a principal escolha para a maioria das bibliotecas de PLN modernas (BIRD; KLEIN; LOPER, 2009). Com interfaces intuitivas e documentação extensiva, essas bibliotecas são acessíveis tanto para iniciantes quanto para especialistas na área, facilitando a adoção e o desenvolvimento de soluções avançadas de PLN em diversos contextos e aplicações.

2.1.5.2 Modelos de Linguagem Pré-Treinados (PLM)

Os PLM proporcionam um entendimento mais profundo e uma representação mais rica da linguagem humana. Esses modelos são baseados em arquiteturas de rede neural que foram treinadas em grandes volumes de texto, permitindo que eles capturem uma ampla gama de nuances linguísticas e padrões contextuais. A seguir, apresentamos dois dos principais modelos pré-treinados que têm definido o estado da arte no PLN contemporâneo.

- **BERT:** utiliza técnicas para o pré-treinamento de representações de linguagem, o BERT captura contextos bidirecionais e tem demonstrado alto desempenho em diversas tarefas de PLN (DEVLIN *et al.*, 2019).
- **GPT:** A série de modelos GPT, desenvolvida pela OpenAI, oferece capacidades impressionantes de geração de texto e compreensão da linguagem, sendo treinada em vastas quantidades de texto para aprender uma representação abrangente da linguagem natural (RADFORD *et al.*, 2019).

2.1.5.3 *Corpus e Datasets*

O PLN se beneficia significativamente da disponibilidade de *corpus* e *datasets* anotados, que são usados para treinar e testar modelos, entre eles destacam-se:

- **Dataset Sentiment140:** Um *dataset* contendo 1,6 milhões de *tweets* anotados para análise de sentimentos, útil para treinar modelos de classificação de sentimentos (GO; BHAYANI; HUANG, 2009).
- **Corpus do Projeto Gutenberg:** Uma coleção de textos de domínio público, que serve como um recurso valioso para modelagem de linguagem e análise literária (LEBERT, 2008).
- **GLUE e SuperGLUE:** *Benchmarks* que fornecem uma variedade de tarefas de PLN para avaliar o desempenho dos modelos em uma ampla gama de domínios (WANG *et al.*, 2018; WANG *et al.*, 2019).

Estes recursos, entre muitos outros, formam a base sobre a qual os pesquisadores e profissionais de PLN podem construir e avaliar suas soluções, contribuindo para o avanço contínuo do campo.

2.2 Análise de Sentimentos

A análise de sentimentos é uma área de pesquisa que se situa na intersecção entre a CD e o PLN, dedicada a compreender e classificar as opiniões expressas em textos. Esta seção detalhará sua definição, importância, e aplicação em redes sociais, como no contexto da F1.

2.2.1 Definição e importância

A análise de sentimentos, frequentemente equiparada à mineração de opiniões ou *opinion mining*, é uma disciplina do PLN que se concentra na identificação e categorização das opiniões expressas em um texto para determinar a atitude do autor em relação a temas específicos ou a generalidade de uma narrativa. A detecção e a análise de sentimentos permitem a avaliação das reações emocionais e subjetivas das pessoas, oferecendo informações valiosas para o entendimento de suas perspectivas e predileções (LIU, 2022).

A importância da análise de sentimentos transcende a simples categorização de textos em positivos, negativos ou neutros. Estende-se ao campo da inteligência de negócios e à gestão de relacionamento com o cliente, onde a capacidade de monitorar e responder às opiniões dos consumidores em tempo real pode significar uma vantagem competitiva importante (PANG; LEE *et al.*, 2008). No âmbito da saúde pública, a análise de sentimentos em discussões online pode ajudar a identificar tendências de bem-estar mental ou a percepção pública de campanhas de saúde (PAUL; DREDZE, 2017).

A área de *opinion mining* é ainda mais abrangente, envolve não apenas a detecção de sentimentos, mas também a atribuição destes a entidades específicas e aspectos. Por exemplo, é possível discernir não apenas se um comentário sobre um automóvel é positivo, mas também se essa positividade é direcionada ao design do veículo, à sua performance ou ao serviço de atendimento ao cliente da marca (HU; LIU, 2004). Este nível de granularidade é particularmente valioso para as empresas que buscam entender as múltiplas facetas das impressões dos clientes.

Além de suas aplicações em negócios e saúde, a análise de sentimentos é crucial em áreas como a política, onde a opinião pública expressa em redes sociais pode ser um preditor de comportamentos eleitorais e tendências sociais (TUMASJAN *et al.*, 2010). Em domínios como o entretenimento e o esporte, a análise de sentimentos permite às entidades acompanhar a recepção do público a eventos e performances, ajudando a moldar estratégias promocionais e de engajamento (MOSTAFA, 2013).

2.2.2 Análise de sentimentos em redes sociais

A análise de sentimentos em redes sociais, e particularmente no Twitter, assume um papel de destaque devido à natureza imediata e abrangente da plataforma. O Twitter permite aos usuários expressarem opiniões e reações a eventos em tempo real, tornando-se uma fonte rica e dinâmica de dados para *opinion mining*. Segundo Pak, Paroubek *et al.* (2010), a concisão dos *tweets*, limitados a um número específico de caracteres, exige que os usuários sejam diretos em suas comunicações, o que muitas vezes leva a expressões claras de sentimentos, tornando a análise mais acessível e menos ambígua.

Além disso, o Twitter é caracterizado pela sua heterogeneidade de usuários e pela diversidade de tópicos discutidos, que vão desde o cotidiano até eventos globais significativos, abrangendo uma ampla gama de domínios e indústrias. Esta variedade oferece aos pesquisadores e organizações a oportunidade de capturar um espectro representativo de opiniões e sentimentos da população (BIFET; FRANK, 2010).

Outro fator que torna o Twitter uma excelente plataforma para a análise de sentimentos é a disponibilidade de dados em tempo real, que permite análises ágeis e atualizadas, refletindo as mudanças na percepção pública quase instantaneamente. Isso é especialmente útil em situações que exigem respostas rápidas, como gerenciamento de crises ou acompanhamento de eventos ao vivo (KOULOUMPIS; WILSON; MOORE, 2011).

Por último, o Twitter fornece uma série de metadados associados a cada *tweet*, como geolocalização, *hashtags* e menções, que enriquecem a análise e permitem a segmentação e o estudo detalhado de tendências e padrões dentro de grupos específicos ou em relação a eventos particulares (TSYTSARAU; PALPANAS, 2012).

2.2.3 Aplicação da análise de sentimentos no contexto da Fórmula 1 (F1)

A F1, a mais prestigiada categoria do automobilismo mundial, apresenta uma plataforma global para o espetáculo das corridas e a inovação tecnológica. Com um calendário que percorre diversos países e culturas, a F1 não é apenas um esporte, mas um evento de significativa visibilidade e impacto social, econômico e tecnológico (COBBS *et al.*, 2017). A natureza competitiva e a paixão que envolve equipes, pilotos e fãs criam um ambiente repleto de emoções e opiniões, que são expressas abertamente nas redes sociais, especialmente durante as corridas e eventos importantes da temporada (SYLT; REID, 2011).

A análise de sentimentos dos fãs de F1 durante as corridas e eventos significativos é uma aplicação particularmente rica e desafiadora do processamento de linguagem natural. O dinamismo das corridas, com a ocorrência de incidentes, ultrapassagens, estratégias de equipe e outras variáveis, provoca respostas imediatas e emotivas dos seguidores do esporte, que são capturadas em tempo real no Twitter e outras plataformas de mídia social (STURM, 2014).

Essa análise proporciona uma compreensão detalhada do sentimento dos espectadores, permitindo às equipes de marketing e comunicação das empresas envolvidas, bem como aos organizadores da F1, captar o pulso da opinião pública. Assim, podem ajustar suas estratégias para aumentar o engajamento dos fãs e otimizar a experiência do espectador, tanto virtual quanto fisicamente presente nos eventos (HUTCHINS, 2008).

Durante eventos importantes da temporada, como o anúncio de mudanças de pilotos, lançamentos de carros, ou polêmicas regulatórias, a análise de sentimentos torna-se ainda mais crucial. As reações dos fãs a esses anúncios podem influenciar a percepção da marca e as decisões de negócios das equipes e patrocinadores (JENKINS, 2006). Por exemplo, a resposta positiva à estética de um novo carro ou à contratação de um piloto pode ser um indicativo valioso para campanhas promocionais futuras.

Além disso, a análise de sentimentos aplicada aos dados das redes sociais durante a temporada de F1 pode revelar tendências e padrões no comportamento dos fãs, fornecendo informações para a melhoria contínua da transmissão das corridas e do conteúdo relacionado. Isso pode incluir ajustes na cobertura televisiva, nas narrativas utilizadas durante a transmissão e na oferta de conteúdos adicionais às corridas (BOYLE, 2009).

Essas análises também podem ser úteis para a gestão de crises. As reações dos fãs a incidentes de corrida, como acidentes ou controvérsias, podem ser monitoradas e avaliadas, permitindo uma resposta rápida e informada por parte das equipes e organizadores, mitigando possíveis danos à imagem e reputação do esporte (ROSTAMI *et al.*, 2015). Dessa forma, os dados de sentimentos recolhidos e analisados oferecem uma oportunidade para as partes interessadas da F1 compreenderem melhor e se conectarem com sua base de fãs global, promovendo um esporte mais inclusivo e alinhado com as expectativas e desejos de seu público (HORTON *et al.*, 2019).

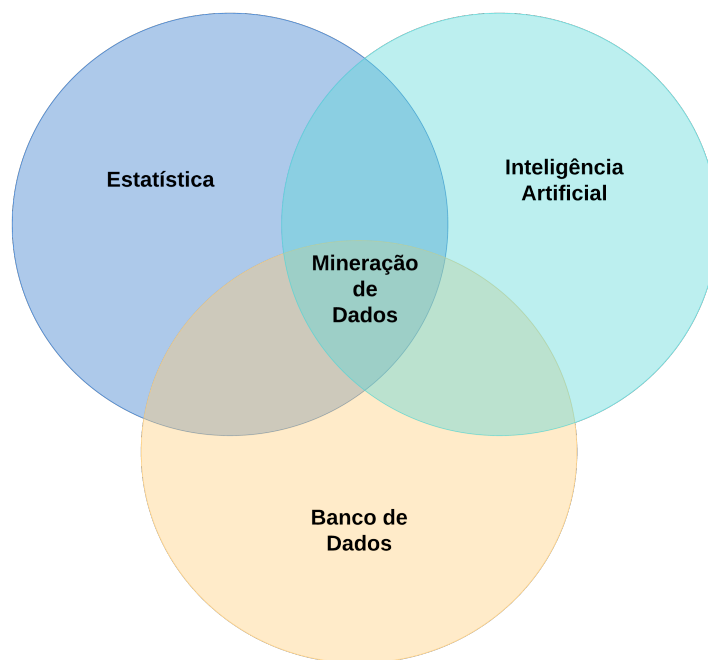
A aquisição da F1 pela Liberty Media em janeiro de 2017 marcou uma mudança significativa na gestão e na direção estratégica da F1. Antes da aquisição, a F1 enfrentava um declínio na audiência global e no engajamento dos fãs, em grande parte devido à sua reputação de ser um "clube de velhos", como descrito por Wood e Burkhalter (2023), e à falha em se adaptar à era digital, perdendo patrocinadores e relevância entre o público mais jovem.

A estratégia da Liberty envolveu uma comunicação aprimorada e conteúdo informativo sobre a F1, enquanto as iniciativas para atrair o público mais jovem, como a série documental *Drive to Survive* da Netflix, se concentraram em criar novos pontos de engajamento. As interações geradas nas redes sociais, especialmente no Twitter, permitiu capturar em tempo real a euforia, as decepções e as expectativas dos fãs durante e após cada evento. A votação no "Piloto do Dia", a interação com *hashtags* das corridas e a participação nos jogos de *e-sports* geraram dados quantitativos e qualitativos que, quando analisados, oferecem informações sobre a experiência dos fãs e sua conexão emocional com a F1 (CANTALUPPI; DAVINO, 2021).

2.3 Mineração e Processamento de Dados

A mineração de dados, como exemplificado na Figura 2, emerge da interseção entre Estatística, Banco de Dados e IA, essenciais para extrair conhecimento a partir de vastas quantidades de dados. A habilidade de coletar, processar e analisar dados eficientemente tornou-se um alicerce para avanços em diversas áreas do conhecimento, desde negócios até ciências sociais e biológicas (HAN; PEI; TONG, 2022).

Figura 2 – Disciplinas que constituem a Mineração de Dados.



Fonte: Autoria própria (2023).

A mineração de dados, ou *data mining*, é o processo de descobrir padrões, correlações e anomalias dentro de grandes conjuntos de dados. Utiliza-se de métodos de aprendizado de máquina, estatística e sistemas de banco de dados para encontrar informações úteis que muitas vezes não são imediatamente evidentes (WITTEN *et al.*, 2005). Em particular, a mineração de textos tornou-se uma área de interesse crescente com a explosão de dados textuais disponíveis nas redes sociais e outras plataformas online (PAK; PAROUBEK *et al.*, 2010).

O processamento de dados, por outro lado, refere-se ao conjunto de técnicas usadas para transformar dados brutos em um formato mais adequado para análise. No contexto de dados textuais, isso geralmente inclui a limpeza de dados, onde erros são corrigidos e informações irrelevantes são removidas, e a normalização, onde os dados são convertidos em um formato padrão. Métodos de pré-processamento, como tokenização, stemming e lematização, são empregados para reduzir a complexidade do texto e aumentar a eficácia das técnicas de mineração de dados subsequentes (INZALKAR; SHARMA, 2015).

A mineração e o processamento de dados também apresentam seus próprios desafios. A qualidade e a integridade dos dados coletados, a escolha de técnicas de processamento adequadas e a interpretação correta dos padrões descobertos são fundamentais para o sucesso da análise. Além disso, questões éticas relacionadas à privacidade e ao uso de dados pessoais exigem consideração cuidadosa (BOYD; CRAWFORD, 2012).

2.4 Aprendizado de Máquina (AM)

O Aprendizado de Máquina (AM), ou em inglês *Machine Learning* é uma subárea da IA que fornece aos sistemas a capacidade de aprender automaticamente e melhorar a partir da experiência sem ser explicitamente programado (MITCHELL, 1997). Esta disciplina científica está focada no desenvolvimento de algoritmos que podem encontrar padrões em dados e fazer previsões ou decisões baseadas nesses padrões, em vez de seguir instruções rígidas.

Figura 3 – Categorias do aprendizado de máquina.



Fonte: Adaptado de Almeida, Carvalho e Menino (2017).

2.4.1 Tipos de Algoritmos de AM

Alpaydin (2020) divide o aprendizado de máquinas em categorias baseadas na natureza do sinal ou *feedback* disponível para o sistema de aprendizagem, como aprendizado supervisionado, não supervisionado, semi-supervisionado e por reforço, como exemplificado na Figura 3.

- **Aprendizado Supervisionado:** no aprendizado supervisionado os algoritmos são treinados em um conjunto de dados rotulado e o objetivo é aprender a mapear as entradas para as saídas. É amplamente utilizado para tarefas de classificação e regressão (BISHOP; NASRABADI, 2006). Por exemplo, na análise de sentimentos, um modelo de aprendizado supervisionado pode ser treinado em um *corpus* de texto onde cada exemplo está marcado com um sentimento positivo ou negativo.
- **Aprendizado Não-Supervisionado:** o aprendizado não-supervisionado lida com dados não rotulados. O objetivo é descobrir estruturas ocultas nos dados. Métodos comuns de aprendizado não supervisionado incluem agrupamento e redução de dimensionalidade, que podem revelar agrupamentos naturais ou padrões nos dados (HASTIE *et al.*, 2009).
- **Aprendizado Semi-Supervisionado:** o aprendizado semi-supervisionado é utilizado quando os algoritmos utilizam tanto dados rotulados quanto não rotulados. Esta abordagem é particularmente útil quando o rótulo de dados é caro ou difícil de obter. O aprendizado semi-supervisionado pode ser utilizado tanto em tarefas de classificação, quanto em tarefas de agrupamento (LEARNING, 2006).
- **Aprendizado por Reforço:** no aprendizado por reforço, um agente aprende a tomar decisões sequenciais recebendo sinais de recompensa ao interagir com o ambiente e busca maximizar a recompensa acumulada ao longo do tempo. Este tipo de aprendizado é muito utilizado em jogos, robótica, decisões em tempo real e métodos heurísticos (SUTTON; BARTO, 2018).

Além destas, existem abordagens híbridas que combinam elementos de dois ou mais dos tipos básicos de aprendizado de máquina. Este tipo de aprendizado é conhecido como aprendizado conjunto ou, em inglês, *ensemble learning* (LAPPALAINEN; MISKIN, 2000). Estes métodos tentam tirar proveito das forças de cada abordagem e mitigar suas fraquezas.

2.4.2 Algoritmos de AM

Entre os algoritmos classificadores mais notáveis e amplamente utilizados no AM estão o *Naïve Bayes* (NB), *Logistic Regression* (LR), *Decision Tree* (DT), RF e *Support Vector Machine*

(SVM). Cada um desses modelos possui características únicas que os tornam apropriados para diferentes tipos de conjuntos de dados e problemas.

Os classificadores NB são uma família de classificadores probabilísticos simples baseados na aplicação do teorema de Bayes com a suposição de independência entre as características. Existem diferentes tipos de classificadores NB, incluindo o Bernoulli NB e o Gaussian NB, que são adequados para diferentes tipos de distribuição de dados.

O Bernoulli NB é recomendado quando as características são variáveis binárias. Cada característica é modelada como uma distribuição de Bernoulli independente (SCHÜTZE; MANNING; RAGHAVAN, 2008). Já o Gaussian NB a probabilidade das características é assumida como sendo gaussiana, ou normal, o que significa que se calcula a média e a variância dos valores das características para realizar a classificação dos dados (MCCALLUM; NIGAM *et al.*, 1998).

O LR é um modelo estatístico que, apesar do nome, é usado para classificação binária, não regressão (JR; LEMESHOW; STURDIVANT, 2013). O LR modela a probabilidade de uma característica binária ser verdadeira em função de uma ou mais variáveis independentes. É um modelo linear para classificação que utiliza a função logística para modelar a probabilidade condicional da classe positiva.

O modelo DT aprendem a partir dos dados de entrada a realizar uma série de decisões baseadas em questões sim/não ou critérios numéricos, levando a uma decisão final ou previsão. Esses modelos são intuitivos e a tomada de decisão é transparente. Os pontos de decisão ou nós são escolhidos com base na homogeneidade das amostras dentro dos nós resultantes, buscando maximizar a separação entre as diferentes classes de sentimentos nos dados de treinamento (QUINLAN, 1986).

O modelo RF é um conjunto de DT, geralmente treinadas com o método de *bagging*¹². A ideia por trás do RF é combinar várias DT para obter uma previsão mais precisa e estável. Cada DT no modelo é construída a partir de uma amostra do conjunto de dados, e a melhor divisão em cada nó é escolhida de um subconjunto aleatório de características (BREIMAN, 2001).

O SVM é um conjunto de métodos de aprendizado supervisionado utilizados para classificação e regressão. Em problemas de classificação, o objetivo do SVM é encontrar o hiperplano que melhor separa as classes de dados em um espaço de características de alta dimensão. O SVM busca maximizar a margem entre as classes, o que pode resultar em uma melhor generalização para dados não vistos (CORTES; VAPNIK, 1995).

Cada um desses modelos pode ser aplicado a dados de texto para análise de sentimentos após um processo adequado de pré-processamento e vetorização das características.

¹² *Bagging*, é um acrônimo para *Bootstrap Aggregating*, é um método de agregação projetado para melhorar a estabilidade e a precisão de algoritmos de AM. Ele funciona treinando cada modelo no conjunto usando uma amostra aleatoriamente selecionada do conjunto de dados de treinamento, este processo é conhecido como *bootstrap*. Após a geração das previsões de todos os modelos, o *bagging* combina essas previsões para formar uma única previsão final através de métodos como votação majoritária para classificação ou média para regressão (BREIMAN, 1996).

A escolha do modelo muitas vezes depende das características específicas dos dados e do problema, bem como das preferências do pesquisador e das limitações computacionais.

2.4.3 Aplicação de AM na Análise de Sentimentos

O campo da análise de sentimentos beneficia-se amplamente do uso de AM na busca de interpretar e classificar emoções expressas em textos. A união desses campos permite discernir se a expressão textual é positiva, negativa ou neutra, e, em contextos mais avançados, identificar emoções específicas como alegria, tristeza, raiva, entre outras. O uso de AM nesta área é primordial, pois permite a automatização da classificação de grandes volumes de dados textuais, que seriam impraticáveis de serem analisados manualmente (PANG; LEE *et al.*, 2008).

Os modelos de AM, como os já mencionados, têm desempenhado um papel vital na análise de sentimentos. Cada um desses modelos oferece uma abordagem única na interpretação de dados textuais, permitindo a extração de percepções e emoções das pessoas.

Na análise de sentimentos, um passo crucial é o pré-processamento dos dados, que inclui a limpeza e transformação de textos brutos em um formato que possa ser efetivamente processado pelos modelos. Isso geralmente envolve a remoção de ruídos (como URLs e menções de redes sociais), normalização (como a conversão de texto para minúsculas), e a vetorização das características (por exemplo, usando o TF-IDF ou Word2Vec). Após este pré-processamento, os dados estão prontos para serem utilizados pelos modelos de AM (AGGARWAL; AGGARWAL, 2015).

O NB, por exemplo, é frequentemente usado devido à sua simplicidade e eficiência, sendo particularmente eficaz em conjuntos de dados grandes e dispersos, comuns em textos de redes sociais. O LR e as SVMs são preferidas por sua robustez e capacidade de lidar com conjuntos de dados de alta dimensão, enquanto o DT e o RF oferecem uma abordagem intuitiva e são úteis quando a interpretabilidade do modelo é uma prioridade (LIU, 2022).

Cada modelo tem suas próprias vantagens e limitações, e a escolha do modelo adequado depende do contexto específico do conjunto de dados e dos objetivos da análise. Em muitos casos, uma combinação de diferentes modelos (aprendizado conjunto) é empregada para melhorar a precisão e a robustez das previsões.

2.5 Aprendizado Profundo (AP)

O advento do Aprendizado Profundo (AP) marca um ponto de inflexão significativo na evolução da IA. Surgindo como uma evolução do AM tradicional, o AP oferece maior capacidade de aprendizado e flexibilidade na modelagem de dados complexos. Diferenciando-se pela capacidade de aprender representações de dados em múltiplos níveis de abstração, consegue

capturar padrões muitas vezes imperceptíveis para técnicas mais simples (LECUN; BENGIO; HINTON, 2015).

Este campo se apoia substancialmente em redes neurais artificiais, especialmente aquelas com múltiplas camadas ocultas, denominadas redes neurais profundas (MIIKKULAINEN *et al.*, 2019). Estas redes são compostas por uma série de camadas, onde cada camada transforma sua entrada em uma representação mais abstrata e, frequentemente, mais informativa. Ao contrário das redes neurais tradicionais, que tendem a ter poucas camadas ocultas, as redes neurais profundas podem ter dezenas ou até centenas de camadas, o que lhes confere uma capacidade excepcional de aprender características complexas dos dados (BENGIO; GOODFELLOW; COURVILLE, 2017).

O aprendizado profundo demonstrou ser particularmente eficaz em tarefas que envolvem grandes quantidades de dados não estruturados, como imagens, áudio e texto. Em áreas como visão computacional, processamento de linguagem natural e reconhecimento de fala, os modelos de AP superaram consistentemente os métodos de AM tradicionais em termos de precisão e eficiência (KRIZHEVSKY; SUTSKEVER; HINTON, 2012; HINTON *et al.*, 2012).

A capacidade de operar diretamente em dados brutos é uma das principais vantagens do AP. Modelos tradicionais de AM muitas vezes dependem de um passo de engenharia de recursos manual e intensivo, onde os especialistas devem identificar e codificar as características dos dados relevantes para a tarefa em questão. Em contraste, redes neurais profundas são capazes de aprender automaticamente essas características representativas a partir dos próprios dados, reduzindo a necessidade de intervenção humana e aumentando a generalização dos modelos (BENGIO; COURVILLE; VINCENT, 2013).

Entretanto, as vantagens do AP vem acompanhado de desafios. Estes modelos requerem grandes quantidades de dados e recursos computacionais substanciais, especialmente GPUs¹³ e TPUs¹⁴, para treinamento efetivo. Além disso, a interpretabilidade dos modelos profundos permanece um desafio, uma vez que a complexidade das redes torna difícil entender como exatamente as decisões são tomadas (CASTELVECCHI, 2016).

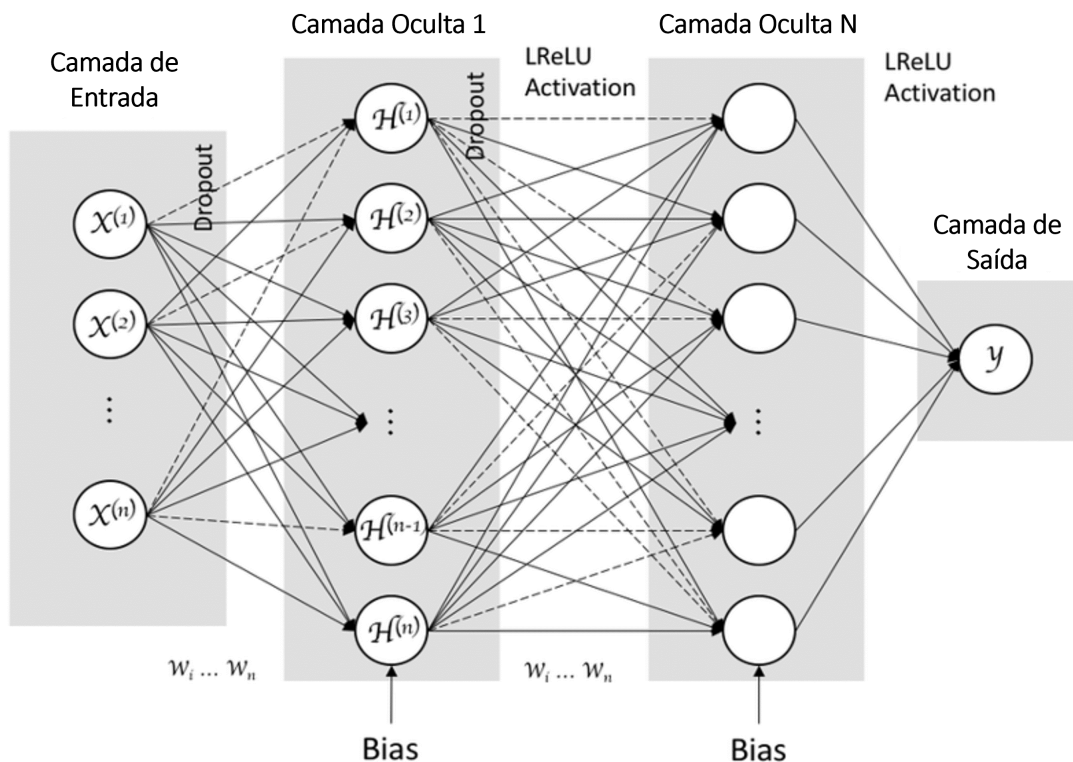
¹³ GPU é a sigla para Unidade de Processamento Gráfico, do inglês *Graphics Processing Unit*, é um processador otimizado para aceleração gráfica e cálculos de processamento paralelo. Em comparação com as Unidades de Processamento Central (CPUs) tradicionais, as GPUs são capazes de realizar uma quantidade significativamente maior de operações por unidade de tempo, tornando-as ideais para tarefas que exigem processamento de alto desempenho, como aprendizado de máquina, processamento de imagens e simulações complexas (OWENS *et al.*, 2008).

¹⁴ TPU é a sigla para Unidade de Processamento de Tensor, do inglês *Tensor Processing Unit*, é um circuito integrado de aplicação específica (ASIC) desenvolvido especificamente para acelerar operações de aprendizado de máquina. Desenvolvido pela Google em 2016, as TPUs são otimizadas para processar operações de matrizes de baixa precisão comumente usadas em redes neurais, oferecendo um aumento significativo de desempenho e eficiência energética em comparação com as GPUs tradicionais (JOUPI *et al.*, 2017).

2.5.1 Modelos de AP

O campo do aprendizado profundo apresenta uma gama diversa de modelos, cada um com características únicas e adequados para diferentes tipos de tarefas. Entre os modelos mais proeminentes, encontram-se as Redes Neurais Multicamadas, como o Perceptron de Multicamadas (*Multilayer Perceptron* - MLP), Redes Neurais Convolucionais (*Convolutional Neural Networks* - CNN), Redes Neurais Recorrentes (*Recurrent Neural Networks* - RNN), Memória de Curto e Longo Prazo (*Long Short-Term Memory* - LSTM) e Unidades de Portão Recorrente (*Gated Recurrent Unit* - GRU).

Figura 4 – Arquitetura básica do MLP.

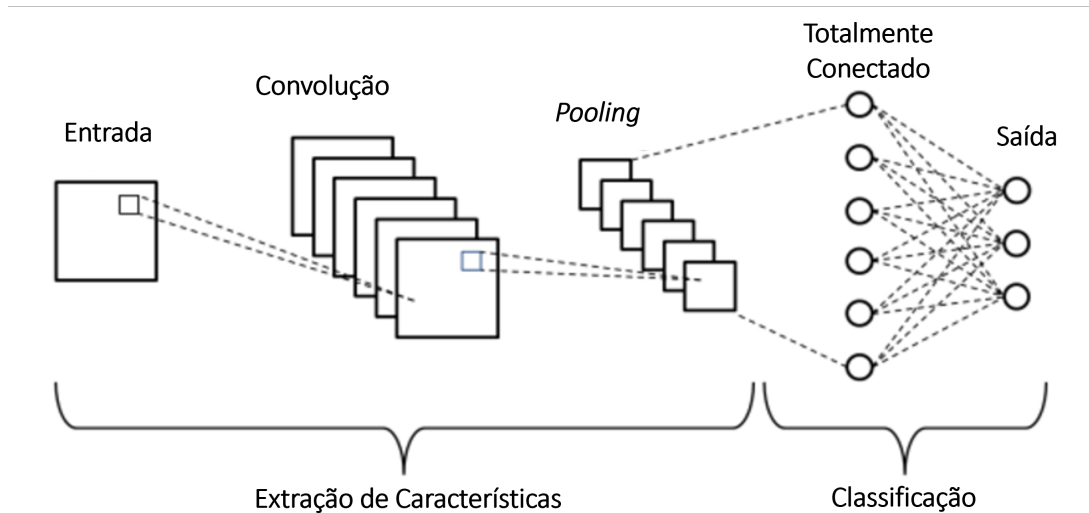


Fonte: Adaptado de Permanasari, Chamsudin e Wahyunggoro (2013).

O MLP é uma das formas mais básicas de redes neurais, como apresentado na Figura 4, é uma rede composta por múltiplas camadas de neurônios em uma estrutura de alimentação direta. Entre as camadas temos uma camada de entrada, uma ou múltiplas camadas ocultas e uma camada de saída. Cada neurônio em uma camada está conectado a todos os neurônios na camada seguinte, permitindo que a rede aprenda relações complexas entre os dados de entrada e saída. MLPs são particularmente úteis para problemas de classificação e regressão, mas podem ter dificuldade com dados sequenciais ou espaciais devido à sua natureza estática (BENGIO; GOODFELLOW; COURVILLE, 2017).

As CNNs são especializadas em processamento de dados com grade topológica, como imagens. Estas redes utilizam operações de convolução para extrair recursos hierárquicos dos dados, tornando-as extremamente eficientes no reconhecimento de padrões visuais. Além de

Figura 5 – Arquitetura básica de uma CNN.



Fonte: Adaptado de Phung e Rhee (2019).

aplicações em visão computacional, CNNs também mostraram ser eficazes em tarefas de processamento de linguagem natural, especialmente na análise de sentimentos de textos curtos, como *tweets* (KRIZHEVSKY; SUTSKEVER; HINTON, 2012). A Figura 5 apresenta a arquitetura básica de uma CNN, esta representação esquemática apresenta uma rede composta por cinco camadas distintas: camada de entrada, camada de convolução, camada de *pooling*, camada com neurônios totalmente conectados¹⁵ e camada de saída. Estas camadas são categorizadas em duas partes principais: extração de características e classificação (PHUNG; RHEE, 2019).

Na etapa de extração de característica, a camada de entrada define um tamanho fixo para os dados, que são redimensionados se necessário. Em seguida, os dados passam pela camada de convolução, onde são convoluídos com múltiplos núcleos de aprendizado, utilizando pesos compartilhados. A camada de *pooling* é responsável pela criação dos mapas de características, tentando preservar o máximo de informações contidas nos dados.

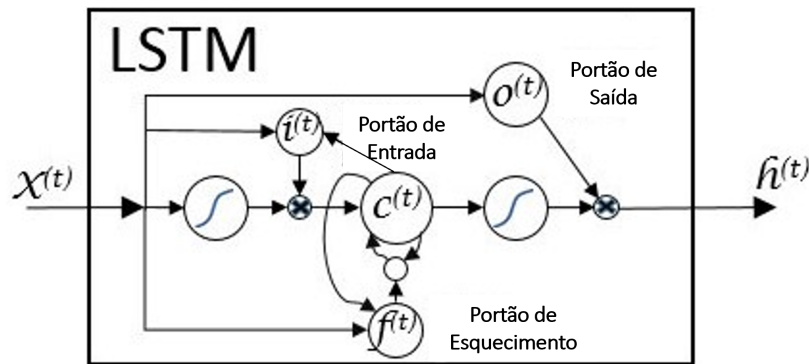
Na parte de classificação, as CNNs se assemelham a estrutura das MLPs, onde todos os neurônios da camada *fully-connected* se conectam a todos os neurônios da camada de saída. Por fim, a camada de saída apresenta o resultado da classificação feita pela CNN.

As RNNs são projetadas para lidar com dados sequenciais, como séries temporais ou texto. Uma das vantagens das RNNs é a capacidade de manter uma “memória” de seus estados anteriores, permitindo-lhes manter um contexto ao longo do tempo. No entanto, as RNNs tradicionais muitas vezes lutam com dependências de longo prazo devido ao problema do desaparecimento do gradiente (MEDSKER; JAIN, 2001).

Durante o treinamento de uma rede neural através da retropropagação, os gradientes dos pesos são calculados e propagados de volta da saída para as camadas de entrada. Este processo de aprendizado ao longo do tempo das RNNs é ilustrado na Figura 6, onde $X^{(t)}$ representa o vetor de características na camada de entrada, $Y^{(t)}$ o resultado da classificação no

¹⁵ Em inglês, *fully-connected layer*.

Figura 7 – Arquitetura dos portões de uma rede LSTM.

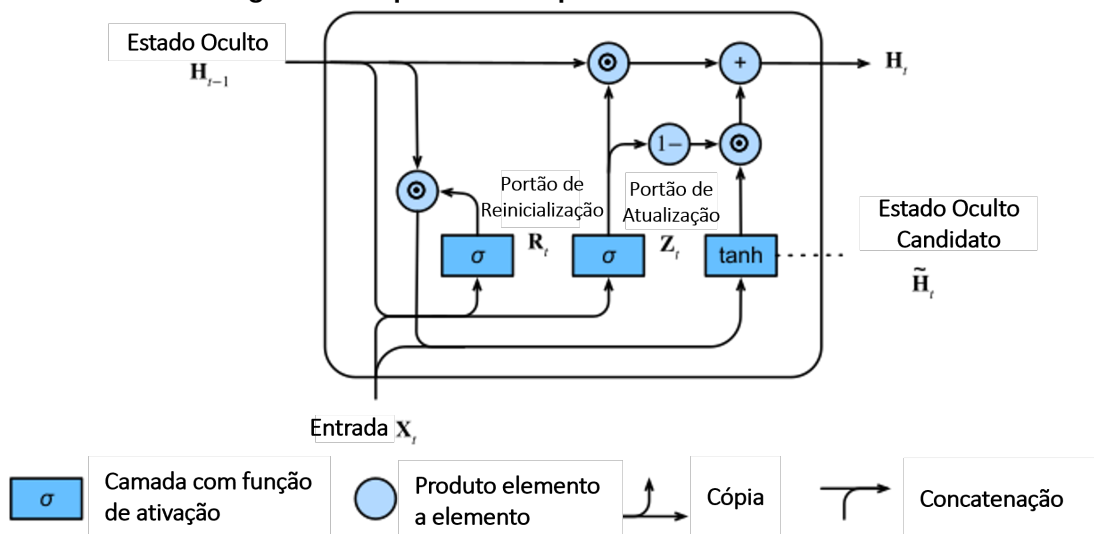


Fonte: Adaptado de Loukas *et al.* (2017).

a adição de novas informações ao estado da célula e o portão de saída determina quais partes do estado da célula devem ser passadas para a saída da rede. Este mecanismo facilita a aprendizagem de dependências de longo prazo, tornando as LSTMs ideais para tarefas complexas de sequência como previsão de séries temporais, reconhecimento de fala e, claro, análise de sentimentos em textos (HOCHREITER; SCHMIDHUBER, 1997).

A GRU, proposta por Cho *et al.* (2014), apresenta-se como uma alternativa simplificada ao LSTM. A GRU unifica o portão de esquecimento e o portão de entrada em um único "portão de atualização". A memória da GRU é gerenciada através de estados ocultos, que armazenam informações relevantes ao longo do treinamento. Estes incluem o estado oculto candidato e o estado oculto anterior. Essa simplificação resulta em uma redução da complexidade computacional e do número de parâmetros, conferindo à GRU maior velocidade e eficiência em certos contextos. Embora mais simples, a GRU alcança desempenho comparável ao do LSTM em diversas tarefas, como na análise de sentimentos. A escolha entre LSTM e GRU depende das especificidades e exigências computacionais da aplicação, enquanto a GRU oferece uma

Figura 8 – Arquitetura dos portões de uma rede GRU.



Fonte: Adaptado de Zhang *et al.* (2021).

abordagem mais simplificada, o LSTM proporciona um controle mais refinado do fluxo de informações, o que pode ser vantajoso conforme a natureza e complexidade dos dados sequenciais envolvidos.

A Figura 8 ilustra a arquitetura e o funcionamento de uma GRU. Seus elementos-chave são o portão de reinicialização, o portão de atualização, o estado oculto e o estado oculto candidato. O portão de reinicialização determina quanta informação do estado anterior deve ser retida, enquanto o portão de atualização decide se o novo estado oculto será baseado no estado oculto candidato ou no estado oculto anterior. Por exemplo, a GRU pode aprender a não atualizar o estado oculto após observações iniciais importantes, a ignorar informações temporárias irrelevantes ou até a redefinir o estado oculto conforme necessário.

Os portões são projetados como vetores com entradas variando entre 0 e 1, permitindo combinações convexas. O estado oculto candidato é gerado a partir da entrada atual e do estado oculto anterior, sendo modulado pela ação do portão de reinicialização. O novo estado oculto é formado pela combinação do estado oculto atual e do candidato, com o grau desta combinação regulado pelo portão de atualização.

2.5.2 Aplicação de AP na Análise de Sentimentos

A Análise de Sentimentos se beneficia imensamente das capacidades avançadas das redes neurais profundas, especialmente em tarefas como a interpretação de nuances contextuais e a detecção de sentimentos implícitos em dados textuais (ZHANG; ZHAO; LECUN, 2015).

Uma das principais vantagens do aprendizado profundo na análise de sentimentos é a sua habilidade em aprender representações de alto nível a partir de dados brutos. Modelos como CNNs e RNNs, incluindo suas variantes como LSTM e GRU, demonstraram ser particularmente eficazes na captura de padrões sequenciais e contextuais em textos (ZHANG; WANG; LIU, 2018). Por estas redes neurais conseguirem aprender a identificar padrões e dependências em longas sequências de palavras, capturando assim o contexto e o significado mais profundo nos textos, elas são particularmente eficazes na análise de sentimentos.

As CNNs, inicialmente destacadas no campo da visão computacional, têm ganhado reconhecimento significativo em aplicações de análise de sentimentos. Esta transição de domínios ilustra a versatilidade das CNNs em lidar com diferentes tipos de dados e tarefas. Na análise de sentimentos, as CNNs demonstram uma capacidade notável de extrair características locais de textos, incluindo frases ou sentenças individuais, e integrar essas informações para formar uma interpretação completa do texto (KIM, 2014). Esta habilidade é particularmente útil na identificação de palavras ou expressões chave que são indicativas de sentimentos positivos ou negativos.

Além disso, as CNNs são capazes de reconhecer padrões complexos em dados textuais, o que é crucial na análise de sentimentos, onde a expressão de emoções e opiniões pode ser sutil e variada. Esta eficácia das CNNs em identificar nuances em textos é destacada por (ZHANG; WANG; LIU, 2018), que evidenciam a competência dessas redes em capturar con-

textos específicos e variações linguísticas que são essenciais para compreender a polaridade sentimental de um texto.

Outra vantagem significativa das CNNs na análise de sentimentos é a sua capacidade de processar dados de entrada de forma hierárquica, identificando primeiro características de baixo nível e, posteriormente, integrando essas informações para identificar características de mais alto nível. Esta abordagem hierárquica é particularmente benéfica na análise de sentimentos, pois permite uma compreensão mais profunda e contextual do texto, que vai além da simples detecção de palavras-chave. Johnson e Zhang (2014) realçam essa capacidade das CNNs, demonstrando como podem eficientemente capturar tanto características locais quanto globais em textos, o que é fundamental para uma análise de sentimentos precisa e abrangente.

Por outro lado, as RNNs e, suas variantes, são especializadas em lidar com a natureza sequencial do texto. Por manter um "estado" ou "memória" ao longo do tempo, essas redes neurais conseguem entender o contexto e a sequência em que as palavras aparecem, elementos fundamentais para a interpretação correta dos sentimentos (HOCHREITER; SCHMIDHUBER, 1997; CHO *et al.*, 2014).

As LSTMs e GRUs, são empregadas em tarefas que dependem da captura de dependências de longo prazo em textos, o que é fundamental para entender a intenção e o sentimento em frases ou parágrafos mais extensos. A capacidade de lembrar informações passadas e usá-las para influenciar a interpretação atual é também útil na análise de sentimentos, onde o significado e a emoção podem ser construídos e evoluir ao longo do texto (HOCHREITER; SCHMIDHUBER, 1997; CHO *et al.*, 2014). Por esse motivo, são amplamente utilizadas em aplicações como monitoramento de opiniões em mídias sociais, análise de *feedback* de clientes e em sistemas de resposta automática, onde é vital a compreensão precisa do sentimento e da intenção através da linguagem.

2.6 Métricas de Desempenho para Modelos de Análise de Sentimentos

Na avaliação de modelos de AM e AP para análise de sentimentos, diversas métricas de classificação são empregadas para mensurar a eficácia e precisão dos modelos. Entre estas métricas, destacam-se a Matriz de Confusão, a Acurácia, a Precisão, o *Recall*, o *F1-Score*, a Curva *Receiver Operating Characteristic* (ROC) (Curva Característica de Operação do Receptor, ou apenas Curva ROC) e a Curva *Precision-Recall*, cada uma fornecendo percepções distintas sobre o desempenho do modelo.

A Matriz de Confusão é uma ferramenta essencial na avaliação de modelos de aprendizado de máquina, especialmente em tarefas de classificação. Ela fornece uma representação visual e tabular das classificações realizadas por um modelo, comparando as previsões do modelo (previstas) com os valores reais (verdadeiros). Esta matriz permite uma análise detalhada dos erros de classificação, destacando as ocorrências de Verdadeiro Positivo (VP), Falso Posi-

tivo (FP), Verdadeiro Negativo (VN) e Falso Negativo (FN), que são utilizados na definição das demais métricas como Acurácia, Precisão, *Recall* e *F1-Score* (LIANG, 2022).

Para ilustrar, considere uma tarefa de classificação binária, como a análise de sentimentos, onde as classes são 'positivo' e 'negativo'. A Matriz de Confusão para este cenário pode ser representada como a matriz na Tabela 1.

Tabela 1 – Exemplo de uma Matriz de Confusão

		Previsão	
		Negativo	Positivo
Real	Negativo	VN	FP
	Positivo	FN	VP

Fonte: Autoria própria (2023).

Nesta matriz, VP e VN representam as classificações corretas, enquanto FP e FN representam as classificações incorretas. A análise da Matriz de Confusão ajuda a identificar se o modelo está tendencioso para uma classe em particular e fornece informações sobre os tipos de erros que o modelo está cometendo. Isso é particularmente útil para ajustar o modelo e melhorar sua precisão (STEHMAN, 1997). Frequentemente, a Matriz de Confusão é apresentada com uma escala de cores, que serve para realçar visualmente os quadrantes com valores mais significativos e facilitar a interpretação dos resultados.

A Acurácia representa a proporção de previsões corretas (tanto positivas quanto negativas) em relação ao total de previsões feitas pelo modelo. Matematicamente, a acurácia é calculada pela fórmula:

$$Acurácia = \frac{VP + VN}{VP + VN + FP + FN}$$

A Acurácia é uma medida útil para obter uma visão geral rápida do desempenho do modelo, mas é importante notar que ela pode ser enganosa em conjuntos de dados com distribuição de classe desequilibrada (JAPKOWICZ; SHAH, 2011). Dessa forma, a Acurácia é frequentemente utilizada como um ponto de partida na avaliação de modelos, mas deve ser complementada com métricas mais detalhadas como Precisão, *Recall*, *F1-Score*, e outras, para obter uma compreensão mais completa e precisa do desempenho do modelo (POWERS, 2020).

A Precisão é uma métrica que indica a proporção de predições positivas corretas realizadas pelo modelo em relação ao total de valores positivos (POWERS, 2020). Matematicamente, é representada pela fórmula:

$$Precisão = \frac{VP}{VP + FP}$$

Em contrapartida, o *Recall*, ou Sensibilidade, mede a proporção de positivos reais corretamente identificados pelo modelo (POWERS, 2020), sendo calculado por:

$$Recall = \frac{VP}{VP + FN}$$

Já o *F1-Score* é uma métrica que harmoniza a Precisão e o *Recall*, e é particularmente útil em cenários de distribuição de classes desequilibradas (POWERS, 2020). É calculado pela média harmônica destas duas métricas, conforme a equação:

$$F1 - Score = 2 * \frac{Precisão * Recall}{Precisão + Recall}$$

A Curva ROC e a métrica da *Area Under the Curve* (AUC) são fundamentais para avaliar a eficácia de modelos de classificação binária. A Curva ROC é um gráfico que mostra o desempenho do modelo de classificação em todos os limiares possíveis, plotando a taxa de verdadeiros positivos, ou *Recall*, contra a taxa de falsos positivos para diferentes limiares de corte (FAWCETT, 2006).

A taxa de falsos positivos é calculada como sendo $1 - especificidade$, sendo que a especificidade, ou taxa de verdadeiros negativos, é a proporção de negativos reais que são corretamente identificados (POWERS, 2020). A especificidade é calculada pela fórmula:

$$Especificidade = \frac{VN}{FP + VN}$$

A Curva ROC é uma ferramenta poderosa porque permite a visualização do *trade-off* entre sensibilidade e especificidade para diferentes limiares, sem se comprometer com um limiar específico. A AUC, por sua vez, é um resumo numérico da Curva ROC. Ela representa a probabilidade de que um classificador classifique corretamente um exemplo positivo aleatório mais alto do que um exemplo negativo aleatório. Uma AUC igual a 1 representa um modelo perfeito que classifica todos os exemplos corretamente, enquanto uma AUC de 0,5 indica um desempenho não melhor do que o acaso. Em outras palavras, quanto maior a AUC, melhor é o modelo na distinção entre as classes positivas e negativas (BRADLEY, 1997).

A Curva *Precision-Recall*, é um gráfico bidimensional que mostra a relação entre a Precisão e o *Recall* para diferentes limiares de decisão (SAITO; REHMSMEIER, 2015). Ela fornece uma visão abrangente de como a eficácia do modelo varia com a alteração do limiar de classificação, permitindo uma análise detalhada da performance do modelo em diferentes níveis de sensibilidade e especificidade.

Um aspecto importante da Curva *Precision-Recall* é que, ao contrário da Curva ROC, ela se concentra apenas nas previsões positivas do modelo. Isso a torna particularmente útil em situações onde a classe positiva (por exemplo, a presença de um determinado sentimento) é de interesse especial e é relativamente rara em comparação com a classe negativa. Em tais casos, mesmo um modelo com alta acurácia pode ter uma performance enganadora se a maioria das previsões corretas for para a classe negativa. A Curva *Precision-Recall* ajuda a destacar a

eficácia do modelo em identificar a classe positiva em meio a um grande número de negativos (DAVIS; GOADRICH, 2006).

Essas métricas são cruciais para comunicar de forma eficaz os resultados de análise de sentimentos, fornecendo uma compreensão detalhada do desempenho do modelo em termos de precisão, capacidade de recuperação de instâncias positivas e generalização para dados não vistos. A interpretação correta dessas métricas é essencial para tomar decisões informadas sobre a seleção, o aperfeiçoamento e a aplicação prática dos modelos.

3 TRABALHOS RELACIONADOS

O estudo da análise de sentimentos em redes sociais, especialmente no Twitter, tem atraído considerável interesse acadêmico e comercial devido ao seu potencial em revelar padrões de comportamento, opiniões e tendências emergentes em tempo real. Esta área de pesquisa é interdisciplinar, abrangendo campos como Processamento de Linguagem Natural (PLN), Aprendizado de Máquina (AM), e Ciências Sociais. O presente capítulo visa contextualizar o trabalho atual em relação aos desenvolvimentos existentes, apresentando uma revisão criteriosa de estudos relevantes que exploram metodologias e aplicações variadas na análise de sentimentos no Twitter.

Um dos trabalhos pioneiros no campo de análise de sentimentos utilizando o Twitter como fonte de dados, foi desenvolvido por Go, Bhayani e Huang (2009). Eles introduziram uma abordagem inovadora para classificar automaticamente o sentimento de mensagens no Twitter, uma técnica que se tornou fundamental na análise de sentimentos em mídias sociais. Eles utilizaram o que é conhecido como supervisão distante, uma técnica de aprendizado de máquina que emprega dados rotulados de forma indireta. Neste caso, os dados consistiam em mensagens do Twitter contendo *emoticons*, usados como rótulos "ruidosos". O estudo explorou diferentes classificadores de aprendizado de máquina, como Naïve Bayes (NB), Máxima Entropia e SVM, e empregou uma série de extratores de características, incluindo unigramas, bigramas e combinações dos dois. O processo de pré-processamento desempenhou um papel crucial, envolvendo a remoção de nomes de usuários, URLs e a normalização de repetições de letras. Os resultados mostraram que esses algoritmos podem atingir uma precisão de 80% quando treinados com dados baseados em *emoticons*, destacando a eficácia da supervisão distante em contextos de mídia social.

Ainda nos primórdios dos estudos de análise de sentimentos em redes sociais, Pak, Paroubek *et al.* (2010) concentraram-se em utilizar o Twitter como um *corpus* para análise de sentimentos e mineração de opiniões. Eles destacaram a relevância do *microblogging* como uma ferramenta de comunicação popular e uma fonte rica de dados para a análise de opiniões. O trabalho envolveu a coleta automática de um *corpus* a partir do Twitter, seguida por uma análise linguística detalhada do mesmo. Para a construção do classificador de sentimentos, foram adotadas estratégias como filtragem, tokenização, remoção de *stopwords* e a construção de unigramas, bigramas e trigramas. Este estudo também enfatizou a importância da utilização de trigramas, melhorando a precisão da classificação dos sentimentos. O classificador de sentimentos foi construído usando o classificador NB multinomial, resultando em uma alta precisão na classificação de sentimentos positivos, negativos e neutros.

Em um estudo focado em trazer um contraponto aos modelos treinados com n-gramas e analisar a influência das características dos *tweets*, Agarwal *et al.* (2011) utilizaram algoritmos de SVM e mostraram a importância de incorporar elementos como *hashtags*, URLs e menções, juntamente com a polaridade e etiquetagem de partes da fala na análise de sentimentos. Eles

demonstraram como esses fatores poderiam influenciar a precisão do modelo, com um ganho de 4% na acurácia em comparação com os modelos baseados em n-gramas.

No estudo desenvolvido por Wang *et al.* (2011), foi explorado o uso apenas de *hashtags* para a análise de sentimentos em *tweets*. Em vez de analisar a polaridade sentimental de cada *tweet*, o estudo se concentrou na classificação de sentimentos no nível das *hashtags* utilizando algoritmos de SVM. A pesquisa revelou que três tipos de informações são úteis para essa tarefa: (1) a polaridade sentimental dos *tweets* contendo a *hashtag*, (2) a relação de coocorrência de *hashtags* e (3) o significado literal das *hashtags*. Para incorporar essas informações em um modelo de classificação, os autores propuseram um novo modelo gráfico e investigaram três algoritmos de classificação. Além disso, eles demonstraram que o desempenho pode ser significativamente aprimorado usando um ambiente de classificação aprimorado por meio de *boosting*, no qual empregam o significado literal das *hashtags* como informação semi-supervisionada. Os resultados experimentais em um conjunto de dados consistindo em 29.195 *tweets* e 2.181 *hashtags*, mostraram a eficácia do modelo e dos algoritmos propostos, chegando a uma acurácia alta de 83,13%, porém com uma precisão e um *recall* baixos, com valores de 59,45% e 36,59%, respectivamente.

Saif *et al.* (2013) realizaram um estudo comparativo de vários conjuntos de dados em dimensões como número total de *tweets*, tamanho do vocabulário e grau de dispersão. Além disso, investigaram as correlações entre essas dimensões e o desempenho da classificação de sentimentos em diferentes conjuntos de dados. Os resultados obtidos através de modelos de SVM, mostraram que a capacidade de detectar *tweets* positivos é mais desafiadora do que detectar *tweets* negativos, com uma precisão média de 80,37% para *tweets* positivos contra 84,53% para *tweets* negativos. Curiosamente, eles descobriram que não existe uma correlação significativa entre o desempenho da classificação e o grau de dispersão dos dados entre os diferentes conjuntos de dados.

Koto e Adriani (2015) investigaram 41 atributos textuais, a partir de nove conjuntos diferentes e analisaram a sua eficácia em determinar os sentimentos expressos nos *tweets*. Entre os dados analisados, incluíam elementos como pontuação, opinião léxica, partes do discurso, *emoticons*, e várias métricas de sentimentos. Os resultados revelaram que os léxicos de sentimento AFINN¹ e *Senti-Strength*², junto com a opinião léxica, foram frequentemente identificados

¹ O léxico AFINN é um conjunto de listas de palavras com anotações de sentimentos associadas, desenvolvido por Finn Årup Nielsen. Cada palavra no léxico é atribuída a uma pontuação de sentimento, variando de -5 (muito negativo) a +5 (muito positivo), que é utilizada para avaliar o sentimento geral de textos curtos, como *tweets*. Este léxico é conhecido por sua simplicidade e eficácia em contextos de mídia social (NIELSEN, 2011).

² *Senti-Strength* é um método de classificação de sentimentos desenvolvido por Mike Thelwall. Ele é projetado especificamente para trabalhar com textos curtos da Internet, como *posts* de mídia social e comentários. *Senti-Strength* estima a força do sentimento positivo e negativo em um texto, dando duas pontuações separadas para cada dimensão (positiva e negativa). Sua abordagem é baseada no reconhecimento de palavras e expressões sentimentais, sendo muito útil na análise de sentimentos em plataformas como o Twitter (THELWALL *et al.*, 2010).

como os melhores em termos de precisão em cada conjunto de dados. Surpreendentemente, o léxico *SentiWordNet*³, amplamente conhecido, não se mostrou tão eficaz para a análise de sentimentos no Twitter. Além disso, características como emoção e pontuação demonstraram ser úteis, especialmente na classificação da polaridade.

No estudo de Sosa (2017), são propostos dois modelos híbridos de redes neurais, CNN-LSTM e LSTM-CNN, para análise de sentimentos em dados do Twitter. Estes modelos visam combinar as capacidades das duas redes neurais profundas para melhorar a classificação de sentimentos. O modelo CNN-LSTM inicia com uma camada convolucional que extrai características locais dos *embeddings* de palavras, seguida por uma camada LSTM que aprende a ordem das características. Inversamente, o modelo LSTM-CNN começa com uma camada LSTM que codifica a sequência de entrada, seguida por uma camada convolucional para detecção de padrões locais. Os resultados mostraram que o modelo LSTM-CNN superou os modelos CNN e LSTM regulares em 8,5% e 2,7%, respectivamente, indicando a eficácia dessa abordagem híbrida na análise de sentimentos no Twitter.

Goularas e Kamis (2019) avaliaram diferentes configurações de redes neurais profundas para a análise de sentimentos em dados do Twitter. Este trabalho investigou a combinação de CNN e LSTM, comparando também diferentes sistemas de *embeddings* de palavras, como Word2Vec e GloVe. Utilizando dados do *workshop* internacional SemEval, os autores observaram que a utilização do GloVe melhorou o desempenho em 5%-7% na maioria das configurações, e que configurações múltiplas de CNN e LSTM superaram configurações mais simples. O estudo também revelou que essa combinação é mais eficaz do que o uso isolado de cada uma, devido à redução eficaz da dimensionalidade pelas CNNs e à preservação das dependências de palavras pelas redes LSTMs.

O trabalho de Swathi, Kasiviswanath e Rao (2022) representa uma contribuição significativa na intersecção da análise de sentimentos em redes sociais e previsão de preços de ações, um tópico de crescente interesse na engenharia financeira. Os autores propõem um modelo inovador que integra a Otimização Baseada em Ensino e Aprendizagem (TLBO) com LSTM para realizar a análise de sentimentos. O modelo TLBO-LSTM é aplicado para classificar os *tweets* em sentimentos positivos ou negativos relacionados aos preços das ações, com o objetivo de investigar a correlação entre as opiniões expressas no Twitter e as variações no mercado de ações. Os resultados experimentais, obtidos a partir de dados do Twitter, são notáveis: o modelo TLBO-LSTM exibe um desempenho superior em comparação com métodos do estado da arte ao classificar o sentimento de um *tweet*, alcançando uma precisão máxima de 95,33%, *recall* de 85,28% e *F1-Score* de 90%.

³ *SentiWordNet* é uma extensão do recurso léxico *WordNet*, onde cada conjunto de sinônimos é associado a três pontuações numéricas: objetividade, positividade e negatividade. Estas pontuações indicam quão objetiva, positiva ou negativa é uma palavra. Desenvolvido por Esuli e Sebastiani, o *SentiWordNet* é frequentemente utilizado em tarefas de análise de sentimentos, oferecendo uma abordagem baseada em léxico para identificar e quantificar sentimentos em textos (SEBASTIANI; ESULI, 2006).

No último ano, de 2023, alguns trabalhos se destacaram no campo da análise de sentimentos no Twitter. O estudo de Saranya e Usha (2023) se destaca pela sua abordagem inovadora no tratamento de dados de mídia social, que são frequentemente curtos, informais, e repletos de ambiguidades linguísticas e ruídos. O artigo propõe uma técnica baseada em aprendizado de máquina que se concentra na extração de características a partir da Frequência do Termo e Frequência Inversa do Documento (TF-IDF), combinada com a aplicação de um lematizador inteligente *WordNet* para refinar a qualidade dos *tweets* ao eliminar ruídos irrelevantes. O algoritmo de RF é escolhido para a tarefa de classificação multi-classe das emoções expressas nos *tweets*, obtendo uma acurácia média de 89,92%, mantendo esse mesmo valor no *recall* e na precisão.

Ainda temos o estudo realizado por Aslan, Kızıloluk e Sert (2023), que estuda o impacto psicológico da pandemia de COVID-19 através da análise de sentimentos no Twitter. Utilizando uma API personalizada, os autores coletaram 173.638 *tweets* relacionados à COVID-19, cobrindo um período crítico de disseminação global da doença. A metodologia proposta, denominada TSA-CNN-AOA, combina CNN com um algoritmo de otimização aritmética (AOA) para melhorar a extração e seleção de características significativas dos *tweets*. Para a classificação dos *tweets* em categorias de sentimentos positivos, negativos e neutros, foram empregadas técnicas como *K-Nearest Neighbors* (KNN), SVM e DT. O destaque da pesquisa é a performance do modelo TSA-CNN-AOA, especialmente quando utilizando KNN, que alcançou uma precisão de classificação de 95,098%.

Por fim, o estudo de Mann *et al.* (2023), explora o aprimoramento da análise de sentimentos utilizando modelos avançados utilizando o BERT para a classificação de sentimentos expressos em *tweets*. Os autores utilizam o conjunto de dados *SMILE*⁴, que fornece uma variedade de expressões emocionais, como felicidade, raiva e tristeza, presentes em *tweets*. O objetivo é categorizar com precisão esses sentimentos, aproveitando a robustez do modelo proposto. Os resultados do estudo são notáveis, com o modelo aprimorado alcançando uma precisão de 96% na classificação de sentimentos.

Embora uma vasta gama de estudos sobre análise de sentimentos no Twitter tenha sido revisada, abrangendo diferentes abordagens e técnicas, notou-se uma lacuna significativa na literatura no que diz respeito à análise de sentimentos focada em públicos específicos de esportes, particularmente na Fórmula 1 (F1) ou outros esportes automobilísticos. Essa omissão destaca uma oportunidade única de pesquisa e inovação. A F1, sendo um esporte global com uma base de fãs diversificada e engajada, oferece um campo fértil para análises sentimentais devido às diversas opiniões e emoções expressas pelos fãs em plataformas de mídia social. Portanto, a incorporação de técnicas avançadas de análise de sentimentos no contexto da F1 pode não apenas enriquecer a compreensão dos padrões de comportamento e percepções dos fãs, mas também abrir novos caminhos para pesquisas relacionadas à interação entre esportes,

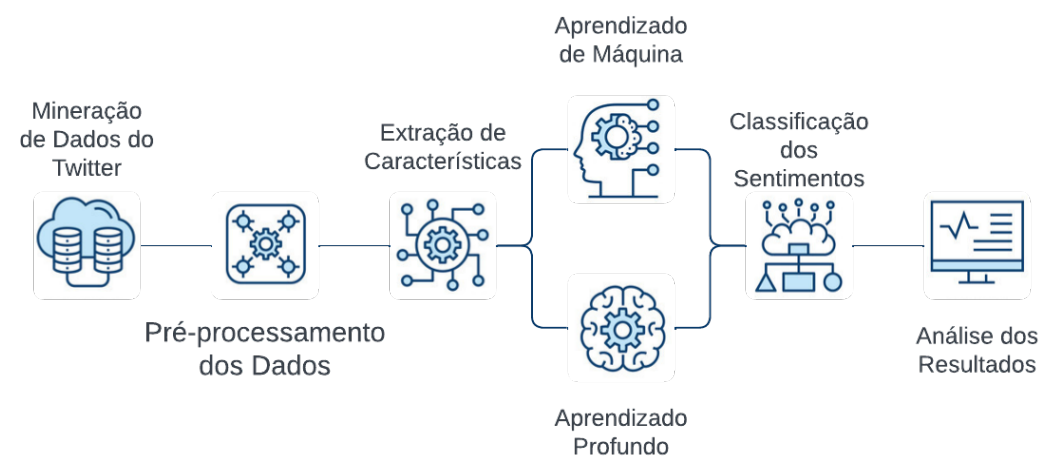
⁴ O *dataset SMILE* pode ser encontrado em: <https://www.kaggle.com/datasets/ashkhagan/smile-twitter-emotion-dataset>

mídia social e análise de dados. Este foco inexplorado, portanto, estabelece uma margem substancial para contribuições significativas por meio do trabalho em questão, visando preencher essa lacuna e explorar as nuances específicas da comunicação e expressão de sentimentos no universo dos esportes automobilísticos.

4 METODOLOGIA EXPERIMENTAL

No desenvolvimento deste trabalho, a aplicação de técnicas apropriadas e rigorosas para a análise de dados é fundamental para alcançar resultados confiáveis e significativos. A Figura 9 apresenta uma visão geral da metodologia experimental empregada na condução deste trabalho. Na figura é descrito o *pipeline* completo para geração das classificações e análises referentes ao sentimentos dos usuários do Twitter ao longo da temporada 2022 da Fórmula 1 (F1). As subseções a seguir detalham cada uma das sub-etapas envolvidas no experimento.

Figura 9 – Pipeline desenvolvida para a análise de sentimentos.



Fonte: Autoria própria (2023).

4.1 Dataset

No desenvolvimento deste trabalho, inicialmente, a mineração de dados via API do Twitter foi planejada, mas devido a alterações na política de acesso da plataforma, esta abordagem tornou-se inviável. Em substituição, optou-se pela utilização de dois conjuntos de dados públicos. O principal conjunto de dados empregado foi o Sentiment140, disponível na plataforma Kaggle¹. Este *dataset*, originalmente desenvolvido por Go, Bhayani e Huang (2009), é composto por aproximadamente 1,6 milhões de *tweets*. Notavelmente, o Sentiment140 foi criado com uma abordagem automatizada de geração de rótulos, assumindo que *tweets* contendo *emojicons* positivos, como ':)', eram positivos, e aqueles com *emojicons* negativos, como ':(', eram negativos.

¹ O dataset Sentiment140 pode ser encontrado em: <https://www.kaggle.com/datasets/kazanova/sentiment140?select=training.1600000.processed.noemoticon.csv>

Tabela 2 – Colunas presentes no *dataset* Sentiment140.

Nome da Coluna	Descrição
target	Polaridade do <i>tweet</i> , com rótulos 0 para negativo, 2 para neutro, e 4 para positivo.
id	Número inteiro que identifica o <i>tweet</i> .
date	Data de postagem do <i>tweet</i> .
flag	Termo de busca para encontrar o <i>tweet</i> , se não há nenhum termo é utilizado o valor "NO_QUERY".
user	Nome do usuário.
text	Texto do <i>tweet</i> .

Fonte: Autoria própria (2023).

A Tabela 2 apresenta as colunas presentes no *dataset* Sentiment140. É importante mencionar que, embora o *dataset* originalmente incluía uma categoria neutra (rótulo 2), a versão utilizada neste trabalho consiste apenas de *tweets* classificados como negativos (800.000) ou positivos (800.000), com a categoria neutra ausente. Para fins de padronização e simplificação, o rótulo da classe de sentimentos positivos foi alterado de 4 para 1.

No desenvolvimento dos modelos de AM e AP neste trabalho, foram utilizadas apenas as colunas de *target* e *text* do *dataset*. A escolha por limitar as variáveis utilizadas visa focar exclusivamente no conteúdo textual dos *tweets* e sua polaridade sentimental, essenciais para o treinamento e validação dos modelos propostos.

Além do Sentiment140, este trabalho também incorporou o uso de um segundo conjunto de dados, especificamente relacionado a *tweets* sobre F1. Este *dataset*, intitulado "*Formula 1 (F1) trending tweets*", abrange os *tweets* postados com a *hashtag* "#f1" no período de 24 de julho de 2021 a 20 de agosto de 2022. Compreendendo um total de 628.360 *tweets*. Contudo, ao contrário do Sentiment140, este conjunto de dados não apresenta uma metodologia específica empregada para a coleta dos *tweets*.

A Tabela 3 apresenta as colunas que constam no *dataset* sobre F1. Para os propósitos deste trabalho, apenas as colunas *text* e *date* foram utilizadas. A coluna *text* é essencial para a aplicação dos modelos de análise de sentimentos, enquanto a coluna *date* permite realizar uma análise exploratória dos sentimentos em relação aos eventos de F1. Assim, o uso restrito a essas duas colunas tem como objetivo focar na identificação de padrões de sentimentos e na correlação destes com eventos específicos no mundo da F1.

Tabela 3 – Colunas presentes no *dataset* de *tweets* sobre Fórmula 1.

Nome da Coluna	Descrição
user_name	Nome do usuário, conforme definido por ele.
user_location	Localização definida pelo usuário para o perfil.
user_description	Descrição definida pelo usuário para sua conta.
user_created	Data e hora de criação da conta do usuário.
user_followers	Número de seguidores do usuário.
user_friends	Número de amigos do usuário.
user_favourites	Número de favoritos do usuário.
user_verified	Indica se a conta do usuário é verificada.
date	Data e hora UTC de postagem do <i>tweet</i> .
text	Texto do <i>tweet</i> .
hashtags	<i>Hashtags</i> adicionais no <i>tweet</i> além de #1.
source	Ferramenta utilizada para postar o <i>tweet</i> .
is_retweet	Indica se o <i>tweet</i> foi repostado pelo usuário autenticado.

Fonte: Autoria própria (2023).

4.2 Pré-processamento

A preparação dos *datasets* para este trabalho envolveu um processo meticuloso e abrangente de pré-processamento de texto, essencial para assegurar a qualidade e a padronização dos dados antes das etapas de modelagem e análise. Este processo incluiu a conversão de todo o texto para letras minúsculas, facilitando a uniformidade e reduzindo a variabilidade nos dados. As abreviações em inglês foram substituídas usando um dicionário específico, construído especificamente para esse trabalho, com mais de 350 traduções, garantindo a clareza e a compreensão do texto. *Emojis* (versões modernas dos *emoticons*), uma forma comum de representar emoções em mídias sociais, foram traduzidos para suas representações textuais correspondentes², o que permitiu a análise desses elementos como parte do conteúdo textual.

URLs presentes nos *tweets* foram extraídas e substituídas por informações detalhadas sobre o domínio, caminho e parâmetros. Isso foi crucial para remover informações irrelevantes para a análise de sentimentos e concentrar-se no conteúdo textual. Da mesma forma, menções

² Por exemplo, o *emoji* 😊, é convertido para o texto *smiley face* ou, em português, rosto sorridente.

e *hashtags* foram extraídas e substituídas por representações textuais claras, empregando expressões regulares para identificar e substituir esses elementos por palavras-chave como "*mention*" e "*hashtag*", seguidas pelos respectivos nomes de usuários e tópicos.

A remoção de pontuações e espaços repetidos foi realizada para limpar o texto e remover elementos que poderiam interferir na análise. Isso incluiu a eliminação de caracteres de pontuação, com exceção de pontos e vírgulas entre números, e a substituição de múltiplos espaços por um único espaço. Essas ações foram fundamentais para garantir a consistência e a legibilidade dos dados de texto.

Após essas etapas iniciais de limpeza, o texto foi submetido a tokenização e lematização. A tokenização dividiu o texto em palavras individuais, enquanto a lematização, converteu as palavras para suas formas base, ou lemas. Palavras consideradas como "*stop words*"³ foram removidas para reduzir o ruído e focar em palavras significativas para a análise.

Esses procedimentos de pré-processamento garantiram que os dados estivessem limpos, padronizados e prontos para a etapa de extração de características.

4.3 Extração de Características

No desenvolvimento deste trabalho, a extração de características dos *datasets* foi um passo fundamental, especialmente considerando as diferenças entre os modelos de Aprendizado de Máquina (AM) e Aprendizado Profundo (AP). Para os modelos de AM, além da utilização de n-gramas (podendo ser unigramas, bigramas, trigramas ou maiores, bem como, a combinação de vários tamanhos em uma matriz), que são sequências de palavras usadas para capturar o contexto dentro do texto, recursos adicionais foram criados. Um desses recursos foi o *embedding* de palavras, realizado utilizando o modelo pré-treinado "Google News *negative 300*" do Word2Vec, uma técnica amplamente reconhecida para gerar representações vetoriais de palavras que capturam relações semânticas e sintáticas. Outro recurso importante foi o *score* de polaridade do texto, obtido por meio do TextBlob, uma biblioteca que fornece uma análise de sentimentos simples mas eficaz.

Para estabelecer a configuração mais eficaz de n-gramas, foi conduzido um experimento sistemático empregando a técnica de TF-IDF. Este experimento envolveu a análise de seis conjuntos de dados distintos, cada um representando uma configuração específica de n-gramas. As configurações testadas incluíram unigramas, bigramas e trigramas de forma isolada, além

³ *Stop Words* são palavras comumente usadas em uma língua que são filtradas antes ou durante o processamento de dados de texto. Geralmente, estas palavras têm pouco valor semântico e são removidas para aumentar a eficiência dos algoritmos de PLN. Exemplos de *stop words* em inglês incluem preposições, conjunções e artigos, como "in", "and", "the", "is", entre outros. Em português, exemplos incluem "e", "ou", "a", "o", "que", "de", "em", entre outros. A remoção de *stop words* é uma prática comum em tarefas de processamento de texto, como análise de sentimentos, classificação de texto e modelagem de tópicos, pois permite que os algoritmos se concentrem em palavras que têm maior potencial de carregar significado importante para a análise (GRISHMAN, 2015).

de suas combinações: 1-2 gramas, 2-3 gramas e, finalmente, uma configuração que integrava unigramas, bigramas e trigramas.

Cada um desses *datasets* foi submetido a uma avaliação utilizando quatro diferentes modelos de AM, que são NB, SVM, LR e RF. Para garantir a consistência e a comparabilidade dos resultados, todos os modelos foram implementados com os hiperparâmetros padrão fornecidos pela biblioteca scikit-learn. Este procedimento metodológico foi escolhido para determinar qual configuração de n-gramas seria mais eficiente na captura das nuances e variações presentes nos dados textuais para a análise de sentimentos.

Em contraste, para os modelos de AP, que são capazes de trabalhar com dados sequenciais, como textos, o método tradicional de n-gramas, TF-IDF (Frequência de Termo-Inverso da Frequência nos Documentos), foi substituído pelo tokenizer e pelo método `fit_on_texts` do Keras, biblioteca de rede neural de código aberto escrita em Python. Esta abordagem é mais adequada para modelos de AP, pois eles requerem uma representação dos dados que preserve a sequência e a estrutura das palavras no texto.

Todas estas características extraídas e processadas foram combinadas em uma única matriz esparsa⁴, que permite uma representação eficiente e compacta dos dados, dada a alta dimensionalidade e a natureza esparsa dos dados textuais. As matrizes resultantes, juntamente com as informações relevantes, foram salvas em arquivos serializados com *pickle*⁵. Estes arquivos foram posteriormente utilizados na etapa de treinamento dos modelos, facilitando o acesso eficiente aos dados e características necessárias para o processo de aprendizado dos modelos.

4.4 Algoritmos de AM

Para a etapa de AM deste trabalho, diversos modelos foram cuidadosamente selecionados e treinados, cada um trazendo características e abordagens distintas para a tarefa de classificação de sentimentos. O modelo de LR foi escolhido para testar a hipótese de que os sentimentos expressos nos textos apresentam características lineares. Esta escolha se baseia na capacidade do LR de modelar relações lineares entre variáveis independentes e a probabilidade de ocorrências de eventos específicos, como a classificação de sentimentos em categorias positivas ou negativas.

⁴ Uma matriz esparsa é uma matriz na qual a maioria dos elementos é zero. Em contextos de processamento de dados, como PLN, matrizes esparsas são comuns, especialmente ao lidar com grandes vocabulários onde muitas palavras não aparecem em cada documento. Elas são eficientes em termos de armazenamento, pois permitem representar dados com muitos zeros de maneira compacta.

⁵ *Pickle* é um módulo em Python utilizado para serializar e deserializar objetos. Serialização refere-se ao processo de converter um objeto em Python em um fluxo de bytes, permitindo que ele seja armazenado em um arquivo ou transmitido pela rede. Isso é particularmente útil para salvar modelos de AM e AP ou grandes conjuntos de dados de forma que possam ser carregados e utilizados posteriormente sem a necessidade de reprocessamento.

Outros modelos consagrados na literatura e com base teórica sólida foram empregados, incluindo o NB, o SVM, o RF e o MLP. Foi utilizado o Naïve Bayes Gaussiano, conhecido por sua eficiência em tarefas de classificação, especialmente em conjuntos de dados com alta dimensionalidade. O SVM é amplamente utilizado devido à sua robustez e eficácia em encontrar a melhor margem de separação entre classes. O RF, por sua vez, é um método de *ensemble* que utiliza múltiplas árvores de decisão para aumentar a precisão e evitar o Sobreajuste (*overfitting*). Por fim, MLP, uma forma simples de rede neural, é utilizado por sua habilidade em capturar relações complexas nos dados.

Na configuração dos hiperparâmetros dos modelos utilizados neste trabalho, adotou-se a decisão de empregar as configurações padrão disponibilizadas pela biblioteca scikit-learn. Essa escolha foi guiada principalmente por considerações práticas relacionadas ao tempo e aos recursos disponíveis. O refinamento de hiperparâmetros, também conhecido como *"tuning"*, é um processo que pode ser bastante demorado e exigir significativa capacidade computacional, especialmente em contextos onde múltiplos modelos e um grande volume de dados estão envolvidos.

Dada essa realidade, e com o objetivo de manter um equilíbrio eficaz entre precisão do modelo e viabilidade prática, optou-se por não realizar o ajuste fino dos hiperparâmetros. Contudo, fez-se uma alteração no hiperparâmetro *"n_jobs"*, definindo-o para -1 nos algoritmos onde tal configuração era aplicável. Esta modificação tem o objetivo de otimizar o processo de treinamento, permitindo que o modelo utilize todos os processadores da máquina disponíveis. Tal ajuste é particularmente útil para acelerar o treinamento sem comprometer a qualidade dos resultados, uma consideração importante em projetos com limitações de tempo e recursos.

4.5 Algoritmos de AP

No desenvolvimento dos modelos de Aprendizado Profundo (AP) para este trabalho, foram empregadas arquiteturas de CNN e LSTM, sendo ambas configuradas para executarem por 100 épocas e com um *early stop* configurado para 10 épocas e baseado no sobreajuste da perda de validação.

A arquitetura do modelo CNN foi cuidadosamente projetada para capturar padrões espaciais nos dados textuais. Inicia-se com uma camada de entrada, que recebe a sequência de *tokens* provenientes do texto. Em seguida, uma camada de *embedding* é utilizada para transformar esses *tokens* em vetores densos, uma representação mais rica e informativa. Posteriormente, uma camada convolucional 1D é aplicada para extrair padrões sequenciais do texto. Esta camada é seguida por uma camada de *max pooling*, que desempenha um papel crucial na redução da dimensionalidade dos dados, e uma camada de achatamento, que prepara os dados para a fase de classificação. O modelo integra camadas densas que processam as representações de word2vec e a polaridade dos sentimentos. Essas informações são combinadas com os recursos extraídos do texto para tomar a decisão final de classificação.

Por outro lado, o modelo LSTM foi estruturado para aproveitar a habilidade desta arquitetura em capturar dependências de longo prazo dentro de sequências textuais. Similar ao CNN, o modelo LSTM começa com uma camada de entrada que recebe *tokens*, seguida por uma camada de *embedding* que os converte em vetores densos. A característica distintiva do LSTM reside em sua camada homônima, que inclui *dropout* para regularização, uma técnica eficaz para prevenir o sobreajuste em redes neurais. Esta camada é essencial para capturar e preservar as informações ao longo da sequência de texto, permitindo que o modelo entenda e processe dependências de longo alcance que são fundamentais na análise de sentimentos. As camadas densas subsequentes processam as representações de word2vec e a polaridade dos sentimentos, culminando em uma camada de saída que utiliza a função de ativação sigmoide, ideal para tarefas de classificação binária.

Ambos os modelos, CNN e LSTM, foram, portanto, desenhados com uma abordagem sequencial e hierárquica, considerando as peculiaridades do processamento de texto e a natureza dos dados de análise de sentimentos. Essas arquiteturas permitem a extração eficiente de características e padrões relevantes dos dados textuais, fundamentais para a classificação precisa dos sentimentos expressos nos textos.

4.6 Setup Experimental

O *setup* experimental para este projeto foi meticulosamente planejado e executado, empregando tecnologias modernas e práticas de desenvolvimento para garantir a consistência e replicabilidade dos resultados. A estruturação do ambiente de desenvolvimento foi realizada utilizando Docker e Docker Compose, proporcionando uma plataforma de execução isolada e consistente para as aplicações em contêineres. Esse *setup* garantiu uma configuração uniforme e reproduzível do ambiente de desenvolvimento, essencial para validar e replicar os experimentos e resultados obtidos.

O desenvolvimento do código foi realizado em duas interfaces de desenvolvimento integrado (IDEs) distintas, JupyterLab e Visual Studio Code (VSCode). O JupyterLab foi utilizado para a experimentação e testes rápidos, devido à sua natureza interativa e facilidade de uso. Por outro lado, o VSCode foi empregado para a criação de *scripts* mais robustos, aproveitando sua ampla gama de funcionalidades e suporte efetivo para desenvolvimento em Python.

O servidor utilizado para execução dos *scripts* foi um HPE ProLiant DL360 Gen10, equipado com 24 CPUs Intel Xeon Silver 4214 a 2.20GHz. Duas máquinas virtuais (VMs) foram configuradas no servidor para a execução dos algoritmos, cada uma com 8 núcleos de processamento, 64 GB de RAM, 96 GB de armazenamento e sistema operacional Ubuntu Server 22.04 LTS.

A programação do projeto foi conduzida na linguagem Python, escolhida por sua ampla aceitação na comunidade científica e suporte extenso para bibliotecas de PLN e AM. Diversos pacotes Python foram utilizados, incluindo bibliotecas para pré-processamento de dados,

Tabela 4 – Ferramentas e bibliotecas utilizadas no projeto

Categoria	Ferramenta	Objetivo
Containerização	Docker	Garantir a consistência na execução do projeto em diferentes máquinas e sistemas operacionais.
IDE	JupyterLab	Desenvolvimento e teste interativo de scripts.
	Visual Studio Code	Criação e desenvolvimento robusto de scripts.
Servidor	HPE ProLiant DL360 Gen10 2.20 GHZ, com 24 CPUs Intel Xeon Silver 4214	Processamento intensivo de grande volume de dados.
Máquinas Virtuais	8 núcleos de processamento 64 GB de RAM 98 GB de armazenamento Ubuntu Server 22.04 LTS	Possibilitar a execução paralela dos scripts.
Linguagem de Programação	Python	Desenvolvimento do projeto e suporte a bibliotecas de CD.
	Re, Emoji, Polars, Numpy, NLTK e URLLib	Limpeza e preparação dos dados.
	TfidfVectorizer	Transformação de textos em matrizes de TF-IDF.
	TextBlob	Análise de polaridade de sentimentos em textos.
Pacotes e Bibliotecas	JobLib, Numpy, Gensim e sciPy	Geração de características para modelos de ML.
	SeaBorn e Matplotlib	Visualização de dados e criação de gráficos.
	Scikit-Learn e Keras	Treinamento de modelos de AM e AP.
Modelos Pré-Treinados	PyTest	Criação de testes automatizados.
	Wor2Vec GoogleNews	Incorporação semântica de palavras.
Algoritmos de AM	LR	<i>Logistic Regression</i>
	NB	<i>Naïve Bayes</i>
	SVM	<i>Support Vector Machine</i>
	RF	<i>Random Forest</i>
	MLP	<i>Multilayer Perceptron</i>
Algoritmos de AP	CNN	<i>Convolutional Neural Network</i>
	LSTM	<i>Long Short-Term Memory</i>

Fonte: Autoria própria (2023).

transformação de textos, análise de sentimentos, manipulação de dados e geração de características, bem como para a visualização de dados e treinamento dos modelos de AM e AP. A Tabela 4 apresenta uma visão geral das bibliotecas e ferramentas utilizadas e seus propósitos específicos no projeto.

Para assegurar a confiabilidade e a robustez do código desenvolvido, especialmente considerando os módulos de pré-processamento, criação de características, salvamento de arquivos e geração de gráficos, bem como a funcionalidade adequada dos *scripts* durante sua execução no servidor, foram desenvolvidos testes automatizados utilizando o PyTest. A abordagem de testes permitiu validar a funcionalidade e precisão dos diferentes componentes do código, assegurando que todos os processos relacionados à manipulação de dados e à análise fossem executados corretamente, sem a necessidade de interromper e reiniciar a execução no servidor.

A estrutura dos *scripts* foi projetada para garantir a continuidade da execução, mesmo em casos de interrupção. Essa funcionalidade foi implementada de forma que, ao reiniciar a execução, os *scripts* automaticamente verificam a presença de arquivos já gerados, como dados pré-processados, arquivos de características e modelos salvos. Se esses arquivos estiverem presentes, o *script* retoma a execução a partir do ponto onde foi interrompido anteriormente, evitando a repetição desnecessária de etapas já concluídas. Esta abordagem não apenas economiza tempo e recursos computacionais, mas também assegura a integridade do fluxo de trabalho, mantendo a consistência dos processos e resultados, mesmo diante de eventuais interrupções.

A estratégia de divisão dos dados para os modelos de AM e AP foi definida para atender não apenas às necessidades específicas de cada tipo de modelo, mas também para considerar as limitações de tempo. Nos modelos de AM, o *dataset* Sentiment140 foi dividido em 60% para treinamento e 40% para teste, enquanto nos modelos de AP, a divisão foi de 40% para treino, 20% para validação e 40% para teste. Essa diferenciação foi crucial para alinhar a divisão dos dados com as particularidades inerentes a cada abordagem de modelagem.

Além das considerações técnicas, essa divisão de dados também foi orientada pela necessidade de garantir que o treinamento dos modelos fosse concluído em um tempo hábil. Dado o grande volume da base de dados, uma porcentagem menor de dados para treinamento foi considerada suficiente para alcançar um desempenho aceitável dos modelos, sem comprometer significativamente a qualidade dos resultados. Isso permitiu otimizar o uso dos recursos computacionais disponíveis, mantendo a eficiência no processo de treinamento.

Para assegurar a gestão efetiva do tempo de treinamento, foi estabelecido um limite de tempo, conhecido como *walltime*, de 300 horas para cada modelo. O *walltime* refere-se ao tempo máximo de execução alocado para um processo ou tarefa em sistemas computacionais. Definir um *walltime* é uma prática comum em ambientes de computação de alto desempenho, servindo como uma medida de controle para prevenir que processos individuais consumam recursos excessivos por períodos prolongados. Essa medida foi fundamental para assegurar que

cada modelo fosse treinado dentro de um período de tempo definido, promovendo um uso eficiente dos recursos computacionais e facilitando o planejamento e a execução dos experimentos dentro do cronograma estabelecido para o projeto.

Em termos de métricas de desempenho, foram utilizadas Matriz de Confusão, Acurácia, Precisão, *Recall*, F1-Score, Curva ROC e Curva *precision-recall*. Estas métricas foram escolhidas por fornecerem uma visão abrangente da performance dos modelos em diferentes aspectos, como a capacidade de previsão correta, o equilíbrio entre precisão e sensibilidade, e a capacidade de discriminação entre classes.

Após a conclusão do treinamento e avaliação dos modelos, o modelo mais eficaz foi aplicado ao conjunto de dados de Fórmula 1 para prever os sentimentos dos fãs. Com os dados de sentimentos obtidos, realizou-se uma análise exploratória detalhada, focando em cada corrida para entender o impacto emocional dos eventos sobre os fãs de Fórmula 1. Esta análise exploratória permitiu não apenas validar o modelo em um cenário real, mas também oferecer uma visão sobre a dinâmica emocional no contexto das corridas de F1.

Para facilitar o acesso e promover a replicabilidade dos procedimentos e análises realizadas neste estudo, um repositório público no GitHub foi estabelecido, contendo todos os *scripts*, módulos, *notebooks* Jupyter e testes automatizados desenvolvidos. A Figura 10 apresenta um código QR que direciona diretamente para o referido repositório. Adicionalmente, o repositório pode ser acessado através do seguinte link: <https://github.com/matheus-mileski/F1TT-Sentiment-Analysis>. Este repositório é um componente chave para assegurar a transparência e a integridade da pesquisa, permitindo que outros pesquisadores e interessados examinem, testem e expandam os métodos e as conclusões apresentadas neste trabalho.

Figura 10 – Código QR para o repositório deste trabalho.



Fonte: Autoria própria (2023).

5 RESULTADOS

Este capítulo apresenta a análise dos resultados alcançados.

5.1 Definição de n-gramas para TF-IDF

No contexto do experimento realizado para definir a configuração ideal do *Term Frequency-Inverse Document Frequency* (TF-IDF), é importante compreender como diferentes combinações de n-gramas podem influenciar a extração de características dos textos. N-gramas são conjuntos contíguos de itens (neste caso, palavras) extraídos de um texto, variando em tamanho desde unigramas (palavras individuais) até bigramas e trigramas, que são conjuntos de duas e três palavras consecutivas, respectivamente.

Por exemplo, considere a frase "Análise de Sentimentos no Twitter". Em um contexto de unigrama, as características extraídas seriam as palavras individuais: ["Análise", "de", "Sentimentos", "no", "Twitter"]. Os bigramas seriam os pares de palavras adjacentes como: ["Análise de", "de Sentimentos", "Sentimentos no", "no Twitter"]. Já os trigramas seriam conjuntos de três palavras consecutivas, tais como ["Análise de Sentimentos", "de Sentimentos no", "Sentimentos no Twitter"].

Cada tipo de n-grama tem suas próprias vantagens na análise de texto. Unigramas são eficientes na identificação da frequência de palavras individuais, embora possam perder o contexto e a relação entre elas. Bigramas capturam melhor essas relações e contextos, adicionando uma camada de compreensão sobre como as palavras são usadas juntas. Trigramas vão além, oferecendo uma visão ainda mais contextualizada das combinações de palavras, mas também aumentam a dimensionalidade do espaço de características, o que pode ser um desafio computacional.

A combinação de unigramas, bigramas e trigramas pode oferecer um equilíbrio entre capturar a frequência de palavras, suas relações binárias e contextos mais amplos. Este equilíbrio é essencial para uma representação abrangente do texto, capturando uma riqueza de informações que pode ser particularmente útil em tarefas como a análise de sentimentos. O experimento com diferentes configurações de n-gramas no TF-IDF teve como objetivo determinar qual combinação forneceria a representação mais eficaz para os propósitos deste projeto, levando em consideração tanto a captura de informações contextuais quanto a eficiência na manipulação da dimensionalidade dos dados.

A saída do TF-IDF é uma representação matricial onde cada linha corresponde a um documento (neste caso, um *tweet*) e cada coluna representa um termo específico dentro do conjunto de n-gramas selecionado. Essa matriz é construída com base na frequência dos termos nos documentos, ponderada pela inversão da frequência desses termos em todo o corpus. O resultado é uma matriz esparsa, onde a maioria dos elementos é zero, especialmente em conjuntos de dados grandes com um vasto vocabulário.

Cada elemento na matriz TF-IDF é um valor numérico que reflete a importância de um termo específico no contexto de um documento particular. Se um termo aparece frequentemente em um documento, mas raramente em outros documentos do corpus, ele recebe um valor TF-IDF alto, indicando sua relevância única para aquele documento. Por outro lado, termos que aparecem frequentemente em muitos documentos recebem um valor mais baixo, refletindo sua menor importância distintiva.

Por exemplo, em uma análise de *tweets*, palavras comuns como ‘o’ e ‘a’ (em português) ou ‘the’ e ‘is’ (em inglês) terão valores TF-IDF baixos, pois são comuns em muitos documentos. Em contraste, palavras ou expressões específicas que são distintivas de certos *tweets* terão valores TF-IDF mais altos. Esta abordagem permite que o TF-IDF destaque as palavras que são críticas para entender o conteúdo específico de um documento, tornando-o uma ferramenta valiosa para tarefas como classificação de texto, análise de sentimentos e modelagem de tópicos.

A matriz TF-IDF resultante, portanto, serve como um conjunto de características ricas e informativas que podem ser utilizadas por algoritmos de Aprendizado de Máquina para realizar classificações ou outras formas de análise de texto. Ao transformar o texto em uma representação numérica ponderada, o TF-IDF facilita a aplicação de técnicas de aprendizado que de outra forma não seriam aplicáveis diretamente aos dados de texto brutos.

Diversos estudos revisados no capítulo de trabalhos relacionados ao projeto demonstraram o uso de diferentes combinações de n-gramas no TF-IDF, evidenciando a necessidade de um experimento dedicado para essa definição no contexto deste trabalho.

Para este experimento, optou-se por utilizar 50% dos *tweets* do *dataset* Sentiment140 (800.000 *tweets*), selecionados de forma balanceada entre as classes. Essa escolha foi feita para acelerar a execução do experimento, considerando que seis conjuntos de dados distintos de TF-IDF foram criados a partir do Sentiment140. Esses conjuntos incluíam um *dataset* para unigramas, um para bigramas, um para trigramas, e outros para combinações de 1-2 gramas, 2-3 gramas e outro para a combinação de 1, 2 e 3 gramas. Cada um desses *datasets* foi posteriormente avaliado com quatro modelos de AM: NB, LR, SVM e RF, totalizando 24 execuções.

Os processos foram configurados e executados em um servidor dedicado, levando aproximadamente quatro semanas para serem concluídos. Esta extensão no tempo refletiu a magnitude e a complexidade dos dados e dos processos envolvidos. Os resultados obtidos e sumarizados na Tabela 5 relevante demonstraram um padrão notável. A configuração que combinava unigramas, bigramas e trigramas (1-3 gramas) mostrou-se superior às configurações com menor complexidade de n-gramas. Esta abordagem mais abrangente resultou em um aprimoramento nas principais métricas de desempenho, como Acurácia, Precisão, *Recall* e *F1-Score*.

Os resultados do experimento revelaram que a combinação de unigramas, bigramas e trigramas (1-3 gramas) destacou-se consistentemente, superando as configurações com menor complexidade de n-gramas, evidenciando um aprimoramento nas métricas-chave de desempenho.

Tabela 5 – Resultados do experimento com o TF-IDF

Dataset	Modelo	Acurácia	Precisão	Recall	F1-Score
1-3 gramas	BernoulliNB	0,79	0,79	0,79	0,79
	LinearSVC	0,80	0,80	0,80	0,80
	LogisticRegression	0,80	0,80	0,80	0,80
	RandomForestClassifier	0,78	0,78	0,78	0,78
1-2 gramas	BernoulliNB	0,79	0,79	0,79	0,79
	LinearSVC	0,80	0,80	0,80	0,80
	LogisticRegression	0,78	0,78	0,78	0,78
	RandomForestClassifier	0,78	0,78	0,78	0,78
2-3 gramas	BernoulliNB	0,74	0,74	0,74	0,74
	LinearSVC	0,74	0,74	0,74	0,74
	LogisticRegression	0,70	0,71	0,70	0,70
	RandomForestClassifier	0,70	0,71	0,70	0,70
Unigrama	BernoulliNB	0,78	0,78	0,78	0,78
	LinearSVC	0,78	0,78	0,78	0,78
	LogisticRegression	0,78	0,78	0,78	0,78
	RandomForestClassifier	0,78	0,78	0,78	0,78
Bigrama	BernoulliNB	0,74	0,74	0,74	0,74
	LinearSVC	0,74	0,74	0,74	0,74
	LogisticRegression	0,70	0,71	0,70	0,70
	RandomForestClassifier	0,70	0,71	0,70	0,70
Trigrama	BernoulliNB	0,62	0,66	0,62	0,60
	LinearSVC	0,62	0,66	0,62	0,60
	LogisticRegression	0,58	0,62	0,58	0,54
	RandomForestClassifier	0,58	0,62	0,58	0,54

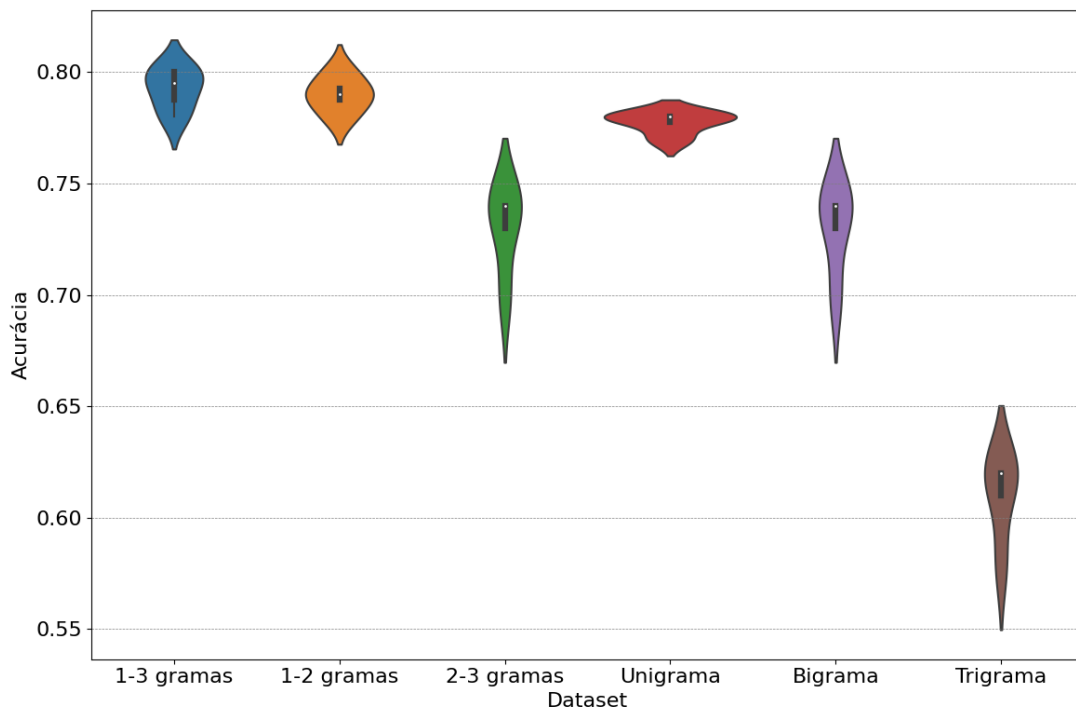
Fonte: Autoria própria (2023).

A superioridade da combinação 1-3 gramas pode ser atribuída à sua capacidade de capturar uma gama mais ampla de informações textuais. Unigramas fornecem uma base sólida ao capturar as palavras mais frequentes e relevantes, enquanto bigramas e trigramas adicionam uma camada contextual ao identificar combinações comuns de palavras e frases que são frequentemente usadas juntas, refletindo nuances específicas e padrões de linguagem. Isso se mostra particularmente útil em tarefas como a análise de sentimentos, onde a compreensão do contexto e a identificação de expressões específicas são fundamentais.

Por outro lado, a combinação de 1-2 gramas também mostrou resultados próximos em termos de eficácia, sugerindo que a inclusão de bigramas juntamente com unigramas já oferece um equilíbrio eficiente entre precisão contextual e complexidade computacional. No entanto, os resultados inferiores das configurações que utilizam apenas bigramas, trigramas ou a combinação de 2-3 gramas indicam que a exclusão dos unigramas resulta na perda de informações cruciais presentes nas palavras individuais. Isso pode levar a uma menor capacidade de capturar a essência dos sentimentos expressos nos *tweets*, especialmente quando termos-chave são considerados isoladamente.

Interessantemente, os trigramas, quando usados isoladamente, apresentaram a menor eficácia, ressaltando que a utilidade de termos de maior comprimento está na sua capacidade de complementar, e não substituir, os unigramas e bigramas. Isso sugere que a sinergia entre as diferentes escalas de n-gramas é um aspecto crucial para capturar com eficiência as nuances dos sentimentos expressos nos *tweets*. A combinação 1-3 gramas, portanto, oferece uma visão mais completa e rica do conteúdo textual, crucial para a análise de sentimentos em dados de mídia social.

Figura 11 – Gráfico de violino da acurácia dos modelos em relação aos *datasets* de n-gramas.



Fonte: Autoria própria (2023).

Os gráficos de violino são instrumentos eficazes para visualizar a distribuição e a probabilidade de variáveis contínuas, e a Figura 27 utiliza esta representação para complementar a análise da performance dos diferentes conjuntos de n-gramas em termos de acurácia. No gráfico de violino, cada conjunto de dados de n-gramas é representado por um “violino” individual, que combina um diagrama de caixa e uma densidade de núcleo para mostrar a distribuição dos resultados de acurácia.

Cada violino no gráfico é construído unindo os resultados de acurácia de todos os quatro modelos testados para o respectivo conjunto de n-gramas. Esta metodologia é adotada devido à realização de uma única execução do experimento, o que significa que, em vez de representar múltiplas execuções de um único modelo, cada violino agrupa os resultados de modelos distintos. Ainda assim, a superioridade do conjunto 1-3 gramas é evidente, pois seu violino apresenta uma densidade mais alta de pontos na extremidade superior da escala de acurácia, indicando que essa configuração de n-gramas é mais eficaz e confiável para a tarefa de classificação de sentimentos. A análise dos gráficos de violino, portanto, reforça a conclusão de que a combina-

ção mais complexa e integrada de n-gramas é vantajosa para obter uma análise mais precisa e detalhada dos sentimentos expressos nos textos.

5.2 Resultado do treinamento dos modelos de AM e AP

No escopo deste trabalho, um conjunto diversificado de modelos foi treinado para a classificação de sentimentos. Entre os modelos de Aprendizado de Máquina (AM) selecionados estavam o *Naïve Bayes* (NB), *Support Vector Machine* (SVM), *Logistic Regression* (LR), *Random Forest* (RF) e *Multilayer Perceptron* (MLP). Adicionalmente, modelos de Aprendizado Profundo (AP), incluindo redes neurais do tipo *Convolutional Neural Network* (CNN) e *Long Short-Term Memory* (LSTM), foram também contemplados.

Durante os experimentos iniciais, foi observado que o tempo de execução para alguns algoritmos estendia-se por várias semanas. Para contornar esse desafio operacional, instituiu-se um limite de tempo, conhecido como *walltime*, fixado em 300 horas para cada processo. Esta medida resultou na exclusão dos algoritmos NB e SVM dos experimentos subsequentes, pois ultrapassaram o *walltime* estabelecido, indicando uma demanda computacional incompatível com as restrições temporais do projeto.

A seleção de características para os modelos de AM foi informada pelos resultados do experimento de TF-IDF, que destacou a superioridade da combinação de unigramas, bigramas e trigramas. Além disso, foram incorporadas representações de word2vec do modelo GoogleNews pré-treinado, bem como análises de polaridade de sentimentos obtidas pelo TextBlob. Esta combinação de características foi projetada para capturar tanto a estrutura semântica quanto a emocional dos dados textuais.

Para os modelos de AP, a preparação dos dados envolveu técnicas especializadas de processamento de texto, tais como tokenização e o método `fit_to_text` do Keras, ambos necessários para adaptar os dados textuais às exigências dos modelos de redes neurais. O word2vec e a análise de polaridade do TextBlob também foram integrados, fornecendo uma camada adicional de informação semântica e emocional para os modelos de AP.

Quanto à divisão dos dados, o *dataset* Sentiment140 foi alocado de maneira diferenciada entre os modelos de AM e AP. Nos modelos de AM, adotou-se uma divisão de 60% dos dados para treinamento e 40% para teste, enquanto para os modelos de AP, o *dataset* foi dividido em 60% para treinamento, com 20% desse montante destinado à validação das épocas durante o treinamento, e os 40% restantes alocados para a fase de testes. Esta estratégia de divisão foi projetada não apenas para acomodar as particularidades de cada tipo de modelo, mas também para agilizar o processo de execução, uma consideração prática dada a escala dos experimentos empreendidos.

Os resultados gerais do treinamento dos modelos, apresentados na Tabela 6, revelaram diferenças notáveis em desempenho entre os algoritmos testados. O LR e a rede LSTM se sobressaíram, apresentando métricas de acurácia, precisão, recall e F1-Score superiores, além

Tabela 6 – Resultados do treinamento dos modelos de AM e AP

Modelo	Acurácia	Precisão	Recall	F1-Score	AUC
Logistic Regression	78,21%	77,19%	80,13%	78,63%	0,86
Random Forest	72,9%	73,5%	71,8%	72,6%	0,81
MLP	73,17%	73,59%	72,35%	72,97%	0,8
CNN	72,29%	71,77%	73,57%	72,66%	0,77
LSTM	78,08%	77,95%	78,4%	78,17%	0,86

Fonte: Autoria própria (2023).

de uma AUC de 0,86, indicativas de um desempenho equilibrado e confiável na classificação de sentimentos.

O LR destacou-se por sua eficiência e rapidez no treinamento, com um tempo de treinamento de 6,86 minutos, e uma acurácia de 78,21%. Este modelo, conhecido por sua eficácia em lidar com variáveis binárias e sua capacidade de fornecer resultados rápidos e robustos, provou ser uma opção viável para a classificação de sentimentos, indicando que as classes de sentimentos positivos e negativos podem ser linearmente separados. O modelo LSTM, apesar de um tempo de treinamento significativamente mais longo, cerca de 433,53 minutos, exibiu um desempenho comparável à LR, com uma acurácia ligeiramente menor de 78,08%, o LSTM se mostrou mais equilibrado em termos de precisão e sensibilidade (*recall*). Em trabalhos futuros, é possível realizar mais execuções desses modelos para uma análise estatística de desempenho e, assim, definir qual o modelo mais eficiente em termos de performance.

Os demais modelos, como RF e MLP, bem como a rede neural CNN, apresentaram desempenhos inferiores em comparação ao LR e ao LSTM. As características intrínsecas destes modelos, que se referem aos seus princípios fundamentais de funcionamento e estrutura, podem explicar esses resultados.

O RF, por exemplo, baseia-se em múltiplas árvores de decisão para realizar classificações, o que pode ser eficaz para dados com atributos claros e distintos, mas pode não ser tão adequado para a complexidade e sutileza dos dados de texto, onde a contextualização e a sequencialidade são cruciais. O MLP, uma forma básica de rede neural, enfrentar desafios ao lidar com dados sequenciais, sendo mais eficiente em padrões estacionários ou dados onde a ordem das entradas não é significativa.

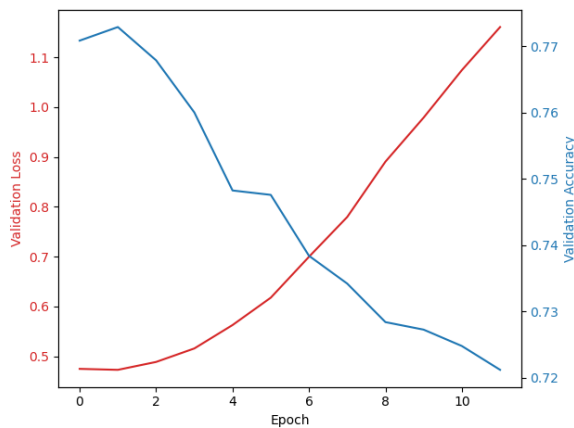
A CNN, que é amplamente reconhecida por sua eficiência em tarefas de processamento de imagem devido à sua habilidade de capturar padrões locais e características espaciais, pode não ter a mesma eficácia em dados sequenciais de texto. Isto pode ser parcialmente devido à arquitetura das camadas da rede, que pode não ter sido otimizada para capturar a estrutura e a semântica do texto. Em imagens, as camadas convolucionais e de *pooling* são excelentes para reduzir a dimensionalidade enquanto preservam características importantes, mas no texto, a im-

portância não está apenas na frequência localizada de palavras, mas também na sua interação ao longo de uma sequência.

Essas limitações indicam que uma revisão das arquiteturas de rede, especialmente para a CNN, pode ser necessária em trabalhos futuros para melhorar a capacidade desses modelos em tarefas de processamento de linguagem natural. A adaptação das camadas e a escolha de hiperparâmetros podem ser pontos de investigação para aprimorar a performance desses algoritmos em contextos onde a sequencialidade e o contexto são de extrema importância.

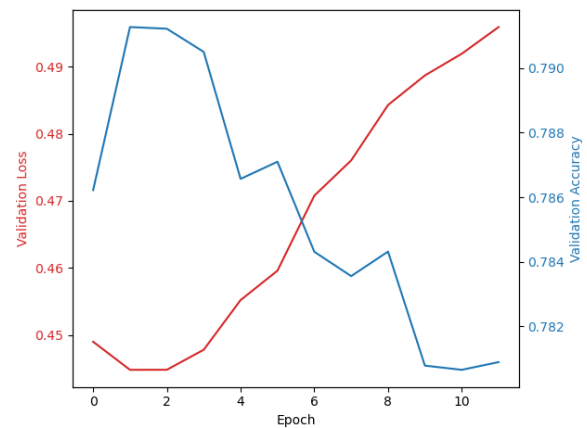
Apesar da LR oferecer uma execução mais rápida, a escolha do modelo LSTM para a análise de sentimentos foi motivada por sua maior capacidade de capturar dependências de longo prazo no texto, o que é particularmente relevante para entender a sequência e o contexto nos dados. Esta característica torna o LSTM um modelo mais equilibrado, apesar do maior custo computacional e do tempo de treinamento alongado. O equilíbrio entre a precisão na classificação de sentimentos e a capacidade de entender estruturas complexas de texto justifica a preferência pelo LSTM, apesar da eficiência do LR.

Figura 12 – Histórico de treinamento da CNN.



Fonte: Autoria própria (2023).

Figura 13 – Histórico de treinamento do LSTM.



Fonte: Autoria própria (2023).

Um comparativo do histórico de treinamento dos modelos de rede neural, CNN e LSTM, revela que ambos foram interrompidos pelo critério de *early stopping* após 12 épocas de treinamento, indicando um rápido alcance de um ponto de estagnação no aprendizado.

O critério de *early stopping* para ambos os modelos foi definido de forma que o treinamento fosse interrompido se a perda de validação não melhorasse após 10 épocas consecutivas. Este mecanismo é essencial para prevenir o treinamento excessivo e garantir que os modelos mantenham uma capacidade adequada de generalização.

Para a CNN, Figura 12, observou-se uma queda na acurácia de validação de 0,77 para 0,72, enquanto a perda de validação aumentou significativamente de 0,47 para 1,16. Este comportamento sugere um claro sobreajuste do modelo: embora a CNN tenha aprendido eficazmente os padrões nos dados de treinamento, sua capacidade de generalizar para os dados de

validação foi comprometida, resultando em um desempenho inferior quando confrontada com novos dados.

No caso do LSTM, a variação na acurácia de validação foi menos pronunciada, passando de 0,786 para 0,791 e, em seguida, caindo ligeiramente para 0,780. No entanto, a perda de validação subiu de 0,44 para 0,5, indicando também uma tendência ao sobreajuste, embora em menor grau em comparação com a CNN. Esta elevação na perda de validação, apesar de mais moderada, revela que o modelo LSTM também começou a se especializar demais nos dados de treinamento, perdendo a eficácia na generalização.

Na sequência, este trabalho apresentará uma comparação detalhada entre os modelos de LR e LSTM, destacando-se pela sua destacada performance entre os algoritmos testados. Esta análise minuciosa visa elucidar as nuances que contribuem para a eficácia desses modelos, focando em particularidades como tempo de treinamento, capacidade de generalização e desempenho nas diversas métricas de avaliação.

Para aqueles interessados em examinar o desempenho dos demais modelos, que não são discutidos extensivamente no corpo principal do texto, os gráficos relevantes, como matrizes de confusão, curvas ROC e curvas *precision-recall*, estão disponibilizados nos apêndices. Estes materiais complementares permitem uma inspeção mais profunda da performance de cada modelo e servem como uma referência valiosa para uma avaliação comparativa mais ampla, enriquecendo assim a discussão e as conclusões derivadas deste estudo.

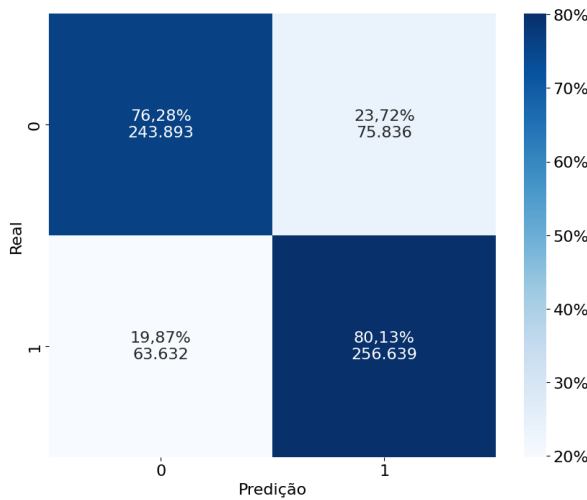
5.2.1 Comparação dos resultados do LR e do LSTM

Na comparação entre o modelo de LR e LSTM, observamos que ambos os modelos alcançaram uma performance notável na tarefa de classificação de sentimentos, com acurácias próximas de 78%. O LR, sendo o mais simples dos dois, surpreendeu positivamente, com métricas de precisão e *recall* que denotam uma tendência a minimizar os falsos negativos, resultando em um *F1-Score* que reflete uma combinação eficaz de precisão e sensibilidade na identificação de sentimentos positivos.

As matrizes de confusão para ambos os modelos, apresentadas nas Figuras 14 e 15, indicam que o LR foi ligeiramente mais eficaz na previsão de verdadeiros positivos do que o LSTM, enquanto o LSTM demonstrou uma distribuição mais equilibrada entre verdadeiros positivos e verdadeiros negativos. Esta diferença sutil revela a capacidade do LSTM de ser mais abrangente em sua classificação, refletindo sua eficácia na captura de relações de longo alcance no texto, o que é essencial para entender a dinâmica dos sentimentos expressos.

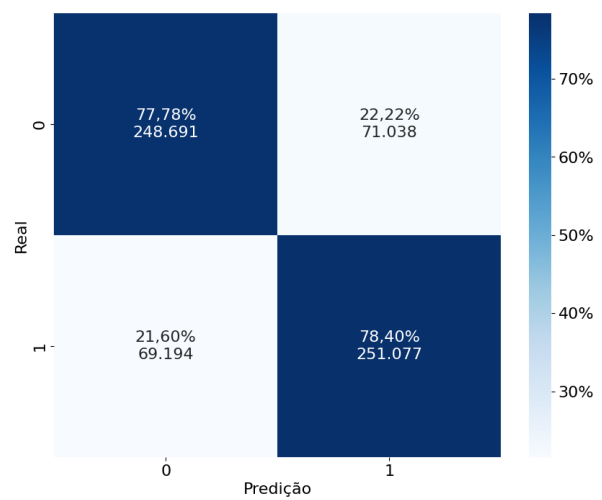
As curvas ROC para ambos os modelos, Figuras 16 e 17, mostraram uma AUC de 0,86, evidenciando uma excelente capacidade discriminatória para separar as classes de sentimentos positivos e negativos. No entanto, é nas curvas *Precision-Recall* que discernimos mais claramente a competência dos modelos. Como mostrado nas Figuras 18 e 19, ambos os modelos apresentaram uma grande área sob a curva *Precision-Recall*, a curva com a concavidade para

Figura 14 – Matriz de confusão do treinamento do LR.



Fonte: Autoria própria (2023).

Figura 15 – Matriz de confusão do treinamento do LSTM.



Fonte: Autoria própria (2023).

baixo é um indicativo de que os modelos conseguem manter níveis elevados de precisão à medida que maximizam a *recall*, um equilíbrio desejável em muitas aplicações práticas.

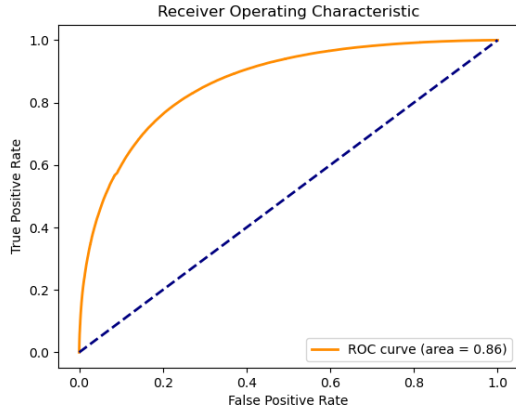
Por fim, ambos os modelos, LR e LSTM, foram aplicados ao *dataset* de *tweets* sobre a F1. Essa aplicação prática tinha como objetivo não apenas identificar os sentimentos expressos, mas também comparar as previsões dos dois modelos.

Essa análise comparativa nos permitiu compreender melhor como cada modelo lidou com as nuances linguísticas e as variações expressivas encontradas no discurso dos usuários sobre eventos da F1. A escolha de aplicar ambos os modelos ao mesmo conjunto de dados reflete um esforço para garantir uma avaliação justa e rigorosa, fornecendo uma base sólida para a seleção do modelo mais adequado para a análise de sentimentos em redes sociais.

A Figura 20 apresenta um comparativo gráfico das previsões dos dois modelos. A região central apresenta as inferências onde ambos os modelos concordaram, sendo os sentimentos negativos na cor preta e os sentimentos positivos na cor cinza. A extremidade esquerda do gráfico apresenta os resultados onde o LSTM identificou sentimentos negativos e o LR sentimentos positivos, o oposto ocorre na extremidade direita do gráfico.

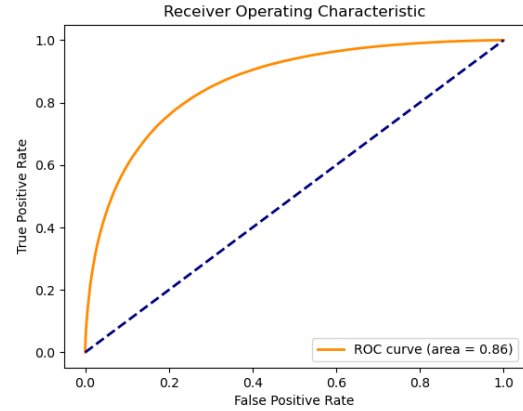
Em 233.582 instâncias, o modelo LR classificou os sentimentos como positivos, enquanto o LSTM os interpretou como negativos. Inversamente, em 141.007 instâncias, o LSTM detectou um sentimento positivo onde o LR havia previsto um negativo. Essas discrepâncias sublinham as diferenças nos mecanismos de aprendizado e na sensibilidade dos modelos às nuances da linguagem utilizada nos *tweets* sobre a F1.

Figura 16 – Curva ROC para o treinamento do LR.



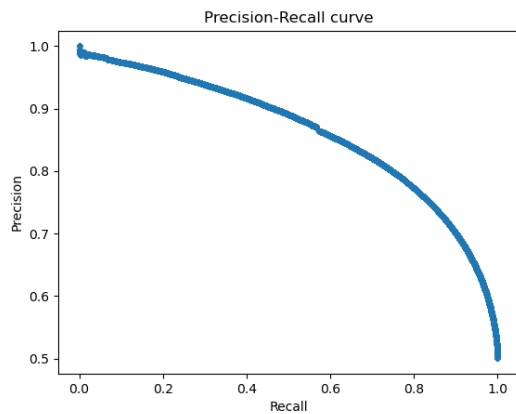
Fonte: Autoria própria (2023).

Figura 17 – Curva ROC para o treinamento do LSTM.



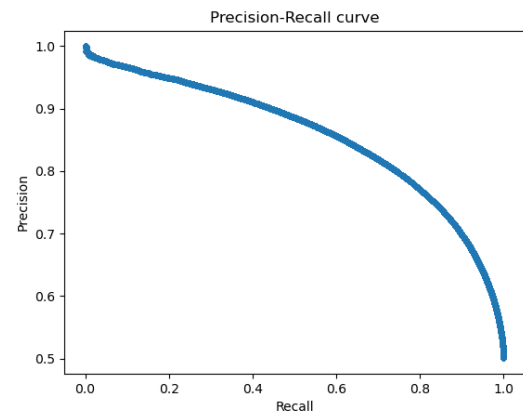
Fonte: Autoria própria (2023).

Figura 18 – Curva Precision-Recall para o treinamento do LR.



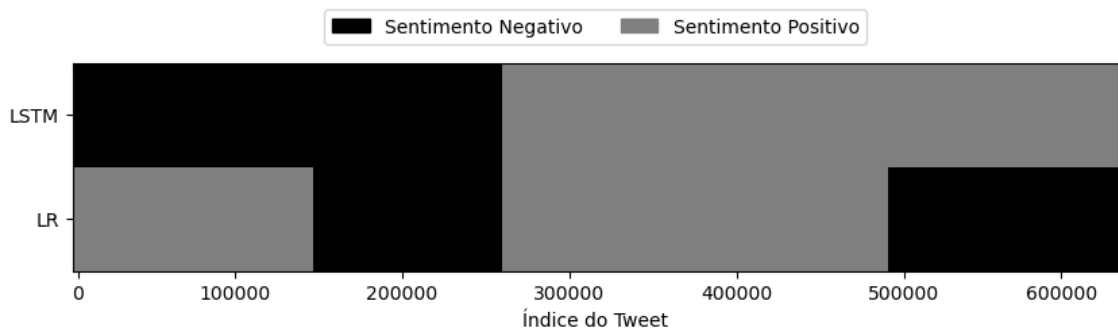
Fonte: Autoria própria (2023).

Figura 19 – Curva Precision-Recall para o treinamento do LSTM.



Fonte: Autoria própria (2023).

Figura 20 – Predições dos modelos de LR e LSTM.



Fonte: Autoria própria (2023).

Tabela 7 – Comparação entre as previsões de sentimentos do LR e do LSTM.

ID	Texto do Tweet	Texto Processado	Predição do LR	Predição do LSTM
1	Qualifying is going to be exciting! #F1 #USGP	qualifying going exciting star struck hashtag f1 hashtag usgp	Positivo	Positivo
2	How can you not like F1!? Brilliant racing. #BrazilGP #F1 #SkyF1	like f1 brilliant racing hashtag brazilgp hashtag f1 hashtag skyf1	Positivo	Positivo
3	To sum up, going unnecessarily slow under safety car (no penalty), catching up to the car in front during a VSC (no penalty), driving into the back of another car (no penalty)... Max gives position back and still receives a 5 second penalty... Seems a bit odd. #SaudiArabiaGP #f1	sum going unnecessarily slow safety car penalty catching car in front vsc penalty driving back another car penalty max give position back still receives 5 second penalty seems bit odd hashtag saudiarabiagp hashtag f1	Negativo	Negativo
4	Every single F1 race the directing is appalling. How did they not have Ocon/Bottas on-screen? #F1	every single f1 race directing appalling ocon bottas screen hashtag f1	Negativo	Negativo
5	What a time to be a #F1 fan!	time hashtag f1 fan	Positivo	Negativo
6	Never have I seen where only ONE car took a race start!! One for the record books? @SkySportsF1 @tedkravitz #F1 @f1 #HungarianGP https://t.co/tiObTmDkj5	never seen one car took race start one record book mention skysportsf1 mention tedkravitz hashtag f1 mention f1 hashtag hungariangp co tiobtmkj5	Positivo	Negativo
7	Can we race in Saudi Arabia every week? This is brilliant. #f1	race saudi arabia every week brilliant hashtag f1	Negativo	Positivo
8	What an epic season and a bizarre ending. Wooooow. Max!!!! Yes!!!! #f1	epic season bizarre ending wooooow max yes hashtag f1	Negativo	Positivo

A Tabela 7 fornece uma comparação direta entre as predições de sentimentos realizadas pelos modelos de LR e LSTM. Esta análise comparativa é auxiliada no entendimento de como cada modelo interpreta e classifica os sentimentos expressos nos *tweets*.

A avaliação dos textos dos *tweets* e seus correspondentes textos processados revela que tanto o LR quanto o LSTM são consistentes em classificar sentimentos positivos em declarações claramente entusiasmadas sobre as qualificações e corridas da F1. Esta consistência é observada em expressões que contêm palavras-chave como "*exciting*" (*tweet 1*) e "*brilliant racing*" (*tweet 2*), indicando uma identificação eficaz dos sentimentos positivos por ambos os modelos.

No entanto, a análise revela também algumas divergências interessantes nas predições de sentimentos entre os dois modelos. Por exemplo, o LR classificou um *tweet* que expressa uma opinião positiva sobre ser fã de F1 como positivo, enquanto o LSTM o interpretou como negativo (*tweet 5*). Isso pode sugerir que o LSTM pode estar ponderando outros aspectos contextuais ou sequenciais que não são capturados pelo modelo de LR. Da mesma forma, em um *tweet* que expressa surpresa com a condução de uma corrida, o LR previu um sentimento positivo, ao passo que o LSTM o interpretou como negativo (*tweet 6*). Esse tipo de divergência aponta para uma sensibilidade contextual mais profunda do LSTM, que pode ser capaz de detectar subtextos ou ironias que o modelo de LR pode não captar.

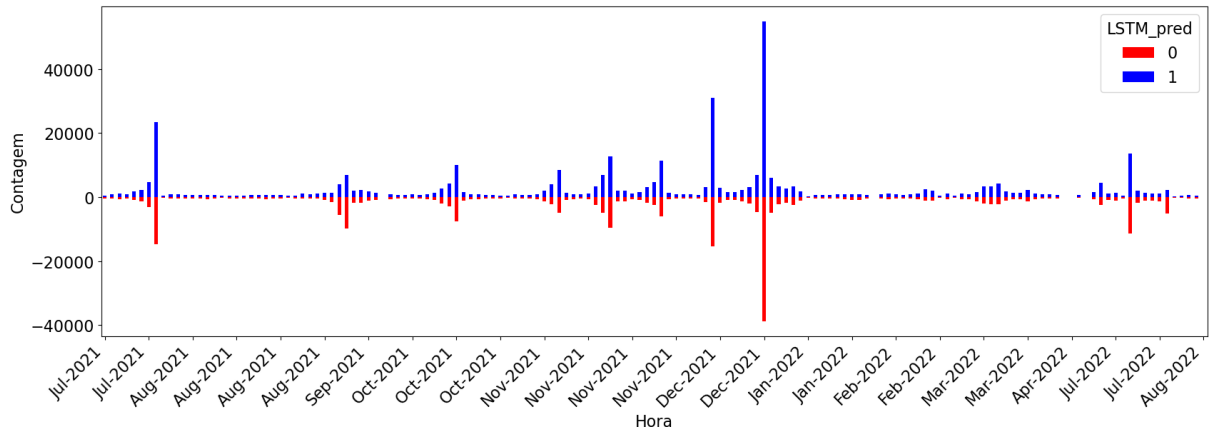
Outra observação notável é a capacidade de ambos os modelos em classificar corretamente os sentimentos negativos em *tweets* que criticam as decisões durante uma corrida ou a direção de uma transmissão. Isso demonstra uma boa interpretação de palavras e frases que contextualizam sentimentos negativos, como "*unnecessarily slow*" (*tweet 3*) e "*appalling*" (*tweet 4*).

No entanto, a precisão dessas predições não é infalível, como ilustrado pelas discrepâncias nas classificações de *tweets* com uma complexidade emocional mais sutil. Em *tweets* que, aparentemente, expressam entusiasmo, o modelo de LR previu um sentimento negativo, enquanto o LSTM o classificou como positivo (*tweets 7 e 8*). Essa diferença pode ser atribuída ao fato de que o LSTM consegue absorver melhor o contexto do texto.

Em termos de desempenho geral, a tabela sugere que, enquanto o LR pode ser mais rápido no treinamento, o LSTM oferece um entendimento mais variado da linguagem natural, justificando a escolha para a análise de sentimentos dos fãs de F1.

5.3 Análise dos sentimentos dos fãs de F1

O histograma apresentado na Figura 21 ilustra a distribuição temporal dos sentimentos dos *tweets*, conforme analisados pela predição do modelo LSTM. Os dados, totalizando 257.795 *tweets* negativos e 374.589 positivos (quatro *tweets* foram removidos por não possuírem a informação de data), foram meticulosamente ordenados e agrupados por dia, com a classe vermelha indicando sentimentos negativos e a azul, os positivos. Observa-se uma concentração

Figura 21 – Histograma de *tweets* positivos e negativos classificados pelo LSTM.

Fonte: Autoria própria (2023).

notável de atividade nas últimas etapas da temporada de Fórmula 1 de 2021, refletida por picos significativos no volume de *tweets*. Este aumento pode estar correlacionado com o clímax da temporada, marcado por eventos decisivos e emocionalmente carregados que naturalmente catalisam um engajamento mais intenso nas redes sociais. Em contraste, o início de 2022 é caracterizado por um volume mais baixo de *tweets*, o que pode ser atribuído ao período de entressafra na Fórmula 1, onde as corridas são menos frequentes e o interesse do público tende a diminuir. Este padrão cíclico no volume de *tweets* sugere uma relação direta entre a intensidade das competições e o engajamento emocional dos espectadores, destacando o valor da análise de sentimentos para compreender as dinâmicas do público em eventos esportivos globais. A subseção a seguir apresenta os resultados de uma análise detalhada de engajamento e expressões por corrida.

5.3.1 Análise dos resultados de sentimento por corrida

A análise da polaridade dos *tweets* relacionados aos Grandes Prêmios de Fórmula 1 revela informações significativas sobre a percepção pública e a reação emocional aos eventos. Esta seção concentra-se na polaridade positiva e negativa dos *tweets*, geradas pelo modelo LSTM, correlacionando-os com os resultados e características específicas das corridas. A Tabela 8 apresenta as corridas que estavam presentes no *dataset* de *tweets* de F1 utilizado neste trabalho. Os dados foram filtrados pela data de cada Grande Prêmio e assim, analisados individualmente.

O principal destaque é o Grande Prêmio de Abu Dhabi, realizado em 12 de dezembro de 2021, se destacou com o maior número de *tweets* tanto positivos (54.877) quanto negativos (38.857). Este elevado volume de *tweets* reflete a intensidade e o impacto emocional deste evento, que foi marcado por uma disputa acirrada pelo título entre Max Verstappen e Lewis Hamilton, culminando com a vitória controversa de Verstappen. A alta polaridade dos *tweets*

Tabela 8 – Corridas presentes no dataset de F1

Grande Prêmio	Data	Vencedor	Equipe	Voltas	Duração	Tweets	
						Neg.	Pos.
Hungary	01/08/21	E. Ocon	Alpine	70	2:04:43.199	14.733	23.447
Netherlands	05/09/21	M. Verstappen	Red Bull	72	1:30:05.395	9.741	6.827
United States	24/10/21	M. Verstappen	Red Bull	56	1:34:36.552	7.493	9.973
Mexico	07/11/21	M. Verstappen	Red Bull	71	1:38:39.086	4.858	8.461
Brazil	14/11/21	L. Hamilton	Mercedes	71	1:32:22.851	9.658	12.811
Qatar	21/11/21	L. Hamilton	Mercedes	57	1:24:28.471	6.121	11.439
Saudi Arabia	05/12/21	L. Hamilton	Mercedes	50	2:06:15.118	15.323	31.036
Abu Dhabi	12/12/21	M. Verstappen	Red Bull	58	1:30:17.345	38.857	54.877
Saudi Arabia	27/03/22	M. Verstappen	Red Bull	50	1:24:19.293	2.237	4.324
Great Britain	03/07/22	C. Sainz	Ferrari	52	2:17:50.311	11.347	13.577

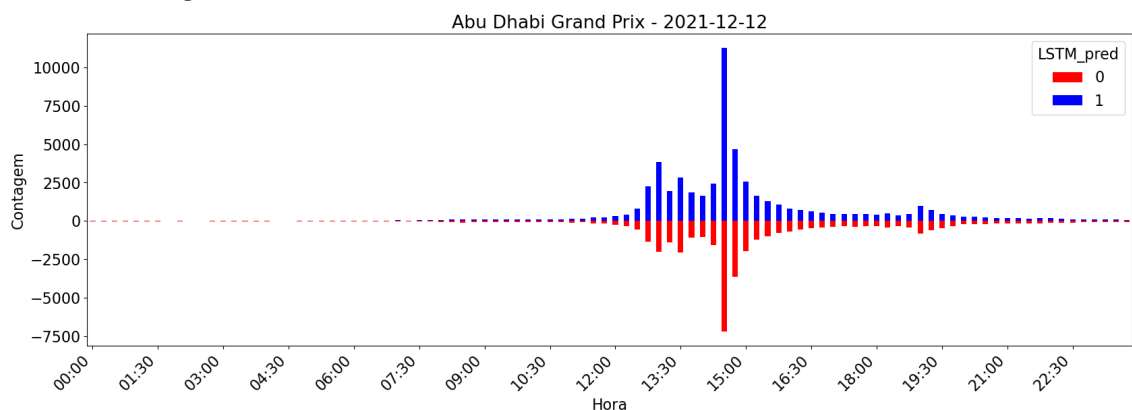
Fonte: Autoria própria (2023).

pode ser atribuída à natureza dramática e ao desfecho da corrida, que gerou debates acalorados entre os fãs e seguidores do esporte. A Figura 22 apresenta o fluxo de *tweets* durante o dia desta corrida. A classe 0 representa sentimentos negativos e a classe 1 os positivos.

Contrastando com Abu Dhabi, o Grande Prêmio do México, ocorrido em 07 de novembro de 2021, apresentou uma menor polaridade nos *tweets*, como mostrado na Figura 23, com 4.858 negativos e 8.461 positivos. A vitória de Verstappen neste evento foi menos controversa, possivelmente contribuindo para uma reação menos engajada e polarizada na comunidade online. Este padrão sugere que eventos com menos controvérsias tendem a gerar um menor volume de interações.

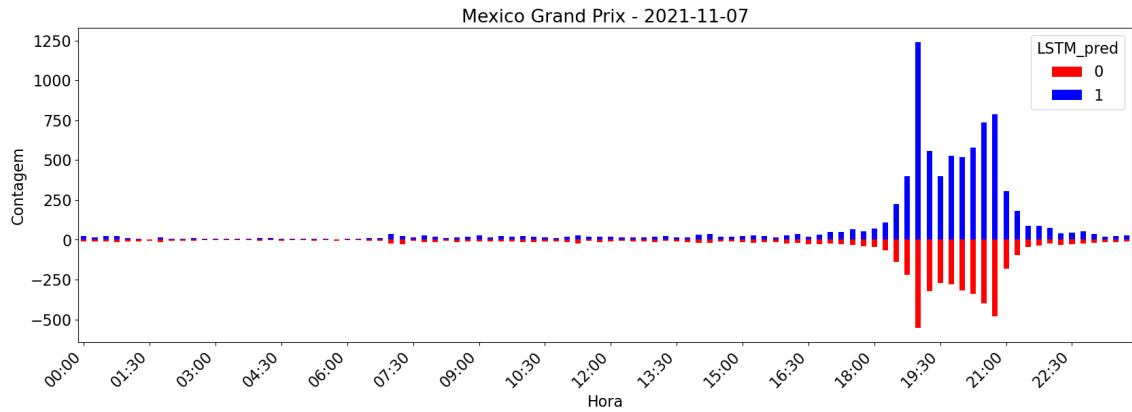
O Grande Prêmio da Arábia Saudita, apresentado na Figura 24, em sua primeira edição em 05 de dezembro de 2021, apresentou uma quantidade significativa de *tweets* positivos (31.036), superando o número de *tweets* negativos (15.323). Este fenômeno pode ser explicado pelo caráter inédito do evento, gerando expectativa e entusiasmo entre os fãs. A vitória de Ha-

Figura 22 – Quantidade de *tweets* durante o GP de Abu Dhabi 2021.



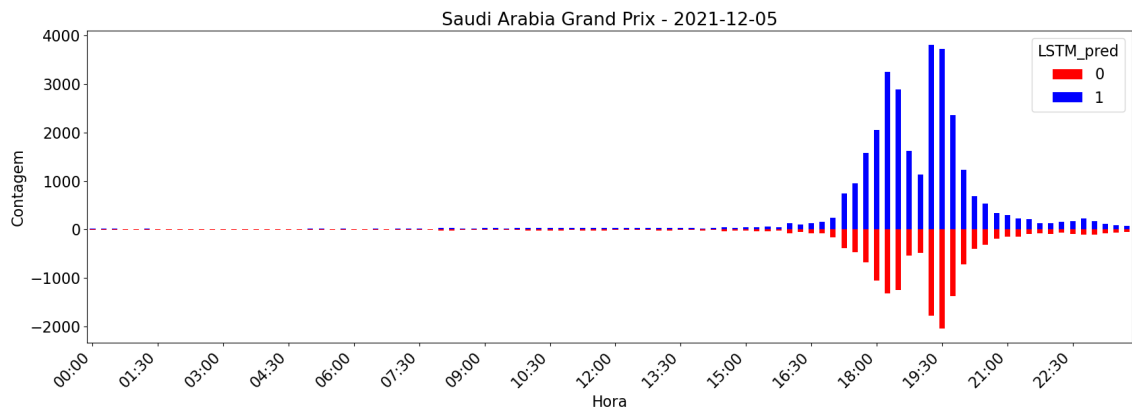
Fonte: Autoria própria (2023).

Figura 23 – Quantidade de *tweets* durante o GP do México 2021.



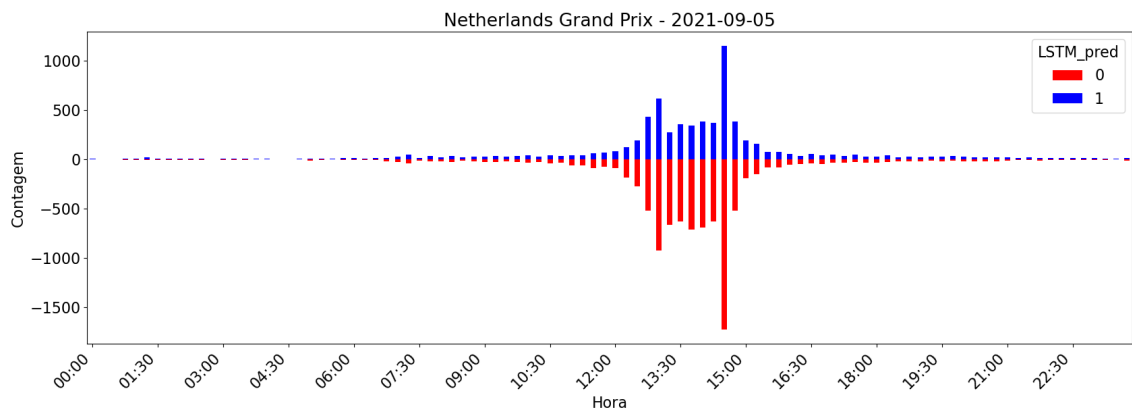
Fonte: Autoria própria (2023).

Figura 24 – Quantidade de *tweets* durante o GP da Arábia Saudita 2021.



Fonte: Autoria própria (2023).

Figura 25 – Quantidade de *tweets* durante o GP da Holanda 2021.



Fonte: Autoria própria (2023).

milton nesta corrida também pode ter contribuído para a predominância de reações positivas, dado o seu amplo suporte e popularidade no cenário da Fórmula 1.

O Grande Prêmio da Holanda de 2021, realizado em 5 de setembro, apresenta um caso singular em nosso dataset: foi a única corrida onde o número de tweets negativos (9.741) superou o de positivos (6.827), a Figura 25 mostra um grande número de tweets negativos durante

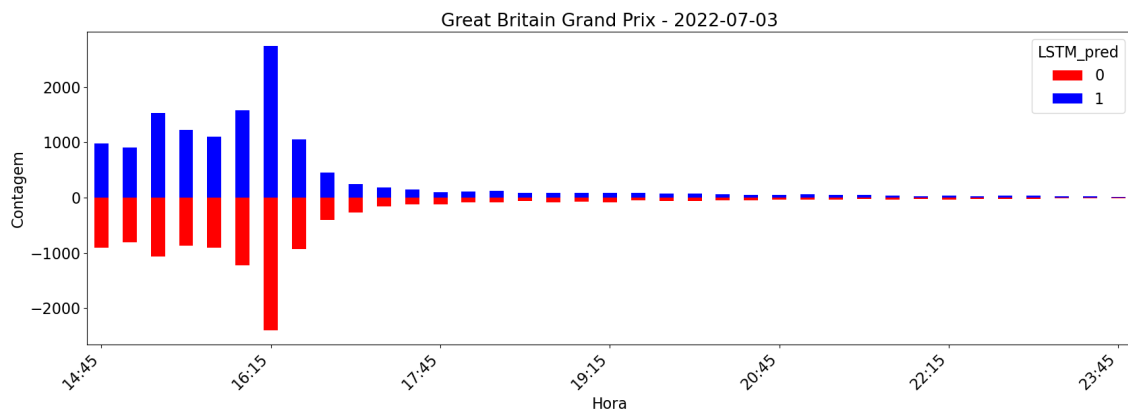
todo o evento, apresentado um pico de mais de 1.500 expressões negativas nos minutos finais da corrida. Esta peculiaridade merece uma análise cuidadosa, considerando o contexto e os eventos associados a esta corrida.

O GP da Holanda marcou o retorno da Fórmula 1 ao Circuito de Zandvoort após uma longa ausência, uma ocasião que, em teoria, poderia gerar um sentimento positivo predominante. No entanto, a vitória de Max Verstappen, embora celebrada por seus fãs, pode ter sido um fator para a polarização, especialmente considerando a intensa rivalidade entre Verstappen e Lewis Hamilton naquela temporada. Essa rivalidade frequentemente se estendia para as bases de fãs, potencialmente gerando um volume maior de reações negativas nas redes sociais. Além disso, aspectos logísticos, como desafios de acesso ao circuito e preocupações ambientais levantadas pela realização do evento, podem ter contribuído para o sentimento negativo. Portanto, enquanto a vitória de um piloto local geralmente gera uma reação positiva, no caso do GP da Holanda de 2021, a combinação de rivalidades esportivas intensas e questões externas ao esporte parece ter influenciado a comunidade online a expressar mais sentimentos negativos do que positivos.

Por outro lado, o Grande Prêmio da Grã-Bretanha, vencido por Carlos Sainz em 03 de julho de 2022, apresentou um equilíbrio mais próximo entre *tweets* positivos (13.577) e negativos (11.347). Este equilíbrio pode ser reflexo da diversidade de opiniões sobre o desempenho de diferentes pilotos e equipes, além da natureza imprevisível da corrida, que frequentemente resulta em reações mistas entre os espectadores. A Figura 26 apresenta o histograma de *tweets* durante o evento, nota-se um fluxo menor de dados, comparados com os outros eventos, isso pode estar ligado diretamente com a construção do dataset utilizado na análise.

Devido a isso, é importante ressaltar que a natureza dos dados presentes no dataset utilizado para esta análise, merece uma consideração especial. Sendo um dataset público, não havia informações disponíveis sobre a metodologia exata de coleta, o que implica em potenciais limitações na representatividade e imparcialidade dos dados. Sem uma métrica de coleta definida, os dados podem estar sujeitos a vieses introduzidos pelos autores do dataset ou por

Figura 26 – Quantidade de *tweets* durante o GP da Grã-Bretanha 2022.



Fonte: Autoria própria (2023).

características inerentes às plataformas de onde foram coletados. Tais vieses podem influenciar a distribuição das polaridades dos *tweets*, refletindo mais as preferências ou comportamentos dos usuários dessas plataformas do que uma visão equilibrada e abrangente das reações ao evento.

Apesar dessas potenciais limitações, para o propósito da análise de sentimentos, o dataset provou ser adequado, permitindo uma investigação significativa sobre as reações emocionais dos fãs de Fórmula 1. No entanto, é crucial abordar os resultados com uma compreensão de que eles podem não representar completamente todas as perspectivas e reações ao evento, e que devem ser interpretados com um certo grau de cautela.

6 CONCLUSÃO

Este trabalho de conclusão de curso investigou a dinâmica emocional da comunidade de fãs de automobilismo no Twitter. Empregando técnicas sofisticadas de Processamento de Linguagem Natural (PLN), o estudo buscou extrair e analisar sentimentos e padrões de engajamento dos usuários durante eventos significativos da temporada de Fórmula 1 de 2022.

6.1 Resultados Gerais

O uso de dois *datasets* públicos provou ser uma solução eficaz, permitindo a condução de análises exploratórias robustas. A preparação dos *datasets* foi realizada com rigor metodológico, adotando processos de limpeza e normalização de dados, essenciais para a precisão do modelo de PLN.

A seleção do melhor TF-IDF através de testes comparativos entre diferentes configurações e modelos foi um passo crítico que resultou na escolha de um arranjo com 1, 2 e 3 gramas para modelos de AM, demonstrando a relevância de uma abordagem combinada de diferentes níveis de análise textual. A criação de características para os *datasets* de aprendizado de máquina, assim como a substituição do TF-IDF pelo *tokenizer* e pelo *fit_on_texts* do Keras para os modelos de AP, foram etapas decisivas que influenciaram diretamente na performance dos modelos.

Os modelos treinados, tanto em AM quanto em AP, foram avaliados através de métricas rigorosas, com destaque para o modelo LSTM, que apresentou desempenho equivalente ao de LR, ambos superiores aos demais modelos.

A aplicação do modelo LSTM ao *dataset* da F1 permitiu uma análise exploratória detalhada dos sentimentos por corrida, revelando a relação entre eventos da F1 e a reação emocional dos fãs. Tal análise não apenas alcançou o objetivo geral de identificar padrões de engajamento e sentimentos, mas também forneceu *insights* sobre as motivações subjacentes às reações dos usuários.

Entre as vantagens do estudo, destaca-se a habilidade de captar e analisar grandes volumes de dados de sentimentos, proporcionando uma visão valiosa sobre a percepção pública de eventos esportivos.

6.2 Limitações e Dificuldades

A mudança nas políticas de acesso à API do Twitter representou uma adversidade substancial, restringindo a coleta de dados inicialmente prevista e forçando a busca por conjuntos de dados públicos. Esse revés inicial evidenciou a volatilidade e a imprevisibilidade associadas à dependência de fontes de dados externas, que são intrínsecas ao campo da ciência de dados.

Adicionalmente, o tempo mostrou-se um fator limitante significativo. Os experimentos e os processos de treinamento dos modelos exigiram aproximadamente dois meses para serem concluídos, um período substancial que destaca a magnitude da computação envolvida e a intensidade dos recursos necessários para tarefas de PLN em grandes conjuntos de dados. Essa extensão do tempo necessário para conduzir os experimentos sublinha a importância do planejamento rigoroso e da gestão de tempo eficiente em projetos de análise de dados de grande escala.

Além disso, a complexidade do PLN apresentou desafios técnicos, desde o pré-processamento e a normalização dos textos até a extração eficiente de características e a interpretação dos resultados dos modelos. Esses desafios ressaltam a necessidade de métodos avançados e ferramentas otimizadas para lidar com a sutileza e a riqueza da linguagem humana, assim como a importância da escolha criteriosa de modelos e algoritmos que se alinhem com os objetivos específicos da pesquisa.

Em resumo, as limitações e dificuldades encontradas no curso deste estudo refletem os desafios intrínsecos ao campo da ciência de dados, particularmente em projetos que envolvem a análise de sentimentos em mídias sociais, e destacam a necessidade de estratégias adaptativas e resilientes para superar obstáculos e alcançar resultados significativos.

6.3 Trabalhos Futuros

Esta pesquisa abre caminho para diversas direções promissoras em trabalhos futuros, visando expandir a compreensão e a precisão na análise de sentimentos, especialmente no contexto de mídias sociais relacionadas à *F1*. Uma das primeiras etapas seria a execução repetida dos experimentos atuais com diferentes *seeds*, possibilitando uma análise estatística mais robusta dos resultados. Essa abordagem ajudaria a assegurar a consistência dos modelos e a confiabilidade dos resultados, além de fornecer informações mais profundas sobre a variação do desempenho dos algoritmos sob diferentes condições iniciais.

Além disso, a aplicação de algoritmos de extração de entidades nomeadas poderia enriquecer significativamente a análise, correlacionando os sentimentos expressos com pilotos, equipes e eventos dentro do universo da *F1*. Isso permitiria uma análise mais detalhada das dinâmicas de sentimentos em relação a entidades específicas, oferecendo uma compreensão mais matizada das reações dos fãs.

Um estudo aprofundado envolvendo técnicas como Análise de Componentes Principais (PCA) e Alocação Latente de Dirichlet (LDA) poderia também ser conduzido para explorar as estruturas subjacentes nos dados de texto. Estas técnicas poderiam ajudar a identificar padrões e tendências ocultas nos dados, além de contribuir para a redução da dimensionalidade e melhoria da interpretabilidade dos conjuntos de dados.

A criação de nuvens de palavras segmentadas por sentimentos expressos pelos fãs ofereceria uma visualização intuitiva e impactante das palavras e frases mais comuns associa-

das a sentimentos específicos. Essa abordagem poderia revelar dados valiosos sobre o léxico utilizado pelos fãs e as tendências emocionais predominantes.

Para capturar a gama completa de emoções humanas, o treinamento de modelos capazes de identificar sentimentos mais complexos, como alegria, tristeza e raiva, seria um avanço significativo. Isso permitiria uma análise emocional mais rica e diferenciada, além de abrir possibilidades para aplicações práticas mais sofisticadas.

Outra direção seria testar modelos mais modernos e avançados encontrados na literatura, como a combinação de CNN com LSTM, que poderiam oferecer um equilíbrio entre a captura de padrões locais e a compreensão de sequências temporais. Além disso, a exploração de modelos de linguagem de última geração, como BERT (Bidirectional Encoder Representations from Transformers) e GPT (Generative Pre-trained Transformer), poderia proporcionar avanços significativos na identificação de sentimentos, dada a sua sofisticação e capacidade de entender nuances complexas na linguagem.

Em resumo, os trabalhos futuros nesta área têm o potencial de aprimorar significativamente a compreensão dos sentimentos dos fãs em mídias sociais, especialmente em contextos dinâmicos e emocionalmente carregados como os eventos de Fórmula 1. A adoção dessas abordagens e tecnologias avançadas poderia levar a descobertas significativas e aplicações práticas inovadoras no campo da análise de sentimentos.

6.4 Considerações Finais

Conclui-se que o trabalho atingiu seu objetivo de empregar técnicas de PLN para análise de sentimentos no Twitter durante a temporada de Fórmula 1 de 2022, identificando com êxito os padrões de engajamento e sentimentos dos usuários. Embora as dificuldades encontradas tenham imposto limitações, estas não comprometeram a integridade e a relevância dos resultados. As descobertas deste estudo contribuem para o entendimento do comportamento dos fãs de F1 nas redes sociais e oferecem uma base sólida para o desenvolvimento de estratégias de marketing e comunicação mais eficazes no contexto esportivo.

REFERÊNCIAS

- AGARWAL, A. *et al.* Sentiment analysis of Twitter data. *In*: NAGARAJAN, M.; GAMON, M. (Ed.). **Proceedings of the Workshop on Language in Social Media (LSM 2011)**. Portland, Oregon: Association for Computational Linguistics, 2011. p. 30–38. Disponível em: <https://aclanthology.org/W11-0705>.
- AGGARWAL, C. C.; AGGARWAL, C. C. **Mining text data**. [S.l.]: Springer, 2015.
- ALMEIDA, A.; CARVALHO, F.; MENINO, F. **Introdução ao Machine Learning**. DATAAT, 2017. Disponível em: <https://dataat.github.io/introducao-ao-machine-learning/index.html#grupo-dataat>.
- ALPAC. **Languages and machines: computers in translation and linguistics**. Washington, D.C, 1966. Disponível em: <https://14509501.fs1.hubspotusercontent-na1.net/hubfs/14509501/ALPAC-1996.pdf>.
- ALPAYDIN, E. **Introduction to machine learning**. [S.l.]: MIT press, 2020.
- ALSHAWI, H. **The Core Language Engine**. CogNet, 1992. (ACL-MIT Press series in natural-language processing). ISBN 9780262011266. Disponível em: https://books.google.com.ag/books?id=_cv4i6heNmWC.
- ASLAN, S.; KIZILOLUK, S.; SERT, E. Tsa-cnn-aoa: Twitter sentiment analysis using cnn optimized via arithmetic optimization algorithm. **Neural Computing and Applications**, Springer, p. 1–18, 2023.
- BENGIO, Y.; COURVILLE, A.; VINCENT, P. Representation learning: A review and new perspectives. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 35, n. 8, p. 1798–1828, 2013.
- BENGIO, Y.; DUCHARME, R.; VINCENT, P. A neural probabilistic language model. **Advances in neural information processing systems**, v. 13, 2000.
- BENGIO, Y.; GOODFELLOW, I.; COURVILLE, A. **Deep learning**. [S.l.]: MIT press Cambridge, MA, USA, 2017. v. 1.
- BENGIO, Y.; SIMARD, P.; FRASCONI, P. Learning long-term dependencies with gradient descent is difficult. **IEEE transactions on neural networks**, IEEE, v. 5, n. 2, p. 157–166, 1994.
- BIFET, A.; FRANK, E. Sentiment knowledge discovery in twitter streaming data. *In*: SPRINGER. **International conference on discovery science**. [S.l.], 2010. p. 1–15.
- BIRD, S.; KLEIN, E.; LOPER, E. **Natural language processing with Python: analyzing text with the natural language toolkit**. [S.l.]: "O'Reilly Media, Inc.", 2009.
- BISHOP, C. M.; NASRABADI, N. M. **Pattern recognition and machine learning**. [S.l.]: Springer, 2006. v. 4.
- BLEI, D. M.; NG, A. Y.; JORDAN, M. I. Latent dirichlet allocation. **Journal of machine Learning research**, v. 3, n. Jan, p. 993–1022, 2003.
- BOLAND, K.; KRÜGER, F. Distant supervision for silver label generation of software mentions in social scientific publications. *In*: **BIRNDL@ SIGIR**. [S.l.: s.n.], 2019. p. 15–27.

- BOYD, D.; CRAWFORD, K. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. **Information, communication & society**, Taylor & Francis, v. 15, n. 5, p. 662–679, 2012.
- BOYLE, R. **Power play: Sport, the media and popular culture**. [S.l.]: Edinburgh University Press, 2009.
- BRADLEY, A. P. The use of the area under the roc curve in the evaluation of machine learning algorithms. **Pattern recognition**, Elsevier, v. 30, n. 7, p. 1145–1159, 1997.
- BREIMAN, L. Bagging predictors. **Machine learning**, Springer, v. 24, p. 123–140, 1996.
- BREIMAN, L. Random forests. **Machine learning**, Springer, v. 45, p. 5–32, 2001.
- BRISCOE, T. *et al.* A formalism and environment for the development of a large grammar of english. *In: IJCAI*. [S.l.: s.n.], 1987. v. 87, p. 703–708.
- CAMBRIA, E. *et al.* New avenues in opinion mining and sentiment analysis. **IEEE Intelligent Systems**, v. 28, n. 2, p. 15–21, 2013.
- CANTALUPPI, E.; DAVINO, G. E. Beyond the grid: the digital side of the drive to survive series. **AVANCA | CINEMA**, p. 886–893, 2021.
- CAO, Y. *et al.* A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt. 2023.
- CASTELVECCHI, D. Can we open the black box of ai? **Nature News**, v. 538, n. 7623, p. 20, 2016.
- CHARNIAK, E. **Statistical Language Learning**. A Bradford Book, 1996. (A Bradford book). ISBN 9780262531412. Disponível em: <https://books.google.com.br/books?id=ps3mqZANrHUC>.
- CHEN, D. *et al.* Reading wikipedia to answer open-domain questions. **arXiv preprint arXiv:1704.00051**, 2017.
- CHO, K. *et al.* Learning phrase representations using rnn encoder-decoder for statistical machine translation. **arXiv preprint arXiv:1406.1078**, 2014.
- CHOWDHURY, G. Natural language processing. **Annual Review of Information Science and Technology**, v. 37, n. 1, p. 51–89, jan. 2003. ISSN 0066-4200.
- CHOWDHURY, G. G. **Introduction to modern information retrieval**. [S.l.]: Facet publishing, 2010.
- CHUNG, J. *et al.* Empirical evaluation of gated recurrent neural networks on sequence modeling. **arXiv preprint arXiv:1412.3555**, 2014.
- COBBS, J. *et al.* Prioritizing sponsorship resources in formula one racing: A longitudinal analysis. **Journal of Sport Management**, Human Kinetics, Inc., v. 31, n. 1, p. 96–110, 2017.
- COLEMAN, J. S. **Introducing speech and language processing**. [S.l.]: Cambridge university press, 2005.
- COLLOBERT, R.; WESTON, J. A unified architecture for natural language processing: Deep neural networks with multitask learning. *In: Proceedings of the 25th international conference on Machine learning*. [S.l.: s.n.], 2008. p. 160–167.
- CORTES, C.; VAPNIK, V. Support-vector networks. **Machine learning**, Springer, v. 20, p. 273–297, 1995.

- DAI, Z. *et al.* Transformer-xl: Attentive language models beyond a fixed-length context. **arXiv preprint arXiv:1901.02860**, 2019.
- DAS, A. *et al.* Automated pipeline for sentiment analysis of political tweets. *In: IEEE. 2021 International Conference on Data Mining Workshops (ICDMW)*. [S.l.], 2021. p. 128–135.
- DAVIS, J.; GOADRICH, M. The relationship between precision-recall and roc curves. *In: Proceedings of the 23rd international conference on Machine learning*. [S.l.: s.n.], 2006. p. 233–240.
- DEVLIN, J. *et al.* BERT: Pre-training of deep bidirectional transformers for language understanding. *In: BURSTEIN, J.; DORAN, C.; SOLORIO, T. (Ed.). Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, 2019. p. 4171–4186. Disponível em: <https://aclanthology.org/N19-1423>.
- FAWCETT, T. An introduction to roc analysis. **Pattern recognition letters**, Elsevier, v. 27, n. 8, p. 861–874, 2006.
- GARDNER, A. *et al.* **Natural Language Understanding**. [S.l.]: Stanford Univ. Department of Computer Science, 1979.
- GO, A.; BHAYANI, R.; HUANG, L. Twitter sentiment classification using distant supervision. **CS224N project report, Stanford**, v. 1, n. 12, p. 2009, 2009.
- GOULARAS, D.; KAMIS, S. Evaluation of deep learning techniques in sentiment analysis from twitter data. *In: 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML)*. [S.l.: s.n.], 2019. p. 12–17.
- GREEN, B. F. *et al.* Baseball: An automatic question-answerer. *In: Papers Presented at the May 9-11, 1961, Western Joint IRE-AIEE-ACM Computer Conference*. New York, NY, USA: Association for Computing Machinery, 1961. (IRE-AIEE-ACM '61 (Western)), p. 219–224. ISBN 9781450378727. Disponível em: <https://doi.org/10.1145/1460690.1460714>.
- GRISHMAN, R. Information extraction. **IEEE Intelligent Systems**, IEEE, v. 30, n. 5, p. 8–15, 2015.
- HAN, J.; PEI, J.; TONG, H. **Data mining: concepts and techniques**. [S.l.]: Morgan kaufmann, 2022.
- HARRIS, Z. S. Distributional structure. **Word**, Taylor & Francis, v. 10, n. 2-3, p. 146–162, 1954.
- HASTIE, T. *et al.* **The elements of statistical learning: data mining, inference, and prediction**. [S.l.]: Springer, 2009. v. 2.
- HEIFT, T.; SCHULZE, M. **Errors and intelligence in computer-assisted language learning: Parsers and pedagogues**. [S.l.]: Routledge, 2007.
- HINTON, G. *et al.* Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. **IEEE Signal processing magazine**, IEEE, v. 29, n. 6, p. 82–97, 2012.
- HIRSCHBERG, J.; MANNING, C. D. Advances in natural language processing. **Science**, v. 349, p. 261 – 266, 2015.
- HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. **Neural computation**, MIT press, v. 9, n. 8, p. 1735–1780, 1997.

- HORTON, R. *et al.* Direct-to-consumer genetic testing. **Bmj**, British Medical Journal Publishing Group, v. 367, 2019.
- HOVY, E. *et al.* Ontonotes: the 90% solution. *In: Proceedings of the human language technology conference of the NAACL, Companion Volume: Short Papers*. [S.l.: s.n.], 2006. p. 57–60.
- HU, M.; LIU, B. Mining and summarizing customer reviews. *In: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. [S.l.: s.n.], 2004. p. 168–177.
- HUTCHINS, B. Signs of meta-change in second modernity: the growth of e-sport and the world cyber games. **New Media & Society**, SAGE Publications Sage UK: London, England, v. 10, n. 6, p. 851–869, 2008.
- HUTCHINS, W. J. **Machine translation: past, present, future**. [S.l.]: Ellis Horwood Chichester, 1986.
- III, H. D.; MARCU, D. Domain adaptation for statistical classifiers. **Journal of artificial Intelligence research**, v. 26, p. 101–126, 2006.
- INZALKAR, S.; SHARMA, J. A survey on text mining-techniques and application. **International Journal of Research In Science & Engineering**, v. 24, p. 1–14, 2015.
- JAPKOWICZ, N.; SHAH, M. **Evaluating learning algorithms: a classification perspective**. [S.l.]: Cambridge University Press, 2011.
- JENKINS, H. Convergence culture: Where old and new media collide. **New York, NY: New**, 2006.
- JOHNSON, R.; ZHANG, T. Effective use of word order for text categorization with convolutional neural networks. **CoRR**, abs/1412.1058, 2014. Disponível em: <http://arxiv.org/abs/1412.1058>.
- JONES, K. S. Natural language processing: a historical review. **Current issues in computational linguistics: in honour of Don Walker**, Springer, p. 3–16, 1994.
- JONES, K. S.; GALLIERS, J. R. **Evaluating Natural Language Processing Systems: An Analysis and Review**. Springer Science & Business Media, 1995. (Lecture Notes in Artificial Intelligence). Disponível em: <https://books.google.com.br/books?id=8xPCC7H9c9oC>.
- JOSEPH, S. I. T.; THANAKUMAR, I. Survey of data mining algorithm's for intelligent computing system. **Journal of trends in Computer Science and Smart technology (TCSST)**, v. 1, n. 01, p. 14–24, 2019.
- JOUPPI, N. P. *et al.* In-datacenter performance analysis of a tensor processing unit. *In: Proceedings of the 44th annual international symposium on computer architecture*. [S.l.: s.n.], 2017. p. 1–12.
- JR, D. W. H.; LEMESHOW, S.; STURDIVANT, R. X. **Applied logistic regression**. [S.l.]: John Wiley & Sons, 2013. v. 398.
- JURAFSKY, D.; MANNING, C. Natural language processing. **Instructor**, v. 212, n. 998, p. 3482, 2012.
- KAMP, H.; REYLE, U. Tense and aspect. *In: From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Dordrecht: Springer Netherlands, 1993. p. 483–689. ISBN 978-94-017-1616-1.

- KATZ, G.; GIESBRECHT, E. Automatic identification of non-compositional multi-word expressions using latent semantic analysis. *In: Proceedings of the Workshop on Multiword Expressions: Identifying and Exploiting Underlying Properties*. [S.l.: s.n.], 2006. p. 12–19.
- KHURANA, D. *et al.* Natural language processing: State of the art, current trends and challenges. **Multimedia tools and applications**, Springer, v. 82, n. 3, p. 3713–3744, 2023.
- KIM, Y. Convolutional neural networks for sentence classification. **arXiv preprint arXiv:1408.5882**, 2014.
- KLEIN, D.; MANNING, C. D. Accurate unlexicalized parsing. *In: Proceedings of the 41st annual meeting of the association for computational linguistics*. [S.l.: s.n.], 2003. p. 423–430.
- KOTO, F.; ADRIANI, M. A comparative study on twitter sentiment analysis: Which features are good? *In: BIEMANN, C. et al.* (Ed.). **Natural Language Processing and Information Systems**. Cham: Springer International Publishing, 2015. p. 453–457.
- KOULOUMPIS, E.; WILSON, T.; MOORE, J. Twitter sentiment analysis: The good the bad and the omg! *In: Proceedings of the international AAAI conference on web and social media*. [S.l.: s.n.], 2011. v. 5, n. 1, p. 538–541.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. **Advances in neural information processing systems**, v. 25, 2012.
- LAPPALAINEN, H.; MISKIN, J. W. Ensemble learning. *In: Advances in Independent Component Analysis*. [S.l.]: Springer, 2000. p. 75–92.
- LEARNING, S.-S. Semi-supervised learning. **CSZ2006.html**, 2006.
- LEBERT, M. **Project Gutenberg (1971-2008)**. [S.l.]: Project Gutenberg, 2008.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **nature**, Nature Publishing Group UK London, v. 521, n. 7553, p. 436–444, 2015.
- LIANG, J. Confusion matrix: Machine learning. **POGIL Activity Clearinghouse**, v. 3, n. 4, 2022.
- LIDDY, E. D. Natural language processing. *In: Encyclopedia of Library and Information Science*. 2. ed. NY, USA: Marcel Decker, Inc, 2001. p. 1218–1222.
- LIU, B. **Sentiment analysis and opinion mining**. [S.l.]: Springer Nature, 2022.
- LORIA, S. *et al.* textblob documentation. **Release 0.15**, v. 2, n. 8, p. 269, 2018.
- LOUKAS, G. *et al.* Cloud-based cyber-physical intrusion detection for vehicles using deep learning. **IEEE Access**, v. 6, p. 3491–3508, 12 2017.
- MANI, I.; MAYBURY, M. T. **Advances in automatic text summarization**. [S.l.]: MIT press, 1999.
- MANN, S. *et al.* Twitter sentiment analysis using enhanced bert. *In: KULKARNI, A. J.; MIRJALILI, S.; UDGATA, S. K.* (Ed.). **Intelligent Systems and Applications**. Singapore: Springer Nature Singapore, 2023. p. 263–271.
- MANNING, C.; SCHUTZE, H. **Foundations of statistical natural language processing**. [S.l.]: MIT press, 1999.

- MAULUD, D. H. *et al.* State of art for semantic analysis of natural language processing. **Qubahan Academic Journal**, v. 1, n. 2, p. 21–28, Mar. 2021. Disponível em: <https://journal.qubahan.com/index.php/qaj/article/view/44>.
- MAYNARD, D. G.; GREENWOOD, M. A. Who cares about sarcastic tweets? investigating the impact of sarcasm on sentiment analysis. *In*: ELRA. **Lrec 2014 proceedings**. [S.l.], 2014.
- MCCALLUM, A.; NIGAM, K. *et al.* A comparison of event models for naive bayes text classification. *In*: MADISON, WI. **AAAI-98 workshop on learning for text categorization**. [S.l.], 1998. v. 752, n. 1, p. 41–48.
- MCTEAR, M. F.; CALLEJAS, Z.; GRIOL, D. **The conversational interface**. [S.l.]: Springer, 2016. v. 6.
- MEDSKER, L. R.; JAIN, L. Recurrent neural networks. **Design and Applications**, v. 5, n. 64-67, p. 2, 2001.
- MIIKKULAINEN, R. *et al.* Evolving deep neural networks. **Artificial Intelligence in the Age of Neural Networks and Brain Computing**, Elsevier, p. 293–312, 2019.
- MIKOLOV, T. *et al.* Distributed representations of words and phrases and their compositionality. **Advances in neural information processing systems**, v. 26, 2013.
- MILLER, G. A. Wordnet: a lexical database for english. **Communications of the ACM**, ACM New York, NY, USA, v. 38, n. 11, p. 39–41, 1995.
- MIN, B. *et al.* Recent advances in natural language processing via large pre-trained language models: A survey. **ACM Computing Surveys**, ACM New York, NY, v. 56, n. 2, p. 1–40, 2023.
- MITCHELL, T. **Machine Learning**. McGraw-Hill, 1997. (McGraw-Hill International Editions). ISBN 9780071154673. Disponível em: <https://books.google.com.br/books?id=EoYBngEACAAJ>.
- MOSTAFA, M. M. More than words: Social networks' text mining for consumer brand sentiments. **Expert systems with applications**, Elsevier, v. 40, n. 10, p. 4241–4251, 2013.
- NADEAU, D.; SEKINE, S. A survey of named entity recognition and classification. **Lingvisticae Investigationes**, John Benjamins, v. 30, n. 1, p. 3–26, 2007.
- NAVIGLI, R. Word sense disambiguation: A survey. **ACM computing surveys (CSUR)**, ACM New York, NY, USA, v. 41, n. 2, p. 1–69, 2009.
- NENKOVA, A.; MCKEOWN, K. *et al.* Automatic summarization. **Foundations and Trends® in Information Retrieval**, Now Publishers, Inc., v. 5, n. 2–3, p. 103–233, 2011.
- NIELSEN, F. Å. A new anew: Evaluation of a word list for sentiment analysis in microblogs. **arXiv preprint arXiv:1103.2903**, 2011.
- OTTEN, M.; PACAK, M. G. Intermediate languages for automatic language processing. *In*: TOU, J. T. (Ed.). **Computer and Information Sciences–1969**. Elsevier, 1971, (SEN Report Series Software Engineering, v. 2). p. 105–118. Disponível em: <https://www.sciencedirect.com/science/article/pii/B9780126962024500140>.
- OWENS, J. D. *et al.* Gpu computing. **Proceedings of the IEEE**, IEEE, v. 96, n. 5, p. 879–899, 2008.
- PAK, A.; PAROUBEK, P. *et al.* Twitter as a corpus for sentiment analysis and opinion mining. *In*: **LREc**. [S.l.: s.n.], 2010. v. 10, n. 2010, p. 1320–1326.

- PANG, B.; LEE, L. *et al.* Opinion mining and sentiment analysis. **Foundations and Trends® in information retrieval**, Now Publishers, Inc., v. 2, n. 1–2, p. 1–135, 2008.
- PARSING, C. Speech and language processing. **Power Point Slides**, 2009.
- PAUL, M. J.; DREDZE, M. **Social monitoring for public health**. [S.l.]: Morgan & Claypool Publishers, 2017.
- PENNINGTON, J.; SOCHER, R.; MANNING, C. D. Glove: Global vectors for word representation. *In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. [S.l.: s.n.], 2014. p. 1532–1543.
- PERMANASARI, A.; CHAMSUDIN, A.; WAHYUNGGORO, O. Utilization of neural network for disease forecasting. *In: .* [S.l.: s.n.], 2013.
- PHAM, Q.-V. *et al.* Artificial intelligence (ai) and big data for coronavirus (covid-19) pandemic: a survey on the state-of-the-arts. **IEEE access**, IEEE, v. 8, p. 130820–130839, 2020.
- PHUNG, V. H.; RHEE, E. J. A high-accuracy model average ensemble of convolutional neural networks for classification of cloud image patches on small datasets. **Applied Sciences**, v. 9, p. 4500, 10 2019.
- PLISSON, J. *et al.* A rule based approach to word lemmatization. *In: Proceedings of IS*. [S.l.: s.n.], 2004. v. 3, p. 83–86.
- PORTER, M. F. An algorithm for suffix stripping. **Program**, Emerald Group Publishing Limited, v. 40, n. 3, p. 211–218, 2006.
- POWERS, D. M. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. **arXiv preprint arXiv:2010.16061**, 2020.
- QUINLAN, J. R. Induction of decision trees. **Machine learning**, Springer, v. 1, p. 81–106, 1986.
- RADFORD, A. *et al.* Improving language understanding by generative pre-training. OpenAI, 2018.
- RADFORD, A. *et al.* Language models are unsupervised multitask learners. **OpenAI blog**, v. 1, n. 8, p. 9, 2019.
- RAE, J. W. *et al.* Compressive transformers for long-range sequence modelling. **arXiv preprint arXiv:1911.05507**, 2019.
- RAFFEL, C. *et al.* Exploring the limits of transfer learning with a unified text-to-text transformer. **Journal of Machine Learning Research**, v. 21, n. 140, p. 1–67, 2020. Disponível em: <http://jmlr.org/papers/v21/20-074.html>.
- ROGERS, A.; KOVALEVA, O.; RUMSHISKY, A. A Primer in BERTology: What We Know About How BERT Works. **Transactions of the Association for Computational Linguistics**, v. 8, p. 842–866, 01 2021.
- ROSTAMI, A. *et al.* Risk management implementation in small and medium enterprises in the uk construction industry. **Engineering, Construction and Architectural Management**, Emerald Group Publishing Limited, v. 22, n. 1, p. 91–107, 2015.
- SAIF, H. *et al.* Evaluation datasets for twitter sentiment analysis: a survey and a new dataset, the sts-gold. 2013.

- SAITO, T.; REHMSMEIER, M. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. **PloS one**, Public Library of Science San Francisco, CA USA, v. 10, n. 3, p. e0118432, 2015.
- SARANYA, S.; USHA, G. A machine learning-based technique with intelligentwordnet lemmatize for twitter sentiment analysis. **Intelligent Automation & Soft Computing**, v. 36, n. 1, 2023.
- SCHMOKER, M. Results: the key to continuous school improvement. **NASSP Bulletin**, v. 84, n. 615, p. 120–122, 2000. Disponível em: <https://doi.org/10.1177/019263650008461515>.
- SCHÜTZE, H.; MANNING, C. D.; RAGHAVAN, P. **Introduction to information retrieval**. [S.l.]: Cambridge University Press Cambridge, 2008. v. 39.
- SEBASTIANI, F.; ESULI, A. Sentiwordnet: A publicly available lexical resource for opinion mining. *In*: EUROPEAN LANGUAGE RESOURCES ASSOCIATION (ELRA) GENOA, ITALY. **Proceedings of the 5th international conference on language resources and evaluation**. [S.l.], 2006. p. 417–422.
- SINCLAIR, J. **Corpus, Concordance, Collocation**. Oxford University Press, 1991. (Describing English language). ISBN 9780194371445. Disponível em: <https://books.google.com.br/books?id=L8l4AAAAIAAJ>.
- SMALL, S. L.; COTTRELL, G. W.; TANENHAUS, M. K. **Lexical ambiguity resolution: Perspective from psycholinguistics, neuropsychology and artificial intelligence**. [S.l.]: Elsevier, 2013.
- SOSA, P. M. Twitter sentiment analysis using combined lstm-cnn models. **Eprint Arxiv**, v. 2017, p. 1–9, 2017.
- STEHMAN, S. V. Selecting and interpreting measures of thematic classification accuracy. **Remote sensing of Environment**, Elsevier, v. 62, n. 1, p. 77–89, 1997.
- STURM, D. A glamorous and high-tech global spectacle of speed: Formula one motor racing as mediated, global and corporate spectacle. *In*: **Sports events, society and culture**. [S.l.]: Routledge, 2014. p. 68–82.
- SUNDHEIM, B. M.; CHINCHOR, N. A. Survey of the message understanding conferences. *In*: **Proceedings of the Workshop on Human Language Technology**. USA: Association for Computational Linguistics, 1993. (HLT '93), p. 56–60. ISBN 1558603247. Disponível em: <https://doi.org/10.3115/1075671.1075684>.
- SUTSKEVER, I.; VINYALS, O.; LE, Q. V. Sequence to sequence learning with neural networks. **Advances in neural information processing systems**, v. 27, 2014.
- SUTTON, R. S.; BARTO, A. G. **Reinforcement learning: An introduction**. [S.l.]: MIT press, 2018.
- SWATHI, T.; KASIVISWANATH, N.; RAO, A. A. An optimal deep learning-based lstm for stock price prediction using twitter sentiment analysis. **Applied Intelligence**, Springer, v. 52, n. 12, p. 13675–13688, 2022.
- SYLT, C.; REID, C. **Formula Money: Formula One's Financial Performance Guide**. Money Sport Media, 2011. Disponível em: <https://books.google.com.br/books?id=kcKfjwEACAAJ>.
- TARWANI, K. M.; EDEM, S. Survey on recurrent neural network in natural language processing. **Int. J. Eng. Trends Technol**, v. 48, n. 6, p. 301–304, 2017.

- THELWALL, M. *et al.* Sentiment strength detection in short informal text. **Journal of the American society for information science and technology**, Wiley Online Library, v. 61, n. 12, p. 2544–2558, 2010.
- TSYTSARAU, M.; PALPANAS, T. Survey on mining subjective data on the web. **Data Mining and Knowledge Discovery**, Springer, v. 24, p. 478–514, 2012.
- TUMASJAN, A. *et al.* Predicting elections with twitter: What 140 characters reveal about political sentiment. *In: Proceedings of the international AAAI conference on web and social media*. [S.l.: s.n.], 2010. v. 4, n. 1, p. 178–185.
- TURNEY, P. D. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. **arXiv preprint cs/0212032**, 2002.
- VASILIEV, Y. **Natural language processing with Python and spaCy: A practical introduction**. [S.l.]: No Starch Press, 2020.
- VASWANI, A. *et al.* Attention is all you need. *In: GUYON, I. et al.* (Ed.). **Advances in Neural Information Processing Systems**. Curran Associates, Inc., 2017. v. 30. Disponível em: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.
- VASWANI, A. *et al.* Attention is all you need. **Advances in neural information processing systems**, v. 30, 2017.
- VINODHINI, G.; CHANDRASEKARAN, R. Sentiment analysis and opinion mining: a survey. **International Journal**, v. 2, n. 6, p. 282–292, 2012.
- WANG, A. *et al.* Superglue: A stickier benchmark for general-purpose language understanding systems. **Advances in neural information processing systems**, v. 32, 2019.
- WANG, A. *et al.* Glue: A multi-task benchmark and analysis platform for natural language understanding. **arXiv preprint arXiv:1804.07461**, 2018.
- WANG, W.; GANG, J. Application of convolutional neural network in natural language processing. *In: 2018 International Conference on Information Systems and Computer Aided Education (ICISCAE)*. [S.l.: s.n.], 2018. p. 64–70.
- WANG, X. *et al.* Topic sentiment analysis in twitter: A graph-based hashtag sentiment classification approach. *In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management*. New York, NY, USA: Association for Computing Machinery, 2011. (CIKM '11), p. 1031–1040. ISBN 9781450307178. Disponível em: <https://doi.org/10.1145/2063576.2063726>.
- WERBOS, P. Backpropagation through time: what it does and how to do it. **Proceedings of the IEEE**, v. 78, n. 10, p. 1550–1560, 1990.
- WIDIASTUTI, N. Convolution neural network for text mining and natural language processing. *In: IOP PUBLISHING. IOP Conference Series: Materials Science and Engineering*. [S.l.], 2019. v. 662, n. 5, p. 052010.
- WILBANKS, B. A.; LANGFORD, P. A. A review of dashboards for data analytics in nursing. **CIN: Computers, Informatics, Nursing**, v. 32, p. 545–549, 2014.
- WINOGRAD, T. **Procedures as a Representation for Data in a Computer Program for Understanding Natural Language**. M.I.T. Project MAC, 1971. (AI-TR). Disponível em: <https://books.google.com.br/books?id=afh4oAEACAAJ>.

WITTEN, I. H. *et al.* Practical machine learning tools and techniques. *In*: ELSEVIER AMSTERDAM, THE NETHERLANDS. **Data mining**. [S.l.], 2005. v. 2, n. 4, p. 403–413.

WOOD, N. T.; BURKHALTER, J. N. A formula for success: How formula one racing embraced digital and social media to engage fans. **Journal of Digital & Social Media Marketing**, Henry Stewart Publications, v. 11, n. 1, p. 43–59, 2023.

WOODS, W. Semantics and quantification in natural language question answering. *In*: YOVITS, M. C. (Ed.). Elsevier, 1978, (*Advances in Computers*, v. 17). p. 1–87. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0065245808603903>.

WORATSCHEK, H.; HORBEL, C.; POPP, B. Value co-creation in sport management. **European Sport Management Quarterly**, v. 14, 02 2014.

WU, Y. *et al.* **Google’s Neural Machine Translation System: Bridging the Gap between Human and Machine Translation**. 2016.

YOUNG, S. J.; CHASE, L. L. Speech recognition evaluation: a review of the u.s. csr and lvcsr programmes. **Computer Speech & Language**, v. 12, n. 4, p. 263–279, 1998. ISSN 0885-2308. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0885230898901012>.

YUAN, W.; NEUBIG, G.; LIU, P. **BARTScore: Evaluating Generated Text as Text Generation**. 2021.

ZHANG, A. *et al.* **Dive into Deep Learning**. [s.n.], 2021. Disponível em: <https://d2l.ai/>.

ZHANG, L.; WANG, S.; LIU, B. Deep learning for sentiment analysis: A survey. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, Wiley Online Library, v. 8, n. 4, p. e1253, 2018.

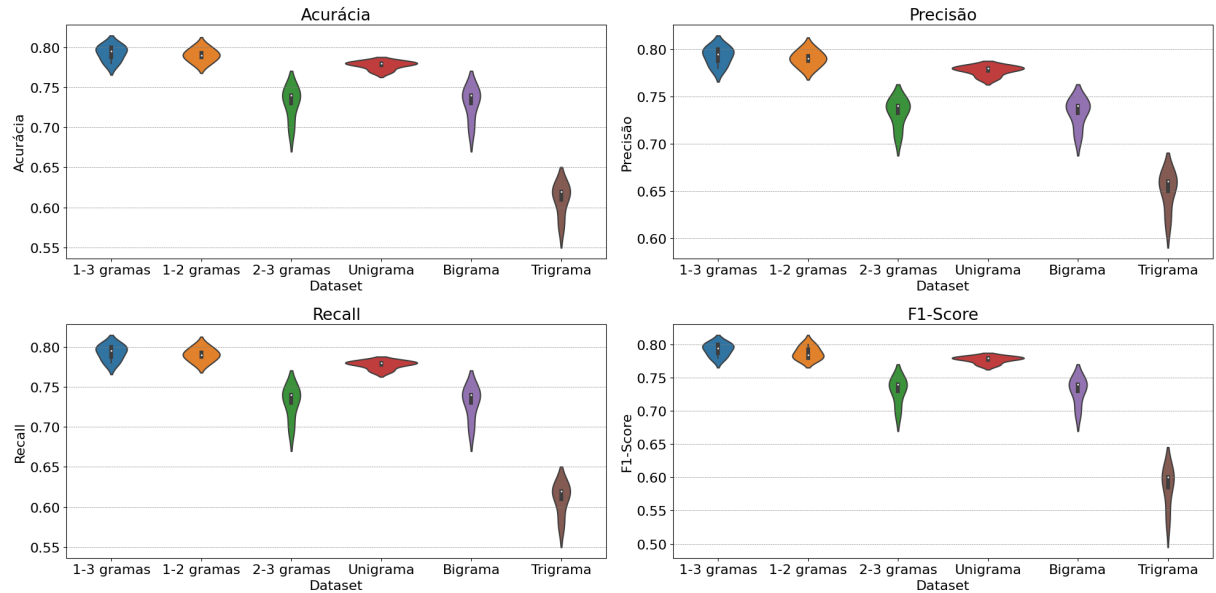
ZHANG, X.; ZHAO, J.; LECUN, Y. Character-level convolutional networks for text classification. **Advances in neural information processing systems**, v. 28, 2015.

APÊNDICE A – Gráficos Complementares

A.1 Resultados do Experimento de Definição de N-gramas para o TF-IDF

A Figura 27 apresenta a comparação entre as diferentes métricas analisadas unindo a execução dos modelos por *dataset* de n-gramas.

Figura 27 – Gráfico de Violino das métricas dos modelos em relação aos *datasets* de n-gramas.

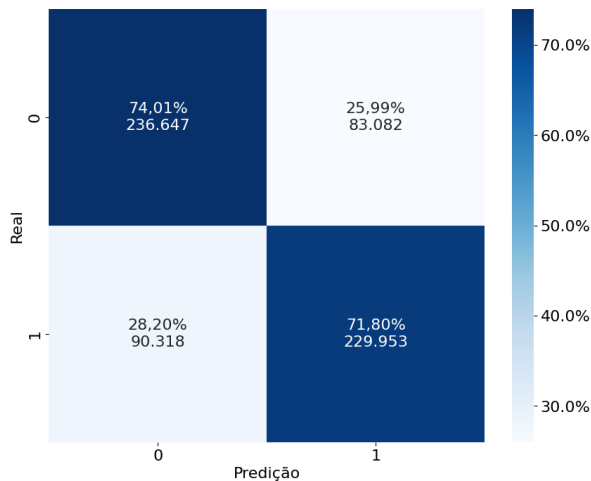


Fonte: Autoria própria (2023).

A.2 Resultados do Treinamento dos Modelos de AM e AP

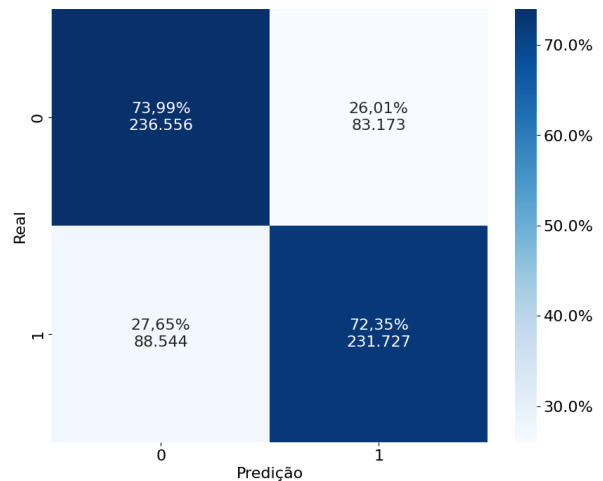
Essa seção apresenta as matrizes de confusão, curva ROC e curva *Precision-Recall* dos modelos RF, MLP e CNN.

Figura 28 – Matriz de Confusão do RF.



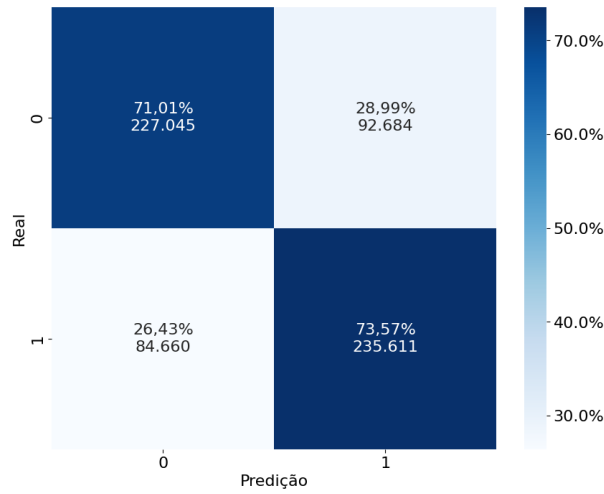
Fonte: Autoria própria (2023).

Figura 29 – Matriz de Confusão do MLP.



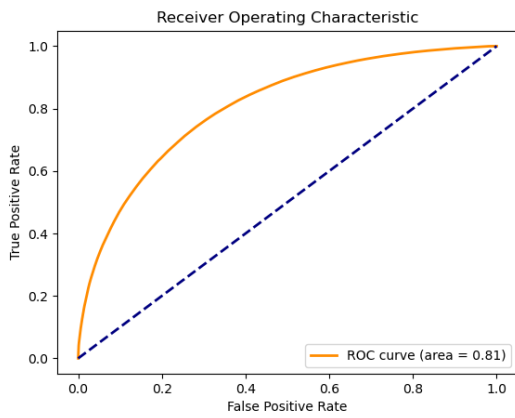
Fonte: Autoria própria (2023).

Figura 30 – Matriz de Confusão do CNN.



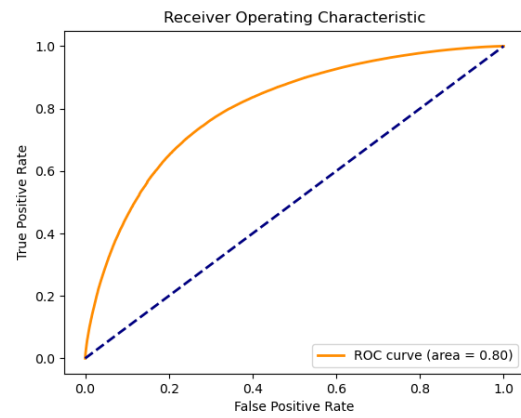
Fonte: Autoria própria (2023).

Figura 31 – Curva ROC do RF.



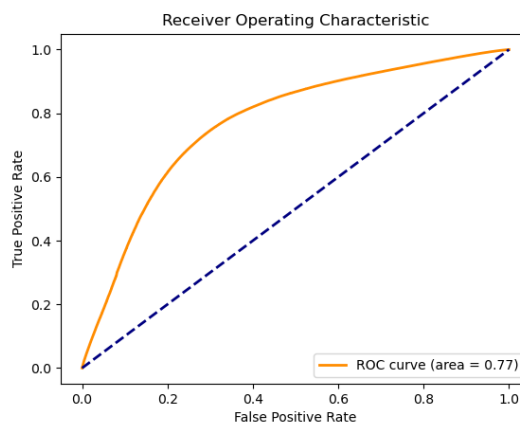
Fonte: Autoria própria (2023).

Figura 32 – Curva ROC do MLP.



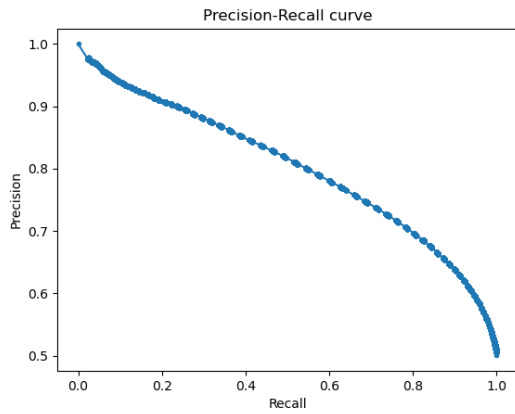
Fonte: Autoria própria (2023).

Figura 33 – Curva ROC do CNN.



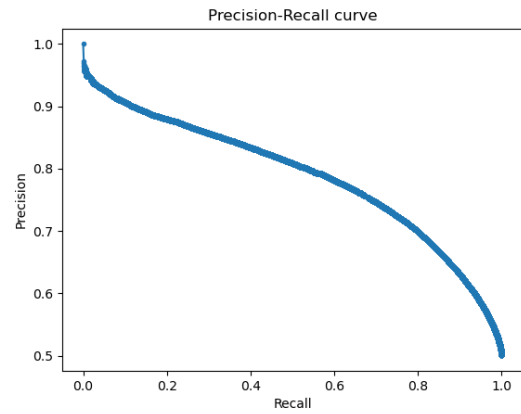
Fonte: Autoria própria (2023).

Figura 34 – Curva *Precision-Recall* do RF.



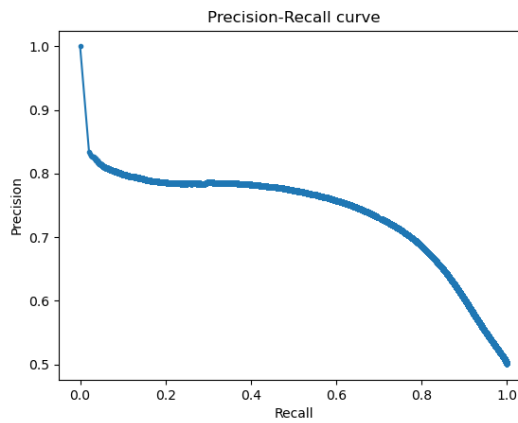
Fonte: Autoria própria (2023).

Figura 35 – Curva *Precision-Recall* do MLP.



Fonte: Autoria própria (2023).

Figura 36 – Curva *Precision-Recall* do CNN.



Fonte: Autoria própria (2023).