

Visualização Interativa para Modelos de Inteligência Artificial Explicáveis



Alice Pereira de Aguiar Penido,
André Luiz Baptista Esteves Bassini

Pontifical Catholic University of Minas Gerais – PUC Minas

October, 2024

- ▶ Objetivo Geral
- ▶ Fundamentação Teórica
- ▶ Trabalhos Relacionados
- ▶ Referências Bibliográficas

Desenvolver visualizações interativas que melhorem a interpretabilidade de modelos de IA explicáveis, permitindo uma exploração intuitiva de seus mecanismos e facilitando a compreensão dos processos decisórios, tornando-os mais acessíveis e transparentes.

Explainable AI (XAI)

- ▶ Um conjunto de técnicas que permite aos usuários humanos compreender e confiar em modelos mais explicáveis [3].

SHAP (SHapley Additive exPlanations)

- ▶ SHAP é uma abordagem baseada na teoria dos jogos que explica os resultados de modelos de machine learning [7].

Proveniência

- ▶ Proveniência refere-se ao registro que descreve a origem e o processamento dos dados [8].

Interação Humano-Computador

- ▶ A interação humano-computador combina IA e engajamento humano para criar sistemas inteligentes interativos [9].

[2] propõe:

“XAutoML, an interactive visual analytics tool for explaining arbitrary AutoML optimization procedures and ML pipelines constructed by AutoML. XAutoML combines interactive visualizations with established techniques from explainable artificial intelligence (XAI) to make the complete AutoML procedure transparent and explainable.”

[6], por outro lado, propõe um framework que

“combines an iterative XAI pipeline with eight global monitoring and steering mechanisms, including quality monitoring, provenance tracking, model comparison, and trust building.”

Referências Bibliográficas

- [1] OOGÉ, J.; VERBERT, K. Explaining Artificial Intelligence with Tailored Interactive Visualisations. In: COMPANION PROCEEDINGS OF THE 27TH INTERNATIONAL CONFERENCE ON INTELLIGENT USER INTERFACES (IUI '22 COMPANION), 2022, New York. Proceedings [...]. New York: ACM, 2022. p. 120–123. Disponível em: <https://doi.org/10.1145/3490100.3516481>.
- [2] ZÖLLER, M.-A.; TITOV, W.; SCHLEGEL, T.; HUBER, M. F. XAutoML: A Visual Analytics Tool for Understanding and Validating Automated Machine Learning. ACM Transactions on Interactive Intelligent Systems, v. 13, n. 4, Artigo 28, 39 p., Dez. 2023. Disponível em: <https://doi.org/10.1145/3625240>.
- [3] DWIVEDI, R. et al. Explainable AI (XAI): Core Ideas, Techniques, and Solutions. ACM Computing Surveys, v. 55, n. 9, Artigo 194, 33 p., Set. 2023. Disponível em: <https://doi.org/10.1145/3561048>.

Referências Bibliográficas

- [4] WANG, J.; LIU, S.; ZHANG, W. Visual Analytics for Machine Learning: A Data Perspective Survey. IEEE Transactions on Visualization and Computer Graphics, 2024. Disponível em: <https://doi.org/10.1109/TVCG.2024.3357065>.
- [5] CHOO, J.; LIU, S. Visual Analytics for Explainable Deep Learning. IEEE Computer Graphics and Applications, v. 38, n. 4, p. 84-92, Jul./Ago. 2018. Disponível em: <https://doi.org/10.1109/MCG.2018.042731661>.
- [6] SPINNER, T. et al. explAiner: A Visual Analytics Framework for Interactive and Explainable Machine Learning. IEEE Transactions on Visualization and Computer Graphics, 2019. Disponível em: <https://doi.org/10.1109/TVCG.2019.2934629>.
- [7] MARCÍLIO, W. E.; ELER, D. M. From explanations to feature selection: assessing SHAP values as feature selection mechanism. In: 33RD SIBGRAPI CONFERENCE ON GRAPHICS, PATTERNS AND IMAGES (SIBGRAPI), 2020, Porto de Galinhas. Proceedings [...]. Porto de Galinhas: IEEE, 2020. p. 340-347. Disponível em: <https://doi.org/10.1109/SIBGRAPI51738.2020.00053>.

Referências Bibliográficas

- [8] WERDER, K.; RAMESH, B.; ZHANG, R. Establishing Data Provenance for Responsible Artificial Intelligence Systems. ACM Transactions on Management Information Systems, v. 13, n. 2, Artigo 22, 23 p., Jun. 2022. Disponível em: <https://doi.org/10.1145/3503488>.

- [9] NAZAR, M.; ALAM, M. M.; YAFI, E.; SU'UD, M. M. A Systematic Review of Human-Computer Interaction and Explainable Artificial Intelligence in Healthcare With Artificial Intelligence Techniques. IEEE Access, v. 9, p. 153316-153348, 2021. Disponível em: <https://doi.org/10.1109/ACCESS.2021.3127881>.