

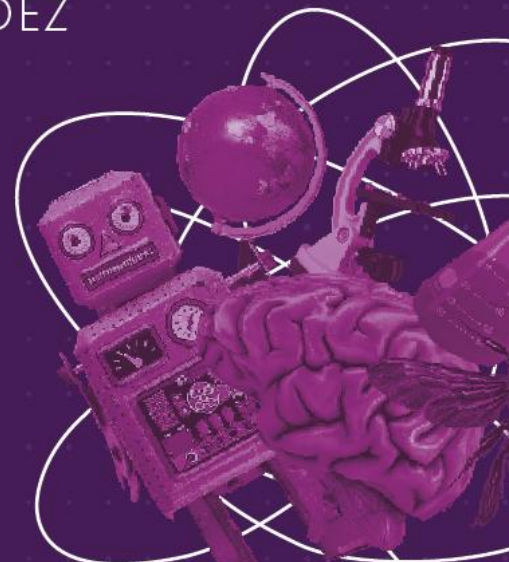
DESVENDANDO AS PREFERÊNCIAS DOS GAMERS: ANÁLISE PREDITIVA E MERCADOLÓGICA DA INDÚSTRIA DE JOGOS

Professor orientador: Abner Santos Belém

Alunos: Leonardo de Lima Amaral e Marcelo Saraiva
Cavalcanti

PROGRAMA DE
INICIAÇÃO CIENTÍFICA
PIC/CEUB

RELATÓRIOS DE PESQUISA
VOLUME 10 Nº 1- JAN/DEZ
2024



**CENTRO UNIVERSITÁRIO DE BRASÍLIA - CEUB
PROGRAMA DE INICIAÇÃO CIENTÍFICA**

**LEONARDO DE LIMA AMARAL
MARCELO SARAIVA CAVALCANTI**

**DESVENDANDO AS PREFERÊNCIAS DOS GAMERS: ANÁLISE PREDITIVA E
MERCADOLÓGICA DA INDÚSTRIA DE JOGOS**

Relatório final de pesquisa de Iniciação Científica apresentado à Assessoria de Pesquisa e Extensão.
Orientação: Abner Santos Belem

**BRASÍLIA
2025**

“Aprender é a única coisa que a mente nunca se cansa, nunca tem medo e nunca se arrepende”

Leonardo da Vinci

RESUMO

A presente pesquisa teve como objetivo central investigar as preferências dos jogadores em plataformas de jogos digitais, como a Steam, estabelecendo uma relação entre o feedback textual fornecido em suas avaliações e seu comportamento de jogo. Para tal, adotou-se uma abordagem metodológica quantitativa e exploratória, baseada na análise de dados extraídos da própria plataforma via APIS (*Application Programming Interface*). O estudo empregou uma série de técnicas de processamento e modelagem. A análise de sentimentos, utilizando o léxico VADER da biblioteca NLTK, revelou uma tendência significativa de avaliações positivas para os jogos mais populares (como maior número de avaliações no período estudado). Em seguida, a análise de redes foi utilizada para mapear os padrões de co-avaliação entre jogos, onde arestas foram ponderadas pela soma do tempo de jogo, evidenciando afinidades e comunidades de jogadores com interesses em comum. O passo mais significativo foi a construção de um modelo de classificação preditiva. O texto das avaliações foi vetorizado usando a técnica de TF-IDF, e um modelo de classificação K-Nearest Neighbors (KNN) foi treinado para prever se uma avaliação seria classificada como positiva (*voted_up*). Os resultados obtidos foram significativos e validaram a tese da pesquisa. O modelo de classificação alcançou métricas de desempenho relevantes, com uma Acurácia de 0.91, Precisão de 0.92, Recall de 0.99 e F1-Score de 0.95. Tais pontuações indicam que o texto das avaliações é um preditor altamente poderoso e confiável para determinar o sentimento do jogador. A contribuição da pesquisa reside na demonstração da eficácia de técnicas de *text mining* na compreensão do feedback de usuários em larga escala, oferecendo uma ferramenta robusta para o monitoramento da satisfação dos jogadores com os jogos que lhe foram apresentados, e é uma abordagem fundamental para decifrar as preferências de jogo.

Palavras-chave: preferências dos gamers; processamento de linguagem natural; modelagem preditiva; indústria de jogos; jogos eletrônicos.

SUMÁRIO

1	Introdução	5
1.1	Objetivo Geral e Específicos	7
2	Fundamentação teórica	9
2.1	Setor de Jogos	9
2.2	Comportamento do consumidor de jogos	9
2.3	Tecnologia	11
2.3.1	API - Interface de Programação de Aplicativos	11
2.3.2	<i>WebScraping</i>	11
2.3.3	<i>Machine Learning</i>	12
2.3.4	<i>Deep Learning</i>	13
3	Método	15
3.1	Tipificação	15
3.2	Caracterização do local de pesquisa	15
3.3	Objeto de estudo	15
3.4	Delimitação e universo da amostra	16
3.5	Instrumento de coleta ou de geração de dados	16
3.6	Procedimentos metodológicos	16
4	Resultados e discussão	18
4.1	Análise Descritiva	18
4.2	Análise de Sentimento via Processamento de Linguagem Natural – PLN	24
4.3	Aplicação de modelagem preditiva sobre textos de avaliação buscando estimar a aceitação dos jogos	25
5	Considerações finais	27
	REFERÊNCIAS	29

1 Introdução

Apesar de toda adversidade trazida pelo Covid-19, este surto acelerou mudanças no comportamento do consumidor e a adoção de recursos digitais de uma maneira que afetará as trajetórias de crescimento futuras (Clemente e Stoppa, 2020). Desde este acontecimento, a cada ano que se passa a sociedade dedica mais tempo, atenção e dinheiro a experiências de Entretenimento e Mídia - E&M, cada vez mais imersivas. E neste momento, observa-se uma estratégia interessante no setor de entretenimento com empresas tentando aproveitar e “surfando nesta onda”. Por exemplo, segundo a pesquisa global de entretenimento e mídia 2022–2026 (PwC, 2022, p. 8) “segmentos antes considerados de nicho, como o de jogos, estão ganhando mais destaque nos últimos anos, enquanto outros antes dominantes estão vendo suas posições competitivas caírem”.

Adicionalmente, dados da Pesquisa Game Brasil de 2024 (PGB, 2024) trazem informações pouco conhecidas pelo público em geral, como a de que 42,1% dos jogadores do Brasil são compostos por pais e mães, e não pelos filhos como esperado. E, 40,2% do público total tem ensino superior completo. Em paralelo, segundo Cardoso *et al.* (2023), o mercado de jogos no Brasil teve faturamento de R\$ 1,3 bilhão em 2022. Além disso, estima-se que no mesmo ano existiam 1.042 desenvolvedoras de jogos ativos no país, e 93% destas desenvolvem jogos de propriedade intelectual própria. A Figura 1, apresenta os tipos de jogos produzidos por categoria ao longo dos últimos anos.

JOGOS PRODUZIDOS POR CATEGORIA	2020	2021	2022
Jogos de entretenimento	36%	44%	58%
Advergames	9%	6%	18%
Jogos educacionais	35%	26%	17,5%
Outros tipos de jogos digitais	3%	5%	3%
Jogos de treinamento corporativo	15%	16%	2%
Simuladores com uso de hardware específico	2%	2%	1%
Jogos para saúde	1%	0%	0,5%
TOTAL DE JOGOS POR ANO	509	643	670

Figura 1: Pesquisa Capacidade de Produção da Indústria Brasileira de Games (2023)

Com o crescimento do mercado de jogos nos últimos anos, percebe-se uma possibilidade de desenvolvimento de pesquisas em um vasto campo de análise. No entanto, tendo em vista as pesquisas mencionadas, observa-se uma quantidade maior trabalhos e pesquisas voltados para o próprio desenvolvimento de jogos, e para a utilização de ferramentas mais modernas de linguagem computacional e inteligência artificial que possam auxiliar nesse desenvolvimento, e em contrapartida, observa-se uma menor quantidade de pesquisas voltadas para o desenvolvimento de ferramentas estatísticas, de *machine learning*/inteligência artificial e estratégias para analisar e compreender melhor as nuances e preferências do público-alvo que consome jogos eletrônicos (Kaufman, 2022).

O público de jogos digitais não é muito pesquisado, pois o setor ainda é considerado por muitos como um setor de trivialidades (Alexandre, 2018). No entanto, a indústria de jogos demonstra claros sinais de expansão (Clement e Statista, 2024), inclusive para públicos antes não alcançados como mulheres, crianças e idosos (Fleury; Nakano e Cordeiro, 2014).

Desse modo, este projeto visou responder o seguinte problema de pesquisa: Quais categorias e características de jogos digitais que contribuem para melhores perspectivas de novos produtos a serem comercialmente rentáveis e gerarem maior fidelização de clientes?

O presente projeto de pesquisa é de enorme importância para o setor de jogos digitais porque à medida que se mostre eficaz em apresentar soluções para que as preferências dos jogadores (consumidores) sejam mais bem apreendidas pelas empresas desenvolvedoras de jogos (fornecedores), contribuirá para o crescimento da indústria de games no Brasil (Amélio, 2018; Fleury; Nakano e Cordeiro, 2014).

Para a academia, este estudo demonstra sua relevância em face da proposta de utilizar, aprimorar e difundir os conhecimentos ministrados em diversas disciplinas do curso de Ciência de Dados, relativos a métodos estatísticos preditivos, linguagens de programação e inteligência artificial nos seus subcampos de *machine learning* e *deep learning* via redes neurais. O que carrega consigo grande significado em vista da constatação de Hoenen, *et al.* (2024, p. 1) de que “[...] há evidências que a academia perde espaço na pesquisa sobre IA para as big techs e o mundo corporativo. Essa questão ganha relevância à luz da trajetória de figuras influentes como Geoffrey Hinton,

que transitou da academia para uma carreira próspera no setor privado”. Portanto, este trabalho ganha mérito ao treinar, fortalecer e difundir o conhecimento para o meio acadêmico.

Para a sociedade, dentre vários aspectos, a influência positiva do presente trabalho pode ser vislumbrada pelo retorno que traz ao público brasileiro na medida em que consiga mostrar aos consumidores de jogos digitais nacionais quais características e tipos de jogos que atendem suas vontades e necessidades da melhor maneira. O que contribuirá para maior diversidade de jogos a serem produzidos por desenvolvedores nacionais.

Por fim, a elevação na oferta de jogos nacionais, aliada à atual importação de jogos estrangeiros, possibilitará o aumento do leque de jogos disponíveis aos gamers. O que ganha maior escopo quando se percebe que os tipos de jogos produzidos no Brasil em 2022, além dos voltados para entretenimento, 17,5% eram de jogos educacionais (Cardoso *et al.*, 2023). O perfil completo dos jogos produzidos no Brasil pode ser visualizado na Figura 2:

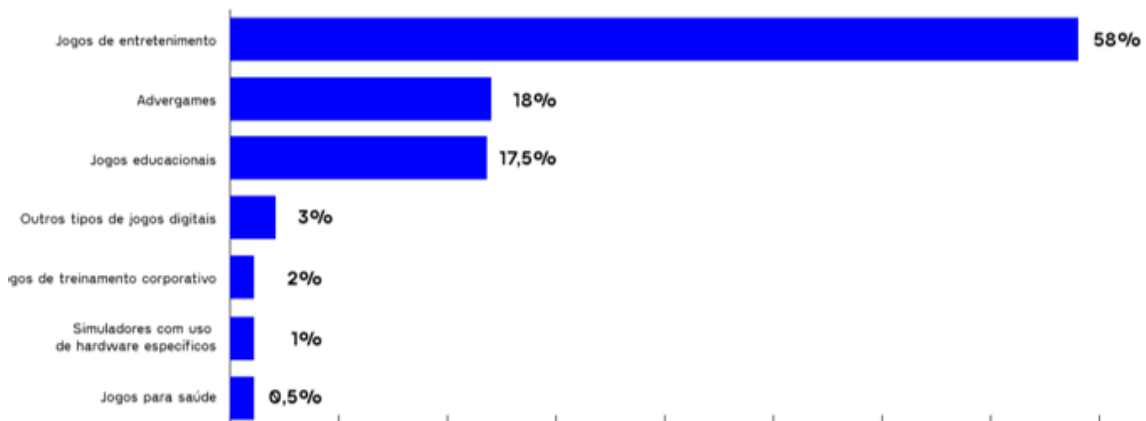


Figura 2: Pesquisa Capacidade de Produção da Indústria Brasileira de Games (2023)

1.1 Objetivo Geral e Específicos

Tendo em vista o contraste entre o elevado crescimento do mercado de jogos digitais no Brasil e o menor número de trabalhos acadêmicos/científicos sobre as motivações que suportam esta elevada demanda, o objetivo desta pesquisa foi o de mapear, através de modelos preditivos, as características e categorias de jogos digitais

que contribuem para melhores perspectivas de novos produtos a serem comercialmente rentáveis e gerarem maior fidelização de clientes. O futuro atingimento deste objetivo busca responder o problema de pesquisa e com seus resultados auxiliar decisões do segmento empresarial de desenvolvimento de jogos nacionais.

Para que este objetivo seja alcançado vislumbrou-se o atingimento de três objetivos específicos:

1. Analisar as características demográficas e de consumo dos jogadores brasileiros: Este objetivo busca identificar e analisar os padrões de comportamento dos jogadores, incluindo faixa etária, gênero, educação e frequência de jogo.

2. Desenvolver modelos preditivos para identificar as preferências de jogo e a viabilidade comercial de novos produtos: Utilizando técnicas avançadas de *Machine Learning* e *Deep Learning*, este objetivo visou criar modelos que possam prever quais características de jogos (como gênero, tema, mecânicas de jogo) são mais atraentes para diferentes segmentos de jogadores.

3. Investigar o impacto das tecnologias emergentes na personalização da experiência de jogos digitais: Explorar como as inovações tecnológicas, como inteligência artificial, podem ser utilizadas para criar experiências de jogo mais envolventes e personalizadas.

2 Fundamentação teórica

2.1 Setor de Jogos

Apesar de sua aceitação crescente, o público de jogos digitais não é muito pesquisado, pois o setor ainda é considerado por muitos como um setor de trivialidades, porém é uma área extremamente ampla a ser explorada. O setor de jogos digitais vem sendo cada vez mais aceito pela sociedade (Alexandre, 2018).

Conforme Fleury, Nakano e Cordeiro (2014, p. 32), “Um dos motivos do crescimento da importância dos jogos digitais é que, atualmente, eles não são consumidos somente por jovens do sexo masculino, como se pensa tradicionalmente, mas também por crianças, mulheres e idosos. Além disso, o uso dos Jogos Digitais e as tecnologias que tal setor desenvolve ultrapassaram o entretenimento, adquirindo caráter “sério”, ao serem incorporados a atividades de educação, a pesquisas científicas, a treinamentos (desde o âmbito corporativo até a Defesa Nacional), além de terem utilização na capacitação relacionada ao atendimento de saúde, na escolha e desenvolvimento de vocações, e na arquitetura e construção civil”.

Aliado ao incremento na diversidade de público, é um mercado em ampla expansão. Segundo dados da Revista VEJA, era esperado que no ano de 2023 este setor ultrapassasse a marca de 3,3 bilhões de jogadores - quase metade da população mundial (VEJA, 2023). Além disso, estima-se que a receita do mercado de jogos digitais em todo o mundo atingiu no ano de 2022 o patamar de 347 bilhões de dólares americanos. E estima-se que em 2029 atingirá a cifra de US\$ 666,69 bilhões (Clement e Statista, 2024).

2.2 Comportamento do consumidor de jogos

Urban *et al.* (2019) argumenta que a capacidade de prever a escolha do cliente é o primeiro passo para melhorar as decisões que envolvem o design do produto, a alocação de recursos de mídia, como promover o produto e quem atingir. Saber o que as pessoas mais valorizam requer experimentação e modelagem de escolha preditiva.

Segundo estes autores, uma das vantagens de utilizar métodos baseados em aprendizado profundo de máquina se dá porque enquanto os métodos estatísticos tradicionais utilizam a capacidade de descrever relações históricas como critério de sucesso, a aprendizagem profunda, por outro lado, utiliza a capacidade de prever as

escolhas e respostas dos consumidores através de novos dados, o que melhora a probabilidade de que as previsões realmente ocorram (Urban *et al.*, 2019).

Capturar com precisão as necessidades e dos usuários é fundamental para melhorar a competitividade dos fabricantes de jogos no mercado (Cui, Zirui. 2023). Segundo estes autores, os produtores pensam em como obter mais lucros, mas ignoram a qualidade do jogo em si. Isso se reflete no tipo de jogo preferido dos fabricantes de jogos online: o jogo de compra no aplicativo. Por isso, a produção de videogames apresenta problemas em razão da falta de construção teórica e da falta de observação dos fatores de influência que interferem na demanda no setor.

Segundo Corey *et al.* (2021) entender a motivação do jogador é indiscutivelmente valioso para aqueles que desejam ter sucesso na crescente e altamente competitiva indústria de desenvolvimento de jogos.

Nesse contexto, empresas precisam se valer de modelos que busquem captar as motivações de jogadores. Estudos nesse sentido vem sendo propostos há anos, como o modelo proposto por Yee N. (2006), que em seu modelo apontou 10 subcomponentes de motivação que se agruparam em três componentes abrangentes: conquista, social e imersão.

Como exemplo, Bertens *et al.* (2018) perceberam que players geram enormes quantidades de dados, pois tudo o que eles fazem dentro de um jogo *free-to-play* é registrado. Neste tipo de plataforma, os jogadores podem selecionar ou comprar uma profusão de itens durante o jogo para progredir e aproveitar totalmente sua experiência. Nesse contexto, pode-se usar um sistema de recomendação para apresentar aos jogadores itens que podem ser interessantes para eles. Em sua pesquisa testaram dois tipos de modelos, um modelo baseado árvores aleatórias e outro em redes neurais. Ambos capazes de ajudar os desenvolvedores a melhorarem o jogo e, também, ajudarem a fazer recomendações personalizadas durante o jogo que se mostram atraentes para os jogadores.

Em resumo, ao permitir uma adaptação mais ágil às mudanças no comportamento do consumidor, sendo capaz de aprender e se adaptar continuamente com novos dados, há um aumento significativo na perspectiva de novos produtos serem comercialmente rentáveis e, também, gerarem maior fidelização de clientes.

2.3 Tecnologia

No presente trabalho de iniciação científica, dentre as principais tecnologias, pretendeu-se utilizar para a captação de dados ferramentas como APIs (*Application Programming Interface*) e *Web Scraping*. E para a análise e utilização de modelos preditivos técnicas de *Machine Learning* e *Deep Learning*. Segue abaixo melhor detalhamento:

2.3.1 API - Interface de Programação de Aplicativos

Uma API é um conjunto de regras, protocolos e ferramentas que permitem a comunicação entre diferentes aplicativos. Ela define a maneira pela qual os componentes de software devem interagir. As APIs são usadas para permitir que diferentes sistemas, plataformas ou aplicativos se comuniquem entre si de forma eficiente. Até o ano de 2000, não havia um padrão sobre como o design de API deveria ser. A integração necessária, uso de protocolos, como SOAP, era notoriamente conhecido como algo complexo a ser construído e manuseado (Schulthess, 2017).

Com uma API é possível que um aplicativo receba e envie informações para outro aplicativo ou uma plataforma de serviços de forma totalmente programável através de requisições HTTP, onde o aplicativo principal envie uma solicitação para um servidor de API e este servidor responda com os dados solicitados. De acordo com modelagem de REST, as requisições HTTP (Fielding, 2004) podem ter alguns recursos, tais como (Get, Post, Put, Delete) onde cada um tem sua respectiva função, como obter, publicar, inserir e deletar.

2.3.2 *WebScraping*

Na era da informação digital, há uma variedade de métodos para automatizar a busca por informações na internet, alguns dos quais remontam aos primórdios da própria rede. Um termo amplamente conhecido e em ascensão nos dias de hoje é o *webscraping*. O método envolve a criação de scripts organizados para percorrer o código-fonte das páginas da web desejadas e extrair as informações consideradas

pertinentes para a pesquisa em questão (Graciano, 2023). Essa técnica tem sido cada vez mais valorizada no contexto da pesquisa científica, proporcionando acesso rápido a dados relevantes que podem ser utilizados para análises e estudos acadêmicos.

Apesar de existirem preocupações éticas e legais associadas ao *webscraping*, é sempre bom seguir conforme o termo de serviço do sistema por meio do qual será aplicado o método (Oliveira, 2023), assim, é possível utilizar essa tecnologia de forma legal e benéfica. Quando empregado corretamente, o *webscraping* pode contribuir significativamente para a obtenção de dados necessários para pesquisas científicas, permitindo a coleta e análise das informações de maneira eficiente e abrangente.

O uso deste método pode servir como uma ferramenta poderosa para coletar dados relevantes, fundamentar teoricamente o estudo e fornecer insights valiosos para a pesquisa em questão. Desde a captação de tendências em redes sociais até a extração de dados de publicações acadêmicas, o *webscraping* oferece uma gama de possibilidades para impulsionar o progresso científico e aprofundar o conhecimento em diversas áreas do saber (Graciano, 2023).

2.3.3 Machine Learning

Uma pesquisa realizada com dados da Hungria, no ano de 2022, utilizando técnicas de *machine learning* com modelos de classificação em algoritmos conhecidos com ‘máquinas de vetores de suporte - SVMs’ e ‘árvores de decisão - Random Forest’ apresentou resultados importantes que contribuem para a análise do mercado de gamers (Gosztonyi, 2023).

Gosztonyi (2023) chegou a algumas conclusões, como: jogadores ocasionais preferem jogos de cartas, palavras e tabuleiro, e conseguem os jogos através de amigos/conhecidos, além de possuírem renda alta, e pertencem a uma faixa etária mais avançada. Por outro lado, os jogadores frequentes adotam uma vasta gama de jogos, desde jogos de ação a jogos de estratégia, bem como jogos de azar e esporte. Normalmente, jogadores mais frequentes, obtêm seus jogos de sistemas ilegais de compartilhamento de arquivos ou de lojas online, jogam não apenas para recreação, mas também para autoaperfeiçoamento, sensação de sucesso, e desenvolvimento de

suas habilidades e criatividade. E, predominantemente, estão na faixa etária de solteiros e adultos jovens.

Além disso, foi comprovado que na Hungria os jogadores frequentes eram majoritariamente jovens adultos do sexo masculino, com um elevado nível de educação e status no mercado de trabalho. No entanto, observou-se que os jogadores ocasionais tinham uma estrutura de rendimentos e um status no mercado de trabalho ainda mais elevados (Gosztonyi, 2023).

2.3.4 Deep Learning

A aprendizagem profunda, ou *deep learning*, é um subcampo da inteligência artificial que tem transformado significativamente várias indústrias através de suas capacidades avançadas de modelagem e previsão. Segundo LeCun, Bengio e Hinton (2015), os avanços em *deep learning* permitem que modelos computacionais compostos por múltiplas camadas de processamento aprendam representações de dados com múltiplos níveis de abstração. Essa capacidade de extrair automaticamente características relevantes dos dados, sem a necessidade de intervenção humana direta, torna a aprendizagem profunda particularmente eficaz em tarefas complexas de percepção e classificação (Schmidhuber, 2015). No contexto da indústria alimentícia, a implementação de técnicas de *deep learning* tem demonstrado ser crucial para otimizar processos e adaptar produtos às preferências detalhadas do consumidor, como evidenciado pelo estudo de Eunyong *et al.* (2023).

Em um estudo recente Eunyong *et al.* (2023), observou a aplicação de técnicas de aprendizagem profunda para a análise das preferências dos consumidores de carne suína na Coreia do Sul, um mercado que tem visto um aumento significativo no consumo nos últimos anos. Utilizando equipamentos de ultrassom para captar características específicas da carne, como a proporção de carne magra e a distribuição de músculo e gordura, os pesquisadores empregaram modelos baseados em redes neurais profundas para prever os escores de preferência de sabor e aparência dos consumidores.

A abordagem adotada permitiu uma análise detalhada e a longo prazo das preferências dos consumidores, baseada em dados sensoriais e físicos da carne. A precisão dos modelos de *deep learning* alcançou 75,3%, superando a eficácia da

avaliação manual por uma margem de 5,5%. Este avanço não apenas destaca a eficiência das técnicas de aprendizagem profunda em fornecer previsões mais acuradas, mas também sublinha o potencial de tais metodologias para otimizar processos produtivos e de distribuição no setor alimentício (Eunyoung *et al.*, 2023).

Os resultados do estudo indicaram que as carcaças de porco com altos escores de sabor, geralmente com maior teor de gordura, eram preferencialmente destinadas a restaurantes para processamento, enquanto aquelas com alta pontuação de aparência e, portanto, menor teor de gordura, eram mais adequadas para venda direta em mercados (Eunyoung *et al.*, 2023). Este sistema não só melhora a alocação e distribuição dos produtos de acordo com a demanda do consumidor, mas também oferece vantagens econômicas significativas, aumentando a eficiência na cadeia de suprimentos e potencializando o aumento das vendas tanto para empresas quanto para produtores.

A aplicação de *deep learning*, portanto, representa uma ferramenta valiosa na tomada de decisões baseada em dados concretos, permitindo a personalização da produção e a oferta de produtos que atendem especificamente às preferências do consumidor. A precisão e a capacidade de adaptação desses modelos tornam-nos especialmente relevantes para indústrias que dependem de um entendimento aguçado das necessidades e desejos dos clientes, como é o caso da indústria de jogos eletrônicos. Em nosso contexto, pretendia-se explorar essas técnicas para prever as preferências dos jogadores, visando a criação de jogos que não apenas engajem, mas também sejam comercialmente bem-sucedidos.

3 Método

3.1 Tipificação

A pesquisa é de natureza quantitativa, pois se baseia na análise de dados numéricos adquiridos da plataforma da Steam como *playtime_forever* (tempo total de jogo), *playtime_at_review* (tempo jogado até o momento da avaliação), *voted_up* (avaliação positiva ou não), *votes_up* (avaliação considerada útil ou não), *appid* (id do jogo), *name* (nome do jogo), *ano* (ano da avaliação sobre o jogo), dentre outras variáveis, e na aplicação de modelos estatísticos e de aprendizado de máquina. A abordagem é exploratória e descritiva, buscando identificar e descrever as relações entre as variáveis, como a análise de sentimento das avaliações e o tempo de jogo. O(s) procedimento(s) técnico(s) envolveram também processamento de linguagem natural (PLN) utilizando o léxico VADER da biblioteca NLTK, estatística descritiva e modelagem preditiva por meio da aplicação da técnica técnicas como TF-IDF para extração de entidades do texto, aliada ao algoritmo K-Nearest Neighbors (KNN).

3.2 Caracterização do locús de pesquisa

O local da pesquisa é a plataforma Steam, um serviço de distribuição digital de jogos, pertencente à Valve Corporation. O ambiente virtual da plataforma serviu como fonte para a coleta de dados de avaliações de usuários e metadados de jogos. A pesquisa se concentrou na área de engajamento e feedback de usuários, analisando como as opiniões expressas influenciam e se correlacionam com as métricas de uso da plataforma.

3.3 Objeto de estudo

O objeto de estudo é a relação entre o feedback textual dos jogadores, manifestado em suas avaliações, e seu comportamento de jogo, medido por variáveis como tempo de jogo (*playtime_forever* e *playtime_at_review*), seu grau de engajamento utilizando a variável *votes_up*, a análise de sentimento sobre os jogos mais jogados da amostra utilizando as variáveis *voted_up* e *review_text*. Bem como se

procedeu à análise preditiva utilizando também as mesmas variáveis *review_text* e *voted_up*.

3.4 Delimitação e universo da amostra

Devido a dificuldades de ordem técnica e capacidade de armazenagem, não foi possível abranger todas as avaliações de jogos publicadas na plataforma da Steam. A base de dados ficou restrita ao período entre 25/05/2010 a 10/10/2025 (aproximadamente dez anos e 4 meses) para todas as avaliações publicadas na plataforma Steam para os jogos com as seguintes características: os 10 mais vendidos, os 10 especiais, os 10 lançamentos e os 20 mais aguardados. E filtrada para incluir apenas avaliações escritas em inglês. O universo da amostra final é composto por 854.505 avaliações.

3.5 Instrumento de coleta ou de geração de dados

Os dados foram coletados de forma secundária, por meio de extrações de dados via *endpoints* de APIs fornecidas pela própria plataforma de jogos da Steam, e baixados no formato de arquivos CSV. O principal instrumento de análise foi a linguagem Python, utilizando bibliotecas como *pandas*, *intertools*, e *collections* para manipulação de dados, *nltk* para Processamento de Linguagem Natural - PLN e *scikit-learn* (*TfidfVectorizer* e *NeighborsClassifier*) para modelagem estatística e predição.

3.6 Procedimentos metodológicos

O desenvolvimento da pesquisa seguiu um passo a passo detalhado:

1. Pré-processamento de Dados;
2. Análise descritiva sobre desempenho dos jogos frente às preferências dos jogadores;
3. Análise de Sentimento via Processamento de Linguagem Natural (PLN);
4. Aplicação de Modelagem preditiva sobre textos de avaliação buscando estimar a aceitação dos jogos;

Os dados brutos dos dois arquivos CSV `steam_store_app_reviews.csv` (862.275 linhas e 26 colunas) e `steam_apps_list.csv` (261.778 linhas e 4 colunas) foram lidos e unidos em um único *dataframe*. Valores duplicados foram removidos, nomes de jogos foram padronizados, e as avaliações foram filtradas para o idioma inglês.

Foram criados diversos gráficos utilizando bibliotecas `matplotlib`, `seaborn`, `plotly`, `networkx`, `intertools` e `collections` para mostrar tendências quanto a jogos com mais avaliações (mais populares), mais bem avaliados, com maior engajamento entre comunidades de jogadores, que tiveram maior capacidade de gerar fidelização do cliente, e mais interligados.

Utilizando o léxico VADER da biblioteca NLTK, cada texto de avaliação foi classificado em Positiva, Negativa ou Neutra com base no `compound score`, uma pontuação de polaridade do texto. Verificando se os cinco jogos com maior número de avaliações no período também obtiverem maior classificação de sentimentos positivos

Vetorização TF-IDF: A técnica de Term Frequency-Inverse Document Frequency (TF-IDF) foi aplicada ao texto das avaliações (coluna `review_text`). O `TfidfVectorizer` do `scikit-learn` foi configurado para remover stopwords em inglês e limitar o vocabulário às 1000 palavras mais importantes, gerando uma matriz TF-IDF esparsa de (854505, 1000) que serviu como base para os modelos.

Modelagem Preditiva (KNN): O dataset foi dividido em conjuntos de treino e teste (80/20). O algoritmo `KNeighborsClassifier` foi utilizado para construir um modelo de classificação com base na matriz TF-IDF. O modelo foi treinado para prever a variável `voted_up` (avaliação positiva), e sua performance foi avaliada usando métricas de classificação.

4 Resultados e discussão

As observações e análises realizadas revelaram que o feedback dos jogadores é um recurso valioso para compreender o engajamento na plataforma. Os resultados abaixo serão expressos em gráficos e métricas que sustentam as discussões teóricas. Os resultados serão apresentados em três tópicos principais:

- Análise descritiva sobre desempenho dos jogos frente às preferências dos jogadores;
- Análise de Sentimento via Processamento de Linguagem Natural - PLN;
- Aplicação de modelagem preditiva sobre textos de avaliação buscando estimar a aceitação dos jogos.

4.1 Análise Descritiva

Como informado anteriormente, a amostra analisada abrangeu o período entre 25/05/2010 a 10/10/2025 (aproximadamente dez anos e 4 meses) para todas as avaliações publicadas na plataforma Steam para jogos com as seguintes características: os 10 mais vendidos, os 10 especiais, os 10 lançamentos e os 20 mais aguardados. Foi filtrada para incluir apenas avaliações escritas em inglês. De forma que o universo da amostra final é composto por 854.505 linhas de avaliações.

Desse conjunto, apresentamos abaixo os cinco jogos com mais avaliações entre os anos de 2020 até 2025, são eles: Cyberpunk 2077, It Takes Two, Peak, Dead by Daylight, e Battlefield. Cabe ressaltar que não se trata dos mais bem avaliados, mas sim os mais populares, que foram objeto de mais avaliações no período, sejam positivas ou negativas.

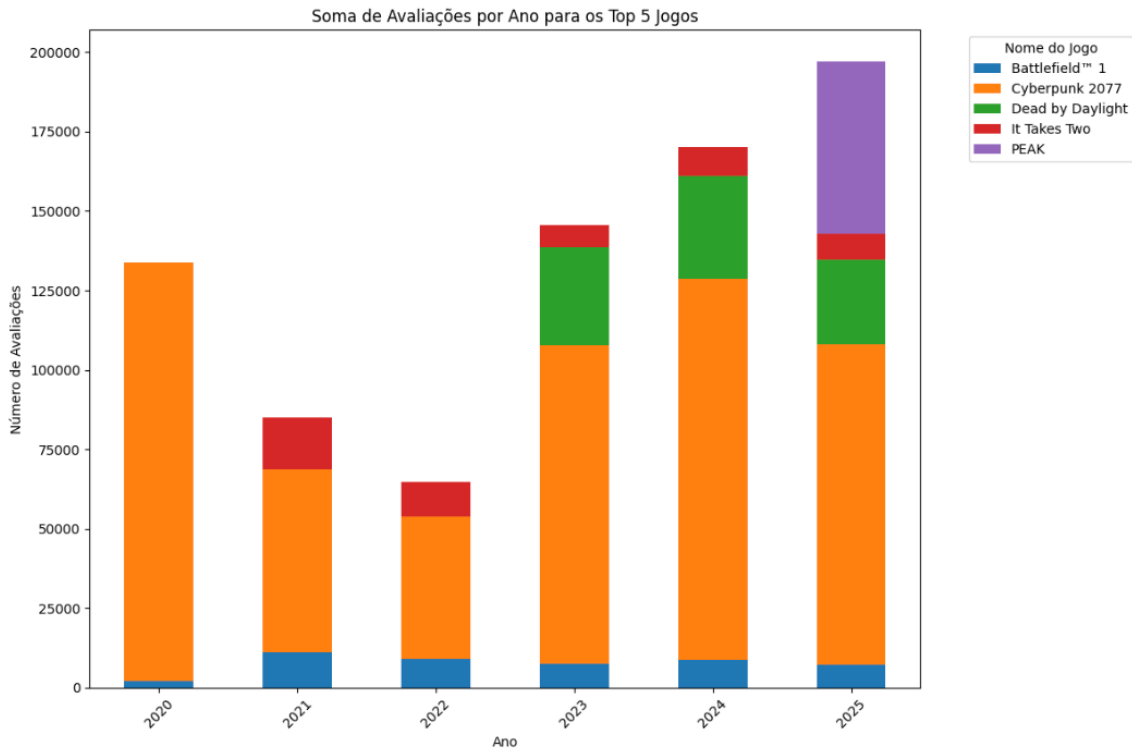


Figura 3: Jogos mais avaliados (mais populares) entre os anos de 2020 a 2025

Por outro lado, os cinco que recebem mais votos positivos (variável `voted_up`) no mesmo período tendem a ser os mesmos, saindo da lista anterior o jogo Battlefield e entrando em seu lugar o jogo STAR WARS, conforme a Figura 4. Ou seja, jogos preferidos dos *gamers*, tendem a se manter na lista dos mais comentados.

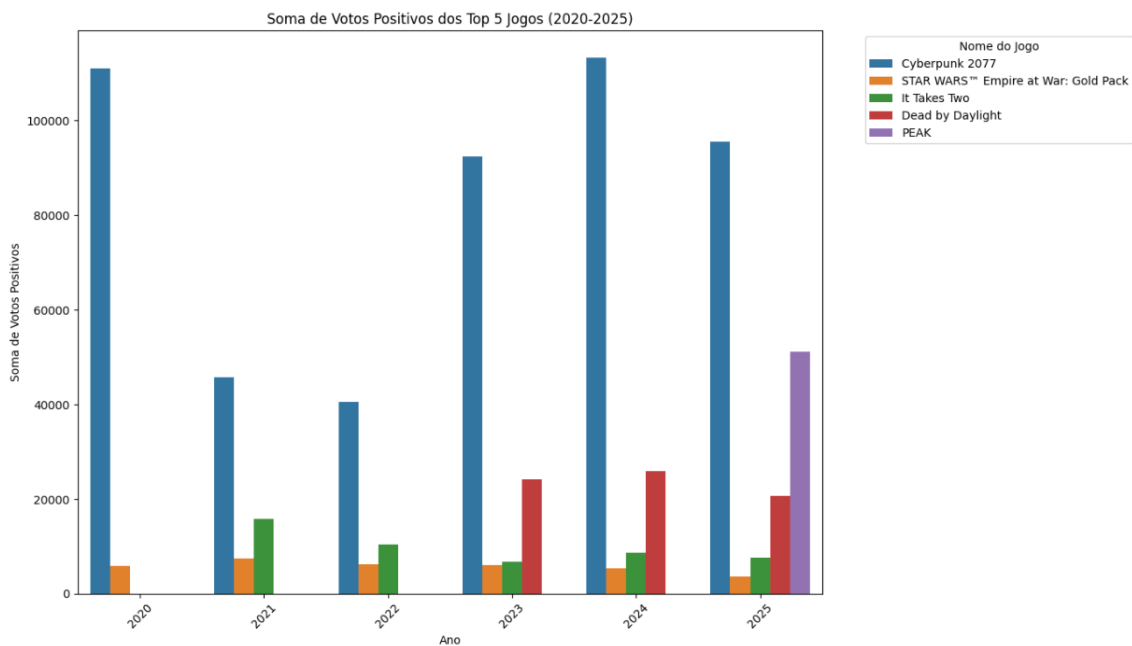


Figura 4: Jogos mais bem avaliados no período

Nesse sentido, a base traz a variável *votes_up*, que indica quantos jogadores consideraram útil determinada avaliação. É interessante utilizar essa variável como uma *proxy* de quais jogos geram maior engajamento, têm mais alcance entre as diversas comunidades de gamers. E, novamente, a Figura 5 demonstra que os mesmos jogos são os que apresentam mais alcance.

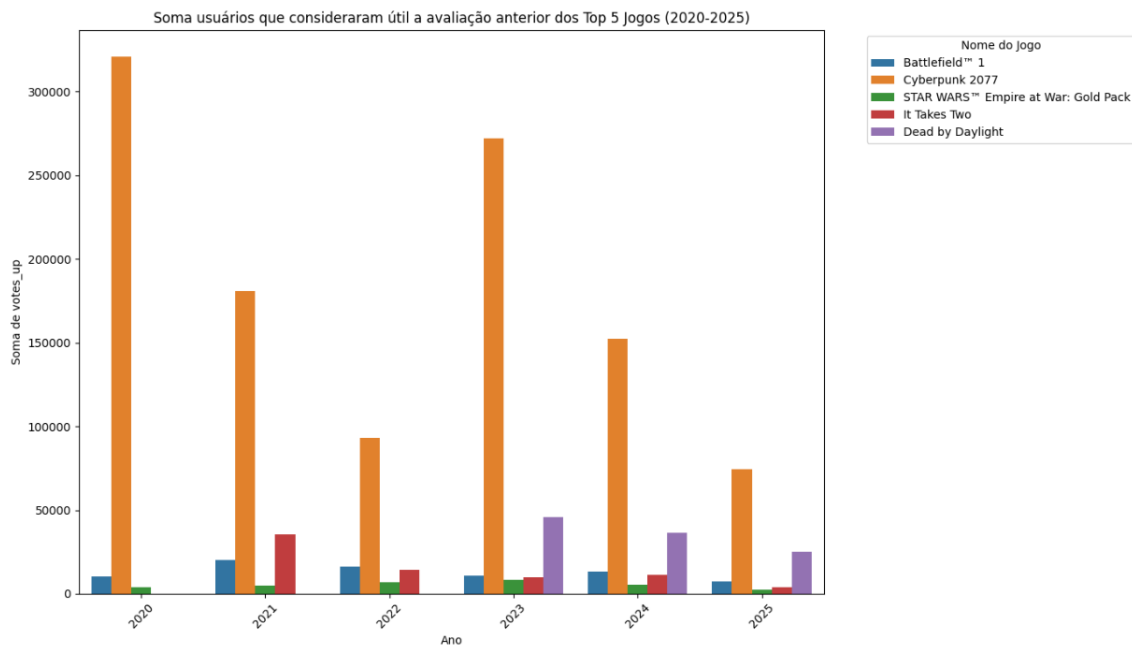


Figura 5: Jogos que geraram mais engajamento no período

Por outro ângulo, procedeu-se também à análise de redes que revelou padrões de afinidade entre os jogadores, mostrando que o comportamento de co-avaliação não é aleatório. Os dados mostram o quanto as comunidades de cada jogo interagem no que diz respeito aos jogos considerados mais populares. Isto pode ser visto na Figura 6, que apresenta grafos, onde cada aresta indica o número de vezes que um mesmo usuário avaliou dois jogos diferentes, os quais são representados por cada nó (círculo). Foi adicionado um filtro para que fosse mostradas arestas com no mínimo 300 co-avaliações.

Grafo de co-avaliação entre jogos

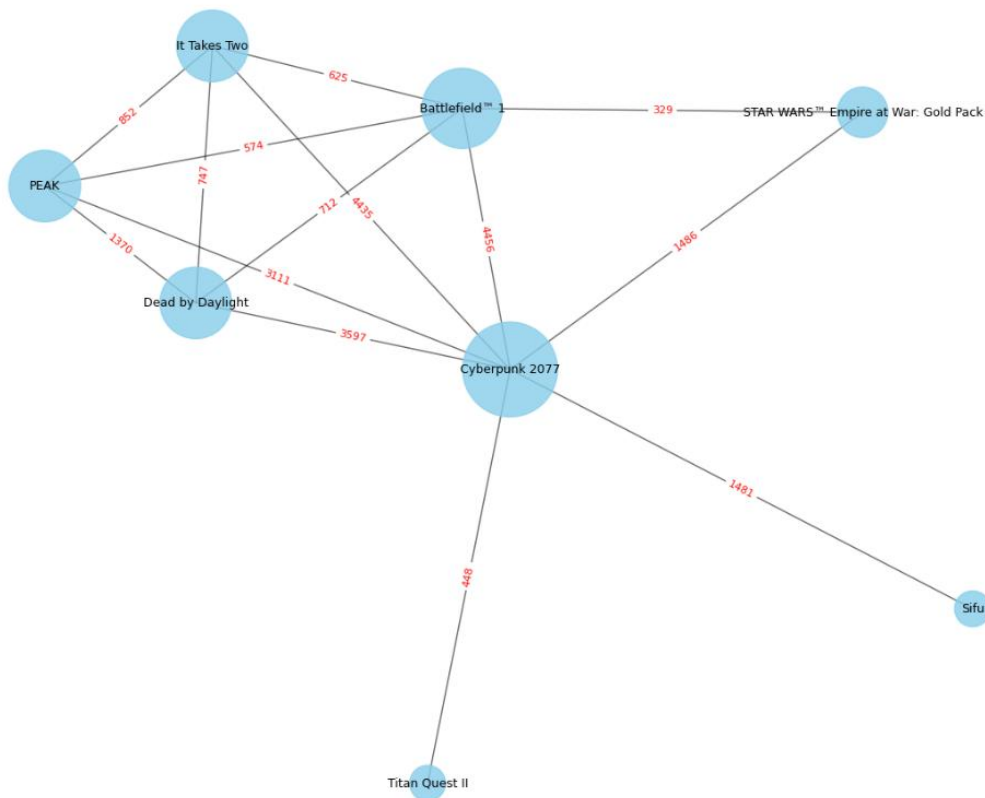


Figura 6: Gráfico de co-avaliação entre os jogos mais avaliados

Pode-se perceber, que os jogos mais populares citados anteriormente, também são os que geraram maior número de co-avaliações no período. Por exemplo, 4.456 pessoas avaliaram simultaneamente os jogos Cyberpunk 2077 e Battlefield, enquanto apenas 448 pessoas avaliaram simultaneamente Cyberpunk 2077 e Titan Quest II. O que é corroborado pela Figura 7, onde se verifica que enquanto a soma de horas jogadas simultaneamente entre Cyberpunk 2077 foi de 971.877h, entre Cyberpunk 2077 e Titan Quest II somou-se apenas 76.172 horas jogadas simultaneamente pelos mesmos usuários.

Soma de Tempo de Jogo para co-avaliações (acima de 10000 horas)

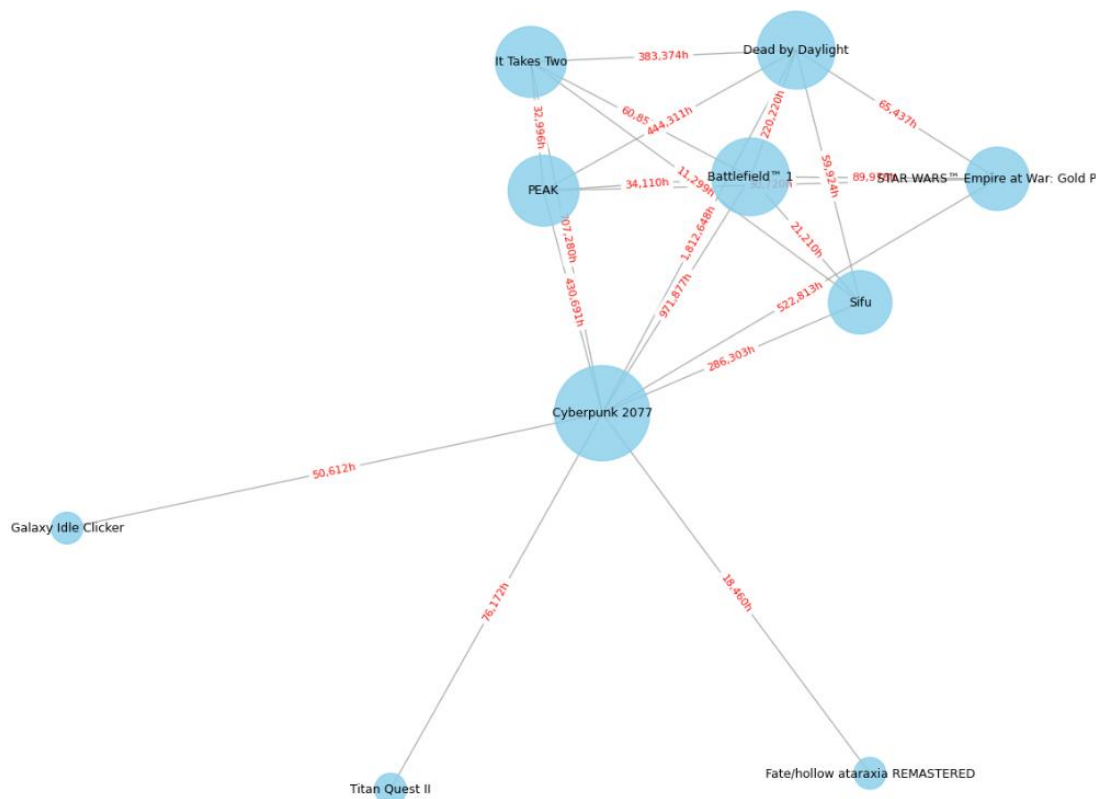


Figura 7: Soma de tempo de jogo para co-avaliações entre mesmos jogadores

Noutro sentido, procuramos verificar quais jogos da amostra geraram maior fidelização. Para isto, procedemos à técnica de feature_engineer criando a variável *fidelização_jogo*, que representa a razão entre as variáveis *playtime_forever* (que indica o tempo total jogado) e *playtime_at_review* (tempo total jogado até o momento da avaliação).

Quanto maior o resultado, resulta a compreensão de que o jogador se manteve jogando mesmo após a avaliação textual que fez do jogo. Indicando o potencial deste em “fidelizar” os gamers após o estágio inicial de conhecimento e avaliação. A Figura 8 apresenta os cinco jogos de toda a base que conseguiram melhores fidelizações.

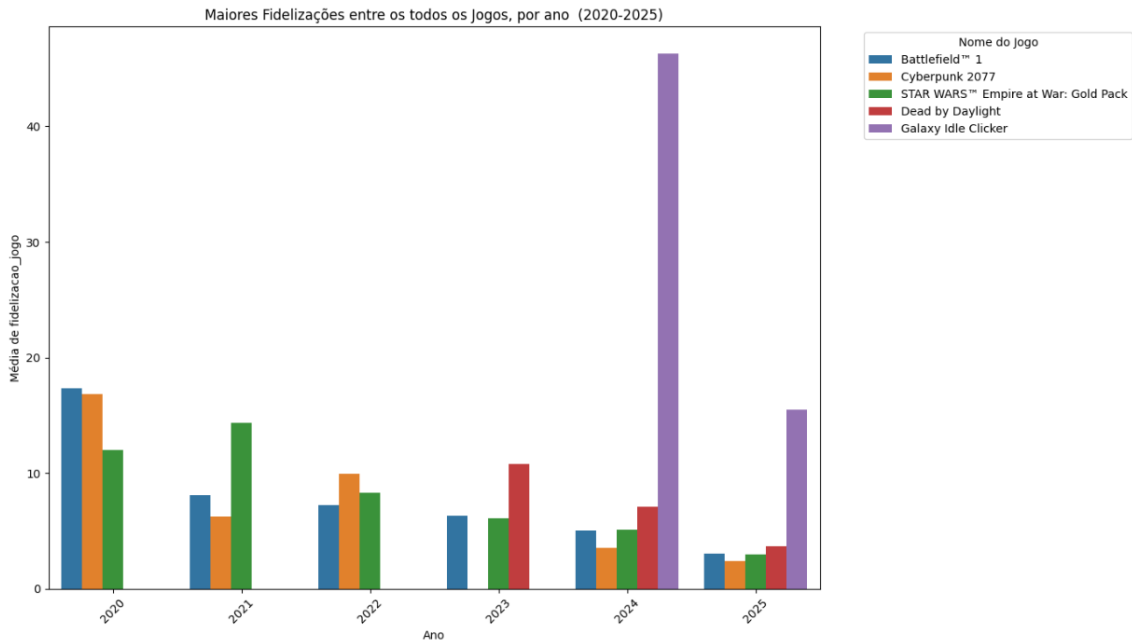


Figura 8: Jogos que geraram melhores fidelizações de gamers no período

Na Figura 8 ganha destaque o jogo Galaxy Idle Clicker, que não constava anteriormente dos mais avaliados ou como melhores avaliações. Provando que não somente os jogos mais populares são os que geram maior fidelização dos clientes. Talvez por ser um jogo oferecido de forma gratuita na plataforma (com possibilidade de compras dentro do game), manuseado à base de cliques e ações simples. E apresentando como chamativos progressões de cenário e de fases.

O que é reforçado pela Figura 9, que apresenta os jogos com maiores médias de horas jogadas (*playtime_forever*) entre 2020 e 2025, onde o Galaxy Idle Clicker se destaca novamente.

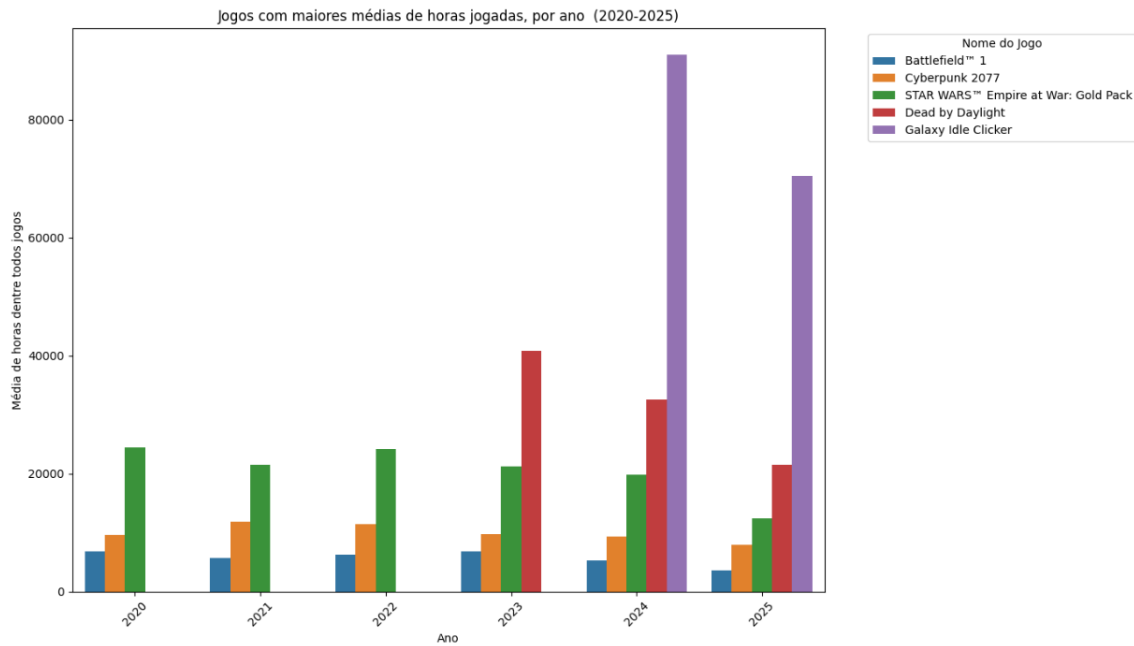


Figura 9: Jogos com mais horas jogadas no período

4.2 Análise de Sentimento via Processamento de Linguagem Natural – PLN

Foi procedida a análise de sentimento utilizando o léxico VADER da biblioteca NLTK do Python, sobre os cinco jogos com maior número de avaliações no período. Por meio dessa técnica, cada texto (*review_text*) de avaliação foi classificado em Positivo, Negativo ou Neutro com base no *compound score*, uma pontuação de polaridade do texto. Os resultados são apresentados na Tabela 1.

Distribuição de Sentimentos

Jogo	Sentimento da Avaliação		
	Positivo	Neutro	Negativo
Cyberpunk 2077	384.456	105.693	64.874
Dead by Daylight	45.297	24.781	19.516
PEAK	29.531	19.777	5.013
It Takes Two	40.771	7.354	3.328
Battlefield™ 1	26.625	10.519	8.803

Tabela 1: Análise de Sentimento dos cinco jogos com maior número de avaliações registradas no período

O modelo baseado na análise dos textos de avaliação postados pelos jogadores se mostrou notavelmente eficaz em prever se uma avaliação seria positiva (*voted_up*).

As métricas de desempenho do modelo foram:

- Acurácia: 0.91;
- Precisão: 0.92;
- Recall: 0.99;
- F1-Score: 0.95.

Esses valores validam a tese de que o texto da avaliação (*review_text*) é um preditor poderoso do sentimento do jogador.

A Acurácia de 0.91 significa que o modelo de classificação acertou a previsão do sentimento da avaliação em 91% das vezes. Em outras palavras, de todas as avaliações no conjunto de teste, o modelo foi capaz de classificar corretamente a grande maioria, tanto as que eram positivas quanto as que eram negativas. É uma pontuação alta e indica que o modelo está com um bom desempenho geral.

O alto Recall (0.99) é particularmente notável, pois indica que o modelo é excelente em identificar todas as avaliações que são realmente positivas, com uma alta taxa de acerto. Em termos mais simples, o Recall responde à pergunta: "De todas as avaliações que são realmente positivas, quantas o meu modelo identificou corretamente?".

O valor de 0.92 para a Precisão significa que, quando o modelo previu que uma avaliação era positiva, ele acertou 92% das vezes. Indicando que o modelo tem uma baixa taxa de "falsos positivos", ou seja, ele raramente confunde uma avaliação negativa com uma positiva.

O F1-Score de 0.95 é uma pontuação muito alta. Este é a média harmônica da Precisão e do Recall. Um valor de 0.95 demonstra que, para a tarefa de classificar avaliações positivas, o modelo não apenas acerta a maioria de suas previsões (Precisão: 0.92), mas também é eficaz em encontrar quase todas as avaliações que são realmente positivas (Recall: 0.99).

Depreende-se que a análise do texto das avaliações postadas pelos gamers, mesmo se apresentando como um modelo relativamente simples, é uma ferramenta robusta para a compreensão do feedback do usuário.

5 Considerações finais

A presente pesquisa atingiu seu objetivo central de investigar as preferências dos jogadores em plataformas digitais, estabelecendo relações claras entre o conteúdo textual das avaliações e o comportamento de consumo na plataforma Steam. A utilização de métodos quantitativos, aliados a técnicas avançadas de Processamento de Linguagem Natural e modelagem preditiva, permitiu não apenas mapear padrões de comportamento, mas também demonstrar a viabilidade do uso de modelos como o K-Nearest Neighbors, alimentados por representações TF-IDF, para prever com alta acurácia o sentimento dos jogadores. O desempenho obtido, com métricas robustas de acurácia, precisão, recall e F1-Score, confirma a relevância do texto das avaliações como preditor confiável da satisfação do usuário, reforçando o potencial de tais abordagens para o monitoramento automatizado da experiência do cliente em larga escala.

Além disso, os resultados corroboram a hipótese de que comunidades de jogadores apresentam padrões consistentes de coavaliação e fidelização, revelando que as preferências não se limitam aos títulos mais populares, mas também se manifestam em jogos específicos capazes de engajar nichos e manter alto tempo médio de jogo. A análise de redes e a mensuração de indicadores como o índice de fidelização trouxeram insights relevantes para estratégias de segmentação e retenção no mercado de jogos, alinhando-se aos objetivos específicos propostos no início da investigação.

Entretanto, algumas limitações devem ser reconhecidas. A coleta de dados esteve restrita ao idioma inglês, o que pode ter excluído nuances culturais e linguísticas relevantes do público brasileiro e de outros mercados. Da mesma forma, o recorte temporal e a seleção de categorias específicas de jogos limitaram a abrangência das conclusões, podendo não representar integralmente o universo da plataforma Steam ou de outros ecossistemas de distribuição de jogos. Outro ponto diz respeito à dependência de dados obtidos via APIs e métodos de webscraping, sujeitos a mudanças estruturais ou restrições de acesso impostas pelas plataformas, o que pode impactar a replicabilidade de estudos semelhantes.

Para pesquisas futuras, recomenda-se ampliar o escopo linguístico, incorporando avaliações em múltiplos idiomas, especialmente o português, a fim de captar as particularidades do público nacional. Também se sugere a aplicação e comparação de diferentes algoritmos de aprendizado de máquina e deep learning, como redes neurais

convolucionais e transformers especializados em análise de sentimento, para avaliar possíveis ganhos preditivos. Adicionalmente, a integração de dados provenientes de múltiplas plataformas de jogos e redes sociais pode enriquecer a compreensão das preferências dos gamers, permitindo uma visão mais holística do comportamento desse público. Por fim, estudos longitudinais que monitorem a evolução das preferências ao longo do tempo, considerando o impacto de lançamentos, atualizações e tendências tecnológicas, poderão oferecer subsídios ainda mais sólidos para estratégias de mercado e desenvolvimento de produtos na indústria de jogos digitais.

Assim, o presente trabalho contribui tanto para a literatura acadêmica quanto para a prática empresarial, evidenciando que a análise preditiva baseada em dados textuais é uma ferramenta estratégica para compreender e antecipar o comportamento de consumo no setor de games. Ao alinhar metodologias estatísticas e computacionais de ponta com questões mercadológicas concretas, esta pesquisa abre caminho para abordagens mais precisas e adaptativas na tomada de decisão, fortalecendo o elo entre ciência de dados e inovação na indústria de entretenimento digital.

REFERÊNCIAS

ALEXANDRE, Vinicius R. **Comportamento do Consumidor de Jogos Multiplayer Online**. Americana, SP. 2018.

AMÉLIO, Camila de Oliveira. A Indústria e o mercado de Jogos Digitais no Brasil. **XVII SBGames**, Foz do Iguaçu, Paraná, Brasil, p. 1497-1506, 2018.

BATTA, V. (2024). **Human Language Data Processing using NLTK**. *International Journal of Advanced Research in Science, Communication and Technology*.
<https://doi.org/10.48175/ijarsct-17685>.

BERTENS, P. *et al.* (2018). **A Machine-Learning Item Recommendation System for Video Games**. 2018 IEEE Conference on Computational Intelligence and Games (CIG), 1-4. <https://doi.org/10.1109/CIG.2018.8490456>.

BHOIR, H. *et al.* (2021). **Web Crawling on News Web Page using Different Frameworks**. *International Journal of Scientific Research in Science and Technology*.
<https://doi.org/10.32628/cseit2174120>.

BISHOP, Christopher. **Pattern recognition and machine learning**. Cambridge: Springer. 2007. Disponível em: <https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf>. Acesso em: 28 de abril de 2024.

CARDOSO, Marcos V. *et al.* **Pesquisa da indústria brasileira de games 2023**. ABRAGAMES: São Paulo, 2023. Disponível em: https://www.abragames.org/uploads/5/6/8/0/56805537/2023_rel%C3%B3rio_final_v4.3.3_ptbr.pdf. Acesso em: 02 de maio de 2024.

CLEMENT, J. STATISTA. **Video game market revenue worldwide from 2019 to 2029**. Disponível em: <https://www.statista.com/statistics/1344668/revenue-video-game-worldwide/>. Acesso em: 28 de abril de 2024.

CLEMENTE, Ana Cristina Fernandes; STOPPA, Edmur Antonio. Lazer doméstico em tempos de pandemia da Covid-19. **LICERE-Revista do Programa de Pós-graduação Interdisciplinar em Estudos do Lazer**, v. 23, n. 3, p. 460-484, 2020.

COREY, Thomas. *et al.* (2021). **10. The Spheres of Player Motivation: Towards a New Typology for Player Motivation in Digital Games**.
<https://doi.org/10.1145/3450337.3483491>.

CUI, Zirui. (2023). **1. Analysis of the Development and Spread of Video Games from the Demand Level**. *BCP business & management*,
<https://doi.org/10.54691/bcpbm.v43i.4619>

DAUMÉ III, Hal. **A Course of machine learning**. Online book, 2013. Disponível em: <http://www.deeplearningbook.org/>.

DE OLIVEIRA, João Vitor Lemes; GRACIANO, Fernando Cesar. Análise de Dados e Desempenho de Jogadores em Games Utilizando API de Dados. **Revista Interface Tecnológica**, v. 19, n. 2, p. 105-116, 2022.

EUNYOUNG Ko *et al.* **A deep learning-based framework for predicting pork preference**. Current Research in Food Science, Volume 6. 2023. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2665927123000631>. Acesso em: 01 de maio de 2024.

FLEURY, Afonso; NAKANO, Davi; CORDEIRO, J. H. D. O. **Mapeamento da indústria brasileira e global de jogos digitais**. São Paulo: GEDIGames/USP, p. 32-33, 2014. Disponível em: https://www.bndes.gov.br/wps/wcm/connect/site/807f877e-22d7-44e5-b11c-c6bd1cf65b66/chamada_publica_FEP0211_mapeamento_da_industria.pdf?MOD=AJPERES&CACHEID=ROOTWORKSPACE.Z18_7QGCHA41LORVA0AHO1SIO51085-807f877e-22d7-44e5-b11c-c6bd1cf65b66-lxKa6Gq. Acesso em: 15 de abril de 2024.

GENEROSI, Andrea *et al.* **A deep learning-based system to track and analyze customer behavior in retail store**. Disponível em: https://www.researchgate.net/publication/329750616_A_deep_learning-based_system_to_track_and_analyze_customer_behavior_in_retail_store. Conference: 2018 IEEE 8th International Conference on Consumer Electronics - Berlin. Acesso em: 01 de maio de 2024.

GERON, Aurélien. **Mãos à obra: aprendizado de máquina com Scikit-learn, Keras & TensorFlow: conceitos, ferramentas e técnicas para a construção de sistemas inteligentes**. Rio de Janeiro: Alta Books. 2021.

GEWERS, F. *et al.* (2018). **Principal Component Analysis**. ACM Computing Surveys (CSUR), 54, 1 - 34. <https://doi.org/10.1145/3447755>.

GIL, Antonio Carlos. **Como Elaborar Projetos de Pesquisa**. São Paulo: Editora Atlas S.A. 2002.

GOLDAMEIR, N. *et al.* (2021). **Classification of the Human Development Index in Indonesia Using the Bootstrap Aggregating Method**. SinkrOn. <https://doi.org/10.33395/sinkron.v6i1.11173>.

GOSZTONYI, Márton. **Who are the gamers? Profiling adult gamers using machine learning approaches**. 2023. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2772503023000348#abs0001>. Acesso em: 28 de abril de 2024.

GRACIANO, Helton L. dos S. et al. **ScraperCI: um web scraper para coleta de dados científicos**. Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação, v. 28, p. 1-18, 2023. Disponível em:

<https://periodicos.ufsc.br/index.php/eb/article/view/92471/53205>. Acesso em: 02 de maio de 2024.

HAIR Jr., Joseph F., *et al.* **Análise multivariada de dados**. Porto Alegre: Bookman, 2009.

HOENEN, Felipe *et al.* **A universidade e os novos muros da inteligência artificial**. *Jornal da USP*. 2024. Disponível em: <https://jornal.usp.br/artigos/a-universidade-e-os-novos-muros-da-inteligencia-artificial/>. Acesso em: 02 de maio de 2024.

JUN LEE, Sang; SIAU, Keng. A review of data mining techniques. **Industrial Management & Data Systems**, v. 101, n. 1, p. 41-46, 2001.

KAUFMAN, Dora. **Desmistificando a inteligência artificial**. Autêntica Editora, 2022.

KHDER, Moaiad Ahmad. Web scraping or web crawling: State of art, techniques, approaches and application. **International Journal of Advances in Soft Computing & Its Applications**, v. 13, n. 3, 2021.

KUMAR, Dr. Chitrlekha *et al.* **Leveraging Deep Learning for Customer Segmentation: Patterns and Preferences Unveiled**. www.ijisae.org. 2024. Disponível em: <https://www.ijisae.org/index.php/IJISAE/article/view/4677/3351>. Acesso em: 01 de maio de 2024.

KURNIAWAN, C. *et al.* (2018). **Analysis of Student Demographic Information Using Data Mining classification with Decision Tree**. . <https://doi.org/10.4108/eai.23-4-2018.2277561>. <https://eudl.eu/pdf/10.4108/eai.23-4-2018.2277561>

LECUN, Y., BENGIO, Y., e HINTON, G. (2015). Deep learning. **Nature**, 521, 436–444.

MANZATO, Antonio José *et al.* **A Elaboração de Questionários de Pesquisa Quantitativa**. Departamento de Ciência de Computação e Estatística – IBILCE – UNESP.

MARQUES, Daniel C. **Sistema de Recomendação baseado em Reinforcement Learning: uma prova de conceito aplicada ao Video on Demand**. Dissertação: Mestrado em Gestão Comercial. Faculdade de Economia, Universidade do Porto. 2021. Disponível em: <https://repositorio-aberto.up.pt/bitstream/10216/138127/2/518066.pdf>. Acesso em: 01 de maio de 2024.

MARTINS, L., & Araújo, F. (2021). **Mineração de Texto para a Análise do Perfil Emocional de Usuários de Jogo Empático**. *Anais do XII Computer on the Beach - COTB '21*. <https://doi.org/10.14210/COTB.V12.P370-377>.

MYATT, Glenn J. **Making sense of data: a practical guide to exploratory data analysis and data mining**. John Wiley & Sons, 2007.

OLIVEIRA, Alyfer R. S. de O. *et al.* **Web scraping aplicado a cibersegurança. 2023.** Disponível em:

https://ric.cps.sp.gov.br/bitstream/123456789/15126/1/tecnologiaemgestaodatecnologia-dainformacao_2023_1_%20Alyfer%20Ricardo%20Sousa%20de%20Oliveira_%20Web%20scraping%20aplicado%20a%20ciberseguran%c3%a7a.pdf. Acesso em: 02 de maio de 2024.

OMOL, E. *et al.* (2024). **Application Of K-Means Clustering For Customer Segmentation In Grocery Stores In Kenya.** International Journal of Science, Technology & Management. <https://doi.org/10.46729/ijstm.v5i1.1024>.

PWC - PricewaterhouseCoopers Brasil Ltda. **Pesquisa Global de Entretenimento e Mídia 2022–2026.** Disponível em: <https://www.pwc.com.br/pt/estudos/setores-atividades/entretenimento-midia/2022/GEMO-2022.pdf>. Acesso em: 02 de maio de 2024.

PYKES, Kurtis (2024). **Primeiros passos com solicitações HTTP Python para APIs REST** (<https://www.datacamp.com/pt/tutorial/making-http-requests-in-python>).

PGB 2024 - **Pesquisa Game Brasil**, Go Gamers, SX Group. 2024. Disponível em: <https://www.pesquisagamebrasil.com.br/pt/>. Acesso em: 02 de maio de 2024.

RAMASWAMY, Sridhar *et al.* **Customer Perception Analysis Using Deep Learning and NLP.** Elsevier B.V. 2018. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1877050918319999>. Acesso em: 01 de maio de 2024.

RED HAT, **O que é uma API REST?** [S.l.], 17 ago. 2023. Disponível em: <https://www.redhat.com/pt-br/topics/api/what-is-a-rest-api>. Acesso em: 01 maio de 2024.

SAPIENZA, A. *et al.* (2018). **Deep Neural Networks for Optimal Team Composition.** Frontiers in Big Data, 2. <https://doi.org/10.3389/fdata.2019.00014>.

SCHMIDHUBER, J. (2015). **Deep learning in neural networks: An overview.** Neural Networks, 61, 85-117.

SCHULTHESS, Coline. **History of REST APIs.** [S.l.], 26 jan. 2017. Disponível em: <https://www.mobapi.com/history-of-rest-apis/>. Acesso em: 01 maio de 2014.

SILVA, Ana Rita M. C. **Previsão de preços de mercado baseada em Deep Learning.** Dissertação: Mestrado Integrado em Engenharia Electrotécnica e de Computadores. Universidade do Porto. 2021. Disponível em: <https://repositorio-aberto.up.pt/bitstream/10216/136721/2/504900.pdf>. Acesso em: 01 de maio de 2024.

SQLITE. **SQLite Home Page.** Disponível em: <https://sqlite.org/index.html>. Acesso em: 03 mar. 2025.

SULISTYOWATI, D. *et al.* (2020). **Implementation of Data Mining Algorithm for Predicting Popularity of Playstore Games in the Pandemic Period of Covid-19.** ,95-100. <https://doi.org/10.33480/jitk.v6i1.1425>.

SUN, Q., Feng, X., Zhao, S. *et al.* **Deep Learning Based Customer Preferences Analysis in Industry 4.0 Environment.** *Mobile Netw Appl* 26, 2329–2340 (2021). Disponível em: <https://doi.org/10.1007/s11036-021-01830-5> (licença: <http://creativecommons.org/licenses/by/4.0/>). Acesso em: 01 de maio de 2024.

SUPRIYATNA, B., & Putri, F. (2024). **Optimized support vector machine for sentiment analysis of game reviews.** *International Journal of Informatics and Communication Technology (IJ-ICT)*. <https://doi.org/10.11591/ijict.v13i3.pp344-353>.

TEJA, A. *et al.* (2023). **Predicting Steam Games Rating with Regression.** *E3S Web of Conferences*. <https://doi.org/10.1051/e3sconf/202338802001>.

THONGTAN, T., & Phienthrakul, T. (2019). **Sentiment Classification Using Document Embeddings Trained with Cosine Similarity.** , 407-414. <https://doi.org/10.18653/v1/P19-2057>.

TRINDADE, Joana F.V. **Interpretabilidade em Modelos de Sistemas de Recomendação.** Dissertação: Mestrado em Ciências de Dados (Data Science). Universidade do Porto. 2020. Disponível em: <https://repositorio-aberto.up.pt/bitstream/10216/132168/2/442109.pdf>. Acesso em: 01 de maio de 2024.

ULLMANN, Gabriel C. *et al.* **Sistema de sugestão de produtos para e-commerce utilizando Inteligência Artificial.** Universidade Regional do Noroeste do Estado do Rio Grande do Sul (UNIJUI). Santa Rosa – RS. 2021. Disponível em: <https://sol.sbc.org.br/index.php/eradrs/article/view/14773/14618>. Acesso em: 01 de maio de 2024.

URBAN, Glen *et al.* **Is Deep Learning a Game Changer for Marketing Analytics?** *MIT Sloan Management Review* 61, 2 (November 2019): 71-76. Disponível em: https://dspace.mit.edu/bitstream/handle/1721.1/130439/Urban-et-al_MG_2019-Deep-Learning-SMR.pdf?sequence=2&isAllowed=y. Acesso em: 01 de maio de 2024.

VEJA. Portal de notícias da VEJA. **Mudou de fase: mercado de games já fatura mais que o de cinema.** 2023. Disponível em: <https://veja.abril.com.br/tecnologia/mudou-de-fase-mercado-de-games-ja-fatura-mais-que-o-de-cinema>. Acesso em: 28 de abril de 2024.

VELKUMAR, K. *et al.* (2020). **Web Crawler and Web Crawler Algorithms: A Perspective.** *International Journal of Engineering and Advanced Technology*. <https://doi.org/10.35940/ijeat.e9362.069520>.

YEE, N. (2006). **Motivations for Play in Online Games. Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society**, 9 6, 772-5 . <https://doi.org/10.1089/cpb.2006.9.772>.

YING, X. (2019). **An Overview of Overfitting and its Solutions**. Journal of Physics: Conference Series, 1168. <https://doi.org/10.1088/1742-6596/1168/2/022022>.