Demo: an Interactive Visualization Combining Rule-Based and Feature Importance Explanations

Eleonora Cappuccio
Università di Pisa and Università degli Studi di Bari Aldo Moro
Italy
eleonora.cappuccio@phd.unipi.it

Daniele Fadda
CNR
Pisa, Italy
daniele.fadda@isti.cnr.it

Rosa Lanzilotti
Università degli Studi di Bari Aldo Moro
Bari, Italy
rosa.lanzilotti@uniba.it

Salvatore Rinzivillo
CNR
Pisa, Italy
rinzivillo@isti.cnr.it

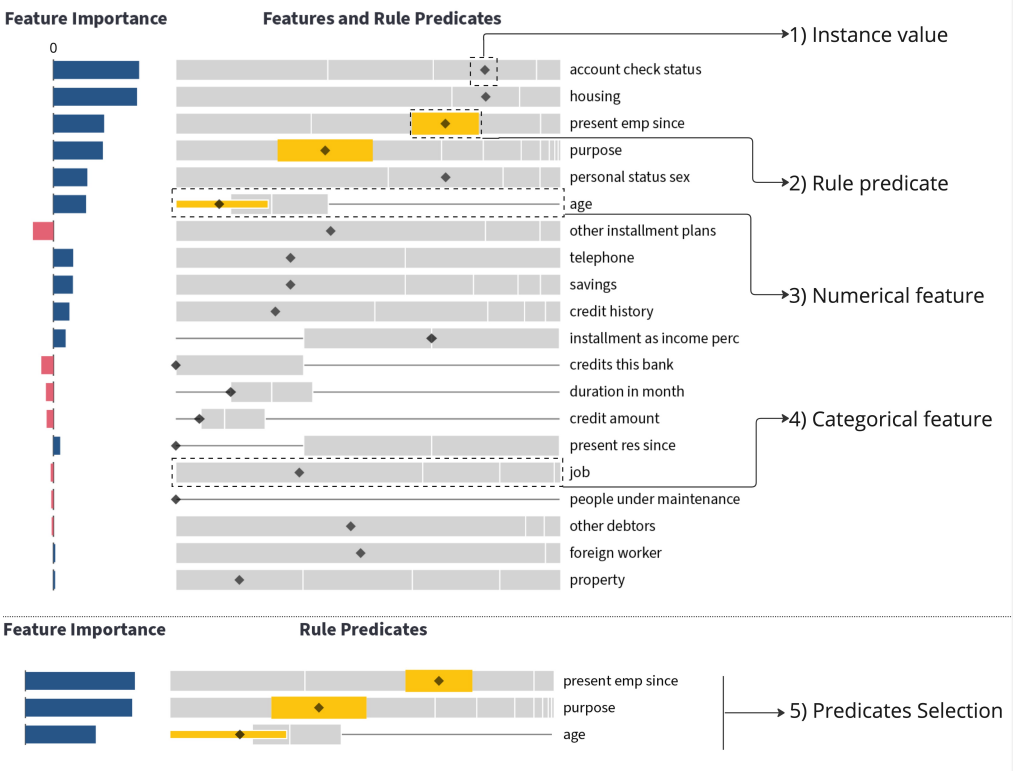


Figure 1: FIPER visualization of one instance of the German Credit Risk dataset.

ABSTRACT

The Human-Computer Interaction (HCI) community has long stressed the need for a more user-centered approach to Explainable Artificial Intelligence (XAI), a research area that aims at defining algorithms and tools to illustrate the predictions of the so-called black-box models. This approach can benefit from the fields of user-interface, user experience, and visual analytics. In this demo, we propose a visual-based tool, "F.I.P.E.R.", that shows interactive explanations combining rules and feature importance.

CHItaly 2023, September 20–22, 2023, Torino, Italy

ACM ISBN 979-8-4007-0806-0/23/09.
https://doi.org/10.1145/3605390.3610811

CCS CONCEPTS

• **Human-centered computing** → *Human computer interaction, Visualization* ; • **Computing methodologies** → *Machine learning*.

KEYWORDS

Explainable AI, User-interface

ACM Reference Format:

Eleonora Cappuccio, Daniele Fadda, Rosa Lanzilotti, and Salvatore Rinzivillo. 2023. Demo: an Interactive Visualization Combining Rule-Based and Feature Importance Explanations. In *15th Biannual Conference of the Italian SIGCHI Chapter (CHItaly 2023), September 20–22, 2023, Torino, Italy*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3605390.3610811

1 INTRODUCTION

Explainable AI aims to make the internal mechanisms of decision-making algorithms interpretable by humans. One of the main criticisms of Explainable AI algorithm designers has been about creating effective explanations only for those already familiar with the model, but not for lay users [8]. A paradigm shift has long been advocated by the HCI community, which can provide knowledge and techniques for defining user-centered explanations [1, 6]. Several authors have stressed the need to design systems through which the user can interact with the explanation through an interface: in 2018 [5] has separated the explainable AI algorithm from the means of disclosing it. A key area of research in this context concerns the design of Explainable User interfaces (XUIs). The user in this way is no longer faced with a static explanation but is able to interact with the system and gather more meaningful insights. Visualizations can be used to present explanations to users. As pointed out by [2] text can be used for simple explanations, while visualizations are better suited for communicating more complex concepts. Different visualizations can influence how users perceive an explanation [9]. However, not many studies are investigating the role of visualization in conveying algorithmically generated explanations. Therefore, it is important to conduct empirical studies with users to test the effectiveness of proposed visual explanations. This Demo introduces a visualization technique and a tool, called F.I.P.E.R. for rules-based explanations, aiming to provide users with valuable insights regarding prediction explanations.

2 THE FIPER TOOL

FIPER, an acronym for Feature Importance Plot for Explanatory Rules, is a visual rule-based explanation that combines and visualizes evidence extracted from two distinct explanation methods: factual rules [4, 11], and Feature Importance (FI) [7, 10]. The visualization consists of two panels that combine the outcome of both methods 1. The left panel displays and arranges the absolute values of the weights of the FI method. Positive contributions of the corresponding features are represented by blue color coding, while negative contributions are represented by magenta. The right panel visualizes the rule predicates following the ordering imposed by the Feature Importance. The panel displaying rule predicates visually represents all features using a specific chart aligned with the elements in the FI panel. Different types of charts are used depending on the feature type. When dealing with categorical data, a stacked bar chart is employed to illustrate the relationship of each possible value within the whole. This representation allows the user to grasp the internal distribution of values, with a diamond point positioned at the center of the observed value for the attribute x. Numerical data types are represented using a box plot chart, which provides a compact visualization of the data distribution, including the minimum, maximum, first quartile, third quartile, and median. The observed value for x is represented by a diamond point within the scale of the box plot. Additionally, for attributes that have a predicate p in the rule r, a second layer is added to highlight the intervals stated by the rule. The visualization of the intervals varies based on the data type. In the case of categorical data, the intervals contained in the rule premise are highlighted in yellow. For numerical data, a yellow bar represents the range of the predicate values.

Two forms of interactivity are implemented for the F.I.P.E.R. visualization 2:

- By dynamically limiting the view to only the attributes mentioned in the rule, the user's attention can be directed specifically toward the predicates (Figure 2). This interaction aligns with the principles of providing users with convenient access to pertinent and significant information [12].
- By hovering the pointer over the visualization, the user can access more detailed information about the distribution of each attribute. Figure 2illustrates two distinct styles of tooltips for different data types. In the case of categorical data (Top), the selected value and its cardinality are displayed. For numerical data, a set of representative values (such as min, max, Q1, Q3, median) and the corresponding feature value are shown.

To stress the efficacy of FIPER, we implemented two other visual representations of factual rules: LORE output is the text-based raw output of the algorithm; XAI library visualization provides visual formatting of the content of the rule by enhancing the readability of each predicate with a sequence of graphical blocks, see fig:3. As a demonstrator of the interface, we trained a random forest on the instances from the UCI *German Credit Risk* data set [3]. The data set is commonly used in education environments and it contains 20 columns for 1000 loan applications. The classification task consists in understanding whether an applicant is a Good or Bad credit risk. For each instance, the users see the predicted class and two expandable containers with detailed information on the data set columns and the value of the inspected instance. Each outcome of the classifier is augmented with the corresponding explanation. During the demonstration, the user can interact with FIPER and choose and compare the three representations described above. To present our demo, we will use our computers, giving people the opportunity to browse the application and explore the features classified by the Random Forest. We may need a wider external screen to allow participants to interact with the tool.

ACKNOWLEDGMENTS

This work has been supported by the European Community Horizon 2020 programme under the funding scheme ERC-2018-ADG G.A. 834756 *XAI: Science and technology for the eXplanation of AI decision making*, by the European Union's Horizon Europe Programme under the CREXDATA project, grant agreement no. 101092749, by the Next Generation EU: NRRP Initiative, Mission 4, Component 2, Investment 1.3, PE0000013 – "Future Artificial Intelligence Research

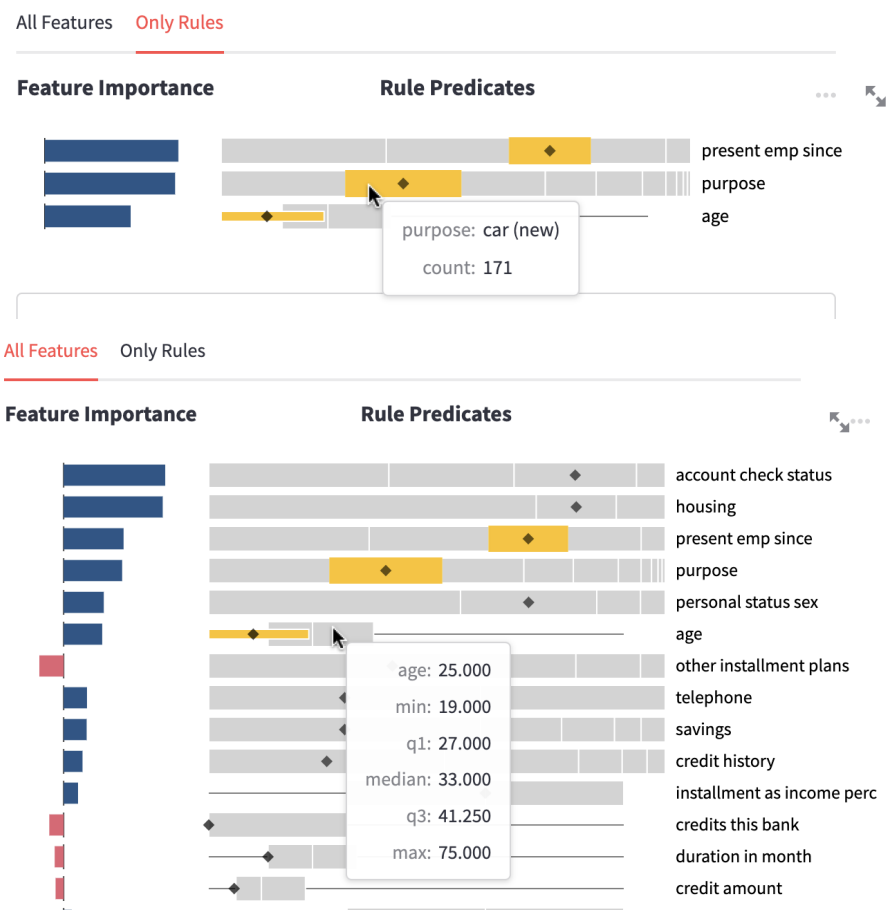
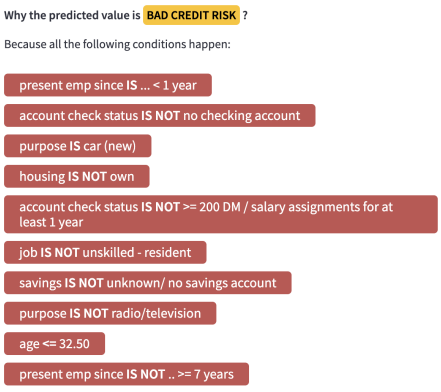


Figure 2: Finer details of a specific feature, selected by hovering the mouse on the corresponding row. (Top) Tooltip for a categorical data type, where the feature's actual value is shown with its class's cardinality. (Bottom) Tooltip for a numerical data type, where statistical central values are shown: min, max, median, Q1, and Q3.

r = { present_emp_since=... < 1 year > 0.79, account_check_status=no checking account <= 0.37, purpose=car (new) > 0.78, housing=own <= 0.50, account_check_status=>= 200 DM / salary assignments for at least 1 year <= 0.50, job=unskilled - resident <= 0.50, savings=unknown/ no savings account <= 0.50, purpose=radio/television <= 0.50, age <= 32.50, present_emp_since=.. >= 7 years <= 0.50 }

(a) LORE Output



(b) XAI Library

Figure 3: An instance visualized as LORE output and XAI library visualization

– FAIR" - CUP: H97G22000210007 and "SoBigData.it - Strengthening the Italian RI for Social Mining and Big Data Analytics" - Prot. IR0000013.

REFERENCES

[1] Ashraf M. Abdul, Jo Vermeulen, Danding Wang, Brian Y. Lim, and Mohan S. Kankanhalli. 2018. Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda. (2018), 582. https://doi.org/10.1145/3173574.3174156

[2] Marina Danilevsky, Kun Qian, Ranit Aharonov, Yannis Katsis, Ban Kawas, and Prithviraj Sen. 2020. A Survey of the State of Explainable AI for Natural Language Processing. (2020), 447–459. https://aclanthology.org/2020.aacl-main.46/

[3] Dheeru Dua and Casey Graff. 2017. UCI Machine Learning Repository. http://archive.ics.uci.edu/ml

[4] Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Dino Pedreschi, Franco Turini, and Fosca Giannotti. 2018. Local Rule-Based Explanations of Black Box Decision Systems. *CoRR* abs/1805.10820 (2018). arXiv:1805.10820 http://arxiv.org/abs/1805.10820

[5] David Gunning and David W. Aha. 2019. DARPA's Explainable Artificial Intelligence (XAI) Program. *AI Mag.* 40, 2 (2019), 44–58. https://doi.org/10.1609/aimag.v40i2.2850

[6] Q. Vera Liao and Kush R. Varshney. 2021. Human-Centered Explainable AI (XAI): From Algorithms to User Experiences. *CoRR* abs/2110.10790 (2021). arXiv:2110.10790 https://arxiv.org/abs/2110.10790

[7] Scott M. Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett (Eds.). 4765–4774. https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

[8] Tim Miller, Piers Howe, and Liz Sonenberg. 2017. Explainable AI: Beware of Inmates Running the Asylum Or: How I Learnt to Stop Worrying and Love the Social and Behavioural Sciences. *CoRR* abs/1712.00547 (2017). arXiv:1712.00547 http://arxiv.org/abs/1712.00547

[9] Henrik Mucha, Sebastian Robert, Rüdiger Breitschwerdt, and Michael Fellmann. 2021. Interfaces for Explanations in Human-AI Interaction: Proposing a Design Evaluation Approach. (2021), 327:1–327:6. https://doi.org/10.1145/3411763.3451759

[10] Marco Túlio Ribeiro, Sameer Singh, and Carlos Guestrin. 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, San Francisco, CA, USA, August 13-17, 2016*, Balaji Krishnapuram, Mohak Shah, Alexander J. Smola, Charu C. Aggarwal, Dou Shen, and Rajeev Rastogi (Eds.). ACM, 1135–1144. https://doi.org/10.1145/2939672.2939778

[11] Marco Túlio Ribeiro, Sameer Singh, and Carlos Guestrin. 2018. Anchors: High-Precision Model-Agnostic Explanations. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, Sheila A. McIlraith and Kilian Q. Weinberger (Eds.). AAAI Press, 1527–1535. https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16982

[12] James Schaffer, Prasanna Giridhar, Debra Jones, Tobias Höllerer, Tarek F. Abdelzaher, and John O'Donovan. 2015. Getting the Message?: A Study of Explanation Interfaces for Microblog Data Analysis. (2015), 345–356. https://doi.org/10.1145/2678025.2701406