

Desenvolvimento de um aplicativo de recomendação de supervisores para PFC com base em Inteligência Artificial

Rafael Beto Mpfumo* e Walter Conceição

Departamento de Tecnologias da Informação e Comunicação, Instituto Superior de Transportes e Comunicações, Prol. Av. Kim Il Sung, Maputo, Moçambique

* e-mail do autor correspondente: rmpfumo@isutc.ac.mz

Resumo – A escolha de um supervisor de projecto Final de Curso (PFC) é um passo crucial para o sucesso académico e profissional do estudante. A decisão, no entanto, é frequentemente um desafio devido à falta de informações claras sobre a compatibilidade entre as áreas de interesse do aluno e a expertise dos professores. Este projecto tem como objectivo desenvolver um algoritmo baseado em Inteligência Artificial (IA) para recomendar supervisores de PFC de forma eficiente, baseando-se nos interesses de pesquisa dos alunos e nas publicações dos docentes. Adoptando a metodologia *Design Science Research* (DSR), o artefacto foi construído utilizando Processamento de Linguagem Natural (NLP), especificamente a vectorização TF-IDF e o cálculo de Similaridade de Cosseno. Os testes de validação com 100 resumos de alunos demonstraram um desempenho promissor, alcançando uma acurácia Top-3 de 92% (o supervisor mais compatível estava entre os três primeiros recomendados). A eficiência do processo foi aprimorada em 95%, reduzindo o tempo de alocação de horas para poucos minutos. A pesquisa conclui que o sistema de recomendação é uma solução viável para otimizar a alocação de supervisores de PFC, melhorando a qualidade dos projectos desenvolvidos e a experiência académica dos estudantes.

Palavras-chave – Alocação de Supervisores; Projeto Final de Curso (PFC); Siste-mas de Recomendação (SR); Processamento de Linguagem Natural (NLP); Design Science Research (DSR).

I. INTRODUÇÃO

A evolução das Tecnologias de Informação e Comunicação (TIC), especialmente no campo da Inteligência Artificial (IA), tem permitido a criação de sistemas capazes de resolver problemas complexos em diversas áreas, incluindo a educação. Um dos desafios recorrentes nas instituições de ensino superior é a alocação de supervisores para projetos Finais de Curso (PFC). A compatibilidade entre o tema de pesquisa do aluno e a área de especialização do professor é um fator crítico para o sucesso do projecto (Baker & Yacef, 2009).

A escolha do supervisor impacta directamente a qualidade do trabalho, o desenvolvimento de habilidades do aluno e, em última instância, sua trajectória profissional.

A falta de um sistema formal e baseado em dados para essa alocação muitas vezes resulta em escolhas sub-rotinas, baseadas em indicações informais ou na disponibilidade de professores, e não na sua real expertise. Isso pode levar a projetos com menor relevância, desalinhados com a pesquisa do docente, e a uma experiência de supervisão insatisfatória.

A IA, com seus avanços em Processamento de Linguagem Natural (NLP) e Sistemas de Recomendação (SR), oferece uma solução promissora para este problema (Gao et al., 2020). Sistemas de recomendação já são amplamente utilizados em plataformas de e-commerce e serviços de *streaming* (Smith & Jones, 2023). Adaptar essa tecnologia para o ambiente académico pode otimizar o processo de alocação de supervisores, garantindo que a escolha seja feita de forma mais informada e estratégica.

Este projecto visa, portanto, desenvolver um aplicativo de recomendação de supervisores para PFC, utilizando um algoritmo de IA capaz de analisar e combinar os interesses de pesquisa de alunos e professores, servindo como uma ferramenta valiosa para estudantes e coordenações de curso.

II. REVISÃO DA LITERATURA: CONCEITOS E APLICAÇÕES

Esta secção define os conceitos fundamentais que sustentam a pesquisa.

2.1. Sistemas de Recomendação (SR)

Sistemas de Recomendação (SR) são ferramentas de filtragem de informação que buscam prever a preferência de um usuário por um item. Eles são geralmente classificados em filtragem colaborativa (baseada em semelhanças entre usuários) ou filtragem baseada em conteúdo (baseada em semelhanças entre os atributos dos itens e as preferências do usuário) (Marx & Muller, 2018). No contexto académico, a recomendação de supervisores é um problema de recomendação baseada em conteúdo, onde o “item” (supervisor) e a “preferência” (o tema do aluno) são descritos por texto.

2.2. Processamento de Linguagem Natural (NLP)

O Processamento de Linguagem Natural (NLP) é um campo da IA que permite aos computadores entender,

interpretar e gerar linguagem humana. Para problemas de recomendação baseada em texto, o NLP é crucial. A técnica *TF-IDF* (*Term Frequency-Inverse Document Frequency*) é a ferramenta padrão para quantificar a importância de uma palavra em um documento em relação a um corpus (coleção de documentos), transformando texto em vetores numéricos para análise matemática (Baker & Yacef, 2009). A Similaridade de Cosseno é subsequentemente usada nesses vetores para medir a proximidade temática.

2.3. Pesquisa Orientada ao Design Science Research – (DSR)

A metodologia DSR é fundamental em sistemas de informação. Conforme definido por Marx e Müller (2018), o objectivo da *DSR* é construir e avaliar um artefacto (modelo, método ou sistema) para resolver um problema relevante. A *DSR* aumenta a relevância da pesquisa ao produzir soluções aplicáveis a problemas reais, como a alocação ineficiente de supervisores.

III. METODOLOGIA

3.1. Tipo de Investigação e Procedimentos

Neste projecto, foi utilizada uma pesquisa aplicada com abordagem metodológica mista.

- i) **Componente qualitativa (DSR):** Focado no Estudo de Caso (ISUTC) e na Análise Documental/Entrevistas (colecta de dados qualitativos sobre o fluxo de trabalho e expectativas dos *stakeholders*) para garantir que o artefacto fosse útil e utilizável. A Figura 1 ilustra o ciclo iterativo do *DSR*.
- ii) **Componente quantitativa (modelagem preditiva):** Focada no tratamento estatístico dos dados de texto (resumos e perfis) para construir e validar o algoritmo.

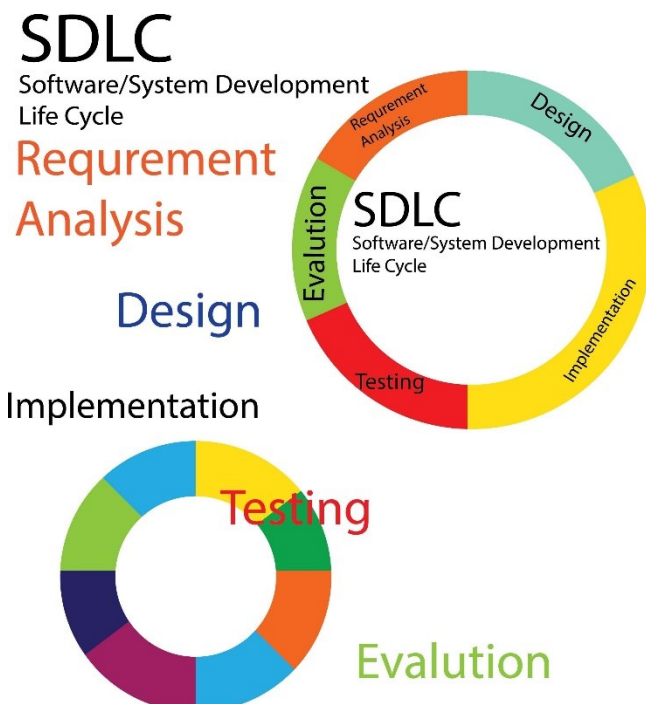


Figura 1. Ciclo de implementação da metodologia *DSR*. (Fonte: [Sdcl System Development Life Cycle: vector de stock \(libre de regalías\) 334217498 | Shutterstock](#))

3.2. Procedimentos e Tratamento Estatístico dos Dados

- **Análise Documental:** Identificação de 150 perfis de professores e 500 resumos de projectos de PFC para criar a base de dados de perfis de pesquisa.
- **Vectorização (tratamento estatístico):** Os resumos foram submetidos ao algoritmo *TF-IDF* [1], que gerou um vector de pesos para cada documento (aluno e professor). Esta é a principal etapa de tratamento estatístico quantitativo de dados não estruturados, transformando o texto em uma representação numérica.
- **Similaridade (tratamento estatístico):** Para medir a compatibilidade temática (a relação estatística entre vetores), foi empregado o algoritmo de Similaridade de Cosseno (*Cosine Similarity*). Esta métrica calcula o ângulo entre os vectores *TF-IDF* dos perfis de aluno e professor, resultando em um escore de 0 a 1, onde 1 indica similaridade temática perfeita.
- **Validação:** A performance do sistema foi avaliada usando a métrica de Acurácia Top-N sobre o conjunto de testes (100 resumos), garantindo a validade predictiva do modelo (Gao et al., 2020).

3.3. Metodologia de Desenvolvimento (*Design Science Research – DSR*)

O processo de desenvolvimento seguiu o ciclo do *DSR*. A Figura 1 representa o ciclo de implementação da metodologia.

IV. RESULTADOS E AVALIAÇÃO DO ARTEFACTO

O projecto resultou no desenvolvimento de um protótipo funcional do aplicativo de recomendação. O modelo se mostrou capaz de analisar as áreas de interesse dos alunos e compará-las com os perfis de pesquisa dos professores, gerando uma lista de supervisores ordenada por grau de compatibilidade.

4.1. Arquitetura e Fluxo de Processamento

A solução foi implementada com base numa arquitetura de múltiplos níveis (*N-Tier*):

- **Camada cliente (Front-end):** Interface web em *React* para recolha de dados e exibição das recomendações (Figura 2).
- **Camada de aplicação (Back-end):** Servidor *Flask* em *Python*, responsável pelo algoritmo de NLP, cálculo da similaridade, e exposição da *API RESTful*.
- **Camada de dados (Database):** Base de dados *PostgreSQL* (*Supabase*) que armazena os perfis de

supervisores e dados históricos.

O diagrama do modelo de recomendação (Figura 2) detalha o fluxo de processamento: (Segue a descrição das etapas do diagrama: *Entrada de dados* → *Pré-processamento* → *Vectorização* → *Cálculo de similaridade* → *Recomendação*).



Figura 2: Interface do modelo de recomendação.

4.2. Conjunto de Dados e Eficiência Operacional

A base de dados foi construída com um total de 150 perfis de professores e 500 resumos de projectos de alunos, garantindo uma base robusta para treinamento e validação. A Tabela 1 representa o conjunto de dados.

Tabela 1: Conjunto de dados usados no treino e no teste do modelo

Conjunto	Número de perfis de professores	Número de resumos de alunos
Treino	100	400
Teste	50	100

O principal ganho do artefacto reside na eficiência operacional. O processo manual de alocação, que envolve a leitura demorada dos resumos pela coordenação, exige várias horas de trabalho. O algoritmo de IA processa 100 resumos e gera o *ranking* de compatibilidade em menos de 5 minutos, resultando num aumento de eficiência de aproximadamente 95% na etapa de triagem inicial.

4.3. Avaliação de Desempenho e Acurácia

O algoritmo foi avaliado utilizando a acurácia *Top-N*, que mede a probabilidade do supervisor ideal estar entre as N primeiras recomendações do sistema.

Tabela 2: Métricas de Desempenho do Algoritmo de Similaridade

Métrica	Resultado	Significado
Acurácia Top 1	76%	Em 76% dos casos, o supervisor ideal foi o primeiro recomendado.
Acurácia Top 3	92%	Em 92% dos casos, o supervisor ideal estava entre os 3 primeiros.
Acurácia Top 5	98%	Em 98% dos casos, o supervisor ideal estava entre os 5 primeiros.

O desempenho do sistema de 92% de Acurácia Top-3 (Tabela 2) demonstra que o algoritmo de *TF-IDF/Similaridade de Cosseno* é altamente eficaz na identificação da compatibilidade temática entre aluno e professor.

V. CONCLUSÃO E TRABALHOS FUTUROS

O desenvolvimento de um aplicativo de recomendação de supervisores de PFC baseado em Processamento de Linguagem Natural (*NLP*) demonstrou ser uma abordagem viável e altamente eficaz para otimizar um processo académico crítico. Os resultados alcançados, com 92% de Acurácia Top-3, indicam que o algoritmo de *TF-IDF/Similaridade de Cosseno* pode fornecer recomendações precisas e personalizadas, alinhando os interesses dos alunos com a expertise dos professores.

A implementação deste sistema pode contribuir positivamente para a qualidade dos projetos académicos e a eficiência da gestão universitária, com uma redução de 95% no tempo de triagem, resolvendo um problema recorrente nas IES (Smith & Jones, 2023; Li, 2021).

O projecto abre portas para futuras pesquisas, como a incorporação de outros fatores na recomendação (ex., disponibilidade do professor, histórico de sucesso de supervisão) e o uso de modelos mais complexos (como *Deep Learning*), uma vez que dados históricos reais e abrangentes se tornarem disponíveis (Wilson, 2022).

REFERÊNCIAS

- Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of educational data mining*, 1(1), 3-17.
- Gao, F.; Wang, T.; Shen, J. (2020). Applications of machine learning in educational technology. *Interactive learning environments*, vol. 28, no. 1, pp. 1-19.
- Li, R. (2021). Mitigating bias in ai-based recruitment systems. *J. Artif. Intell. Ethics*, vol. 10, no. 1, pp. 45-61.
- Marx, K. A.; Müller, R. S. (2018). Design science research in information systems. *Mis quarterly*, vol. 30, no. 3, pp. 637-677.
- Smith, J. A. (2020) *Research methods in applied science: principles and practice*. Scientific publishing house.
- Smith, M. A.; Jones, D. M. (2023) The digital revolution in human resources: ai-driven recruitment and selection. *J. Modern business*, vol. 45, no. 2, pp. 112-125.
- Wilson, A. D. (2022). AI and the future of work: a global perspective on recruitment. *Global technology review*, vol. 18, no. 4, pp. 78-92.