

**UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO**

Isabelly de Almeida Silva

**Sistemas de Recomendação: Uma Abordagem Sobre
Jogos de Tabuleiro Modernos**

São Carlos

2026

Isabelly de Almeida Silva

Sistemas de Recomendação: Uma Abordagem Sobre Jogos de Tabuleiro Modernos

Trabalho de conclusão de curso apresentado
ao Centro de Ciências Matemáticas Aplicadas
à Indústria do Instituto de Ciências Matemá-
ticas e de Computação, Universidade de São
Paulo, como parte dos requisitos para conclu-
são do MBA em Ciências de Dados.

Área de concentração: Ciências de Dados

Orientador: Prof. Dr. Adriano Kamimura Su-
zuki

Versão original

**São Carlos
2026**

Isabelly de Almeida Silva

Sistemas de Recomendação: Uma Abordagem Sobre Jogos de Tabuleiro Modernos

Trabalho de conclusão de curso apresentado
ao Centro de Ciências Matemáticas Aplicadas
à Indústria do Instituto de Ciências Matemá-
ticas e de Computação, Universidade de São
Paulo, como parte dos requisitos para conclu-
são do MBA em Ciências de Dados.

Data de defesa: 07 de Fevereiro de 2026

Comissão Julgadora:

Prof. Dr. Adriano Kamimura Suzuki
Orientadora

Professor
Convidado1

Professor
Convidado2

São Carlos
2026

*Dedico este trabalho a todas as pessoas que caminharam comigo ao longo dessa jornada.
Em especial, aos alunos da turma de 2025 do MBA em Ciência de Dados,
pelas trocas, aprendizados e reflexões compartilhadas.
E aos pesquisadores do Instituto de Ciências Matemáticas e de Computação (ICMC).*

AGRADECIMENTOS

A realização deste trabalho só foi possível graças ao apoio, incentivo e compreensão de pessoas muito importantes ao longo dessa jornada.

Agradeço, em primeiro lugar, ao Andre, meu companheiro no amor e na vida, pela parceria constante, pelo cuidado e pela paciência, especialmente nos períodos de ausência dedicados ao estudo e à escrita deste trabalho. Seu apoio diário, sua presença e sua compreensão tornaram esse caminho mais leve e possível.

Aos meus amigos, deixo meu sincero agradecimento pelo incentivo contínuo, pela confiança em mim e por sempre me lembrarem das pequenas alegrias compartilhadas, mesmo nos momentos mais desafiadores. O apoio de vocês foi fundamental para manter a motivação ao longo de todo o processo.

Agradeço também ao Professor Adriano Kamimura Suzuki, pela orientação atenta, pelas contribuições valiosas e pela condução ao longo do desenvolvimento deste trabalho, que foram essenciais para a construção de um resultado consistente e significativo.

Por fim, agradeço a todos que, direta ou indiretamente, contribuíram para a realização deste trabalho e para minha formação acadêmica e pessoal ao longo dessa trajetória.

Muito obrigada a todos!

*“Prefiro ter perguntas que não podem ser respondidas
do que respostas que não podem ser questionadas.”*

Richard P. Feynman

RESUMO

SILVA, I. **Sistemas de Recomendação: Uma Abordagem Sobre Jogos de Tabuleiro Modernos.** 2026. 73p. Monografia (MBA em Ciências de Dados) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2026.

Este trabalho analisa a aplicação de Sistemas de Recomendação no contexto de jogos de tabuleiro modernos, a partir de dados reais da plataforma *BoardGameGeek*. São avaliadas diferentes abordagens, incluindo Filtragem Colaborativa, modelos Baseados em Conteúdo e modelos híbridos, com o objetivo de comparar seu desempenho preditivo e qualidade de ranqueamento. A metodologia envolve análise exploratória, construção da matriz usuário–item, ajuste de hiperparâmetros e avaliação por métricas de acurácia (RMSE e MAE) e de ranqueamento (Precision@k e Recall@k). Os resultados indicam que os modelos híbridos apresentam desempenho superior às abordagens individuais, destacando-se a combinação entre Filtragem Colaborativa e Baseada em Conteúdo, que oferece melhor equilíbrio entre precisão, qualidade das recomendações e mitigação do problema de *cold-start*. Conclui-se que a integração de múltiplas fontes de informação é uma estratégia eficaz para Sistemas de Recomendação aplicados a jogos de tabuleiro modernos.

Palavras-chave: Sistemas de recomendação. Jogos de tabuleiro. Filtragem colaborativa. Método *baseline*. Filtragem baseada em conteúdo. Modelos híbridos.

ABSTRACT

SILVA, I. **Recommendation Systems: A Approach On Modern Board Games.** 2026. 73p. Monografia (MBA em Ciências de Dados) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2026.

This work analyzes the application of Recommendation Systems in the context of modern board games using real-world data from the *BoardGameGeek* platform. Different approaches are evaluated, including Collaborative Filtering, Content-Based models, and hybrid models, with the aim of comparing their predictive performance and ranking quality. The methodology comprises exploratory data analysis, construction of the user-item matrix, hyperparameter tuning, and evaluation using accuracy metrics (RMSE and MAE) and ranking metrics (Precision@k and Recall@k). The results indicate that hybrid models outperform individual approaches, with the combination of Collaborative Filtering and Content-Based methods providing a better balance between accuracy, recommendation quality, and mitigation of the *cold-start* problem. It is concluded that integrating multiple sources of information is an effective strategy for Recommendation Systems applied to modern board games.

Keywords: Recommendation Systems. Board Games. Collaborative Filtering. Baseline Method. Content-Based Filtering. Hybrid Models.

LISTA DE FIGURAS

Figura 1 – Distribuição das Notas Atribuídas pelos Usuários	51
Figura 2 – Distribuição de Avaliações por Usuário e por Jogo (Cauda Longa)	53
Figura 3 – Distribuição de Avaliações por Usuário e por Jogo (Cauda Longa)	54
Figura 4 – Relação entre Popularidade e Nota Média dos Jogos	55
Figura 5 – Relação entre Popularidade (Escala Logarítmica) e Nota Média dos Jogos	55
Figura 6 – Sensibilidade do hiperparâmetro k — Métricas de acurácia do modelo FC User–User.	58
Figura 7 – Sensibilidade do hiperparâmetro k — Métricas de ranking do modelo FC User–User.	58
Figura 8 – Análise de sensibilidade do hiperparâmetro k para o modelo BC Item– Item considerando métricas de acurácia.	60
Figura 9 – Análise de sensibilidade do hiperparâmetro k para o modelo BC Item– Item considerando métricas de ranking.	60
Figura 10 – Comparação entre os modelos híbridos FC+Baseline e FC+BC em termos de métricas de acurácia para diferentes valores do coeficiente de hibridização β	64
Figura 11 – Comparação entre os modelos híbridos FC+Baseline e FC+BC em termos de métricas de ranking para diferentes valores do coeficiente de hibridização β	64
Figura 12 – Comparativo das métricas de acurácia dos modelos avaliados.	65
Figura 13 – Comparativo das métricas de ranking dos modelos avaliados.	65

LISTA DE TABELAS

Tabela 1 – Resumo das métricas de desempenho do modelo FC User–User para diferentes valores de k .	57
Tabela 2 – Resumo das métricas de desempenho do modelo BC para diferentes valores de k .	59
Tabela 3 – Comparativo de desempenho dos modelos avaliados	61
Tabela 4 – Validação do Modelo: Previsões para Jogos Não Avaliados	66
Tabela 5 – Métricas de desempenho dos modelos híbridos FC + BL e FC + BC para diferentes valores de β .	73

LISTA DE QUADROS

Quadro 1 – Filtragem temporal e consolidação das avaliações	50
Quadro 2 – Aplicação do Dense Sampling	51

LISTA DE ABREVIATURAS E SIGLAS

ABRINQ	Associação Brasileira de Fabricantes de Brinquedos
BC	Filtragem Baseada em Conteúdo
BGG	BoardGameGeek
BL	Modelo Baseline
CF	Filtragem Colaborativa (<i>Collaborative Filtering</i>)
DL	Aprendizagem Profunda (<i>Deep Learning</i>)
DNN	Redes Neurais Profundas (<i>Deep Neural Networks</i>)
EDA	Análise Exploratória de Dados
FC	Filtragem Colaborativa
FC+BC	Modelo Híbrido: Filtragem Colaborativa + Conteúdo
FC+BL	Modelo Híbrido: Filtragem Colaborativa + Baseline
KNN	K-Vizinhos Mais Próximos (<i>K-Nearest Neighbors</i>)
MAE	Erro Absoluto Médio (<i>Mean Absolute Error</i>)
ML	Aprendizagem de Máquina (<i>Machine Learning</i>)
MIT	Instituto de Tecnologia de Massachusetts (<i>Massachusetts Institute of Technology</i>)
P@10	Precisão no Top-10 (<i>Precision@10</i>)
R@10	Revocação no Top-10 (<i>Recall@10</i>)
RMSE	Raiz do Erro Quadrático Médio (<i>Root Mean Squared Error</i>)
SR	Sistema de Recomendação
SRs	Sistemas de Recomendação
SVD	Decomposição em Valores Singulares (<i>Singular Value Decomposition</i>)
SGD	Gradiente Descendente Estocástico (<i>Stochastic Gradient Descent</i>)
SVM	Máquina de Vetores de Suporte (<i>Support Vector Machine</i>)

TF-IDF Frequênciā do Termo - Frequênciā Inversa do Documento (*Term Frequency - Inverse Document Frequency*)

LISTA DE SÍMBOLOS

β	Coeficiente de hibridização do modelo híbrido
μ	Média global das avaliações
b_u	Viés associado ao usuário u
b_i	Viés associado ao item i
R	Matriz de avaliações usuário–item
$\mathbb{R}^{m \times n}$	Espaço das matrizes reais com m linhas e n colunas
\mathbb{R}^n	Espaço dos vetores reais com n dimensões
$sim(u, v)$	Similaridade entre os usuários u e v
$sim(i, j)$	Similaridade entre os itens i e j
r_{ui}	Avaliação do usuário u para o item i
\hat{r}_{ui}	Avaliação prevista para o usuário u e item i
m	Número total de usuários
n	Número total de itens
k	Número de vizinhos considerados no modelo
$N(u)$	Conjunto de vizinhos do usuário u
I_u	Conjunto de itens avaliados pelo usuário u
\mathcal{K}	Conjunto de avaliações conhecidas
$O(m^2)$	Complexidade computacional quadrática
$O(m)$	Complexidade computacional linear
$\hat{r}_{u,i}^{SR_1}$ e $\hat{r}_{u,i}^{SR_2}$	Avaliações previstas pelos Sistemas de Recomendação 1 e 2, respectivamente
$\hat{r}_{u,i}^H$	Avaliação prevista pelo modelo híbrido
Σ	Somatório

SUMÁRIO

1	INTRODUÇÃO	25
1.1	Os Jogos de Tabuleiro Modernos	25
1.2	Contribuições Educacionais e Sociais dos Jogos de Tabuleiro	26
1.3	O Mercado dos Jogos de Tabuleiro	27
1.4	Os Sistemas de Recomendação	28
1.5	Objetivos do Trabalho	28
2	METODOLOGIA	31
2.1	O que são Sistemas de Recomendação	31
2.2	Objetivos dos Sistemas de Recomendação	32
2.3	Breve Histórico e Evolução	33
2.4	Tipos e Abordagens de Sistemas de Recomendação	33
2.4.1	Filtragem Colaborativa	34
2.4.1.1	Abordagem baseada em usuários	34
2.4.1.2	Abordagem baseada em itens	34
2.4.2	Sistemas Baseados em Conteúdo	35
2.4.2.1	Representação de Atributos e Similaridade	35
2.4.3	Predição Baseada em Vizinhança de Conteúdo	35
2.4.4	Modelo Baseline (Ajuste por Viés)	36
2.4.4.1	Estimação dos Parâmetros	36
2.4.4.1.1	1. Média Global (μ):	36
2.4.4.1.2	2. Viés do Item (b_i):	37
2.4.4.1.3	3. Viés do Usuário (b_u):	37
2.4.4.2	Vantagens do Modelo	37
2.4.5	Sistemas Híbridos	37
2.4.5.1	Integração de Previsões	37
2.5	Desafios Comuns em Sistemas de Recomendação	38
2.6	Algoritmos e Métricas em Sistemas de Recomendação	39
2.6.1	Medidas de Similaridade	39
2.6.1.0.1	Similaridade do Cosseno:	39
2.6.1.0.2	Distância Euclidiana:	39
2.6.1.0.3	Coeficiente de Correlação de Pearson:	40
2.6.1.0.4	Coeficiente de Jaccard:	40
2.7	Metodologia de Avaliação	40
2.7.1	Avaliação Online e Offline	41
2.7.2	Matriz de Avaliação e Divisão dos Dados	41

2.7.3	Métricas de Acurácia	42
2.7.3.1	Métricas de predição.	42
2.7.3.1.1	Métricas de ranqueamento.	42
2.8	Métricas Complementares	43
3	DADOS E VARIÁVEIS	45
3.1	Origem dos dados	45
3.2	Estrutura da Base de Dados	45
3.3	Seleção e Integração das Tabelas Utilizadas	46
4	DESENVOLVIMENTO E AVALIAÇÃO DO SISTEMA DE RECOMENDAÇÃO	49
4.1	Implementação e Análise Prática	49
4.2	Análise Exploratória e Pré-processamento dos Dados	49
4.2.1	Filtragem Temporal e Limpeza dos Dados	50
4.2.2	Estratégia de Amostragem Densa	50
4.2.3	Análise Exploratória do Conjunto de Dados	51
4.3	Definição da Matriz de Avaliações R	56
4.4	Estratégia de Divisão dos Dados para Treinamento e Teste	56
4.5	Treinamento e Parametrização dos Modelos	56
4.5.1	Treinamento da Filtragem Colaborativa (FC)	57
4.5.2	Treinamento da Filtragem Baseada em Conteúdo (BC)	59
4.5.3	Treinamento do Modelo Baseline	61
4.6	Estratégia de Integração: O Modelo Híbrido	61
4.6.1	Avaliação dos Modelos Híbridos	63
4.7	Modelo em Produção	65
4.7.1	Validação do Modelo com Jogos Não Avaliados pelo Usuário	66
5	CONCLUSÃO E TRABALHOS FUTUROS	67
	REFERÊNCIAS	69
	APÊNDICES	71
	APÊNDICE A – MÉTRICAS DE DESEMPENHO DOS MODELOS HÍBRIDOS	73

1 INTRODUÇÃO

Este trabalho propõe o desenvolvimento e análise de sistemas de recomendação aplicados ao contexto dos jogos de tabuleiro modernos, com base em dados reais extraídos do BoardGameGeek e disponibilizados no Kaggle. A pesquisa busca compreender como diferentes abordagens — colaborativa, baseada em conteúdo e híbrida — podem auxiliar jogadores na descoberta de novos títulos alinhados aos seus interesses e preferências. Ao final, espera-se obter um modelo capaz de recomendar jogos relevantes, avaliando seu desempenho por meio de métricas quantitativas e qualitativas e discutindo suas limitações e potenciais de aplicação prática.

1.1 Os Jogos de Tabuleiro Modernos

Não é possível determinar com precisão o período e tampouco qual foi o primeiro jogo de tabuleiro a ser criado na história da humanidade. Entretanto, há indícios de que esses jogos remontam às primeiras civilizações já que as peças de jogo mais antigas descobertas foram encontradas no sudoeste da Turquia há cerca de 5.000a.C.

Um dos candidatos mais reconhecidos é o Royal Game of Ur, originário da antiga Mesopotâmia há cerca de 4.600 anos, com exemplares datados entre 2600 e 2400 a.C. Outro jogo muito retratado nos registros históricos é o Mancala. De origem africana, o jogo é caracterizado pelo uso de tabuleiro e de sementes como peças. Estima-se que o Mancala já fosse praticado por volta de 2.500 a.C., tendo se transformado ao longo do tempo em mais de 200 variações, resultado de adaptações e inovações culturais ([Wikipedia, 2025](#)).

Outros jogos clássicos também marcaram a história e permanecem presentes até os dias atuais como o Senet, jogado no Antigo Egito há mais de 5.000 anos; o Go, de origem chinesa, que data de aproximadamente 2.500 a.C.; e o Xadrez, surgido na Índia por volta do século VII d.C. Já em períodos mais recentes, outros jogos se tornaram igualmente icônicos e seguem como parte de nossa história civilizatória. Exemplos disso são a Damas, difundida a partir da Idade Média; o Dominó, com raízes na China e popularizado no Ocidente a partir do século XVIII; e jogos do século XX que alcançaram projeção global, como o Banco Imobiliário (Monopoly) e o Jogo da Vida (The Game of Life). Todos esses jogos citados são chamados pelos amantes do hobby de “jogos clássicos”.

O termo “jogos de tabuleiro modernos” passou a ser utilizado principalmente a partir da década de 1990, período marcado pelo sucesso internacional Colonizadores de Catan (lançado em 1995 pelo designer Klaus Teuber). Considerado um divisor de águas no setor, Catan introduziu elementos inovadores, como a forte interação entre os jogadores por meio de trocas de recursos, mecânicas de desenvolvimento estratégico e uma dinâmica

que equilibrava competição e cooperação ([LUDOPEDIA, 2023](#)).

Embora não exista uma definição oficial para o conceito, a expressão “jogo de tabuleiro moderno” costuma ser empregada para designar títulos que se diferenciam dos clássicos por suas características próprias. Esses jogos geralmente apresentam regras, que favorecem a acessibilidade a diferentes públicos, ao mesmo tempo em que oferecem profundidade estratégica. Também se caracterizam por partidas mais curtas e dinâmicas, nas quais os jogadores conseguem permanecer engajados até o final da partida, evitando eliminações precoces e longos períodos de espera para o início de um próximo jogo.

Outro traço marcante é a ênfase nas decisões estratégicas em detrimento do acaso, de modo que a vitória depende muito mais do planejamento e das escolhas feitas ao longo da partida do que da sorte em rolagens de dados ou sorteios de cartas. Além disso, as interações entre jogadores tendem a ser menos violentas e mais indiretas, centradas na disputa por recursos, rotas ou territórios. A estética também ocupa lugar de destaque, já que os jogos modernos são concebidos para proporcionar uma experiência visualmente agradável, com componentes de alta qualidade e temáticas imersivas.

Entre os estilos mais relevantes dentro desse movimento, destacam-se os chamados ameritrash, que valorizam narrativas fortes, conflito direto e alta imersão temática, e os eurogames, ou estilo alemão, que priorizam regras acessíveis, equilíbrio entre jogadores e uma abordagem mais estratégica e inclusiva. Este último, em particular, tem sido amplamente reconhecido como o principal responsável pelo crescimento recente do hobby em escala global ([PAPERGAMES, 2021; PAPERGAMES, 2022](#)).

1.2 Contribuições Educacionais e Sociais dos Jogos de Tabuleiro

Os jogos de tabuleiro vêm ganhando destaque não apenas como forma de entretenimento, mas também pelos benefícios educacionais, sociais e formativos que proporcionam. Do ponto de vista educacional e do desenvolvimento pessoal, eles são eficazes na formação de mentalidades empreendedoras em jovens, ao estimular habilidades como pensamento estratégico, tomada de decisões e disposição para assumir riscos. Ao propor desafios que exigem planejamento e resolução de problemas, os jogos de tabuleiro se tornam ferramentas valiosas para o desenvolvimento de competências empreendedoras ([SHAIKH, 2023](#)).

Além disso, os jogos de tabuleiro desempenham um papel importante na promoção da interação social e no fortalecimento dos vínculos familiares. Eles criam um ambiente propício para a convivência, a cooperação e o diálogo entre pais e filhos, contribuindo para a construção de relações mais harmoniosas e significativas no núcleo familiar ([RAHMATILANI; PATRIANSAH, 2024](#)).

1.3 O Mercado dos Jogos de Tabuleiro

O mercado global de jogos de tabuleiro, também chamados de *board games*, é um setor em constante crescimento, com estimativas de movimentação significativas. Segundo dados da consultoria *Statista*, o mercado de jogos de tabuleiro no mundo movimentou mais de R\$ 53 bilhões apenas em 2024 ([Statista, 2025](#)).

De acordo com um relatório da *Grand View Research* ([Grand View Research, 2025](#)), o mercado global de jogos de tabuleiro foi avaliado em aproximadamente 19,9 bilhões de dólares em 2024 e projeta-se que alcance cerca de 31,93 bilhões de dólares até 2030, com uma taxa de crescimento anual composta (CAGR) de cerca de 8,3% durante de 2025 a 2030. No Brasil, o mercado de jogos de tabuleiro também tem mostrado um crescimento notável. Segundo dados do Anuário Estatístico de Brinquedos de 2025, elaborado pela Associação Brasileira de Fabricantes de Brinquedos (ABRINQ), os jogos de tabuleiros representaram 14,8% das vendas de brinquedos no país em 2024, um aumento de 5,7 p.p. em comparação aos 9,1% registrados em 2017, refletindo uma tendência de alta no segmento ([ABRINQ, 2025](#)).

Esse crescimento é resultado de várias condições. A globalização e a internet facilitaram o acesso a uma variedade maior de títulos, permitindo que jogadores de diferentes regiões descubram e compartilhem suas experiências. Plataformas de financiamento coletivo, como o Kickstarter, também desempenharam um papel importante ao possibilitar que designers independentes lancem seus jogos diretamente para o público, sem depender exclusivamente das grandes editoras. A pandemia de COVID-19 também contribuiu para esse crescimento, já que muitas pessoas buscaram alternativas de entretenimento em casa durante os períodos de isolamento social. Esses fatores impulsionaram a popularização dos jogos de tabuleiro modernos e a expansão das comunidades de jogadores.

O avanço desse mercado trouxe também novos desafios, especialmente para novos jogadores, que muitas vezes se deparam com uma vasta quantidade de títulos disponíveis. Embora existam plataformas especializadas que auxiliam na organização e divulgação desses jogos, como a Ludopedia, principal comunidade online de jogadores de tabuleiro no Brasil, e o *BoardGameGeek* (BGG), maior base de dados internacional sobre o tema, os recursos de avaliação, rankings e recomendações oferecidos por esses sites podem não ser suficientes para orientar as escolhas dos consumidores.

Nesse contexto, a aplicação de sistemas de recomendação surge como uma solução promissora para reduzir a sobrecarga de opções e facilitar a descoberta de jogos que atendam melhor aos diferentes perfis de público. Tal ferramenta pode não apenas potencializar o engajamento dos jogadores, mas também ampliar o acesso a experiências lúdicas que promovam o aprendizado, o convívio e o desenvolvimento pessoal.

1.4 Os Sistemas de Recomendação

Desde o inicio da internet, o volume de informações disponíveis cresceu exponencialmente, tornando cada vez mais difícil para os usuários encontrar conteúdos relevantes. Nesse contexto, os sistemas de recomendação surgiram como uma solução eficaz para filtrar e sugerir itens com base nas preferências e comportamentos dos usuários. Esses sistemas são amplamente utilizados em diversas plataformas, como serviços de streaming (*Netflix, Spotify*), e-commerce (*Amazon, eBay*) e redes sociais (*Facebook, Instagram*), ajudando os usuários a descobrir produtos, filmes, músicas e outros conteúdos que possam ser do seu interesse ([RICCI; ROKACH; SHAPIRA, 2010](#)).

Os sistemas de recomendação são algoritmos projetados para sugerir itens relevantes a usuários com base em seus históricos de interação, preferências declaradas ou comportamentos semelhantes de outros usuários. Eles podem ser classificados em três categorias principais: filtragem colaborativa, filtragem baseada em conteúdo e sistemas híbridos ([AGGARWAL et al., 2016](#)).

1.5 Objetivos do Trabalho

Este trabalho tem por objetivo explorar a aplicação de sistemas de recomendação no contexto dos jogos de tabuleiro modernos, visando auxiliar os jogadores na descoberta de novos títulos que se alinhem com seus gostos e preferências. A seguir, serão apresentados os conceitos fundamentais dos sistemas de recomendação, suas técnicas e metodologias. Além disso, será feita uma análise comparativa sobre os desempenho das três categorias de algoritmos no contexto da base de dados selecionada para este trabalho. Ao final, serão apresentadas as conclusões e sugestões para trabalhos futuros, destacando o potencial impacto de sistemas de recomendação na experiência dos jogadores de jogos de tabuleiro modernos.

De maneira mais específica, este trabalho busca:

- Compreender os fundamentos dos sistemas de recomendação e suas abordagens principais;
- Pré-processar os dados da base Kaggle de forma adequada para o desenvolvimento dos sistemas de recomendação;
- Avaliar o desempenho de diferentes algoritmos de recomendação: Filtragem Colaborativa, Baseada em Conteúdo e Híbrida;
- Identificar as limitações e potenciais de aplicação prática dos sistemas de recomendação;

- Avaliar o desempenho dos modelos com métricas de erro preditivo (como RMSE e MAE) e métricas complementares de experiência do usuário (como diversidade e novidade).

2 METODOLOGIA

Na era da sobrecarga de informações, onde explosão de conteúdos e produtos online torna desafiador para os usuários encontrar itens relevantes, os *Sistemas de Recomendação* (SRs) emergem como ferramentas essenciais para personalizar a experiência do usuário. Eles utilizam algoritmos avançados para analisar dados de comportamento e preferências, oferecendo sugestões que facilitam a tomada de decisão. Esta seção explora os fundamentos dos SRs, suas abordagens principais, desafios comuns, algoritmos utilizados e métricas de avaliação, além de destacar suas aplicações em diversos domínios.

2.1 O que são Sistemas de Recomendação

Sistemas de Recomendação (SR) são ferramentas de software e algoritmos, frequentemente baseadas em inteligência artificial que fornecem sugestões personalizadas de itens relevantes aos usuários. Nesse sentido, o termo “item” é usado para se referir a qualquer entidade que possa ser recomendada, como produtos, serviços, conteúdos digitais, ou até mesmo pessoas (em redes sociais ou plataformas de namoro) ([ADOMAVICIUS; TUZHILIN, 2005](#); [RICCI; ROKACH; SHAPIRA, 2010](#); [AGGARWAL et al., 2016](#)).

O principal objetivo dos SRs é apoiar os usuários em processos de tomada de decisão, atuando como uma estratégia eficaz para mitigar a sobrecarga de informações, especialmente em ambientes online com grande volume de dados.

A tarefa de recomendação pode ser formulada de diferentes maneiras, sendo os dois enfoques mais comuns:

1. Versão Preditiva: a abordagem consiste na solução de um problema de regressão onde o sistema prevê os valores da classificação para uma combinação usuário-item, como a previsão de uma nota de 1 a 5 estrelas que um usuário pode atribuir a um filme ou produto. Por suposição, tem-se que os dados de treinamento são compostos por avaliações explícitas ou implícitas fornecidas pelos usuários. Para m usuários e n itens, tem-se uma matriz $m \times n$ incompleta, o objetivo então é prever a avaliação r_{ui} desconhecida que o usuário u daria ao item i . Esse problema é conhecido como “problema de complementação de matriz” ([AGGARWAL et al., 2016](#));
2. Versão de Classificação: nesta abordagem não é necessária a predição de uma nota específica, mas sim a identificação dos k principais itens que um usuário pode gostar ou determinar os k principais usuários a serem segmentados a um item específico, sendo a primeira abordagem mais comum.

Apesar de no seu surgimento os SRs serem predominantemente desenvolvidos para auxiliar os usuários no problema da “Tirania da Escolha” como descreveu Ricci, Rokach e Shapira (2010), os SRs representam um instrumento estratégico para empresas que buscam aumentar o engajamento do usuário, a retenção de clientes e, consequentemente, suas receitas. Plataformas como *Amazon*, *Netflix*, *Spotify*, *YouTube*, *Instagram* e *TikTok* são exemplos notáveis de como os SRs podem transformar a experiência do usuário, oferecendo recomendações personalizadas que facilitam a descoberta de novos produtos ou conteúdos, diversificando seu catálogo e fidelizando os consumidores.

2.2 Objetivos dos Sistemas de Recomendação

Segundo Aggarwal et al. (2016) o objetivo de um SR é o aumento da receita sobre a venda de produtos, serviços ou conteúdos. Ao sugerirem itens cuidadosamente selecionados de acordo com as preferências do usuário, esses sistemas direcionam a atenção para produtos relevantes, elevando o volume de consumo e, consequentemente, a receita do provedor. Para atingir esse objetivo mais amplo, Aggarwal et al. (2016) destaca quatro metas operacionais e técnicas mais comuns em SRs: relevância, novidade, serendipidade e diversidade.

A relevância refere-se à capacidade do sistema de recomendar itens que realmente interessam ao usuário, aumentando a probabilidade de interação e satisfação. A novidade diz respeito à introdução de itens novos ou pouco conhecidos, incentivando a exploração e descoberta de conteúdos ou produtos que o usuário ainda não experimentou. A serendipidade envolve a recomendação de itens inesperados, mas que acabam sendo apreciados pelo usuário, proporcionando uma experiência agradável e surpreendente. Por fim, a diversidade busca oferecer uma gama variada de recomendações, evitando a monotonia e ampliando o leque de opções disponíveis para o usuário.

Entretanto, em complemento a visão acima, Ricci, Rokach e Shapira (2010) destacam que a função de um sistema de recomendação vai além da busca pelo lucro, podendo desempenhar multiplas funções dentro de um ecossistema de recomendação. Além de recomendar itens isolados (“*find some good items*”), eles podem propor conjuntos coerentes de itens (“*recommend a bundle*”), sequências ordenadas (“*recommend a sequence*”), auxiliar na navegação exploratória de catálogos (“*just browsing*”), permitir que o usuário expresse suas preferências (“*express self*”) ou contribua com avaliações que beneficiem a comunidade (“*help others*”). Essas diferentes funções revelam que os sistemas de recomendação devem equilibrar os interesses de provedores e usuários, oferecendo valor mútuo e sustentando a dinâmica de personalização que caracteriza a economia da informação.

Enquanto Aggarwal et al. (2016) enfatiza a maximização de receita por meio da experiência do usuário, Ricci et al. ampliam a compreensão sobre o papel dos sistemas de recomendação, destacando funções que nem sempre estão diretamente ligadas ao lucro,

mas que apoiam a interação, exploração e satisfação do usuário

2.3 Breve Histórico e Evolução

O campo de investigação científica em SR evoluiu significativamente a partir dos anos 1990. O crédito pelo pioneirismo é frequentemente atribuído a Jussi Karlgren, que em 1990 elaborou uma álgebra para recomendação, visando melhorar a interação humano-computador com modelos estatísticos ([KARLGREN, 1994](#)).

Em 1992, David Goldberg et al. desenvolveram e publicaram o software Tapestry, que utilizava o conceito de “filtragem colaborativa” para recomendar documentos de notícias por e-mail, sendo considerado o primeiro SR computacional. Em 1994, o projeto GroupLens, do MIT e da Universidade de Minnesota, apoiou investigações em SR, preocupando-se com a sobrecarga de informações e propondo uma estrutura para que os usuários pudessem avaliar mensagens em newsgroups ([GOLDBERG et al., 1992](#)).

Em 2006, a *Netflix* lançou o *Netflix Prize*, oferecendo um milhão de dólares para quem desenvolvesse um sistema de recomendação com pelo menos 10% mais acurácia do que o Cinematch, seu sistema da época, evidenciando a importância da pesquisa na área ([AGGARWAL et al., 2016](#)).

Após o impulso dado pelo *Netflix Prize*, tornou-se comum a utilização de técnicas de Aprendizagem de Máquina (ML) em SRs. Por exemplo, em se tratando da abordagem usando classificação, algoritmos como Máquinas de Vetores de Suporte (SVM), Árvores de Decisão, Classificadores Bayesianos e K-Vizinhos Mais Próximos (KNN) passaram a ser empregados para melhorar a precisão das recomendações.

Nos últimos anos, SRs têm sido fortemente associados a técnicas de Aprendizagem de Máquina (ML), incluindo Aprendizagem Profunda (Deep Learning - DL) e Redes Neurais Profundas (Deep Neural Networks - DNN), para construir recomendações relevantes e eficientes ([AZAMBUJA; MORAIS; FILIPE, 2021](#)).

2.4 Tipos e Abordagens de Sistemas de Recomendação

Os sistemas de recomendação utilizam dois tipos principais de dados: (i) interações usuário-item, como avaliações, histórico de compras ou comportamento de navegação, e (ii) atributos de usuários e itens, como perfis textuais, características demográficas ou palavras-chave relevantes.

Os métodos que se baseiam nas interações são chamados de filtragem colaborativa, enquanto os que exploram atributos recebem o nome de baseados em conteúdo. Embora esses sistemas frequentemente usem matrizes de classificação, os métodos baseados em conteúdo concentram-se nas avaliações de um único usuário, em vez de considerar todo o conjunto de usuários. Já os sistemas baseados em conhecimento utilizam requisitos

explicitamente fornecidos pelo usuário e informações externas para gerar recomendações, sem depender de dados históricos de avaliação ou compra. Por fim, os sistemas híbridos combinam diferentes abordagens, integrando os pontos fortes de cada método para fornecer recomendações mais precisas e consistentes em diferentes contextos.

2.4.1 Filtragem Colaborativa

A Filtragem Colaborativa (FC) é uma abordagem central em sistemas de recomendação, baseada na ideia de que *usuários com preferências semelhantes tendem a gostar de itens semelhantes*. Essa abordagem utiliza interações históricas entre usuários e itens, como avaliações, compras ou cliques, que são organizadas em uma matriz de interações:

$$R = \begin{bmatrix} r_{1,1} & r_{1,2} & \dots & r_{1,n} \\ r_{2,1} & r_{2,2} & \dots & r_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m,1} & r_{m,2} & \dots & r_{m,n} \end{bmatrix}, \quad (2.1)$$

onde $R \in \mathbb{R}^{m \times n}$ representa as avaliações de m usuários em n itens, e $r_{u,i}$ é a avaliação ou interação do usuário u com o item i . A matriz R é geralmente esparsa, pois nem todos os usuários interagem com todos os itens.

2.4.1.1 Abordagem baseada em usuários

Nesta abordagem, a similaridade entre usuários é calculada a partir das linhas da matriz R . Usuários com comportamentos similares são considerados “vizinhos”, e recomenda-se a um usuário itens que seus vizinhos avaliaram positivamente. A previsão da avaliação $\hat{r}_{u,i}$ de um usuário u para um item i é dada por:

$$\hat{r}_{u,i} = \bar{r}_u + \frac{\sum_{v \in N(u)} \text{sim}(u, v) \cdot (r_{v,i} - \bar{r}_v)}{\sum_{v \in N(u)} |\text{sim}(u, v)|}, \quad (2.2)$$

onde:

- \bar{r}_u é a média das avaliações do usuário u ;
- $N(u)$ é o conjunto de vizinhos mais similares a u ;
- $\text{sim}(u, v)$ é a similaridade entre usuários u e v ;
- $r_{v,i}$ é a avaliação do usuário v para o item i .

2.4.1.2 Abordagem baseada em itens

Na abordagem baseada em itens, a similaridade é medida entre itens, considerando as colunas da matriz R . A previsão da avaliação de um usuário u para um item i é:

$$\hat{r}_{u,i} = \frac{\sum_{j \in I_u} \text{sim}(i, j) \cdot r_{u,j}}{\sum_{j \in I_u} |\text{sim}(i, j)|}, \quad (2.3)$$

onde:

- I_u é o conjunto de itens previamente avaliados pelo usuário u ;
- $\text{sim}(i, j)$ é a similaridade entre os itens i e j .

2.4.2 Sistemas Baseados em Conteúdo

De acordo com [Aggarwal et al. \(2016\)](#), os Sistemas de Recomendação Baseados em Conteúdo (BC) fundamentam-se na premissa de que a relevância de um item para um usuário pode ser determinada pela similaridade entre os atributos desse item e as preferências históricas manifestadas pelo próprio usuário. Diferentemente da filtragem colaborativa, esta abordagem não utiliza informações da comunidade de usuários (*peer users*), tornando-se robusta ao problema de *cold-start* de novos itens, desde que seus atributos sejam conhecidos.

2.4.2.1 Representação de Atributos e Similaridade

Nesta abordagem, a representação dos itens é crucial. Cada item i é descrito por um vetor de características \mathbf{x}_i extraído de seus metadados (como palavras-chave, gênero ou, neste trabalho, mecânicas de jogo). A relação entre dois itens é quantificada por uma função de similaridade $\text{Sim}(i, j)$.

Segundo [Aggarwal et al. \(2016\)](#), a escolha da métrica de similaridade deve refletir a natureza dos dados:

- **Similaridade do Coseno:** Padrão para dados vetoriais esparsos e contínuos (como representações de texto TF-IDF).
- **Coeficiente de Jaccard:** Recomendado para atributos binários ou categóricos (conjuntos), onde o interesse reside na sobreposição de características presentes.

2.4.3 Predição Baseada em Vizinhança de Conteúdo

Embora seja possível treinar modelos de classificação complexos (como Bayes ou Árvores de Decisão) para cada usuário, uma abordagem eficiente e amplamente utilizada é a Classificação de Vizinhos Mais Próximos (k -NN) aplicada ao Conteúdo, conforme descrito na Seção 4.4 de [Aggarwal et al. \(2016\)](#).

Neste método, a predição da nota $\hat{r}_{u,i}$ para um item alvo i é calculada através da média ponderada das avaliações que o usuário u atribuiu aos k itens mais similares a i em seu próprio histórico.

Define-se $N_u(i)$ como o conjunto de vizinhos de conteúdo do item i restrito às avaliações do usuário u . Formalmente:

$$\hat{r}_{u,i} = \frac{\sum_{j \in N_u(i)} \text{sim}(i, j) \cdot r_{u,j}}{\sum_{j \in N_u(i)} |\text{sim}(i, j)|}, \quad (2.4)$$

onde:

- $N_u(i)$ é o subconjunto de itens previamente avaliados pelo usuário u que possuem maior similaridade de conteúdo com o item alvo i .
- $r_{u,j}$ é a nota real dada pelo usuário ao item histórico j .

Essa formulação permite que a recomendação seja diretamente explicável pelos atributos compartilhados entre o item recomendado e o histórico do usuário.

2.4.4 Modelo Baseline (Ajuste por Viés)

Frequentemente, as avaliações em Sistemas de Recomendação não refletem apenas a interação específica entre um usuário e um item, mas também carregam tendências sistemáticas — ou vieses — que independem dessa interação. Segundo Aggarwal et al. (2016), o Modelo Baseline assume que a classificação r_{ui} pode ser decomposta em uma componente global e vieses específicos.

A previsão da avaliação $\hat{r}_{u,i}$ é dada pela soma de três termos:

$$\hat{r}_{u,i} = \mu + b_u + b_i, \quad (2.5)$$

onde μ é a média global, b_u é o viés do usuário e b_i é o viés do item.

2.4.4.1 Estimação dos Parâmetros

Embora esses parâmetros possam ser estimados através da minimização de mínimos quadrados regularizados (ALS ou SGD), este trabalho adota a abordagem heurística de desacoplamento de médias, que oferece alta eficiência computacional sem perda significativa de precisão para bases densas. O cálculo é realizado em três etapas sequenciais:

2.4.4.1.1 1. Média Global (μ):

Calculada como a média aritmética de todas as avaliações conhecidas no conjunto de treinamento \mathcal{K} :

$$\mu = \frac{\sum_{(u,i) \in \mathcal{K}} r_{u,i}}{|\mathcal{K}|}. \quad (2.6)$$

2.4.4.1.2 2. Viés do Item (b_i):

Representa o desvio médio das notas recebidas pelo item i em relação à média global. Um valor positivo indica um item popular (acima da média). É calculado para cada item i como:

$$b_i = \frac{\sum_{u \in U_i} (r_{u,i} - \mu)}{|U_i|}, \quad (2.7)$$

onde U_i é o conjunto de usuários que avaliaram o item i .

2.4.4.1.3 3. Viés do Usuário (b_u):

Representa o desvio médio do usuário u em relação à média, após descontar o efeito da popularidade do item. Isso impede que um usuário pareça generoso apenas porque avaliou jogos muito populares. É calculado sobre o resíduo:

$$b_u = \frac{\sum_{i \in I_u} (r_{u,i} - \mu - b_i)}{|I_u|}, \quad (2.8)$$

onde I_u é o conjunto de itens avaliados pelo usuário u .

2.4.4.2 Vantagens do Modelo

A principal vantagem desta abordagem reside na sua eficiência computacional. Diferentemente de modelos de vizinhança que exigem cálculos quadráticos $O(m^2)$, o Modelo Baseline possui complexidade linear $O(N)$ em relação ao número de avaliações. Além disso, por possuir poucos parâmetros, o modelo é robusto contra *overfitting*, servindo como um excelente *benchmark* de precisão.

2.4.5 Sistemas Híbridos

Os Sistemas Híbridos combinam diferentes técnicas de recomendação para aproveitar os pontos fortes de cada abordagem e mitigar limitações.

2.4.5.1 Integração de Previsões

Sejam $\hat{r}_{u,i}^{SR_1}$ e $\hat{r}_{u,i}^{SR_2}$ as previsões de dois SRs distintos. Uma combinação linear simples desses dois modelos gera a previsão híbrida:

$$\hat{r}_{u,i}^H = \beta \hat{r}_{u,i}^{SR_1} + (1 - \beta) \hat{r}_{u,i}^{SR_2}, \quad (2.9)$$

onde $\beta \in [0, 1]$ define a importância relativa de cada abordagem.

2.5 Desafios Comuns em Sistemas de Recomendação

O desempenho de um Sistema de Recomendação é fortemente dependente da qualidade, da estrutura e da quantidade dos dados disponíveis, e vários desafios podem surgir nesse contexto. A seguir, são discutidos alguns dos problemas mais comuns enfrentados por esses sistemas, bem como suas implicações e possíveis estratégias de mitigação.

Um dos problemas mais recorrentes é a esparsidade dos dados. Em bases reais, o número de itens disponíveis tende a ser muito maior do que o número de avaliações ou interações observadas, produzindo uma matriz usuário–item $R \in \mathbb{R}^{m \times n}$ em que a fração de valores conhecidos é pequena:

$$\text{densidade}(R) = \frac{|\{(u, i) \mid r_{ui} \text{ observado}\}|}{m \times n} \ll 1. \quad (2.10)$$

Essa característica dificulta a estimativa de preferências não observadas, prejudicando a generalização dos modelos e aumentando a incerteza nas previsões. Estratégias como fatoração matricial regularizada ou modelos probabilísticos latentes são comumente aplicadas para mitigar esse problema, impondo restrições de baixa dimensionalidade sobre as representações de usuários e itens.

Outro desafio é o problema do novo usuário ou novo item, conhecido como *cold start*. Nesse cenário, não há interações suficientes para inicializar o perfil de um novo elemento no sistema. No caso de um novo usuário, o vetor de preferências \mathbf{p}_u é inicialmente indefinido; para um novo item, o vetor de características \mathbf{q}_i não pode ser estimado a partir de avaliações históricas. Soluções típicas envolvem abordagens híbridas que incorporam atributos de conteúdo (como gênero, descrição textual ou metadados) para complementar a ausência de feedback explícito.

Em contextos de larga escala, surge o problema da escalabilidade computacional. O crescimento exponencial do número de usuários (m) e de itens (n) impõe restrições severas à viabilidade de execução de algoritmos clássicos. Métodos baseados em vizinhança, por exemplo, exigem cálculos de similaridade $O(m^2)$ ou $O(n^2)$, tornando-se inviáveis para bases massivas. Mesmo modelos de fatoração matricial, cuja complexidade típica de treinamento é $O(kmn)$ para k fatores latentes, podem tornar-se onerosos quando m e n alcançam a ordem de milhões.

Estratégias de mitigação incluem a adoção de técnicas de *sampling* e decomposição incremental, além do uso de arquiteturas distribuídas de processamento, como *MapReduce* ou *parameter servers*.

A dinâmica temporal das preferências é outro aspecto crítico. As preferências dos usuários variam ao longo do tempo, de modo que um modelo estático $\hat{r}_{ui} = f(\mathbf{p}_u, \mathbf{q}_i)$ pode tornar-se obsoleto. Modelos temporais introduzem uma dependência explícita de

tempo, $f_t(\mathbf{p}_{u,t}, \mathbf{q}_{i,t})$, ou incorporam esquemas de decaimento exponencial nas observações históricas, ponderando mais fortemente as interações recentes. Essa adaptação temporal é especialmente desafiadora em sistemas com fluxos contínuos de dados e milhões de transações diárias.

2.6 Algoritmos e Métricas em Sistemas de Recomendação

Os Sistemas de Recomendação (SRs) empregam diferentes algoritmos para estimar o grau de similaridade entre usuários ou itens, etapa fundamental para gerar previsões e recomendações personalizadas. A escolha adequada da métrica influencia diretamente a precisão das recomendações e a eficiência computacional do sistema, especialmente diante dos desafios de esparsidade, heterogeneidade e escalabilidade dos dados.

2.6.1 Medidas de Similaridade

As medidas de similaridade buscam quantificar o grau de proximidade entre vetores de características — que podem representar perfis de usuários, itens ou interações entre ambos. Seja $A, B \in \mathbb{R}^n$ dois vetores correspondentes às avaliações ou representações latentes de usuários ou itens. As métricas mais utilizadas incluem a *similaridade do cosseno*, a *distância euclidiana*, a *correlação de Pearson* e o *coeficiente de Jaccard*, cada qual adequada a diferentes tipos de dados e contextos de recomendação.

2.6.1.0.1 Similaridade do Cosseno:

A similaridade do cosseno mede o ângulo entre dois vetores no espaço \mathbb{R}^n , sendo uma métrica independente da magnitude, ideal para dados esparsos e avaliações normalizadas. É definida como:

$$\text{similaridade}_{\cos}(A, B) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}. \quad (2.11)$$

O valor resultante pertence ao intervalo $[-1, 1]$, onde 1 indica alinhamento perfeito (máxima similaridade), 0 ausência de correlação e -1 oposição total.

2.6.1.0.2 Distância Euclidiana:

A distância euclidiana mede a proximidade geométrica entre dois vetores e é expressa por:

$$d(A, B) = \sqrt{\sum_{i=1}^n (A_i - B_i)^2}. \quad (2.12)$$

Trata-se de uma métrica intuitiva, simétrica e não negativa. Entretanto, é sensível à escala e à magnitude dos dados, sendo menos robusta em contextos de alta esparsidade.

2.6.1.0.3 Coeficiente de Correlação de Pearson:

A correlação de Pearson avalia a relação linear entre dois vetores, removendo o efeito de diferenças de escala e de médias individuais. É amplamente empregada em filtragem colaborativa, principalmente quando as avaliações são normalizadas em torno da média do usuário:

$$r_{A,B} = \frac{\sum_{i=1}^n (A_i - \bar{A})(B_i - \bar{B})}{\sqrt{\sum_{i=1}^n (A_i - \bar{A})^2} \sqrt{\sum_{i=1}^n (B_i - \bar{B})^2}}. \quad (2.13)$$

onde \bar{A} e \bar{B} representam as médias das avaliações de A e B , respectivamente. Valores próximos de 1 indicam alta correlação positiva, enquanto valores próximos de -1 indicam correlação inversa.

2.6.1.0.4 Coeficiente de Jaccard:

Quando as interações são binárias (por exemplo, presença ou ausência de uma compra, clique ou avaliação), utiliza-se o coeficiente de Jaccard, definido como:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}. \quad (2.14)$$

Essa métrica mede a proporção de itens compartilhados entre dois conjuntos, sendo ideal para bases implícitas ou altamente esparsas.

Cada métrica apresenta vantagens específicas. A similaridade do cosseno é eficiente e robusta para dados esparsos e vetores normalizados; a distância euclidiana é mais intuitiva, mas sensível à escala; a correlação de Pearson captura variações de preferência relativas, tornando-a mais interpretável em avaliações subjetivas; e o coeficiente de Jaccard é particularmente útil para sistemas com dados binários ou interações implícitas. Em ambientes de larga escala, a escolha da métrica deve considerar também o custo computacional, visto que o cálculo de similaridade entre milhões de vetores pode se tornar o principal gargalo de desempenho dos SRs.

2.7 Metodologia de Avaliação

A avaliação de sistemas de recomendação (SR) é uma etapa essencial para compreender a efetividade e a robustez dos algoritmos empregados. De acordo com Aggarwal et al. (2016), os SR podem ser avaliados por meio de abordagens *online* ou *offline*, sendo

esta última a mais utilizada em pesquisas devido à disponibilidade de conjuntos de dados históricos e à possibilidade de reproduções dos experimentos.

2.7.1 Avaliação Online e Offline

A distinção entre avaliação *online* e *offline* é fundamental, constituindo os dois paradigmas principais de análise de desempenho, conforme discutido por Aggarwal et al. (2016).

A avaliação online é conduzida em sistemas implantados e acessados por usuários reais, frequentemente utilizando testes do tipo *A/B*. Seu objetivo é medir métricas operacionais e de comportamento, como a taxa de conversão (frequência com que um item recomendado é selecionado), engajamento, receita ou lucro. Trata-se do método mais fidedigno para aferir a eficácia do sistema a longo prazo, embora apresente custo elevado e dificuldade de controle experimental.

Já a avaliação offline baseia-se em conjuntos de dados históricos que simulam o comportamento do usuário. Nessa abordagem, parte das avaliações conhecidas é ocultada, e o algoritmo é testado quanto à sua capacidade de prever os valores omitidos. É o método mais comum em pesquisas e competições (como o *Netflix Prize*), permitindo testar a acurácia de diferentes algoritmos sob condições controladas e reproduzíveis.

2.7.2 Matriz de Avaliação e Divisão dos Dados

Para garantir uma avaliação imparcial, é essencial que as observações utilizadas no treinamento não sejam reaproveitadas na fase de teste. Seja $R \in \mathbb{R}^{m \times n}$ a matriz de avaliações, em que cada elemento r_{uj} representa a nota atribuída pelo usuário u ao item j . O sistema gera uma estimativa \hat{r}_{uj} , e a diferença entre valores reais e previstos define o erro:

$$e_{uj} = \hat{r}_{uj} - r_{uj}. \quad (2.15)$$

Para evitar sobreajuste, o conjunto de entradas observadas $S = \{(u, j) \mid r_{uj} \text{ conhecido}\}$ é dividido em:

- **Treinamento:** usado para ajustar o modelo e aprender parâmetros latentes;
- **Validação:** utilizado para ajuste de hiperparâmetros e seleção de modelos;
- **Teste:** reservado exclusivamente para mensurar a acurácia final.

As principais estratégias de particionamento incluem o método *Hold-Out*, em que uma fração dos dados é retida para teste, e a Validação Cruzada (*Cross-Validation*), na qual o conjunto de observações é dividido em q subconjuntos e a média de desempenho é calculada ao longo das iterações.

2.7.3 Métricas de Acurácia

A acurácia constitui o critério mais fundamental de avaliação, sendo dividida conforme o objetivo do sistema: prever o valor exato de uma avaliação (*rating prediction*) ou ordenar itens por relevância (*ranking prediction*).

2.7.3.1 Métricas de predição.

Essas métricas medem o erro entre o valor previsto \hat{r}_{uj} e o valor real r_{uj} . As mais utilizadas incluem:

- **Erro Absoluto Médio (MAE):**

$$MAE = \frac{1}{|E|} \sum_{(u,j) \in E} |r_{uj} - \hat{r}_{uj}|. \quad (2.16)$$

É mais robusto a valores extremos, pois não amplifica grandes erros.

- **Raiz do Erro Quadrático Médio (RMSE):**

$$RMSE = \sqrt{\frac{1}{|E|} \sum_{(u,j) \in E} (r_{uj} - \hat{r}_{uj})^2}. \quad (2.17)$$

Penaliza mais fortemente grandes discrepâncias, sendo a métrica oficial do *Netflix Prize*.

2.7.3.1.1 Métricas de ranqueamento.

Quando o sistema gera listas ordenadas (*Top-k Recommendation*), avalia-se a posição dos itens relevantes. Algumas dessas métricas são:

- **Precisão a k (Precision@k):** Mede a fração de itens recomendados no topo da lista que são relevantes. É uma métrica de pureza da recomendação. Matematicamente:

$$P@k = \frac{|\{\text{itens relevantes}\} \cap \{\text{top-}k \text{ itens recomendados}\}|}{k}. \quad (2.18)$$

Uma alta precisão indica que o usuário não precisa navegar por muitos itens ruins para encontrar um bom.

- **Revocação a k (Recall@k):** Mede a fração de *todos* os itens relevantes conhecidos do usuário que o sistema conseguiu incluir na lista dos top-*k*. É uma métrica de cobertura.

$$R@k = \frac{|\{\text{itens relevantes}\} \cap \{\text{top-}k \text{ itens recomendados}\}|}{|\{\text{total de itens relevantes do usuário}\}|}. \quad (2.19)$$

Um alto recall indica que o sistema é capaz de encontrar a maioria dos jogos que o usuário gosta, evitando que boas opções passem despercebidas.

2.8 Métricas Complementares

Embora a acurácia seja o critério operacional primário, ela não captura plenamente a qualidade percebida das recomendações. Segundo Aggarwal et al. (2016) algumas métricas complementares são necessárias para avaliar aspectos subjetivos da experiência do usuário, destacam-se:

- **Novidade (Novelty):** mede a probabilidade de o sistema recomendar itens ainda não conhecidos pelo usuário, estimulando a descoberta de novos conteúdos.
- **Serendipidade (Serendipity):** quantifica a frequência de recomendações inesperadas, mas relevantes, refletindo a capacidade do sistema de surpreender positivamente.
- **Diversidade (Diversity):** avalia a variedade entre os itens recomendados, geralmente pela dissimilaridade média entre pares de itens na lista Top-k.
- **Cobertura (Coverage):** indica a proporção de usuários ou itens para os quais o sistema consegue gerar recomendações válidas, podendo ser definida como cobertura de usuário ou de catálogo.
- **Confiança (Confidence) e Confiança do Usuário (Trust):** a primeira reflete a incerteza do modelo em sua predição, enquanto a segunda refere-se à percepção subjetiva de credibilidade pelo usuário.
- **Robustez e Estabilidade:** medem a capacidade do sistema em manter desempenho consistente frente a ruídos, mudanças temporais ou ataques intencionais (*shilling attacks*).
- **Escalabilidade:** avalia a eficiência computacional em termos de tempo de treinamento, tempo de predição e uso de memória, especialmente em contextos de grandes volumes de dados.

Algumas dessas métricas — como relevância, novidade, serendipidade e diversidade — já foram discutidas neste trabalho como objetivos secundários dos sistemas de recomendação, conforme apresentado na Seção 2.2. Na etapa de avaliação, esses mesmos conceitos podem ser quantificados por meio de diferentes abordagens métricas, permitindo analisar o quanto o sistema consegue gerar recomendações úteis, diversificadas e interessantes a longo prazo.

Neste trabalho, essas métricas são apresentadas apenas de forma conceitual, já que o cálculo prático exigiria testes adicionais e maior capacidade computacional. Ainda assim, elas são importantes para mostrar que a qualidade de um sistema de recomendação vai além da simples precisão das previsões.

3 DADOS E VARIÁVEIS

3.1 Origem dos dados

O *Kaggle* ([Kaggle, 2025](#)) é uma plataforma online voltada para ciência de dados e aprendizado de máquina. Ela oferece competições com dados reais, notebooks interativos, datasets públicos e um fórum para colaboração entre cientistas de dados de todo o mundo. Também serve como repositório gratuito de dados para projetos acadêmicos e profissionais.

Entre os muitos conjuntos de dados hospedados no *Kaggle*, destaca-se a base derivada do *BoardGameGeek* (BGG) — um dos maiores bancos de dados e comunidades online dedicadas a jogos de tabuleiro. No BGG, os usuários registram suas coleções, atribuem avaliações, compartilham resenhas e discutem estratégias, o que torna o site uma referência consolidada tanto para a indústria quanto para a academia em análises envolvendo jogos de tabuleiro.

3.2 Estrutura da Base de Dados

A base de dados “*Board Games Database from BoardGameGeek*”, hospedada no próprio *Kaggle* e organizada pelo usuário *threnjen* reune dados públicos do site *BoardGameGeek*, incluindo informações a cerca de 22.000 jogos, 411.000 usuários e aproximadamente 19 milhões de avaliações. Segundo o próprio autor ([threnjen, 2025](#)) a base é ideal para análises exploratórias, modelagem preditiva e sistemas de recomendação.

Este conjunto de dados inclui nove arquivos potenciais para exploração e/ou modelagem que são:

1. **games.csv**: o arquivo de informações básicas com 47 classificações sobre cada um dos 22 mil jogos de tabuleiro. A chave primária é **BGGId**, que é o ID do jogo.
2. **ratings_distribution.csv**: inclui a distribuição completa das classificações para cada **BGGId**
3. **themes.csv**: tabela de temas para cada **BGGId**
4. **mechanics.csv**: tabela de mecânicas com sinalizadores binários por **BGGId**
5. **subcategories.csv**: tabela de subcategorias com sinalizadores binários por **BGGId**
6. **artists_reduce**: fornece informações sobre os artistas para cada **BGGId**. Este arquivo é reduzido para artistas com menos de 3 obras, com um sinalizador binário indicando que o jogo incluiu um artista com 3 ou mais obras.

7. **designers_reduced**: fornece informações do designer para cada BGGId. Este arquivo é reduzido para designers com menos 3 obras, com um sinalizador binário indicando que o jogo incluiu um designer com 3 ou mais obras.
8. **publishers_reduced**: fornece informações da editora para cada BGGId. Este arquivo é reduzido para editoras com menos de 3 obras, com um sinalizador binário indicando que o jogo incluiu uma editora com 3 ou mais obras.
9. **user_ratings.csv**: contém todas as avaliações para todos os BGGId com nome de usuário. Existem mais de 411 mil usuários únicos e aproximadamente 19 milhões de avaliações.

3.3 Seleção e Integração das Tabelas Utilizadas

Embora a base do *BoardGameGeek* contenha nove arquivos distintos, nem todos são igualmente relevantes para a construção de um sistema de recomendação eficiente. Considerando os objetivos deste trabalho — desenvolver um modelo de recomendação que une viabilidade computacional e representatividade dos dados — foram selecionadas apenas três tabelas principais: **user_ratings**, **games** e **mechanics**.

A escolha se justifica pelos seguintes motivos:

- **user_ratings**: Essencial para recomendações baseadas em filtragem colaborativa (FC). Como essa tabela contém as avaliações dos usuários, ela permite construir a matriz usuário–item, que é fundamental para entender quem avaliou o quê e com qual nota já que contém a ID do jogo (BGGId), o nome do usuário (Username) e a avaliação (Rating).
- **games**: Fornece atributos globais dos jogos, como nome (Name), ano de publicação (YearPublished), avaliação média (AvgRating), categorias principais, entre outros. Esses atributos são úteis na identificação dos jogos baseadas nos BGGId presentes na tabela de avaliações dos usuários, além de permitir a extração de características úteis para limpeza e análise exploratória dos dados.
- **mechanics**: É uma tabela de características binárias que indicam se um jogo possui determinadas mecânicas (como “Deck Building”). Essas características são extremamente úteis para recomendações por similaridade de conteúdo (BC), pois permitem identificar jogos com atributos semelhantes, facilitando a recomendação de jogos que compartilham elementos específicos apreciados pelos usuários.

A combinação dessas três fontes é suficiente para representar, de forma equilibrada, as dimensões de interação (usuário–item) e conteúdo (atributos dos jogos), preservando o potencial analítico sem comprometer a eficiência computacional do sistema. Outros

arquivos da base (*Themes*, *Publishers*, *Designers*, entre outros) foram descartados por apresentarem menor relevância direta para o modelo ou por aumentarem significativamente o volume de dados sem ganho proporcional de informação.

4 DESENVOLVIMENTO E AVALIAÇÃO DO SISTEMA DE RECOMENDAÇÃO

4.1 Implementação e Análise Prática

Neste capítulo, será apresentada a aplicação prática dos conceitos discutidos anteriormente em 2. O objetivo principal é desenvolver e validar quatro algoritmos de recomendação distintos para jogos de tabuleiro modernos, avaliando sua precisão preditiva e eficiência na geração de listas de sugestões personalizadas.

A abordagem prática divide-se em três etapas fundamentais: primeiramente, o tratamento exploratório e a amostragem estratégica dos dados brutos; em seguida, a parametrização e o cálculo dos modelos de vizinhança e viés; e, por fim, a análise comparativa de desempenho através de métricas de acurácia (RMSE e MAE) e de ranqueamento (Precision e Recall). Através desta metodologia, busca-se identificar qual dos modelos propostos oferece uma melhor experiência de recomendação.

Foram implementados os seguintes modelos:

- **Filtragem Colaborativa (FC) - User-User:** a similaridade foi calculada usando a Correlação de Pearson que escolhida por sua capacidade computacionalmente eficiente de capturar relações lineares entre usuários, ajustando-se bem às variações individuais de escala nas avaliações.
- **Filtragem Baseada em Conteúdo (BC) - Item-Item:** neste modelo, foi escolhida a similaridade utilizando a Distância de Jaccard uma vez que ela é particularmente eficaz para medir a similaridade entre conjuntos binários, como os vetores binários de mecânicas de jogo, permitindo identificar jogos com características semelhantes.
- **Modelo Baseline - Viés Global, de Item e de Usuário:** este é um modelo de fácil implementação que permite capturar os vieses inerentes aos usuários e itens, proporcionando uma linha de base sólida para comparação.
- **Modelo Híbrido - Combinação Linear Ponderada entre dois Modelos:** esta implementação visou aproveitar as forças de dois os modelos e foi implementado com a finalidade de mitigar as limitações individuais de cada abordagem, especialmente em cenários de *cold-start*, além da tentativa de melhorar a precisão geral das recomendações.

4.2 Análise Exploratória e Pré-processamento dos Dados

Para a implementação do Sistema de Recomendação, foram utilizados três conjuntos de dados principais extraídos da plataforma BoardGameGeek (BGG): `user_ratings.csv`

(notas dos usuários para determinados jogos), `games.csv` (metadados dos jogos) e `mechanics.csv` (atributos de conteúdo).

O volume original dos dados brutos (aproximadamente 380 MB no arquivo de avaliações) impõe desafios de memória e custo computacional para operações iterativas de modelagem. Desta forma, antes da aplicação dos algoritmos, os dados foram submetidos a um *pipeline* de pré-processamento composto por etapas de limpeza, filtragem temporal e redução de dimensionalidade, visando garantir tanto a viabilidade computacional quanto a relevância das recomendações.

4.2.1 Filtragem Temporal e Limpeza dos Dados

No âmbito da filtragem temporal, optou-se por manter apenas jogos publicados a partir de 1990 (*Modern Board Games*). Esta decisão concentra a análise nos ditos jogos modernos de tabuleiro.

Em seguida, observou-se a presença de múltiplas avaliações para um mesmo par usuário-item na base bruta. Para resolver essa inconsistência, os dados foram agrupados por usuário e jogo, calculando-se a média das notas atribuídas, garantindo assim a unicidade nas entradas da futura matriz de utilidade.

Quadro 1 – Filtragem temporal e consolidação das avaliações

```
# Filtragem para manter apenas jogos lançados a partir de 1990
games_modern = games[games['YearPublished'] >= 1990]

# Consolidação de notas duplicadas (média por usuário/jogo)
ratings = ratings.groupby(['Username', 'BGGId'],
    as_index=False)['Rating'].mean()
```

4.2.2 Estratégia de Amostragem Densa

Dada a esparsidate da base original, foi necessária uma estratégia robusta de redução de dimensionalidade para viabilizar o cálculo das matrizes de similaridade. Em vez de uma amostragem aleatória simples, que poderia agravar a esparsidate, optou-se por uma abordagem de *Dense Sampling* (Amostragem Densa).

Esta técnica consiste em selecionar o subnúcleo mais ativo da plataforma, maximizando a densidade de interações:

- **Seleção de Usuários:** Foram mantidos os 5.000 usuários com maior número de avaliações registradas.

- Seleção de Itens:** Dentre os jogos avaliados por esses usuários, foram mantidos os 2.000 jogos mais populares (maior volume de notas).

Quadro 2 – Aplicação do Dense Sampling

```
# Selecionar os top 5.000 usuarios mais ativos
top_users = ratings['Username'].value_counts().nlargest(5000).index
ratings = ratings[ratings['Username'].isin(top_users)]

# Selecionar os top 2.000 jogos mais populares neste subgrupo
top_games = ratings['BGGId'].value_counts().nlargest(2000).index
ratings = ratings[ratings['BGGId'].isin(top_games)]
```

Após o cruzamento dessas restrições, a base final de trabalho resultou em 2.217.792 avaliações observadas. A densidade resultante desta amostra é significativamente superior aos padrões de bases esparsas de sistemas de recomendação, favorecendo a convergência dos algoritmos.

4.2.3 Análise Exploratória do Conjunto de Dados

Foram gerados gráficos e estatísticas descritivas com o objetivo de compreender melhor a distribuição das variáveis após o pré-processamento. A Figura 1 apresenta o histograma das notas atribuídas pelos usuários aos jogos.

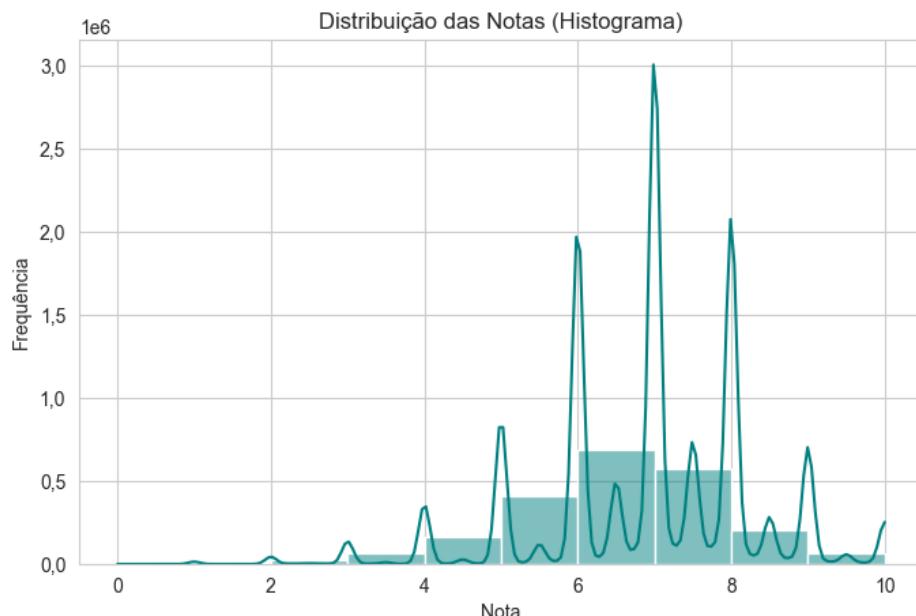


Figura 1 – Distribuição das Notas Atribuídas pelos Usuários

Observa-se que a distribuição das avaliações é fortemente concentrada em valores elevados, principalmente no intervalo entre 6 e 8 refletindo o viés de positividade durante o processo de avaliação, no qual os usuários tendem a avaliar predominantemente jogos que apreciam.

A distribuição das notas mostra ainda que os usuários tendem a avaliar os jogos de forma predominantemente discreta ou, em menor proporção, com incrementos de 0,5 ponto indicando que as notas representam preferências subjetivas, e não uma escala contínua precisa. Isto implica dizer que pequenas diferenças numéricas entre notas previstas e reais podem não ser perceptíveis para o usuário no momento da recomendação. Essa condição reforça a necessidade de utilizar métricas complementares às de erro absoluto, como métricas de ranqueamento, uma vez que pequenas diferenças numéricas entre notas podem não representar diferenças relevantes na percepção real dos usuários.

As estatísticas descritivas corroboram essa observação: a média das notas é igual a 6,93, enquanto a mediana é 7,00, indicando uma distribuição centrada em valores altos. A assimetria negativa moderada ($-0,53$) revela a presença de uma cauda mais longa à esquerda, evidenciando que avaliações muito baixas são menos frequentes. Além disso, a curtose positiva (0,98) indica uma maior concentração de observações em torno da média quando comparada a uma distribuição normal.

Essa característica da distribuição possui implicações diretas para os modelos de recomendação utilizados neste trabalho. A concentração das notas em torno da média global pode favorecer modelos baseados em vieses, como o Modelo Baseline visto em [2.4.4](#), e pode dificultar a obtenção de grandes ganhos absolutos em métricas de erro, como RMSE e MAE. Por esse motivo, métricas de ranking, como *Precision@K* e *Recall@K*, são fundamentais para uma avaliação mais adequada da qualidade das recomendações.

A Figura [2](#) apresenta a distribuição do número de avaliações realizadas por usuário. Observa-se que a maioria dos usuários possui um volume moderado de avaliações, concentrando-se entre aproximadamente 200 e 600 registros. Entretanto, nota-se a presença de uma cauda longa à direita, indicando a existência de usuários extremamente ativos, responsáveis por um grande número de avaliações.

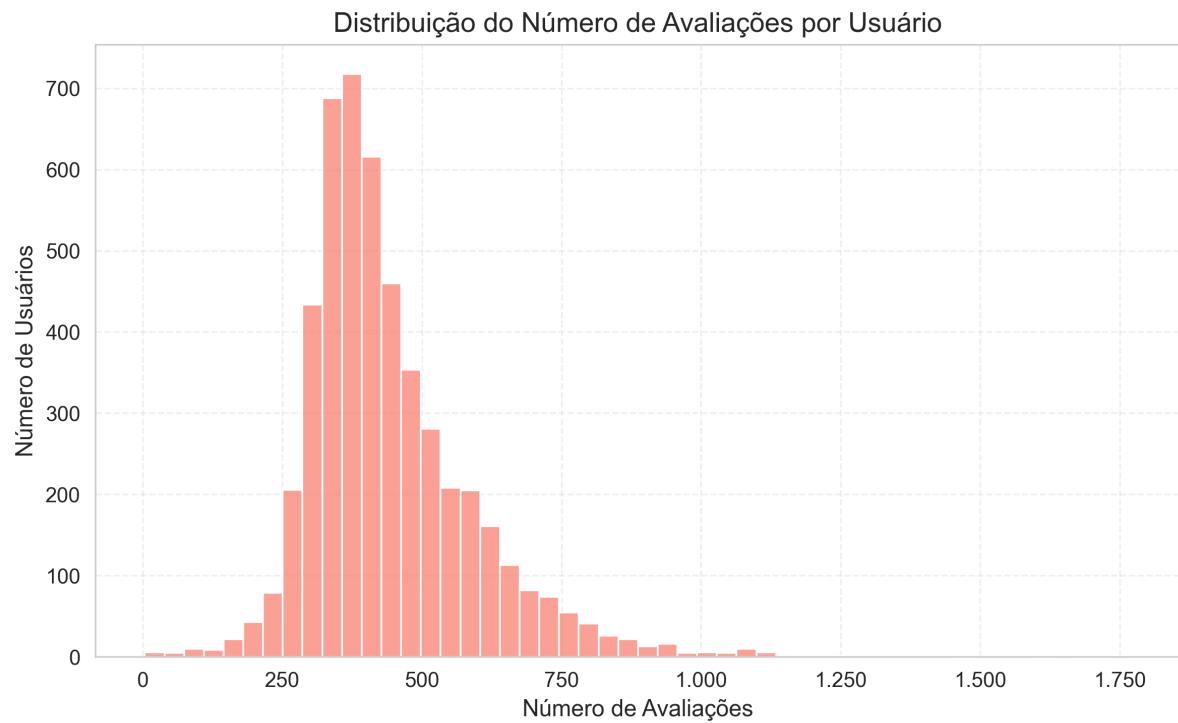


Figura 2 – Distribuição de Avaliações por Usuário e por Jogo (Cauda Longa)

Esse comportamento revela uma distribuição assimétrica à direita e evidencia a heterogeneidade no engajamento dos usuários. Tal característica pode impactar diretamente nos modelos de Filtragem Colaborativa baseados em usuários, uma vez que usuários com poucas avaliações tendem a gerar estimativas de similaridade menos estáveis, enquanto usuários mais ativos contribuem para perfis mais robustos.

A Figura 3 apresenta a distribuição do número de avaliações por jogo. Observa-se um padrão típico de sistemas de recomendação reais, caracterizado por uma forte concentração de jogos com poucas avaliações e uma cauda longa à direita, representando um pequeno conjunto de jogos altamente populares.

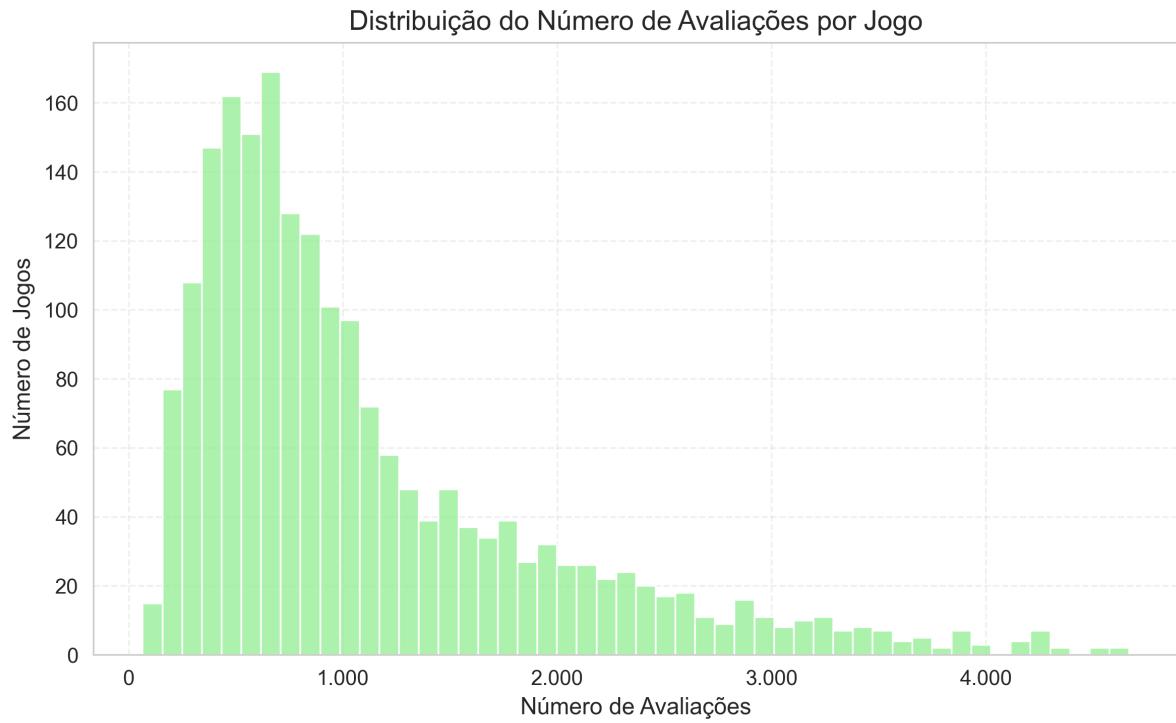


Figura 3 – Distribuição de Avaliações por Usuário e por Jogo (Cauda Longa)

Essas características reforçam a natureza altamente esparsa da matriz usuário–item e justificam a adoção de estratégias complementares, como modelos Baseados em Conteúdo e abordagens híbridas, capazes de mitigar os efeitos do *cold-start* e da concentração de popularidade.

A Figura 4 apresenta a relação entre a popularidade dos jogos, medida pelo número de avaliações recebidas, e a respectiva nota média. No gráfico em escala linear, observa-se elevada concentração de pontos na região de baixa popularidade, reflexo do fenômeno de cauda longa visto anteriormente.

Para uma análise mais informativa, a Figura 5 apresenta a mesma relação utilizando escala logarítmica no eixo da popularidade. Essa representação evidencia que jogos com poucas avaliações apresentam elevada variabilidade na nota média, enquanto jogos mais populares tendem a apresentar estimativas mais estáveis, com valores concentrados em torno da média global.

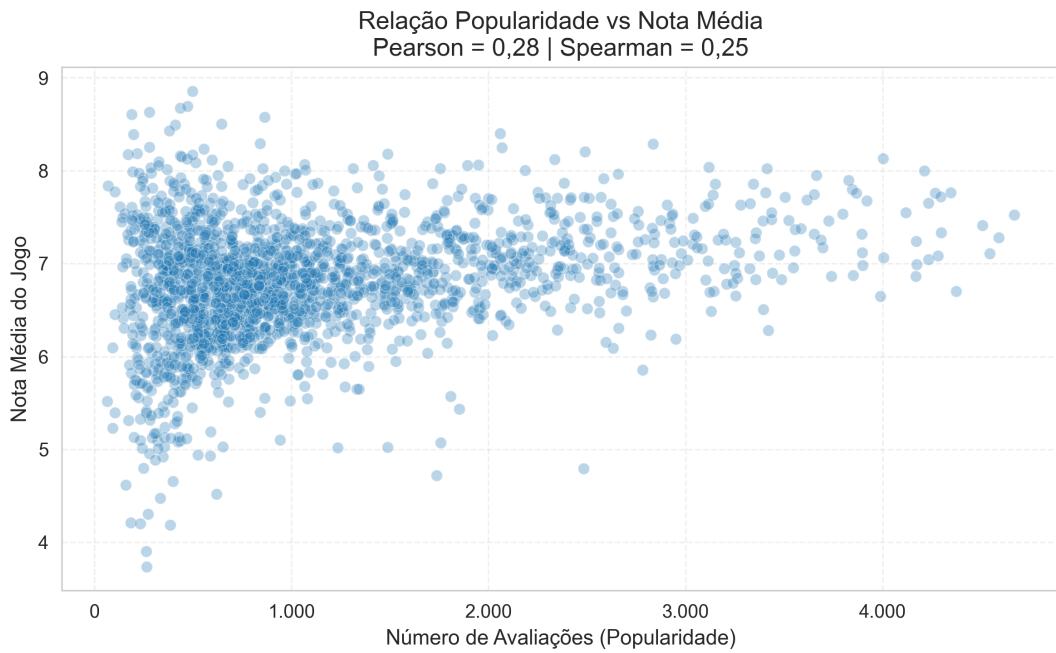


Figura 4 – Relação entre Popularidade e Nota Média dos Jogos

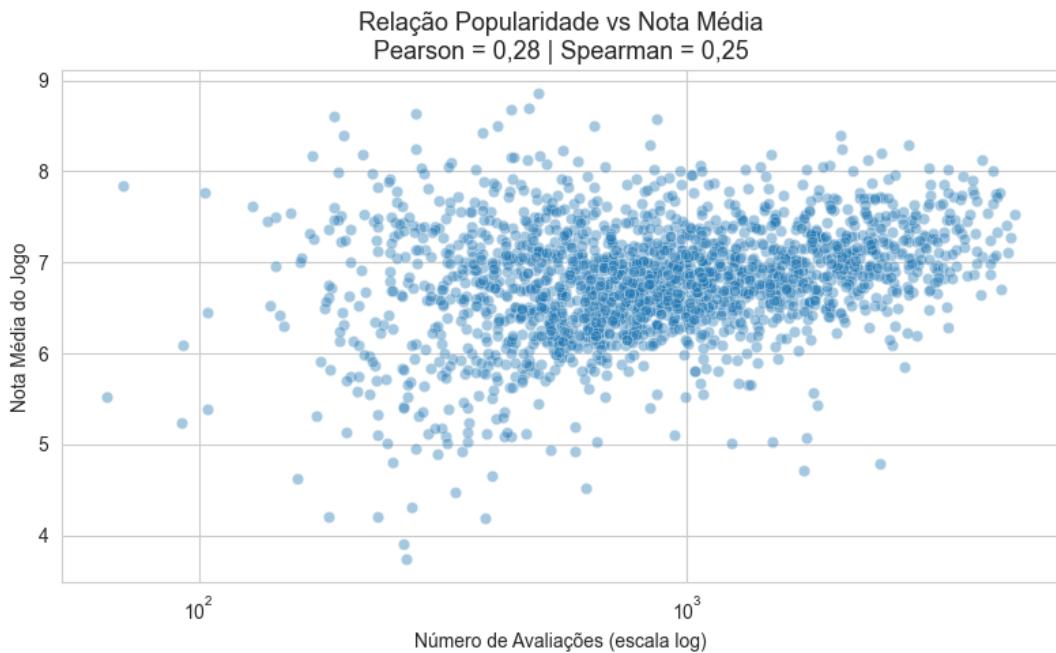


Figura 5 – Relação entre Popularidade (Escala Logarítmica) e Nota Média dos Jogos

A correlação entre popularidade e nota média é positiva, porém fraca, com coeficientes de Pearson igual a 0,28 e Spearman igual a 0,25. A proximidade entre essas medidas indica a ausência de uma relação linear forte, sugerindo que a popularidade, por si só, não é um fator determinante da avaliação média dos jogos.

Esses resultados evidenciam um fenômeno de regressão à média, no qual o aumento do número de avaliações reduz a variabilidade da nota média. Tal característica possui

implicações diretas para os modelos de recomendação avaliados neste trabalho, reforçando a importância de abordagens que combinem informações colaborativas e baseadas em conteúdo, bem como o uso de métricas de ranking para uma avaliação mais adequada da qualidade das recomendações.

4.3 Definição da Matriz de Avaliações R

Após as etapas de filtragem e amostragem, os dados foram estruturados na matriz de avaliações R , fundamental para os modelos de recomendação. Define-se $R \in \mathbb{R}^{m \times n}$, onde $m = 5.000$ (usuários) e $n = 2.000$ (jogos), tal que:

$$R = [r_{uj}], \quad r_{uj} = \begin{cases} \text{nota atribuída pelo usuário } u \text{ ao jogo } j, & \text{se observada;} \\ 0, & \text{caso contrário.} \end{cases} \quad (4.1)$$

Esta matriz serve como entrada principal para os cálculos de similaridade (Pearson e Coseno) e para a decomposição em fatores latentes nos capítulos subsequentes.

4.4 Estratégia de Divisão dos Dados para Treinamento e Teste

A divisão dos dados para treinamento e teste foi realizada de forma estratificada por usuário, e não por uma divisão aleatória global (*global shuffle*). A abordagem foi escolhida afim de evitar o vazamento de dados (*data leakage*) e garantir que o teste simule um cenário real de recomendação.

O processo de divisão seguiu as seguintes regras:

1. **Separação por Usuário:** Para cada usuário, 20% de suas avaliações foram separadas para o conjunto de teste e 80% mantidas no conjunto de treino.
2. **Critério de Histórico Mínimo:** Apenas usuários com mais de 5 avaliações na base final tiveram dados movidos para o conjunto de teste. Usuários com 5 ou menos interações foram mantidos integralmente no treino ou descartados do teste, visto que a remoção de dados poderia inviabilizar o cálculo de similaridade ou a própria validação (problema de *cold-start* no teste).

Essa estratégia garante que, no momento da validação, o modelo tenha acesso a um histórico parcial do usuário no treino (os 80%) para inferir suas preferências e tentar prever as interações ocultas (os 20% do teste).

4.5 Treinamento e Parametrização dos Modelos

A fase de treinamento consistiu no pré-cálculo das matrizes de similaridade e na estimativa dos parâmetros de viés (*model fitting*). Os métodos implementados utilizam

aprendizado baseado em memória e heurísticas estatísticas, onde o modelo treinado é representado pelas estruturas de dados resultantes descritas abaixo.

4.5.1 Treinamento da Filtragem Colaborativa (FC)

Para o modelo de Filtragem Colaborativa (abordagem User-User), conforme visto na 2.4.1.1, o treinamento focou na identificação de padrões de vizinhança entre os usuários com base em seus históricos de avaliação compartilhados.

Foi gerada uma matriz densa e simétrica de dimensões 5.000×5.000 , onde cada célula (u, v) armazena o coeficiente de correlação de Pearson descrito em 2.13, a medida foi usada para quantificar a similaridade entre o usuário u e o usuário v . Esta etapa incorreu em um custo computacional significativo, dada a complexidade quadrática em relação ao número de usuários.

Para esta abordagem, buscou-se identificar o melhor valor de k (número de vizinhos mais próximos) através de experimentação empírica. O número de vizinhos testados foram $k = \{10, 20, 30, 40, 50, 60, 80, 100\}$ e, após testes, foram contabilizadas métricas de acurácia (RMSE, MAE) e ranking (Precision@K e Recall@K) no conjunto de teste.

Ao final desses testes, obteve-se os seguinte resultados resumidos na Tabela 1 a seguir:

Tabela 1 – Resumo das métricas de desempenho do modelo FC User–User para diferentes valores de k .

Número de Vizinhos (k)	RMSE	MAE	Precision@10	Recall@10
10	1,1225	0,8451	0,8509	0,1727
20	1,1024	0,8275	0,8631	0,1762
30	1,0959	0,8220	0,8662	0,1769
40	1,0931	0,8195	0,8682	0,1775
50	1,0916	0,8181	0,8700	0,1780
60	1,0907	0,8173	0,8699	0,1779
80	1,0898	0,8163	0,8706	0,1781
100	1,0895	0,8159	0,8709	0,1781

A Figura 6 ilustra a sensibilidade das métricas RMSE e MAE em relação ao valor de k . Já a Figura 7 apresenta a variação das métricas de ranking (Precision@10 e Recall@10) conforme o valor de k .

Figura 6 – Sensibilidade do hiperparâmetro k — Métricas de acurácia do modelo FC User–User.

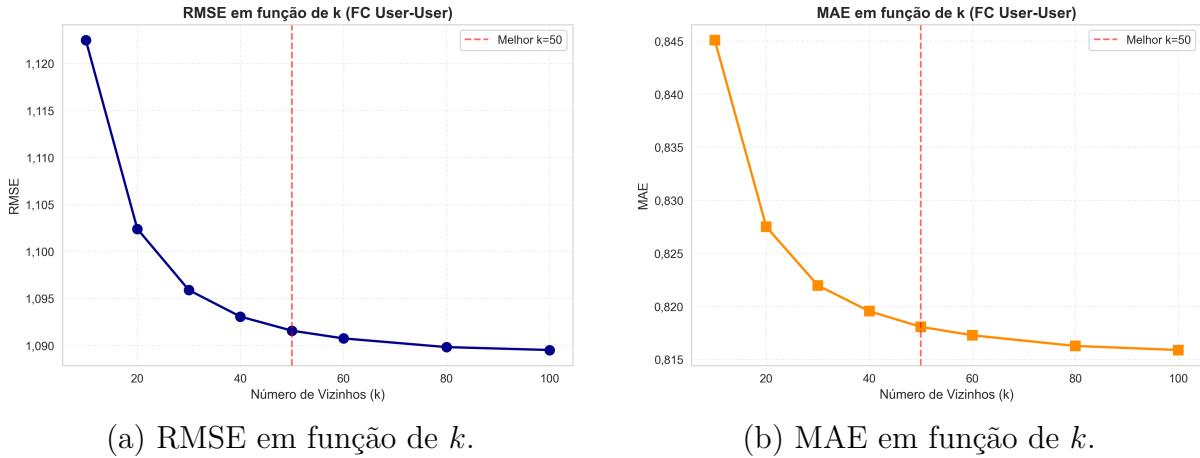
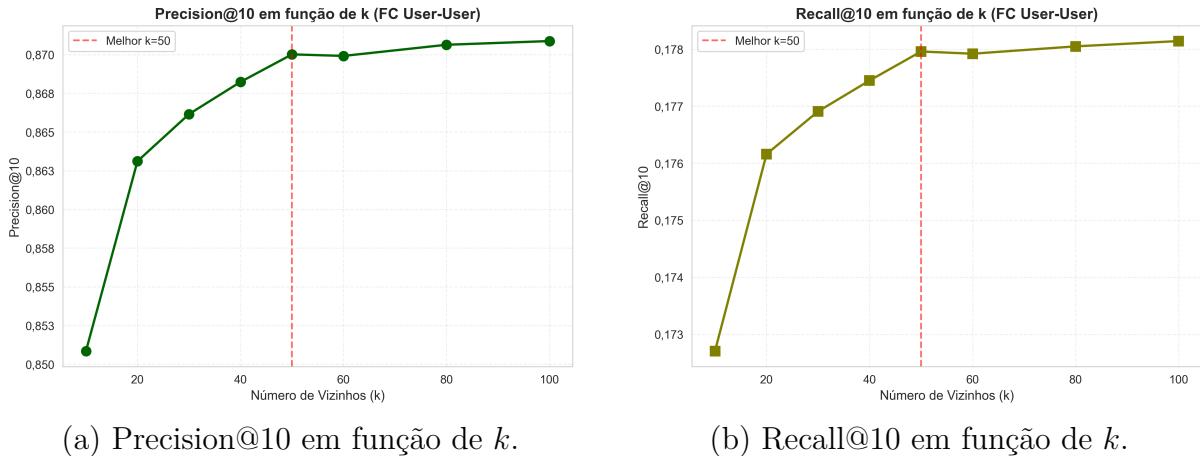


Figura 7 – Sensibilidade do hiperparâmetro k — Métricas de ranking do modelo FC User–User.



A análise de sensibilidade do hiperparâmetro k , apresentada nas Figuras 6 e 7, indica que o desempenho do modelo melhora significativamente para valores iniciais de k , especialmente até $k = 50$. A partir desse ponto, observa-se um regime de saturação, no qual os ganhos em termos de RMSE e MAE tornam-se marginais.

Em relação às métricas de ranking, nota-se que tanto a Precision@10 quanto o Recall@10 apresentam crescimento consistente até aproximadamente $k = 50$, estabilizando-se para valores superiores. Esse comportamento sugere que a inclusão de vizinhos adicionais além desse limiar tende a incorporar usuários com menor similaridade, contribuindo pouco para a qualidade das recomendações e potencialmente introduzindo ruído.

Considerando conjuntamente o desempenho preditivo, a estabilidade das métricas de ranking e o custo computacional associado ao aumento do número de vizinhos, adotou-se

$k = 50$ como valor final para o modelo FC User–User, representando um compromisso adequado entre acurácia, robustez e eficiência computacional.

4.5.2 Treinamento da Filtragem Baseada em Conteúdo (BC)

O modelo Baseado em Conteúdo (abordagem Item-Item) foi construído utilizando os atributos descritivos dos jogos, isto é, as mecânicas presentes em cada jogo descritas no arquivo `mechanics.csv`.

Diferente da FC, a similaridade aqui é estática e baseada em metadados. Utilizou-se a distância de Jaccard conforme visto na Seção 2.6.1.0.4 sobre os vetores binários de mecânicas para gerar uma matriz de dimensões 2.000×2.000 .

Assim como na abordagem FC, o valor de k (número de vizinhos mais próximos) foi determinado através de experimentação empírica. Foram testados os mesmos valores de $k = \{10, 20, 30, 40, 50, 60, 80, 100\}$ e as métricas de desempenho foram avaliadas no conjunto de teste.

A Tabela 2 resume os resultados obtidos:

Tabela 2 – Resumo das métricas de desempenho do modelo BC para diferentes valores de k .

Número de Vizinhos (k)	RMSE	MAE	Precision@10	Recall@10
10	1,2051	0,9034	0,7810	0,1575
20	1,1891	0,8903	0,7898	0,1592
30	1,1865	0,8877	0,7917	0,1595
40	1,1871	0,8878	0,7924	0,1595
50	1,1885	0,8887	0,7919	0,1593
60	1,1899	0,8896	0,7913	0,1594
80	1,1925	0,8915	0,7920	0,1597
100	1,1949	0,8932	0,7899	0,1593

Seguindo a mesma lógica adotada na FC, as Figuras 8 e 9 ilustram a sensibilidade das métricas RMSE, MAE, Precision@10 e Recall@10 em relação aos valores de k testados.

A análise de sensibilidade do hiperparâmetro k para o modelo BC, apresentada nas Figuras 8 e 9, evidencia a existência de um ponto ótimo bem definido em torno de $k = 30$. Observa-se uma redução significativa das métricas de erro (RMSE e MAE) até esse valor, seguida por um aumento gradual do erro para valores superiores de k .

Figura 8 – Análise de sensibilidade do hiperparâmetro k para o modelo BC Item–Item considerando métricas de acurácia.

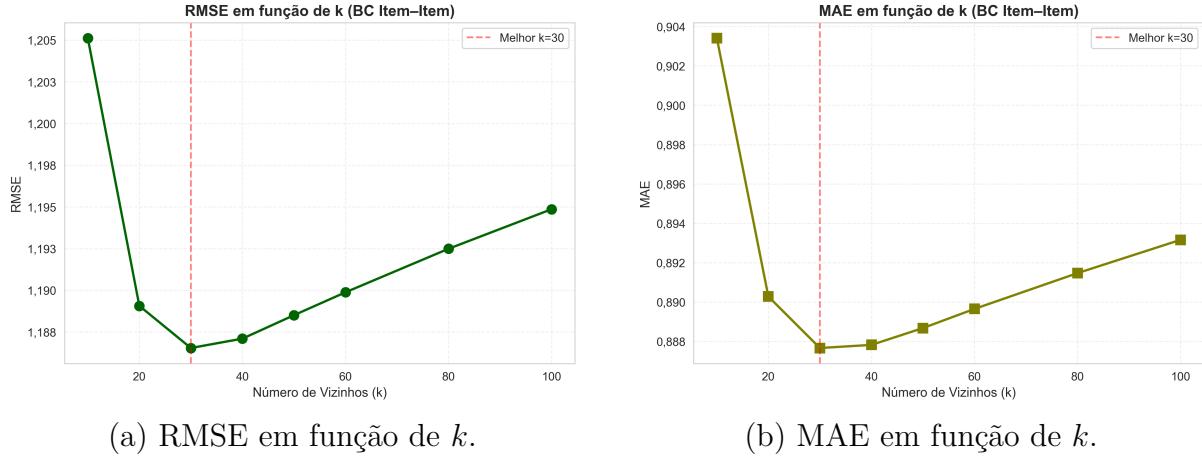
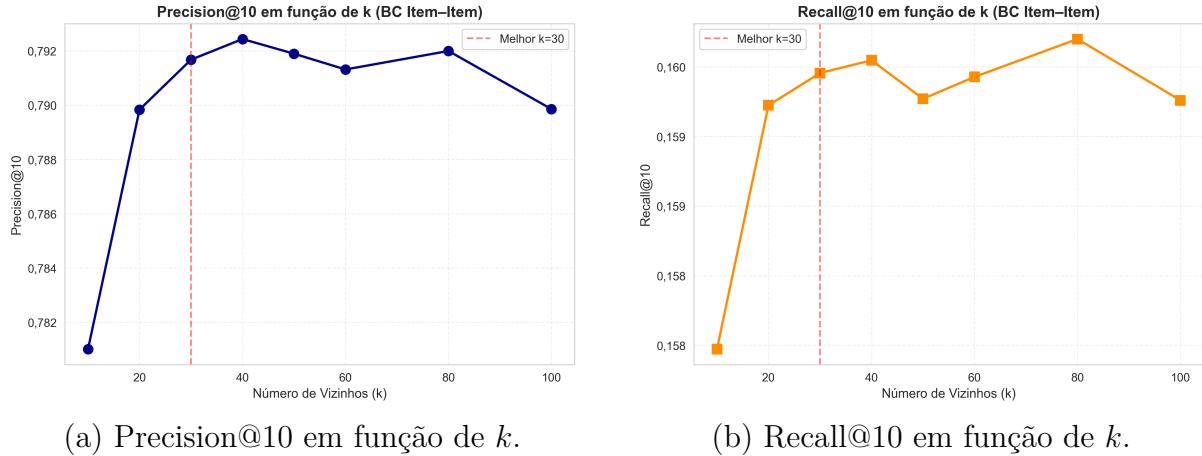


Figura 9 – Análise de sensibilidade do hiperparâmetro k para o modelo BC Item–Item considerando métricas de ranking.



Em relação às métricas de ranking, tanto a Precision@10 quanto o Recall@10 apresentam melhora até aproximadamente $k = 30$, tornando-se instáveis ou apresentando ganhos marginais para valores mais elevados. Esse comportamento sugere que a inclusão de um número excessivo de itens vizinhos tende a incorporar jogos semanticamente menos relacionados, introduzindo ruído no processo de recomendação.

Dessa forma, considerando o desempenho preditivo, a estabilidade das métricas de ranking e o custo computacional, adotou-se $k = 30$ como valor final para o modelo BC, representando um compromisso adequado entre qualidade das recomendações e eficiência computacional.

4.5.3 Treinamento do Modelo Baseline

O Modelo Baseline foi treinado através de uma decomposição sequencial das médias para isolar os vieses globais, de item e de usuário. Os parâmetros aprendidos a partir do conjunto de treino foram:

1. **Média Global (μ)**: O valor base calculado sobre todas as interações de treino foi $\mu \approx 6,93$, servindo como ponto de partida para todas as previsões.
2. **Viés do Item (b_i)**: Para cada um dos 2.000 itens, calculou-se o desvio médio em relação a μ . Itens populares receberam vieses positivos, enquanto itens com recepção mista receberam vieses negativos.
3. **Viés do Usuário (b_u)**: Para cada um dos 5.000 usuários, calculou-se o desvio médio dos resíduos. Esta etapa final captura a tendência individual (indulgência ou rigor) após descontar a qualidade intrínseca dos jogos avaliados.

4.6 Estratégia de Integração: O Modelo Híbrido

A implementação de um Modelo Híbrido não constitui uma escolha arbitrária, mas resulta da análise das limitações observadas nos modelos individuais avaliados até o momento.

A análise exploratória dos dados feitas na Seção 4.2.3 evidenciou características como a forte assimetria na distribuição das avaliações, a presença do fenômeno de cauda longa e a concentração das notas em torno de valores elevados. Esses aspectos podem impactar o desempenho dos modelos, especialmente em cenários de esparsidade e *cold-start*.

A Tabela 3 apresenta um comparativo de desempenho entre os modelos individuais implementados, evidenciando que cada abordagem possui pontos fortes e limitações específicas.

Tabela 3 – Comparativo de desempenho dos modelos avaliados

Modelo	RMSE	MAE	P@10	R@10
FC User–User (Pearson)	1,0895	0,8159	0,8709	0,1781
Baseline ($\mu + b_u + b_i$)	1,1141	0,8352	0,8489	0,1721
BC Item–Item (Jaccard)	1,1865	0,8877	0,7917	0,1595

A partir desses resultados, é possível destacar as seguintes observações:

- A **Filtragem Colaborativa (FC) usuário–usuário** apresentou desempenho competitivo em termos de RMSE quando comparada aos demais modelos, indicando boa capacidade preditiva quando há histórico suficiente de interações. Entretanto, por

depender exclusivamente de avaliações passadas, o modelo torna-se inadequado em contextos com dados históricos insuficientes, nos quais novos usuários não dispõem de informação suficiente para o cálculo de similaridades confiáveis.

- **O Modelo Baseline**, fundamentado na média global e nos vieses de usuários e itens, este modelo demonstrou desempenho robusto em termos de erro médio, uma vez que se beneficia da concentração das notas em torno da média global. Ainda assim, sua capacidade de personalização é limitada, uma vez que, nesta implementação, não incorpora informações adicionais sobre as características dos itens ou preferências específicas dos usuários.
- **O Modelo Baseado em Conteúdo (BC)** apresentou desempenho inferior ao Baseline em termos de RMSE, porém mostrou-se particularmente eficaz em cenários de *cold-start*, uma vez que sua capacidade preditiva independe diretamente do volume de avaliações históricas. A análise de sensibilidade do hiperparâmetro k indicou a existência de um ponto ótimo bem definido, reforçando a necessidade de limitar o número de itens vizinhos para evitar a introdução de ruído.

Dessa forma, embora a Filtragem Colaborativa tenha apresentado o menor erro médio em determinados cenários, sua adoção isolada não é suficiente para garantir a robustez do sistema de recomendação como um todo. A ausência de mecanismos para lidar com *cold-start* e a sensibilidade à esparsidade dos dados limitam sua aplicabilidade em contextos reais.

Nesse contexto, a combinação dos modelos em uma abordagem híbrida surge como uma estratégia que permite conciliar desempenho preditivo, capacidade de personalização e cobertura do sistema. Ao integrar diferentes fontes de informação, o modelo híbrido permite mitigar as limitações individuais de cada abordagem, tornando o sistema mais robusto e adequado a cenários reais de recomendação.

A fim de potencializar as forças de cada abordagem e reduzir seus impactos negativos, optou-se por implementar dois modelos híbridos distintos: Filtragem Colaborativa combinada ao modelo Baseline (FC + BL) e Filtragem Colaborativa combinada ao modelo Baseado em Conteúdo (FC + BC).

As principais características de cada abordagem híbrida são:

- **FC + BL:** combina a capacidade de personalização da Filtragem Colaborativa com a estabilidade numérica do modelo Baseline, atuando como um mecanismo de suavização das previsões e reduzindo a variância dos erros, especialmente em cenários com dados ruidosos.

- **FC + BC:** integra a Filtragem colaborativa com informações baseadas em conteúdo, ampliando a cobertura do sistema e mitigando o problema de *cold-start*, uma vez que a predição não depende exclusivamente de interações históricas.

4.6.1 Avaliação dos Modelos Híbridos

Para um ajuste eficiente do modelo híbrido, foi conduzida uma análise de sensibilidade do coeficiente de hibridização β , que controla a contribuição relativa de cada componente no cálculo da nota prevista. Foram testados valores de β no intervalo $[0, 1]$ com incrementos de 0,1, onde:

A Tabela 5 com os resultados completos da análise de sensibilidade do coeficiente de hibridização β encontram-se apresentados no Apêndice A deste trabalho. As Figuras 10 e 11 ilustram o comportamento dos modelos híbridos em termos de métricas de acurácia e de ranking, respectivamente.

A análise das métricas de acurácia indica comportamentos distintos entre os dois modelos híbridos. No caso do modelo FC + BL, valores baixos de β resultam em erros elevados, refletindo a influência predominante do modelo Baseado em Conteúdo. À medida que o peso da Filtragem Colaborativa aumenta, observa-se uma redução rápida do erro, evidenciando que a componente colaborativa corrige de forma eficiente as imprecisões iniciais do modelo híbrido. Para valores elevados de β , em especial a partir de $\beta = 0,8$, nota-se novamente o aumento do erro, sugerindo que a contribuição exclusiva da Filtragem Colaborativa não é suficiente para manter ganhos adicionais de acurácia. Esse comportamento indica que a combinação entre FC e BC é mais eficaz do que a utilização isolada de qualquer uma das abordagens.

Já no modelo FC + BL, a variação do erro ao longo dos diferentes valores de β é mais suave. Observa-se uma estabilização das métricas de erro à medida que o peso da Filtragem Colaborativa aumenta, indicando que o Baseline atua principalmente como um mecanismo de suavização das previsões, sem contribuir de forma significativa para melhorias adicionais de acurácia. Nesse cenário, os resultados sugerem que a inclusão do Baseline no modelo híbrido oferece ganhos limitados em relação ao uso exclusivo da Filtragem Colaborativa.

Quando comparados os dois modelos híbridos, nota-se que para $\beta = 0,6$ ambos apresentam desempenhos bastante semelhantes com relação aos erros, o que reduz a vantagem prática de um modelo em relação ao outro para esse valor específico do coeficiente de hibridização.

A análise de sensibilidade do coeficiente de hibridização β revelou que valores intermediários proporcionam o melhor compromisso entre desempenho preditivo e qualidade do ranking. Em particular, adotou-se $\beta = 0,7$ que, apesar de não apresentar o menor erro

médio absoluto, oferece um RMSE e MAE competitivos aliados a melhorias significativas nas métricas de ranking além de não retirar completamente a contribuição do modelo Baseado em Conteúdo.

Ainda considerando as métricas de ranking, o modelo FC + BC apresenta desempenho superior com resultados mais consistentes para valores β no intervalo $0,5 \leq \beta \leq 0,7$. A partir desse ponto, observa-se uma tendência de queda nas métricas, indicando que a redução do peso do modelo Baseado em Conteúdo compromete a qualidade do ranqueamento.

Figura 10 – Comparação entre os modelos híbridos FC+Baseline e FC+BC em termos de métricas de acurácia para diferentes valores do coeficiente de hibridização β .

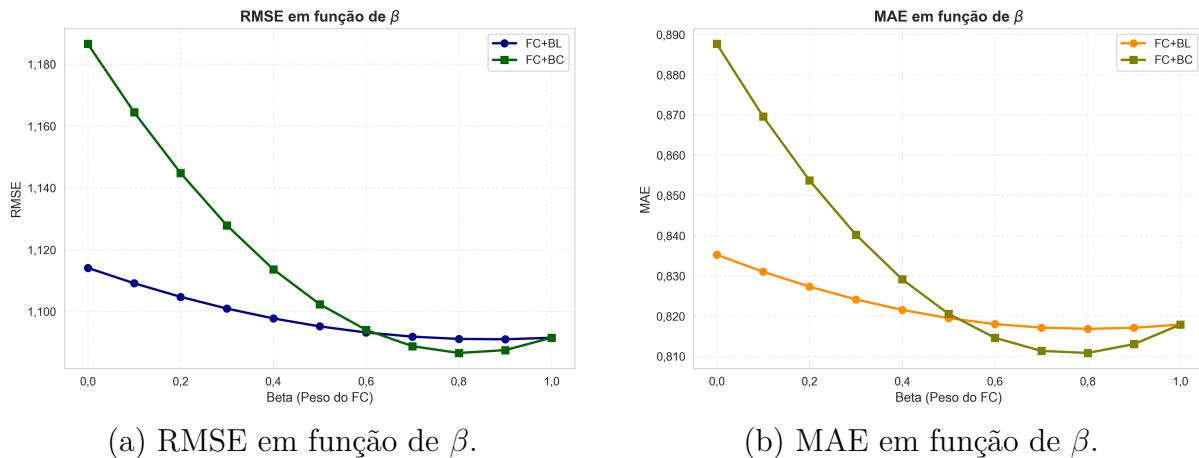
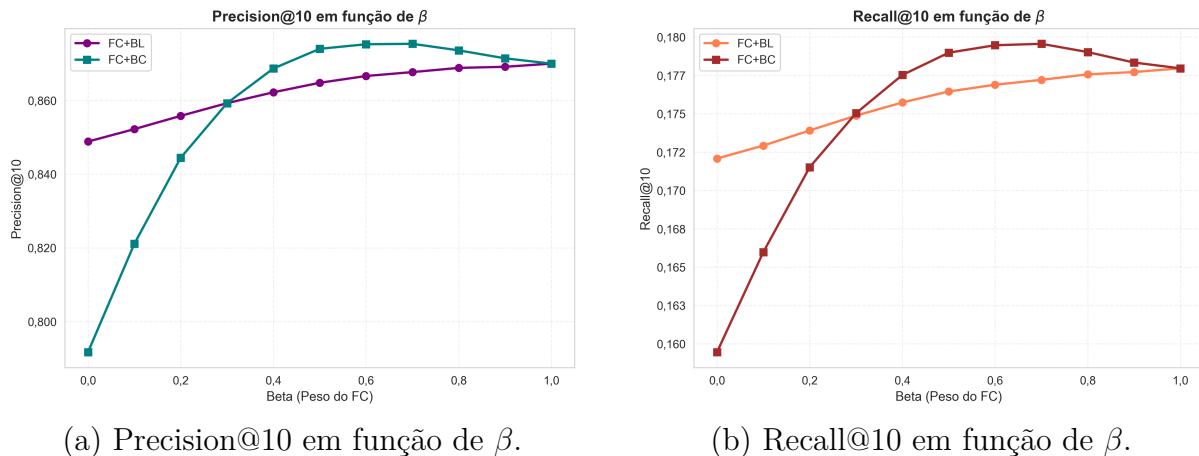


Figura 11 – Comparação entre os modelos híbridos FC+Baseline e FC+BC em termos de métricas de ranking para diferentes valores do coeficiente de hibridização β .



As Figuras 12 e 13 sintetizam o desempenho dos modelos avaliados, destacando a superioridade do modelo híbrido ponderado por $\beta = 0,7$ em termos de equilíbrio entre precisão preditiva e qualidade do ranqueamento.

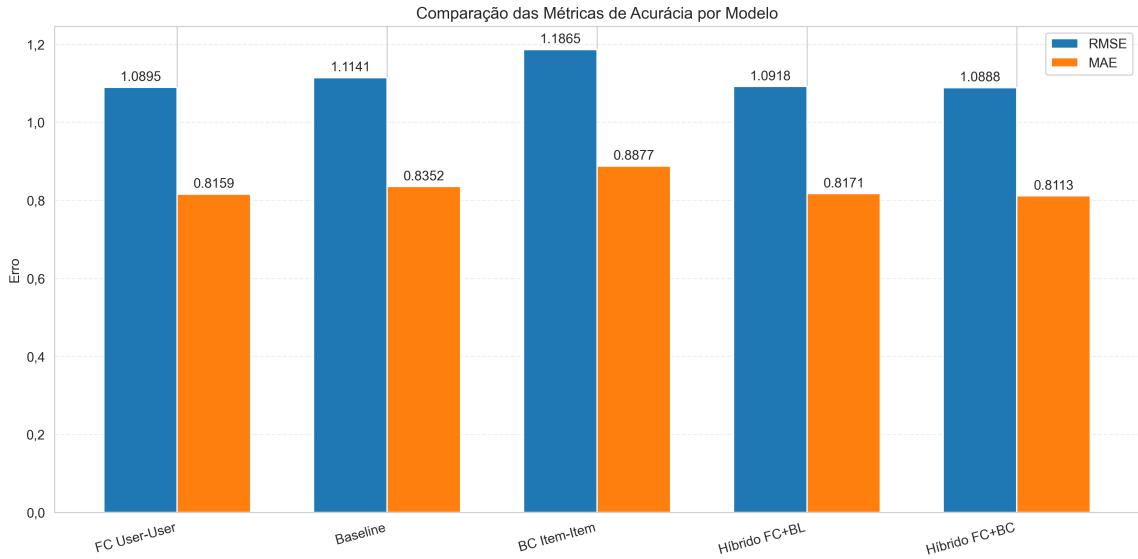


Figura 12 – Comparativo das métricas de acurácia dos modelos avaliados.

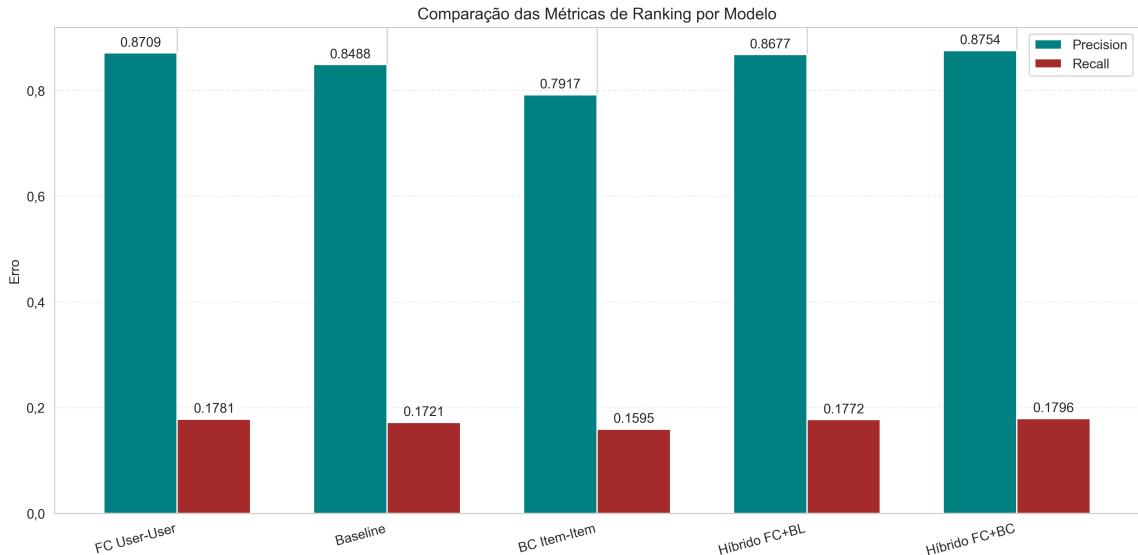


Figura 13 – Comparativo das métricas de ranking dos modelos avaliados.

4.7 Modelo em Produção

A fim de validar a aplicabilidade do modelo proposto em um cenário real, realizou-se um experimento adicional de como o sistema de recomendação híbrido funciona ao tentar recomendar jogos para um usuário.

Foi filtrado da base de usuário que não participou do treinamento e que tinha pelo menos 20 avaliações feitas. Das notas dadas por esse usuário foram ocultadas 5 avaliações aleatórias para que o sistema pudesse tentar recomendá-las. Com base nas avaliações restantes, o sistema utilizou o modelo híbrido para prever as notas dos jogos que ele “ainda não havia avaliado”. A seguir, são apresentadas as recomendações geradas pelo modelo, ordenadas pela nota prevista e os valores reais (ocultos) para comparação:

4.7.1 Validação do Modelo com Jogos Não Avaliados pelo Usuário

Para avaliar a capacidade de generalização do modelo de recomendação, realizou-se um experimento adicional: prever a nota de jogos que não estavam presentes no histórico de avaliações de um usuário de teste (denominado aqui como "**monots**"). O modelo utilizou suas 310 avaliações conhecidas para estimar as notas de cinco jogos aleatoriamente selecionados considerando apenas aqueles presentes no conjunto de treinamento, permitindo medir o erro absoluto individual em cada predição. A aleatoriedade do processo foi controlada por meio da fixação de uma semente, assegurando a reproduzibilidade dos resultados.

Tabela 4 – Validação do Modelo: Previsões para Jogos Não Avaliados

Jogo	Nota Prevista	Nota Real	Erro Absoluto
Deep Sea Adventure	7,3050	6,5000	0,8050
BANG!	7,0340	8,0000	0,9660
La Granja	7,2450	8,0000	0,7550
Stone Age	7,2650	8,0000	0,7350
Dungeon Lords	7,0500	8,7000	1,6500

Os resultados mostram que o erro absoluto variou entre aproximadamente 0,7353 e 1,6500 pontos em uma escala de 0 a 10. Valores próximos de 1 ponto de erro são compatíveis com o RMSE global do modelo, indicando consistência entre desempenho agregado e desempenho em itens individuais. O maior erro ocorreu em *Dungeon Lords*, sugerindo que jogos com características semânticas ou mecânicas não suficientemente representadas no perfil do usuário podem gerar maior incerteza preditiva. Ainda assim, a maior parte das previsões permaneceu dentro de margens aceitáveis.

Conforme discutido na Seção 4.2.3, vale ressaltar que o modelo produz previsões de notas contínuas, enquanto as avaliações reais de alguns usuários tendem a ser discretas (com ou sem incrementos de 0,5). Isso implica que pequenas discrepâncias podem ser atribuídas à natureza intrínseca do sistema de avaliação, onde diferenças mínimas na percepção do usuário podem resultar em variações na nota atribuída.

5 CONCLUSÃO E TRABALHOS FUTUROS

Com base na análise multidimensional (Erro, Ranking e eficiência computacional), conclui-se que o **Modelo Híbrido Ponderado FC+BC** com fator de hibridização $\beta = 0,7$ é a solução definitiva para o problema proposto neste trabalho. Sua superioridade justifica-se por três pilares:

1. **Equilíbrio entre desempenho preditivo e generalização:** a combinação entre Filtragem Colaborativa e informações Baseadas em Conteúdo permite alcançar níveis de RMSE e MAE competitivos, sem depender exclusivamente de interações históricas, o que resulta em maior robustez frente à esparsidão dos dados.
2. **Mitigação do problema de *cold-start*:** ao preservar a contribuição do modelo Baseado em Conteúdo, o híbrido FC+BC mantém capacidade preditiva para itens com poucas avaliações, ampliando a cobertura do sistema e tornando-o mais adequado a cenários reais de aplicação.
3. **Melhoria na qualidade do ranking:** a integração de padrões colaborativos com similaridade semântica entre itens produz listas de recomendação mais relevantes e bem ordenadas, refletidas em ganhos nas métricas de Precision@10 e Recall@10 quando comparado aos modelos individuais e ao híbrido FC+BL.

Este trabalho pode ser ampliado em várias direções para fortalecer tanto a metodologia quanto a eficiência prática do sistema de recomendação. Embora o Baseline tenha sido aqui utilizado como uma abordagem heurística simples por conta de limitações computacionais, Aggarwal et al. (2016) destaca que ele pode ser aprimorado por modelos mais robustos. Uma evolução natural consiste em introduzir regularização aos vieses de usuários e itens, substituindo o cálculo por médias simples pela solução de um problema de mínimos quadrados regularizados. Essa reformulação reduz sobreajuste e melhora a estabilidade do modelo, especialmente para usuários e itens com poucas avaliações.

A partir desse estágio, o sistema pode avançar para um modelo de Fatoração de Matrizes, incorporando fatores latentes capazes de capturar padrões mais complexos de interação. Nesse caso, a predição deixa de depender apenas dos vieses e passa a incluir o produto escalar entre vetores latentes de usuários e itens, resultando em ganhos de desempenho já consolidados na literatura. Por fim, o modelo híbrido empregado neste trabalho pode ser fortalecido ao substituir o Baseline heurístico por essa versão regularizada ou pelo SVD completo, combinando o ganho estatístico da fatoração com a capacidade temática do componente Baseado em Conteúdo.

Do ponto de vista das bases dados usadas, há informações adicionais que podem ser exploradas para enriquecer o perfil dos jogos. A inclusão de atributos como número de jogadores, tempo de jogo, faixa etária recomendada e complexidade pode fornecer dimensões extras para a similaridade baseada em conteúdo.

REFERÊNCIAS

- ABRINQ. **Anuário Estatístico de Brinquedos 2025**. 2025. Acesso em: 15 set. 2025. Disponível em: <<http://www.abrinq.com.br/wp-content/uploads/2025/03/Anu%C3%A1rio-2025.pdf>>.
- ADOMAVICIUS, G.; TUZHILIN, A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. **IEEE transactions on knowledge and data engineering**, IEEE, v. 17, n. 6, p. 734–749, 2005.
- AGGARWAL, C. C. et al. **Recommender systems**. [S.l.]: Springer, 2016. v. 1.
- AZAMBUJA, R. X. d.; MORAIS, A. J.; FILIPE, V. Teoria e prática em sistemas de recomendação. **Revista de Ciências da Computação**, Universidade Aberta, p. 23–46, 2021.
- GOLDBERG, D. et al. Using collaborative filtering to weave an information tapestry. **Communications of the ACM**, ACM New York, NY, USA, v. 35, n. 12, p. 61–70, 1992.
- Grand View Research. **Playing Cards and Board Games Market Size, Share and Trends Analysis Report**. 2025. Acesso em: 15 set. 2025. Disponível em: <<https://www.grandviewresearch.com/industry-analysis/playing-cards-board-games-market>>.
- Kaggle. **Kaggle: Your Machine Learning and Data Science Community**. 2025. <<https://www.kaggle.com>>. Acesso em 15 jul. 2025.
- KARLGREN, J. **A Recommendation Algebra**. [S.l.], 1994. Tech Report supported by the IBM Electrum Scholarship 1992. Disponível em: <<mailto:jussi@sics.se>>.
- LUDOPEDIA. **Jogos modernos: o que são e de onde vieram**. 2023. Acesso em: 14 set. 2025. Disponível em: <<https://ludopedia.com.br/topicos/25361/jogos-modernos-o-que-sao-e-de-onde-vieram>>.
- PAPERGAMES. **Principais características de um jogo de tabuleiro moderno**. 2021. Acesso em: 14 set. 2025. Disponível em: <<https://papergames.com.br/principais-caracteristicas-jogos-de-tabuleiro-moderno/#:~:text=Os%20jogos%20de%20tabuleiro%20modernos,originais%20e%20data%20de%20cria%C3%A7%C3%A3o>>.
- _____. **O que é boardgame?** 2022. Acesso em: 14 set. 2025. Disponível em: <<https://papergames.com.br/o-que-e-boardgame/>>.
- RAHMATILANI, C.; PATRIANSAH, M. Board game sebagai media komunikasi visual pendekatan hubungan harmonis keluarga antar orang tua dan anak. **Besaung: Jurnal Seni Desain dan Budaya**, v. 9, n. 1, p. 116–128, 2024.
- RICCI, F.; ROKACH, L.; SHAPIRA, B. Introduction to recommender systems handbook. In: **Recommender systems handbook**. [S.l.]: Springer, 2010. p. 1–35.
- SHAIKH, S. Exploring the role of board games on cultivating entrepreneurial mindsets in youth. **Sports Sciences and Physical Education Review**, v. 2, n. 1, p. LXXI–LXXXI, 2023.

Statista. **Board Games - Brazil**. 2025. Acesso em: 15 set. 2025. Disponível em: <<https://www.statista.com/outlook/cmo/toys-games/board-games/brazil>>.

threnjen. **Board Games Database from BoardGameGeek**. 2025. Kaggle dataset. Acesso em 13 de julho de 2025. Disponível em: <<https://www.kaggle.com/datasets/threnjen/board-games-database-from-boardgamegeek>>.

Wikipedia. **Board game — Wikipedia, The Free Encyclopedia**. 2025. <https://en.wikipedia.org/wiki/Board_game>. [Online; accessed 16-September-2025].

Apêndices

APÊNDICE A – MÉTRICAS DE DESEMPENHO DOS MODELOS HÍBRIDOS

Tabela 5 – Métricas de desempenho dos modelos híbridos FC + BL e FC + BC para diferentes valores de β .

β	Modelo	RMSE	MAE	P@10	R@10
0.0	FC+BL	1,1141	0,8352	0,8488	0,1721
	FC+BC	1,1865	0,8877	0,7917	0,1595
0.1	FC+BL	1,1091	0,8310	0,8522	0,1729
	FC+BC	1,1644	0,8696	0,8211	0,1660
0.2	FC+BL	1,1047	0,8273	0,8558	0,1739
	FC+BC	1,1448	0,8537	0,8444	0,1715
0.3	FC+BL	1,1009	0,8241	0,8593	0,1749
	FC+BC	1,1278	0,8402	0,8593	0,1750
0.4	FC+BL	1,0977	0,8215	0,8622	0,1757
	FC+BC	1,1136	0,8291	0,8687	0,1775
0.5	FC+BL	1,0952	0,8195	0,8648	0,1764
	FC+BC	1,1023	0,8205	0,8740	0,1790
0.6	FC+BL	1,0932	0,8180	0,8667	0,1769
	FC+BC	1,0940	0,8146	0,8753	0,1795
0.7	FC+BL	1,0918	0,8171	0,8677	0,1772
	FC+BC	1,0888	0,8113	0,8754	0,1796
0.8	FC+BL	1,0911	0,8168	0,8689	0,1776
	FC+BC	1,0866	0,8108	0,8736	0,1790
0.9	FC+BL	1,0910	0,8171	0,8692	0,1777
	FC+BC	1,0875	0,8130	0,8715	0,1783
1.0	FC+BL	1,0915	0,8179	0,8700	0,1780
	FC+BC	1,0915	0,8179	0,8700	0,1780