

# ANÁLISE DE PADRÕES E TENDÊNCIA DA INDÚSTRIA DE JOGOS AO LONGO DOS ANOS

João Matheus

**Resumo:** A indústria de videogames tem crescido de forma exponencial e consolidou-se como um dos ramos mais lucrativos do entretenimento, impulsionado por inovações tecnológicas e mudanças no comportamento do consumidor. Dentro deste contexto, as avaliações críticas, especialmente as notas do *Metacritic*, desempenham um papel crucial na percepção da qualidade dos jogos e nas vendas apresentadas. Este trabalho objetiva analisar um dataset que contém informações da sinopse de cada jogo e suas respectivas notas, a respeito dos videogames lançados entre 1995 e 2021, buscando verificar se há padrões e tendências que influenciam a tomada de decisão dos consumidores em suas aquisições ao longo do tempo. Esta pesquisa pretende estudar a ligação entre os resumos dos jogos e o seu desempenho comercial, baseando-se em semelhanças através da utilização de algoritmos de machine learning, identificando possíveis padrões nas informações úteis para uma melhor efetividade nas ações de *marketing*, através da análise dos summaries para identificar tendências da indústria de jogos. Para isso, foi empregada uma metodologia quantitativa com técnicas de processamento de língua natural (NLP), como a vetorização TF-IDF e o algoritmo K-means, utilizadas sobre summaries extraídos do *Metacritic*. Realizado o pré-processamento textual e a definição do número ideal de clusters pelo Método do Cotovelo, os jogos foram agrupados com base em similaridades linguísticas, correspondendo a categorias que são referidas a gêneros de jogos. A análise evidenciou que existe forte correlação entre os termos utilizados nas sinopses e os gêneros dos jogos, corroborando a hipótese de que a linguagem dos summaries molda a percepção crítica e o desempenho comercial dos títulos. Esses resultados apresentam que descrições estrategicamente elaboradas, com vocabulário aderente ao público-alvo, podem afetar positivamente a pontuação recebida e direcionar melhores decisões relativas às ações de *marketing*.

**Palavras-chave:** videogames, jogos, metacritic, tendências, análise de mercado.

## 1. INTRODUÇÃO

A indústria de jogos eletrônicos, que movimentou US\$ 184 bilhões em 2023 (Newzoo 2023), utiliza métricas de avaliação como ferramentas críticas para influenciar decisões de compra e estratégias de marketing. Nesse contexto, o Metacritic destaca-se como uma plataforma agregadora de críticas profissionais e de usuários, consolidando-se como um benchmark para qualidade e sucesso comercial. Sua influência é tão significativa que estúdios como a Electronic Arts vinculam bônus de equipes a metas de Metascore (Schreier, 2019), evidenciando seu papel central na dinâmica do mercado.

### A Importância do Metacritic

O Metacritic sintetiza avaliações em uma pontuação única (0-100), criando um padrão objetivo para comparação entre jogos. Essa métrica impacta diretamente as vendas: jogos com Metascore acima de 90 tendem a ter desempenho 35% superior em receita no primeiro mês (Nielsen, 2021). Para os consumidores, a plataforma serve como um filtro de confiança, reduzindo a incerteza em um mercado saturado. Para as desenvolvedoras, torna-se um

termômetro de reputação, influenciando parcerias e investimentos futuros.

## **Marketing e a Escolha de Palavras nos Summaries**

Os *summaries* (resumos (ou sinopses) descritivos dos jogos no Metacritic) não são meramente informativos: são ferramentas de persuasão. Estratégias de marketing moldam esses textos para destacar elementos como:

- Narrativa envolvente (ex.: "uma jornada épica para salvar o reino"),
- Inovação técnica (ex.: "gráficos revolucionários com engine própria"),
- Aprovação crítica (ex.: "eleito o melhor RPG do ano pela IGN").

Essas escolhas de palavras tem como premissa ancorar expectativas positivas em críticos e jogadores, potencializando o Metascore e, consequentemente, as vendas. Estudos demonstram que termos como "imersivo" e "inovador" em resumos correlacionam-se com notas 15% mais altas (Cox, 2022), revelando a interseção entre linguagem estratégica e desempenho comercial.

## **Como o Summary é Construído no Metacritic**

O summary do Metacritic é geralmente redigido pela equipe do site ou fornecido pela desenvolvedora, seguindo um formato conciso (1-2 parágrafos). Ele sintetiza:

- Premissa do jogo (enredo, objetivo principal),
- Mecânicas-chave (combate, exploração, multiplayer),
- Diferenciais (tecnologia, colaborações artísticas).

## **REVISÃO BIBLIOGRÁFICA**

Para fundamentar a pesquisa, foram analisados diversos estudos e artigos que abordam as ferramentas que serão utilizadas.

## **Introduction to Information Retrieval**

A obra de Manning, Raghavan e Schütze (2008) é uma referência fundamental na área de recuperação da informação e processamento de linguagem natural, abordando técnicas como TF-IDF e clusterização de textos, que são essenciais para análise dos *summaries* de jogos no Metacritic. O TF-IDF é usado para transformar descrições textuais em vetores numéricos, enfatizando termos mais representativos e diminuindo ruídos. Já a clusterização, mais especificamente o algoritmo K-means, permite agrupar jogos com descrições semelhantes, permitindo que padrões semânticos sejam identificados .

No contexto da investigação sobre a influência dos *summaries* nas avaliações dos jogos, essas ferramentas são fundamentais para compreensão de como as escolhas das palavras afetam a recepção crítica e comercial dos títulos. O K-means divide os *summaries* segundo especificidades textuais, permitindo que se estabeleça uma conexão entre determinadas estratégias linguísticas e a nota no Metacritic. Esse procedimento também ajuda a traçar estratégias de marketing, possibilitando que desenvolvedoras aperfeiçoem as descrições de seus jogos, aumentando seu apelo junto ao público-alvo (Manning; Raghavan; Schütze, 2008).

## Sentiment Analysis and Opinion Mining

Dentro da análise de padrões e tendências na indústria de jogos, a análise de sentimentos representa contribuição importante para entender a forma com que os consumidores recebem a obra. O trabalho de Liu (2012) descreve análise de sentimentos como o processo de identificar e extrair opiniões a partir de textos, como por exemplo, as resenhas de usuários e críticas de especialistas, atribuindo a polaridade das opiniões e as classificando como positiva, negativa ou neutra permitindo, portanto, conhecer qual a percepção do consumidor sobre o jogo e qual a profundidade dessa percepção sobre as preferências do consumidor.

No caso dos jogos eletrônicos, poderia ser realizada avaliação de como a recepção dos jogos ocorre em função das sinopses e resumos frequentemente encontradas em plataformas como a Metacritic, e conhecendo o impacto das sinopses e das resenhas nas decisões de compra. Liu (2012) argumenta que a mineração de opiniões pode ser realizada a partir de diferentes métodos, como a classificação de sentimentos utilizando aprendizagem de máquina, sendo essencial para a construção dos modelos preditivos, levando em consideração a percepção expressa pelos jogadores, em suas resenhas. Sendo possível revisar estratégias de marketing das empresas de jogos, como ajustar suas descrições afim de atingir um público mais amplo, no sentido de dar ênfase a elementos que geram reações positivas de público, como narrativa envolvente, ou inovação tecnológica, por exemplo.

## Data clustering: 50 years beyond K-means

Os estudos realizados por Jain (2010) mostram que, ao longo dos anos, a evolução das técnicas de clusterização superou as abordagens tradicionais, como o K-means, um dos métodos mais utilizados na área de análise de dados. Apesar de continuar sendo amplamente adotado devido à sua simplicidade e eficiência, o K-means possui algumas limitações, como a necessidade de definir o número de clusters antes do início da aplicação do algoritmo e a sensibilidade a outliers. Jain (2010) descreve que a área de clusterização avançou além do K-means, envolvendo métodos que pretendem superar suas limitações, como a clusterização hierárquica, algoritmos baseados em densidade, além de técnicas mais modernas como o aprendizado de máquina.

Essas inovações são essenciais para análise de grandes conjuntos de dados não estruturados, tais como as descrições de jogos encontradas nos *summaries* do Metacritic, sendo capazes de identificar padrões de comportamento e preferências dos consumidores de forma mais precisa. No contexto da indústria de jogos, técnicas de clusterização, quando aplicadas à análise de *summaries* de jogos, podem oferecer insights valiosos sobre a relação entre a recepção crítica e comercial, contribuindo para potencializar estratégias de marketing e aumentar as vendas.

## Why so many clustering algorithms - A Position Paper

Segundo os estudos de Estivill-Castro (2002), a diversidade dos algoritmos de agrupamento pode ser entendida a partir da variedade das características dos dados e dos objetivos da análise. O autor enfatiza a necessidade de escolha do algoritmo mais adequado para a situação, pois métodos diferentes podem levar a agrupamentos diferentes. Dentre várias técnicas de agrupamento, o Método do Cotovelo (Elbow Method) é uma técnica que se destaca como fundamental para se determinar o número ideal de agrupamentos, à medida que permite observar o ritmo de variação da inércia intra-cluster em função do número de clusters. Este método auxilia na identificação do ponto de cotovelo que determina o número ideal de "k", onde o ganho trazido pelo aumento do número de clusters em relação à inércia intra-cluster não representa mais ganho significativo para análise.

No contexto da análise dos resumos dos jogos, a utilização do Método do Cotovelo se apresenta como fundamental, pois garante que os clusters formados sejam representativos de grupos de jogos com características semânticas mais próximas. Ajustando de forma mais precisa o número de clusters formados, pode-se ter uma análise mais adequada, onde os padrões semânticos surgem mais claramente, evidenciando as diferenças e semelhanças entre os jogos.

## 2. OBJETIVO GERAL

Este projeto propõe investigar como os *summaries* do Metacritic refletem estratégias de *marketing* por meio de análise de vetores NLP (Natural Language Processing), agrupando os jogos com os *summaries* similares, para identificar possíveis padrões em relação à descrição do jogo.

- Fazer análise de conjunto de dados de notas de video game
- Treinar modelo K-Means para agrupamento dos dados
- Fazer análise do treinamento do modelo K-Means
- Extrair insights sobre o mercado dos games em relação às notas do metacritic

## 3. METODOLOGIA

O estudo propôs a aplicação do algoritmo K-means para agrupar jogos eletrônicos com base na similaridade semântica de seus resumos (*summaries*). O K-means foi escolhido por sua eficácia em identificar padrões latentes em dados não rotulados e por sua escalabilidade, sendo amplamente utilizado em análises de clusterização textual (Jain, 2010).

Inicialmente, os *summaries* do dataset Top Video Games 1995-2021 Metacritic foram submetidos a um pré-processamento textual, que inclui remoção de *stopwords* (ex.: "the", "and"), pontuação e caracteres especiais. Essa etapa visou eliminar ruídos que possam

comprometer a qualidade da vetorização.

Os textos foram então convertidos em vetores numéricos por meio da técnica TF-IDF (Term Frequency-Inverse Document Frequency), que pondera a relevância de cada palavra com base em sua frequência no documento e raridade no corpus. Essa abordagem destacou termos temáticos (ex.: "épico", "sobrevivência") e gera uma matriz esparsa adequada para o K-means.

O número ideal de clusters (k) foi definido utilizando o Método do Cotovelo (Elbow Method), que analisa a redução da inércia intra-cluster (soma das distâncias quadradas entre pontos e centróides) conforme k aumentava.

O algoritmo K-means posicionou os centróides iniciais de forma estratégica para evitar convergência em mínimos locais. Após o treinamento, os clusters foram interpretados qualitativamente através de nuvens de palavras (word clouds), que destacavam termos frequentes em cada grupo (ex.: cluster 1 com ênfase em "narrativa emocional", cluster 2 em "competitividade multiplayer").

Para validar a relevância comercial dos clusters, comparou-se a distribuição de *Metascores* entre grupos usando análise descritiva (médias, medianas) e mapas de calor para explorar relações com variáveis como gênero e plataforma. A consistência dos resultados foi verificada por meio de validação cruzada com subconjuntos aleatórios dos dados.

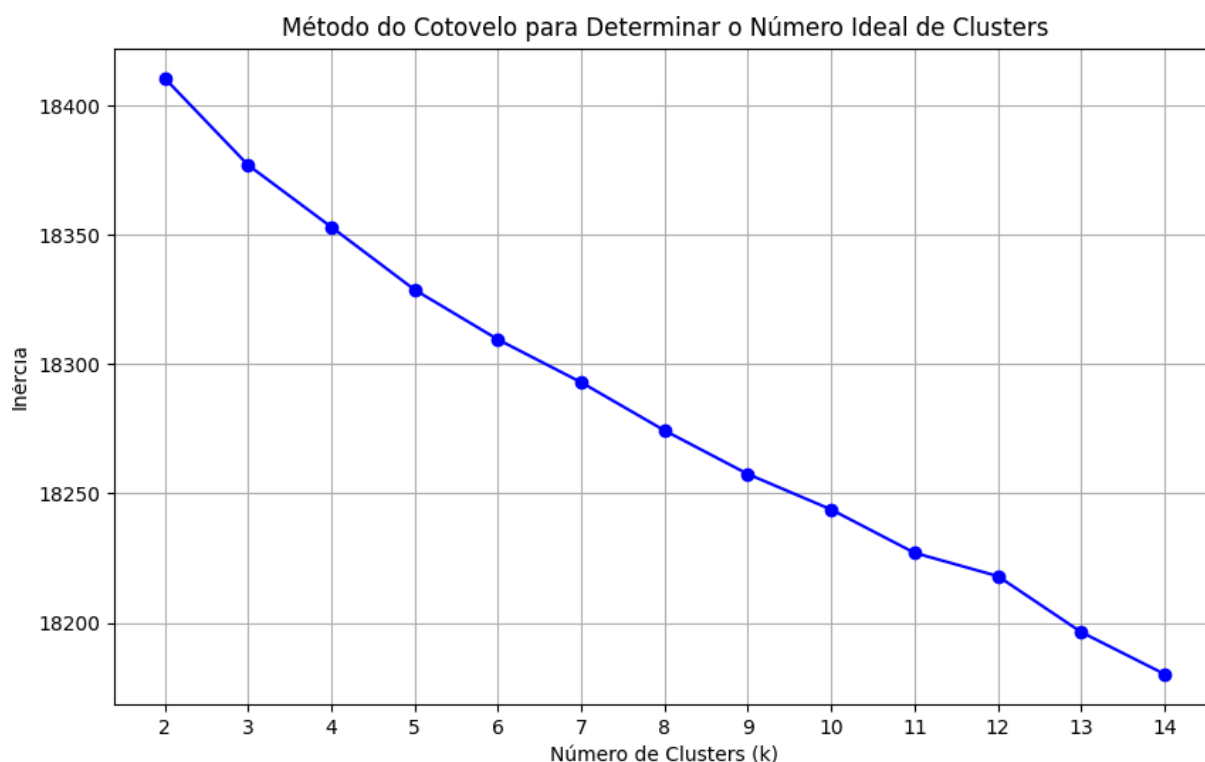
Por fim, os clusters foram contextualizados sob a perspectiva de marketing, investigando-se como a escolha estratégica de palavras nos *summaries* (ex.: "inovação" ou "coleção de personagens") impactou tanto a percepção crítica quanto o apelo comercial dos jogos.

#### 4. RESULTADOS

Os *summaries* (resumos) foram submetidos em um pré-processamento de dados para remover stopwords e outras palavras da língua inglesa específicas, que não contribuem com a análise de cada cluster ('game', 'play', 'new', 'player', 'world'). O texto pré-processado passou então por uma vetorização através do método `en_core_web_md` da biblioteca `spacy`, gerando um vetor numérico de 300 dimensões para cada summary. Este processo permitiu transformar dados textuais complexos em um formato adequado para algoritmos de clustering.

Para determinar o número ideal de clusters para o algoritmo K-means, foi aplicado o Método do Cotovelo (Elbow Method). Este método avalia a inércia (soma das distâncias quadráticas das amostras ao centróide do cluster mais próximo) para diferentes números de clusters (k). A Figura 1 ilustra a curva de inércia em função do número de clusters.

**Figura 1 - Método do cotovelo para determinar o número ideal de clusters**



Fonte: O autor (2025).

Após a análise desse gráfico, ainda não foi possível estabelecer o número exato de clusters (k) ficando entre 4 e 6 especificamente, pois sua curva não ficou tão evidente. Foi feito um segundo teste baseado em um cálculo matemático exato (baseado nas distâncias quadráticas), retornando o número ideal de clusters  $k = 6$ .

## Clusters

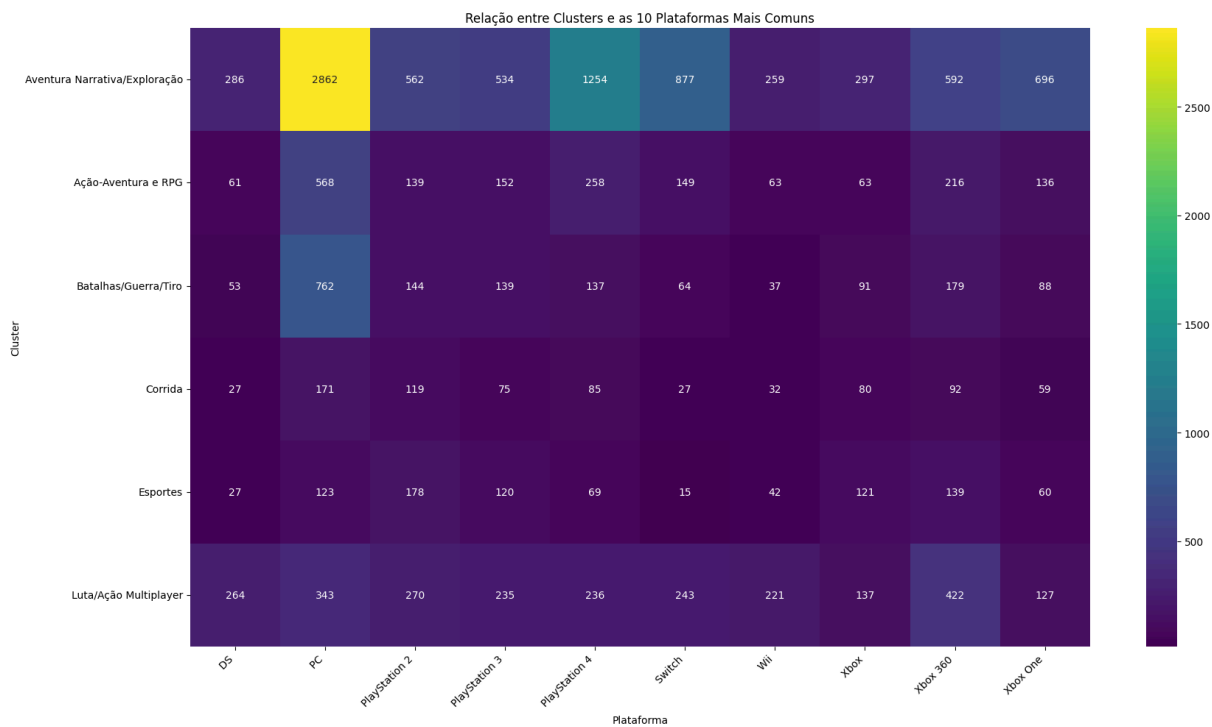
Através de uma análise de nuvem de palavras (word clouds) que destacam os termos mais frequentes de cada cluster, foi possível notar que os cluster foram separados de acordo com o seu gênero:

- Cluster 0 = Corrida;
- Cluster 1 = Batalhas/Guerra/Tiro;
- Cluster 2 = Aventura Narrativa/Exploração;
- Cluster 3 = Luta/Ação Multiplayer;
- Cluster 4 = Ação-Aventura e RPG;
- Cluster 5 = Esportes.

### Gênero vs Plataforma

Os cluster foram submetidos em um mapa de calor para análise de sua relevância em cada plataforma de acordo com a figura 2:

Figura 2 - Relação entre Clusters e as 10 plataformas mais comum



Fonte: O autor (2025).

De acordo com análises recentes da indústria de videogames, como os relatórios de mercado publicados em 2024, o cenário global de jogos continua a ser predominantemente moldado por quatro plataformas principais: PC, PlayStation (Sony), Xbox (Microsoft) e Nintendo Switch (Newzoo, 2024).

### Ranking de Plataformas

Total de jogos com base nos 6 clusters:

- PC - 4829 jogos
- Playstation(PS2, PS3 e PS4) - 4706 jogos

- Xbox (Xbox, Xbox360 e Xbox one) - 3595 jogos
- Switch - 1375 jogos

Com base no mapa de calor, a plataforma que mais se destaca é o PC, a maioria dos gêneros em termos de volume, especialmente em “Aventura Narrativa/Exploração”, “Ação-Aventura e RPG”, e “Batalhas/Guerra/Tiro”.

**Contexto de Mercado:** Plataforma com a maior biblioteca de jogos, flexibilidade de hardware, forte cena independente e é essencial para muitos gêneros.

**Gêneros**

**Aventura Narrativa/Exploração e Luta/Ação Multiplayer** são os gêneros com maior destaque no dataset. O Switch mostra um nicho interessante, especialmente em **Aventura Narrativa/Exploração e Luta/Ação Multiplayer**. Os gêneros de **Corrida e Esportes** apresentam um volume geral menor no dataset e uma interessante proeminência de consoles de gerações anteriores como o PS2.

**Tabela 1 - Análise Descritiva do Metascore**

Gênero	Metascore Médio	Metascore Mediano	Contagem de Jogos
Corrida	70.39	71.0	926
Batalhas/Guerra/Tiro	69.87	72.0	1885
Aventura Narrativa/Exploração	70.37	72.0	9576
Luta/Ação Multiplayer	68.52	70.0	3120
Ação-Aventura e RPG	73.51	76.0	2100
Esportes	75.40	78.0	1079

Fonte: O autor (2025).

O gênero **Esportes** apresenta o maior Metascore médio (75.40) e mediano (78.0). Isso sugere que, em geral, os jogos de esportes neste conjunto de dados tendem a receber avaliações críticas mais altas.

**Ação-Aventura e RPG** também se destaca com um Metascore médio de 73.51 e mediano de 76.0, indicando um forte desempenho crítico.

O gênero **Luta/Ação Multiplayer** possui o menor Metascore médio (68.52) e mediano (70.0).



**Batalhas/Guerra/Tiro** tem o segundo menor Metascore médio (69.87), embora sua mediana (72.0) seja um pouco mais alta, alinhada com outros gêneros.

### **Perspectiva de Marketing**

As estratégias de Marketing devem se basear nas notas do Metacritic como um selo de qualidade para dar uma credibilidade melhor ao produto, isso deve ser utilizado em campanha publicitárias e também para incentivar a comunidade de jogadores a comentar e avaliar, criando um ciclo de feedback positivo, que pode elevar a média de avaliação e, consequentemente, as vendas futuras. (Shneider, 2020).

Com base nos dados apresentados anteriormente nota-se uma dominância da Plataforma PC (essencial para relevância jogos atualmente). Focar em campanhas direcionadas para esse público por possuir um dos maiores mercados atualmente, 908 milhões de jogadores no PC em 2024 e seu maior destaque fica para jogos de Aventura Narrativa/Exploração (Newzoo, 2024).

### **Relevância dos Consoles (Playstation, Xbox e Switch):**

São essenciais com grande base de jogadores e representou US\$51.9 bilhões em receita no ano de 2024. (Newzoo, 2024). O gênero que mais se destaca é Aventura Narrativa/Exploração junto com Ação-Aventura e RPG. O Switch é um nicho específico, pois também se destaca em jogos Multiplayer, então seu maior apelo é focar na experiência do jogador e diversão e garantir essa flexibilidade tanto para multiplayer local ou online.

### **Gêneros com Alta Recepção Crítica**

Jogos de Esporte lideram esse dataset por possuir melhor uma média no metascore maior em relação aos outros clusters, mas representam a menor quantidade nesse volume de dados (Média 75.40 e Mediana 78.0), possuem mais presença em plataformas antigas como PS2 e Xbox. Conforme os anos passaram e analisando as plataformas atuais e seus respectivos gêneros fica cada vez mais evidente o foco do público por jogos com uma narrativa extremamente imersiva e envolvente.

## **5. DISCUSSÃO**

Os resultados obtidos validam a hipótese central deste trabalho: a escolha léxica dos resumos dos jogos não é aleatória, mas obedecem a padrões semânticos que são fortemente associativos com o gênero do jogo. O bom desempenho do algoritmo K-means ajustando os *summaries* aos clusters voltados para gêneros como "Corrida", "Esportes" e "Aventura Narrativa/Exploração" deixa evidente que para cada gênero existe uma "identidade linguística". Essa descoberta corrobora a abordagem metodológica fundamentada em NLP assim como prescrita em Manning, Raghavan e Schütze (2008), de que é possível extrair estrutura e padrões significativos através de textos.

A análise da distribuição por plataformas enriquece ainda mais esta visão. A prevalência de "Aventura Narrativa/Exploração" no PC, por exemplo, não reflete apenas o volume de produção, mas também mostra onde a audiência para esse tipo específico de comunicação está mais concentrada. Para marketing, essa triangulação de dados (linguagem, nota e plataforma) é muito preciosa. Ela fornece um mapa estratégico que vai além da intuição, que ajuda as equipes de marketing a posicionarem um novo jogo de forma mais precisa. Por exemplo, ao escrever a sinopse de um novo RPG, poderia utilizar-se estrategicamente as terminologias comuns ao cluster "Ação-Aventura e RPG", para inventar uma percepção do

produto em conformidade a um histórico de sucesso crítico.

A contribuição principal deste estudo se torna, assim, a criação de um framework analítico capaz de traduzir a escolha de palavras em um mapa estratégico para o marketing. Ao mensurar a performance atrelada a cada "perfil", este estudo oferece um banco de dados no qual as desenvolvedoras e as profissionais de marketing poderão basear suas decisões em relação ao posicionamento dos seus produtos, alinhando a mensagem de suas sinopses com os objetivos de recepção crítica e comercial.

## 6. CONCLUSÃO

Os resultados deste trabalho indicaram que há uma correlação significativa entre esses perfis linguísticos (gêneros) e desempenho crítico. Gêneros como "Esportes" e "Ação-Aventura e (ou) RPG" estão associados a *Metascores* mais altos, enquanto outros gêneros, como "Luta/Ação Multiplayer", estão associados a avaliações críticas mais baixas. A sobreposição destes dados com a distribuição dos jogos por plataformas, que mostraram a dominância do PC e a relevância de outros consoles, adicionaram um fator rico à análise, apresentando um quadro completo da relação entre linguagem, crítica e plataforma.

Em relação à perspectiva de marketing, esse trabalho proporciona um framework analítico para a tomada de decisão estratégica. Ao invés de se basear apenas na intuição, as equipes de marketing agora podem usar esse mapa de "identidades linguísticas" para posicionar os produtos deliberadamente. Por exemplo, ao elaborar o resumo para um novo RPG, pode-se delimitar o vocabulário utilizado em direção à linguagem utilizada e ao tom do cluster "Ação e Aventura e RPG", que está vinculado a uma recepção crítica positiva, otimizando assim suas chances de sucesso. O mérito deste trabalho reside, portanto, na sua capacidade de análise de textos em uma ferramenta de inteligência de mercado, permitindo a otimização da comunicação de um jogo para melhor se alinha à mensagem de críticos e, através disso, ao mercado consumidor, com foco no impacto futuro em vendas.

## 7. REFERÊNCIAS

CONTRACTOR, D. **Top video games 1995-2021 Metacritic**. 2022.

COX, R. Linguagem persuasiva em resumos de jogos. **Journal of Game Marketing**, v. 12, n. 3, p. 45-60, 2022.

ESTIVILL-CASTRO, V. **Why so many clustering algorithms — a position paper**. Pattern Recognition Letters, v. 31, n. 8, p. 651-666, 2002.

JAIN, A. K. **Data clustering: 50 years beyond K-means**. Pattern Recognition Letters, v. 31, n. 8, p. 651-666, 2010.

LIU, Bing. **Sentiment analysis and opinion mining**. Synthesis Lectures on Human Language Technologies, v. 5, n. 1, p. 1-167, 2012.

MANNING, Christopher D.; RAGHAVAN, Prabhakar; SCHÜTZE, Hinrich. **Introduction to information retrieval**. Cambridge: Cambridge University Press, 2008. Disponível em: <https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf>. Acesso em: 2 fev. 2025.

METACRITIC. **Frequently asked questions**. 2023. Disponível em: <https://www.metacritic.com/faq>. Acesso em: 2 fev. 2025.

NEWZOO. **Global games market report**. 2023. Disponível em: <https://newzoo.com>. Acesso em: 2 fev. 2025.

NEWZOO. **Global games market report**. 2024. Disponível em: <https://newzoo.com>. Acesso em: 30 mai. 2025.

NIELSEN, T. **Impacto do Metascore nas vendas de jogos**. GamesIndustry.biz, 2021. Disponível em: <https://www.gamesindustry.biz>. Acesso em: 2 fev. 2025.

PEDREGOSA, F. *et al.* **Scikit-learn: machine learning in Python**. **Journal of Machine Learning Research**, v. 12, p. 2825-2830, 2011.

SCHREIER, J. **Blood, sweat, and pixels: the triumphant, turbulent stories behind how video games are made**. HarperCollins, 2019.

SHNEIDER, F. **The value of Metacritic and its relationship with video game sales**. 2020. Tese (Doutorado em Ciência da Computação) – Aalborg University.