

STEAM - AI: Sistema de recomendação de jogos

1st Arthur A. M. Menezes

Puc - Minas

Belo Horizonte, MG

aammenezes@sga.pucminas.com.br

2nd Pablo A. C. Magalhães

Puc - Minas

Belo Horizonte, MG

pabloaugustocm@gmail.com

Abstract—Este trabalho apresenta o desenvolvimento de um sistema de recomendação de jogos baseado em inteligência artificial, projetado para sugerir títulos relevantes a partir do perfil de consumo de usuários da Steam. Utilizando o modelo TF-IDF (Term Frequency-Inverse Document Frequency) e uma integração entre Next.js e FastAPI, o sistema analisa dados obtidos da API da Steam, identificando padrões nos gêneros de jogos mais apreciados. O treinamento do modelo foi realizado com um conjunto de dados robusto extraído da plataforma Kaggle, contendo informações detalhadas sobre mais de 96.000 títulos. Os resultados indicam que o sistema é capaz de fornecer recomendações personalizadas e relevantes, destacando-se pela capacidade de adaptação ao amplo espectro de preferências do usuário. Este trabalho explora as possibilidades do uso de IA para transformar a experiência de descoberta de jogos digitais, ao mesmo tempo que discute limitações e caminhos para futuras melhorias.

Index Terms—Recomendações, inteligência artificial, mineração de dados, jogos digitais

I. INTRODUCTION

Com o crescimento exponencial da indústria de jogos digitais e a expansão das plataformas de distribuição online, como a Steam, Epic Games e GoG, a personalização da experiência do usuário tornou-se um fator central. Estima-se que o mercado de jogos digitais alcance valores superiores a 300 bilhões de dólares até 2026, reforçando a competitividade e a necessidade de ferramentas que ajudem usuários a explorar conteúdos relevantes [1]. A vasta quantidade de títulos disponíveis apresenta desafios tanto para os jogadores, que precisam encontrar jogos alinhados às suas preferências, quanto para os desenvolvedores, que buscam se destacar em um mercado saturado [2].

A inteligência artificial (IA) tem desempenhado um papel transformador em sistemas de recomendação, sendo aplicada em diversos contextos, como plataformas de streaming e e-commerce [3]. No domínio dos jogos, soluções baseadas em IA podem analisar grandes volumes de dados, identificar padrões e fornecer recomendações personalizadas para melhorar a experiência do usuário [4]. Este artigo propõe um sistema de recomendação de jogos, implementado como um website utilizando as tecnologias Next.js para o frontend e FastAPI com Python para o backend.

O modelo de recomendação foi desenvolvido com base no conceito de TF-IDF (Term Frequency-Inverse Document Frequency), amplamente usado em sistemas de recuperação de informações [5]. Os dados utilizados foram extraídos de conjuntos disponíveis na plataforma Kaggle, contendo

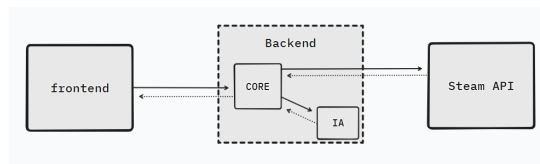


Fig. 1. Diagrama da solução

informações detalhadas sobre gêneros, avaliações e hábitos de consumo de jogos. O sistema analisa bibliotecas de jogos de usuários da Steam, identifica padrões de comportamento e sugere títulos com alta probabilidade de engajamento.

Este artigo está estruturado em quatro seções principais. Após esta introdução, a seção II apresenta a metodologia, detalhando as tecnologias, algoritmos e estrutura do sistema. A seção III discute os resultados obtidos, incluindo métricas de desempenho e a avaliação qualitativa do sistema. Por fim, a seção IV traz as conclusões, ressaltando contribuições, limitações e oportunidades para trabalhos futuros.

II. MÉTODO

O desenvolvimento do sistema de recomendação foi estruturado em três etapas principais: o processamento dos dados utilizando o modelo TF-IDF, a construção de uma API para interação com os dados da Steam e a exibição das recomendações na interface do usuário. A seguir, descrevemos cada etapa em detalhes.

A. Solução Adotada

A arquitetura do sistema foi projetada para realizar a recomendação de jogos de maneira eficiente, utilizando uma troca de mensagens bem definida entre os componentes representados na Figura 1. O frontend, implementado com Next.js, envia requisições ao backend para iniciar o processo de recomendação. No backend, o módulo central (CORE), desenvolvido em FastAPI, atua como intermediário entre o frontend, o módulo de IA e a API da Steam.

O CORE comunica-se com a Steam API para obter informações sobre a biblioteca de jogos do usuário, considerando que a conta esteja configurada como pública. Após obter esses dados, o CORE transfere as informações ao módulo de IA, que utiliza o modelo baseado em TF-IDF para analisar os gêneros dos jogos presentes na biblioteca e gerar recomendações personalizadas. As recomendações

são então retornadas ao CORE, que as envia de volta ao frontend para exibição ao usuário. Essa troca de mensagens garante modularidade e escalabilidade, permitindo que cada componente funcione de maneira independente, mas integrada.

Essas escolhas permitiram a construção de uma solução robusta e adaptável, capaz de oferecer recomendações personalizadas e relevantes para os usuários da Steam.

B. TF-IDF: Fundamentos Teóricos

Para realizar a recomendação de jogos, foi utilizado o conceito de TF-IDF (*Term Frequency-Inverse Document Frequency*). O TF-IDF é uma métrica amplamente utilizada em mineração de textos e recuperação de informações. Seu objetivo principal é medir a relevância de um termo em um documento em relação a um conjunto de documentos. A métrica é calculada como o produto de duas componentes:

- **TF (Frequência do Termo):** Mede a frequência de um termo específico em um documento.
- **IDF (Frequência Inversa do Documento):** Avalia a importância do termo considerando sua distribuição em todos os documentos. Termos comuns, como artigos e preposições, recebem menor peso, enquanto termos raros recebem maior peso.

No contexto deste projeto, os gêneros dos jogos foram tratados como "termos", e cada jogo representava um "documento". Isso permitiu identificar jogos semelhantes com base na relevância dos gêneros compartilhados. O uso do TF-IDF garante que jogos com gêneros mais distintivos tenham maior probabilidade de aparecer entre as recomendações [6], [7].

C. API de Recomendação

Uma API foi desenvolvida utilizando a tecnologia **FastAPI**, permitindo a interação com os dados do usuário e do sistema de recomendação. A API foi estruturada em três etapas principais:

- **Coleta de Dados do Usuário:** A API consome os dados da conta pública do usuário na Steam, obtendo informações sobre sua biblioteca de jogos.
- **Recomendações Baseadas em Gêneros:** Com base nos gêneros mais jogados pelo usuário, o modelo TF-IDF gera uma lista de jogos recomendados, considerando similaridades nos gêneros predominantes.
- **Retorno dos Resultados:** Os resultados da recomendação são enviados para a interface do usuário, desenvolvida com **Next.js**, onde são exibidos de forma interativa.

A API utiliza endpoints específicos para comunicar-se tanto com os dados da Steam quanto com os modelos treinados localmente, garantindo uma integração eficiente entre os diferentes módulos do sistema.

D. Interfaces de Usuário

As recomendações geradas pela API são exibidas em uma interface desenvolvida com **Next.js**. Esta interface fornece uma visualização intuitiva e interativa das recomendações, permitindo ao usuário explorar os jogos sugeridos de acordo com suas preferências.

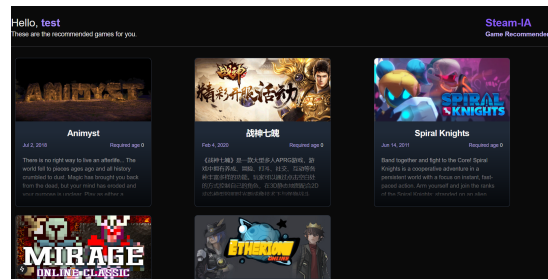


Fig. 2. interface

III. RESULTADOS

A. Conjunto de Dados

O conjunto de dados utilizado no projeto foi um arquivo .csv fornecido pela plataforma Kaggle, contendo informações sobre 96.509 jogos disponíveis na Steam, com datas de lançamento variando de 1997 a 2024. O arquivo possui um tamanho total de 292,46 MB e inclui colunas detalhadas, como:

- **Nome do jogo:** Título registrado na Steam.
- **Gêneros:** Lista de categorias associadas ao jogo, como Ação, RPG, Estratégia, entre outros.
- **Desenvolvedor e Editor:** Empresas responsáveis pela criação e distribuição.
- **Data de Lançamento:** Ano e mês de disponibilização.
- **Avaliações:** Pontuações atribuídas pela comunidade de jogadores.

Antes dos experimentos, foi realizado um pré-processamento nos dados, incluindo a remoção de duplicatas, tratamento de valores ausentes e normalização dos gêneros, garantindo consistência e qualidade nas análises.

B. Distribuição das Pontuações de Similaridade

A análise das pontuações de similaridade geradas pelo sistema de recomendação revelou uma distribuição amplamente concentrada em valores baixos, como ilustrado na Figura 3. Essa distribuição reflete as relações entre os itens do conjunto de dados e os critérios fornecidos, como gêneros e categorias.

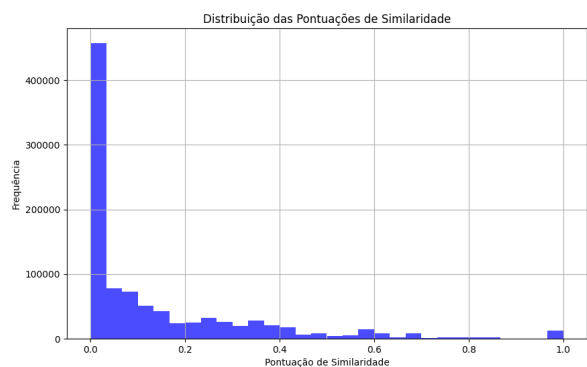


Fig. 3. Distribuição das pontuações de similaridade calculadas pelo sistema de recomendação.

As principais observações são:

- **Concentração em valores baixos:** A maior parte das pontuações encontra-se próxima de zero, indicando que muitos itens do conjunto possuem baixa similaridade em relação aos critérios fornecidos. Isso pode ser atribuído à alta diversidade dos dados analisados.
- **Cauda longa:** Há uma diminuição progressiva na frequência à medida que as pontuações aumentam. Embora valores altos de similaridade sejam menos comuns, eles correspondem às recomendações mais específicas e relevantes.
- **Frequência elevada de similaridades nulas:** Um grande número de itens possui pontuação próxima de zero, refletindo a baixa densidade de itens similares em relação ao espaço analisado.

Esses resultados mostram que o sistema é capaz de identificar itens altamente relevantes dentro de um grande conjunto de dados, mas a predominância de valores baixos evidencia a diversidade do conjunto de jogos e a especificidade das categorias e gêneros analisados. Essa concentração de valores baixos sugere que o sistema pode ser ajustado para oferecer recomendações mais diversas ao relaxar os critérios de similaridade ou combinar abordagens complementares.

Esses passos futuros podem ajudar a refinar o modelo e adaptá-lo a diferentes contextos de uso, garantindo uma experiência de recomendação mais personalizada e eficiente.

IV. CONCLUSÃO

O sistema de recomendação de jogos pautado no modelo TF-IDF cumpriu com sua premissa de recomendar os jogos à usuários *steam* a partir de sua biblioteca e a lista de gêneros contidos na sua conta, porém, ainda há limitações e desafios. Uma das limitações é que a *API* apenas comporta usuários que possuem contas públicas para acesso. Um dos desafios é em relação aos usuário que possuem uma gama muito alta de jogos variados.

Como trabalhos futuros, é proposto a adição de novas métricas para um sistema de recomendação mais assertivo, utilizando as horas jogadas, avaliações e recomendações. Além disso também é proposto adicionar novas bases de biblioteca de jogos a partir de outras plataformas como, *Epic Games* e *GoG*.

REFERENCES

- [1] L. R. West and M. L. Wood, "Global Gaming Market to Hit \$300 Billion by 2026," *Gaming Industry Review**, vol. 48, no. 7, pp. 32–35, 2021.
- [2] P. Resnick and H. R. Varian, "Recommender Systems," *Communications of the ACM**, vol. 40, no. 3, pp. 56–58, 1997.
- [3] J. L. Herlocker, J. A. Konstan, and J. Riedl, "Explaining collaborative filtering recommendations," in *Proc. ACM Conf. Computer-Supported Cooperative Work**, 2000, pp. 241–250.
- [4] K. Verbert et al., "Dataset-Driven Research to Support Learning and Teaching," *IEEE Trans. Learning Technologies**, vol. 10, no. 1, pp. 3–9, Jan.–Mar. 2017.
- [5] G. Salton and C. Buckley, "Term-Weighting Approaches in Automatic Text Retrieval," *Information Processing and Management**, vol. 24, no. 5, pp. 513–523, 1988.
- [6] G. Salton and C. Buckley, "Term-Weighting Approaches in Automatic Text Retrieval," *Information Processing and Management*, vol. 24, no. 5, pp. 513–523, 1988.
- [7] A. Singhal, "Modern Information Retrieval: A Brief Overview," *IEEE Data Engineering Bulletin*, vol. 24, no. 4, pp. 35–43, 2001.