Explicações de Inteligência Artificial Conscientes de Contexto em Aplicações de Mercado de Trabalho

Gabriel B. S. Pinto[1], **Carlos E. R. de Mello** [1], **Ana Cristina B. Garcia**[1]

[1]Programa de Pós-Graduação em Informática (PPGI)
Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
CEP 2290-255 – Av. Pasteur, 458, Urca, RJ – Rio de Janeiro – RJ – Brasil

gabrielbspinto@edu.unirio.br, {mello,cristina.bicharra}@uniriotec.br

Abstract. *This study design develops a context-aware explainability model to enhance user satisfaction, trust, and comprehension in AI-driven labor market applications. Addressing the limitations of SHAP and LIME, it integrates domain-specific variables to reduce informational asymmetries. Using a Design Science Research (DSR) approach, the study consists of two sequential experiments and the findings will provide empirical evidence on contextualization's role in AI explanations, improving transparency and usability.*

Resumo. *Este projeto de pesquisa desenvolve um modelo de explicabilidade sensível ao contexto para aprimorar a satisfação, confiança e compreensão dos usuários em aplicações de IA no mercado de trabalho. Para superar as limitações do SHAP e LIME, o modelo integra variáveis específicas do domínio, reduzindo assimetrias informacionais. Seguindo a abordagem de Design Science Research (DSR), a pesquisa consiste em dois experimentos sequenciais, cujos resultados fornecerão evidências empíricas sobre o impacto da contextualização na explicabilidade da IA, promovendo transparência e usabilidade.*

1. Introdução

A Inteligência Artificial Explicável (XAI) tem sido amplamente estudada para tornar modelos de aprendizado de máquina mais interpretáveis, permitindo que usuários compreendam as decisões automatizadas. No mercado de trabalho, os principais esforços para tornar a IA explicável concentram-se em três processos essenciais: recrutamento, recomendação de habilidades e avaliação de desempenho. Todos atuando em processos da empresa pela área de recursos humanos. No recrutamento, os modelos são empregados para análise de perfis e entrevistas, previsão de compatibilidade entre candidatos e vagas, processamento de currículos, detecção de fraudes e previsão da demanda por contratações [Bhattacharya et al. 2022, Sogancioglu et al. 2023, Marra and Kubiak 2024, Pessach et al. 2020, Peña et al. 2023, Naudé et al. 2023, Zhang et al. 2021].

Alguns métodos de XAI, como SHAP, que quantifica a contribuição de cada variável para as previsões, e LIME, que explica decisões localmente ao modificar entradas do modelo, são amplamente utilizados [Lundberg 2017, Ribeiro et al. 2016]. Além disso, explicações contrafactuais demonstram como alterações nos dados poderiam modificar os resultados, enquanto modelos intrinsecamente interpretáveis, como regressão linear e árvores de decisão, expõem diretamente a influência das variáveis [Arrieta et al. 2020].

Frameworks de XAI combinam essas abordagens para aprimorar a transparência e fortalecer a confiança dos usuários [Linardatos et al. 2020]. No entanto, uma limitação fundamental dessas abordagens é que elas não consideram suficientemente o contexto específico dos usuários e a natureza do processo decisório no ambiente de trabalho, o que reduz a efetividade das explicações [Daudt et al. 2021].

A explicabilidade eficaz de IA no mercado de trabalho deve ir além da interpretabilidade técnica e incorporar três pilares essenciais: negociação, mitigação de vieses e justiça. A negociação, quando apoiada por explicações com contexto personalizado, permite que candidatos e funcionários compreendam como características específicas de seu perfil — como escolaridade, setor, tempo de experiência ou localização — influenciam as decisões algorítmicas, reduzindo assimetrias informacionais e favorecendo interações mais transparentes [Das Swain et al. 2023]. Quando esse processo é também reflexivo, as explicações reconhecem distorções históricas e contextos sociais desiguais que afetam determinados grupos, permitindo que o indivíduo questione e negocie decisões de forma mais informada [Lashkari and Cheng 2023]. Já a justiça requer abordagens que levem em conta o contexto organizacional e social para evitar a perpetuação de desigualdades estruturais [Sánchez-Monedero et al. 2020]. Sem essas considerações, os sistemas de IA podem falhar em gerar explicações úteis e confiáveis, prejudicando sua aceitação e impacto positivo no mercado de trabalho.

1.1. Problema de Pesquisa

Apesar dos avanços da XAI no mercado de trabalho, os métodos explicativos tradicionais não contextualizam adequadamente as explicações, limitando a compreensão, satisfação e confiança dos usuários nesses sistemas automatizados.

Essa limitação se torna ainda mais relevante quando se considera que o mercado de trabalho é caracterizado por assimetrias de informação, onde candidatos e funcionários frequentemente não possuem o mesmo nível de acesso e entendimento dos processos decisórios algorítmicos que empregadores e recrutadores. A inclusão de contexto nas explicações pode, portanto, potencializar a adoção eficaz dos sistemas de inteligência artificial, colocando os usuários no centro do desenvolvimento dos métodos explicativos. Embora a XAI tenha avançado significativamente em diversos setores, ainda há grandes oportunidades para seu aprimoramento no domínio do mercado de trabalho. Explorar as percepções dos usuários por meio da avaliação de satisfação, confiança e compreensão – medidas essenciais para validar as explicações fornecidas pelos sistemas – pode favorecer uma adoção mais ampla e eficaz das aplicações de IA nesse contexto.

2. Trabalhos Relacionados

Na revisão sistemática realizada para fundamentar a pesquisa [Pinto et al.], ao mapear os métodos de explicabilidade utilizados no mercado de trabalho e ao analisar os requisitos sociotécnicos destacados pelos usuários, foram identificados 17 métodos de explicabilidade, com destaque para LIME e SHAP. A revisão foi conduzida em bases de dados da ciência da computação, resultando na identificação de 266 estudos elegíveis, dos quais 29 artigos foram selecionados para análise detalhada sobre explicabilidade e justiça em sistemas de IA aplicados ao mercado de trabalho.

O SHAP foi empregado em seis estudos, sendo dois voltados à avaliação de perfis de candidatos [Bhattacharya et al. 2022, Sogancioglu et al. 2023], três para predição de

rotatividade [Das et al. 2022, Makanga et al. 2024, Sekaran and Shanmugam 2022] e um para avaliação de desempenho, combinado com LIME [Sampath et al. 2024]. Já LIME foi utilizado exclusivamente para recomendação de habilidades [Akkasi 2024].

Outras abordagens também foram identificadas, como Variable-Order Bayesian Network (VOBN) para recrutamento [Pessach et al. 2020], raciocínio contrafactual para recomendação de habilidades [Tran et al. 2021], e um modelo log-linear estilizado para avaliações de funcionários [Maurya et al. 2018]. Técnicas baseadas na importância das variáveis foram aplicadas para detecção de fraudes [Naudé et al. 2023] e mitigação de viés de gênero [Marra and Kubiak 2024].

Entre os métodos mais avançados, destacam-se o Self-Explaining Skill Recommendation Framework, voltado para a modelagem de habilidades [Sun et al. 2024], e a Talent Demand Attention Network (TDAN), utilizada para previsão da demanda por talentos [Zhang et al. 2021].

Os achados dessa revisão evidenciam a variedade de abordagens empregadas para tornar a IA explicável no mercado de trabalho, destacando tanto métodos consolidados, como LIME e SHAP, quanto modelos emergentes que ampliam a transparência e confiabilidade das explicações. Esses resultados reforçam a necessidade de contextualização das explicações para aprimorar a aceitação e usabilidade dos sistemas de IA nesse domínio.

3. Procedimentos Metodológicos

A crescente adoção de sistemas de inteligência artificial no mercado de trabalho exige explicações mais compreensíveis e alinhadas ao contexto específico dos usuários. No entanto, os métodos tradicionais de explicabilidade, como SHAP e LIME, não consideram variáveis contextuais que podem ser essenciais para garantir maior satisfação, confiança e compreensão dos usuários. Para abordar essa lacuna, esta pesquisa propõe o desenvolvimento de um método de explicabilidade descrito nas seções a seguir.

3.1. Caracterização da Pesquisa

A hipótese central deste estudo estabelece que a inclusão de variáveis contextuais nas explicações fornecidas por sistemas de IA no mercado de trabalho melhora a satisfação, a confiança e a compreensão dos usuários. Como ponto de partida, parte-se do pressuposto de que a eficácia das explicações deve ser avaliada sob a perspectiva do usuário, e que a ausência de variáveis do domínio nas bases de dados pode limitar a qualidade das explicações geradas. Assim, a integração dessas variáveis representa uma oportunidade para aprimorar a transparência e a interpretabilidade dos sistemas de IA, garantindo que as decisões automatizadas sejam mais compreensíveis e alinhadas às necessidades dos usuários.

O modelo proposto consiste em um artefato conceitual que combina abordagens tradicionais de explicabilidade com variáveis contextuais derivadas do conhecimento do domínio. A estrutura do modelo inclui etapas para identificar e extrair variáveis que não estão presentes nas bases de dados, mas que são essenciais para a interpretação dos resultados. Essas variáveis são então integradas aos métodos tradicionais, permitindo a geração de explicações mais robustas, personalizadas e contextualizadas. Com essa abordagem, busca-se promover um modelo holístico e centrado no usuário, que alia técnicas já consolidadas a informações qualitativas específicas do contexto laboral.

A literatura ainda apresenta uma lacuna significativa na explicabilidade da IA aplicada ao mercado de trabalho, uma vez que não foram identificados estudos que explorem a integração de variáveis contextuais no aprimoramento da satisfação, confiança e compreensão dos usuários. Assim, esta pesquisa contribui para o avanço da área ao propor um framework inovador para a explicabilidade contextualizada, com potencial para ser aplicado em diversas áreas de gestão de talentos e recursos humanos.

A pesquisa adota um paradigma epistemiológico baseado no Design Science Research (DSR), com abordagem empírica, mista e explanatória. O método de pesquisa é experimental, permitindo a avaliação iterativa do modelo proposto a partir da interação dos usuários com o artefato.

3.2. Avaliação

Para validar o modelo proposto, a pesquisa será conduzida por meio de dois experimentos sequenciais, fundamentados nos princípios do Design Science Research (DSR) [Peffers et al. 2007]. Esses experimentos permitirão a avaliação iterativa do artefato, possibilitando sua adaptação com base no feedback dos usuários.

No primeiro experimento, os participantes interagirão com a versão inicial do sistema e responderão a três questionários para mensurar satisfação, confiança e compreensão em relação à explicação fornecida. Em seguida, será conduzido um grupo focal para identificar quais variáveis contextuais podem ser incorporadas ao modelo para aprimorar a qualidade das explicações.

No segundo experimento, os participantes testarão a versão aprimorada do artefato, agora incorporando as variáveis identificadas no primeiro grupo focal. Os usuários responderão novamente aos três questionários, permitindo uma comparação quantitativa entre os dois momentos. Além disso, um segundo grupo focal será realizado para analisar qualitativamente o impacto das variáveis contextuais na percepção dos usuários.

Essa abordagem, que combina métodos quantitativos e qualitativos, possibilita uma avaliação robusta e iterativa, assegurando que o artefato seja continuamente refinado para maximizar sua eficácia e aceitação pelos usuários.

A análise dos resultados será conduzida por meio de métodos estatísticos e qualitativos. Os dados quantitativos dos questionários serão analisados estatisticamente para verificar diferenças significativas entre os dois experimentos, utilizando técnicas como análise de variância (ANOVA) [Gelman 2005] e testes de hipóteses apropriados. Paralelamente, os dados qualitativos dos grupos focais serão submetidos a uma análise de conteúdo, permitindo a identificação de padrões e insights sobre a percepção dos usuários em relação à explicabilidade do sistema.

4. Considerações Finais

Os achados desta pesquisa fornecerão evidências empíricas sobre a importância do contexto na explicabilidade de sistemas de IA, destacando sua influência na transparência, confiança e usabilidade. Além disso, os resultados contribuirão para o desenvolvimento de frameworks mais avançados de explicabilidade, servindo como referência para futuras pesquisas e aplicações no mercado de trabalho e em ambientes organizacionais.

Referências

Akkasi, A. (2024). Job description parsing with explainable transformer based ensemble models to extract the technical and non-technical skills. *Natural Language Processing Journal*, 9:100102.

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., et al. (2020). Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information fusion*, 58:82–115.

Bhattacharya, P., Zuhair, M., Roy, D., Prasad, V. K., and Savaliya, D. (2022). Aajeevika: Trusted explainable ai based recruitment scheme in smart organizations. In *2022 5th International Conference on Contemporary Computing and Informatics (IC3I)*, pages 1002–1008. IEEE.

Das, S., Chakraborty, S., Sajjan, G., Majumder, S., Dey, N., and Tavares, J. M. R. (2022). Explainable ai for predictive analytics on employee attrition. In *International Conference on Soft Computing and its Engineering Applications*, pages 147–157. Springer.

Das Swain, V., Gao, L., Wood, W. A., Matli, S. C., Abowd, G. D., and De Choudhury, M. (2023). Algorithmic power or punishment: Information worker perspectives on passive sensing enabled ai phenotyping of performance and wellbeing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–17.

Daudt, F., Cinalli, D., and Garcia, A. C. B. (2021). Research on explainable artificial intelligence techniques: An user perspective. In *2021 IEEE 24th international conference on computer supported cooperative work in design (CSCWD)*, pages 144–149. IEEE.

Gelman, A. (2005). Analysis of variance—why it is more important than ever.

Lashkari, M. and Cheng, J. (2023). "finding the magic sauce": Exploring perspectives of recruiters and job seekers on recruitment bias and automated tools. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–16.

Linardatos, P., Papastefanopoulos, V., and Kotsiantis, S. (2020). Explainable ai: A review of machine learning interpretability methods. *Entropy*, 23(1):18.

Lundberg, S. (2017). A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874*.

Makanga, C., Mukwaba, D., Agaba, C. L., Murindanyi, S., Joseph, T., Hellen, N., and Marvin, G. (2024). Explainable machine learning and graph neural network approaches for predicting employee attrition. In *Proceedings of the 2024 Sixteenth International Conference on Contemporary Computing*, pages 243–255.

Marra, T. and Kubiak, E. (2024). Addressing diversity in hiring procedures: a generative adversarial network approach. *AI and Ethics*, pages 1–25.

Maurya, A., Akoglu, L., Krishnan, R., and Bay, D. (2018). A lens into employee peer reviews via sentiment-aspect modeling. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 670–677. IEEE.

Naudé, M., Adebayo, K. J., and Nanda, R. (2023). A machine learning approach to detecting fraudulent job types. *AI & SOCIETY*, 38(2):1013–1024.

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, 24(3):45–77.

Peña, A., Serna, I., Morales, A., Fierrez, J., Ortega, A., Herrarte, A., Alcantara, M., and Ortega-Garcia, J. (2023). Human-centric multimodal machine learning: Recent advances and testbed on ai-based recruitment. *SN Computer Science*, 4(5):434.

Pessach, D., Singer, G., Avrahami, D., Ben-Gal, H. C., Shmueli, E., and Ben-Gal, I. (2020). Employees recruitment: A prescriptive analytics approach via machine learning and mathematical programming. *Decision Support Systems*, 134:113290.

Pinto, G. B. S., Mello, C. E., and Garcia, A. C. B. Explainable ai in labor market applications.

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). "why should i trust you?"explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144.

Sampath, K., Devi, K., Ambuli, T., and Venkatesan, S. (2024). Ai-powered employee performance evaluation systems in hr management. In *2024 7th International Conference on Circuit Power and Computing Technologies (ICCPCT)*, volume 1, pages 703–708. IEEE.

Sánchez-Monedero, J., Dencik, L., and Edwards, L. (2020). What does it mean to'solve'the problem of discrimination in hiring? social, technical and legal perspectives from the uk on automated hiring systems. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*, pages 458–468.

Sekaran, K. and Shanmugam, S. (2022). Interpreting the factors of employee attrition using explainable ai. In *2022 International Conference on Decision Aid Sciences and Applications (DASA)*, pages 932–936. IEEE.

Sogancioglu, G., Kaya, H., and Salah, A. A. (2023). Using explainability for bias mitigation: A case study for fair recruitment assessment. In *Proceedings of the 25th International Conference on Multimodal Interaction*, pages 631–639.

Sun, Y., Ji, Y., Zhu, H., Zhuang, F., He, Q., and Xiong, H. (2024). Market-aware long-term job skill recommendation with explainable deep reinforcement learning. *ACM Transactions on Information Systems*.

Tran, H. X., Le, T. D., Li, J., Liu, L., Liu, J., Zhao, Y., and Waters, T. (2021). Recommending the most effective intervention to improve employment for job seekers with disability. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 3616–3626.

Zhang, Q., Zhu, H., Sun, Y., Liu, H., Zhuang, F., and Xiong, H. (2021). Talent demand forecasting with attentive neural sequential model. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 3906–3916.