

Grace: Sistema de Recomendação de Currículos com Inteligência Artificial*

Ana Clara B. Medeiros¹, João Fernando V. Franciscon¹,
Julia Brugnari P. Longo¹, Marcos Vinícius R. de Moraes¹,
Mariana de O. Macedo¹, Robson P. Bonidia¹

¹Grupo BioFatecou - Faculdade de Tecnologia de Ourinhos (Fatec)
Ourinhos - SP - Brasil

ana.medeiros40fatec.sp.gov.br, mariana.macedo@fatec.sp.gov.br

julia.longo@fatec.sp.gov.br, joao.franciscon01@fatec.sp.gov.br

marcos.moraes30@fatec.sp.gov.br, bonidia@usp.br

Resumo. *O recrutamento e alocação de funcionários são desafios constantes para as empresas. Diante disso, soluções inovadoras baseadas em Inteligência Artificial (IA) têm surgido para otimizar esses processos. No entanto, IAs podem seguir vieses sociais escondidos nos dados, levando a tomada de decisões injustas, danosas e/ou preconceituosas. Com base nisso, esta pesquisa desenvolveu um sistema de recomendação automatizado de candidatos, baseado em IA, chamado Grace, que visa remover o viés exclusivo de candidatos, como exclusão por gênero, cor, etnia, orientação sexual ou qualquer outro que interfira na escolha dos melhores currículos. Espera-se que essa aplicação resulte em uma seleção imparcial de candidatos, com maior precisão e redução significativa de vieses sociais.*

Abstract. *Recruitment and allocation of employees are constant challenges for companies. In light of this, innovative solutions based on Artificial Intelligence (AI) have emerged to optimize these processes. However, AIs can perpetuate hidden social biases within data, leading to unjust, harmful, and prejudiced decision-making. Based on this, this research has developed an automated candidate recommendation system, based on AI, called Grace, which aims to eliminate exclusive biases against candidates, such as gender, race, ethnicity, sexual orientation, or any other factor that might influence the selection of the best resumes. We hope that this application will result in an unbiased selection of candidates, with greater accuracy and a significant reduction in social biases.*

1. Introdução

O processo de recrutamento e seleção desempenha um papel vital no desenvolvimento dos negócios, levando as organizações a buscar departamentos de Recursos Humanos (RH) inteligentes que possam tomar decisões-chave de forma eficaz [Alsharef et al. 2023]. Durante o processo de contratação, é essencial analisar

* Acesso: <http://grace.fatecourinhos.edu.br/>
Orientador: Robson P. Bonidia - rpbonidia@gmail.com ou bonidia@usp.br

cuidadosamente cada informação presente nos currículos e avaliar o quão bem o documento se alinha com a vaga em questão [Vamsi Naidu et al. 2021], antes de prosseguir com as demais partes de um processo seletivo. Nessa análise, muitas empresas usam métodos manuais, levando uma quantidade expressiva de tempo para que os profissionais de RH escolham um grupo de candidatos para uma possível entrevista [Nurhopipah et al. 2021, Shovon et al. 2023].

De acordo com [Tiago Machado 2021], CEO da Rocketmat, inúmeras empresas visitadas apresentam uma grande dificuldade em conduzir processos seletivos, devido ao volume de currículos. Diante desse cenário, algumas empresas procuram maneiras de agilizar o processo, através da tecnologia e de algoritmos que realizem a filtragem das informações relevantes, por exemplo, abordagens de IA e aprendizado de máquina [Nurhopipah et al. 2021, Shovon et al. 2023].

O mercado de tecnologia oferece ferramentas capazes de auxiliar o setor de RH em seus processos seletivos, recomendando candidatos, conduzindo entrevistas por meio de salas de conversação e oferecendo outros serviços. No entanto, muitas dessas soluções são desenvolvidas no exterior e frequentemente não são adaptadas para atender às necessidades de pequenas ou médias empresas, que são o foco do nosso projeto. Além disso, essas soluções muitas vezes têm um custo elevado por vaga anunciada. Um ponto adicional de preocupação é que, com a disseminação generalizada e a adoção da IA, surge uma crescente inquietação de que esses sistemas possam perpetuar preconceitos de pessoas na organização [Mujtaba and Mahapatra 2019, Tilmes 2022].

Em busca de uma solução mais justa, ética e responsável, desenvolvemos um sistema de recomendação de currículos baseado em IA, chamado **Grace** (homenagem a Grace Murray Hopper), que visa analisar cada currículo e recomendar uma lista com os melhores candidatos para uma determinada vaga. O sistema **Grace** tem a finalidade de auxiliar as empresas na busca por melhores colaboradores, a fim de gerar mais assertividade e evitar vieses sociais que podem estar escondidos nos dados. Além dessas contribuições, **Grace** é disponibilizado sem custo para as empresas.

2. Trabalhos Relacionados

Nos estudos sobre recomendação de candidatos, algumas abordagens diferem das aplicadas pelo sistema **Grace**. Por exemplo, [Hong et al. 2013] propôs o agrupamento em *clusters* dos candidatos, utilizando parâmetros como atividade do usuário, coleta de informação (habilidades, experiências, etc.), frequência de cliques e frequência de buscas e comentários. [Kurdija et al. 2020a] e [Kurdija et al. 2020b] também aplicaram técnicas de agrupamento em conjunto das habilidades dos candidatos, utilizando representação vetorial da correlação entre candidatos e vagas. Outra aplicação foi o uso do modelo de árvore de decisão apresentado por [Gupta and Garg 2014], o qual classifica os candidatos de acordo com as características extraídas com o auxílio de métodos de Processamento de Linguagem Natural (PLN). Em contraste, nossa abordagem é baseada na extração de conteúdo dos currículos, evitando informações que podem conter vieses e a descrição da vaga. Além disso, **Grace** foi testado em cenários reais para avaliar suas decisões. Em relação às IAs responsáveis, várias iniciativas têm discutido propostas para lidar com esse problema, abordando questões éticas, de justiça, confiabilidade, sustentabilidade, transparência e reprodutibilidade.

Contudo, a maioria dos estudos é de natureza teórica, carecendo de experimentos, e frequentemente não inclui implementação prática e validação junto a empresas e/ou candidatos.

3. Desenvolvimento

Como mencionado anteriormente, para o desenvolvimento do sistema **Grace**, foram empregadas técnicas de sistemas de recomendação. Essas técnicas recebem as informações principais de uma vaga específica e percorrem todos os currículos enviados em busca das características desejadas. Ao conhecer as características dos candidatos que se deseja encontrar, é possível filtrar dados irrelevantes e tomar decisões mais assertivas. A Figura 1 apresenta o fluxo de trabalho do **Grace**.

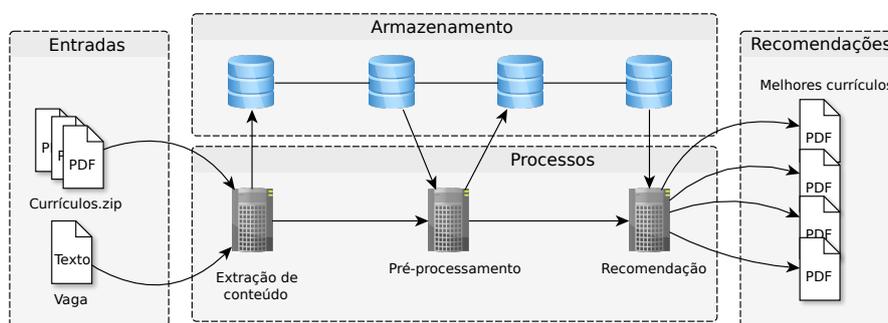


Figura 1. Fluxo de Trabalho do Sistema Grace

O sistema proposto basicamente recebe uma coleção de currículos, que podem estar separados ou compactados em um arquivo *zip*. Esses currículos podem conter dados estruturados ou não. Além disso, a descrição da vaga desejada é inserida na interface do sistema pelo recrutador. A descrição de um cargo tem como foco apresentar uma oportunidade em aberto, visando encontrar profissionais com as características compatíveis com o descrito para o processo seletivo. Tendo em vista o tamanho do mercado, atualmente para encontrar a pessoa ideal para determinado cargo, é necessário descrever com clareza quais serão as atividades propostas e as habilidades interpessoais desejadas. O primeiro passo envolve a extração do conteúdo desses documentos, seguida pelo pré-processamento dos textos utilizando bibliotecas em Python. Esse pré-processamento é crucial para otimizar o desempenho do sistema. Durante todo o processo, os dados intermediários são armazenados, permitindo sua recuperação para auxiliar na tomada de decisão. Por fim, o sistema realiza a recomendação utilizando três métricas diferentes: Contagem simples de palavras, Distância de Levenshtein e Similaridade por cosseno. O resultado é a lista dos melhores currículos para a vaga em questão. Os detalhes de todos esses processos estão apresentados na Figura 1 e serão explorados nas próximas seções.

3.1. Sistema de Recomendação Baseado em Conteúdo

Os sistemas de recomendação utilizam-se dos dados coletados dos usuários para sugerir novos produtos, auxiliando na tomada de decisão e, simultaneamente, a lidar com a sobrecarga de informação. Para implementar o sistema **Grace**, foi utilizada a filtragem baseada em conteúdo, na qual a similaridade de um item recomendado será

medida pela similaridade com as propriedades do item que o usuário tenha adquirido ou pesquisado anteriormente [Isinkaye et al. 2015]. No caso deste projeto, em que o foco é nas propriedades dos candidatos, o sistema de recomendação avaliará as necessidades do departamento de Recrutamento e Seleção da empresa, sugerindo o currículo com maior afinidade de conteúdo para atender tais requisitos.

3.2. Extração de Texto

O sistema **Grace** recebe dados em PDF, DOCX, arquivos compactados e descrições em texto, que muitas vezes são dados não estruturados. Para uma melhor compreensão e análise deles é necessário transformá-los em dados estruturados, para assim realizar a extração do conteúdo, e então dar início ao tratamento dos valores introduzidos. Para isso, foi utilizada a biblioteca *Apache Tika* [Tika 2022], escrita em Java, responsável por retirar o conteúdo (metadados e textos) de diversas entradas de arquivos. A aplicação da biblioteca *Apache Tika* facilitou o manuseio dos currículos para serem então aplicadas as demais técnicas apresentadas nas seções posteriores.

3.3. Processamento de Linguagem Natural

O PLN é usado para compreender a linguagem do ser humano [Meera and Geerthik 2022], e neste trabalho foi aplicada para auxiliar o algoritmo de recomendação a compreender o conteúdo dos currículos para então recomendar os candidatos mais adequados. PLN capacita os dispositivos a responderem a busca de recrutadores, fornecendo respostas por alternativas que fazem sentido com as iterações. O sistema **Grace** utiliza diversas técnicas de tratamento dos dados, tais como padronização dos caracteres para minúsculo que visa facilitar a comparação entre as palavras dos currículos com as das vagas, retirada dos espaços em branco, quebra de linha, pontuação, *stopwords*, remoção de numerais, símbolos e aplicação de corretores automáticos.

Nessa fase também visamos remover viés exclusivo de candidatos, como exclusão por gênero, cor, etnia, orientação sexual ou qualquer outro que interfira na escolha dos melhores currículos. Em seguida, foi executada a lematização [Akhmetov et al. 2020], um método que não desconsidera o significado da palavra, mas altera os verbos para o presente e unifica os sinônimos [Barbosa et al. 2017]. Ao fim, foi aplicado um vetorizador que converte as palavras em números, aumentando o grau da análise. O sistema proposto utiliza o vetorizador *Term Frequency–Inverse Document Frequency* (TF-IDF), o qual atribui um peso a cada palavra conforme sua frequência e importância no texto [Rangel 2019]. Para construir as recomendações do sistema, três métodos de similaridade entre a vaga e a massa de currículos submetida foram implementados, baseado em bons resultados da literatura [Shovon et al. 2023, Alsharif et al. 2023], entre eles: Contagem simples de palavras, Distância Levenshtein e Similaridade por Cosseno.

4. Validação e Resultados

O sistema **Grace** tem o objetivo de automatizar tarefas de recomendação de currículos com uma interface simples e intuitiva, simplificando o acesso a funções com o menor número de cliques possível. Para melhor demonstração, a Figura 2 apresenta a tela de recomendação dos currículos. O sistema busca fazer sua seleção

sem viés ou interferências externas, proporcionando uma escolha mais assertiva do candidato. Além disso, o **Grace** passou por um processo de validação com vagas de empresas parceiras para avaliar as sugestões recomendadas pelo sistema, verificando a compatibilidade com as decisões reais, principalmente para garantir que o algoritmo seja responsável em suas escolhas.

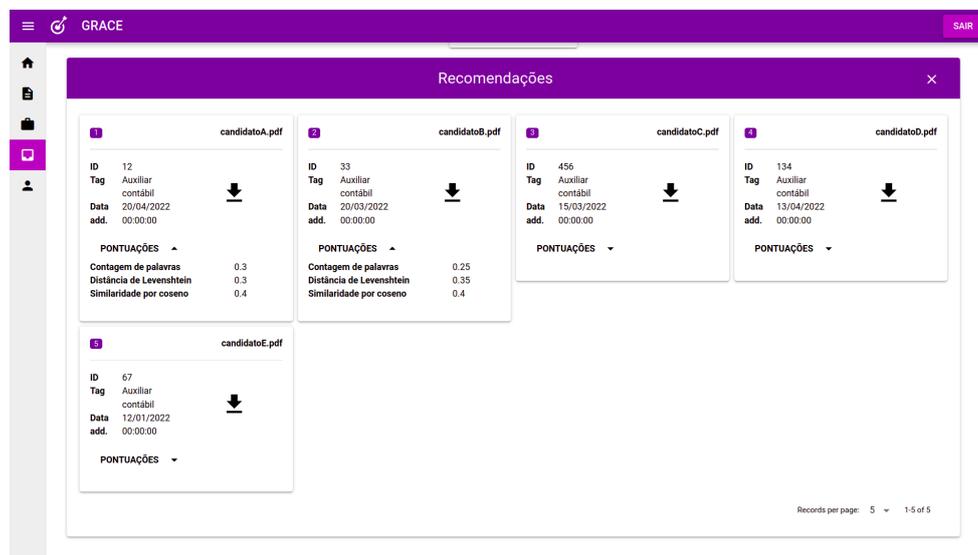


Figura 2. Screenshot de recomendação dos candidatos.

Para validar as recomendações do sistema, três estudos de casos reais foram realizados em parceria com duas empresas, uma do segmento de vale-alimentação e outra do segmento de energia renovável. Basicamente, as empresas forneceram currículos de processos seletivos de cargos já ocupados na empresa e a descrição da vaga divulgada. O objetivo principal do experimento era verificar se o candidato contratado pela organização foi apresentado entre os melhores currículos recomendados no **Grace**. O experimento seguiu todas as normas estabelecidas pela Lei Geral de Proteção de Dados Pessoais (LGPD).

As duas empresas forneceram três processos seletivos: (1) Analista de Gestão de Projetos, (2) Desenvolvedor de Python Júnior e (3) Executivo Comercial. Para a vaga de analista de gestão de projetos, o sistema **Grace** recomendou 10 candidatos, no qual o candidato contratado pela empresa ocupou a sétima posição entre as recomendações. Na vaga de desenvolvedor de Python júnior foram analisados 11 currículos, e a empresa neste caso selecionou dois candidatos para ocuparem a vaga. Ambos os contratados estavam entre os recomendados pelo sistema, um deles ocupou a primeira posição do ranking, sendo totalmente assertivo e o outro, a décima posição.

Para a vaga de executivo comercial, com 53 currículos encaminhados, dos quais foram selecionados para a entrevista 9 candidatos, e dois foram aprovados. Na recomendação realizada pelo sistema **Grace**, apenas o candidato da oitava colocação foi selecionado para a entrevista, os demais entrevistados ficaram espalhados no ranking. Já os dois aprovados pela empresa ocuparam a décima segunda e vigésima primeira posição. Nesse último experimento, percebeu-se que a influência para respostas tão divergentes foi a descrição da vaga, pois estava muito abrangente e

com poucas palavras-chave, o que dificultou o trabalho do algoritmo.

No entanto, é importante ressaltar que o sistema **Grace** foi construído para auxiliar na filtragem dos melhores currículos e não indicar qual contratação deveria ser feita, ou seja, é uma ferramenta para recomendar uma número x (definido pelo usuário) de currículos para análise. A posição dos candidatos não deveria influenciar na avaliação. Como pode ser observado nos experimentos, os candidatos contratados estavam na filtragem do sistema **Grace**, mas em posições distintas em todas as vagas testadas. Assim, vale reforçar que os possíveis usuários devem considerar a filtragem, avaliando todos os currículos recomendados e não enviesar sua análise pela posição da recomendação. Finalmente, durante a validação do sistema, incorporou-se um mapeamento de seu comportamento, buscando um algoritmo que, por meio de mecanismos de representação adequados, atenda às complexidades sociais e contribua para um mercado de trabalho mais justo.

5. Conclusão

O sistema proposto visa ter um papel social quando aplicado ao recrutamento de pessoas, tornando os processos mais assertivos, ao avaliar a qualidade do conteúdo dos currículos apresentados, buscando evitar vieses exclusivos de candidatos, como exclusão por gênero, cor, etnia, orientação sexual ou qualquer outro que interfira na escolha dos melhores currículos. Em conclusão aos resultados, o sistema **Grace** performou como esperado, recomendando, em quase todos os estudos de caso, o candidato contratado entre os melhores. Embora em alguns casos o candidato contratado ocupe as últimas posições, a ideia não é fornecer o candidato mais indicado em primeiro lugar. Referente às limitações, percebe-se que um fator limitante do desempenho do sistema **Grace** é o tipo de descrição fornecida para a vaga analisada. Vagas com mais requisitos técnicos, comuns em vagas para desenvolvedores, profissionais contábeis, analistas, entre outras, tendem a possuir um desempenho superior a vagas mais abstratas ou muito amplas. Por isso, é recomendado o uso de palavras-chave na descrição das vagas, em vez de uma descrição muito abrangente, pois elas representam um formato mais amigável para os algoritmos utilizados, que buscam uma similaridade entre os conteúdos. Levando isso em consideração, as empresas poderiam usar o **Grace** para verificar a precisão da descrição de futuras vagas, auxiliando na definição dos melhores requisitos técnicos para a posição.

Referências

- Akhmetov, I., Pak, A., Ualiyeva, I., and Gelbukh, A. (2020). Highly language-independent word lemmatization using a machine-learning classifier. *Computación y Sistemas*, 24(3):1353–1364.
- Alsharif, A., Nassour, H., Sharma, J., et al. (2023). Exploring the efficiency of text-similarity measures in automated resume screening for recruitment. In *2023 10th International Conference on Computing for Sustainable Global Development (INDIACom)*, pages 36–42. IEEE.
- Barbosa, J., Vieira, J. P. A., Santos, R., Junior, G. V. M., MUNIZ, M. d. S., and Moura, R. S. (2017). Introdução ao processamento de linguagem natural usando python. *III Escola Regional de Informatica do Piauí*, 1:336–360.

- Gupta, A. and Garg, D. (2014). Applying data mining techniques in job recommender system for considering candidate job preferences. pages 1458–1465.
- Hong, W., Zheng, S., Wang, H., and Shi, J. (2013). A job recommender system based on user clustering. *J. Comput.*, 8(8):1960–1967.
- Isinkaye, F. O., Folaajimi, Y. O., and Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian informatics journal*, 16(3):261–273.
- Kurdija, A. S., Afric, P., Sikic, L., Plejic, B., Silic, M., Delac, G., Vladimir, K., and Srbljic, S. (2020a). Building vector representations for candidates and projects in a cv recommender system. In *Artificial Intelligence and Mobile Services–AIMS 2020: 9th International Conference, Held as Part of the Services Conference Federation, SCF 2020, Honolulu, HI, USA, September 18-20, 2020, Proceedings 9*, pages 17–29. Springer.
- Kurdija, A. S., Afric, P., Sikic, L., Plejic, B., Silic, M., Delac, G., Vladimir, K., and Srbljic, S. (2020b). Candidate classification and skill recommendation in a cv recommender system. In *International Conference on AI and Mobile Services*, pages 30–44. Springer.
- Meera, S. and Geerthik, S. (2022). Natural language processing. *Artificial Intelligent Techniques for Wireless Communication and Networking*, pages 139–153.
- Mujtaba, D. F. and Mahapatra, N. R. (2019). Ethical considerations in ai-based recruitment. In *2019 IEEE International Symposium on Technology and Society (ISTAS)*, pages 1–7. IEEE.
- Nurhopipah, A., Ceasar, Y., and Priadana, A. (2021). Improving machine learning accuracy using data augmentation in recruitment recommendation process. In *2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT)*, pages 203–208. IEEE.
- Rangel, M. d. M. (2019). Categorização automática de conjuntos de dados de portais de dados abertos utilizando aprendizado de máquina supervisionado.
- Shovon, S. S. F., Mohsin, M. M. A. B., Tama, K. T. J., Ferdaous, J., and Momen, S. (2023). Cvr: An automated cv recommender system using machine learning techniques. In *Data Science and Algorithms in Systems: Proceedings of 6th Computational Methods in Systems and Software 2022, Vol. 2*, pages 312–325. Springer.
- Tiago Machado, C. d. R. (2021). Papel social da inteligência artificial é tornar visíveis os profissionais invisíveis para o mercado.
- Tika, A. (2022). Apache tika - a content analysis toolkit.
- Tilmes, N. (2022). Disability, fairness, and algorithmic bias in ai recruitment. *Ethics and Information Technology*, 24(2):21.
- Vamsi Naidu, P., Bommu, V. M. R., Pallapothu, V. D., Janapamula, S. R. N., and Kommuri, N. L. (2021). Resume screening using machine learning. In *International Conference on Image Processing and Capsule Networks*, pages 745–751. Springer.