



**UNIVERSIDADE
FEDERAL RURAL
DE PERNAMBUCO**



Júlia de Melo Albuquerque

Análise de Sentimentos em Reviews de Jogos Digitais da Plataforma Steam

Recife

2024

Júlia de Melo Albuquerque

Análise de Sentimentos em Reviews de Jogos Digitais da Plataforma Steam

Monografia apresentada ao Curso de Bacharelado em Sistemas de Informação da Universidade Federal Rural de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

Universidade Federal Rural de Pernambuco – UFRPE

Departamento de Estatística e Informática

Curso de Bacharelado em Sistemas de Informação

Orientador: Gabriel Alves

Recife

2024

Dados Internacionais de Catalogação na Publicação
Sistema Integrado de Bibliotecas da UFRPE
Bibliotecário(a): Ana Catarina Macêdo – CRB-4 1781

A345a Albuquerque, Júlia de Melo.
Análise de sentimentos em reviews de jogos digitais da
plataforma Steam / Júlia de Melo Albuquerque. – Recife,
2024.
49 f.; il.

Orientador(a): Gabriel Alvez.

Trabalho de Conclusão de Curso (Graduação) –
Universidade Federal Rural de Pernambuco, Bacharelado
em Sistemas da Informação, Recife, BR-PE, 2024.

Inclui referências e anexo(s).

1. Teoria dos autômatos. 2. Jogos eletrônicos. 3.
Interação homem-máquina. I. Alvez, Gabriel, orient. II.
Título

CDD 004

JÚLIA DE MELO ALBUQUERQUE

Análise de Sentimentos em Reviews de Jogos Digitais da Plataforma Steam

Monografia apresentada ao Curso de Bacharelado em Sistemas de Informação da Universidade Federal Rural de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

Aprovada em: 26 de Setembro de 2024.

BANCA EXAMINADORA

Gabriel Alves (Orientador)
Departamento de Estatística e Informática
Universidade Federal Rural de Pernambuco

Cleviton Monteiro
Departamento de Estatística e Informática
Universidade Federal Rural de Pernambuco

Agradecimentos

Agradeço primeiramente à minha família, minha mãe, Débora Dias, e ao meu pai, José Plínio Filho, por acreditarem em mim em todos os momentos e me dar forças para tudo.

Ao meu orientador, Professor Gabriel Alves, pela dedicação e paciência nesse momento de conclusão de curso e em toda minha trajetória como estudante.

Uma menção honrosa aos meus colegas de faculdade, Carlos Henrique, Isadora Tavares e Everton Matheus, que compartilharam comigo tantos momentos de estudo, desafios e conquistas. A colaboração e amizade de vocês foram fundamentais para tornar essa jornada mais leve.

“Me esforço para ser melhor a cada dia. Pois bondade também se aprende.”
(Cora Coralina)

Resumo

A análise de sentimento é uma área que investiga as expressões emocionais da linguagem humana, buscando compreender as necessidades e opiniões subjacentes expressas em textos. Sua complexidade existe na habilidade de discernir não apenas o conteúdo textual, mas também as matrizes emocionais implícitas. Com o avanço tecnológico, a facilidade de expressar a opinião publicamente se dissemina por diversos meios, sendo campo dos jogos online uma vertente que atrai inúmeras publicações de jogadores sobre os diversos títulos disponíveis. No entanto, essa diversidade de públicos e temas torna desafiador compreender o sentimento expresso que permeia esse universo. O objetivo deste estudo é aplicar técnicas de análise de sentimento em avaliações de jogos digitais, adotando uma abordagem focada em algoritmos de aprendizado de máquina supervisionados e bibliotecas pre-polarizadas, a fim de identificar o melhor caminho de classificação capaz de discernir os sentimentos expressos pelos usuários nas avaliações. Nessa operação está sendo considerado uma abordagem com todas as opiniões e outra focada em gênero específico de cada jogo. Essa análise foi conduzida por meio da exploração de dados provenientes de uma empresa de distribuição de jogos online (Steam), seguindo com uma preparação desses dados devido às particularidades presentes nos registros. Os resultados revelam que os modelos de aprendizado de máquina superam as abordagens tradicionais, como a utilização da biblioteca VADER, apresentando uma precisão maior em aproximadamente 10% nas capturas, observou-se uma diferença de 20% a mais em métricas como recall e F1-score. Esse estudo representa uma contribuição analítica para a área da análise de sentimento, destacando a capacidade do modelo construído em lidar com a complexidade da linguagem humana.

Palavras-chave: Aprendizado de Máquina, Sentimentos, Jogos, Opinião, Avaliações, Classificação, Bibliotecas.

Abstract

Sentiment analysis is an area that investigates the emotional expressions of human language, aiming to understand the underlying needs and opinions expressed in texts. Its complexity lies in the ability to discern not only the textual content but also the implicit emotional matrices. With technological advancements, the ease of publicly expressing opinions is disseminated through various means, with online gaming being a sector that attracts numerous player posts about various available titles. However, this diversity of audiences and topics makes it challenging to understand the expressed sentiment that pervades this universe. The aim of this study is to apply sentiment analysis techniques to digital game reviews, adopting an approach focused on supervised machine learning algorithms and pre-polarized libraries, in order to identify the best classification path capable of discerning the sentiments expressed by users in the reviews. This operation considers an approach with all opinions and another focused on each game's specific genre. This analysis was conducted by exploring data from an online game distribution company (Steam), followed by data preparation due to the peculiarities present in the records. The results reveal that machine learning models outperform traditional approaches, such as using the VADER library, showing a higher precision by approximately 10% in captures. A difference of 20% more was observed in metrics such as recall and F1-score. This study represents an analytical contribution to the field of sentiment analysis, highlighting the model's ability to deal with the complexity of human language.

Keywords: Machine Learning, Sentiment, Games, Opinion, Reviews, Classification, Libraries.

Lista de ilustrações

Figura 1 – Faturamento das desenvolvedoras brasileiras de jogos em 2022 . . .	10
Figura 2 – Faturamento anual global do mercado de games em 2021.	14
Figura 3 – Ilustração do funcionamento da árvore de decisão.	20
Figura 4 – Comportamento dos hiperplanos nas margens.	22
Figura 5 – Ilustração da matriz de confusão.	23
Figura 6 – Ilustração gráfica da curva ROC-AUC.	25
Figura 7 – Etapas de desenvolvimento do modelo.	29
Figura 8 – Comentário avaliativo retirado da Steam.	30
Figura 9 – Quantidade de recomendações positivas e negativas.	35
Figura 10 – Ilustração dos valores referentes a curva ROC.	37
Figura 11 – Métricas referentes a curva ROC por gênero usando SVM.	39
Figura 12 – Métricas valores referentes a curva ROC por gênero usando RF. . .	39
Figura 13 – Valores referentes a curva ROC do gênero Terror.	40
Figura 14 – Valores referentes a curva ROC do gênero Anime.	41
Figura 15 – Ilustração das features mais importantes na previsão do sentimento (terror).	42
Figura 16 – Ilustração das features mais importantes na previsão do sentimento (anime).	43

Lista de tabelas

Tabela 1 – Exemplo gerado dos dados da Steam ao executar a função de remoção e formatação.	16
Tabela 2 – Exemplo gerado dos dados da Steam ao executar a biblioteca NLTK.	17
Tabela 3 – Exemplos das diferenças entre a lematização e a stemização.	17
Tabela 4 – Frase original e modificada após o pré-processamento.	32
Tabela 5 – Métricas dos algoritmos trabalhados.	36
Tabela 6 – Métricas dos modelos para identificar a categoria dos jogos.	38
Tabela 7 – Métricas de identificação dos sentimentos em reviews de jogos de terror.	40
Tabela 8 – Métricas de identificação dos sentimentos em reviews de jogos de anime.	41

Lista de abreviaturas e siglas

PLN	Processamento de linguagem natural
MEI	Microempreendedor individual
API	Application Programming Interface
HTTP	Hypertext Transfer Protocol
RF	Random Forest
SVM	Support Vector Machine
VADER	Valence Aware Dictionary and Sentiment Reasoner
ROC	Receiver Operating Characteristic
AUC	Area Under the Curve
NLTK	Natural Language Toolkit
TF	Term Frequency
IDF	Inverse Document Frequency

Sumário

Lista de ilustrações	6
1 INTRODUÇÃO	10
1.1 Objetivo Geral	11
1.2 Objetivo Específicos	11
2 REFERENCIAL TEÓRICO	13
2.1 Jogos Digitais	13
2.2 Análise de Sentimento	14
2.3 Processamento de Linguagem Natural	15
2.3.1 Pré-processamento	16
2.4 Aprendizado de máquina	19
2.5 Classificadores	20
2.5.1 Random Forest	20
2.5.2 Support Vector Machine (SVM)	21
2.5.3 VADER	22
2.5.4 Métricas de Avaliação	23
3 TRABALHOS RELACIONADOS	26
4 FERRAMENTAS E MÉTODOS	29
4.1 Coleta dos Dados	30
4.2 Pré-processamento	31
4.3 Modelagem	32
4.4 Avaliação	33
5 RESULTADOS	35
5.1 Experimento 1: Análise de sentimento em todas as classes	35
5.2 Experimento 2: Análise de sentimento aplicada em casos de gênero específico	37
6 CONSIDERAÇÕES FINAIS	44
REFERÊNCIAS	47

1 Introdução

A indústria de games, ao longo de décadas, está emergindo como uma forma de cultura altamente influente e transformativa, exercendo não apenas um impacto significativo no cenário do entretenimento, mas também reverberando positivamente em diversos setores financeiros (MBAKIRTZIS, 2023). Em um mundo cada vez mais digital sua perspectiva é o crescimento ao longo do tempo por se aplicar em diversos cenários como consoles, computadores, smartphones, aparelhos de realidade virtual e outros. No Brasil, a movimentação econômica dos jogos levanta mais de 100 milhão de dolares por ano, se tornando um dos principais mercados no país, com um cenário em perspectiva de crescimento.

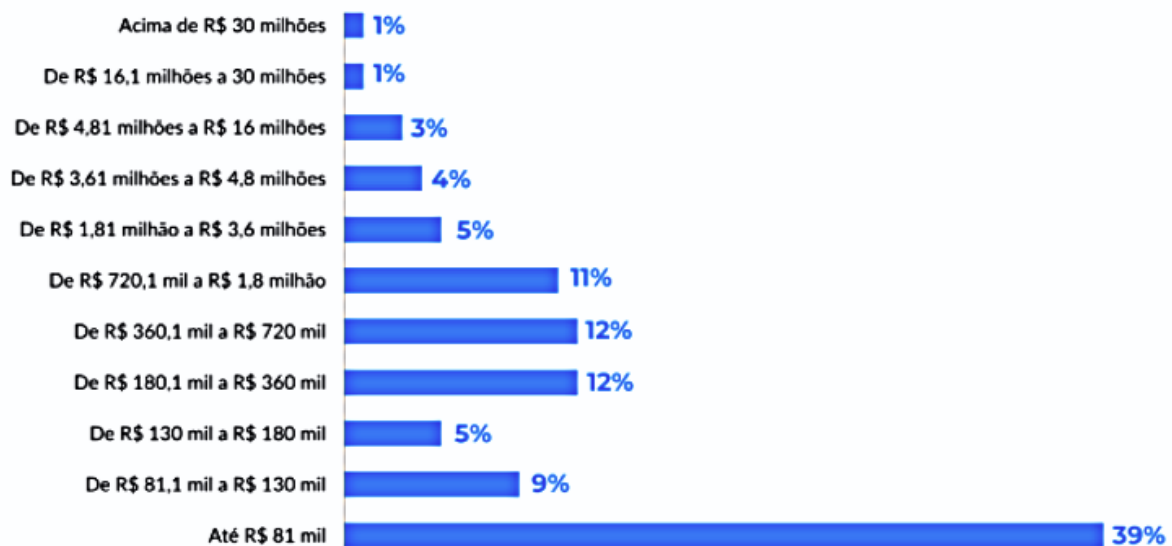


Figura 1 – Faturamento das desenvolvedoras brasileiras de jogos em 2022
Fonte: Censo de Jogos Digitais, Brazil Games (2022)

Na Figura 1, é apresentado o faturamento anual de desenvolvedores de jogos no Brasil na categoria MEI, ME e LTDA. Observa-se que, apesar da maioria ainda registrar um faturamento inferior a 100 mil, há uma variação considerável nos valores, indicando uma relação direta com o amadurecimento do cenário de jogos no país. Até 2018, mais de 73% dos casos apresentavam um faturamento abaixo dos 80 mil (JÚNIOR; CAVALCANTE,). Esta representação visual também evidencia o avanço significativo das empresas brasileiras na indústria de jogos, tanto em termos quantitativos quanto qualitativos.

Em um mercado altamente competitivo, com uma vasta gama de opções disponíveis, o sucesso de um jogo depende de ter um diferencial positivo que atenda

às necessidades do jogador. Esse ambiente é envolvido pela experiência do usuário, sendo um dos meios para potencializar o engajamento dos jogos (FELIX et al., 2020). Esse potencial existe no momento em que o jogador externaliza essa sensação de satisfação, meios comunicativos online são suas principais vias, como comunidades em sites, redes sociais e plataformas de jogos.

O caso do jogo No Man's Sky reflete como é importante realizar essa análise, no seu lançamento em 2016 os jogadores ficaram desapontados com a experiência proporcionada e realizaram diversas críticas nas plataformas de jogos a respeito de promessas não cumpridas sobre ambientação, multiplayer, recursos e jogabilidade. (LU et al., 2020). A empresa arcou com diversos prejuízos, e para conquistar o público foram necessárias diversas análises estratégicas sobre o que os usuários necessitavam para melhorar sua experiência. Com as atualizações implementadas, as críticas do público começaram a melhorar e hoje o jogo é reconhecido como uma referência em suporte ao usuário.

Abordagens alternativas na análise de sentimento são uma forma de solução nesse cenário. A análise de sentimento vem como um estudo que busca identificar o sentimento que o texto transparece, o intuito é realizar essa atividade de forma automatizada, sendo uma ferramenta valiosa institucionalmente para avaliação do mercado através da coleta de opiniões e interesses do público. Para realização desse projeto, foi necessária a coleta dos dados opinativos pela API da plataforma de jogos Steam, para transformar as informações disponíveis em campos de aprendizado para o modelo. O resultado dessa construção é uma solução eficiente na análise de sentimento a respeito de jogos, podendo facilitar estratégias de negócios em relação a experiência do jogador.

1.1 Objetivo Geral

O objetivo principal desta pesquisa é construir modelos de classificação de sentimento utilizando abordagens sobre a influência do gênero voltado especificamente para a área de jogos digitais.

1.2 Objetivo Específicos

Visando o desenvolvimento acerca dos resultados dos modelos criado para a identificação de sentimento por texto, este artigo tem como objetivos:

- Analisar um histórico de avaliações de jogos para identificação de padrões e elementos textuais que caracterizam expressão de sentimento;

- Avaliar métodos de lidar com classificação binária e multiclases;
- Classificar a categoria dos jogos pelo texto de recomendação.
- Processar grandes volumes de textos para garantir a qualidade da análise para os algoritmos;
- Comparar modelos de classificação de aprendizado de máquina e bibliotecas pré-processadas.
- Realizar uma avaliação dos resultados obtidos, utilizando métricas de desempenho.

2 Referencial Teórico

Nesta seção serão apresentados os principais conceitos relacionados à análise de sentimento e suas aplicações. Contextualizando as técnicas adotadas para construção dos modelos de aprendizado de máquina, pré-processamentos, fórmulas de metrificação e discussões de estratégias para o uso desses recursos.

2.1 Jogos Digitais

Um dos meios de manifestação mais moderno da atualidade, os jogos eletrônicos, tem um posicionamento mundial consolidado como forma de entretenimento e expressão artística. Sendo assim, o livro *Homo ludens: o jogo como elemento da cultura* (HUIZINGA, 1971) explica que os jogos podem ser definidos como uma atividade lúdica muito mais ampla que um fenômeno físico ou reflexo psicológico, ou seja, um movimento cultural que tem a evasão da vida real como principal característica.

O impacto global dos jogos transcende as fronteiras e constrói um ecossistema livre de barreiras da distância com suas singularidades, o elemento que traz força a essa globalização são os jogos online e todo o conteúdo público que gira ao redor de um jogo.

Com o crescimento do mercado a geração de receita é superior a 100 bilhões de dolares em jogos pelo mundo inteiro, como demonstra a Figura 2. No Brasil, o cenário não é diferente, sendo a décima potência do mundo o país chega a acumular números como 66,7% da sociedade brasileira consumindo jogos eletrônicos de diversas plataformas (FORBES, 2022). A consolidação dos jogos eletrônicos como uma forma de entretenimento respeitada é evidente nas abordagens artísticas, desde narrativas emocionantes e complexas até experiências mais competitivas e educativas, os jogos oferecem uma paleta rica de expressão criativa.

Como todo mercado, os jogos tem uma ampla popularidade em pautas envolvendo a opinião do público, gerando debates que podem decidir o rumo lucrativo daquela obra. A influência das opiniões pode ser propagada em várias formas como, análises de jogadores, comentários em redes sociais e vídeos em plataformas de audiovisual. As opiniões proporcionam uma visão mais autêntica e pessoal que incentiva outras pessoas a consumir determinado produto, no geral, a influência das análises destaca a importância da transparência na indústria de jogos.

A Steam é uma plataforma digital de distribuição de jogos desenvolvida pela Valve Corporation. Ela permite aos usuários comprar, baixar e jogar uma ampla vari-

OS PRINCIPAIS MERCADOS DE GAME NO MUNDO

Faturamento anual (em US\$ bilhões)

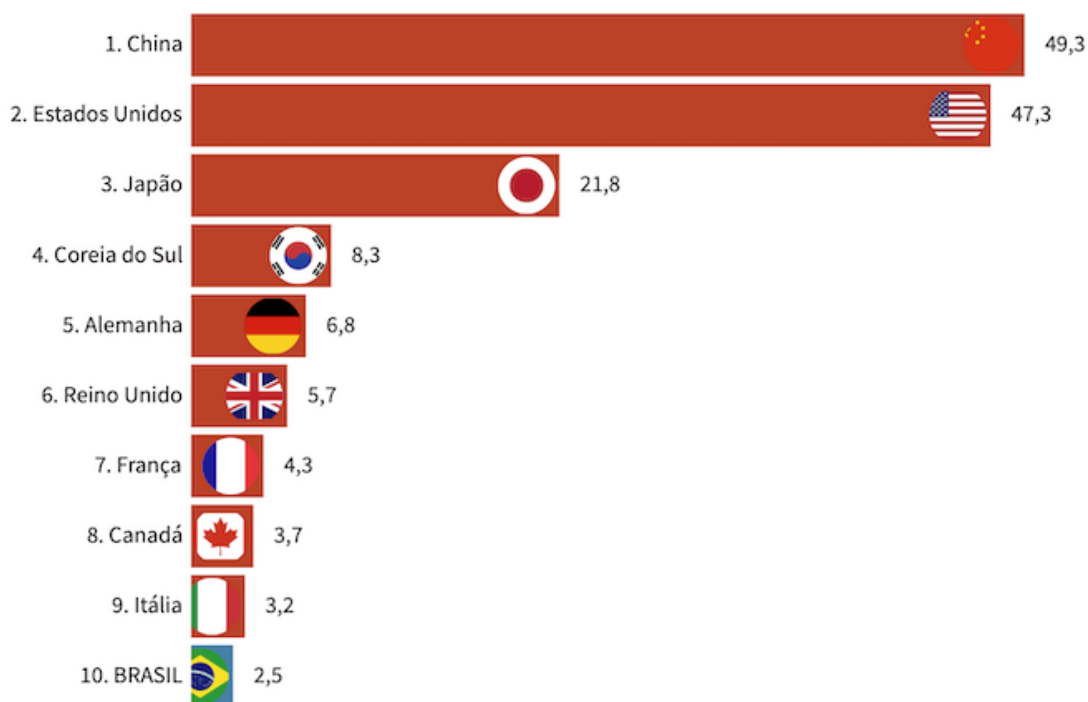


Figura 2 – Faturamento anual global do mercado de games em 2021.

Fonte: Newzoo, The Games Market and Beyond in 2021.

idade de jogos, oferecendo também recursos como multiplayer online, streaming e comunidades. O serviço conta com um sistema de reviews dos jogadores, permitindo que os usuários publiquem análises e avaliações sobre os jogos que compraram e jogaram. Essas informações são disponíveis publicamente para outros usuários acessarem, incluindo a opinião escrita sobre o jogo, o tempo que o avaliador passou jogando e se ele recomenda ou não o produto.

2.2 Análise de Sentimento

A Análise de Sentimento, também denominada como mineração de opinião, é uma área de estudo dedicada à análise das opiniões, sentimentos, avaliações e atitudes expressas pelas pessoas em relação a um produto, serviço, indivíduo ou entidades. Utilizando como base as postagens e informações compartilhadas pelos usuários em redes sociais, fóruns e outras plataformas online, a análise de sentimento emprega técnicas de processamento de linguagem natural e aprendizado de máquina para extrair informações valiosas sobre a percepção pública e as tendências das pessoas a aquele produto (KARAMIBEKR; GHORBANI, 2012).

Com o objetivo de externalizar as suas emoções a respeito de algo, atualmente

as pessoas designam o seu tempo para contextualizar seus sentimentos em plataformas digitais. A opinião tem grande poder sobre o mercado, é a partir desses textos que outras pessoas podem se influenciar e tirar conclusões sobre o produto/serviço sem consumi-lo. Tomando como exemplo o lançamento de um novo smartphone, é comum encontrar uma variedade de feedbacks, alguns usuários exaltam as inovações presentes, enquanto outros apontam críticas e sugestões de melhoria. Diante desse contexto, a análise de sentimentos desempenha o papel de extrair e compreender essas informações, onde as empresas vão utilizar no momento de inovação e adaptação dos seus produtos às necessidades do mercado, em relação as novas demandas de smartphones.

A análise de sentimento é um campo de estudo em diversas áreas, mas nessa pesquisa o foco é examinar o conteúdo emocional em comentários digitais. Na computação esse processo de extração de informações subjetivas e opiniões expressas em textos para compreensão da emoção compartilhada e estruturadas, podem gerar modelos de análise automática. (PANG; LEE et al., 2008).

A complexidade da emoção humana traz consigo uma variedade de níveis de expressão. Embora palavras positivas como “ótimo”, “bom” e “ideal”, possam ser interpretadas como indicadores de satisfação por parte do autor, a construção do contexto, ou seja, a frase completa, ganha uma relevância crucial para o sentido final. Ignorar essa contextualização pode nos levar a interpretações equivocadas, como ironia, comparações e outras nuances. Há situações de interpretação neutra também, como “O filme foi lançado em 2011”, essa frase não expressa sentimento, indica somente uma afirmação por parte do escritor.

Em conclusão, a análise de sentimento emerge como uma ferramenta crucial na era atual, por desempenhar um papel fundamental em diversos setores, desde a indústria de jogos até a tomada de decisões empresariais. A capacidade de compreender e interpretar as emoções expressas em dados textuais fornece informações valiosas sobre a percepção do produto, expandindo assim as expectativas futuras do negócio.

2.3 Processamento de Linguagem Natural

O processamento de linguagem natural é essencial na área da ciência da computação que estuda o desenvolvimento de programas que analisam, reconhecem, classificam e interpretam textos em linguagem humana, uma vertente da inteligência artificial que se aproxima do nosso estilo de comunicação (VIEIRA; LOPES, 2010).

O objetivo principal é permitir que os computadores compreendam e gerem linguagem de forma semelhante à comunicação humana. Esse campo abrange uma ampla variedade de tarefas, desde o reconhecimento de fala/sentimento até a tradu-

ção automática, passando pela extração de informações e geração de textos próprios. Todas essas capacidades são possíveis com o processamento.

Neste trabalho, o PLN foi usado para processar os dados oriundos da plataforma Steam, dados esses como feedbacks escritos por usuários a respeito de jogos específicos em linguagem natural. Com o objetivo de classificar determinado sentimento e características do jogo como o gênero, a partir somente do texto opinativo. Houve a necessidade de processá-los para que algoritmos de aprendizado de máquina e bibliotecas pudessem trabalhar com eles.

2.3.1 Pré-processamento

O pré-processamento é uma das etapas essencial na análise de sentimento, para converter os textos crus em estruturas interpretáveis por um algoritmo (BARBOSA et al., 2017). Nesta fase tendemos a melhorar a performance destes, removendo informações irrelevantes para a tarefa de classificação textual e aplicando técnicas em conjunto nesses dados. Vamos explorar alguns dos principais aspectos do pré-processamento textual realizados no trabalho.

Como primeira etapa, a conversão para caixa baixa é realizada, esse processo refere-se à transformação de todas as letras maiúsculas em minúsculas. Isso é feito para garantir consistência no tratamento das palavras, evitando que o modelo interprete palavras diferentes só por conta das letras maiúsculas e minúsculas presentes no texto. Nesse processamento também é aplicado a remoção de caracteres especiais e pontuação, eliminar símbolos que não contribuem diretamente para a análise semântica do texto, como apresenta a tabela 1, simplificando o texto e mantendo apenas as palavras.

Antes	WoW! The Game really caught vulnerable nostalgia and beauty.
Depois	wow the game really caught vulnerable nostalgia and beauty

Tabela 1 – Exemplo gerado dos dados da Steam ao executar a função de remoção e formatação.

Na segunda etapa focamos na remoção das stopwords, que são grupo de palavras cuja o uso é constante nas produções textuais, porém não dependem de um tópico e não carregam informação, como pronomes, artigos e conjunções (UYSAL; GUNAL, 2014). A tabela 2 demonstra a remoção de algumas dessas palavras.

Nas construções de códigos é possível criar sua própria lista ou encontrar pré-determinada as stopwords. A biblioteca NLTK (Natural Language Toolkit) (BIRD; KLEIN; LOPER, 2009), é construída em linguagem Python e projetada para facilitar o trabalho com dados de linguagem natural. O próprio NLTK fornece uma lista pré-definida de stop words para várias línguas, incluindo o inglês utilizado nessa pesquisa.

Antes	wow the game really caught vulnerable nostalgia and beauty
Depois	game really caught vulnerable nostalgia beauty

Tabela 2 – Exemplo gerado dos dados da Steam ao executar a bibliotexa NLTK.

Dessa forma temos algumas vantagens para interpretação dessas informações. Temos a redução de ruído, as *Stopwords* como "and", "of", "or", muitas vezes aparecem com frequência em textos, mas têm pouco valor semântico, removê-las ajuda a reduzir o ruído e focar nas palavras-chave mais importantes. Temos a economia de espaço, com essa remoção ocorre a diminuição dos conjuntos de dados, economizando armazenamento, facilitando a coleta e destacando as palavras que agregam.

A lematização é a terceira etapa do pré-processamento e pode ser definida como, um modo de agrupamento padrão das diversas variantes de um mesmo signo, com a finalidade de simplificar e apresentar de modo facilitado à consulta dos extratos lexicais em geral (COSTE; GALLISON, 1983).

O processo consiste em reduzir uma palavra para a sua forma mais básica, onde é preservado o sentido semântico, porém é reduzido a uma unidade significativa livre de plurais, conjugações e outras regras da literatura. Existe também outro processo de simplificação chamado stemming (stemização), que se caracteriza por cortar ou remover afixos, deixando somente a raiz das palavras. A lematização se caracteriza por envolver a análise morfológica das palavras, preservando o sentido e integradida, enquanto o stemming simplifica as palavras cortando afixos. Conseguimos perceber essa diferença com algumas palavras usadas pela reviews e executando o processo de stemming e lematização na tabela 3.

Palavra	Stemização	Lematização
Computers	Comput	Computer
Studies	Studi	Study
Changes	Chang	Change

Tabela 3 – Exemplos das diferenças entre e lematização e stemização.

Para construir um modelo computacional é necessário processar o texto para uma forma que seja capaz da máquina interpretar, a conversão de dados textuais para numéricos é uma adaptação ao modelo para desempenhar a tarefa de classificação. A última etapa do pré-processamento consiste na vetorização dos textos, esses vetores numéricos quanto maior a dimensão, maior o poder computacional necessário para processá-lo.

Existem várias formas de codificar um texto, neste trabalho iremos focar na técnica conhecida como *Term frequency–Inverse Document Frequency* (TF-IDF), por ser o modo mais utilizado nas modelagens com literatura. (JONES, 1972)

O TF-IDF é uma medida estatística para indicar a importância de uma palavra em um documento. Um dos principais usos do TF-IDF na área de modelagem é para seleção de parâmetros, neste trabalho usaremos com o mesmo sentido, para buscar as palavras com maior importância para cada texto e montar uma matriz. A construção dessa matriz, consiste em cada linha e coluna correspondem a um documento específico, e os valores são as pontuações TF-IDF que refletem a importância das palavras. O cálculo do TF (*Term frequency*) calcula a proporção do número de vezes que um termo específico aparece no documento em relação ao número total de termos.

$$\text{TF}(t, D) = \frac{f_{t,D}}{n_D} \quad (2.1)$$

- $f_{t,D}$ é o número de vezes que o termo t aparece no documento D ,
- n_D é o número total de termos no documento D .

O IDF (*Inverse Document Frequency*) é uma medida que indica a raridade de um termo em um conjunto de documentos.

$$\text{IDF}(t, D) = \log \left(\frac{N}{|\{d \in D : t \in d\}|} \right) \quad (2.2)$$

Onde:

- t é o termo que está sendo analisado,
- D é o conjunto de documentos,
- N é o número total de documentos no corpus,
- $|\{d \in D : t \in d\}|$ é o número de documentos que contêm o termo t .

$$\text{TF-IDF} = \text{TF} \times \text{IDF}$$

Para fins demonstrativos, vamos supor o seguinte conjunto de frases:

- Eu amo gato
- Eu adoro cachorro

Após a execução do TF-IDF a codificação identifica que as palavras (“gato”, “cachorro”, “amo”, “adoro”) são importantes para as classes de seus textos, pois são únicas em suas respectivas frases.

2.4 Aprendizado de máquina

Aprendizado de máquina, ou *machine learning*, é classificada como uma área da inteligência artificial que fornece ao computador a habilidade de aprender uma determinada tarefa sem ser explicitamente programada (HOMEM; UFES, 2020). O objetivo principal do ML é permitir que os sistemas façam previsões ou tomem decisões com base em padrões identificados nos dados.

Usando essas técnicas é possível construir modelos de aprendizagem, que normalmente são classificadores, regressores, algoritmos de recomendação ou outro tipo de sistema inteligente com a máquina. Um caso de sucesso é o sistema de recomendação da Netflix, empresa de entretenimento que oferece um catálogo de streaming de vídeo sob demanda, onde o algoritmo com os dados do usuário identifica os gêneros que ele tem interesse com base em suas interações anteriores, como filmes assistidos, avaliações dadas e padrões de visualização. O algoritmo pode sugerir filmes e séries que têm maior probabilidade de agradar ao usuário após essa análise, isso é feito com a construção de um modelo de machine learning específico para a área de recomendação ao cliente (CORREIA, 2021).

Existem três tipos principais de aprendizado na área de ML, o supervisionado, não supervisionado e o por reforço. O aprendizado supervisionado se caracteriza por envolver dados rotulados, ou seja, é fornecido dados a entrada e saída esperada ao modelo, durante o treinamento é utilizado essas informações para o aprendizado e dessa forma o modelo mapeia corretamente as entradas para as saídas esperadas. O aprendizado não supervisionado, significa que o modelo é treinado para encontrar padrões nos dados sem ter acesso a informações sobre as saídas desejadas. O aprendizado por reforço é um processo em que um agente interage com um ambiente dinâmico e aprende a tomar decisões sequenciais para maximizar uma recompensa, o agente não é explicitamente instruído sobre como realizar uma tarefa, mas é recompensado ou penalizado com base em suas ações.

A construção de um modelo ML também tem certos desafios em relação ao consumo de dados, existem fenômenos que podem afetar significativamente a eficiência do algoritmo. O overfitting é uma preocupação central em ML, ocorre quando um modelo se adapta excessivamente aos dados de treinamento, capturando o ruído ou características específicas desse conjunto de dados que não são gerais. Também tem casos onde a quantidade de dados rotulados é muito desbalanceada, nos quais uma ou mais classes são significativamente mais representadas do que outras, nesse caso é aplicado o balanceamento de dados para capacitar o modelo a reconhecer as duas classes de forma igual.

2.5 Classificadores

Os classificadores desempenham um papel fundamental no processo de construção de um modelo, classificar é uma forma de atribuir uma categoria ou rótulo a um item com base em suas características. Em termos de aprendizado de máquina, a classificação é uma tarefa de aprendizado supervisionado na qual um algoritmo é treinado para rotular.

O aprendizado supervisionado é o principal meio desse artigo para classificação do sentimento. Os algoritmos utilizados foram Random Forest e SVM, além de usar uma metodologia diferente para comparação, o Vader (Valence Aware Dictionary and Sentiment Reasoner), que é um dicionário de palavras previamente rotuladas com polaridades, juntamente com regras heurísticas para classificar o sentimento do texto.

2.5.1 Random Forest

O conceito do algoritmo de Random Forest, deriva da técnica de árvore de decisão. Elas são uma representação de suporte a decisões com base nas condições sequenciais. É praticamente uma estrutura semelhante a uma árvore, organizada de forma hierárquica na tomada de decisões para chegar ao resultado. O RF é um método de aprendizado de máquina responsável por executar atividades de classificação e regressão. O seu funcionamento consiste em construir uma série de árvores de decisão durante o treinamento do algoritmo de forma aleatória (HO, 1995). Após a construção de todas as árvores na floresta, o algoritmo faz previsões combinando os resultados de cada árvore. No caso da classificação, a classe final é determinada por votação majoritária, enquanto na regressão, é calculada a média das previsões de todas as árvores.



Figura 3 – Ilustração do funcionamento da árvore de decisão.

Fonte: Didática-tech, Como funciona o algoritmo Árvore de Decisão, 2021.

Com uma base de dados adequada o algoritmo consegue chegar a uma conclusão de forma eficaz, ele começa a perceber padrões de comportamento dos dados para cada resultado conforme a análise processa. Essa abordagem é fundamentada no conceito de aprendizado em conjunto. Na figura 3 temos o seguinte comportamento, se o dia for (Sol: Não, Vento: Sim) as respostas remetem a uma negativa de não ir para a praia, se estiver Sol e Vento a decisão permanece a de não ir a praia, enquanto que sem o vento e com sol a ideia é ir a praia. O algoritmo percebe esse padrão e melhora suas decisões, é dessa forma que funciona o comitê onde cada árvore de decisão classifica individualmente e vota em uma classe, a classe com mais votos se torna a previsão do nosso modelo.

Durante a construção de cada árvore, uma amostra aleatória do conjunto de treinamento é selecionada, para evitar a ocorrência de overfitting. Por ser uma técnica que lida facilmente em interpretabilidade de multiparâmetros e manejo de grandes quantidades de dados, o random forest é um dos algoritmos escolhidos para experimentação do modelo.

2.5.2 Support Vector Machine (SVM)

O Support Vector Machine (SVM), ou Máquina de Vetores de Suporte, é um algoritmo de aprendizado de máquina, que tem como objetivo encontrar um hiperplano em um espaço N-dimensional, sendo N o número de características, que classifica distintamente os pontos de dados (BOSER; GUYON; VAPNIK, 1992).

Uma das características distintivas do SVM é a sua abordagem geométrica para encontrar a melhor separação entre as classes de dados. O algoritmo busca identificar um hiperplano de separação que maximize a margem entre os pontos de dados mais próximos de cada classe, chamados de vetores de suporte. Esta maximização da margem ajuda a garantir uma boa generalização do modelo para novos dados e ajuda a reduzir o risco de sobreajuste.

O SVM é capaz de lidar com problemas de classificação não lineares através do uso de funções de kernel, permitindo que seja aplicado em uma ampla variedade de cenários, mesmo quando os dados não são linearmente separáveis. O seu poder computacional é bastante adaptável.

O algoritmo demonstra melhor desempenho em cenários onde a distinção entre classes é clara, evidenciada pela sua função de perda. Adicionalmente, o modelo tende a performar melhor quando a dimensionalidade dos dados é superior ao número de amostras. Apesar desses benefícios, o SVM pode apresentar algumas limitações, como sua ineficácia em lidar com conjuntos de dados volumosos e sua baixa performance quando há presença significativa de ruídos.

No cenário da Figura 4, é possível observar todos os três hiperplanos (A, B e C), estão definindo a classificação corretamente. Porém, A e B têm uma margem de erro muito menor, ou seja, estão muito próximos dos pontos de dados. Se houver alguma flutuação nos valores dos pontos de dados, eles poderão cair em erros. O hiperplano C, por outro lado, possui uma margem significativa para erros, pois possui a distância máxima dos pontos de dados de ambas as classes.

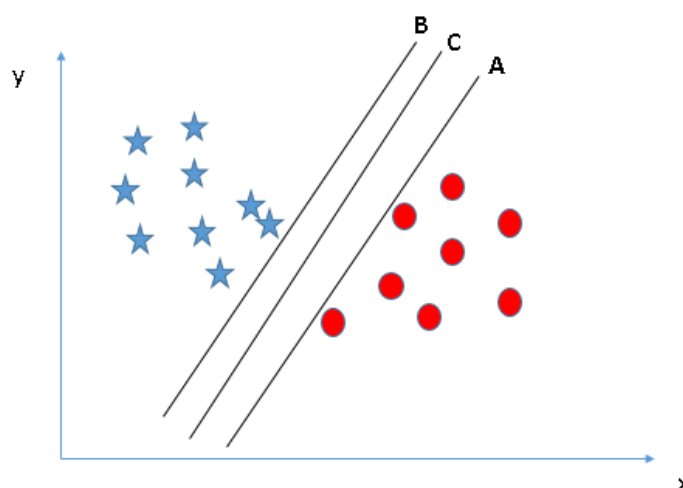


Figura 4 – Comportamento dos hiperplanos nas margens.
Fonte: UFPO, Entendendo Sua Matemática, 2016.

2.5.3 VADER

Valence Aware Dictionary and Sentiment Reasoner (VADER), é uma ferramenta construída com o intuito de definir uma polaridade em relação ao sentimento de um texto, a utilização é gratuita e conta com suporte para 8 das principais linguagens de programação (HUTTO; GILBERT, 2014).

O objetivo é lidar com a natureza complexa e variada da linguagem humana, incluindo gírias, negações, sarcasmo e outras nuances. Ele usa um dicionário de palavras previamente rotuladas com polaridades, juntamente com regras heurísticas para avaliar a polaridade do texto. A principal vantagem do VADER é sua capacidade de lidar com textos curtos e informais, comuns em mídias sociais, avaliações de produtos e comunicações online.

Como o VADER utiliza de uma biblioteca a ferramenta conta com algumas limitações. Pode não ser tão preciso em textos longos ou técnicos, e pode falhar em entender sarcasmo ou ironia em alguns casos. Além disso, sua precisão pode variar dependendo do domínio ou contexto específico em que é aplicado.

Após analisar todas as palavras e seus contextos na frase, o VADER calcula uma pontuação de sentimento global para o texto. Neste caso, a frase "I LOVE THIS

GAME!!!”, provavelmente seria classificada como positiva pelo VADER, devido a palavra positiva “love”. A ferramenta apresenta 4 campos de pontuação: positivo, negativo, neutro e o compound. O compound é uma pontuação numérica que varia de -1 a +1, quando o valor está próximo de -1 indica que o texto é negativo, um valor próximo de +1 indica que o texto é positivo, sendo a média que harmoniza as pontuações positivas e negativas de determinado texto.

2.5.4 Métricas de Avaliação

O desempenho para um modelo de aprendizagem de máquina para classificação, define o quão confiável é a tomada de decisões. Seus valores demonstram a qualidade de um modelo, logo, se ocorrer geração de resultados ruins, o modelo está em má condições para atuar (TRINDADE, 2023).

Após o treinamento dos modelos, a segunda fase é a classificação dos sentimentos. Nessa etapa é possível observar a quantidade de previsões corretas e erradas através de uma matriz de confusão. A matriz de confusão é uma tabela que compara as previsões do modelo com os valores reais (MORETTIN; SINGER, 2022). Essa matriz é útil para entender a capacidade do modelo e analisar o nível de acerto e erros, a partir desses valores que geramos as métricas, para facilitar temos a ilustração da figura 5 mostrando os componentes de uma matriz de confusão.

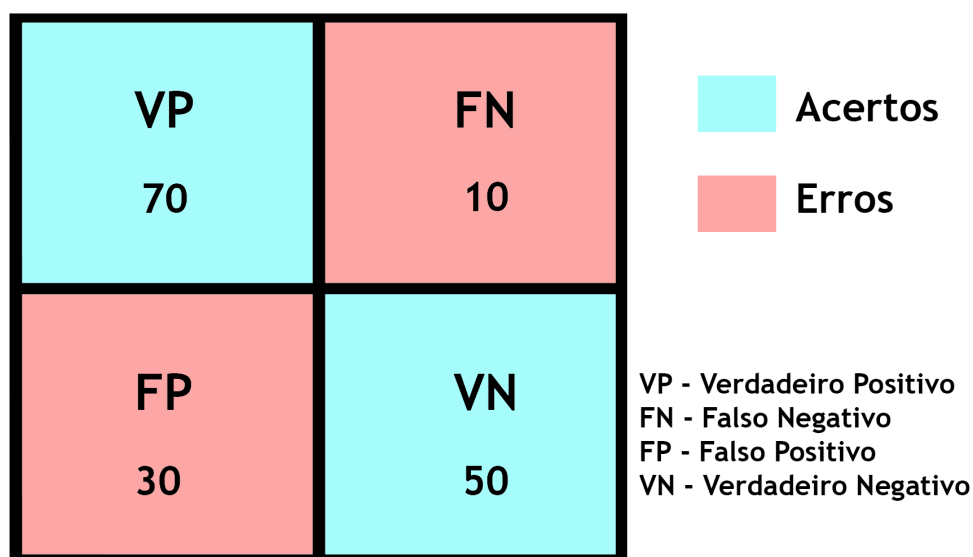


Figura 5 – Ilustração da matriz de confusão.
Fonte: Própria Autora.

Essa matriz é útil para entender a capacidade do modelo e analisar o nível de acerto e erros, a partir desses valores que geramos as métricas de desempenho.

Com o modelo construído e classificando, começamos a etapa de estudo das métricas de validação. Coletando as informações obtidas pela matriz de confusão, conseguimos obter medidas como a acurácia, que fornece a taxa de acerto do nosso sistema.

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

Em quantidades desbalanceadas de amostras a acurácia não é o método mais aconselhado a se seguir, por conta disso a precisão, recall e F1-Score é levada em conta. A precisão mede a proporção de instâncias classificadas corretamente como positivas em relação ao total de instâncias classificadas como positivas.

$$\text{Precisão} = \frac{VP}{VP + FP}$$

O uso das métricas Recall e F1-score é especialmente em situações em que há desequilíbrio entre as classes ou quando é importante equilibrar a precisão e a capacidade de encontrar exemplos positivos. O recall é uma medida em que todos os verdadeiros positivos sejam identificados, principalmente em situações em que identificar falsos negativos é muito prejudicial, em resumo, o recall é diferente da precisão por focar em minimizar os falsos negativos e garantindo que o modelo identifique corretamente o maior número possível de exemplos.

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

O F1-score é uma média harmônica que combina precisão e recall em um único valor. Ele é útil quando há um desequilíbrio entre as classes no conjunto de dados, ou seja, quando uma classe é muito mais prevalente do que a outra.

$$\text{F1-score} = 2 \times \frac{\text{Precisão} \times \text{Recall}}{\text{Precisão} + \text{Recall}}$$

Como as classificações de sentimento nesse cenário são binárias e multiclases, existe uma métrica útil chamada de curva característica de operação do receptor, ou curva ROC. A curva ROC é uma representação gráfica da relação entre duas taxas, taxa de verdadeiros positivos e taxa de falsos positivos. Para medir os valores dessa curva usamos o a área sob a curva ROC-AUC, é uma medida numérica da qualidade do seu modelo, quanto maior a AUC, melhor é o desempenho do modelo em discriminar entre as classes positiva e negativa. A Figura 6 demonstra um exemplo de um bom comportamento do modelo com o valor alto de AUC.

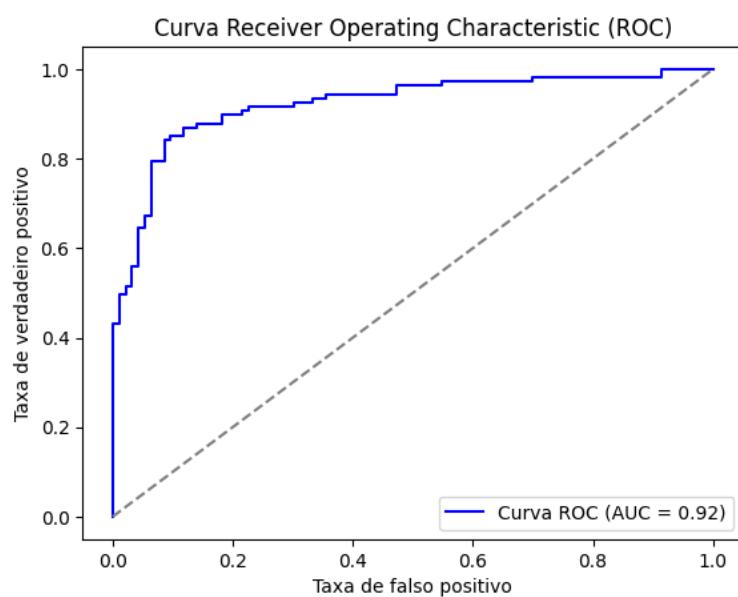


Figura 6 – Ilustração gráfica da curva ROC-AUC.
Fonte: Própria Autora.

3 Trabalhos Relacionados

Este capítulo apresenta um compilado de trabalhos relacionados às abordagens de análise de sentimento utilizando algoritmos. O intuito é estabelecer uma base sobre o assunto, e mesclar metodologias para a construção do modelo.

Nos trabalhos de Pang e Lee (PANG; LEE et al., 2008), temos a metodologia que envolve como realizar a mineração de opinião e a partir desses dados analisar o sentimento de diferentes domínios. O artigo retrata a importância da análise de sentimento para diferentes nichos, como críticas de filmes, resenhas de produtos, discussões políticas e busca compreender as nuances que o ser humano representa no momento em que textualiza a sua ideia. Seus métodos de análises foram construídos com modelos supervisionados como o Naive Bayes, SVM e MaxEnt foram utilizadas para a análise binária de sentimentos, alcançando precisões de 81% a 84% sendo o SVM o mais eficiente para os cenários abordados. O trabalho atual, no entanto, foca na análise de opiniões em jogos, utilizando o SVM e Random Forest como modelos supervisionados e o VADER um método de classificação pré-estabelecida para análise de sentimento.

No artigo de Ravi (RAVI; RAVI, 2015), entramos na área da análise de sentimento com foco no meio computacional. A pesquisa reuniu diversos artigos de todo o mundo entre os anos de 2002 e 2015, nesses estudos a aplicação dos algoritmos é preditiva em relação ao sentimento expresso, é possível estatisticamente gerar resultados sobre os métodos mais utilizados e como são aplicados. Dessa forma, é explorado as diferentes abordagens para encontrar os sentimentos do usuário, sendo mostrado desde bibliotecas léxicas para predição até algoritmos de aprendizado de máquina. É possível observar a evolução das bibliotecas em relação a acurácia de acertos, chegando a resultados superiores a 80%, em relação a exploração supervisionada por aprendizagem de máquina temos a demonstração dos mais utilizados para captação de sentimento que são o SVM, Naive Bayes, Radom Forest e regressão linear.

As abordagens computacionais usadas para explorar essa área da classificação sentimental são amplas, O autor Lucas Avanço (AVANÇO, 2015) avalia três tipos de técnicas para fundamentações textuais a baseada em léxicos, baseada em algoritmos de aprendizado de máquina e uma abordagem híbrida. Essa comparação de avaliações utilizando base de dados do Buscapé e Mercado Livre em relação a produtos vendidos dentro da plataforma, constrói uma estrutura linear de experimentação e demonstra os resultados utilizando cada tipo de classificação, o autor destaca os algoritmos com fluxo de árvores ou vetores na etapa de modelagem. O Random Forest,

Naive Bayes e o SVM apresentam métricas de acurácia e f1-score acima da média nas avaliações do buscapé e do mercado livre.

As abordagens de Mardjo e Choksuchat ([MARDJO; CHOKSUCHAT, 2022](#)) propuseram uma temática mais exclusiva do uso do processamento de linguagem natural para análise de sentimento. O objetivo é a construção de um modelo mais induzido ao tema proposto, aprofundando suas técnicas de treinamento para criação de algo exclusivo para área de moeda virtual. O foco em questão são os Bitcoins, e os dados utilizados para gerar a inteligência do modelo são das redes sociais, especificamente o Twitter (atual X). O estudo investe na interpretação híbrida, usando elementos do dicionário léxico para rotulações das opiniões, e das técnicas supervisionadas para qualificar corretamente os sentimentos com base na construção anterior. Esse modelo alcançou valores acima de 75% em de precisão, acurácia e F1-score utilizando o algoritmo de RF, enquanto no Recall fica abaixo somente do modelo SVM e Decision Tree com uma diferença de 10% em cima da base de dados escalada, essa atitude de especificar o sua predição para somente um assunto auxilia na cooperação de abordagens mistas.

No Brasil, a pesquisa de Moreira ([MOREIRA, 2021](#)) realizou uma comparação entre três tipos distintos de codificação de palavras em opiniões críticas relacionadas a filmes retiradas da IMDB (Internet Movie Database), depois aplicando essa codificação em classificadores mistos. Os classificadores recebem as representações numéricas dos textos e treinam usando uma validação cruzada, sendo utilizado o Random Forest, SVM e a biblioteca Wizard. Os resultados demonstrou que a análise de sentimento dependendo da combinação, codificador e classificador, tem ganhos significativos nas métricas e tempo de execução. Visto isso os melhores resultados abrangem os classificadores com algoritmos de aprendizagem de máquina, especificamente a abordagem com árvores aleatórias foi superior pelo suporte a grandes valores de dados e relação rápida a codificadores de binarização das informações, porém é demonstrado que bibliotecas léxicas tem um ganho de tempo maior e performam com qualquer tipo de codificador.

Estudos recentes como o apresentado no artigo de Krugmann e Hartmann ([KRUGMANN; HARTMANN, 2024](#)), demonstram a evolução das *Large Language Models*, modelos de IA treinados com uma grande quantidade de texto para compreender padrões complexos da literatura, assim gerando e processando novos textos. Neste estudo é utilizado alguns LLMs atuais no mercado, como o ChatGPT, GPT-3.5, GPT-4 e Llama 2 para a análise de sentimentos. O cenário do estudo analisa a classificação binária e multiclases, os dados são retirados de diversas plataformas (Twitter, Amazon, IMDb) com opiniões de assuntos divergentes como filmes, produtos, política e música. A principal métrica de avaliação escolhida para comparar com os modelos de aprendizado

de máquina foi a acurácia. Os LLMs alcançaram acurácias consistentemente acima de 80%, apresentando um bom desempenho diante de diferentes assuntos, destacando o Llama 2 que chegou a um valor de 93% nas opiniões relacionadas a filmes mesmo com bons resultados os modelos de aprendizagem de máquina ficaram com valores maiores em todas as outras categorias. Esse uso das LLMs abre novas perspectivas para a pesquisas na parte de sentimentos, levando a uma constante evolução dos modelos para classificações nessa área.

Apresentando contribuições à literatura existente sobre análise de sentimento, observa-se uma tendência de comparar o desempenho de diversos métodos avaliativos, visando identificar a ferramenta mais eficaz. Essa abordagem se reflete no tipo de classificador, codificador ou base de dados adotada. Em contraste com os estudos anteriores, este trabalho se destaca com o objetivo de construir um modelo de análise de sentimento intrinsecamente relacionado as categorias com as respectivas opiniões e analisando a melhor forma de classificar.

4 Ferramentas e Métodos

Este capítulo apresenta os métodos propostos para concretizar a construção dos modelos para classificação de sentimento em textos opinativos para a área de jogos digitais. O intuito é construir modelos mais eficazes para prever o comportamento sentimental dos usuários desse ramo.

As ferramentas utilizadas empregam um processo onde conseguimos coletar as informações de opinião a respeito do jogo pela plataforma Steam, e usamos os dados disponibilizados para construir o algoritmo. Para entendermos o processo melhor, dividimos o desenvolvimento em etapas na Figura 7.

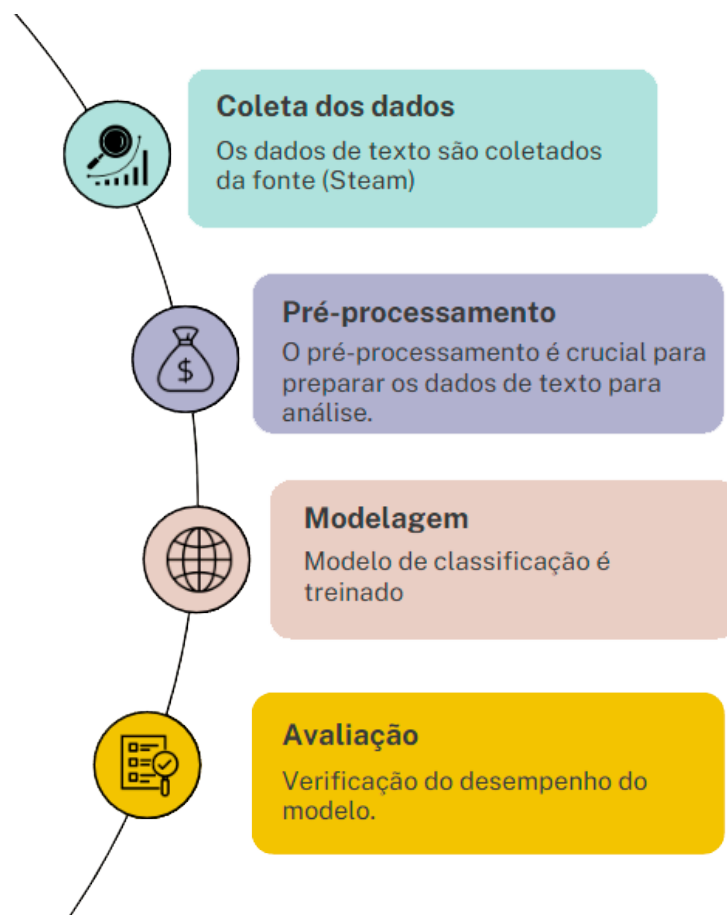


Figura 7 – Etapas de desenvolvimento do modelo.
Fonte: Próprio da autora.

A proposta é seguir esse esquema de desenvolvimento de modelo em todos os algoritmos, e a partir dessa lógica montar os distintos cenários de experimentação. O objetivo é explorar as maneiras diferentes de classificar os textos no ambiente específico de jogos. Como a linguagem nesse caso é repleta de expressões específicas dependendo da categoria que está inserido, por exemplo, se o usuário tem preferência

por temas de terror sua avaliação pode ser repleta de palavra com conotação negativa como “assustador” ou “medo” que podem significar algo positivo em jogos de terror e horror, mas em outro gênero não.

Essas nuances classificatórias podem afetar o resultado do tipo de sentimento, por isso levamos em consideração todos os fatores como métricas, trabalho computacional, abrangência dos dados e métricas geradas do resultado do modelo.

4.1 Coleta dos Dados

Quando falamos do tópico jogo virtual, precisamos entender que este jogo precisa está em algum lugar para que o jogador consiga adquiri-lo, e atualmente existem múltiplas plataformas online onde é feita a comercialização deste produto. Uma das plataformas mais conhecidas mundialmente é a Steam, que conta com 10 milhões de usuários ativos em computadores e consoles.

A Steam conta com um sistema de avaliação bem robusto, onde os usuários ganham pontos por expor sua opinião a respeito do produto, porém a avaliação só pode ser feita se você obteve o jogo e jogou. Além de selecionar se recomenda ou não, o jogador escreve um comentário dizendo o que achou e esse avaliação é exibida na página principal do jogo, como demonstra Figura 8:



Figura 8 – Comentário avaliativo retirado da Steam.

Fonte: Steam.

Todas essas informações são disponibilizadas pela Steam em formato de API (Application Programming Interface), um serviço disponibilizado para acessar os recursos do sistema usando apenas requisições HTTP, encontradas no site oficial da empresa Steam Web API. Através desse site que coletamos as features que vamos utilizar: Categoria do jogo, se o usuário recomendou ou não e o comentário opinativo.

As *reviews* dos jogos são textos escritos pelos usuários que expressam suas opiniões sobre os jogos. Essas opiniões podem incluir avaliações de aspectos especí-

ficos, como jogabilidade, gráficos, história, e até mesmo críticas sobre bugs ou problemas técnicos. É o principal conteúdo que torna possível a classificação do sentimento do jogo e de sua categoria.

Neste trabalho, foram coletados mais de 14 mil *reviews* de jogos, abrangendo um total de 64 títulos distintos. Todos os *reviews* foram fornecidos por diferentes usuários e englobam uma variedade de estilos.

4.2 Pré-processamento

Como campo de desenvolvimento de todas as atividades envolvendo os dados e construção do modelo, utilizamos o *Google Colab*¹, plataforma gratuita oferecida pelo Google que permite a execução de código Python² (linguagem de programação) de forma colaborativa na nuvem. Então no ambiente, incluímos a leitura dos dados utilizando a biblioteca *Pandas*³ que disponibiliza de diversos recursos para construir o dataset elegível para receber mudanças em Python.

Depois da aquisição dos dados vem a fase do pré-processamento, consiste em passar por funções de manipulação que deixam a base de dados pronta para a fase de testes. O pré-processamento foi construído em sua maior parte pela biblioteca *NLTK*⁴ (biblioteca em Python), que fornece uma variedade de módulos e recursos para facilitar tarefas como tokenização, lematização, análise sintática, extração de informações e muito mais na área de textos. Nessa fase executamos os processos de:

- Função de remoção das stopwords;
- Função de remoção de maiúsculas e caracteres especiais;
- Função de lematização;
- Função de vetorização com TF-IDF.

Para a primeira função utilizamos a biblioteca *NLTK* que conta com sua própria lista de palavras em várias línguas. Utilizamos a inglês e conseguimos remover praticamente todas as palavras que não agregam sentido às frases. Na segunda função, percorremos todo o texto e começamos a etapa de padronização. Nessa fase todas as palavras ficam minúsculas, para que o modelo não diferencie palavras iguais. Foram removidos também os caracteres especiais que não acrescentam sentido para aquele texto. Toda essa modificação é utilizando funções nativas do próprio Python.

¹ <https://colab.google/>

² <https://www.python.org/>

³ <https://pypi.org/project/pandas/>

⁴ <https://www.nltk.org/>

No quesito lematização, esse processo linguístico é definido por reduzir as palavras flexionadas, normalmente verbos conjugados e plurais. O objetivo é agrupar palavras variantes para que possam ser tratadas como uma única forma, e agrupando essas palavras podemos simplificar a análise e melhorar a precisão do modelo. Esse processamento foi realizado com a função *WordNetLemmatizer* da biblioteca NLTK.

Por último, temos a vetorização, realizada com as funções *TfidfVectorizer* e *TfidfTransformer*, ambas da biblioteca do scikit-learn popular fornecer funções ligadas a aprendizado de máquina. Essa técnica é utilizada para transformar documentos de texto em representações numéricas, que são mais adequadas para análise, o TF significa frequência do termo enquanto o IDF significa frequência inversa do documento. Essa abordagem é útil para destacar a importância relativa de palavras em relação ao contexto do documento e da coleção total de textos, dessa forma cada termo é transformado em um vetor numérico ligado a sua frequência e dessa jeito é formado uma matriz TF-IDF. Essa matriz é formada pela parte de frequência do termo (TF), onde representa a frequência de um termo específico dentro de um documento e a frequência inversa do documento (IDF), mede a importância de um termo em todo o conjunto de documentos. O resultado é uma pontuação que representa a importância de um termo em um documento específico em relação a todo o conjunto de documentos. Sendo assim a entrada adequada para o algoritmo de classificação que interpretam entradas numéricas. A tabela 4 apresenta uma review retirada da base de dados e passando por todos os processos textuais de pré-processamento.

Antes	The game was really exciting and fun!!! but I found it too difficult to progress.
Depois	game really exciting fun find too difficult progress.

Tabela 4 – Frase original e modificada após o pré-processamento.

Fonte: Steam.

4.3 Modelagem

Para iniciar a classificação, começamos a etapa de criação do modelo. Utilizando a biblioteca scikit-learn, citada acima por implementar métodos eficientes de aprendizado de máquina em Python, importamos as classes que envolvem os algoritmos de random forest, SVM e a biblioteca VADER.

O *RandomForestClassifier* é a função que implementa a interface da lógica de classificação e com isso construímos o modelo, nessa função conseguimos personalizar alguns parâmetros como: número de árvores geradas, a profundidade de cada árvore e o número de amostras. Para definir os melhores valores usamos um processo de hiper parametrização para ajustar os hiperparâmetros de um modelo de aprendizado de máquina para otimizar seu desempenho, realizamos esse processo com a função

GridSearchCV também do *scikit-learn* que facilita a busca por hiperparâmetros ideais passando os dados vetorizados das *review*.

O SVC é uma implementação do algoritmo SVM focada na área de classificação, é importado pela biblioteca *sklearn* para utilização no código. Ele oferece suporte a vários parâmetros que permitem ajustar o comportamento do modelo, incluindo o tipo de kernel. O SVC procura encontrar um hiperplano que melhor separa os exemplos de diferentes classes em um espaço de características. Este hiperplano é escolhido de forma a maximizar a margem entre os exemplos mais próximos de cada classe, conhecidos como vetores de suporte.

O *SentimentIntensityAnalyzer* é uma classe da biblioteca NLTK, que implementa o método VADER para análise de sentimentos em textos. Essa classe tem como objetivo avaliar a polaridade dos textos e resultar em um score para cada tipo de sentimento. Utilizando um dicionário de palavras previamente rotuladas com suas polaridades juntamente com regras heurísticas é possível classificar as *reviews*.

Após coletar os melhores parâmetros dos algoritmos, a atividade seguinte envolve separar os testes e treinos. Os conjuntos de treinamento foram coletados do dataset e envolvia a parte dos dados que vai ser usada no modelo, dessa forma o modelo aprende com esses treinos. O conjunto de testes é formado por outra parte do dataset, que não foi usada no modelo, para avaliar o seu desempenho classificatório.

4.4 Avaliação

Para avaliar o desempenho dos modelos propostos, utilizaremos um conjunto de métricas que nos permitirão entender como cada modelo se comporta em relação ao problema abordado. Gerando uma matriz de confusão com o conjunto de teste classificado pelos modelos, separamos os resultados em uma tabela e realizamos o cálculo para encontrar as medidas de acurácia, precisão, recall, F1-score e a AUC-ROC de cada resultado.

A pesquisa tem como objetivo explorar duas alternativas: uma abordagem mais abrangente e geral e outra mais específica e segmentada. Na abordagem geral, todas as *reviews* são consideradas, independentemente do gênero do jogo. Nesse caso, aplica-se a análise de sentimento sobre todas as opiniões disponíveis, buscando identificar tendências gerais e o sentimento predominante em relação aos jogos como um todo. Por outro lado, a abordagem segmentada foca em analisar as opiniões de um gênero específico de jogo. O objetivo é explorar como a análise de sentimento se comporta ao considerar somente as *reviews* de um determinado tipo de jogo, como ação, aventura, ou RPG, por exemplo. Isso permite observar nuances de sentimento relacionadas a características próprias de cada gênero, oferecendo uma visão mais

detalhada e focada sobre a percepção dos usuários nesses diferentes contextos.

Ao analisar essas métricas e abordagens de cada modelo, poderemos identificar qual deles apresenta o melhor desempenho em relação ao nosso problema específico de análise de sentimentos. Um modelo com alta acurácia, precisão, recall, AUC-ROC e F1-score será considerado o mais adequado para nossa aplicação. Em específico um modelo com F1-score maior é melhor para essa análise de sentimento, porque ele é mais eficaz em identificar corretamente as instâncias positivas e a taxa de captura do modelo, sendo uma média harmonica entre a precisão e o recall proporcionando uma visão mais precisa do sentimento geral.

5 Resultados

Neste capítulo, apresentamos os resultados obtidos a partir das análises realizadas no estudo. Nosso objetivo é avaliar o desempenho dos modelos de classificação de sentimento aplicados aos dados coletados e pré-processados, bem como verificar a eficácia das técnicas e metodologias.

5.1 Experimento 1: Análise de sentimento em todas as classes

Nesse caso está incluso todo o volume de dados presente no dataset e não é realizada nenhuma filtragem por categoria, como apresenta a figura 9 dividindo por sentimento.

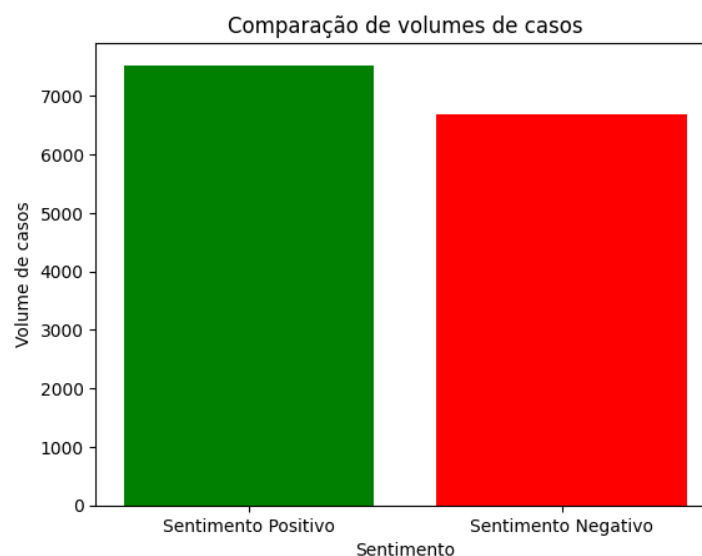


Figura 9 – Quantidade de recomendações positivas e negativas.

Fonte: Própria Autora.

O campo de recomendação serve como o target, variável que representa o rótulo do sentimento. As partes de review, os dados são os parâmetros de treinamento principal, e passam pelo pré-processamento se tornando uma parte da entrada do modelo. Nos campos de recomendação, esses dados são representados de forma binária, afinal o usuário só tem a opção de “Sim, eu recomendo” e “Não, eu não recomendo”. Separamos o conjunto de dados em treinamento e teste, nessa etapa 20% dos dados estão como conjunto de teste para garantir que tenhamos uma quantidade suficiente para treinar o modelo de forma eficaz. O VADER não realiza um treinamento, a ferramenta recebe uma pontuação composta que representa a polaridade geral do sentimento no texto, levando em consideração a construção gramatical.

Para a abordagem com o modelo Random Forest, foram utilizados parâmetros como: **n_estimators=200**, utilizado para definir o número de árvores geradas, nesse caso gerando 200 árvores, **max_depth=None**, utilizado para controlar a profundidade máxima de cada árvore de decisão na floresta, quando definido como None, as árvores são expandidas até que todas as folhas sejam puras ou até que todas as folhas contêm menos do que o número mínimo de amostras. No SVM o **Kernel** escolhido foi o linear, sendo o parâmetro mais adequado para dados rotulados binários, alcançando boas performances no momento da divisão do hiperplano. Todos os hiperparâmetros foram escolhidos a partir de uma análise com a ferramenta Grid Search, que explora exaustivamente um conjunto de hiperparâmetros predefinidos para encontrar a combinação que otimiza o desempenho do modelo.

Modelos	Acurácia	Precisão	Recall	F1-Score	AUC
SVM	0.8122%	0.7993%	0.7868%	0.7917%	0.8664%
Random Forest	0.8318%	0.8182%	0.7206%	0.7672%	0.8480%
VADER	0.6890%	0.6711%	0.4395%	0.5310%	0.7056%

Tabela 5 – Métricas dos algoritmos trabalhados.

Na tabela 5 temos a divisão dos modelos e os valores presentes de cada métrica usada na análise. Com base nas métricas, podemos observar que no quesito desempenho os algoritmos de aprendizado de máquina foram superiores ao uso do VADER. Ambos tiveram resultados satisfatórios, chegando a valores elevados a 80% em algumas medidas, o Random Forest teve uma maior precisão indicando menos falsos positivos, porém o SVM conseguiu um F1-score superior significando que teve uma melhor captura, sendo mais vantajoso para esse problema de encontrar o real sentimento do cliente. Por outro lado, o VADER não alcançou resultados bons, mesmo a base dos textos sendo com linguagem informal, os usuários fornecem explicações detalhadas sobre os motivos de gostarem ou não do conteúdo, o que pode ter dificultado a previsão da polaridade devido ao grande volume de dados disponíveis.

A Figura 10 ilustra como a sensibilidade do modelo varia em relação à sua especificidade, quanto mais perto a linha estiver do canto superior ao lado esquerdo indica altas taxas de verdadeiros positivos insinuando uma boa performance do modelo. O SVM e RF mostraram valores altamente capazes de distinguir entre exemplos positivos e negativos, independentemente do limiar de classificação utilizado. O Vader representou um resultado inferior, porém ainda acima da limiar de 50%, e abaixo dos 80% se encontrando em uma faixa boa de classificação.

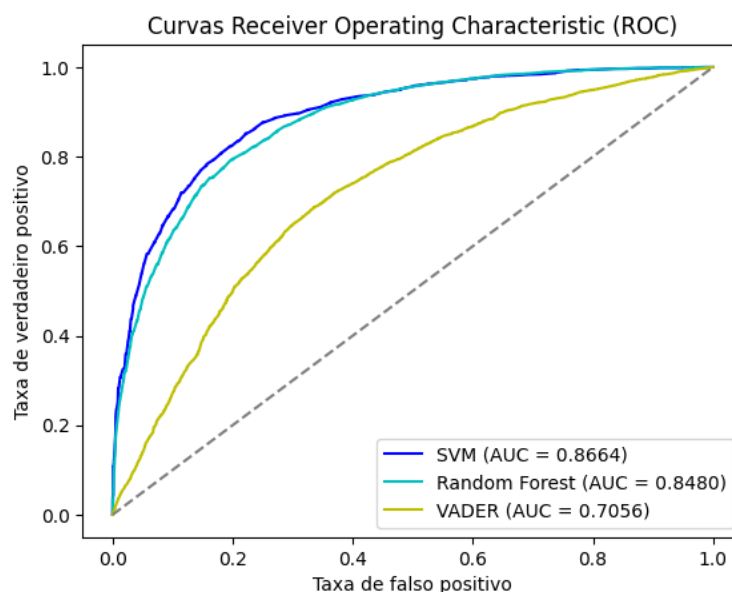


Figura 10 – Ilustração dos valores referentes a curva ROC.

Fonte: Própria Autora.

5.2 Experimento 2: Análise de sentimento aplicada em casos de gênero específico

Neste experimento, nosso foco está em uma abordagem mais específica, onde primeiro buscamos classificar a categoria do jogo e em seguida, nos aprofundamos na análise de sentimento filtrada pelo gênero. Para isso, surge a necessidade de desenvolver abordagens mais eficientes e precisas para classificar jogos em diferentes tipos, a perspectiva adequada é a classificação de multiclasse.

A abordagem escolhida para classificar o gênero são os campos de reviews e tipo principal disponíveis na tabela. A análise dessas reviews permite identificar palavras, expressões e contextos específicos que estão associados a diferentes categorias de jogos. Essas informações de características linguísticas presentes nos textos dos jogadores são os insígnias principais na busca pelo gênero.

O conjunto de dados apresenta os títulos de jogos e seus gêneros, podendo conter até 4 categorias diferentes por jogo. Porém cada título tem o seu gênero principal e são neles que vamos trabalhar o classificador. Para essa operação cada algoritmo usa uma estratégia, o RF utiliza o chamado “One-vs-All”, nesse contexto um conjunto de árvores de decisão é treinado para distinguir uma classe das demais, durante a previsão cada árvore de decisão vota na classe prevista, a classe com mais votos é escolhida. O SVM utiliza algo similar, chamado “One-vs-Rest”, onde o classificador binário é treinado para cada classe, considerando-a como a classe positiva e todas as outras como negativas, durante a previsão o modelo SVM com a maior pontuação de decisão é escolhido como a previsão final.

Nesta operação, não foi viável utilizar o Vader, pois sua capacidade se limita à classificação de sentimentos. Portanto, optamos por utilizar o Vader na segunda etapa do processo, na qual as análises já estarão filtradas de acordo com suas respectivas categorias.

Os gêneros presentes no conjunto de dados foram ação, anime, terror, rpg (jogo de interpretação de personagem), tiro, estratégia e corrida. Totalizando 7 tipos diferentes, todos com o balanceamento dos dados similares, dessa forma é possível compreender mais profundamente os dados e se torna uma escolha poderosa na capacidade de classificar textos. A tabela 6 apresenta todos os resultados presentes por modelo, categoria e métricas calculadas.

Gênero	Modelos	Acurácia	Precisão	Recall	F1-Score	AUC
Terror	SVM	0.7968	0.8400	0.9110	0.8741	0.8709
Ação	SVM	0.9502	0.9630	0.4483	0.6118	0.7206
RPG	SVM	0.9668	0.9570	0.5841	0.6956	0.8203
Tiro	SVM	0.9122	0.9343	0.5768	0.7133	0.8470
Estratégia	SVM	0.8880	0.9010	0.5024	0.6491	0.7144
Anime	SVM	0.8990	0.8712	0.7529	0.8077	0.8142
Corrida	SVM	0.6613	0.8227	0.3951	0.5311	0.5506
Terror	Random Forest	0.6967	0.8071	0.5098	0.6554	0.8460
Ação	Random Forest	0.8909	0.9311	0.5022	0.7390	0.8501
RPG	Random Forest	0.7606	0.9024	0.3938	0.4919	0.7110
Tiro	Random Forest	0.6990	0.9103	0.5058	0.6184	0.7198
Estratégia	Random Forest	0.8350	0.8571	0.5199	0.6900	0.7881
Anime	Random Forest	0.9297	0.9059	0.8229	0.8884	0.8655
Corrida	Random Forest	0.6700	0.8133	0.4598	0.6009	0.6019

Tabela 6 – Métricas dos modelos para identificar a categoria dos jogos.

Com base nas métricas apresentadas, o modelo SVM demonstra um desempenho superior em grande parte das medidas diante do modelo Random Forest. O SVM mostra ter a melhor capacidade na identificação do tipo de jogo em relação as reviews de textos, com sua abordagem de treinar vários modelos binários separados, um para cada classe. O RF encontra boas métricas, porém apresenta valores abaixo da estratégia do SVM. Lidar com uma identificação de multiclases é um desafio para modelos que foram construídos para problemáticas de binários, é importante reconhecer que essas métricas podem ser influenciadas pela distribuição de classe, presença de classes raras e outras complexidades inerentes.

As Figuras 11 e 12 apresentam o comportamento dos gêneros diante do algoritmo de classificação. O SVM tem os tipos terror, tiro, rpg e anime acima de 80%, foram os melhores para distinguir entre exemplos positivos e negativos, enquanto no RF foram os tipos ação, terror e anime.

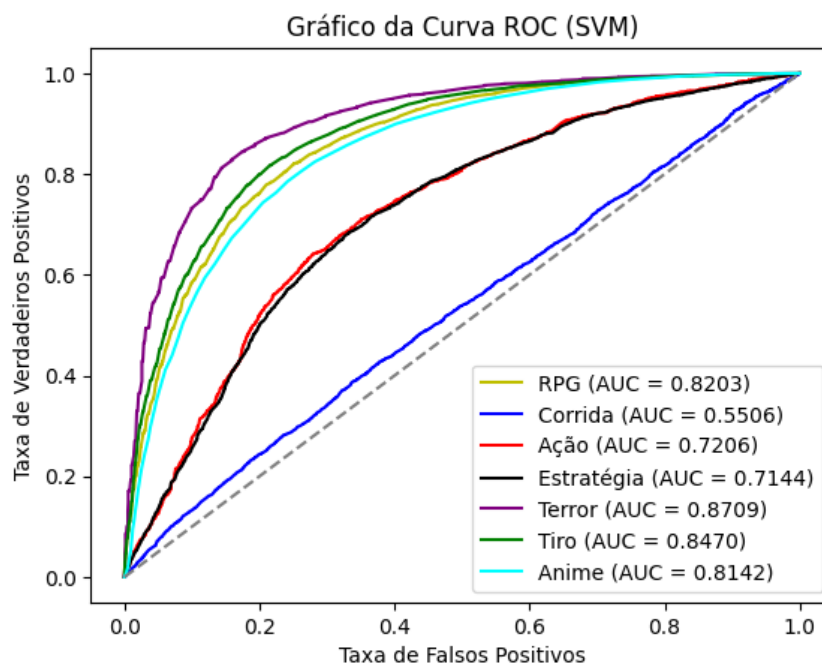


Figura 11 – Métricas referentes a curva ROC por gênero usando SVM.
Fonte: Própria Autora.

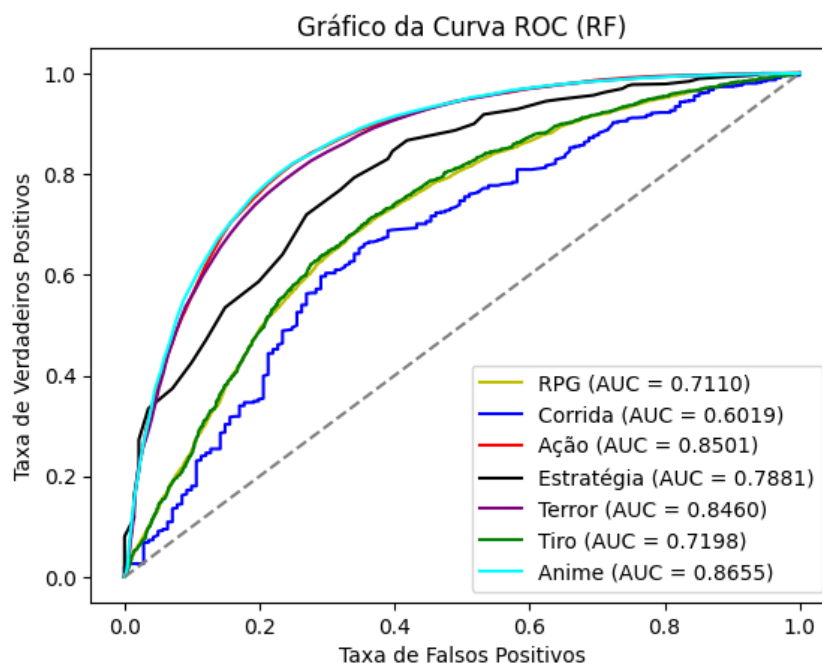


Figura 12 – Métricas valores referentes a curva ROC por gênero usando RF.
Fonte: Própria Autora.

A segunda etapa da experimentação é filtrar as reviews pela sua categoria de acordo com o classificador anterior. Dessa forma temos uma análise de sentimento mais especializada para o determinado tipo de jogo, tornando uma quantidade menor de dados e padronizando ainda mais o conteúdo das opiniões. Logo para representar a análise dessa etapa, os gêneros terror e anime foram escolhidos, por serem 2 tipos

com maior volume de dados e divergencia de palavras utilizadas na review. Na tabela 7 temos a inclusão do Vader nessa etapa e os resultados obtidos por cada método.

Modelos	Acurácia	Precisão	Recall	F1-Score	AUC
SVM	89,93%	89,69%	88,15%	84,96%	90,94%
Random Forest	86,43%	85,05%	83,75%	83,40%	89,07%
VADER	73,87%	69,32%	63,29%	66,18%	70,55%

Tabela 7 – Métricas de identificação dos sentimentos em reviews de jogos de terror.

No contexto em que a classificação se concentra mais nas categorias específicas dos jogos para depois analisar o sentimento, observamos um incremento na precisão para os três modelos avaliados. Essa melhoria nas métricas reflete a vantagem de uma abordagem mais especializada na classificação dos dados. Notavelmente, o SVM continua a demonstrar um desempenho superior em comparação com os outros modelos, seguido pelo Random Forest e, por fim, o VADER. Esses resultados reforçam a eficácia de adaptar a classificação dos algoritmos às características específicas dos dados, proporcionando uma análise mais precisa e refinada do sentimento dos jogos.

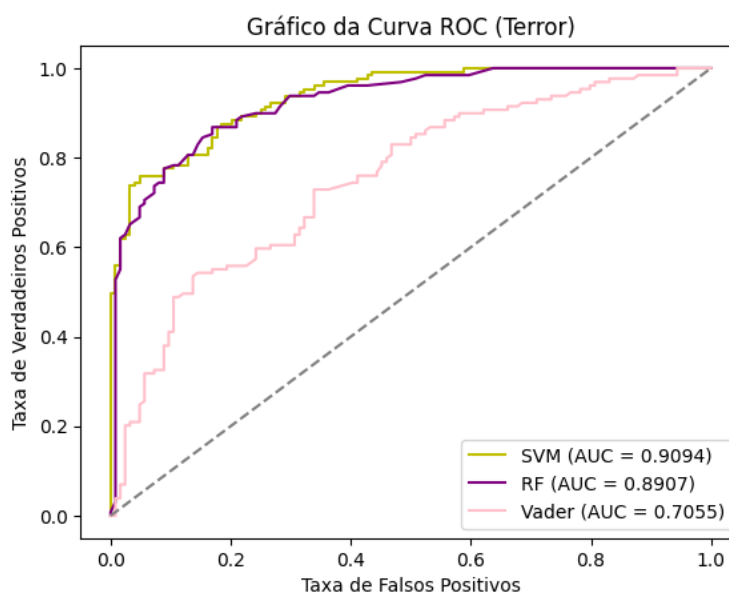


Figura 13 – Valores referentes a curva ROC do gênero Terror.

Fonte: Própria Autora.

A Figura 13 apresenta levemente uma superioridade do algoritmo SVM na capacidade de classificar corretamente, porém o RF ainda se demonstra um ótimo algoritmo com 1% a menos de diferença. O Vader é a métrica que se encontra mais abaixo, como demonstra a tabela 8.

Nessa segmentação de cenários, é evidente que o algoritmo consegue captar com mais precisão as palavras que têm maior relevância para cada tipo de jogo. Essa distinção na seleção de características se torna perceptível ao analisarmos o destaque

Modelos	Acurácia	Precisão	Recall	F1-Score	AUC
SVM	89,62%	87,90%	85,01%	85,88%	87,82%
Random Forest	86,02%	85,11%	84,31%	84,28%	85,19%
VADER	71,49%	65,99%	62,71%	63,44%	66,90%

Tabela 8 – Métricas de identificação dos sentimentos em reviews de jogos de anime.

das features por meio do método de importância de features (feature importance) dos algoritmos de aprendizado de máquina.

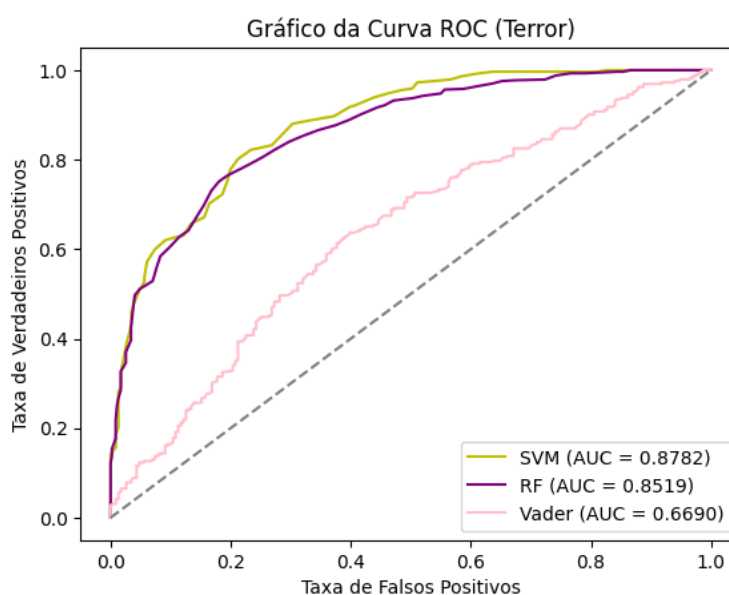


Figura 14 – Valores referentes a curva ROC do gênero Anime.

Fonte: Própria Autora.

Na Figura 14 temos valores inferiores aos calculados na Figura 13 pelo gênero Anime, porém ainda apresentando um bom desempenho em SVM e RF. O Vader continua sendo o inferior, dessa vez ficando abaixo dos 70%.

Na captura das palavras mais relevantes, representados na Figura 15, no cenário de terror conseguimos coletar expressão com conotação as características desse estilo de jogo como “horror”, “scare”/”scary” (susto/pavor) e “spooky” (assustador), como alguns desses jogos disponibilizados no dataset são online temos o uso de palavras como “server” (servidor) e “access” (acesso).

Essa análise revela como as palavras mais consideráveis refletem não apenas as temáticas específicas dos jogos de terror, mas também aspectos práticos à experiência do jogo. Essas descobertas oferecem insights valiosos para entender melhor as características distintivas e os elementos predominantes nos jogos de terror presentes no conjunto de dados.

Olhando para uma área distinta, percebemos uma diferença na abordagem do gráfico:

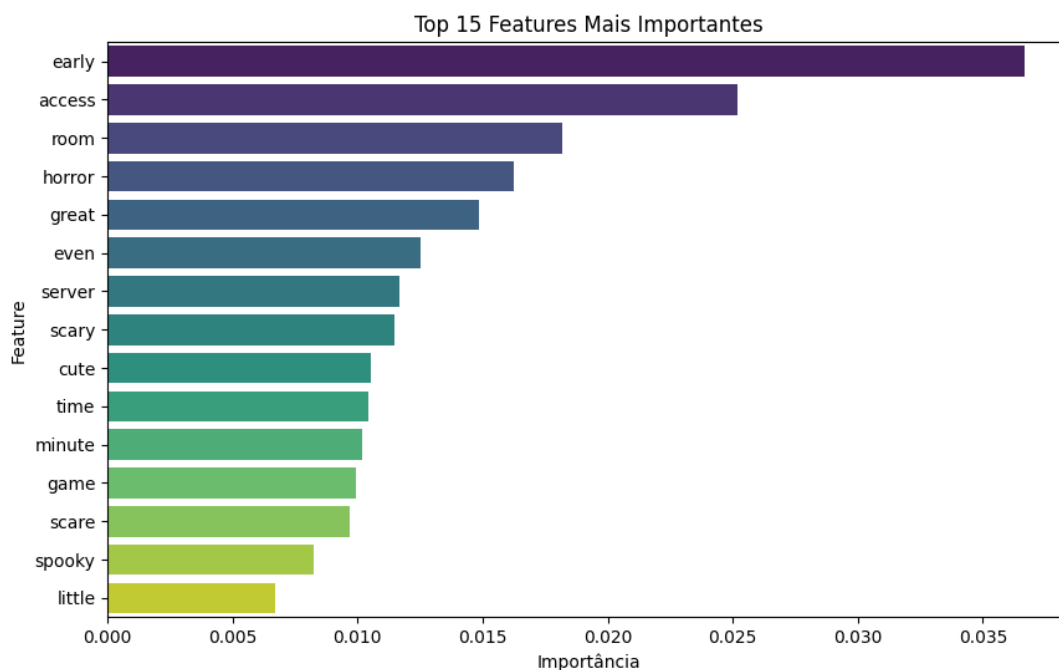


Figura 15 – Ilustração das features mais importantes na previsão do sentimento (terror).

Nesse outro aspecto analítico, na Figura 16, temos o uso de outras palavras relacionadas aos estilos de jogos dessa categoria como “shonen”, “love”, “cute” enquanto temos muitas palavras relacionadas a qualidade do jogo como “great”, “bad” e até número “10”.

Diante das duas experimentação podemos concluir que a experimentação 1 apresentou bons resultados mesmo com alta quantidade de reviews e misturando os tipos de textos (sem filtragem por gênero), tivemos valores acima de 70% em todas as métricas analisadas nos algoritmos de SVM e RF, sendo o SVM o mais eficaz por apresentar o recall e f1-score superiores. Essa abordagem exige menos força de computacional, afinal só construímos 1 modelo para cada algoritmo, porém sobrecarregamos os modelos com a alta quantidade de volumes de textos. O VADER não obteve bons resultados nessa primeira experimentação pela quantidade de textos com muitos caracteres, a biblioteca não se mostrou eficiente nesse caso. A experimentação 2 apresentou ótimos resultados em ambas as abordagens, desde a classificação multiclasse dos gêneros como a classificação do sentimento a partir da review filtrada por tipo. Esse cenário exige mais poder computacional por se tratar da construção de modelos com propostas diferentes, porém usando o mesmo review como base, é possível obter percepções mais específicas e direcionadas sobre como diferentes tipos de conteúdo estão sendo recebidos pelos usuários. O VADER teve uma melhora de resultados em todas as métricas nessa segunda fase, caracterizado pela filtragem que distribuiu uma menor quantidade de dados. Em resumo, a escolha entre as abordagens de análise de sentimento depende dos objetivos específicos do projeto e da profundidade de in-

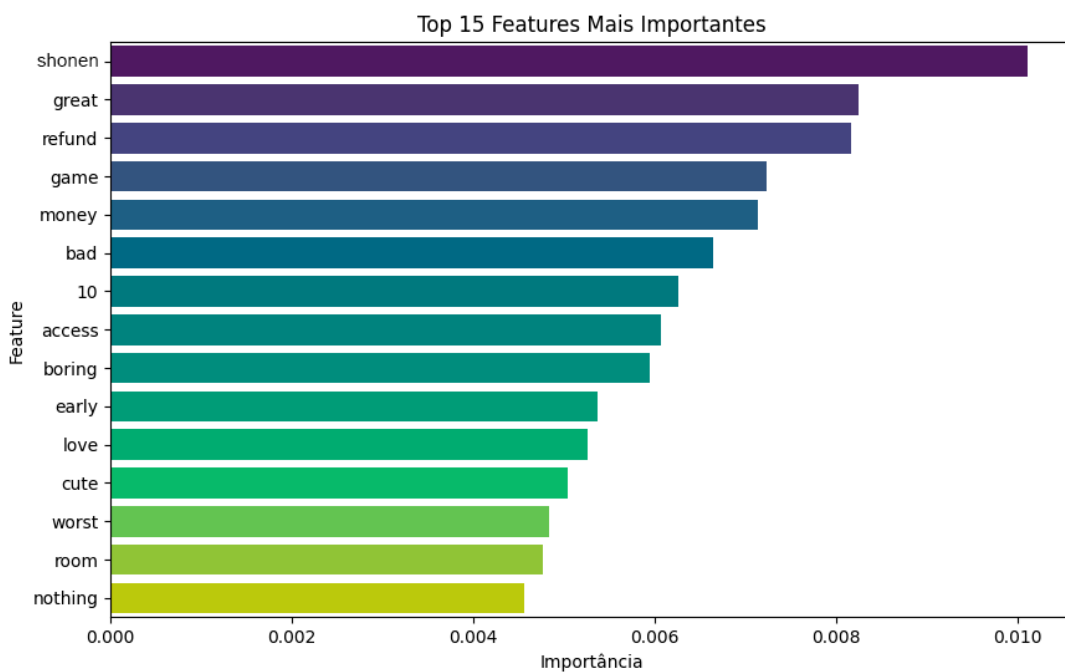


Figura 16 – Ilustração das features mais importantes na previsão do sentimento (anime).

sights desejada. Ambas as abordagens têm seus méritos e podem ser aplicadas de forma eficaz dependendo do contexto da análise.

6 Considerações Finais

O objetivo deste trabalho foi investigar diversas abordagens para a construção de um modelo eficiente para análise de sentimento em reviews de jogos eletrônicos. Com o intuito de encontrar os melhores resultados na classificação, a pesquisa se dividiu em etapas de desenvolvimento para preparar esse modelo, sendo elas a coleta dos dados, em seguida o pré-processamento, após isso a aplicação em 3 algoritmos diferentes para uma análise comparativa e depois o estudo das métricas.

Ressaltando os experimentos realizados temos, uma abordagem geral, onde não filtramos os casos e analisamos o sentimento de forma ampla, e uma abordagem específica, onde filtramos pela categoria do jogo e depois realizamos a análise de sentimento. Os resultados gerados pelo primeiro experimento revelaram que um modelo construído com aprendizagem de máquina é mais eficaz diante de outros métodos sem inteligência computacional. O modelo baseado em SVM demonstrou melhor performance, alcançando um recall e F1-score de 79%, enquanto a acurácia e precisão chegaram em 80%. O RF, por sua vez, também se mostrou eficaz, apresentando um recall e F1-score de 73%, mas com acurácia e precisão superiores, de 81%. O VADER teve uma pontuação significativamente menor, com a maior métrica sendo a precisão, em 68%. Neste contexto, o F1-Score se destaca como a métrica mais relevante do trabalho, pois oferece uma visão balanceada do desempenho do modelo ao combinar os efeitos de precisão e recall, a precisão refere-se à proporção de previsões corretas entre todas as previsões feitas, enquanto o recall mede a capacidade do modelo de identificar todos os verdadeiros positivos, ou seja, a média dessas métricas. Por conta disso o modelo SVM se demonstrou o mais efetivo.

No segundo experimento, onde primeiro era realizado a classificação das categorias usando as reviews como target, nessa etapa utilizamos só os modelos com aprendizagem de máquina. Nossa classificação multiclases apresentou resultados superiores com o modelo SVM na maioria das categorias, alcançando mais de 90% em precisão e acurácia. No entanto, os valores de recall e F1-score foram mais divergentes, com pontuações acima de 80% em encontrar determinadas categorias como terror, anime, tiro e rpg. O RF teve valores mais baixos, alcançando até 85% acurácia e precisão, enquanto no recall e F1-score os valores chegaram a 73% em determinadas categorias, tendo como destaque a identificação dos tipos estratégia, anime e terror. Na parte seguinte onde classificamos o sentimento com as gêneros filtrados, ou seja, usando só as reviews daquele tipo em específico, foram selecionadas as reviews de jogos de terror e anime que foram as melhores resultados na análise anterior. O SVM teve os melhores valores de métricas acima de 88% e com um AUC de 90% em am-

bas categorias, enquanto o RF ficou um pouco abaixo com métricas até 86%. O Vader nessa última etapa chegou a valores de 73% em acurácia e desempenhos acima de 68% nos restos das métricas nos dois cenários, mostrando um melhor comportamento diante dos valores da primeira experimentação.

Os resultados deste trabalho revelaram que um modelo construído com aprendizagem de máquina é mais eficaz diante de outros métodos sem inteligência computacional. Ao ressaltar os experimentos realizados temos, uma abordagem geral, onde não filtramos os casos e analisamos o sentimento de forma ampla, e uma abordagem específica, onde filtramos pela categoria do jogo e depois realizamos a análise de sentimento. Observamos que, embora a abordagem geral apresente métricas ligeiramente inferiores à abordagem filtrada, o esforço computacional é bem inferior, pois só realizamos a construção de um único modelo. No entanto, é importante destacar que a empresa poderia explorar um potencial de análise multidimensional muito maior com um modelo multiclases, indo além da análise de sentimento.

Com os resultados abaixo da biblioteca VADER em relação a algoritmos computacionais, há algumas limitações da ferramenta. Falta de adaptação para assuntos específicos, no caso o vocabulário da biblioteca treinado em mídias sociais e notícias não tem um bom desempenho com comentários relacionados a jogos. Como vantagem o VADER é uma biblioteca pronta e fácil de utilizar.

Para trabalhos futuros nesta temática, é possível explorar ainda mais a análise de sentimento em diferentes domínios, algoritmos e contextos específicos além dos jogos. Outra linha de pesquisa seria investigar técnicas avançadas de processamento de linguagem natural e aprendizado de máquina, como o uso de redes neurais profundas ou LLMs, para melhorar a precisão e a generalização dos modelos de análise de sentimento. Imagine um cenário onde é possível atingir várias opiniões de divergentes assuntos com um mesmo LLM, interpretando ironia e palavras sensíveis de todos os contextos de somente uma máquina. Algumas pesquisas como o de Krugmann em *Sentiment analysis in the age of generative ai*, (KRUGMANN; HARTMANN, 2024), mostram o potencial de prompts focados na exploração de sentimentos em LLMs com a demonstração de resultados que chegam próximos dos desenvolvidos no trabalho atual e até superiores como o Llama 2 que apresentou uma precisão de 92% e recall superiores a 80% na identificação do sentimento. Além disso podemos utilizar as reviews dos usuários e construir um novo sistema a partir do sentimento, como o de recomendação, com base nas opiniões do jogador diante de diferentes estilos de jogos, esse foco mais no usuário final é onde a classificação de sentimento tem vantagem. Essas direções podem contribuir para avançar o campo da análise de sentimento e formentar a indústria com novas ideias.

Em suma, este estudo contribui para o avanço no campo da análise de senti-

mentos em reviews de jogos eletrônicos, fornecendo informações valiosos sobre as diferentes abordagens metodológicas e suas respectivas implicações. Essas descobertas podem orientar futuras pesquisas relacionadas a estudo de jogadores e auxiliar empresas na tomada de decisões ou construções de estratégias diante do sentimento do público.

Referências

- AVANÇO, L. V. *Sobre normalização e classificação de polaridade de textos opinativos na web*. Tese (Doutorado) — Universidade de São Paulo, 2015. Citado na página 26.
- BARBOSA, J. et al. Introdução ao processamento de linguagem natural usando python. *III Escola Regional de Informatica do Piauí*, v. 1, p. 336–360, 2017. Citado na página 16.
- BIRD, S.; KLEIN, E.; LOPER, E. *NLTK: The Natural Language Toolkit*. 2009. <http://www.nltk.org/>. Accessed: 2024-09-28. Citado na página 16.
- BOSER, B. E.; GUYON, I. M.; VAPNIK, V. N. A training algorithm for optimal margin classifiers. In: *Proceedings of the fifth annual workshop on Computational learning theory*. [S.l.: s.n.], 1992. p. 144–152. Citado na página 21.
- CORREIA, M. C. *A Importância do Sistema de Recomendação no Serviço Streaming: O Caso Netflix*. Tese (Doutorado) — ISCTE-Instituto Universitario de Lisboa (Portugal), 2021. Citado na página 19.
- COSTE, R. G. D.; GALLISON, R. Dicionário de didáctica das línguas. *Coimbra, Editora Almedina*, v. 22, 1983. Citado na página 17.
- FELIX, Z. C. et al. Os modelos pens e gameflow na avaliação da satisfação do jogador: Uma análise com o jogo “caixa de pandora” mobile. *Revista Brasileira de Informática na Educação*, v. 28, p. 664–692, 2020. Citado na página 11.
- FORBES. *Promissor: mercado de games ultrapassará US\$ 200 bi até 2023*. 2022. Leia mais em <https://forbes.com.br/forbes-tech/2022/01/com-2022-decisivo-mercado-de-games-ultrapassara-us-200-bi-ate-2023/>. Citado na página 13.
- HO, T. K. Random decision forests. In: IEEE. *Proceedings of 3rd international conference on document analysis and recognition*. [S.l.], 1995. v. 1, p. 278–282. Citado na página 20.
- HOMEM, W. L.; UFES, P. Apostila de machine learning. *PET Engenharia Mecânica, UFES*, 2020. Citado na página 19.
- HUIZINGA, J. *Homo ludens: o jogo como elemento da cultura*. [S.l.]: Editora da Universidade de S. Paulo, Editora Perspectiva, 1971. v. 4. Citado na página 13.
- HUTTO, C.; GILBERT, E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: *Proceedings of the international AAAI conference on web and social media*. [S.l.: s.n.], 2014. v. 8, n. 1, p. 216–225. Citado na página 22.
- JONES, K. S. A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, MCB UP Ltd, v. 28, n. 1, p. 11–21, 1972. Citado na página 17.

JÚNIOR, A. L. dos S.; CAVALCANTE, I. F. Distribuição da força de trabalho por gênero no desenvolvimento de jogos eletrônicos no Brasil. Citado na página 10.

KARAMIBEKR, M.; GHORBANI, A. A. Sentiment analysis of social issues. In: IEEE. *2012 international conference on social informatics*. [S.l.], 2012. p. 215–221. Citado na página 14.

KRUGMANN, J. O.; HARTMANN, J. Sentiment analysis in the age of generative ai. *Customer Needs and Solutions*, Springer, v. 11, n. 1, p. 1–19, 2024. Citado 2 vezes nas páginas 27 e 45.

LU, C. et al. Patches and player community perceptions: Analysis of no man's sky steam reviews. In: DIGRA. *DiGRA'20-Proceedings of the 2020 DiGRA International Conference*. [S.l.], 2020. Citado na página 11.

MARDJO, A.; CHOKSUCHAT, C. Hyvadr: Hybrid vader–random forest and gwo for bitcoin tweet sentiment analysis. *IEEE Access*, IEEE, v. 10, p. 101889–101897, 2022. Citado na página 27.

MBAKIRTZIS, M. M. Análise do crescimento da indústria dos jogos digitais. Universidade Estadual Paulista (Unesp), 2023. Citado na página 10.

MOREIRA, D. J. d. C. Classificação de texto usando codificação de palavras. Universidade Federal do Rio de Janeiro, 2021. Citado na página 27.

MORETTIN, P. A.; SINGER, J. d. M. Estatística e ciência de dados. 2022. Citado na página 23.

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. Citado 2 vezes nas páginas 15 e 26.

RAVI, K.; RAVI, V. A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-based systems*, Elsevier, v. 89, p. 14–46, 2015. Citado na página 26.

TRINDADE, M. O. S. d. *Exploração e comparação de algoritmos de classificação em Machine Learning: uma abordagem estatística*. Dissertação (B.S. thesis) — Universidade Federal do Rio Grande do Norte, 2023. Citado na página 23.

UYVAL, A. K.; GUNAL, S. The impact of preprocessing on text classification. *Information processing & management*, Elsevier, v. 50, n. 1, p. 104–112, 2014. Citado na página 16.

VIEIRA, R.; LOPES, L. Processamento de linguagem natural e o tratamento computacional de linguagens científicas. *Em corpora*, p. 183, 2010. Citado na página 15.