PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS

Bacharelado em Ciência da Computação

Alice Pereira de Aguilar Penido,
André Luiz Baptista Esteves Bassini

Visualização Interativa para Modelos de Inteligência Artificial Explicáveis

Belo Horizonte

2024

Alice Pereira de Aguilar Penido,
André Luiz Baptista Esteves Bassini

Visualização Interativa para Modelos de Inteligência Artificial Explicáveis

Projeto de Pesquisa apresentado na disciplina Trabalho Interdisciplinar III - Pesquisa Aplicada do curso de Ciência da Computação da Pontifícia Universidade Católica de Minas Gerais.

Belo Horizonte

2024

RESUMO

Texto do resumo.

Palavras-chave: .

SUMÁRIO

1 INTRODUÇÃO

A crescente complexidade dos modelos de Inteligência Artificial (IA) e o aumento de seu uso em diversas áreas, como saúde, finanças e indústria, levantam a necessidade de tornar esses modelos mais compreensíveis para os usuários. A Explainable AI (XAI) busca resolver essa questão ao oferecer técnicas que tornam os modelos mais transparentes, permitindo que os usuários humanos compreendam e confiem nas decisões automatizadas. Entre essas técnicas, destaca-se o SHAP (SHapley Additive exPlanations), que utiliza a teoria dos jogos para explicar as previsões dos modelos, atribuindo a cada variável uma contribuição para o resultado final.

Entretanto, uma das principais limitações dessas abordagens é a dificuldade de interação e exploração dos dados pelos usuários, o que impede uma compreensão profunda dos processos decisórios subjacentes. Assim, a criação de plataformas de visualização interativas torna-se essencial, permitindo aos usuários não apenas observar, mas também interagir com os dados e os modelos, escolhendo variáveis e vendo como suas alterações afetam os resultados. Isso melhora a confiança nas decisões geradas pela IA e facilita a compreensão dos mecanismos dos modelos, tornando-os mais acessíveis e transparentes.

Este projeto visa desenvolver uma plataforma de visualização interativa para modelos de IA explicáveis, utilizando metodologias como o SHAP para ilustrar os impactos das variáveis nas previsões, permitindo uma exploração intuitiva e visual dos modelos e seus processos decisórios.

Este trabalho está organizado da seguinte forma.

1.1 Objetivos

O objetivo geral deste projeto é desenvolver uma plataforma de visualização interativa que permita a explicação de modelos de Inteligência Artificial, utilizando técnicas como SHAP para melhorar a interpretabilidade e acessibilidade dos modelos, possibilitando que usuários explorem e compreendam os impactos das variáveis nos resultados.

1.1.1 Objetivos específicos

Os objetivos específicos deste projeto são:

1. Implementar visualizações gráficas interativas que permitam a exploração dos impactos das variáveis nos modelos de IA.
2. Integrar a metodologia SHAP à plataforma para explicações baseadas em teoria dos jogos.
3. Facilitar a seleção e manipulação de dados pelos usuários na plataforma.
4. Avaliar a usabilidade da plataforma por meio de testes com usuários de diferentes áreas.
5. Validar a plataforma utilizando modelos de IA aplicados a diferentes setores, como saúde e finanças.

2 REVISÃO BIBLIOGRÁFICA

Este capítulo apresenta definições necessárias para compreender o tema do trabalho, bem como a solução proposta.

2.1 Inteligência Artificial (IA)

Campo da ciência da computação que trata da criação de sistemas ou máquinas capazes de realizar tarefas que normalmente exigem inteligência humana (RUSSELL; NORVIG, 2016). Esses sistemas são projetados para simular capacidades cognitivas como aprendizagem, raciocínio, tomada de decisão, reconhecimento de padrões e resolução de problemas. A IA inclui subcampos como aprendizado de máquina (machine learning), processamento de linguagem natural (NLP), visão computacional e robótica.

2.2 Explainable AI (XAI)

A *Explainable AI* refere-se a métodos que tornam os modelos de aprendizado de máquina mais interpretáveis, tanto para desenvolvedores quanto para usuários finais. Técnicas como *LIME* fornecem explicações locais, explicando previsões individuais ao destacar a influência de variáveis específicas (DWIVEDI et al., 2023). Por outro lado, *SHAP* fornece uma abordagem consistente baseada na teoria dos jogos, atribuindo valores Shapley a cada variável e medindo sua contribuição para a previsão final (MARCíLIO; ELER, 2020).

2.3 LIME

O *LIME* (Local Interpretable Model-agnostic Explanations) é uma técnica de explicabilidade de IA que visa interpretar modelos complexos (conhecidos como "caixa-preta"). Ele gera explicações locais, ou seja, explica previsões individuais ao aproximar o comportamento do modelo com um modelo mais simples (como uma regressão linear) em torno de um ponto específico. Isso ajuda a entender como cada variável influencia uma previsão específica, sem depender da arquitetura do modelo.

2.4 SHAP

O *SHAP* (SHapley Additive exPlanations), introduzido por (LUNDBERG; LEE, 2017), é amplamente utilizado para fornecer explicações locais e globais em modelos complexos, como redes neurais e árvores de decisão (MARCíLIO; ELER, 2020). A técnica permite uma explicação robusta e detalhada, baseada em teoria dos jogos, sobre como cada variável contribui para a saída do modelo. Aplicações de *SHAP* são observadas em setores como a saúde, onde a compreensão das variáveis de um paciente é crucial para prever diagnósticos. (LUNDBERG; ERION; LEE, 2020)

2.5 Visualização Interativa em XAI

Ferramentas de visualização interativa têm sido desenvolvidas para facilitar a compreensão de modelos explicáveis de IA. Interfaces que permitem a manipulação de variáveis e observação de seus impactos em tempo real sobre os resultados do modelo são essenciais para explorar e entender o comportamento do sistema (OOGE; VERBERT, 2022; ZöLLER et al., 2023).

2.6 Proveniência de Dados e Transparência

A proveniência de dados é um elemento importante na explicação de modelos de IA. A rastreabilidade do ciclo de vida dos dados, desde sua origem até o uso no modelo, garante transparência e permite auditoria dos resultados (WERDER; RAMESH; ZHANG, 2022). A integração de dados de proveniência com XAI oferece uma visão mais ampla dos processos decisórios e possibilita maior confiança nos modelos.

2.7 Seleção de Características com SHAP

Além de fornecer explicações sobre as previsões, *SHAP* tem sido utilizado como uma ferramenta eficaz para a seleção de características. (MARCíLIO; ELER, 2020) demonstraram que os valores *SHAP* podem ser usados para identificar as variáveis mais relevantes de um modelo, otimizando seu desempenho. Essa técnica tem aplicações práticas em cenários com grandes volumes de dados, onde é necessário identificar e utilizar apenas as variáveis mais importantes.

2.8 Interação Humano-Computador (IHC)

A *Interação Humano-Computador* (IHC) é fundamental no desenvolvimento de plataformas de XAI. A usabilidade e acessibilidade das interfaces são essenciais para que usuários possam explorar os modelos de IA e entender seus processos decisórios (NAZAR et al., 2021). Em setores como a saúde, sistemas de IHC que integram explicações claras e acessíveis são especialmente importantes para a adoção de IA em ambientes de tomada de decisão (NAZAR et al., 2021).

3 TRABALHOS RELACIONADOS

Nesta seção, discutimos os principais trabalhos relacionados à visualização interativa aplicada à explicabilidade de modelos de inteligência artificial (IA). A visualização interativa em IA explicável tem como objetivo tornar os modelos de aprendizado de máquina mais acessíveis, permitindo que os usuários explorem, compreendam e ajustem os modelos de maneira intuitiva.

3.1 Visualização Interativa e Explicabilidade

O desenvolvimento de ferramentas de visualização interativa tem avançado significativamente, com a intenção de melhorar a compreensão dos usuários sobre os modelos de IA. O estudo de (OOGE; VERBERT, 2022) apresenta uma plataforma de visualização que oferece explicações personalizadas de modelos de IA, adaptando-se ao público-alvo e ao contexto de uso. A pesquisa sugere que interfaces visuais customizadas podem facilitar a compreensão de modelos complexos por diferentes tipos de usuários, como especialistas e leigos, ao fornecer insights sobre como o modelo toma suas decisões.

3.2 XAutoML: Visual Analytics para AutoML

O trabalho de (ZöLLER et al., 2023) introduz a plataforma XAutoML, que combina análise visual interativa com técnicas de AutoML (Automated Machine Learning). A ferramenta permite que os usuários visualizem e ajustem os parâmetros dos modelos de aprendizado de máquina ao longo do pipeline. O XAutoML é particularmente relevante para a explicabilidade, pois oferece uma maneira de visualizar cada etapa do processo de construção de modelos, permitindo uma exploração detalhada de como as variáveis e configurações impactam o desempenho do modelo. A integração de visualizações interativas com AutoML possibilita que os usuários explorem modelos complexos sem a necessidade de conhecimentos profundos em aprendizado de máquina.

3.3 ExplAIner: Visualização Interativa e Explicações de IA

O trabalho de (SPINNER et al., 2019) apresenta o framework explAIner, uma ferramenta de visual analytics que visa fornecer explicações interativas para modelos de aprendizado de máquina. A plataforma permite que os usuários explorem diferentes aspectos dos modelos, incluindo a importância das variáveis e o impacto de cada variável sobre as previsões. O framework utiliza gráficos interativos que facilitam a exploração dos resultados, tornando mais fácil para o usuário identificar as variáveis que mais influenciam o desempenho do modelo. A visualização é um elemento-chave para tornar os modelos explicáveis e acessíveis a públicos com diferentes níveis de experiência técnica.

3.4 Aplicações em Saúde com Visualização Interativa

Ferramentas de visualização interativa têm encontrado aplicações em áreas como a saúde, onde a explicabilidade de IA é fundamental para garantir a confiança nas decisões automáticas. O estudo de (LUNDBERG; ERION; LEE, 2020) aplica SHAP em um sistema clínico, utilizando visualizações interativas para explicar as previsões de um modelo que monitora pacientes durante a cirurgia. A visualização do impacto das variáveis foi crucial para que os profissionais de saúde pudessem entender como o modelo tomava decisões e para aumentar a confiança nas previsões de risco. A combinação de visualização interativa com técnicas de explicabilidade, como o SHAP, demonstra o valor dessas ferramentas em cenários sensíveis.

3.5 Personalização e Flexibilidade nas Visualizações

O trabalho de (OOGE; VERBERT, 2022) também destaca a importância da personalização em interfaces de visualização interativa. Ferramentas que permitem ajustar a interface às necessidades do usuário podem melhorar significativamente a acessibilidade e a utilidade dos modelos de IA. A flexibilidade nas visualizações permite que usuários de diferentes contextos e níveis de expertise, como cientistas de dados ou médicos, explorem os modelos de maneira personalizada, adaptando as explicações ao seu nível de entendimento e ao seu contexto de decisão

4 METODOLOGIA

Este capítulo .... Apresentar uma classificação da pesquisa.

4.1 Atividades a serem realizadas

Esta seção apresenta ....

4.1.1 Atividade 1: xxxx

Descrição

4.1.2 Atividade 2: xxxx

Descrição

4.1.3 Atividade n: xxxx

Descrição

4.2 Cronograma

Esta seção apresenta ... (Tabela 1).

Tabela 1 – Cronograma

	Meses 1-3	**Meses 4-6**	**Meses 7-9**	**Meses 10-11**
Pesquisa asdads	X	X		
Coleta de dados		X	X	
sdfsdf	X		X	X
nova linha	X		X	X

5 PRIMEIRO CAPÍTULO DE EXEMPLO

A seguir serão apresentados alguns comandos do LaTex usados comumente para formatar textos de dissertação baseados na normalização da PUC (2011).

Para as citações a norma estabelece duas formas de apresentação. A primeira delas é empregada quando a citação aparece no final de um parágrafo. Neste caso, o comando cite é usado para formatar a citação em caixa alta, como é mostrado no exemplo a seguir. (DUATO; YALAMANCHILI; LIONEL, 2002).

Outra forma de apresentação da citação é a que ocorre no decorrer do texto, essa situação é exemplificada na próxima frase. Conforme Bjerregaard e Mahadevan (2006), o estudo mencionado revela progressos no desempenho dos processadores. Para a formatação da citação em caixa baixa deve ser usado o comando citeonline.

Nas citações que aparecem mais de uma referência as mesmas devem ser separadas por vírgulas, como neste exemplo. (KEYES, 2008; ZHAO, 2008; GANGULY et al., 2011). Se houver necessidade de especificar a página ou que foi realizada uma tradução do texto deve ser feito da seguinte maneira. (SASAKI et al., 2009, p. 2, tradução nossa). A citação direta deve ser feita de forma semelhante. "[...] A carga de trabalho de um sistema pode ser definida como o conjunto de todas as informações de entrada." (MENASCE; ALMEIDA, 2002, p. 160).

O arquivo dissertacao.bib mostra exemplos de representação para vários tipos de referências (artigos de conferências, periódicos, relatórios, livros, dentre outros). Cada um desses tipos requer uma forma diferente de representação para que a referência seja formatada conforme as exigências da normalização.

5.1 Primeira seção

Para gerar a lista de siglas automaticamente deve ser usado o pacote *acronym*. Para tanto, toda vez que uma sigla for mencionada no texto deve ser usado o comando ac{sigla}. Dessa forma, se for a primeira ocorrência da sigla a mesma será escrita por extenso conforme descrição feita no arquivo lista-siglas.tex. Caso contrário, somente a sigla será mostrada. Ex

5.1.1 Primeira subseção

As enumerações devem ser geradas usando o pacote *compactitem*. Cada item deve terminar com um ponto final. Abaixo um exemplo de enumeração é apresentado:

a) Coletar e analisar.
b) Configurar e simular.
c) Definir a metodologia.
d) Avaliar o desempenho.
e) Analisar e avaliar características.

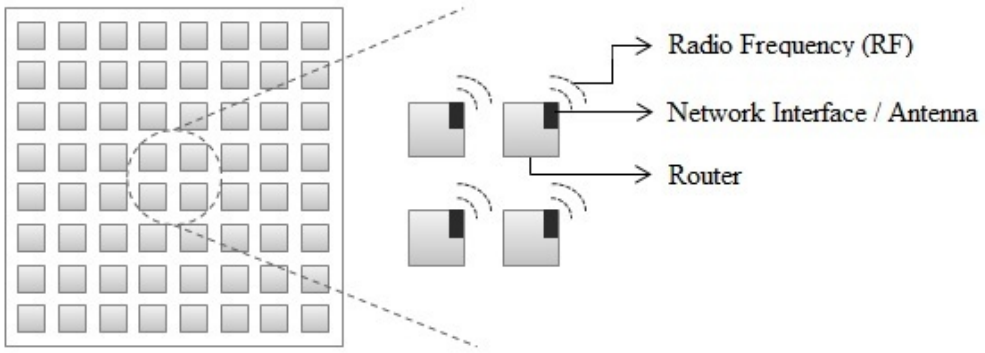
5.2 Segunda seção

Para referenciar um capítulo, seção ou subseção basta definir um label para o mesmo e usar o comando ref para referênciá-lo no texto. Exemplo: Como pode ser visto no Capítulo 5 ou na Seção 5.1.

6 SEGUNDO CAPÍTULO DE EXEMPLO

As figuras devem ser apresentadas pelos comandos abaixo. O parâmetro *width* determina o tamanho que a figura será exibida. No parâmetro *caption* o texto que aparece entre colchetes será o exibido no índice de figuras e o texto contido entre chaves será exibido na legenda da figura. Para citar a figura o comando ref deve ser usado juntamente com o label, como é mostrado nesse exemplo da Figura 1.

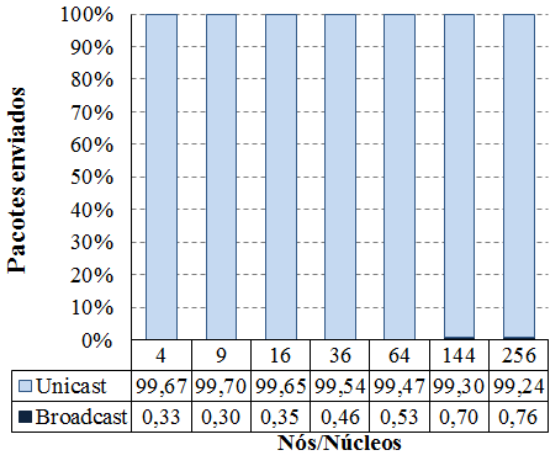
Figura 1 – Principais componentes de WiNoCs



Fonte: (OLIVEIRA et al., 2011)

Os comandos abaixo são usados para apresentação de gráficos. A diferença está apenas na definição do tipo “grafico“ que permite a adição dos itens no índice de gráficos de forma automática. Os parâmetros são semelhantes aos usados para representação de figuras. O parâmetro *width* determina o tamanho do gráfico. O texto entre colchetes no *caption* será o exibido no índice de gráficos e o texto contido entre chaves será exibido na legenda.

Gráfico 1 – Percentual de pacotes enviados



	4	9	16	36	64	144	256
Unicast	99,67	99,70	99,65	99,54	99,47	99,30	99,24
Broadcast	0,33	0,30	0,35	0,46	0,53	0,70	0,76

Fonte: Dados da pesquisa

Um exemplo de criação de tabela é mostrado a seguir. As colunas são separadas por elementos & e as linhas por duas barras invertidas. Os comandos *hline* e | definem a criação de linhas e colunas para separar os conteúdos, respectivamente. A tabela pode ser referenciada usando o comando ref juntamente com o label, como na Tabela 2.

Tabela 2 – Parâmetros definidos por classe

Benchmark	Parâmetro	Classe S	Classe W	Classe A	Classe B	Classe C	Classe D
BT	*Grid*	12^3	24^3	64^3	102^3	162^3	408^3
CG	Linhas	1400	7000	14000	75000	150000	1500000
EP	Pares	2^{24}	2^{25}	2^{28}	2^{30}	2^{32}	2^{36}
FT	*Grid*	64^3	$128^2 * 32$	$256^2 * 128$	$512 * 256^2$	512^3	$2048 * 1024^2$
IS	Chaves	2^{16}	2^{20}	2^{23}	2^{25}	2^{27}	2^{31}
LU	*Grid*	12^3	33^3	64^3	102^3	162^3	408^3
MG	*Grid*	32^3	128^3	256^3	256^3	512^3	1024^3
SP	*Grid*	12^3	36^3	64^3	102^3	162^3	408^3

Fonte: Adaptado de (NPB, 2011)

7 OBSERVAÇÕES IMPORTANTES

Este documento foi compilado em ambiente linux (Ubuntu 10.04) usando o programa Kile - an Integrated LaTeX Environment - Version 2.0.85. Para correta formatação os seguintes arquivos do pacote *abntex* devem ser alterados.

a) Arquivo abnt.cls
   No Ubuntu o arquivo fica armazenado em */usr/share/texmf/tex/latex/abntex*. Comentar a linha 967: Linha comentada para reduzir o espaçamento entre o topo da página e o título. Alterar a linha 1143: Parâmetro alterado de 30pt para -30pt para reduzir o espaçamento entre o top da página e o título do apêndice. Alterar a linha 985: Parâmetro alterado de 0pt para -30pt para reduzir o espaçamento entre o top da página e o título. Alterar a linha 991: Parâmetro alterado de 45pt para 30pt para reduzir o espaçamento entre o texto e o título.
b) Arquivo acronym.sty
   No Ubuntu o arquivo fica armazenado em */usr/share/texmf-texlive/tex/latex/acronym*. Alterar a linha 225: Inserir o separador – entre acrônimo/descrição e remover o negrito com o *normalfont*.

REFERÊNCIAS

BJERREGAARD, T.; MAHADEVAN, S. A survey of research and practices of network-on-chip. **Computing Surveys**, ACM, New York, USA, v. 38, n. 1, p. 1–51, Jun. 2006. ISSN 0360-0300.

DUATO, J.; YALAMANCHILI, S.; LIONEL, N. **Interconnection networks**: an engineering approach. San Francisco: Morgan Kaufmann Publishers, 2002. 515 p. ISBN 1558608524.

DWIVEDI, R. et al. Explainable ai (xai): Core ideas, techniques, and solutions. ACM Computing Surveys, v. 55, n. 9, p. Article 194, 33 pages, Sep. 2023.

GANGULY, A. et al. Scalable hybrid wireless network-on-chip architectures for multi-core systems. **Journal Transactions on Computers**, IEEE Computer Society, Los Alamitos, USA, v. 60, n. 10, p. 1485–1502, 2011. ISSN 0018-9340.

KEYES, R. W. Moore's law today. **Circuits and Systems Magazine**, IEEE Computer Society, Los Alamitos, USA, v. 8, n. 2, p. 53–54, 2008.

LUNDBERG, S.; LEE, S.-I. A Unified Approach to Interpreting Model Predictions. 2017. Disponível em: <https://arxiv.org/abs/1705.07874>.

LUNDBERG, S. M.; ERION, G. G.; LEE, S. I. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nature Biomedical Engineering, p. 1–17, 2020.

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, 33., 2020, Porto de Galinhas, Brazil. **Proceedings...** Porto de Galinhas: IEEE, 2020. p. 340–347.

MENASCE, D. A.; ALMEIDA, V. A. F. **Planejamento de capacidade para serviços na web**: métricas, modelos e métodos. Rio de Janeiro: Campus, 2002. 472 p. ISBN 8535211020.

NAZAR, M. et al. A systematic review of human–computer interaction and explainable artificial intelligence in healthcare with artificial intelligence techniques. IEEE Access, v. 9, p. 153316–153348, Nov. 2021.

NPB. **NAS Parallel Benchmarks**. Disponível em http://www.nas.nasa.gov/publications/npb.html. Acesso em jun. 2011.

OLIVEIRA, P. A. C. et al. Performance evaluation of winocs for parallel workloads based on collective communications. In: IADIS APPLIED COMPUTING, 8., 2011, Rio de Janeiro, Brasil. **Proceedings...** Rio de Janeiro: IADIS Applied Computing, 2011. p. 307–314.

OOGE, J.; VERBERT, K. Explaining artificial intelligence with tailored interactive visualisations. In: 27TH INTERNATIONAL CONFERENCE ON INTELLIGENT USER INTERFACES (IUI '22 COMPANION), 27., 2022, New York, USA. **Proceedings...** New York: ACM, 2022. p. 120–123.

RUSSELL, S. J.; NORVIG, P. Artificial Intelligence: A Modern Approach. 3rd. ed. Upper Saddle River, NJ: Pearson Education, 2016.

SASAKI, N. et al. A single-chip ultra-wideband receiver with silicon integrated antennas for inter-chip wireless interconnection. **Journal of Solid-State Circuits**, IEEE Computer Society, Los Alamitos, USA, v. 44, n. 2, p. 382–393, Feb. 2009. ISSN 0018-9200.

SPINNER, T. et al. explainer: A visual analytics framework for interactive and explainable machine learning. IEEE Transactions on Visualization and Computer Graphics, 2019.

WERDER, K.; RAMESH, B.; ZHANG, R. Establishing data provenance for responsible artificial intelligence systems. ACM Transactions on Management Information Systems, v. 13, n. 2, p. Article 22, 23 pages, Jun. 2022.

ZHAO, D. Ultraperformance wireless interconnect nanonetworks for heterogeneous gigascale multi-processor SoCs. In: 2TH WORKSHOP ON CHIP MULTIPROCESSOR, MEMORY SYSTEMS AND INTERCONNECTS, 3., 2008, Beijing, China. **Proceedings...** Beijing: CMP-MSI, 2008. p. 1–3.

ZöLLER, M.-A. et al. Xautoml: A visual analytics tool for understanding and validating automated machine learning. ACM Transactions on Interactive Intelligent Systems, v. 13, n. 4, p. Article 28, 39 pages, Dec. 2023.