

Random Forest na Análise de Sentimentos em Jogos Digitais

Guilherme Cruz Moreira Pinto
Pós Graduação Data Science e
Analytics
Senai-Cimatec
Salvador, Brasil
guilhermecruzmp@gmail.com

Èldman de Oliveira Nunes
Pós Graduação Data Science e
Analytics
Senai-Cimatec
Salvador, Brasil
eldman.nunes@fieb.org.br

Abstract—This study describes the application of the Random Forest algorithm in sentiment analysis of digital games, categorizing text into positive, negative, or neutral sentiments. The research aims to deepen the understanding of player interactions with games, providing relevant insights for companies in a constantly evolving industry. The methodology employed includes the collection of a dataset containing 640,987 player reviews and meticulous preprocessing to ensure data quality. The dataset was divided into 80% for training (512,790 reviews) and 20% for testing (128,197 reviews), with a sample of 109,075 reviews analyzed. The Random Forest model achieved an accuracy of 89%, a precision of 99% for positive sentiments, and 63% for negative sentiments, and a recall of 88% for positive sentiments and 94% for negative sentiments. The confusion matrix results indicate an overall accuracy of 89%. These metrics demonstrate the model's effectiveness in accurately classifying player sentiments. In summary, this study provides valuable data-driven insights into player feedback, offering a practical tool for companies to enhance their products and strategies in the competitive gaming market.

Keywords—Random Forest, sentiment analysis, digital games, text categorization, player interactions, data preprocessing, model evaluation.

I. INTRODUÇÃO

Os jogos digitais tornaram-se uma parte significativa da cultura moderna, oferecendo entretenimento e engajamento para milhões de pessoas ao redor do mundo [1]. Com o crescimento exponencial da indústria, compreender as interações dos jogadores com esses jogos tornou-se essencial para desenvolvedores e empresas do setor [2]. Uma questão central é como categorizar as opiniões dos jogadores, que podem variar de positivas e negativas. Isso é crucial porque a opinião dos jogadores impacta diretamente o desenvolvimento dos jogos, as estratégias de marketing e o sucesso geral do produto [3]. Opiniões negativas podem indicar problemas de jogabilidade, bugs ou falhas que precisam ser corrigidos, enquanto opiniões positivas podem destacar os pontos fortes do jogo que devem ser mantidos ou aprimorados [4]. Além disso, as opiniões podem fornecer insights sobre áreas que precisam de melhorias para aumentar a satisfação do jogador [5].

O objetivo desta pesquisa é analisar os sentimentos em jogos digitais da plataforma Steam [6], visando categorizar as opiniões dos jogadores de maneira precisa e eficiente.

Este trabalho se concentra especificamente no contexto dos jogos digitais e utiliza um conjunto de dados extenso e diversificado coletado da principal plataforma de jogos digitais, a Steam [6]. Além disso, a pesquisa inclui uma avaliação abrangente dos resultados utilizando métricas específicas, o que fornece uma análise detalhada e prática dos sentimentos dos jogadores.

A metodologia adotada envolve várias etapas: a coleta de dados, o pré-processamento meticuloso dos textos e a implementação do modelo Random Forest para a análise dos sentimentos. A coleta de dados foi realizada na plataforma de avaliação de jogos Steam [6], resultando em um conjunto de dados com 640,987 opiniões de jogadores. O pré-processamento incluiu a remoção de stopwords, normalização e tokenização dos textos. O modelo Random Forest foi escolhido pela sua robustez e capacidade de lidar com grandes volumes de dados, sendo implementado utilizando a biblioteca Scikit-Learn em Python [7].

A estrutura deste artigo está organizada da seguinte forma: a Seção II apresenta a análise de sentimentos, a Seção III descreve os jogos digitais, a Seção IV detalha a metodologia utilizada, a Seção V discute os resultados e a Seção VI aborda as conclusões e sugestões para trabalhos futuros.

II. ANÁLISE DE SENTIMENTOS

A análise de sentimentos, também conhecida como mineração de opiniões, é o campo de estudo que analisa automaticamente as opiniões, sentimentos, avaliações, atitudes e emoções das pessoas em relação a produtos, serviços, organizações, indivíduos, questões, eventos, tópicos e seus atributos [8]. Esta análise pode ser realizada em textos provenientes de diversas fontes, como avaliações de produtos, redes sociais, fóruns de discussão, blogs, entre outros. O objetivo principal é determinar a polaridade do texto, categorizando-o como positivo, negativo ou neutro [8].

A análise de sentimentos é fundamental no contexto dos jogos digitais por várias razões. Primeiramente, permite às empresas entender melhor as opiniões dos jogadores, fornecendo uma visão clara e detalhada sobre como os jogadores percebem diferentes aspectos dos jogos, como

gráficos, jogabilidade, enredo e experiência geral. Isso é essencial para identificar pontos fortes que devem ser mantidos e promovidos, bem como áreas problemáticas que necessitam de melhorias urgentes [9].

Em segundo lugar, a análise de sentimentos ajuda a direcionar as estratégias de marketing. Compreender o que os jogadores gostam ou não gostam pode orientar campanhas de marketing que ressoem melhor com o público-alvo. Além disso, permite a segmentação de mercado baseada em preferências reais dos jogadores, potencializando as estratégias de comunicação e promoção [10]. Em terceiro lugar, a análise de sentimentos é crucial para o desenvolvimento e a evolução contínua dos jogos. Feedback negativo pode apontar para problemas técnicos ou de jogabilidade que precisam ser resolvidos rapidamente para evitar a perda de jogadores. Por outro lado, feedback positivo pode destacar elementos do jogo que são particularmente bem recebidos e que poderiam ser explorados ou ampliados em futuras atualizações ou novos jogos [11].

Finalmente, a análise de sentimentos permite monitorar a satisfação dos jogadores em tempo real. Isso é particularmente importante em um mercado altamente competitivo, onde a capacidade de responder rapidamente ao feedback dos jogadores pode determinar o sucesso ou o fracasso de um jogo [12]. A análise contínua das opiniões dos jogadores permite ajustes rápidos e melhorias contínuas, garantindo que os jogos permaneçam atrativos e competitivos.

III. JOGOS DIGITAIS

A análise de sentimentos em jogos digitais é fundamental para entender a experiência do jogador e melhorar a interação com o jogo. Empresas do setor têm utilizado essa técnica para monitorar a satisfação dos jogadores em tempo real, identificar problemas na jogabilidade e até prever tendências de mercado, segmentando e entendendo melhor o público-alvo. Por meio dessa análise, estratégias de marketing e desenvolvimento de produtos podem ser direcionadas de maneira mais precisa. Além disso, a análise de sentimentos pode ser aplicada no desenvolvimento de narrativas e personagens. Compreender as emoções dos jogadores em relação aos diferentes aspectos do jogo permite a criação de histórias envolventes e personagens capazes de gerar uma conexão emocional. No entanto, é importante ressaltar que a análise de sentimentos em jogos digitais ainda enfrenta desafios, pois as pessoas expressam suas emoções de maneiras diferentes. É fundamental utilizar abordagens e algoritmos robustos que levem esses aspectos em consideração.

A análise de sentimentos em jogos digitais possui um enorme potencial para melhorar a experiência dos jogadores e impulsionar o sucesso das empresas do ramo. A compreensão dos sentimentos dos jogadores é uma estratégia que permite identificar quais os pontos fortes e fracos, direcionando estratégias para aprimorar continuamente os jogos e fornecer uma experiência de usuário excepcional.

Além disso, a análise de sentimentos desempenha um papel fundamental na antecipação das tendências do mercado, identificando desejos e demandas emergentes do público.

Outro aspecto importante é a detecção de comportamentos negativos, como o cyberbullying e a toxicidade. Monitorar as interações dos jogadores possibilita a identificação de padrões de linguagem ofensivos e prejudiciais, permitindo intervenções proativas para criar um ambiente de jogo mais seguro e agradável.

Em suma, a análise de sentimentos em jogos digitais se torna uma ferramenta decisiva para empresas que desejam se destacar em um mercado cada vez mais competitivo. Com ela, é possível direcionar estratégias, desenvolver e promover jogos que atendam às necessidades e desejos dos jogadores, além de criar uma base de fãs engajados. A análise de sentimentos não é apenas uma tendência no setor de jogos, mas sim uma prática essencial para o sucesso e crescimento das empresas nesse segmento.

IV. RANDOM FOREST NA ANÁLISE DE SENTIMENTOS

O Random Forest é um algoritmo de aprendizado de máquina que pertence à classe de métodos ensemble, onde múltiplos modelos de aprendizado são treinados para resolver o mesmo problema e suas previsões são combinadas para obter uma melhor performance do que qualquer modelo individual. Foi introduzido por Leo Breiman em 2001 como uma extensão das árvores de decisão [13].

A técnica de Random Forest deriva da ideia de árvores de decisão, que são estruturas hierárquicas de perguntas baseadas nas características dos dados. No entanto, árvores de decisão individuais podem ser instáveis, pois pequenas variações nos dados podem resultar em árvores muito diferentes. Para contornar essa limitação, o Random Forest utiliza o conceito de bootstrap aggregating, ou bagging, para criar várias árvores de decisão independentes a partir de diferentes subconjuntos do conjunto de dados de treinamento [13].

O algoritmo Random Forest segue um processo em várias etapas:

1. **Bootstrapping:** Um grande número de subconjuntos é gerado aleatoriamente do conjunto de dados original, com substituição. Cada subconjunto servirá como um conjunto de dados de treinamento para uma árvore de decisão individual.
2. **Construção das Árvores de Decisão:** Para cada subconjunto, uma árvore de decisão é construída. Durante o processo de construção, a cada nó da árvore, uma amostra aleatória de características é considerada para determinar a melhor divisão, em vez de considerar todas as características. Isso adiciona uma camada adicional de aleatoriedade e contribui para a diversidade entre as árvores.
3. **Agregação:** Depois que todas as árvores de decisão são construídas, o Random Forest combina suas previsões. Para problemas de classificação, a

previsão final é baseada na maioria das previsões (votação majoritária) das árvores individuais. Para problemas de regressão, a média das previsões de todas as árvores é calculada.

Vantagens do Random Forest:

- **Robustez ao Overfitting:** Devido ao processo de bagging e à seleção aleatória de características, o Random Forest é menos suscetível ao overfitting em comparação com árvores de decisão individuais.
- **Desempenho em Dados Grandes e Complexos:** É altamente eficaz em conjuntos de dados grandes e complexos, com muitas características e registros, devido à sua capacidade de paralelizar a construção das árvores.
- **Importância das Características:** O Random Forest permite a avaliação da importância das características, identificando quais variáveis são mais influentes na predição do modelo. Isso é feito calculando a redução média da impureza (Gini impurity) proporcionada por cada variável em todas as árvores.
- **Resistência a Outliers e Ruídos:** A robustez do Random Forest se estende à sua capacidade de lidar com ruídos e outliers nos dados, mantendo uma performance consistente.

Aplicação na Análise de Sentimentos: Na análise de sentimentos, a capacidade do Random Forest de lidar com alta dimensionalidade e variáveis correlacionadas é particularmente útil. Textos de avaliações de jogos digitais, por exemplo, contêm muitas palavras e expressões diferentes que podem estar correlacionadas. O Random Forest processa essas informações de maneira eficaz, construindo um modelo robusto que pode identificar padrões complexos de sentimentos expressos nos textos.

Para esta pesquisa, o algoritmo foi implementado utilizando a biblioteca Scikit-Learn em Python [7], que fornece ferramentas eficientes para a construção de modelos de aprendizado de máquina.

V. MÉTODO

Esta seção detalha a metodologia adotada para a análise de sentimentos em jogos digitais utilizando o algoritmo Random Forest. A metodologia baseia-se em práticas estabelecidas na literatura de aprendizado de máquina e processamento de linguagem natural, adaptadas para o contexto específico dos jogos digitais. As etapas do método incluem a coleta de dados, o pré-processamento dos textos, o tratamento de classes de dados desbalanceadas e a implementação do modelo Random Forest. A seguir, descreve-se cada uma dessas etapas com mais detalhes.

A. Coleta de Dados

A coleta de dados para análise de sentimentos em jogos digitais foi realizada através de Web scraping e

transformação em CSV da plataforma Steam [6]. Textos de análises, comentários de jogadores e discussões relacionadas aos jogos selecionados para o estudo foram extraídos. Os dados foram coletados de forma sistemática, considerando diferentes fontes e tipos de conteúdo, com o objetivo de assegurar uma base ampla e diversificada para a construção do modelo de análise de sentimentos. Durante o processo de coleta, a plataforma de jogos Steam [6] foi explorada, fornecendo informações sobre a percepção dos jogadores em relação aos vários jogos analisados. Além disso, técnicas de processamento de linguagem natural foram aplicadas para o pré-processamento dos textos coletados, a fim de eliminar interferências e padronizar a análise dos sentimentos [8]. Títulos de diferentes gêneros, como jogos de ação, aventura, estratégia e RPG, foram selecionados para garantir uma diversidade de experiências de jogo. Os dados coletados foram organizados em um formato estruturado, facilitando a análise subsequente. Um total de mais de 640,987 textos foi coletado, contendo milhões de palavras, proporcionando uma base de dados robusta para a análise de sentimentos em jogos digitais [9]. Estes dados foram utilizados para treinar e avaliar o modelo de análise de sentimentos, originando informações valiosas sobre a percepção dos jogadores em relação a vários aspectos dos jogos, como gráficos, jogabilidade, enredo e desafio [9].

B. Pré-processamento de Dados

Durante o pré-processamento de dados para análise de sentimentos em jogos digitais, uma série de etapas rigorosas foi seguida para assegurar a qualidade e a integridade dos dados coletados. Em primeiro lugar, realiza-se a remoção de pontuações, stopwords e palavras irrelevantes que poderiam afetar os resultados da análise [10]. Este passo é fundamental para eliminar ruídos e focar apenas nas palavras que carregam significado relevante para a análise de sentimentos. Listas de stopwords específicas do idioma inglês foram utilizadas, considerando que a maioria dos dados coletados estava nesse idioma [10]. A lista de stopwords foi extraída da biblioteca NLTK (Natural Language Toolkit) [10]. Além disso, técnicas de stemming e lematização foram aplicadas para normalizar as palavras. O stemming consiste em cortar os sufixos das palavras, enquanto a lematização utiliza dicionários linguísticos para transformar palavras em suas formas raiz. Para o stemming, utilizou-se o PorterStemmer e o SnowballStemmer, enquanto para a lematização, utilizou-se o WordNetLemmatizer, todas implementadas com a biblioteca NLTK [10].

A normalização dos textos desempenhou um papel crucial no pré-processamento. Todas as palavras foram transformadas para minúsculas, facilitando a comparação e o processamento subsequente. Além disso, caracteres especiais, como emojis e símbolos gráficos, foram removidos para evitar interferências indesejadas na análise. Outro passo importante foi a vetorização dos textos, onde os textos foram representados como vetores numéricos que podem ser processados pelos algoritmos de machine learning [11]. Técnicas de vetorização, como TF-IDF (Term Frequency-Inverse Document Frequency) e CountVectorizer, foram utilizadas, transformando os textos em uma matriz de

frequências ou de pesos, respectivamente [11]. Estas técnicas foram escolhidas devido à sua eficácia em capturar a relevância das palavras nos textos. Além disso, os textos foram divididos em frases ou parágrafos menores para uma análise mais detalhada dos sentimentos expressos. Esta abordagem permitiu identificar padrões e tendências em diferentes partes do texto, fornecendo uma compreensão mais abrangente dos sentimentos dos jogadores [12].

Para assegurar a reprodutibilidade da metodologia, experimentos empíricos foram realizados para avaliar as diferentes técnicas de pré-processamento e identificar a combinação ideal para o conjunto de dados. A eficácia do stemming e da lematização foi testada separadamente e em conjunto, comparando o desempenho do modelo em termos de acurácia, precisão, recall e F1-score. A combinação que apresentou o melhor desempenho foi utilizada na versão final do modelo.

C. Tratamento de Classes de Dados Desbalanceadas

Para lidar com o desbalanceamento de classes no conjunto de dados, métodos de oversampling e undersampling foram utilizados. O SMOTE (Synthetic Minority Over-sampling Technique) foi aplicado para aumentar a quantidade de exemplos da classe minoritária, enquanto o RandomUnderSampler foi utilizado para reduzir a quantidade de exemplos da classe majoritária [14]. Isso foi essencial para evitar que o modelo ficasse tendencioso em relação às classes majoritárias e para garantir uma melhor generalização. O CountVectorizer foi utilizado com um máximo de 3000 características e o TfidfTransformer para transformar os textos [14]. Após a vetorização, o SMOTE foi aplicado para aumentar a quantidade de exemplos da classe minoritária e o RandomUnderSampler para reduzir a quantidade de exemplos da classe majoritária [14]. Este procedimento foi seguido pela aplicação do Random Forest Classifier [7].

D. Implementação do Modelo

O processo de análise de sentimentos envolveu várias etapas. Primeiramente, os textos foram coletados e pré-processados. O pré-processamento incluiu a remoção de stopwords, normalização, tokenização e, em alguns casos, stemming e lematização. Stopwords são palavras comuns, como "e", "o", "a", que foram removidas porque não adicionam significado significativo ao texto. A normalização incluiu a conversão de todos os caracteres para minúsculas para manter a consistência. A tokenização dividiu o texto em unidades menores, como palavras ou frases [12].

Após o pré-processamento, os textos foram transformados em uma representação numérica que pode ser processada por algoritmos de aprendizado de máquina. Uma técnica comum foi a vetorização, onde cada texto foi transformado em um vetor de características utilizando métodos como TF-IDF (Term Frequency-Inverse Document Frequency) [14].

O conjunto de dados foi dividido em treinamento e teste, com uma proporção de 80% para treinamento e 20% para teste. Um total de 400 árvores de decisão foi utilizado no modelo Random Forest. Essa escolha foi baseada em experimentos empíricos que avaliaram diferentes valores de estimadores, variando de 100 a 500, com incrementos de 50. A métrica de acurácia foi utilizada como medida de desempenho durante o treinamento, uma vez que é uma métrica direta e intuitiva para avaliar a proporção de previsões corretas feitas pelo modelo [7]. O valor de 400 estimadores foi selecionado como o melhor compromisso entre precisão e eficiência computacional.

Após o treinamento, o modelo foi avaliado com o conjunto de teste para medir seu desempenho em dados não vistos. Métricas de desempenho como acurácia, precisão, recall e F1-score foram utilizadas para uma avaliação abrangente. Essas métricas fornecem diferentes perspectivas sobre o desempenho do modelo, considerando não apenas a proporção de previsões corretas, mas também a qualidade das previsões para cada classe de sentimento [14]. Além disso, foi realizada a análise de importância das variáveis no modelo, a fim de identificar quais fatores mais influenciam na predição de sentimentos nos jogos digitais. A importância das variáveis foi calculada com base na redução média da impureza (Gini impurity) proporcionada por cada variável em todas as árvores do modelo [7]. Esta análise ajuda a entender quais características dos textos têm maior impacto na classificação dos sentimentos, fornecendo insights valiosos para a interpretação dos resultados.

VI. RESULTADOS E DISCUSSÃO

O modelo Random Forest apresentou um desempenho significativo na classificação das opiniões dos jogadores devido à sua capacidade de lidar com grandes volumes de dados e variáveis complexas. A precisão do modelo foi medida utilizando o conjunto de teste, alcançando uma acurácia de 89%. Este alto nível de precisão indica que o modelo identifica corretamente os sentimentos expressos pelos jogadores nos textos analisados.

Para compreender melhor os dados utilizados e a eficácia do modelo, a Fig. 1 apresenta a análise das palavras mais comuns no dataset de avaliações de jogos digitais. A visualização dessas palavras é essencial para identificar os termos que aparecem com maior frequência nas análises dos jogadores, fornecendo insights sobre os tópicos mais discutidos e as emoções mais expressadas. Isso auxilia na interpretação dos sentimentos predominantes e na compreensão dos principais fatores que influenciam a percepção dos jogadores sobre os jogos

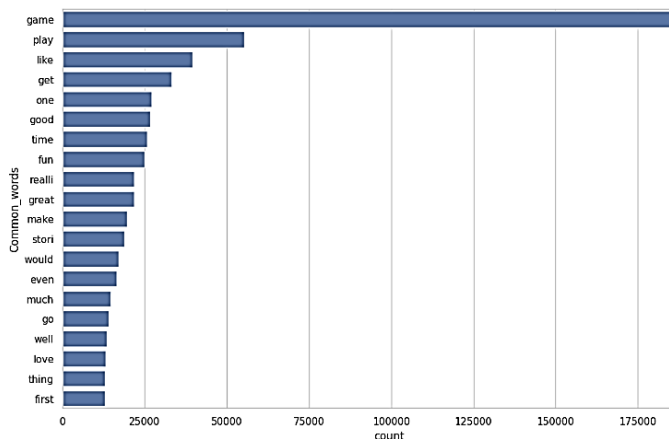


Fig. 1. Palavras mais comuns comentadas por jogadores

A matriz de confusão, representada na Fig. 2, fornece uma visão detalhada do desempenho do modelo na classificação dos sentimentos. Ela mostra a precisão do modelo em identificar corretamente os sentimentos positivos e negativos, com uma precisão de 99% para sentimentos positivos e 63% para sentimentos negativos, e um recall de 88% para sentimentos positivos e 94% para sentimentos negativos. Esta análise é crucial para avaliar a capacidade do modelo em categorizar corretamente os diferentes sentimentos expressos pelos jogadores.



Fig. 2. Matriz de Confusão Random Forest

A análise de importância das variáveis revelou que as palavras mais influentes na predição dos sentimentos foram aquelas associadas a emoções fortes e termos específicos de jogos. Utilizou-se a métrica de importância Gini para identificar quais palavras mais contribuíam para as decisões do modelo. As palavras mais influentes incluíam termos que foram normalizados utilizando de técnicas de stemming e lematização, o que significa que as palavras aparecem em suas formas reduzidas e padronizadas.

A Tabela I resume as principais métricas de desempenho do modelo, incluindo precisão, recall, F1-score e acurácia. Essas métricas fornecem uma visão abrangente sobre a eficácia do

modelo em classificar corretamente os sentimentos dos jogadores. A precisão elevada e o alto valor do F1-score indicam que o modelo Random Forest é confiável e eficiente para a tarefa de análise de sentimentos em jogos digitais.

Tabela I: Resultados das métricas de classificação para Random Forest

Métrica	Random Forest
Precisão	0.89
Recall	0.91
F1-Score	0.88
Acurácia	0.89

VII. CONCLUSÃO

A pesquisa teve como objetivo analisar os sentimentos em jogos digitais especificamente na plataforma Steam, categorizando textos em sentimentos positivos, negativos ou neutros. Para alcançar esse objetivo, foi empregada uma metodologia que incluiu a coleta de um grande conjunto de dados de avaliações de jogos digitais, o pré-processamento detalhado desses dados e a implementação do modelo Random Forest para a análise de sentimentos. A avaliação do desempenho do modelo utilizou métricas específicas de classificação, como precisão, recall, F1-Score e acurácia. Os resultados indicam que o modelo Random Forest apresenta alto desempenho na classificação de sentimentos em textos relacionados a jogos digitais na plataforma Steam, com valores elevados de precisão, recall, F1-Score e acurácia.

As contribuições reais da pesquisa abrangem diversas áreas. Primeiramente, o estudo desenvolveu um procedimento metodológico detalhado para a coleta e pré-processamento de dados de avaliações de jogos digitais. A coleta de dados foi realizada através de scraping da plataforma Steam, resultando em um conjunto de dados extenso e diversificado. O pré-processamento dos textos incluiu a remoção de stopwords, normalização, tokenização, stemming e lematização, assegurando a qualidade e a integridade dos dados utilizados na análise de sentimentos.

A pesquisa apresentou uma análise abrangente dos sentimentos dos jogadores na plataforma Steam, categorizando as opiniões em positivas, negativas. Esta abordagem permitiu identificar padrões de interação dos jogadores e os principais fatores que influenciam suas percepções, como jogabilidade, gráficos e enredo dos jogos. Esses insights fornecem uma base valiosa para compreender melhor as preferências e insatisfações dos jogadores, oferecendo dados concretos para aprimorar o desenvolvimento e a comercialização de jogos digitais.

Além disso, a pesquisa gerou insights práticos para empresas de jogos digitais. A análise detalhada das opiniões dos jogadores possibilita que as empresas foquem nos aspectos

mais valorizados e abordem áreas de insatisfação, aumentando a satisfação dos jogadores e potencialmente o sucesso comercial dos jogos. Esses insights são úteis para orientar decisões estratégicas e táticas no desenvolvimento e promoção de jogos, auxiliando as empresas a se manterem competitivas no mercado.

No âmbito acadêmico, o estudo contribui para o corpo de conhecimento sobre análise de sentimentos ao aplicar um modelo de aprendizado de máquina em um contexto específico de jogos digitais. A pesquisa explora os desafios e oportunidades associados à análise de grandes volumes de dados textuais, fornecendo uma base metodológica que pode ser replicada ou aprimorada em estudos futuros. Foram identificadas limitações, como a restrição do conjunto de dados às avaliações da plataforma Steam e a limitação a textos em inglês, sugerindo a necessidade de expandir a análise para outras plataformas e idiomas para aumentar a generalização dos resultados.

Essas contribuições ressaltam a importância e o impacto da pesquisa, tanto em termos acadêmicos quanto práticos, oferecendo um recurso valioso para o entendimento e melhoria da experiência dos jogadores de jogos digitais.

REFERÊNCIAS

- [1] Entertainment Software Association, "Essential Facts About the Computer and Video Game Industry 2020." [Online]. Available: <https://www.theesa.com/esa-research/2020-essential-facts-about-the-computer-and-video-game-industry/>. [Accessed: Jul. 4, 2024].
- [2] Statista, "Number of active video gamers worldwide from 2015 to 2023." [Online]. Available: <https://www.statista.com/statistics/748044/number-video-gamers-world/>. [Accessed: Jul. 4, 2024].
- [3] A. Anderson and K. E. Dill, "Video games and aggressive thoughts, feelings, and behavior in the laboratory and in life," *Journal of Personality and Social Psychology*, vol. 78, no. 4, pp. 772-790, 2000.
- [4] D. Williams, N. Yee, and S. E. Caplan, "Who plays, how much, and why? Debunking the stereotypical gamer profile," *Journal of Computer-Mediated Communication*, vol. 13, no. 4, pp. 993-1018, 2008.
- [5] J. Kücklich, "Precarious playbour: Modders and the digital games industry," *The Fibreculture Journal*, no. 5, 2005..
- [6] "Steam." [Online]. Available: <https://store.steampowered.com/>. [Accessed: Jul. 4, 2024].
- [7] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011.
- [8] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New avenues in opinion mining and sentiment analysis," *IEEE Intelligent Systems*, vol. 28, no. 2, pp. 15-21, 2013.
- [9] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093-1113, 2014.
- [10] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*, O'Reilly Media Inc., 2009.
- [11] J. Ramos, "Using TF-IDF to Determine Word Relevance in Document Queries," in *Proc. of the First Instructional Conference on Machine Learning*, 2003.
- [12] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, vol. 2, no. 1-2, pp. 1-135, 2008.
- [13] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.
- [14] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427-437, 2009.

**CENTRO UNIVERSITÁRIO SENAI CIMATEC
ESPECIALIZAÇÃO EM DATA SCIENCE & ANALYTICS****ATA DE APRESENTAÇÃO DE PROJETO FINAL DE CURSO**

Ata de apresentação do Projeto Final de Curso, “**Random Forest na Análise de Sentimentos em Jogos Digitais**”, submetido pelo aluno **Guilherme Cruz Moreira Pinto**, como parte dos requisitos para obtenção do Certificado de **Especialista em Data Science & Analytics** pelo Centro Universitário SENAI CIMATEC, às 17h00 do dia 07 de Junho de 2024. Reuniu-se remotamente pela plataforma Google Meet, a Banca Examinadora designada pelo Prof Dr. Éldman de Oliveira Nunes – Orientador, constituída pelo Prof Dr. Éldman de Oliveira Nunes e Prof MSc Braian Varjão Gama Bispo. O Prof Orientador deu início aos trabalhos com as devidas orientações, e a exposição foi realizada pelo estudante dentro do prazo de tempo estabelecido. Ao final da apresentação a banca reuniu-se atribuindo a seguinte nota: **7,0** (sete pontos e zero décimos).

A banca de avaliadores decidiu pela:**(X) Aprovação do trabalho**

Caberá ao aluno apresentar em no máximo em 30 (trinta) dias a contar da data de assinatura desta Ata, uma cópia do trabalho em PDF, constando as considerações pontuadas pela banca. A Ata de Apresentação do Projeto Final de Curso deve ser digitalizada e inserida na terceira página do TCC ou como anexo do artigo.

() Reprovação do trabalho

O aluno terá que se matricular novamente no TCC – Trabalho de Conclusão de Curso e ser submetido a uma banca avaliadora no semestre seguinte.

As ações consequentes ao status de Aprovação deverão obedecer ao prazo proposto acima sob pena do parecer final ser modificado para o status de Reprovado automaticamente e sem possibilidade de recurso.

Para constar, lavrou-se a presente ata que vai assinada por todos os membros da Banca. Por estarem cientes de suas obrigações estão de acordo com os termos desse documento:

Salvador, 07 de Junho de 2024.

gov.br
Documento assinado digitalmente
ELDMAN DE OLIVEIRA NUNES
Data: 08/06/2024 11:28:10-0300
Verifique em <https://validar.iti.gov.br>

Éldman de Oliveira Nunes
Professor Orientador

Assinado eletronicamente por:
Patricia Freitas Tourinho
CPF: ***.733.265-**
Data: 25/06/2024 17:20:35 -03:00

Patricia Freitas Tourinho
Coordenadora de Pós-Graduação *Lato Sensu*

Braian Varjão
Gama Bispo
Assinado de forma digital
por Braian Varjão Gama
Bispo
Dados: 2024.06.13 10:33:47
-03'00'

Braian Varjão Gama Bispo
Membro da banca



MANIFESTO DE ASSINATURAS



Código de validação: KFHKX-QQHxD-AZ3MU-YDCMC

Esse documento foi assinado pelos seguintes signatários nas datas indicadas (Fuso horário de Brasília):

- ✓ Patricia Freitas Tourinho (CPF *****.733.265-****) em 25/06/2024 17:20 - Assinado eletronicamente

Endereço IP	Geolocalização
179.105.137.162	Lat: -13,008630 Long: -38,483951
	Precisão: 20 (metros)
Autenticação	patricia.tourinho@fieb.org.br (Verificado)
Login	
UTetBt0deIYYvPX8PcCPQtwWH4qJUnhgXhg/Ua0lcEk=	
SHA-256	

Para verificar as assinaturas, acesse o link direto de validação deste documento:

<https://assinatura.senaibahia.com.br/validate/KFHKX-QQHxD-AZ3MU-YDCMC>

Ou acesse a consulta de documentos assinados disponível no link abaixo e informe o código de validação:

<https://assinatura.senaibahia.com.br/validate>