# EXTREMUM: A Web-Based Tool to Generate and Explore Counterfactual Explanations on Tabular and Time-Series Data

Athanasios Lakes[1[0009−0005−4803−4722]], Luis Quintero[1[0000−0002−6047−2793]], and Panagiotis Papapetrou (✉)[1[0000−0002−4632−4815]]

Department of Computer and Systems Sciences (DSV), Stockholm University
{athanasio.lakes,luis.quintero,panagiotis}@dsv.su.se

**Abstract.** There is an increasing need to include explainability on the machine learning (ML) models. Among the various approaches, counterfactual (CF) explanations allow the design of what-if scenarios and the interactive exploration of ML model behavior on sensitive decision-making domains. However, the generation of CF for tabular and time-series data requires technical skills that are not always available to the end-users of ML-powered systems. Therefore, we propose a modular web-based tool to easily generate, visualize, and interact with CF on any tabular or time-series dataset. The EXTREMUM platform provides access to state-of-the-art CF algorithms, where users can train ML models and explore CF on their tabular or time-series datasets with an intuitive user interface. The project is instantiated on two tabular datasets within healthcare and five time-series datasets with various domains. The open-source repository lets ML researchers adapt the existing ML tool to new application domains: https://gitea.dsv.su.se/DataScienceGroup/EXTREMUM-demo.

**Keywords:** Counterfactual · Explainability · Interpretability · Time Series · Machine Learning · Interactive · Dashboard · User Interface.

## 1 Introduction

The increasing use of artificial intelligence (AI) tools and predictive machine learning (ML) models also permeates high-stake application domains, such as healthcare. These decision-support systems must strongly emphasize the interpretability of the ML models to ensure accountability, ethical compliance, and trust among end users [9,1]. Several forms of interpretability have emerged to understand the factors that most significantly influence a model's output. For example, feature importance scores provide intuitive insights into which variables affect a prediction; more sophisticated model-agnostic approaches, like LIME and SHAP, allow users to assess the contribution of individual features at a local or global level; and counterfactual (CF) explanations describe how the features from a given data instance would need to change to alter the model's initial prediction.

CF explanations have emerged as a promising direction for AI-powered tools within sensitive domains, as they offer a trade-off between technical (e.g., model performance, explanation validity) and human perspectives (e.g., task performance, explanation understandability). CF explanations remain in the data domain, whereas other explanation techniques often require users to learn ML-specific metrics and visualizations. Moreover, CF reasoning through what-if scenarios is naturally embedded in prescriptive practices like healthcare and banking. Lastly, CF lets end-users keep control of the decision-making process by iteratively examining the factors that influence the model's prediction.

The interest to design ML algorithms for CF generation is increasing for all data modalities [5,10,8,12]. However, running algorithms for CF generation require technical skills that differ vastly from the domain-specific skills held by the end-users exploring CF explanations. Hence, there is a need to simplify the construction of AI-powered tools including state-of-the-art CF explanations, and thus bridge the gap between designers of AI-powered systems and end-users leveraging interpretability in real-life contexts.

Prior **related work** on user-centered explainable AI has identified that web interactive user interfaces are the primary medium to evaluate AI-powered systems with end users [1]. They also outline methodological aspects to assess human-related factors of AI-powered systems such as trust, understanding, usability, and human-AI collaboration performance[9]. There are several open-source implementations of tools related to explainability in AI. Most toolkits aim to simplify the work of data scientists and ML engineers [3], such as the AI explainability 360 [2] that includes ten techniques for non-interactive explanations, or the WebSHAP [13] library that offers model-agnostic explanations for a single technique but with high interactivity. However, these solutions do not offer CF explanations, and their online implementation is restricted to a predefined dataset. Other approaches aimed at end-users, such as What-If Tool [14] and Outcome-Explorer [4], allow interactive exploration of data, ML classifiers, and interpretability with CF reasoning. However, they are aimed to work with tabular data, images, or text; excluding time-series datasets. Interactive user interfaces focused on time series allow imputation techniques [6], and our closest related work [11] allows interactive CF generation tool for univariate time series. However, it involves interpreting projected activation and attributions in deep learning models, which are too technical to be interpreted by real-life practitioners.

Therefore, the main **contribution** of this work is an open-source web-based tool to visualize datasets, train ML classifiers, and explore CF explanations. It allows post-hoc model-agnostic local explanations through CF reasoning using state-of-the-art algorithms. Tabular datasets generate CF with DiCE [8], whereas Glacier [12] generates locally constrained CF where users can favor or discourage specific changes on parts of the time series. EXTREMUM is designed modular and private so that practitioners can deploy the tool locally, upload their dataset, and explore AI-powered decision-making tools within their expertise domain.

## 2 EXTREMUM: Web-Based Tool for Counterfactuals

The tool is publicly accessible on the project's website[1]. The open-source repository[2] contains an example video and installation instructions. The system workflow is divided in three parts, as described below:

### 2.1 Data Selection and Exploratory Data Analysis

The homepage lists available datasets. Users can choose between tabular or time-series data and explore visualizations accordingly. For tabular data, interactive plots are generated asynchronously based on features and classes selected through drop-down menus. For time series, users can browse from predefined generic datasets. Users import and manage their own private data in the locally deployed version. The public version of the tool allows exploration with three example datasets, as shown in Figure 1.

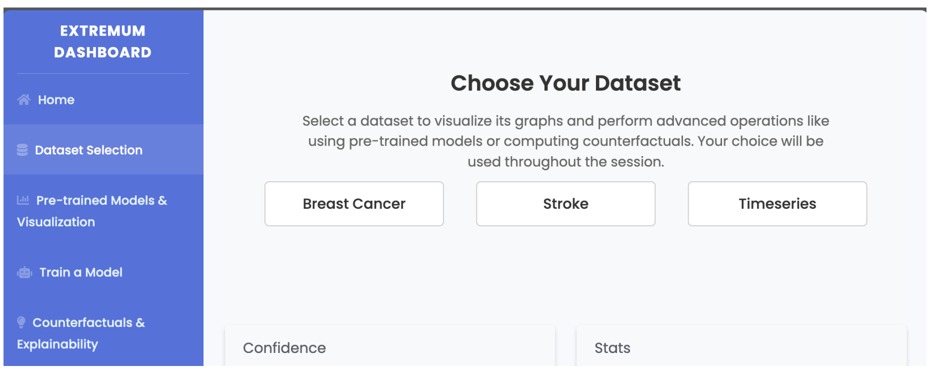


Fig. 1: Example of the EXTREMUM tool allowing interactive exploration of several example datasets, two tabular or five time series.

### 2.2 Model Training

The local version allows training new ML models; in the online version, users choose pre-trained models and see their performance via classification reports, feature importance, and PCA plots. For tabular data, the available models are Linear Regression, Decision Tree, Random Forest, SVM, and XGBoost. For time series, 1D-CNN is used for Glacier, while K-NN and Random Shapelet Forest are used with Wildboar [10]. Users must also define preprocessing, test set ratio, and (for tabular) the target label. After training, a report is shown to help decide to save the model for future interactive sessions.

[1] Online version: https://extremum.dsv.su.se/app
[2] Open-source repo: https://gitea.dsv.su.se/DataScienceGroup/EXTREMUM-demo

### 2.3 Counterfactual Generation and Exploration

CFs are generated through an interactive workflow, where users pick specific data samples to view the CF list from a pre-trained model. The process varies by dataset type. A t-SNE plot is rendered for tabular data, allowing interactive point selection. These 2D points map back to the original data for CF generation. For time series, if Glacier is chosen, the user selects a constraint (either from a pre-computed experiment or a new one) before generating CF. On top of that, sample selection and target label remain the same for both CF algorithms. The Figure 2 displays a counterfactual for an example instance in the ECG dataset.

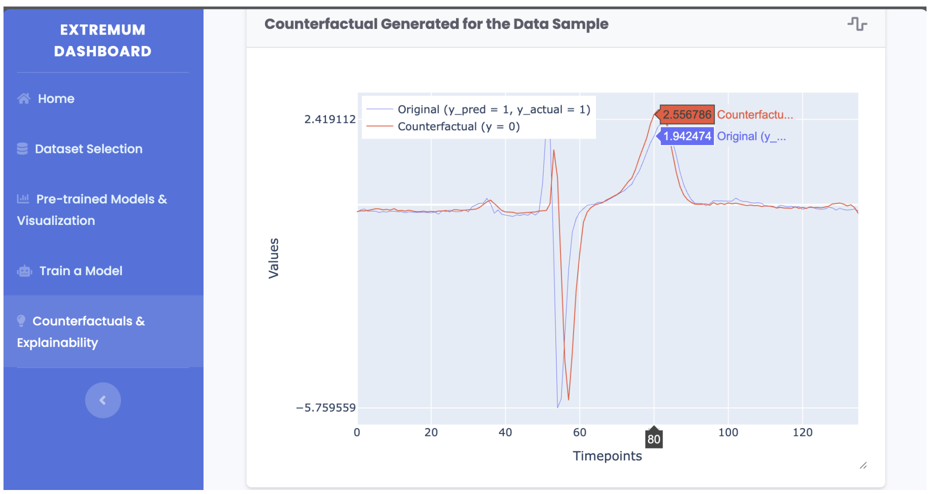


Fig. 2: Example of interactive exploration of counterfactuals for time-series data

## 3 Conclusion and Future Work

EXTREMUM is inspired by the possibilities of CF to support human-AI collaborative decision-making within healthcare [7]. It allows the abstraction of advanced ML algorithms, instantiated on two tabular datasets for medical applications (breast cancer, stroke) and two time-series ECG datasets. The intuitive user interface is an end-to-end workflow that enables users to upload private datasets, train and manage ML classifiers, visualize results, and conduct interactive CF reasoning. It may support assistive tools for practitioners assessing patient data and wanting actionable insights for possible treatment plans. The open-source project may also be adapted to other high-stake applications.

**Acknowledgments.** This study was partly funded by the Digital Futures project titled EXTREMUM, and partly by DSV at Stockholm University.

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## References


1. Al-Ansari, N., Al-Thani, D., Al-Mansoori, R.S.: User-Centered Evaluation of Explainable Artificial Intelligence (XAI): A Systematic Literature Review. Human Behavior and Emerging Technologies **2024**(1), 4628855 (2024). https://doi.org/10.1155/2024/4628855
2. Arya, V., Bellamy, R.K.E., Chen, P.Y.., Zhang, Y.: AI explainability 360: an extensible toolkit for understanding data and machine learning models. J. Mach. Learn. Res. **21**(1), 130:5114–130:5119 (2020)
3. Hohman, F., Park, H., Robinson, C., Polo Chau, D.H.: Summit: Scaling Deep Learning Interpretability by Visualizing Activation and Attribution Summarizations. IEEE Transactions on Visualization and Computer Graphics **26**(1), 1096–1106 (2020). https://doi.org/10.1109/TVCG.2019.2934659
4. Hoque, M.N., Mueller, K.: Outcome-Explorer: A Causality Guided Interactive Visual Interface for Interpretable Algorithmic Decision Making. IEEE Transactions on Visualization and Computer Graphics **28**(12), 4728–4740 (2022). https://doi.org/10.1109/TVCG.2021.3102051
5. Karlsson, I., Rebane, J., Papapetrou, P., Gionis, A.: Locally and globally explainable time series tweaking. Knowledge and Information Systems **62**(5), 1671–1700 (2020). https://doi.org/10.1007/s10115-019-01389-4
6. Khayati, M., Nater, Q., Pasquier, J.: ImputeVIS: An Interactive Evaluator to Benchmark Imputation Techniques for Time Series Data. Proc. VLDB Endow. **17**(12), 4329–4332 (2024). https://doi.org/10.14778/3685800.3685867
7. Lee, M.H., Chew, C.J.: Understanding the Effect of Counterfactual Explanations on Trust and Reliance on AI for Human-AI Collaborative Clinical Decision Making. Proc. ACM Hum.-Comput. Interact. **7**, 369:1–369:22 (2023). https://doi.org/10.1145/3610218
8. Mothilal, R.K., Sharma, A., Tan, C.: Explaining Machine Learning Classifiers through Diverse Counterfactual Explanations. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. pp. 607–617 (2020). https://doi.org/10.1145/3351095.3372850
9. Rong, Y., Leemann, T., Nguyen, T.T., Fiedler, L., Qian, P., Unhelkar, V., Seidel, T., Kasneci, G., Kasneci, E.: Towards Human-Centered Explainable AI: A Survey of User Studies for Model Explanations. IEEE Transactions on Pattern Analysis and Machine Intelligence **46**(4), 2104–2122 (2024). https://doi.org/10.1109/TPAMI.2023.3331846
10. Samsten, I.: isaksamsten/wildboar: wildboar (2020). https://doi.org/10.5281/ZENODO.4264062
11. Schlegel, U., Rauscher, J., Keim, D.A.: Interactive Counterfactual Generation for Univariate Time Series (2024). https://doi.org/10.48550/arXiv.2408.10633
12. Wang, Z., Samsten, I., Miliou, I., Mochaourab, R., Papapetrou, P.: Glacier: guided locally constrained counterfactual explanations for time series classification. Machine Learning **113**(3) (Mar 2024). https://doi.org/10.1007/s10994-023-06502-x
13. Wang, Z.J., Chau, D.H.: WebSHAP: Towards Explaining Any Machine Learning Models Anywhere. In: Companion Proceedings of the ACM Web Conference 2023. pp. 262–266 (2023). https://doi.org/10.1145/3543873.3587362
14. Wexler, J., Pushkarna, M., Bolukbasi, T., Wattenberg, M., Viégas, F., Wilson, J.: The What-If Tool: Interactive Probing of Machine Learning Models. IEEE Transactions on Visualization and Computer Graphics **26**(1), 56–65 (2020). https://doi.org/10.1109/TVCG.2019.2934619