

ANÁLISE DA ARQUITETURA DE IA

A IA projetada para o sistema de recomendações personalizadas foi criada para fornecer informações para os usuários com base nos seus interesses pessoais, profissionais e acadêmicos. O sistema foi desenvolvido através de uma série de etapas que garantiram um bom funcionamento e eficácia, passando por etapas desde a coleta e preparação dos dados fornecidos até o treinamento e validação do modelo final.

Dados como interesse profissional, interesse pessoal e área de interesse são coletados e armazenados em uma base de dados (Excel) e passam por um pré-processamento onde o primeiro passo é a limpeza dos dados, que remove os valores ausentes e inconsistentes.

Para construção do sistema, foi utilizado o modelo Híbrido de Recomendações e combina dois métodos.

- Filtrada baseada em conteúdo: Foi utilizado técnicas como TF-IDF (Term Frequency-Inverse Document Frequency) para transformar os interesses dos usuários em vetores de características, o modelo compara esses vetores com os vetores de interesses dos cursos disponíveis, gerando uma similaridade.
- Filtragem colaborativa: A recomendação dos cursos é feita com base nas informações de outros usuários, utilizando algoritmos de aprendizado para identificar padrões e semelhanças nas escolhas de cursos.

A decisão de usar um modelo híbrido foi tomada para combinar os pontos fortes de cada abordagem: a filtragem baseada em conteúdo é eficaz em gerar recomendações personalizadas, enquanto a colaborativa aproveita o comportamento coletivo dos usuários.

Foram escolhidos dois algoritmos para desenvolver o sistema de recomendações (Regressão Linear e K-Nearest Neighbors) pela eficiência e simplicidade para lidar com dados e encontrar padrões. A Regressão Linear nesse sistema foi usada para prever as notas dos cursos com base nos interesses do usuário e o K-Nearest Neighbors foi utilizado para identificar cursos semelhantes com base na preferência de outros usuários com interesses parecidos.

Após o pré-processamento, os dados são divididos em conjuntos de treino e teste, onde o modelo passa por um treinamento usando os dados de treino e aprende a associar interesses com as notas dos cursos. O desempenho do modelo foi

validado com os dados de teste, onde métricas como o Erro Quadrático Médio (MSE) são utilizadas para avaliar a precisão das previsões.

As recomendações em tempo real é uma das principais características que foram implementadas através de:

- API em tempo real: O sistema expõe um endpoint que processa as entradas dos usuários (interesses) e retorna recomendações imediatas.
- Atualização contínua de dados: O sistema integra os novos dados fornecidos pelos usuários de forma contínua na base de dados garantindo que as recomendações sejam baseadas nas informações mais recentes.

Esse componente é essencial e uma das principais características por garantir que o sistema se adapte dinamicamente a novos perfis de usuários, além disso um bom desempenho é fundamental para garantir a escalabilidade e a eficiência em larga escala. Para otimização do sistema, foram implementadas as estratégias de Indexação e Caching, Paralelização e Escalabilidade Horizontal.

- Indexação e Caching: Foi utilizado técnicas de cache para o armazenamento temporário dos resultados e consultas frequentemente realizadas, reduzindo o tempo de processamento.
- Paralelização: O cálculo de similaridade entre vetores é paralelizado para reduzir o tempo de resposta quando o sistema lida com grandes volumes de dados.
- Escalabilidade Horizontal: O sistema foi projetado para ser escalável horizontalmente, o que permite a adição de novos servidores conforme a demanda cresce.

Com essas otimizações, o sistema pode lidar com grandes volumes de dados e usuários simultâneos sem comprometer a performance.

O feedback dos usuários é utilizado para refinar o sistema continuamente. Cada interação dos usuários (como notas e comentários) é armazenada e usada para melhorar as recomendações. As principais técnicas de melhoria incluem:

- Retreinamento periódico do modelo: O modelo é retreinado regularmente com novos dados para garantir que as recomendações permaneçam relevantes e atualizadas.
- Ajuste hiper parâmetros: O desempenho do modelo é monitorado constantemente, e ajustes automáticos podem ser feitos para melhorar a precisão.

Essa abordagem garante que o sistema de recomendação continue a melhorar com o tempo, adaptando-se às necessidades dinâmicas dos usuários e das tendências de interesse.

Portanto, essa arquitetura foi escolhida pela sua capacidade de lidar com dados e gerar recomendações relevantes com base nos interesses dos usuários. O resultado é a combinação da flexibilidade da filtragem baseada em conteúdo com a robustez da filtragem colaborativa, garantindo recomendações precisas e uma experiência otimizada. A implementação em tempo real, juntamente com a escalabilidade e o feedback contínuo, assegura que o sistema possa crescer e evoluir junto com a base de usuários.