

Recomendação Jurídica Interativa sob Alta Esparsidade: Uma Avaliação Empírica com Algoritmos Bandit

Mateus Macorini Alves
Universidade Estadual de Maringá
(UEM)
Maringá, Paraná Brazil
pg405271@uem.br

Matheus Leme
Universidade Estadual de Maringá
(UEM)
Maringá, Paraná Brazil
pg55308@uem.br

Marcos Aurelio Domingues
Universidade Estadual de Maringá
(UEM)
Maringá, Paraná Brazil
madomingues@uem.br

ABSTRACT

Recommender systems based on Multi-Armed Bandit (MAB) algorithms have been widely investigated in domains such as entertainment and e-commerce, where user-item interactions are relatively recurrent and collaborative signals are more abundant. However, their applicability to the legal domain remains underexplored. This work empirically evaluates the feasibility of bandit-based interactive recommendation for legal documents, a scenario characterized by high sparsity, large item catalogs, and scarce positive feedback. Using the iRec framework, we conducted offline experiments on a real legal interaction dataset, named JurisLog, comparing classical algorithms such as ϵ -Greedy, Upper Confidence Bound (UCB), and Thompson Sampling against a non-personalized popularity baseline. To isolate the effect of sparsity and item recurrence, we constructed three versions of the dataset under different density constraints and compared them with reference datasets commonly used in the literature. The results showed that the legal domain imposes substantial structural challenges, with all evaluated algorithms achieving considerably lower performance than in traditional recommendation domains. Moreover, the global popularity baseline consistently outperformed the bandit-based approaches in the legal datasets, suggesting that sparse collaborative feedback limits the effectiveness of direct interactive learning. These findings indicate that traditional Multi-Armed Bandit methods are not directly sufficient for legal recommendation scenarios and highlight the need for hybrid approaches that incorporate semantic representations and domain-specific information.

KEYWORDS

Recommender Systems, Legal Recommendation, Bandit Algorithms, High Sparsity

1 INTRODUÇÃO

O crescente volume de conteúdos digitais consolidou a importância dos Sistemas de Recomendação (SR) como ferramentas essenciais para mitigar a sobrecarga informacional, beneficiando tanto consumidores quanto fornecedores [2, 3]. Com a evolução da área, abordagens baseadas em aprendizado por reforço, especificamente algoritmos de Multi-Armed Bandits (MAB), têm ganhado destaque em relação aos métodos tradicionais. A eficácia dos algoritmos MAB reside na sua capacidade de lidar com o problema de tomada de

decisão sequencial sob incerteza. Eles adaptam-se dinamicamente às preferências dos usuários ao equilibrar a recomendação de itens conhecidos com a investigação de itens potencialmente relevantes no longo prazo [8, 11].

Apesar dos avanços e do sucesso consolidado dos SR em domínios como o comércio eletrônico e o entretenimento, sua aplicação no setor jurídico ainda é incipiente. Esse domínio impõe desafios intrínsecos severos, como a alta complexidade semântica dos textos, o rigor técnico do vocabulário e a forte dependência de interpretação contextual. Além disso, os dados jurídicos tendem a ser escassos, desbalanceados e sensíveis, limitando a eficácia das abordagens tradicionais de recomendação e exigindo estratégias que aprendam de forma adaptativa e incremental.

O domínio jurídico apresenta características substancialmente distintas das utilizadas na literatura sobre sistemas de recomendação. Diferentemente de bases de dados como MovieLens¹ ou Yahoo Music², os documentos jurídicos formam um catálogo extremamente fragmentado, com baixa recorrência de consumo entre os usuários, elevada esparsidade e escassez de *feedback* positivo compartilhado. Essas propriedades comprometem diretamente o pressuposto colaborativo utilizado pelos algoritmos MAB, cuja aprendizagem depende da observação repetida de recompensas ao longo das interações.

Diante desse cenário, este trabalho investiga a hipótese de que a aplicação direta de algoritmos MAB tradicionais é limitada por características estruturais do domínio jurídico, e não apenas pela escolha do algoritmo. Para testar essa hipótese, conduzimos uma avaliação experimental com dados reais de interação jurídica, comparando os algoritmos ϵ -Greedy, UCB e Thompson Sampling em diferentes níveis de densidade do conjunto de dados. Complementarmente, foram realizados experimentos de sensibilidade nos conjuntos MovieLens-1M e Yahoo Music, empregando diferentes critérios de binarização das avaliações explícitas, com o objetivo de verificar se a definição de relevância influencia as conclusões obtidas.

Assim, as principais contribuições deste trabalho são:

- (1) Uma caracterização experimental do comportamento de algoritmos MAB em um domínio jurídico real;
- (2) Uma análise do impacto da esparsidade e da densidade das interações sobre o desempenho desses algoritmos;
- (3) Uma comparação sistemática entre diferentes estratégias clássicas de MAB e uma linha de base baseada em popularidade; e
- (4) Evidências experimentais de que limitações estruturais do domínio jurídico reduzem significativamente a efetividade

In: Proceedings of the Brazilian Symposium on Multimedia and the Web (WebMedia'2026). Lavras/MG, Brazil. Porto Alegre: Brazilian Computer Society, 2026.
© 2026 SBC – Brazilian Computing Society.
ISSN 2966-2753

¹<https://grouplens.org/datasets/movielens>

²<https://github.com/saranyavr/Music-Recommendations>

da aprendizagem colaborativa empregada por algoritmos MAB tradicionais.

O restante deste artigo está organizado como segue: Na Seção 2 são apresentados os trabalhos relacionados. A metodologia adotada neste trabalho é apresentada na Seção 3. Na Seção 4 são discutidos os resultados da avaliação experimental. Na Seção 5 são abordadas as limitações. Por fim, na Seção 6, são apresentadas as conclusões e os possíveis trabalhos futuros.

2 TRABALHOS RELACIONADOS

A literatura sobre Sistemas de Recomendação (SR) baseados em algoritmos MAB apresenta aplicações bem-sucedidas em diversos domínios, com foco na otimização da experiência do usuário e na tomada de decisão sequencial. Pesquisas recentes têm se concentrado em superar limitações clássicas dos SR, como o problema de *cold-start*. Nesse contexto, os autores em [12, 14] propuseram a integração de agrupamentos aos algoritmos MAB, permitindo a associação rápida de novos usuários a perfis semelhantes e acelerando a personalização inicial.

Em cenários dinâmicos, em que as preferências mudam ao longo do tempo ou as recompensas são defasadas, abordagens contextuais e probabilísticas demonstraram superioridade em relação a métodos estáticos. Estudos aplicados a plataformas de *streaming* musical e de vídeo, bem como ao marketing digital, utilizaram algoritmos como Thompson Sampling (TS) e Upper Confidence Bound (UCB) para adaptar recomendações interativas em tempo real, equilibrando continuamente exploração e aproveitamento [7, 9, 16]

Adicionalmente, avanços estruturais na modelagem MAB foram propostos para capturar relações complexas entre itens. Em [15], os autores desenvolveram modelos com braços dependentes, combinando regressão de tópicos e filtragem colaborativa para influenciar a estimativa de recompensa de itens correlacionados, enquanto em [10], os autores focaram na reprodutibilidade do *e-commerce* por meio de conjuntos de dados públicos para avaliação fora da política.

Apesar da consolidação dessas técnicas em domínios de entretenimento e de comércio eletrônico, evidencia-se uma lacuna expressiva na aplicação de MAB no contexto de documentos técnicos e não estruturados. Revisões sistemáticas na literatura sobre sistemas de recomendação e recuperação de informação não identificaram trabalhos que aplicassem algoritmos MAB especificamente ao domínio jurídico sob a ótica da aprendizagem sequencial por reforço. O domínio jurídico permanece, portanto, inexplorado nesse paradigma. Diferentemente dos estudos existentes, este trabalho não busca apenas aplicar algoritmos MAB em um novo domínio, mas também investigar experimentalmente como propriedades estruturais do domínio jurídico, especialmente a esparsidade extrema, a baixa recorrência de itens e a escassez de *feedback* compartilhado, influenciam a capacidade de aprendizagem desses algoritmos. Dessa forma, a contribuição concentra-se na compreensão dos limites de aplicação dos métodos tradicionais nesse contexto.

Paralelamente à literatura de Sistemas de Recomendação, o domínio jurídico é investigado sob o paradigma da Recuperação de Informação Legal (*Legal Information Retrieval*, LIR). Trabalhos nessa linha identificam desafios estruturais intrínsecos ao domínio, como a extensão dos documentos, a complexidade da linguagem especializada e a definição de relevância fundamentada em relações

jurídicas específicas, que vão além da similaridade semântica superficial [6]. No contexto brasileiro, a escassez de recursos para LIR em língua portuguesa é apontada como uma barreira ao avanço da área, com poucos conjuntos de dados disponíveis que contemplem consultas reais acompanhadas de julgamentos de relevância [4]. Contudo, esses trabalhos operam sob um paradigma estático de recuperação, no qual consultas explícitas são mapeadas para documentos relevantes por meio de modelos treinados *offline*. Este trabalho investiga um paradigma distinto: a recomendação interativa e sequencial, na qual o sistema aprende incrementalmente a partir do *feedback* implícito de usuários reais em tempo de execução. Essa diferença paradigmática é fundamental, pois desloca o problema da recuperação orientada por consulta para a tomada de decisão sequencial sob incerteza, um cenário ainda inexplorado no domínio jurídico.

3 METODOLOGIA

Este trabalho foi concebido como um estudo de caracterização experimental. Em vez de propor um novo algoritmo de recomendação, o objetivo é compreender em que medida as propriedades estruturais do domínio jurídico afetam o comportamento de algoritmos MAB tradicionalmente utilizados na literatura.

Os experimentos foram conduzidos utilizando o *framework* de código aberto iRec³, projetado para a simulação e avaliação *offline* de sistemas de recomendação interativos. No protocolo adotado, o problema é modelado como uma tarefa de decisão sequencial: a cada iteração, o sistema recebe o contexto do usuário, processa os itens candidatos e fornece uma recomendação. Em seguida, o *framework* consulta os dados de teste em busca de um *feedback* real positivo e atualiza o modelo incrementalmente, simulando o ciclo natural de exploração e aproveitamento.

3.1 Pré-processamento e Construção dos Conjuntos de Dados

O conjunto de dados utilizado neste trabalho, denominado JurisLog, é composto por logs reais de interação entre usuários e documentos jurídicos, coletados de uma grande plataforma brasileira de informação jurídica. Cada registro compreende uma tupla contendo o identificador do usuário, o identificador do documento, o *timestamp* e o *feedback* implícito, representado em binário: 1 para conversões positivas e 0 para ausência de conversão. Os dados foram disponibilizados previamente em formato anonimizado e, dessa forma, informações adicionais relacionadas aos documentos jurídicos, bem como dados que permitissem a identificação dos usuários, não estavam disponíveis para a realização deste estudo.

Na etapa de pré-processamento, as interações foram ordenadas cronologicamente. Para mitigar vieses de avaliação, visto que a política do *framework* impede a recomendação repetida de um mesmo item para o mesmo usuário, removeu-se a duplicidade de pares usuário-item, mantendo exclusivamente a interação mais recente. Essa heurística assume que o último contato reflete com maior precisão o estado final de interesse do usuário pelo documento.

Com o objetivo de realizar uma análise de sensibilidade e isolar o impacto da esparsidade no desempenho dos algoritmos, a base deduplicada derivou três cenários experimentais distintos que estão

³<https://github.com/irec-org>

sumarizados na Tabela 1. O JurisLog-E é o retrato fiel do ecossistema jurídico natural. É composto pelos 10.000 usuários com maior volume de interações, mantendo a alta esparsidade original do domínio. Já o JurisLog-Core-3 é proposto para evitar o colapso do volume de dados que ocorreria ao aplicar filtros no JurisLog-E. Este subconjunto foi extraído de uma base inicial mais ampla (100.000 usuários), retendo apenas itens com no mínimo 3 interações e usuários com no mínimo 5 interações. Por fim, o JurisLog-Core-5 apresenta um cenário denso, análogo aos conjuntos de dados acadêmicos tradicionais. Ele é extraído da mesma base ampliada, mantendo apenas itens e usuários com 5 ou mais interações. Como evidenciado na Tabela 1, a comparação quantitativa com bases consolidadas, como *Yahoo Music* e *MovieLens 1M*, ilustra a severidade da esparsidade no domínio jurídico, o que justifica a criação deste cenário de controle.

A adoção de recortes iniciais distintos decorre do efeito da filtragem por recorrência sobre a estrutura do conjunto de dados. O JurisLog-E foi construído a partir dos 10.000 usuários mais ativos, pois esse recorte já fornecia um conjunto viável para avaliar o cenário original de alta esparsidade, sem impor restrições adicionais à frequência dos itens. Entretanto, ao aplicar critérios mínimos simultâneos para usuários e itens, como nas versões Core-3 e Core-5, o mesmo recorte de 10.000 usuários tornava-se excessivamente restritivo: a combinação de poucos usuários iniciais e baixa recorrência dos itens eliminava grande parte da base, reduzindo drasticamente o número de usuários, itens e interações disponíveis. Por esse motivo, as versões Core foram geradas a partir dos 100.000 usuários mais ativos, permitindo que os filtros de recorrência fossem aplicados sem descaracterizar o experimento por falta de massa amostral. Assim, a diferença entre 10k e 100k não representa uma mudança arbitrária de critério, mas sim uma adaptação necessária ao objetivo de comparar um cenário esparsos com cenários progressivamente mais densos.

A partição experimental foi realizada por meio da estratégia temporal global do *framework* iRec. Nesse procedimento, os usuários são ordenados por *timestamp* de sua primeira interação no conjunto de dados, o que permite identificar a sequência temporal de entrada dos usuários no sistema. A partir dessa ordenação, os usuários com as primeiras interações mais recentes são reservados para o conjunto de teste, enquanto os demais usuários compõem a base de treinamento. O ponto de corte é definido de forma a preservar 80% dos usuários para treinamento e 20% para teste. O conjunto de teste é formado por todas as interações dos usuários selecionados para avaliação, enquanto o conjunto de treinamento contém as interações dos usuários restantes que ocorreram até o ponto de corte temporal definido pela primeira interação do último usuário selecionado para teste. Dessa forma, a avaliação simula um cenário temporalmente ordenado, no qual os algoritmos são treinados com dados de usuários observados anteriormente e avaliados sobre usuários que passam a interagir com o sistema em um período posterior.

Embora o iRec suporte rotinas automatizadas para otimização de hiperparâmetros no conjunto de validação, optou-se pela utilização das configurações e dos parâmetros estáticos padrão do *framework*. Esta decisão metodológica garante um ambiente de avaliação rigorosamente controlado, assegurando que as variações de desempenho entre os algoritmos sejam atribuíveis exclusivamente aos

diferentes níveis de esparsidade dos dados, além de favorecer a reprodutibilidade do estudo.

3.2 Algoritmos Avaliados

Neste trabalho, comparou-se a eficácia de abordagens tradicionais de bandits estocásticos e determinísticos em relação a uma linha de base estatística. A linha de base não personalizada implementada foi o modelo MostPopular, encarregado de recomendar os itens de maior frequência histórica no conjunto de treino. A avaliação deste modelo é crucial em domínios de cauda longa, pois permite aferir se o viés de popularidade global supera a eficácia de sinais fracos de personalização em cenários esparsos.

Os algoritmos descritos a seguir foram selecionados por representarem os métodos clássicos mais frequentemente utilizados na literatura de MAB aplicada a sistemas de recomendação.

Para avaliar a viabilidade da tomada de decisão sequencial no domínio jurídico, implementou-se o algoritmo ϵ -Greedy como um dos modelos preditivos desta pesquisa. A escolha desta política justifica-se por sua capacidade direta de controlar o compromisso entre a exploração de itens jurídicos com potencial ainda desconhecido e o aproveitamento das recomendações de maior relevância já mapeadas pelo sistema.

Na modelagem do experimento, a política de seleção operou da seguinte forma: a cada interação t , o algoritmo decide se explorará ou se aproveitará o conhecimento. Com uma probabilidade ϵ pré-definida, o sistema ignora o histórico e recomenda uma ação (item) de forma puramente aleatória, permitindo a descoberta de novas preferências do usuário. Por outro lado, com probabilidade $1 - \epsilon$, o sistema adota uma postura gananciosa, selecionando a ação a_t^* que maximiza a recompensa histórica estimada, conforme estabelecido pela Equação 1.

$$a_t^* = \arg \max_{a \in \mathcal{A}} Q_t(a), \quad (1)$$

onde $\mathcal{A}$ representa o espaço de ações possíveis, em nosso contexto, documentos jurídicos ou modelos candidatos a recomendação no subconjunto avaliado.

A recompensa estimada ($Q_t(a)$) para cada item foi sistematicamente atualizada no *framework*, calculando-se a média das recompensas empíricas obtidas nas interações passadas. O cálculo da recompensa estimada é dado pela Equação 2

$$Q_t(a) = \frac{1}{N_t(a)} \sum_{\tau=1}^{t-1} r_{\tau}, \quad (2)$$

onde $N_t(a)$ corresponde à frequência com que a ação a foi recomendada até o instante t , e r_{τ} representa o *feedback* de relevância (recompensa) aferido na rodada τ . Esse mecanismo de atualização contínua permitiu que o algoritmo ajustasse suas estimativas dinamicamente diante da esparsidade dos dados do JurisLog.

Dado que a eficácia prática desta abordagem depende criticamente do hiperparâmetro ϵ , adotou-se uma taxa de exploração reduzida, com $\epsilon = 0,0001$, mantida fixa em todos os conjuntos de dados avaliados. Essa configuração prioriza o aproveitamento das estimativas já aprendidas pelo modelo, limitando a exploração aleatória a uma fração mínima das interações. A escolha é coerente com o cenário experimental analisado, no qual o grande número

Table 1: Caracterização quantitativa das bases de dados. As bases Yahoo Music e MovieLens 1M são incluídas como linhas de base para evidenciar o contraste de densidade frente ao domínio jurídico.

Característica	JurisLog-E	JurisLog-Core-3	JurisLog-Core-5	MovieLens 1M	Yahoo Music
Usuários	10.000	15.804	9.262	6.040	15.400
Itens	173.670	10.952	4.429	3.706	1.000
Interações	189.970	105.033	59.599	1.000.209	365.703
Interações/Usuário	19,00	6,65	6,43	165,60	23,75
Interações/Item	1,09	9,59	13,46	269,89	365,70
Densidade	0,0109%	0,0607%	0,1453%	4,4684%	2,3747%
Taxa de positivos/relevantes	14,40%	13,99%	14,14%	57,52%	35,50%

Table 2: Caracterização quantitativa das bases de dados MovieLens 1M e Yahoo Music após binarização do feedback. A configuração 4–5 considera avaliações 4 e 5 como positivas, enquanto a configuração somente 5 considera apenas avaliações 5 como positivas.

Característica	MovieLens 1M 4–5	MovieLens 1M 5	Yahoo Music 4–5	Yahoo Music 5
Usuários	6.040	6.040	15.400	15.400
Itens	3.706	3.706	1.000	1.000
Interações	1.000.209	1.000.209	365.703	365.703
Interações/Usuário	165,60	165,60	23,75	23,75
Interações/Item	269,89	269,89	365,70	365,70
Densidade	4,4684%	4,4684%	2,3747%	2,3747%
Taxa de positivos/relevantes	57,52%	22,63%	35,50%	21,32%

de itens e a elevada esparsidade tornam a exploração aleatória excessivamente pouco informativa e potencialmente prejudicial ao desempenho acumulado.

Para lidar com o desafio empírico de equilibrar a recomendação de itens jurídicos já consolidados e a descoberta de novos documentos relevantes, implementou-se o algoritmo Upper Confidence Bound (UCB). Diferentemente do ϵ -Greedy, que realiza exploração puramente aleatória, o UCB fundamenta-se no princípio do otimismo diante da incerteza [5]. No contexto deste trabalho, isso significa que o *framework* atribui um potencial exploratório maior a documentos jurídicos ou a modelos que foram sub-recomendados, penalizando o viés de popularidade.

Durante os experimentos, a seleção da ação a_t em cada rodada t foi determinística e governada pela maximização conjunta da recompensa histórica e do grau de incerteza, conforme a regra estrutural descrita pela Equação 3.

$$a_t = \arg \max_{a \in \mathcal{A}_t} (R_{a,t} + C_{a,t}) \quad (3)$$

onde $\mathcal{A}_t$ representa o conjunto de ações disponíveis no instante t . O termo $R_{a,t}$ denota a estimativa empírica da recompensa média do braço a , refletindo o aproveitamento do conhecimento já adquirido. O diferencial metodológico concentra-se no termo $C_{a,t}$, que atua como o limite superior de confiança (intervalo de incerteza associado àquela estimativa).

O limite superior de confiança é formalmente definido pela Equação 4.

$$C_{a,t} = c \cdot \sqrt{\frac{\ln t}{N_t(a)}}, \quad (4)$$

onde c é o parâmetro de escala que controla a intensidade da exploração e $N_t(a)$ é o número de vezes em que o braço a foi selecionado até o instante t . Nos experimentos, adotou-se $c = 0,01$, mantendo uma configuração conservadora de exploração. Essa escolha reduz o peso atribuído à incerteza em relação à recompensa empírica observada, o que é particularmente relevante em um domínio caracterizado por grande quantidade de itens e baixa recorrência de interações.

Em complemento às estratégias heurísticas e determinísticas, o *framework* de avaliação incorporou o algoritmo Thompson Sampling (TS), introduzindo uma abordagem estocástica e Bayesiana para o problema de recomendação no domínio jurídico. A principal justificativa para a aplicação do TS neste experimento é sua capacidade de modelar a recompensa de cada documento não como um valor fixo estimado, mas como uma distribuição de probabilidade contínua, permitindo uma adaptação mais suave e dinâmica às preferências dos usuários [1].

Nos experimentos, o Thompson Sampling foi configurado com $\alpha_0 = 1$ e $\beta_0 = 100$, o que define uma distribuição a priori conservadora em relação à probabilidade inicial de recompensa positiva. Essa parametrização reduz a propensão inicial do modelo a assumir alta relevância a itens ainda pouco observados, comportamento adequado para um cenário esparsos, no qual recompensas positivas são raras e a exploração indiscriminada tende a produzir baixo retorno.

Na operacionalização do TS, antes da primeira interação, o sistema assume uma distribuição a priori sobre os parâmetros desconhecidos do ambiente (as preferências no conjunto de dados). A cada nova rodada t , o algoritmo não escolhe a ação de maior média, mas sim realiza uma amostragem aleatória do valor dos parâmetros θ a partir da distribuição a posteriori, atualizada com base no histórico de observações D . A política de seleção da recomendação determina

a ação a_t que maximiza a expectativa de recompensa baseada na amostragem, conforme a Equação 5.

$$a_t = \arg \max_{a \in \mathcal{A}} E(r \mid a, x, \theta), \quad (5)$$

em que $E(r \mid a, x, \theta)$ representa o valor esperado da recompensa r , dada a ação a , o contexto x e os parâmetros amostrados θ . Uma vez que o *feedback* do usuário é recebido, o sistema utiliza a regra de Bayes para atualizar a distribuição a posteriori do respectivo item jurídico.

Do ponto de vista experimental, essa característica Bayesiana conferiu ao modelo a capacidade de explorar itens com alta variância, em nosso contexto, documentos jurídicos cujas probabilidades de sucesso ainda eram incertas e, simultaneamente, explorar intensamente os itens cuja distribuição de recompensa já havia convergido para valores altos. Essa atualização sequencial tornou o TS particularmente robusto para lidar com as distribuições de dados complexas observadas nos subconjuntos avaliados.

3.3 Protocolo de Avaliação e Métricas

A condução dos testes baseou-se na política *FixedInteraction*. O limite de 100 interações por usuário foi estabelecido em consonância com o protocolo de avaliação adotado no iRec [13], permitindo observar o comportamento dos algoritmos em horizontes cumulativos de curto (@10), médio (@50) e longo prazo (@100), além de viabilizar a comparação com resultados reportados na literatura que utilizam o mesmo *framework*. Para cada usuário do conjunto de teste, foi simulada uma sessão fixa de 100 recomendações iterativas. A cada etapa, o modelo recomendou um único item ainda não recomendado anteriormente ao mesmo usuário e não presente em seu histórico conhecido pelo agente. Quando o item recomendado estava presente entre as interações relevantes do usuário no conjunto de teste, a recomendação era contabilizada como acerto. Nos conjuntos de dados jurídicos, por se tratar de *feedback* binário, adotou-se um limiar de relevância de 0,999, de modo que apenas interações positivas com valor 1 fossem consideradas relevantes.

Uma diferença importante entre os conjuntos de dados avaliados diz respeito à natureza do *feedback*. Nos conjuntos de dados jurídicos, as interações são representadas de forma binária, com valores 0 e 1, indicando a ausência ou a presença de *feedback* positivo. Em contraste, os conjuntos de dados de referência do *framework*, como MovieLens-1M e Yahoo Music, utilizam avaliações explícitas em escala ordinal, variando de 1 a 5. Por essa razão, a definição de relevância foi ajustada conforme a estrutura de cada conjunto de dados. Além da configuração originalmente utilizada pelo *framework*, foi conduzida uma análise de sensibilidade sobre os conjuntos MovieLens-1M e Yahoo Music. Nessa etapa, os experimentos foram repetidos considerando duas estratégias de binarização das avaliações explícitas. Na primeira, avaliações iguais ou superiores a quatro foram consideradas relevantes, enquanto as inferiores foram consideradas irrelevantes. Essa configuração é equivalente ao limiar originalmente empregado pelo *framework* e foi utilizada para validar se a etapa de binarização preserva o protocolo experimental. Na segunda estratégia, apenas avaliações iguais a cinco foram consideradas relevantes, tornando o conjunto de *feedback* positivo mais restritivo e aumentando a esparsidade efetiva dos dados. O objetivo dessa análise foi verificar se alterações na definição de

relevância modificariam as conclusões obtidas na comparação entre os conjuntos de dados tradicionais e o domínio jurídico.

Dessa forma, a comparação com MovieLens-1M e Yahoo Music deve ser interpretada como uma referência estrutural, e não como uma equivalência direta entre domínios. Esses conjuntos de dados são utilizados para contextualizar diferenças de densidade, de recorrência de itens e de disponibilidade de sinal colaborativo, enquanto as conclusões principais do estudo são derivadas do comportamento observado nos diferentes recortes do JurisLog.

Além dos experimentos reportados na tabela principal, foram realizadas execuções adicionais com diferentes configurações de particionamento, subconjuntos do conjunto de dados e parâmetros de avaliação, com o objetivo de verificar a estabilidade dos resultados e identificar o protocolo experimental mais adequado ao domínio jurídico. No entanto, para manter a objetividade da análise e facilitar a comparação entre os conjuntos de dados, esse artigo reporta apenas os melhores resultados obtidos para cada algoritmo e cenário avaliados. Essa decisão permite apresentar a configuração de maior desempenho observada empiricamente, sem alterar a interpretação central do estudo.

O desempenho consolidado dos modelos foi aferido em pontos de corte progressivos de 10, 50 e 100 interações, permitindo a observação das taxas de adaptação a curto, médio e longo prazo. As métricas primárias de avaliação utilizadas foram Hits, Precision e Recall, que avaliam, respectivamente, o volume absoluto de recomendações corretas acumuladas, refletindo a maximização de recompensa inerente aos bandits, a proporção de acertos sobre o total recomendado e a proporção de itens relevantes do teste efetivamente recuperados.

No protocolo de avaliação adotado, os algoritmos são avaliados de forma interativa e cumulativa. Para cada usuário, o *framework* simula uma sequência de recomendações e registra se o item recomendado pertence ao conjunto de itens relevantes do usuário no conjunto de teste. As métricas são reportadas nos pontos @10, @50 e @100, correspondendo ao desempenho acumulado após 10, 50 e 100 recomendações por usuário, respectivamente. A métrica Hits representa o número médio absoluto de recomendações corretas por usuário ao longo da simulação. Portanto, Hits@100 = 22,113411 significa que, em média, o algoritmo acertou aproximadamente 22,11 itens relevantes a cada 100 recomendações por usuário. Precision mede a proporção de recomendações realizadas que foram relevantes, isto é, no caso de uma recomendação por rodada, Precision@100 = 0,221134 equivale aproximadamente a 22,11% das 100 recomendações sendo relevantes. Já a Recall mede a fração dos itens relevantes disponíveis no conjunto de teste que foi recuperada pelas recomendações acumuladas. Dessa forma, enquanto Hits@K indica a quantidade média de acertos, Precision indica a eficiência das recomendações feitas e Recall indica a cobertura dos itens relevantes disponíveis para cada usuário.

Embora os experimentos tenham sido conduzidos com diferentes níveis de densidade do conjunto de dados jurídicos e comparados a *benchmarks* amplamente utilizados na literatura, os resultados refletem apenas os algoritmos clássicos implementados no *framework* iRec e o conjunto de dados considerado. Assim, os resultados não devem ser generalizados a todas as abordagens de recomendação

jurídica, especialmente a modelos baseados em conteúdo, arquiteturas híbridas ou algoritmos contextuais, cuja avaliação constitui uma direção para trabalhos futuros.

4 RESULTADOS E DISCUSSÃO

Nesta seção, apresentamos os resultados do trabalho, sumarizados nas Tabelas 3 e 4. A avaliação foi realizada nos domínios jurídico, musical e cinematográfico. Além dos conjuntos jurídicos JurisLog, avaliamos MovieLens-1M e Yahoo Music sob duas definições binárias de relevância: uma configuração menos restritiva, na qual avaliações 4 e 5 são consideradas positivas, e uma configuração mais restritiva, na qual apenas avaliações 5 são consideradas positivas.

A análise dos experimentos mostra diferenças expressivas entre os conjuntos de dados jurídicos e os demais conjuntos de dados disponíveis no *framework*. No JurisLog-E, que representa o cenário jurídico mais esparsos, todos os algoritmos apresentaram valores absolutos muito baixos nas três métricas. O melhor desempenho foi obtido pelo MostPopular, com Hits@100 = 0,016500, Precision@100 = 0,000165 e Recall@100 = 0,005365. Esses valores indicam que, após 100 recomendações por usuário, o algoritmo acerta, em média, apenas 0,0165 item relevante por usuário, isto é, muito menos de um acerto por usuário avaliado. A Precision@100 de 0,000165 corresponde a aproximadamente 0,0165% das recomendações sendo relevantes, enquanto o Recall@100 de 0,005365 indica que apenas cerca de 0,54% dos itens relevantes disponíveis no teste foram recuperados.

Nas versões mais densas do conjunto de dados jurídicos, houve melhora progressiva, embora os valores ainda permaneçam baixos em termos absolutos. No JurisLog-Core-3, o MostPopular alcança Hits@100 = 0,092376, Precision@100 = 0,000924 e Recall@100 = 0,048337. No JurisLog-Core-5, os valores aumentam para Hits@100 = 0,131210, Precision@100 = 0,001312 e Recall@100 = 0,069097. Esse crescimento indica que a redução da esparsidade e o aumento da recorrência mínima dos itens tornam o problema menos severo para os algoritmos. Ainda assim, mesmo no Core-5, o melhor resultado corresponde a aproximadamente 0,13 acerto por usuário a cada 100 recomendações, o que evidencia que o domínio jurídico continua sendo substancialmente mais difícil do que os domínios clássicos de recomendação.

A comparação com MovieLens-1M reforça essa diferença estrutural. No mesmo protocolo, o MostPopular atinge Hits@100 = 22,113411, Precision@100 = 0,221134 e Recall@100 = 0,299400. Isso significa que, em média, o algoritmo acerta cerca de 22 itens relevantes a cada 100 recomendações por usuário, com aproximadamente 22,11% de recomendações relevantes. Esse resultado é de várias ordens de magnitude superior ao observado nos conjuntos de dados jurídicos. A diferença sugere que, no MovieLens, há maior sobreposição de consumo entre usuários e maior recorrência dos mesmos itens, o que favorece algoritmos baseados em popularidade e também fornece um melhor sinal para algoritmos interativos.

O Yahoo Music apresenta um comportamento intermediário em Hits e Precision, mas ainda muito superior ao JurisLog. Para o MostPopular, o resultado em Yahoo Music foi Hits@100 = 4,075649, Precision@100 = 0,040756 e Recall@100 = 0,563203. Isso indica cerca de 4,08 acertos médios a cada 100 recomendações por usuário, com aproximadamente 4,08 das recomendações sendo relevantes.

O Recall mais elevado sugere que, nesse conjunto de dados, uma parcela maior dos itens relevantes disponíveis para cada usuário é recuperada, mesmo que a quantidade absoluta de acertos seja menor do que no MovieLens. Em ambos os conjuntos de dados, os valores são consistentemente superiores aos obtidos nos conjuntos de dados jurídicos.

Como verificação adicional da robustez dos resultados, repetiram-se os experimentos em MovieLens-1M e Yahoo Music utilizando duas estratégias de binarização das avaliações explícitas. Quando foram consideradas relevantes as avaliações iguais ou superiores a quatro, os resultados permaneceram praticamente idênticos aos originalmente reportados, corroborando que a etapa de binarização preserva o protocolo de avaliação utilizado pelo *framework*. Já ao considerar apenas avaliações iguais a cinco como relevantes, observou-se uma redução consistente nos valores absolutos de Hits, Precision e Recall, comportamento esperado devido à diminuição da quantidade de *feedback* positivo disponível. Apesar dessa redução, as tendências observadas entre os algoritmos permaneceram inalteradas, reforçando que as conclusões deste trabalho estão associadas principalmente às características estruturais do domínio jurídico, e não à definição adotada de relevância.

Entre os algoritmos avaliados, o MostPopular foi o melhor em todos os conjuntos de dados jurídicos. Esse resultado é metodologicamente importante, pois indica que, no domínio analisado, a popularidade global dos itens oferece um sinal mais estável do que os mecanismos de exploração e aproveitamento personalizado utilizados pelos algoritmos MAB. Em contextos altamente esparsos, algoritmos como ϵ -Greedy, UCB e Thompson Sampling têm poucas oportunidades de observar recompensas positivas suficientes para estimar preferências individuais ou aprender itens bons de forma incremental. Por isso, eles tendem a apresentar desempenho pior do que uma estratégia simples baseada na popularidade agregada.

Esse comportamento é consistente com a hipótese investigada neste trabalho, segundo a qual a aplicação direta de algoritmos MAB tradicionais é limitada pela alta esparsidade, pela baixa recorrência dos itens e pela escassez de *feedback* positivo compartilhado entre os usuários. A melhora observada entre JurisLog-E e as versões Core indica que o aumento da recorrência favorece o processo de aprendizagem. Entretanto, a distância em relação aos conjuntos de dados, MovieLens-1M e Yahoo Music permanece elevada, sugerindo que essas limitações decorrem principalmente das características estruturais do domínio jurídico.

Dessa forma, os resultados não devem ser interpretados apenas como um baixo desempenho dos algoritmos, mas como evidência empírica de que o domínio jurídico apresenta condições menos favoráveis à recomendação interativa baseada em bandits. Diferentemente de filmes ou músicas, documentos jurídicos tendem a estar associados a necessidades informacionais específicas, casos concretos e consultas pontuais, reduzindo a chance de múltiplos usuários consumirem os mesmos itens. Essa fragmentação enfraquece o sinal colaborativo e torna mais difícil a aprendizagem incremental de preferências.

Esses resultados indicam que a principal limitação observada não está associada a um algoritmo específico, mas sim às condições impostas pelo próprio ambiente de aprendizagem. Em domínios altamente fragmentados, nos quais poucos usuários interagem repetidamente com os mesmos itens, diferentes algoritmos MAB tendem

Table 3: Desempenho comparativo dos algoritmos de recomendação nos conjuntos de dados JurisLog. Valores sublinhados indicam o melhor resultado de cada algoritmo, por métrica, entre os subconjuntos avaliados.

		JurisLog-E			JurisLog-Core-3			JurisLog-Core-5		
Algorithms		@10	@50	@100	@10	@50	@100	@10	@50	@100
MostPopular	Hits	0,004000	0,012500	<u>0,016500</u>	0,024676	0,062322	<u>0,092376</u>	0,037797	0,090173	0,131210
	Precision	<u>0,000400</u>	0,000250	0,000165	<u>0,002468</u>	0,001246	0,000924	<u>0,003780</u>	0,001803	0,001312
	Recall	<u>0,001508</u>	0,004018	0,005365	<u>0,013197</u>	0,034383	0,048337	<u>0,020212</u>	0,048983	0,069097
ϵ -Greedy	Hits	0,000000	0,000500	<u>0,001000</u>	0,003796	0,014869	<u>0,028788</u>	0,004860	0,031317	<u>0,053456</u>
	Precision	0,000000	<u>0,000010</u>	<u>0,000010</u>	<u>0,000380</u>	0,000297	0,000288	0,000486	<u>0,000626</u>	0,000535
	Recall	0,000000	<u>0,000125</u>	<u>0,000208</u>	<u>0,002346</u>	0,008378	0,016498	0,002583	<u>0,017203</u>	0,028904
UCB	Hits	0,001000	0,001500	<u>0,002000</u>	0,001582	0,013287	<u>0,027523</u>	0,003780	0,033477	<u>0,059935</u>
	Precision	<u>0,000100</u>	0,000030	0,000020	0,000158	0,000266	<u>0,000275</u>	0,000378	<u>0,000670</u>	0,000599
	Recall	<u>0,000183</u>	0,000246	0,000329	0,001012	0,007023	<u>0,014689</u>	0,002137	<u>0,018129</u>	<u>0,033711</u>
Thompson Sampling	Hits	0,000000	0,000000	<u>0,001000</u>	0,006643	0,014236	<u>0,021512</u>	0,009719	0,028618	<u>0,044816</u>
	Precision	0,000000	0,000000	<u>0,000010</u>	<u>0,000664</u>	0,000285	0,000215	0,000972	0,000572	0,000448
	Recall	0,000000	0,000000	<u>0,000146</u>	0,003071	0,007695	<u>0,012478</u>	0,004383	0,015335	<u>0,022930</u>

Table 4: Desempenho comparativo dos algoritmos de recomendação nos conjuntos de dados MovieLens-1M e Yahoo Music. Valores sublinhados indicam o melhor resultado de cada algoritmo, por métrica, entre os subconjuntos avaliados.

		MovieLens-1M			Yahoo Music		
Algorithms		@10	@50	@100	@10	@50	@100
MostPopular	Hits	3,764901	13,865894	<u>22,113411</u>	1,216558	3,215584	4,075649
	Precision	<u>0,376490</u>	0,277318	0,221134	<u>0,121656</u>	0,064312	0,040756
	Recall	<u>0,060771</u>	0,199897	0,299400	<u>0,179567</u>	0,456034	0,563203
ϵ -Greedy	Hits	2,320364	9,407285	<u>16,377483</u>	0,554221	2,333442	3,439610
	Precision	<u>0,232036</u>	0,188146	<u>0,163775</u>	0,055422	0,046669	0,034396
	Recall	<u>0,034037</u>	0,131227	0,220174	<u>0,081183</u>	0,334423	0,486280
UCB	Hits	2,392384	9,474338	<u>16,330298</u>	0,565584	2,347078	3,404221
	Precision	<u>0,239238</u>	0,189487	<u>0,163303</u>	0,056558	0,046942	0,034042
	Recall	<u>0,034852</u>	0,130464	0,217820	<u>0,078649</u>	0,331088	0,474860
Thompson Sampling	Hits	2,996689	11,591887	<u>19,293874</u>	0,756818	2,568831	3,518182
	Precision	<u>0,299669</u>	0,231838	<u>0,192939</u>	0,075682	0,051377	0,035182
	Recall	<u>0,045835</u>	0,163980	<u>0,260559</u>	<u>0,107766</u>	0,359729	0,488514

Table 5: Desempenho comparativo dos algoritmos de recomendação nos conjuntos de dados MovieLens-1M e Yahoo Music após binarização do feedback. As configurações 4–5 consideram avaliações 4 e 5 como positivas, enquanto as configurações 5 consideram apenas avaliações 5 como positivas. Valores sublinhados indicam o melhor resultado de cada algoritmo, por métrica, em cada subconjunto avaliado.

		MovieLens-1M 4-5			MovieLens-1M 5			Yahoo Music 4-5			Yahoo Music 5		
Algorithms		@10	@50	@100	@10	@50	@100	@10	@50	@100	@10	@50	@100
MostPopular	Hits	3,869205	14,177980	<u>22,419702</u>	2,493377	7,999172	<u>12,049669</u>	1,216558	3,233442	<u>4,134416</u>	0,760065	2,033117	2,610714
	Precision	<u>0,386921</u>	0,283560	0,224197	<u>0,249338</u>	0,159983	0,120497	<u>0,121656</u>	0,064669	0,041344	<u>0,076006</u>	0,040662	0,026107
	Recall	<u>0,062284</u>	0,204963	0,301340	<u>0,101627</u>	0,286865	0,401733	<u>0,179567</u>	0,459559	0,572034	<u>0,162311</u>	0,406571	0,502642
ϵ -Greedy	Hits	2,352649	10,238411	<u>17,517384</u>	2,000000	6,644040	<u>10,406457</u>	0,554870	2,389610	<u>3,497403</u>	0,317857	1,475000	2,171104
	Precision	<u>0,235265</u>	0,204768	0,175174	<u>0,200000</u>	0,132881	0,104065	<u>0,055487</u>	0,047792	0,034974	<u>0,031786</u>	0,029500	0,021711
	Recall	<u>0,034734</u>	0,144348	0,236565	<u>0,076979</u>	0,234192	0,346827	<u>0,078348</u>	0,340691	0,487826	<u>0,068020</u>	0,298913	0,420901
UCB	Hits	2,385762	10,278146	<u>17,629139</u>	1,980132	6,610927	<u>10,379139</u>	0,567208	2,395130	<u>3,501299</u>	0,318506	1,466558	2,164935
	Precision	<u>0,238576</u>	0,205563	0,176291	<u>0,198013</u>	0,132219	0,103791	<u>0,056721</u>	0,047903	0,035013	<u>0,031851</u>	0,029331	0,021649
	Recall	<u>0,035216</u>	0,144952	0,238595	<u>0,078069</u>	0,232067	0,348551	<u>0,079958</u>	0,341623	0,487636	<u>0,067838</u>	0,294790	0,423532
Thompson Sampling	Hits	2,996689	11,591887	<u>19,293874</u>	2,114238	6,908940	<u>10,581954</u>	0,717208	2,462013	<u>3,480195</u>	0,380519	1,439286	2,060065
	Precision	<u>0,299669</u>	0,231838	0,192939	<u>0,211424</u>	0,138179	0,105820	<u>0,071721</u>	0,049240	0,034802	<u>0,038052</u>	0,028786	0,020601
	Recall	<u>0,045835</u>	0,163980	<u>0,260559</u>	<u>0,083205</u>	0,243810	0,353997	<u>0,103562</u>	0,348260	0,483355	<u>0,077519</u>	0,288942	0,405024

a convergir para desempenhos semelhantes devido à disponibilidade limitada de sinais colaborativos. Assim, a estrutura dos dados passa a exercer influência mais significativa sobre o desempenho do que o mecanismo de seleção de ações adotado por cada algoritmo.

Consequentemente, os resultados sugerem que a simples aplicação de algoritmos MAB tradicionais pode não ser suficiente para sistemas de recomendação jurídica. Nesse contexto, abordagens que incorporem informações semânticas dos documentos, contexto das

consultas ou estratégias híbridas representam direções promissoras para pesquisas futuras.

5 LIMITAÇÕES

Apesar dos resultados elucidarem empiricamente os desafios da aplicação de algoritmos de recomendação no ecossistema jurídico, este estudo apresenta limitações inerentes ao desenho experimental que devem ser consideradas na interpretação de seus achados.

Primeiramente, destaca-se a ausência de otimização de hiperparâmetros nos modelos avaliados. Em virtude de restrições de tempo computacional e do alto custo de execução exigido para simular recomendações interativas em grandes volumes de dados, optou-se pela utilização de parâmetros estáticos baseados nas configurações de referência do próprio *framework* iRec. Embora essa decisão metodológica seja justificada pela necessidade de garantir uma comparabilidade direta e controlada entre os diferentes níveis de esparsidade dos dados, é possível que o desempenho absoluto dos algoritmos MAB pudesse ser marginalmente superior caso fossem ajustados individualmente à topologia específica do domínio jurídico.

Em segundo lugar, o estudo baseia-se na avaliação *offline* de um processo iterativo. O problema foi formulado como uma tarefa sequencial em que a recompensa depende da consulta a uma matriz de interações logadas no conjunto de teste. Simulações *offline* carregam um viés de limitação na capacidade do algoritmo de explorar rotas verdadeiramente inéditas, uma vez que só é possível contabilizar como acertos as recomendações de itens com os quais o usuário já interagiu historicamente no conjunto de dados. Além disso, o *feedback* empírico foi tratado de forma estritamente binária, em que 1 representa conversão e 0 representa uma interação negativa ou ausência de conversão. Em um cenário jurídico real, a falta de interação com um documento não implica necessariamente que ele seja irrelevante, mas pode apenas refletir a dificuldade inicial do usuário em encontrá-lo.

Os experimentos adicionais de sensibilidade reforçam essa conclusão. Mesmo sob diferentes critérios de binarização das avaliações explícitas dos conjuntos de dados, as diferenças observadas entre o domínio jurídico e os conjuntos de dados tradicionais permaneceram consistentes, indicando que a hipótese investigada não depende da escolha do limiar de relevância utilizado na avaliação.

6 CONCLUSÕES E TRABALHOS FUTUROS

Este trabalho investigou a viabilidade empírica de algoritmos de recomendação interativa baseados em MAB aplicados ao domínio jurídico, um cenário de cauda longa caracterizado por extrema esparsidade, alta dimensionalidade de catálogo e escassez de *feedback* positivo. Os resultados sugerem que a aplicação direta de algoritmos MAB clássicos, sem uso de contexto semântico ou de representações de conteúdo, é limitada no cenário jurídico avaliado. Essa limitação foi observada em um protocolo *offline* baseado no iRec e parece estar associada à baixa recorrência de itens e à escassez de *feedback* positivo compartilhado.

Constatou-se que a estratégia baseada em popularidade global (modelo MostPopular) forneceu um sinal preditivo mais robusto e estável do que as tentativas de personalização interativa dos algoritmos estocásticos e determinísticos avaliados (ϵ -Greedy, UCB e

Thompson Sampling). Embora o desempenho dos Bandits tenha apresentado uma melhora progressiva nos cenários de esparsidade artificialmente reduzida (JurisLog-Core-3 e JurisLog-Core-5), os resultados permaneceram em ordens de magnitude inferiores aos observados em conjuntos de dados como MovieLens-1M e Yahoo Music. Essa diferença persistiu mesmo após a binarização de MovieLens-1M e Yahoo Music, sob duas definições de relevância, incluindo uma configuração mais restritiva em que apenas as avaliações máximas foram consideradas positivas. Portanto, a principal contribuição desta pesquisa reside em evidenciar cientificamente que a fragmentação informacional e a ausência de um sinal colaborativo denso sugerem que métodos consolidados em outros domínios não podem ser diretamente transpostos para ecossistemas de alta complexidade técnica sem adaptações ao contexto, à semântica dos itens e à estrutura de *feedback* observada no domínio jurídico.

Como desdobramentos naturais desta pesquisa, delineiam-se diversas frentes de trabalho futuras, concebidas para mitigar as limitações identificadas. Primeiramente, recomenda-se a exploração de abordagens híbridas que incorporem representações baseadas em conteúdo. A utilização de técnicas de Processamento de Linguagem Natural e de Modelos de Linguagem de Grande Escala para capturar a semântica, a jurisdição e as similaridades textuais dos documentos jurídicos pode enriquecer o espaço de estados do modelo. A integração dessas features textuais em algoritmos MAB contextuais, como o LinearUCB ou o Linear Thompson Sampling, mostra-se como uma solução mais promissora para contornar a esparsidade e mitigar o problema de *cold-start*.

Adicionalmente, investigações futuras devem contemplar rotinas extensas de otimização de hiperparâmetros, adequando as taxas de exploração estritamente à topologia dos dados legais. Por fim, a modelagem de funções de recompensa contínuas ou de múltiplos níveis, como tempo de leitura ou número de downloads de documentos, em oposição ao *feedback* estritamente binário *offline*, poderá fornecer sinais de aprendizado mais ricos e densos, favorecendo a convergência das recomendações personalizadas no domínio jurídico.

7 CUIDADOS ÉTICOS

Para garantir a conformidade com as diretrizes de privacidade e a integridade da pesquisa, todos os dados foram anonimizados aleatoriamente. As informações nos conjuntos de dados foram removidas e substituídas por códigos identificadores gerados aleatoriamente, assegurando a irreversibilidade do vínculo com as fontes originais e mantendo a consistência dos dados para os experimentos realizados.

8 AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. A revisão textual e o aprimoramento da redação deste manuscrito foram realizados com auxílio do ChatGPT (OpenAI, GPT-5.5, acessado em junho de 2026). A ferramenta não participou da concepção da pesquisa, da realização dos experimentos, da análise dos resultados ou da elaboração das conclusões. Todas as sugestões foram revisadas e aceitas ou rejeitadas pelos autores, que assumem integral responsabilidade pelo manuscrito.

REFERENCES

- [1] Olivier Chapelle and Lihong Li. 2011. An empirical evaluation of thompson sampling. *Advances in neural information processing systems* 24 (2011).
- [2] Xiacong Chen, Lina Yao, Julian McAuley, Guanglin Zhou, and Xianzhi Wang. 2021. A survey of deep reinforcement learning in recommender systems: A systematic review and future directions. *arXiv preprint arXiv:2109.03540* (2021).
- [3] Keunho Choi, Donghee Yoo, Gunwoo Kim, and Yongmoo Suh. 2012. A hybrid online-product recommendation system: Combining implicit rating-based collaborative filtering and sequential pattern analysis. *electronic commerce research and applications* 11, 4 (2012), 309–317.
- [4] Leandro Carisio Fernandes, Leandro dos Santos Ribeiro, Marcos Vinicius Borela de Castro, Leonardo Augusto da Silva Pacheco, and Edans Flávio de Oliveira Sandes. 2026. JurisTCU: A Brazilian Portuguese Information Retrieval Dataset with Query Relevance Judgments. *Language Resources and Evaluation* 60 (2026), 23. <https://doi.org/10.1007/s10579-025-09881-w>
- [5] Tor Lattimore and Csaba Szepesvári. 2020. *Bandit algorithms*. Cambridge University Press.
- [6] Yixiao Ma, Yueyue Wu, Qingyao Ai, Yiqun Liu, Yunqiu Shao, Min Zhang, and Shaoping Ma. 2023. Incorporating Structural Information into Legal Case Retrieval. *ACM Transactions on Information Systems* 42, 2, Article 40 (2023), 28 pages. <https://doi.org/10.1145/3609796>
- [7] Miguel Martín, Antonio Jiménez-Martín, and Alfonso Mateos. 2019. A numerical analysis of allocation strategies for the multi-armed bandit problem under delayed rewards conditions in digital campaign management. *Neurocomputing* 363 (2019), 99–113.
- [8] James McInerney, Benjamin Lacker, Samantha Hansen, Karl Higley, Hugues Bouchard, Alois Gruson, and Rishabh Mehrotra. 2018. Explore, exploit, and explain: personalizing explainable recommendations with bandits. In *Proceedings of the 12th ACM conference on recommender systems*. 31–39.
- [9] Dattaraj Rao. 2020. Contextual bandits for adapting to changing user preferences over time. *arXiv preprint arXiv:2009.10073* (2020).
- [10] Yuta Saito, Shunsuke Aihara, Megumi Matsutani, and Yusuke Narita. 2020. Open bandit dataset and pipeline: Towards realistic and reproducible off-policy evaluation. *arXiv preprint arXiv:2008.07146* (2020).
- [11] Javier Sanz-Cruzado, Pablo Castells, and Esther López. 2019. A simple multi-armed nearest-neighbor bandit for interactive recommendation. In *Proceedings of the 13th ACM conference on recommender systems*. 358–362.
- [12] Sulthana Shams, Daron Anderson, and Douglas Leith. 2021. Cluster-based bandits: Fast cold-start for recommender system new users. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*. 1613–1616.
- [13] Thiago Silva, Nicollas Silva, Heitor Werneck, Carlos Mito, Adriano CM Pereira, and Leonardo Rocha. 2022. Irec: An interactive recommendation framework. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 3165–3175.
- [14] Thiago Silva, Nicollas Silva, Heitor Werneck, Adriano CM Pereira, and Leonardo Rocha. 2020. The impact of first recommendations based on exploration or exploitation approaches in recommender systems' learning. In *Proceedings of the Brazilian Symposium on Multimedia and the Web*. 173–180.
- [15] Qing Wang, Chunqiu Zeng, Wubai Zhou, Tao Li, S Sitharama Iyengar, Larisa Shwartz, and Genady Ya Grabarnik. 2018. Online interactive collaborative filtering using multi-armed bandit with dependent arms. *IEEE Transactions on Knowledge and Data Engineering* 31, 8 (2018), 1569–1580.
- [16] Chunyi Zhou, Yuanyuan Jin, Xiaoling Wang, and Yingjie Zhang. 2020. Conversational music recommendation based on bandits. In *2020 IEEE international conference on knowledge graph (ICKG)*. IEEE, 41–48.