Visualization Taxonomy

Below is our tool reference list followed by our visualization design elements taxonomy developed from a comprehensive set of 26 visualizations from 20 visualization-based ML explainability tools for local feature importance of tabular data models. We split up our tools into sets of 10, 10, and 2, to make the tables easier to parse.

REFERENCES

[1] A. Al-Ebrahim and M. Sharif. explainX. github.com/explainX. 3
[2] H. Baniecki and P. Biecek. modelStudio: Interactive Studio with Explanations for ML Predictive Models. *Journal of Open Source Software*, 4(43):1798, 2019. 2
[3] P. Biecek. DALEX: Explainers for complex predictive models in R. *JMLR*, 19(84):1–5, 2018. 3
[4] P. Biecek and T. Burzykowski. *Explanatory Model Analysis*. Chapman and Hall/CRC, New York, 2021. 2, 3
[5] P. Biecek et al. Effects and importances of model ingredients. https://github.com/ModelOriented/ingredients, 2020. 3
[6] Google. Vertex explainable AI. cloud.google.com/vertex-ai. 4
[7] A. Gosiewska and P. Biecek. Do not trust additive explanations. arxiv.org/abs/1903.11420, 2020. 2
[8] A. Gosiewska et al. An r wrapper of shap python library. https://github.com/ModelOriented/shapper, 2023. 3
[9] J. Klaise, A. V. Looveren, G. Vacanti, and A. Coca. Alibi explain: Algorithms for explaining machine learning models. *JMLR*, 22(1), 2021. 3, 4
[10] M. Korobov and K. Lopuhin. ELI5. github.com/eli5-org/eli5/tree/master, 2016–2017. 2
[11] S. M. Lundberg et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature Biomedical Engineering*, 2(10):749, 2018. 2
[12] S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. In *NeurIPS*, vol. 30, 2017. 2
[13] M. Mayer. *flashlight: Shed Light on Black Box Machine Learning Models*, 2024. R package version 0.9.0.9000. 4
[14] R. K. Mothilal, A. Sharma, and C. Tan. Explaining machine learning classifiers through diverse counterfactual explanations. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 607–617, 2020. 4
[15] H. Nori, S. Jenkins, P. Koch, and R. Caruana. InterpretML: A unified framework for machine learning interpretability. *arXiv:1909.09223*, 2019. 3
[16] E. Ranguelova et al. *dianna*, 2024. zenodo.org/record/5592606. 3
[17] M. T. Ribeiro, S. Singh, and C. Guestrin. “Why should I trust you?”: Explaining the predictions of any classifier. In *KDD*, pp. 1135–1144, 2016. doi: 10.1145/2939672.2939778 2
[18] M. T. Ribeiro, S. Singh, and C. Guestrin. Anchors: High-precision model-agnostic explanations. In *AAAI*, 2018. 2
[19] M. Staniak and P. Biecek. Explanations of model predictions with live and breakDown packages. *The R Journal*, 2018. doi: 10.32614/RJ-2018-072 2
[20] I. Tenney, R. Mullins, B. Du, S. Pandya, M. Kahng, and L. Dixon. Interactive prompt debugging with sequence salience. *arXiv preprint arXiv:2404.07498*, 2024. 4
[21] J. Wexler. The what-if tool: Code-free probing of machine learning models. ai.googleblog.com/2018/09/the-what-if-tool-code-free-probing-of.html, 2018. 4
[22] F. Yang, Z. Zhang, H. Wang, Y. Li, and X. Hu. XDeep: An interpretation tool for deep neural networks. *arXiv:1911.01005*, 2019. 3

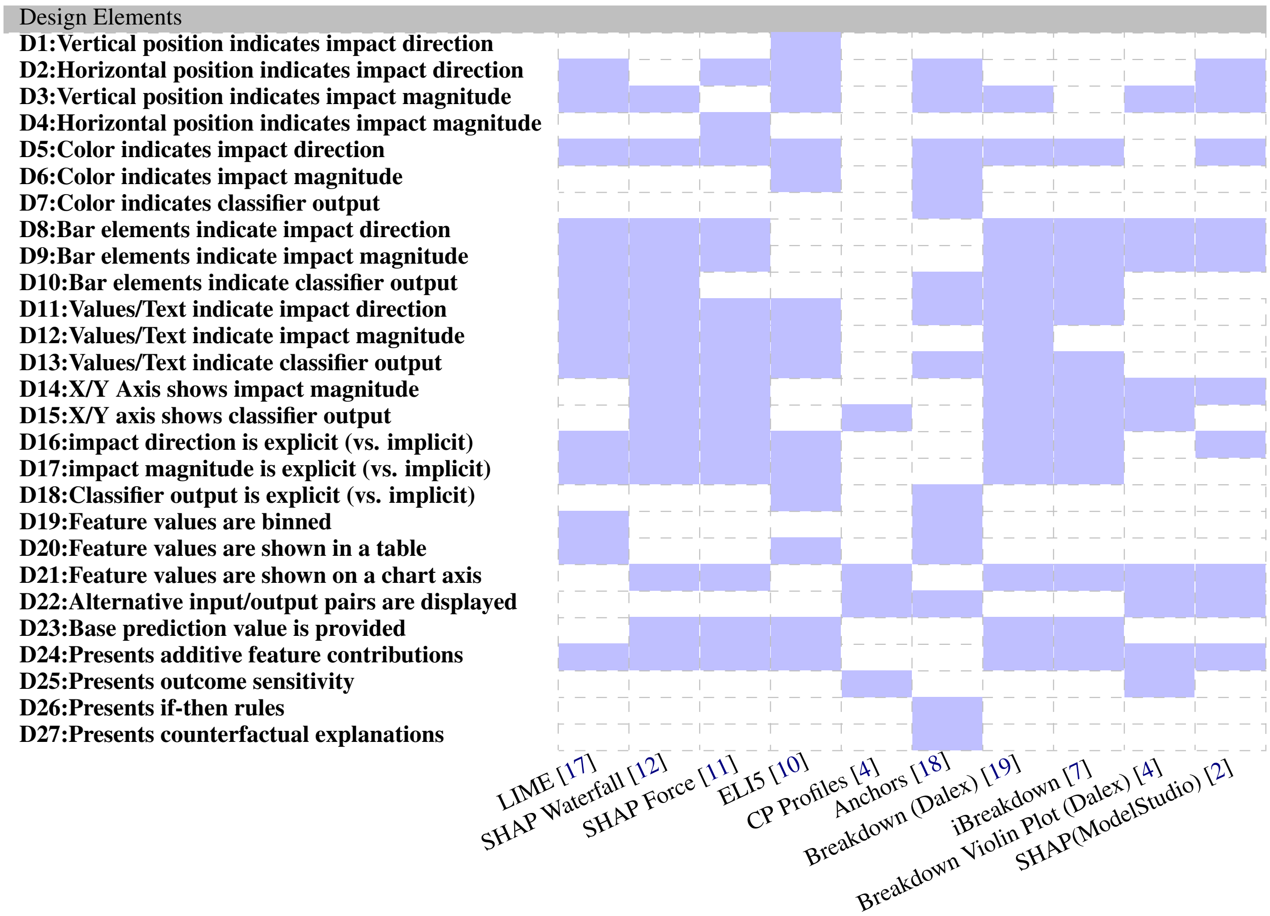

Figure 1: Our visualization design elements taxonomy developed from a comprehensive set of 26 visualizations from 15 visualization-based ML explainability tools for local feature importance of tabular data models.

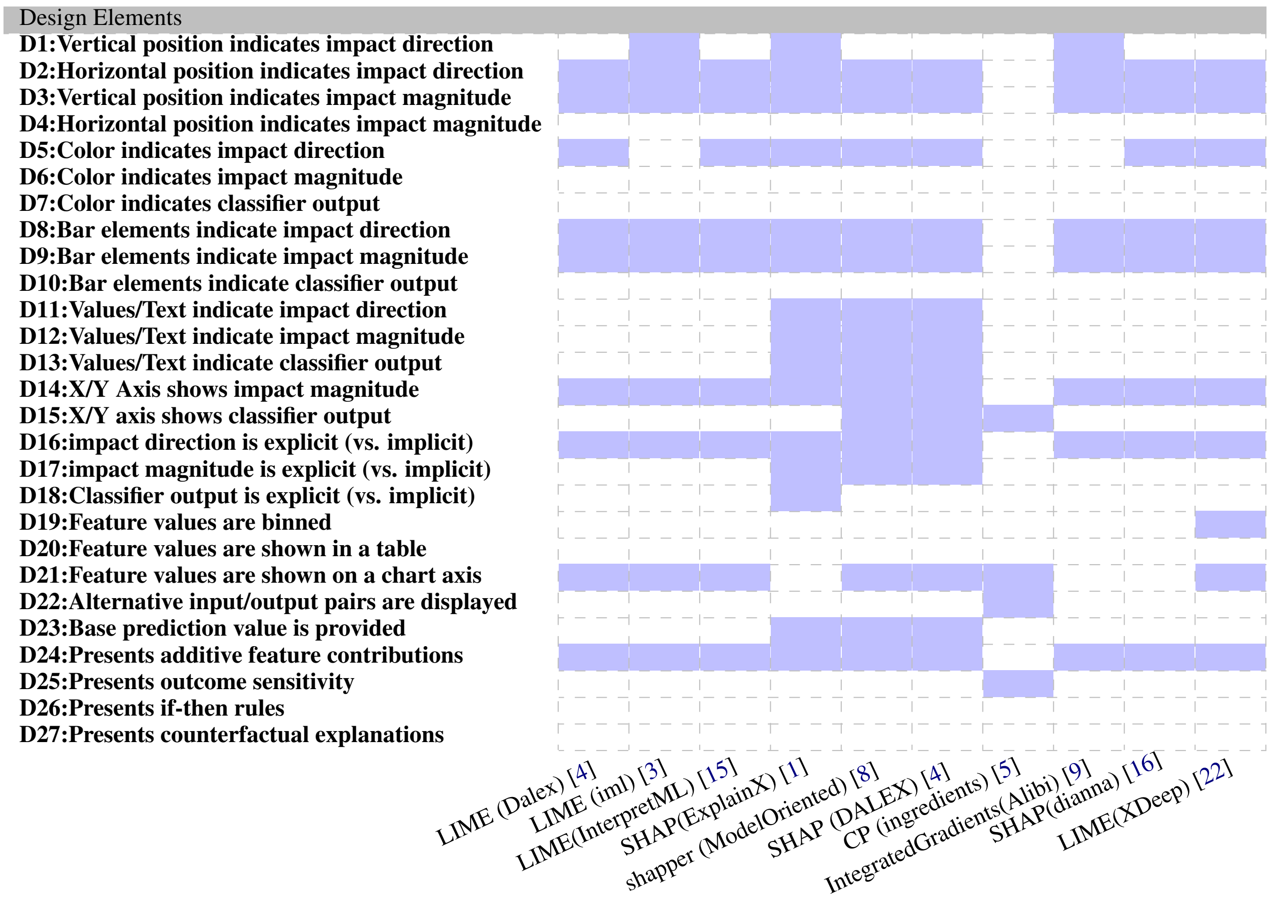

Figure 2: Our visualization design elements taxonomy developed from a comprehensive set of 26 visualizations from 15 visualization-based ML explainability tools for local feature importance of tabular data models.

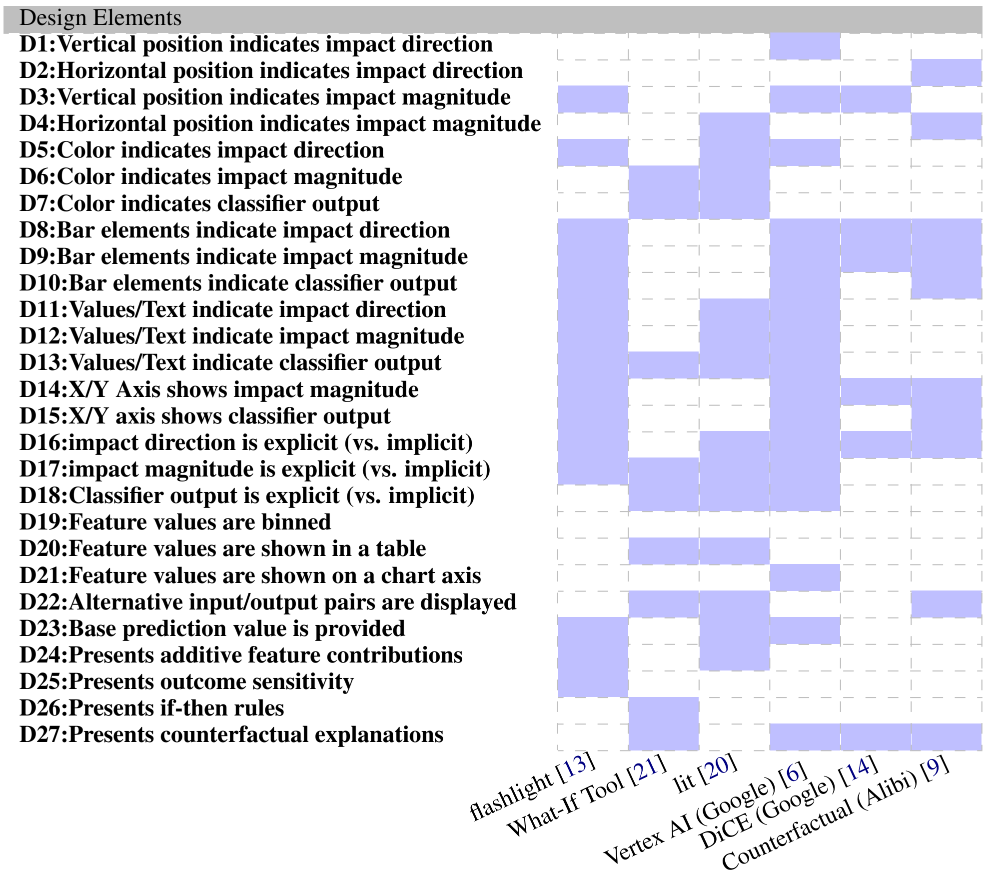

Figure 3: Our visualization design elements taxonomy developed from a comprehensive set of 26 visualizations from 20 visualization-based ML explainability tools for local feature importance of tabular data models.