

# **MBA em Ciências de Dados**

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

**Autor: Isabelly da Silva Almeida**

**Orientador: Adriano Kamimura Suzuki**



# Introdução

# Contexto e Motivação

- ▶ Crescimento dos jogos de tabuleiro modernos nas últimas décadas
- ▶ Ampliação do público e diversidade de perfis de jogadores
- ▶ Contribuições educacionais, sociais e cognitivas
- ▶ Excesso de opções dificulta a escolha de novos jogos

# Problema e Objetivos

## Problema

Como auxiliar usuários na descoberta de jogos de tabuleiro alinhados às suas preferências em um cenário de grande volume de opções?

## Objetivo do Trabalho

Avaliar e comparar diferentes abordagens de sistemas de recomendação aplicadas ao contexto de jogos de tabuleiro modernos, utilizando dados reais da plataforma *BoardGameGeek*.

# Metodologia

# O que são Sistemas de Recomendação

Sistemas de recomendação são ferramentas computacionais projetadas para auxiliar usuários na descoberta de itens de interesse em ambientes com grande volume de opções.

- ▶ Utilizados em domínios como filmes, músicas, produtos e jogos
- ▶ Baseiam-se em dados históricos de interação dos usuários
- ▶ Buscam reduzir a sobrecarga de escolha

# Objetivos dos Sistemas de Recomendação

Os principais objetivos dos sistemas de recomendação incluem:

- ▶ Sugerir itens relevantes de acordo com o perfil do usuário
- ▶ Melhorar a experiência e satisfação do usuário
- ▶ Aumentar o engajamento com a plataforma
- ▶ Auxiliar na descoberta de novos itens

# Tipos e Abordagens de Sistemas de Recomendação

Diferentes abordagens podem ser adotadas na construção de sistemas de recomendação, cada uma com características específicas.

- ▶ Filtragem Colaborativa
- ▶ Sistemas Baseados em Conteúdo
- ▶ Modelos Baseline
- ▶ Sistemas Híbridos



# Filtragem Colaborativa

A filtragem colaborativa explora padrões de avaliação entre usuários e itens, assumindo que usuários com comportamentos semelhantes tendem a compartilhar preferências.

O problema é modelado por uma matriz de avaliações  $R \in \mathbb{R}^{|U| \times |I|}$ , onde  $r_{ui}$  representa a avaliação do usuário  $u$  para o item  $i$ .

A matriz  $R$  é tipicamente esparsa.

# Filtragem Colaborativa: Predição e Top-k

A predição da avaliação é realizada a partir dos  $k$  vizinhos mais similares (Top- $k$ ):

$$\hat{r}_{ui} = \bar{r}_u + \frac{\sum_{v \in N_k(u)} \text{sim}(u, v) (r_{vi} - \bar{r}_v)}{\sum_{v \in N_k(u)} |\text{sim}(u, v)|}$$

## Vantagens

- ▶ Boa capacidade de capturar padrões coletivos

## Desvantagens

- ▶ Sensível à esparsidade e ao problema de *cold-start*

# Sistemas Baseados em Conteúdo

Os sistemas baseados em conteúdo recomendam itens semelhantes àqueles previamente avaliados pelo usuário, utilizando características descritivas dos itens.

A predição é obtida considerando os  $k$  itens mais similares:

$$\hat{r}_{ui} = \frac{\sum_{j \in N_k(i)} \text{sim}(i, j) r_{uj}}{\sum_{j \in N_k(i)} |\text{sim}(i, j)|}$$

## Vantagens

- ▶ Menor impacto do *cold-start* de usuários

## Desvantagens

- ▶ Dependência da qualidade dos atributos dos itens

# Modelo Baseline

O modelo baseline fornece uma estimativa simples das avaliações, considerando efeitos globais e individuais:

$$\hat{r}_{ui} = \mu + b_u + b_i$$

## Objetivo

- ▶ Servir como referência mínima de desempenho

## Vantagens

- ▶ Simples e computacionalmente eficiente

## Desvantagens

- ▶ Não captura interações complexas entre usuários e itens

# Modelo Híbrido

Os sistemas híbridos combinam diferentes abordagens de recomendação, buscando explorar suas vantagens complementares.

$$\hat{r}_{ui}^{(h)} = \alpha \hat{r}_{ui}^{(FC)} + (1 - \alpha) \hat{r}_{ui}^{(BC)}$$

- ▶ Integra filtragem colaborativa e conteúdo
- ▶ Reduz limitações individuais dos modelos

# Medidas de Similaridade

As medidas de similaridade variam conforme o tipo de sistema de recomendação.

- ▶ **Correlação de Pearson:** adequada para filtragem colaborativa
- ▶ **Similaridade do Cosseno:** utilizada em vetores de atributos
- ▶ **Distância Euclidiana:** sensível à magnitude das avaliações
- ▶ **Coeficiente de Jaccard:** apropriado para dados binários

# Métricas de Avaliação

As métricas de avaliação foram divididas em dois grupos:

## Métricas de Acurácia

$$\text{MAE} = \frac{1}{N} \sum |r_{ui} - \hat{r}_{ui}|$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum (r_{ui} - \hat{r}_{ui})^2}$$

## Métricas de Ranqueamento

$$\text{Precision@k} = \frac{|\text{itens relevantes} \cap \text{top-}k|}{k}$$

$$\text{Recall@k} = \frac{|\text{itens relevantes} \cap \text{top-}k|}{|\text{itens relevantes}|}$$

Dados



# Base de Dados

- ▶ Fonte: plataforma *BoardGameGeek*
- ▶ Base amplamente utilizada em estudos sobre jogos de tabuleiro
- ▶ Dados compostos por:
  - ▶ Avaliações de usuários para jogos
  - ▶ Informações descritivas dos jogos (categorias e mecânicas)
- ▶ Estrutura típica de recomendação: matriz usuário-jogo

# Resultados



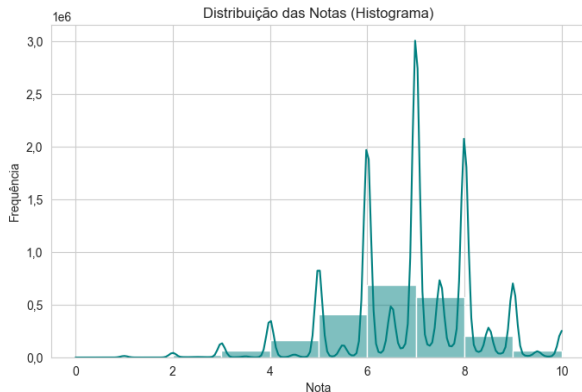
# Implementação e Análise Prática

Este capítulo apresenta os resultados obtidos a partir da implementação dos modelos de recomendação descritos na metodologia.

- ▶ Análise exploratória e pré-processamento dos dados
- ▶ Definição da matriz de avaliações
- ▶ Treinamento, parametrização e avaliação dos modelos

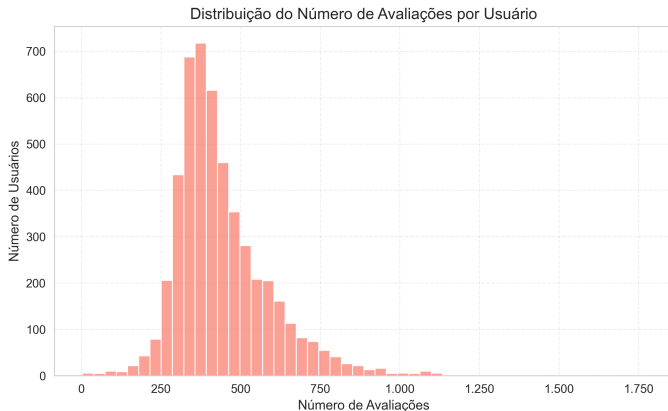
# Distribuição das Notas

A distribuição das avaliações permite compreender o comportamento geral das notas atribuídas pelos usuários.



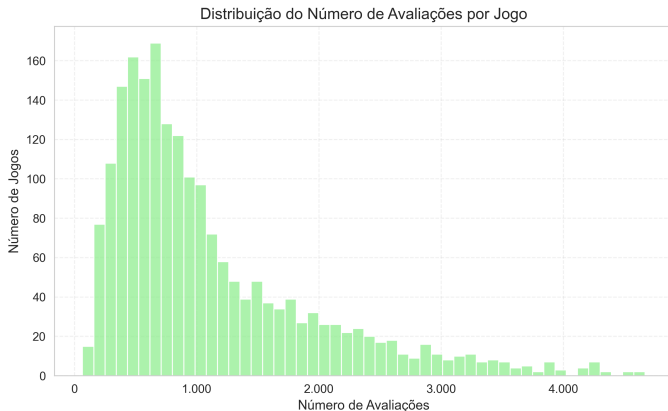
# Distribuição de Avaliações por Usuário

A maioria dos usuários avalia apenas um número reduzido de jogos, o que contribui para a esparsidade da base.



# Distribuição de Avaliações por Jogo

Observa-se concentração de avaliações em um subconjunto reduzido de jogos, indicando assimetria de popularidade.



# Matriz de Avaliações $R$

Após o pré-processamento, os dados são organizados em uma matriz usuário-jogo  $R$ , utilizada como base para os modelos de recomendação.

- ▶ Linhas representam usuários
- ▶ Colunas representam jogos
- ▶ Entradas correspondem às avaliações observadas

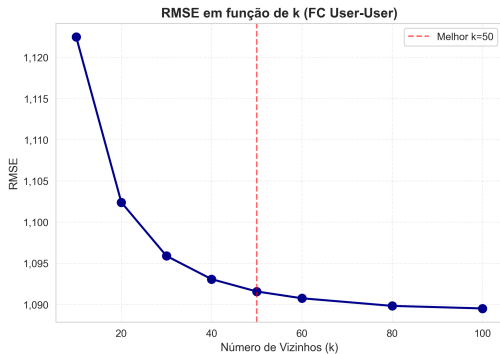
# Divisão dos Dados em Treino e Teste

Os experimentos foram conduzidos por meio de uma divisão dos dados em conjuntos de treinamento e teste.

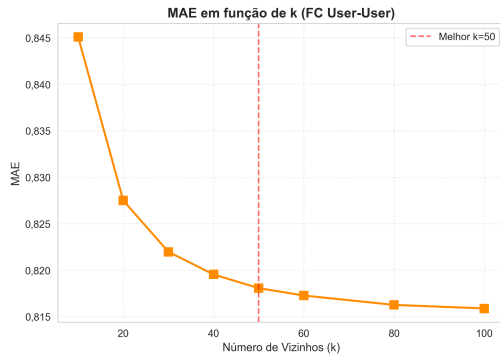
- ▶ Avaliação realizada em cenário offline
- ▶ Garantia de comparabilidade entre os modelos



# Filtragem Colaborativa — Sensibilidade ao Top-k (Acurácia)

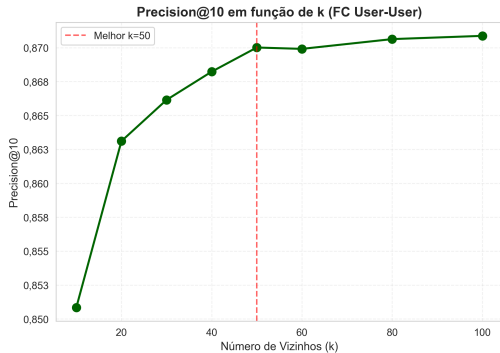


RMSE

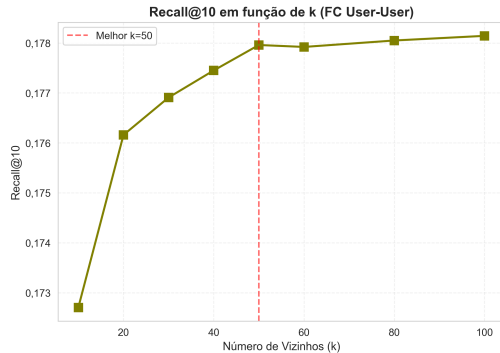


MAE

# Filtragem Colaborativa — Sensibilidade ao Top-k (Ranking)

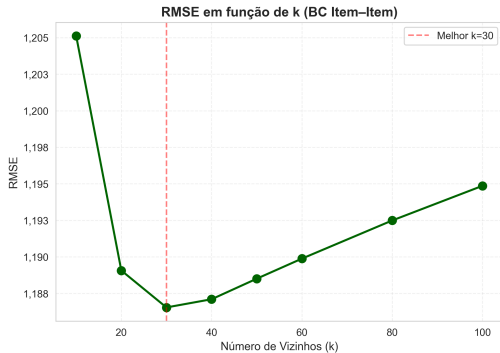


Precision@k

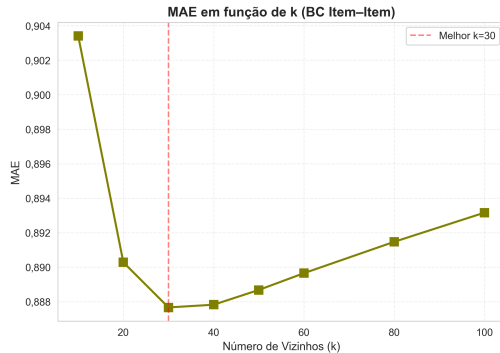


Recall@k

# Baseado em Conteúdo — Sensibilidade ao Top-k (Acurácia)

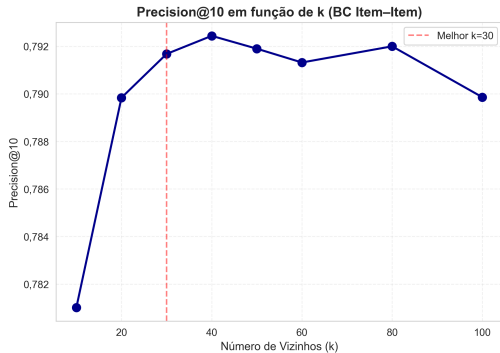


RMSE

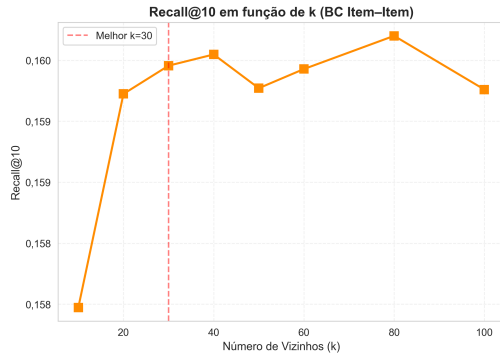


MAE

# Baseado em Conteúdo — Sensibilidade ao Top-k (Ranking)



Precision@k



Recall@k

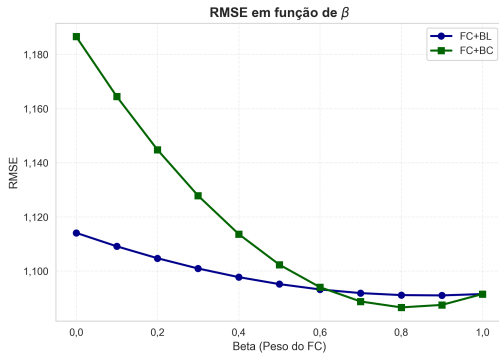
# Modelo Baseline — Resultados

O modelo baseline é utilizado como referência mínima de desempenho, permitindo avaliar os ganhos obtidos pelos modelos mais complexos.

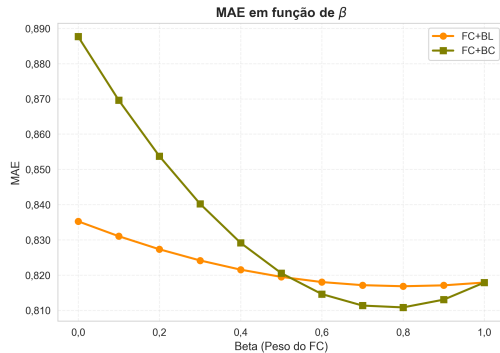
Modelo	MAE	RMSE	Precision@k	Recall@k
Baseline	XX.XXX	XX.XXX	XX.XXX	XX.XXX

- ▶ Não depende de vizinhança nem de parâmetros
- ▶ Serve como linha de base para comparação

# Modelo Híbrido — Sensibilidade ao Parâmetro $\beta$ (Acurácia)

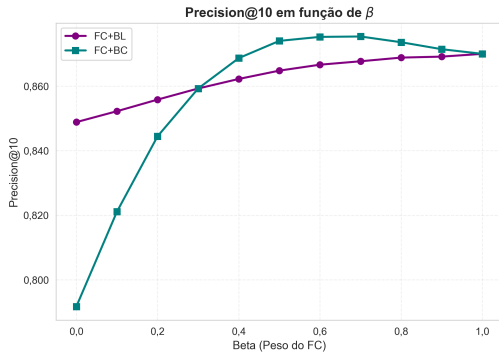


RMSE

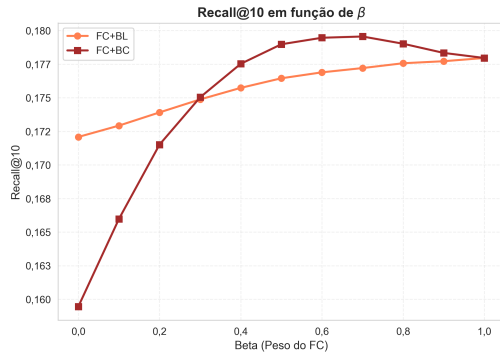


MAE

# Modelo Híbrido — Sensibilidade ao Parâmetro $\beta$ (Ranking)



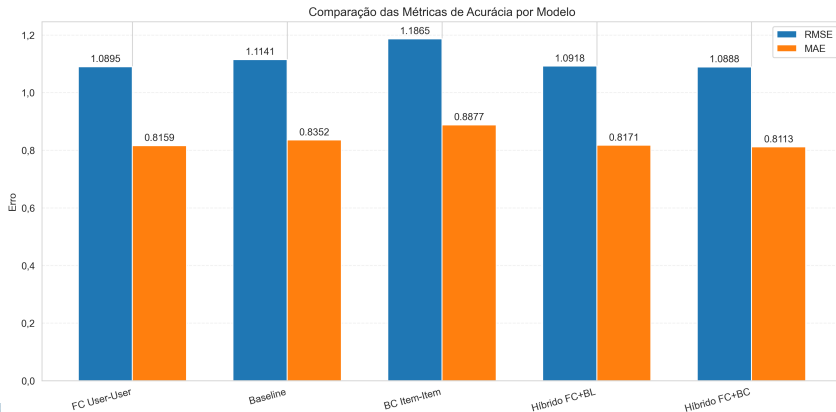
Precision@k



Recall@k

# Comparação Final — Acurácia

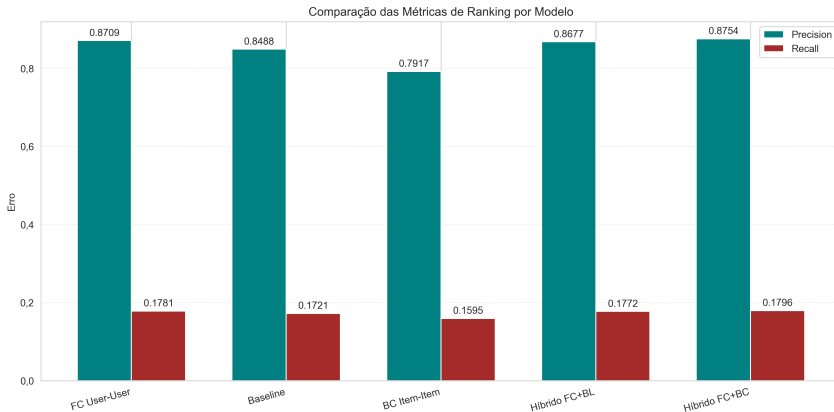
A comparação entre os modelos evidencia diferenças no desempenho preditivo segundo as métricas de acurácia.





# Comparação Final — Ranqueamento

As métricas de ranqueamento permitem avaliar a qualidade das listas de recomendação geradas pelos modelos.



# Conclusão



# Conclusões

- ▶ O trabalho avaliou diferentes abordagens de sistemas de recomendação
- ▶ Modelos híbridos apresentaram melhor desempenho global
- ▶ A integração de previsões reduziu limitações individuais dos modelos
- ▶ Os resultados confirmam a viabilidade da aplicação no contexto de jogos de tabuleiro

# Limitações e Trabalhos Futuros

- ▶ Avaliação restrita a dados históricos (offline)
- ▶ Ausência de validação com usuários reais
- ▶ Possibilidade de inclusão de dados temporais
- ▶ Exploração de modelos baseados em aprendizado profundo

Obrigada!