www.aileaders-project.eu

EXERCÍCIO DE CENÁRIO

- Sistema de recomendação online

Co-funded by the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.

EXERCÍCIO DE CENÁRIO - Sistema de recomendação online

01 Resumo

Tipo de REA

Exercício de cenário usando o Google Colab (riverML)

Objetivo/Finalidade

Este exercício baseado em cenários coloca os alunos no papel de um profissional de aprendizagem automática (ML) encarregado de analisar a imparcialidade do sistema de recomendação online. Ele explora como a adaptação contínua ao comportamento do utilizador pode amplificar os preconceitos existentes ao longo do tempo. Os alunos serão desafiados a identificar preconceitos e avaliar estratégias de mitigação.

Resultados de aprendizagem esperados

O aluno será capaz de:

- Identificar potenciais fontes de viés nas recomendações geradas por ML;
- Comparar métricas de equidade, como paridade demográfica e equidade de exposição;
- Compreender como os ciclos de feedback afetam a equidade ao longo do tempo em sistemas de recomendação online.

Palavras-chave

- Aprendizagem automática
- Recomendação
- Preconceitos
- Equidade

Sugerido Metodológico Abordagem

Aprendizagem baseada em problemas.

02 Introdução

No contexto de experiências digitais cada vez mais personalizadas, os sistemas de recomendação baseados em IA desempenham um papel fundamental na formação do comportamento do consumidor, influenciando as decisões de compra e determinando a visibilidade de produtos e conteúdos em todos os segmentos de utilizadores.

Embora esses sistemas ofereçam vantagens consideráveis para o marketing e a otimização dos negócios, eles também levantam questões éticas e regulatórias urgentes, particularmente no que diz respeito à equidade, transparência e potencial discriminação (Akter et al., 2022; Bozdag, 2013).

Os sistemas de recomendação não são meros artefactos algorítmicos, mas construções sociotécnicas complexas, influenciadas por decisões humanas em todas as etapas — desde a recolha de dados e o design do modelo até a interpretação e implementação dos resultados (Bozdag, 2013). Os preconceitos podem surgir de várias fontes: preconceitos humanos, limitações técnicas ou desalinhamentos contextuais. Especialmente em ambientes de marketing, esses preconceitos podem interagir e se agravar, moldando não apenas o desempenho do sistema, mas também as experiências dos utilizadores e os resultados sociais (Akter et al., 2022).

Uma preocupação significativa reside na natureza iterativa dos sistemas de recomendação, que dependem do feedback do utilizador para retreinar e refinar modelos preditivos. Este ciclo de feedback pode reforçar os desequilíbrios existentes entre entrada e saída, levando a disparidades persistentes ao longo do tempo (Sun et al., 2020). À medida que os modelos incorporam continuamente novos dados e preferências dos utilizadores, a sua equidade e precisão estão sujeitas a mudanças dinâmicas. Se não for controlado, isso pode resultar em um ajuste excessivo aos comportamentos dominantes dos utilizadores, exacerbando a sub-representação de grupos minoritários e restringindo a diversidade de conteúdo — um fenómeno associado ao gatekeeping algorítmico e ao aumento das bolhas de filtro (Harambam et al., 2018).

Motivado por esses desafios, este exercício baseado em cenários convida os alunos a explorar as dimensões éticas dos sistemas de recomendação usando o **conjunto de dados MovieLens 100K** como banco de testes.

Através de experiências práticas, os alunos examinam como os ciclos de feedback algorítmico podem amplificar involuntariamente as desigualdades, particularmente ao distorcer a visibilidade do conteúdo e favorecer as preferências da maioria. Eles são incentivados a investigar os aspetos técnicos e éticos desses sistemas, unindo o conhecimento teórico à análise prática. Essa abordagem pedagógica promove a consciência crítica dos riscos associados à personalização antiética, incluindo a perda da confiança do consumidor, danos à reputação e acesso reduzido a oportunidades para usuários marginalizados. O exercício destaca que a equidade na IA não é uma avaliação estática, mas uma responsabilidade dinâmica e contínua. À medida que as preferências dos utilizadores evoluem e o conteúdo muda, os modelos também devem se adaptar para manter um desempenho equilibrado em todos os grupos de utilizadores. Em última análise, essa experiência de aprendizagem equipa os alunos com as ferramentas para projetar sistemas de IA que não sejam apenas eficazes e eficientes, mas também equitativos, transparentes e socialmente responsáveis.

Métricas de equidade

01 **Erro de estimativa absoluto** — mede o erro absoluto entre a classificação prevista e a classificação real dada pelo utilizador u ao item I, definido por:

$$\varepsilon_a(u,i) = |\hat{r}_{u,i} - r_{u,i}|$$

02 **Erro de sobreestimativa** - mede o quanto a previsão sobreestima a classificação real, definido por:

$$\varepsilon_o(u,i) = \max(\hat{r}_{u,i} - r_{u,i}, 0)$$

03 **Erro de subestimativa** - mede o quanto a previsão subestima a classificação real, definido por:

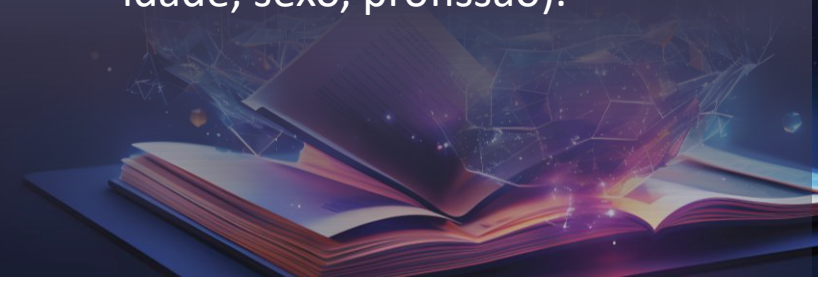
$$\varepsilon_u(u,i) = \max(r_{u,i} - \hat{r}_{u,i}, 0)$$

04 **Erro de estimativa de valor** - mede o erro assinado entre a classificação prevista e a classificação real, definido por:

$$\varepsilon_v(u,i) = \hat{r}_{u,i} - r_{u,i}$$

03 Apresentação das ferramentas

Este exercício utiliza o **conjunto de dados MovieLens 100K**, uma referência amplamente utilizada na investigação de sistemas de recomendação. O conjunto de dados inclui 100 000 classificações, 943 utilizadores, 1682 filmes, juntamente com dados demográficos dos utilizadores (por exemplo, idade, sexo, profissão).



Embora útil para testes de algoritmos

o conjunto de dados contém desafios significativos de equidade

desafios de imparcialidade

01 **Representação desequilibrada entre grupos demográficos;**

02 **Distribuições de classificação distorcidas;**

03 **Representação excessiva de itens populares.**

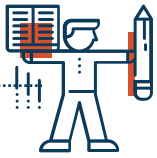
Estas características tornam o MovieLens 100K ideal para explorar o viés e a imparcialidade nos algoritmos de recomendação.

RiverML[1] é uma biblioteca Python de código aberto projetada especificamente para aprendizado de máquina em fluxos de dados. Ela fornece ferramentas para construir, treinar e avaliar modelos que aprendem incrementalmente a partir de dados que chegam em tempo real, em vez de em lotes. O RiverML suporta algoritmos de aprendizagem supervisionados e não supervisionados, oferecendo recursos como aprendizagem adaptativa, deteção de desvio de conceito e pré-processamento de fluxo. O seu design modular facilita a combinação de modelos, etapas de pré-processamento e estratégias de avaliação, tornando-o uma escolha poderosa para aplicações de aprendizagem online, como deteção de fraudes, sistemas de recomendação e análise de dados de sensores. As métricas de equidade disponíveis na biblioteca foram implementadas pela equipa FEP, contribuindo para a avaliação ética dos modelos de streaming.

Esta ferramenta permite a execução de pipelines de avaliação de equidade e a comparação visual do desempenho das recomendações entre grupos e ao longo do tempo.

1. Tutorial Riverml - https://riverml.xyz/latest/introduction/installation/

04 Atividades práticas



Atividade

01 Acesse o Jupyter Notebook fornecido em https://tinyurl.com/5n8vvpx9.

02 Execute todas as células no notebook (clique na guia Runtime e, em seguida, em Run All).

03 Analise e compare as métricas de equidade entre diferentes grupos de utilizadores (por exemplo, por género, idade).

04 Observe como os indicadores de equidade evoluem à medida que mais interações dos utilizadores são recolhidas

Sugestões para discussão

01 Quais grupos de utilizadores são sistematicamente favorecidos ou desfavorecidos?

02 Como a qualidade das recomendações muda ao longo do tempo?

03 Que compromissos surgem quando se tenta equilibrar o envolvimento com a equidade?

05 Conclusão

Este exercício ilustra a natureza dinâmica da equidade nos sistemas de recomendação online.

Embora o desempenho inicial do sistema possa parecer imparcial, a operação a longo prazo pode levar a disparidades significativas devido a ciclos de feedback e viés de popularidade. Esses desequilíbrios afetam a experiência e a confiança do utilizador, especialmente para grupos de utilizadores sub-representados ou minoritários. Ao aplicar métricas de justiça e analisar a evolução do viés em tempo real, os alunos obtêm uma visão prática sobre a implementação ética do ML. Eles também exploram as limitações da avaliação estática e a importância do monitoramento contínuo da justiça em sistemas adaptativos. O exercício destaca a tensão entre otimizar o envolvimento e garantir um tratamento equitativo — um desafio fundamental para o design ético da IA.

06 Referências

- Bozdag, E. (2013). Viés na filtragem algorítmica e personalização. Ética e Tecnologia da Informação, 15(3), 209–227.
- Akter, S., Michael, K., Bandara, R., Wamba, S.F., Foropon, C., & Papadopoulos, T. (2022). Viés algorítmico em modelos de marketing baseados em aprendizagem automática: uma visão da capacidade dinâmica. Jornal de Investigação Empresarial.
- Harambam, J., Helberger, N., & van Hoboken, J. (2018). Democratizando recomendadores algorítmicos de notícias: como materializar a voz em um ecossistema de mídia tecnologicamente saturado. Big Data & Society.
- Sun, T., Gautam, K., & Joachims, T. (2020). Eliminando o viés da aprendizagem a partir da interação do utilizador. Anais da 43ª Conferência Internacional ACM SIGIR sobre Pesquisa e Desenvolvimento em Recuperação de Informação.

leaders

Acompanhe a nossa jornada

www.aileaders-project.eu

Co-funded by the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.