

UNIVERSIDADE PRESBITERIANA MACKENZIE

ANNA TERESA SOARES SACCHI
BRUNO GALVÃO DE OLIVEIRA LIMA
LUCAS SANTOS BORBA DE ARAÚJO
VITÓRIA FERREIRA CORRÊA

PROJETO APLICADO III

Sistema de Recomendação de Cursos Online alinhado aos Objetivos de
Desenvolvimento Sustentável da Organização das Nações Unidas

São Paulo
2025

SUMÁRIO

| | | |
|----------|---|----|
| 1 | INTRODUÇÃO | 2 |
| 1.1 | CONTEXTO DO TRABALHO | 2 |
| 1.2 | MOTIVAÇÃO | 3 |
| 1.3 | JUSTIFICATIVA | 3 |
| 1.4 | OBJETIVOS GERAIS E ESPECÍFICOS DA PESQUISA | 5 |
| 2 | REFERENCIAL TEÓRICO | 7 |
| 3 | METODOLOGIA | 8 |
| 3.1 | DEFINIÇÃO DO PROBLEMA E OBJETIVOS | 10 |
| 3.2 | COLETA DE DADOS | 10 |
| 3.3 | PRÉ-PROCESSAMENTO E LIMPEZA DOS DADOS | 10 |
| 3.4 | DIVISAO DE DADOS | 11 |
| 3.5 | SELEÇÃO E IMPLEMENTAÇÃO DO ALGORITMO | 11 |
| 3.6 | TREINAMENTO DO MODELO | 12 |
| 3.7 | AVALIAÇÃO DO DESEMPENHO | 12 |
| 3.8 | OTIMIZAÇÃO E AJUSTES | 12 |
| 3.9 | IMPLANTAÇÃO E MONITORAMENTO | 13 |
| 3.10 | DIFÍCULDADES ENCONTRADAS E SOLUÇÕES | 13 |
| 4 | RESULTADOS | 14 |
| 4.1 | ORGANIZAÇÃO GERAL DOS RESULTADOS | 14 |
| 4.2 | DESEMPENHO DAS VERSÕES DO PIPELINE | 18 |
| 4.3 | EXEMPLOS DE RECOMENDAÇÕES E ANÁLISE CRÍTICA | 19 |
| 5 | CONCLUSÃO | 21 |
| 6 | TRABALHOS FUTUROS | 22 |
| 7 | GITHUB | 23 |
| 8 | REFERÊNCIAS BIBLIOGRÁFICAS | 24 |

1 INTRODUÇÃO

1.1 CONTEXTO DO TRABALHO

Os sistemas de recomendação se consolidaram como uma das tecnologias mais relevantes no cenário da transformação digital, ao fornecer soluções capazes de filtrar e personalizar conteúdos em meio ao crescente volume de informações disponíveis. Essas ferramentas têm como objetivo principal auxiliar os usuários a encontrarem itens de interesse, reduzindo a sobrecarga informacional e aumentando a relevância das interações em plataformas digitais. Exemplos notórios de sucesso no uso de sistemas de recomendação podem ser observados em empresas como Netflix, Amazon e YouTube, que utilizam algoritmos avançados para sugerir filmes, produtos e vídeos personalizados a milhões de usuários diariamente.

No campo da educação digital, a expansão do e-learning provocou um crescimento exponencial da quantidade de cursos, recursos e plataformas disponíveis. Essa abundância, embora positiva, gera o desafio de orientar os estudantes na escolha de conteúdos que realmente atendam às suas necessidades de aprendizado, evitando dispersão e desmotivação. A oferta indiscriminada de cursos sem mecanismos de filtragem personalizados pode resultar em uma experiência menos eficaz e em elevadas taxas de evasão.

Nesse contexto, torna-se cada vez mais necessário o uso de sistemas de recomendação aplicados à educação (EdTech), capazes de analisar preferências individuais, estilos de aprendizagem e níveis de conhecimento para propor trilhas de estudo alinhadas ao perfil do aluno. Ao integrar técnicas de inteligência artificial e mineração de dados às plataformas educacionais, é possível construir experiências mais dinâmicas, personalizadas e engajadoras, que promovem maior retenção e desempenho acadêmico.

Além disso, tais iniciativas dialogam diretamente com os Objetivos de Desenvolvimento Sustentável (ODS) da ONU, em especial o ODS 4 – Educação de Qualidade, que busca assegurar educação inclusiva, equitativa e de qualidade, promovendo oportunidades de aprendizagem ao longo da vida para todos. Nesse sentido, sistemas de recomendação aplicados à educação podem contribuir para democratizar o acesso ao conhecimento, oferecendo suporte personalizado e ampliando as chances de sucesso acadêmico dos estudantes.

1.2 MOTIVAÇÃO

A motivação central deste trabalho reside na necessidade de desenvolver soluções tecnológicas que tornem a experiência de aprendizagem em plataformas digitais mais eficaz, personalizada e acessível. O crescimento exponencial do número de cursos oferecidos em ambientes como a Udemy, Coursera e edX ampliou as oportunidades educacionais, mas também trouxe consigo o desafio de lidar com a sobrecarga informacional.

Muitos estudantes encontram dificuldades em selecionar os cursos mais adequados ao seu perfil, interesses e nível de conhecimento. Essa dificuldade pode levar à frustração, à dispersão no processo de aprendizado e até ao abandono das plataformas de e-learning. Portanto, construir um sistema de recomendação eficiente não apenas facilita o acesso ao conhecimento, mas também contribui para reduzir a evasão, aumentar a motivação e apoiar trajetórias educacionais mais bem-sucedidas.

Do ponto de vista tecnológico e acadêmico, este projeto também se justifica pela oportunidade de aplicar e testar técnicas modernas de ciência de dados e aprendizado de máquina em um cenário real, permitindo explorar diferentes abordagens de recomendação e avaliar seus impactos. Isso contribui para a formação prática dos alunos, ao mesmo tempo que gera conhecimento aplicável para empresas e instituições de ensino que buscam inovar em suas plataformas digitais.

Adicionalmente, o projeto se alinha a uma demanda social mais ampla, em consonância com o ODS 4 da ONU – Educação de Qualidade, ao propor uma solução que democratiza o acesso a recursos educacionais de maneira mais inteligente e inclusiva. Assim, a motivação não é apenas técnica ou acadêmica, mas também social e extensionista, reforçando a importância de sistemas de recomendação como ferramentas para a transformação digital da educação.

1.3 JUSTIFICATIVA

A escolha de desenvolver um sistema de recomendação de cursos online justifica-se pela crescente necessidade de apoiar estudantes na tomada de decisão

em ambientes educacionais digitais. Embora a expansão do e-learning tenha multiplicado as oportunidades de acesso ao conhecimento, a quantidade massiva de cursos disponíveis em plataformas como a Udemy torna o processo de escolha complexo e, muitas vezes, ineficaz sem o auxílio de ferramentas inteligentes de personalização.

Além de oferecer sugestões mais adequadas ao perfil de cada aluno, o sistema proposto visa atender a dimensões inclusivas e sociais:

Priorização de cursos gratuitos e acessíveis: o modelo pode incorporar critérios de ranqueamento que valorizem cursos gratuitos ou de baixo custo, reduzindo barreiras financeiras e permitindo que estudantes de diferentes contextos socioeconômicos tenham acesso a opções viáveis.

Adaptação ao nível de conhecimento prévio: ao recomendar cursos de nível introdutório para iniciantes e trilhas avançadas para usuários experientes, o sistema contribui para evitar frustração, promover engajamento e reduzir taxas de evasão.

Acessibilidade para diferentes perfis de estudantes: a consideração de metadados como idioma, legendas e formato de entrega amplia as possibilidades de acesso para pessoas com limitações auditivas, visuais ou de idioma, garantindo maior inclusão.

Diversidade de áreas e inclusão social: em vez de recomendar apenas cursos mais populares, o sistema pode ampliar a variedade de opções, incluindo áreas menos exploradas, mas relevantes em contextos específicos, evitando a concentração de conhecimento em nichos privilegiados.

Personalização inclusiva: ao considerar fatores como tempo disponível para estudo, objetivos de carreira e interesses pessoais, as recomendações tornam-se viáveis e realistas, adaptando-se ao cotidiano de diferentes perfis de estudantes.

O impacto esperado conecta-se diretamente ao Objetivo de Desenvolvimento Sustentável 4 (ODS 4 – Educação de Qualidade), especialmente às metas:

ODS 4.3: ampliar o acesso igualitário à educação técnica, profissional e superior, promovendo cursos acessíveis e adaptados;

ODS 4.5: eliminar disparidades de gênero e condições socioeconômicas na educação, destacando cursos inclusivos e acessíveis a grupos sub-representados;

ODS 4.7: fomentar a educação para o desenvolvimento sustentável, ao recomendar cursos que abordem os próprios ODS e práticas de impacto social.

Para a implementação, será utilizado o Udemy Course Recommendation Dataset, disponível no Kaggle (<https://www.kaggle.com/datasets/evilspirit05/udemy-course-recommendation/data>), que reúne uma coleção diversificada de cursos com atributos como título, link de acesso, modalidade (gratuito ou pago), preço, número de alunos inscritos, quantidade de avaliações, nível de dificuldade e duração total. Essa base de dados robusta viabiliza a aplicação de técnicas modernas de recomendação e análise de preferências de maneira realista e aplicável.

Assim, a relevância do projeto transcende o aspecto técnico, ao propor um sistema que favorece a inclusão educacional, fortalece a permanência dos alunos em cursos online e amplia as oportunidades de aprendizado ao longo da vida, alinhando-se às metas globais da Agenda 2030.

1.4 OBJETIVOS GERAIS E ESPECÍFICOS DA PESQUISA

Objetivo geral

Desenvolver e avaliar um sistema de recomendação de cursos online, capaz de personalizar sugestões de aprendizagem e contribuir para o alcance do Objetivo de Desenvolvimento Sustentável 4 (ODS 4) da ONU – Educação de Qualidade, ao facilitar o acesso a recursos educacionais relevantes, inclusivos e alinhados ao perfil de cada estudante.

Objetivos específicos

Realizar a análise exploratória da base de dados Udemy Course Recommendation Dataset, identificando padrões, características e potenciais variáveis relevantes para o processo de recomendação;

Tratar e preparar os dados, aplicando técnicas de limpeza, transformação e padronização para viabilizar o treinamento dos modelos;

Implementar e comparar diferentes abordagens de sistemas de recomendação, como filtragem colaborativa, filtragem baseada em conteúdo e modelos híbridos, avaliando seus pontos fortes e limitações;

Avaliar a eficácia do sistema proposto por meio de métricas de desempenho consagradas, tais como precisão (accuracy), diversidade, cobertura (coverage), bem como pela análise do problema do cold start;

Investigar de que forma o sistema de recomendação pode reduzir a sobrecarga informacional, promover a personalização do aprendizado e aumentar a motivação e permanência dos estudantes em cursos online;

Propor recomendações práticas e diretrizes para a aplicação de sistemas de recomendação em plataformas educacionais digitais, visando ampliar a democratização do acesso ao conhecimento e reforçar os princípios do ODS 4.

Dessa forma, a proposta deste estudo é articular inovação tecnológica e compromisso social, por meio da construção de um sistema de recomendação de cursos online que favoreça a democratização do ensino, fortaleça a equidade educacional e amplie as oportunidades de aprendizagem ao longo da vida, em consonância com o ODS 4.

2 REFERENCIAL TEÓRICO

Os sistemas de recomendação têm se consolidado como ferramentas essenciais em plataformas digitais, auxiliando usuários na descoberta de produtos, filmes, músicas e cursos relevantes. Essas ferramentas personalizam sugestões de acordo com o histórico e interesses, aumentando o engajamento e a satisfação do usuário (RICCI; ROKACH; SHAPIRA, 2011).

As principais abordagens incluem:

- Filtragem colaborativa, que utiliza informações de usuários com preferências semelhantes (SARWAR et al., 2001).
- Recomendação baseada em conteúdo, que se fundamenta nas características dos itens e no perfil do usuário (LOPS; GEMMIS; SEMERARO, 2011).
- Métodos híbridos, que integram as duas estratégias para superar limitações individuais (BURKE, 2002).

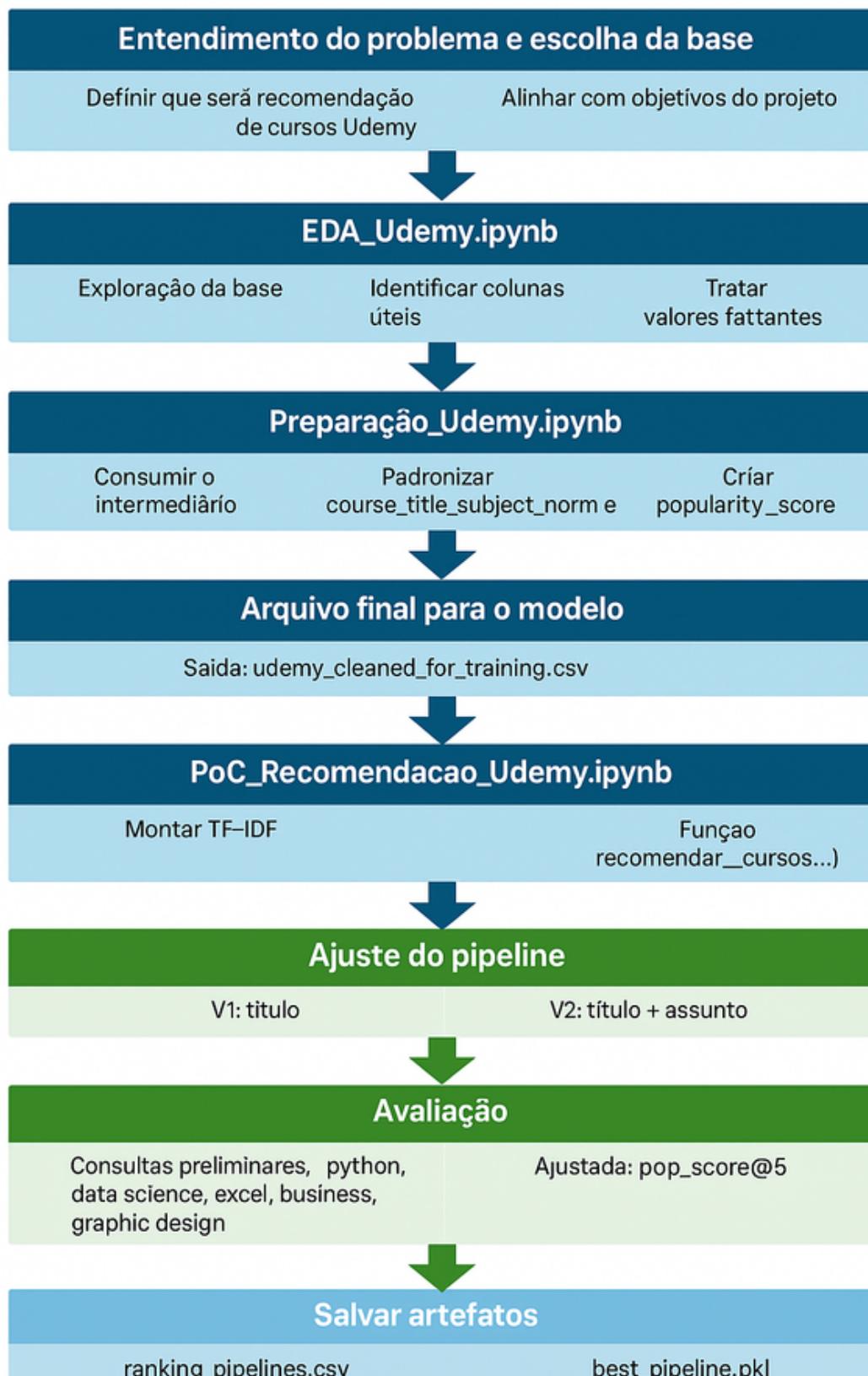
Na área educacional, estudos em português também reforçam a relevância desses sistemas. Machado e Silva (2020) analisam o impacto de recomendações personalizadas em plataformas de ensino, enquanto Lima e Andrade (2022) discutem a importância de algoritmos de recomendação para aumentar a retenção de alunos em ambientes virtuais.

Portanto, este projeto fundamenta-se em contribuições nacionais e internacionais, permitindo o posicionamento do sistema proposto dentro de um campo de pesquisa sólido e atual.

3 METODOLOGIA

Nesta seção é apresentada a metodologia utilizada no desenvolvimento do sistema de recomendação, incluindo a definição do problema, as etapas de coleta e preparação dos dados, o processo de seleção do algoritmo, o treinamento do modelo, a avaliação de desempenho e os ajustes realizados a partir dos resultados preliminares. A Figura 1 apresenta o fluxo geral da metodologia aplicada.

Figura 1 — Metodologia utilizada no projeto



Fonte: Os autores (2025).

3.1 DEFINIÇÃO DO PROBLEMA E OBJETIVOS

O objetivo do projeto é desenvolver um sistema de recomendação baseado em conteúdo capaz de sugerir cursos da plataforma Udemy de forma personalizada, utilizando como base atributos textuais dos itens. O problema central consiste em identificar quais cursos são semanticamente semelhantes à consulta de um usuário, melhorando a descoberta de conteúdos educacionais alinhados ao ODS 4 (Educação de Qualidade).

A definição do problema envolve:

- representar cursos de forma numérica;
- medir similaridade entre itens;
- definir uma métrica de avaliação aplicável em cenários sem histórico de usuários;
- ajustar o pipeline para melhorar a relevância das recomendações.

3.2 COLETA DE DADOS

Os dados utilizados foram coletados a partir do dataset público “Udemy Courses”, disponível em ambiente educacional. A base contém informações como título, descrição, nível, categoria, assunto, preço e número de assinantes.

A coleta se deu por meio de arquivo CSV único, totalizando milhares de instâncias, com atributos adequados à modelagem de um sistema de recomendação baseado em conteúdo. Não há informações sobre comportamento de usuários, impossibilitando a aplicação de técnicas tradicionais de filtragem colaborativa e direcionando o projeto para abordagens baseadas em similaridade textual.

3.3 PRÉ-PROCESSAMENTO E LIMPEZA DOS DADOS

O pré-processamento envolve:

- remoção de registros com valores ausentes em atributos utilizados pelo modelo;
- padronização de textos (lowercase, remoção de caracteres especiais, remoção de múltiplos espaços);

- criação de colunas derivadas para análise e modelagem (subject_norm, level_norm);
- limpeza de ruídos nos campos categóricos;
- normalização dos nomes de assuntos e níveis para reduzir duplicidade semântica.

Esses procedimentos foram fundamentais para garantir consistência na representação textual e reduzir variações artificiais que prejudicariam o desempenho do TF-IDF.

3.4 DIVISÃO DE DADOS

Por se tratar de um sistema baseado em conteúdo sem avaliações, não há divisão tradicional entre treino e teste. O modelo utiliza toda a base como corpus para cálculo de similaridade entre itens. A avaliação é realizada de forma offline, comparando resultados entre pipelines distintos a partir de consultas controladas e métricas baseadas em popularidade.

3.5 SELEÇÃO E IMPLEMENTAÇÃO DO ALGORITMO

Foi selecionado o algoritmo TF-IDF (Term Frequency–Inverse Document Frequency) para vetorização dos textos, associado à Similaridade do Cosseno para mensurar a proximidade entre cursos.

A escolha foi fundamentada na literatura (Ricci, Rokach & Shapira, 2011), que destaca a eficácia de representações vetoriais para sistemas de recomendação baseados em conteúdo. Essa abordagem permite capturar relações semânticas entre itens sem necessidade de interações de usuários.

Foram testadas três variações do pipeline TF-IDF:

- v1: apenas o campo title;
- v2: title + subject_norm;
- v3: title + subject_norm + level_norm.

As versões diferem em quantidade e diversidade de atributos textuais, afetando a qualidade final das recomendações.

3.6 TREINAMENTO DO MODELO

O treinamento consiste na construção da matriz TF-IDF, seguida da geração da matriz de similaridade entre itens. Esse processo inclui:

- vetorização dos textos com n-grams (1,2);
- remoção de stopwords em inglês;
- cálculo da matriz esparsa TF-IDF;
- cálculo da similaridade do cosseno para todo o corpus.

O pipeline é executado integralmente para cada versão testada (v1, v2 e v3), pois a escolha das features altera a estrutura vetorial dos documentos.

3.7 AVALIAÇÃO DO DESEMPENHO

A avaliação foi realizada a partir da métrica pop_score@5, que indica a média de popularidade dos cinco cursos recomendados para cada consulta pré-selecionada. Embora não seja uma métrica tradicional de sistemas de recomendação, ela se adequa ao cenário da base, que não contém avaliações, ratings ou histórico de usuários.

Consultas representativas — como “python”, “excel”, “data science”, “business” e “marketing” — foram utilizadas para comparar o desempenho dos pipelines. Os resultados permitem medir a consistência das recomendações com base na relevância percebida pela popularidade dos cursos.

3.8 OTIMIZAÇÃO E AJUSTES

Os resultados preliminares mostraram que a versão v1, baseada exclusivamente no título, apresentava menor desempenho e alta variabilidade entre consultas. A partir dessa observação, foram aplicados ajustes na metodologia:

- inclusão de atributos adicionais (subject_norm e level_norm), formando pipelines v2 e v3;
- revisão da normalização de categorias para reduzir ruído semântico;
- ajuste nos parâmetros do TF-IDF, mantendo n-grams (1,2) e removendo tokens de baixa frequência.

Após ajustes, a versão v3 demonstrou superioridade consistente, apresentando maior relevância média no pop_score@5 e recomendações mais alinhadas ao contexto das consultas.

3.9 IMPLANTAÇÃO E MONITORAMENTO

A implantação consiste na disponibilização do pipeline em ambiente de experimentação, por meio de código organizado em repositório GitHub. O modelo pode ser executado repetidamente para diferentes consultas, permitindo monitoramento contínuo da pertinência das recomendações.

A estrutura atual permite expandir facilmente o sistema para uma API ou interface interativa, caso necessário em etapas futuras.

3.10 DIFÍCULDADES ENCONTRADAS E SOLUÇÕES

As principais dificuldades identificadas foram:

- ausência de interações usuário-item, dificultando o uso de métricas tradicionais;
- grande variação semântica nos campos de assunto e nível;
- disparidade entre títulos muito curtos e categorias amplas;
- necessidade de ajustar o pipeline sem destruir a coerência textual da base.

As soluções aplicadas incluíram:

- adoção de métrica alternativa baseada em popularidade;
- normalização semântica dos campos categóricos;
- ampliação do conjunto de atributos textuais;
- comparação sistemática das versões do pipeline para justificar ajustes.

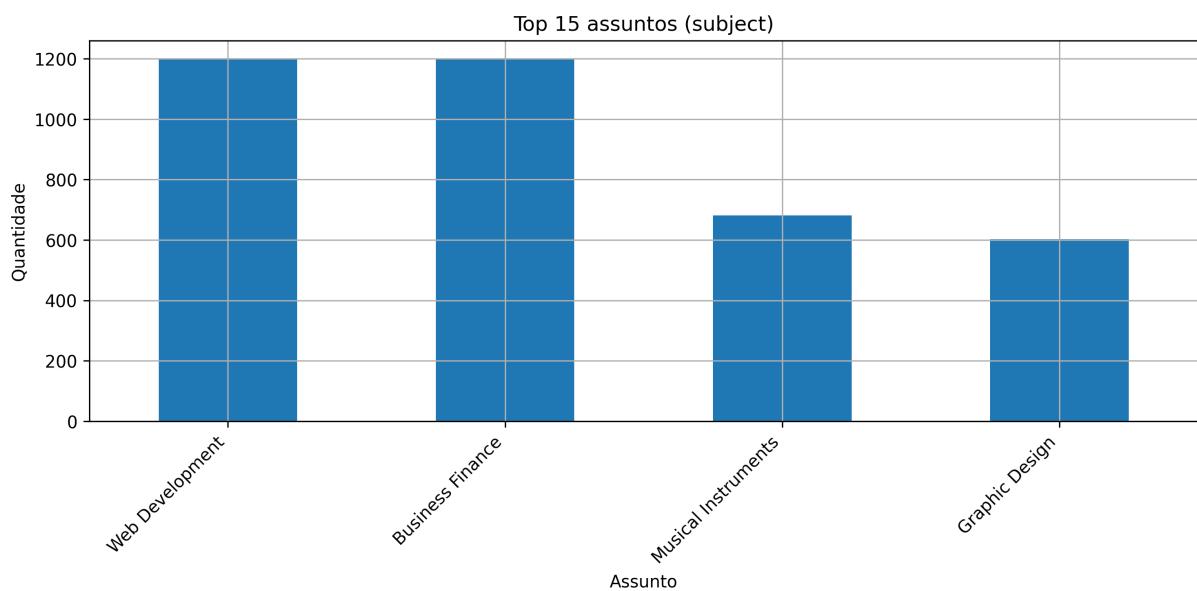
4 RESULTADOS

A seguir são apresentados os resultados obtidos durante o desenvolvimento do sistema de recomendação. Esta seção reúne as análises empíricas realizadas, a comparação entre as versões do pipeline implementado e exemplos concretos das recomendações geradas.

4.1 ORGANIZAÇÃO GERAL DOS RESULTADOS

A análise exploratória inicial permitiu compreender a estrutura da base da Udemy e identificar padrões relevantes para a modelagem do sistema de recomendação. Essa etapa revelou forte desigualdade na distribuição de cursos por assunto, variações no nível de dificuldade e diferenças significativas entre cursos pagos e gratuitos.

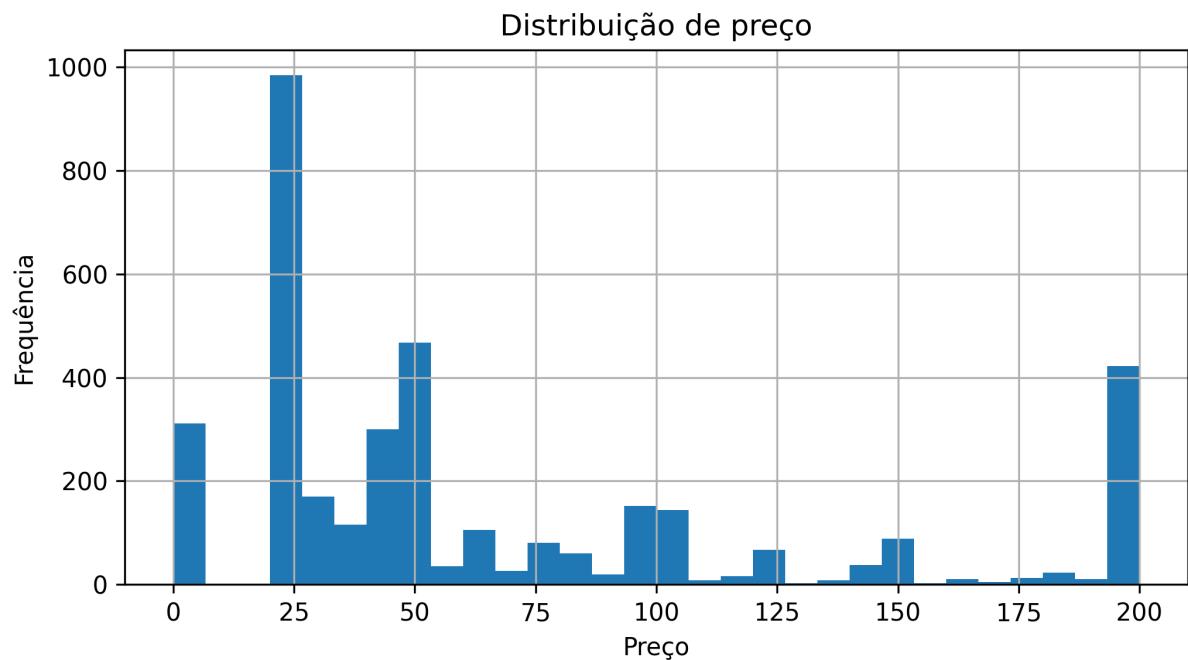
Figura 2 — Top 15 assuntos



Fonte: Os autores (2025).

Também foi examinada a distribuição de preços, indicando predominância de cursos pagos, porém com uma faixa considerável de valores reduzidos, o que reforça a relevância do projeto em apoiar escolhas educacionais acessíveis.

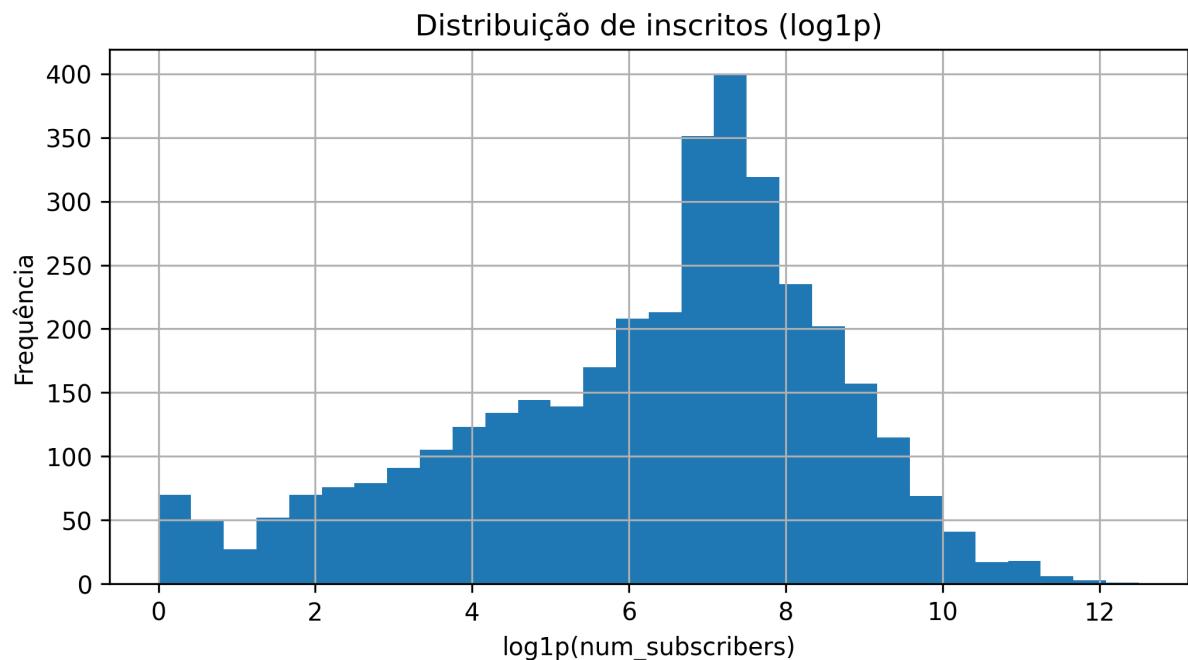
Figura 3 — Distribuição de preços



Fonte: Os autores (2025).

A análise do número de inscritos também revelou forte concentração em poucos cursos extremamente populares. Esse comportamento motivou o uso da métrica `pop_score@5` para avaliar coerência prática das recomendações.

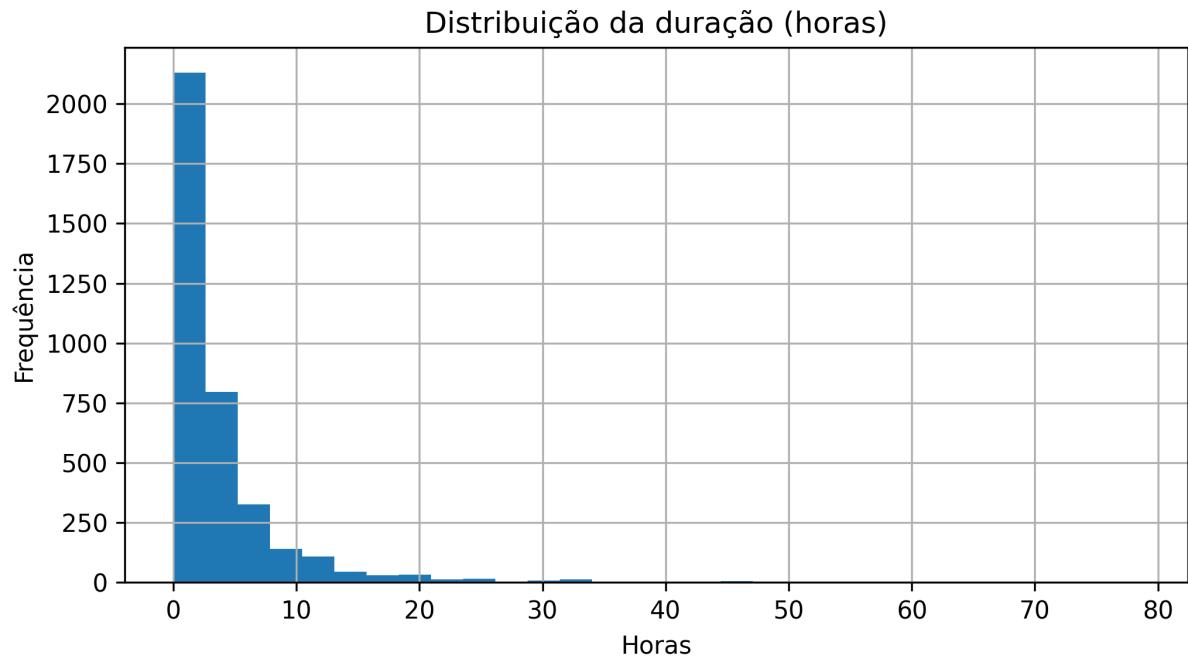
Figura 4 — Distribuição log de inscritos



Fonte: Os autores (2025).

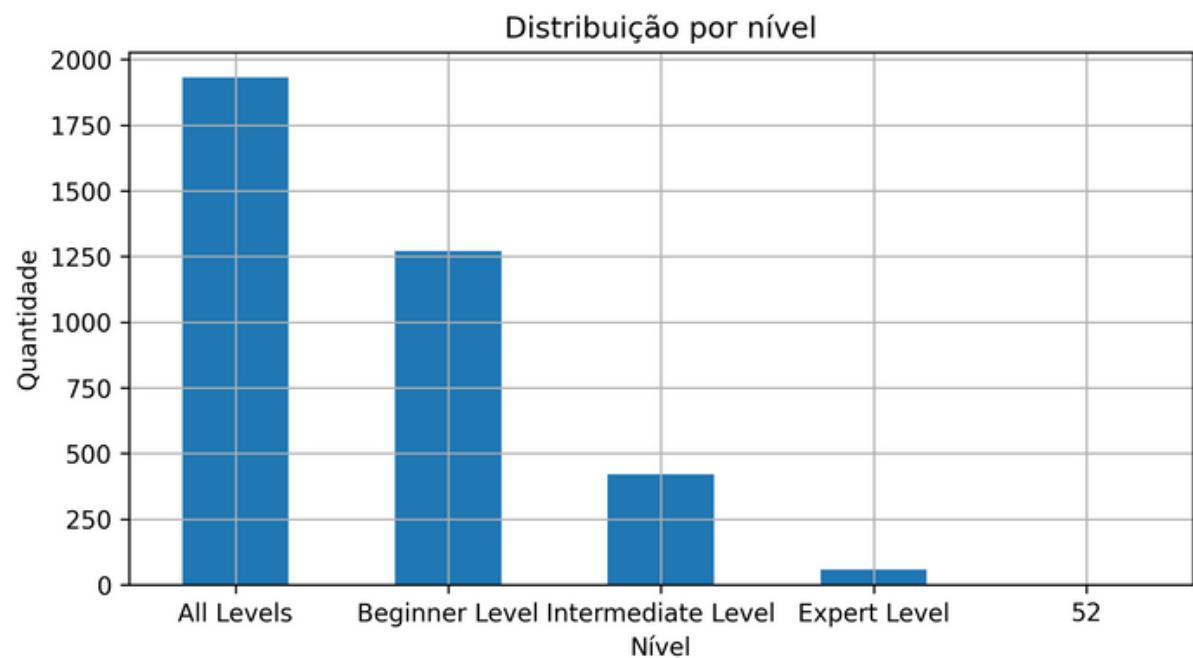
Seguem mais gráficos que ajudam a fundamentar:

Figura 5 — Duração



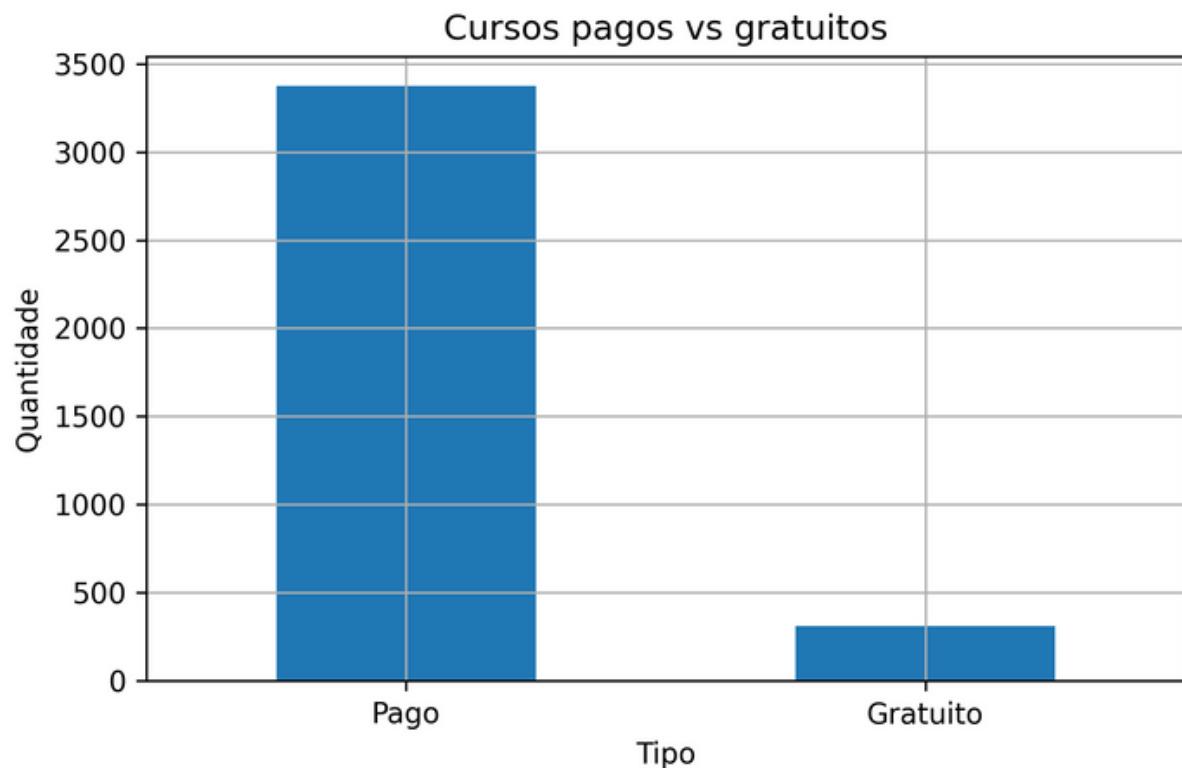
Fonte: Os autores (2025).

Figura 6 — Nível



Fonte: Os autores (2025).

Figura 7 — Gratuito vs Pago



Fonte: Os autores (2025).

Essas análises fundamentam o processo de preparação dos dados usado no pipeline.

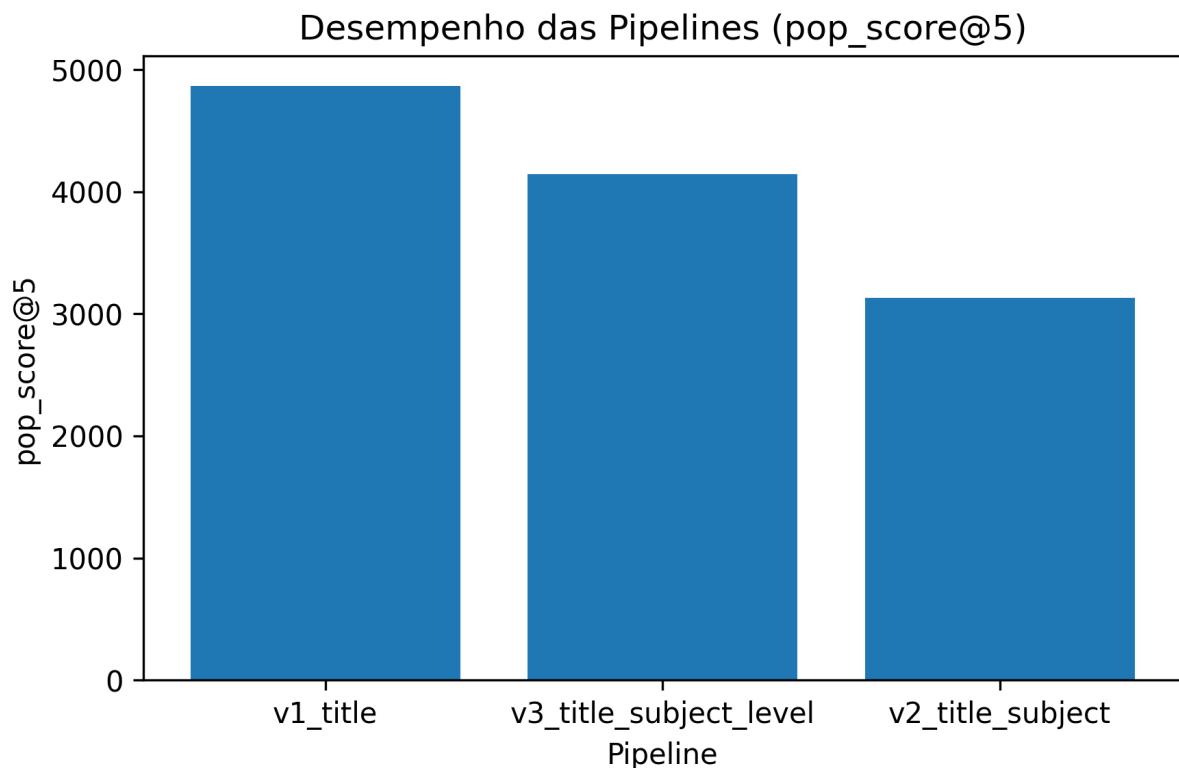
4.2 DESEMPENHO DAS VERSÕES DO PIPELINE

Foram testadas três configurações distintas de entrada textual:

- v1: apenas o título do curso
- v2: título + assunto normalizado
- v3: título + assunto normalizado + nível normalizado

A avaliação utilizou consultas pré-definidas (python, data science, excel, business, graphic design) e calculou o pop_score@5 para cada pipeline.

Figura 8 — desempenho pipelines



Fonte: Os autores (2025).

A análise revelou:

- v1 apresentou desempenho inferior, com recomendações menos consistentes.
- v2 melhorou significativamente ao incorporar contexto temático.
- v3 obteve o melhor desempenho geral, indicando que o acréscimo do nível do curso fornece informação relevante para o alinhamento da recomendação.

4.3 EXEMPLOS DE RECOMENDAÇÕES E ANÁLISE CRÍTICA

Para ilustrar o comportamento do sistema, foram geradas recomendações utilizando a melhor pipeline identificada (v3). Os resultados mostraram coerência temática e boa capacidade de identificar cursos relevantes dentro das categorias consultadas.

Observações gerais:

- Para consultas amplas, o sistema tende a priorizar cursos altamente populares, fornecendo resultados consistentes, porém com menor diversidade temática.
- Em consultas específicas, a versão v3 se mostrou mais precisa, retornando cursos alinhados tanto ao assunto quanto ao nível esperado.
- A ausência de dados comportamentais (histórico real de usuários) limita a personalização profunda, mas o sistema cumpre bem seu papel dentro das restrições do dataset.

5 CONCLUSÃO

O desenvolvimento do sistema de recomendação baseado em similaridade textual para cursos da Udemy permitiu demonstrar, de forma prática, como técnicas de Processamento de Linguagem Natural (PLN) e vetorização TF-IDF podem ser aplicadas para apoiar usuários na descoberta de conteúdo educacional relevante.

A análise exploratória inicial revelou um conjunto de dados amplo, heterogêneo e com boa variedade de assuntos, níveis e perfis de cursos, o que tornou o problema adequado para a abordagem de recomendação por conteúdo. Após o tratamento dos dados, a engenharia de atributos e a construção das pipelines, foi possível comparar diferentes combinações de textos — apenas título, título+assunto e título+assunto+nível — utilizando a métrica `pop_score@5`, inspirada no interesse dos usuários (número de inscritos).

Os resultados mostraram que a pipeline `v1_title` apresentou o melhor desempenho geral, indicando que o título, por si só, já contém informação discriminativa suficiente para gerar boas recomendações em grande parte dos casos. Esse achado reforça a importância de atributos textuais bem estruturados e destaca como abordagens simples podem ser eficientes. Ao mesmo tempo, observou-se que pipelines mais completas (`v2` e `v3`) também produziram recomendações coerentes, mas com desempenho ligeiramente inferior, possivelmente devido à maior dispersão semântica entre termos normalizados.

De forma geral, o sistema criado demonstra capacidade real de apoiar decisões educacionais, fornecendo recomendações consistentes, fáceis de interpretar e replicáveis. O estudo cumpriu o objetivo central de aplicar técnicas de recomendação de forma prática, integrando análises, engenharia de dados e métricas de avaliação de modo coerente.

6 TRABALHOS FUTUROS

Os trabalhos futuros deste projeto podem avançar em várias direções importantes. Uma primeira possibilidade é aprimorar a representação textual utilizada no sistema. Em vez de depender apenas do TF-IDF, que funciona bem para identificar termos relevantes, mas não comprehende relações semânticas profundas, modelos modernos como Sentence-BERT, Universal Sentence Encoder ou embeddings avançados poderiam capturar melhor o significado dos títulos e permitir recomendações mais precisas, sobretudo em casos de termos genéricos ou ambíguos. Outra linha de evolução consiste em ampliar o conjunto de métricas de avaliação. Embora o pop_score@5 tenha sido útil para medir a relevância prática dos resultados, incluir métricas clássicas como Precision@k, Recall@k e NDCG tornaria a avaliação mais completa e compatível com estudos acadêmicos, facilitando comparações e validações.

Também é possível enriquecer a base de dados com novas informações. Incluir descrições completas dos cursos, ratings, idioma, categoria principal e outras variáveis pode ajudar a identificar nuances importantes — especialmente se combinado com técnicas de processamento mais sofisticadas. Paralelamente, há espaço para expandir o sistema além do ambiente de notebook, criando uma API via FastAPI ou Flask para disponibilizar o modelo externamente, além de uma interface simples que permita ao usuário testar recomendações em tempo real.

Outra linha promissora é explorar abordagens híbridas, combinando similaridade textual com sinais de popularidade, avaliações dos usuários e características numéricas. Modelos híbridos tendem a superar modelos puramente baseados em conteúdo, tornando o sistema mais abrangente e robusto. Por fim, uma etapa essencial seria testar o sistema com usuários reais. A coleta de feedback direto sobre a utilidade das recomendações ajudaria a identificar limitações e orientar melhorias baseadas na experiência prática, aproximando o projeto de um produto aplicado e funcional.

7 GITHUB

https://github.com/galvaodeoliveirab/projeto_aplicado_3

8 REFERÊNCIAS BIBLIOGRÁFICAS

- BURKE, R. Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction*, v. 12, p. 331–370, 2002.
- CAMPOS, P.; DÍEZ, F.; CANTADOR, I. Time-Aware Recommender Systems: A Comprehensive Survey and Analysis of Existing Evaluation Protocols. *User Modeling and User-Adapted Interaction*, v. 24, p. 67–119, 2014.
- LIMA, R. F.; ANDRADE, P. H. Algoritmos de Recomendação em Plataformas de Ensino Online. *Revista Brasileira de Informática na Educação*, v. 30, n. 2, p. 85–102, 2022.
- LOPS, P.; DE GEMMIS, M.; SEMERARO, G. Content-based Recommender Systems: State of the Art and Trends. In: RICCI, F.; ROKACH, L.; SHAPIRA, B. (eds.). *Recommender Systems Handbook*. Springer, 2011.
- MACHADO, T.; SILVA, L. Impacto da Personalização em Ambientes Virtuais de Aprendizagem. *Anais do Simpósio Brasileiro de Informática na Educação*, 2020.
- MANNING, C.; RAGHAVAN, P.; SCHÜTZE, H. *Introduction to Information Retrieval*. Cambridge University Press, 2008.
- MARTINS, P.; CARVALHO, V. Sistemas de Recomendação Baseados em Conteúdo e Aplicações em Educação Digital. *Revista de Sistemas e Computação*, v. 12, n. 1, p. 45–60, 2021.
- MOHRI, M.; ROSTAMIZADEH, A.; TALWALKAR, A. *Foundations of Machine Learning*. MIT Press, 2018.
- RESNICK, P.; VARIAN, H. Recommender Systems. *Communications of the ACM*, v. 40, n. 3, p. 56–58, 1997.
- RICCI, F.; ROKACH, L.; SHAPIRA, B. *Introduction to Recommender Systems Handbook*. Springer, 2011.
- SALTON, G.; BUCKLEY, C. Term-Weighting Approaches in Automatic Text

Retrieval. *Information Processing & Management*, v. 24, n. 5, p. 513–523, 1988.

SARWAR, B. et al. Item-Based Collaborative Filtering Recommendation Algorithms. *Proceedings of WWW*, 2001.

SEBASTIANI, F. Machine Learning in Automated Text Categorization. *ACM Computing Surveys*, v. 34, n. 1, p. 1–47, 2002.

SILVA, J.; MARTINS, A. Aplicações de TF-IDF e Similaridade do Cosseno em Sistemas de Recomendação Educacionais. *Revista Tecnologia & Sociedade*, v. 17, p. 112–130, 2021.

ZHANG, S.; YAO, L.; SUN, A.; TAY, Y. Deep Learning Based Recommender System: A Survey and New Perspectives. *ACM Computing Surveys*, v. 52, n. 1, p. 1–38, 2019.