

THE ARTIST TECHNOLOGY FOR TRUSTWORTHY AI - OUR SOLUTION TO DATA BIAS

S. Cléménçon, D. Cortès, A. Dib & D. Moret
Télécom Paris - Hi!Paris Cluster 2030



The ARTIST Project in Brief

A software in the domain of **Trustworthy AI** for:

- solving **data bias** issues (selection/drift), with automatic :
detection - explanation (XAI) - mitigation
- **scoring** : best-in-class technology



Fig. 1: The solution to data drift/selection bias issues

A technological solution **based on LTCI research**

- **>35 publications** in Q1 conferences/journals, e.g. [3], [4], [2], [1]
- based on **proprietary ML algorithms** (patents TreeRank [5], [6])

Data Bias Control - Trusted AI

- ML-based AI learns according to a **frequentist principle** by re-producing training data (massive, high-dimensional)
- **AI performance** (generalization ability) is only guaranteed if **data in the predictive phase is distributed like training data**

$$P_{train} = P_{test}$$

- In most cases, training database **acquisition processes are poorly or not at all controlled** (web, apps, smartphones, etc.) and suffer from **significant representation biases**

The first part of the **AI Regulation** ('AI Act', EU regulation) specifically concerns **bias in data**:

*'against discrimination,
and to ensure appropriate governance of AI Systems'*

ARTIST: an **operational response** to the treatment of bias in data

ARTIST - Main Features

- The user wishes to **test the hypothesis that two data sets are distributed in the same way** ('absence of bias or drift')
- In just a few seconds, ARTIST technology calculates the **probability of the hypothesis being true** (p -value) and provides a *confidence interval*
- The solution makes it possible to **quantify** and **represent visually** the presence of bias (deviation of the optimal ROC curve from the diagonal, confidence region)
- ARTIST identifies the **variables explaining the observed differences**
- ARTIST can also be used to adjust (weight) one of the data sets to **mitigate the detected bias** and that the correction is effective.

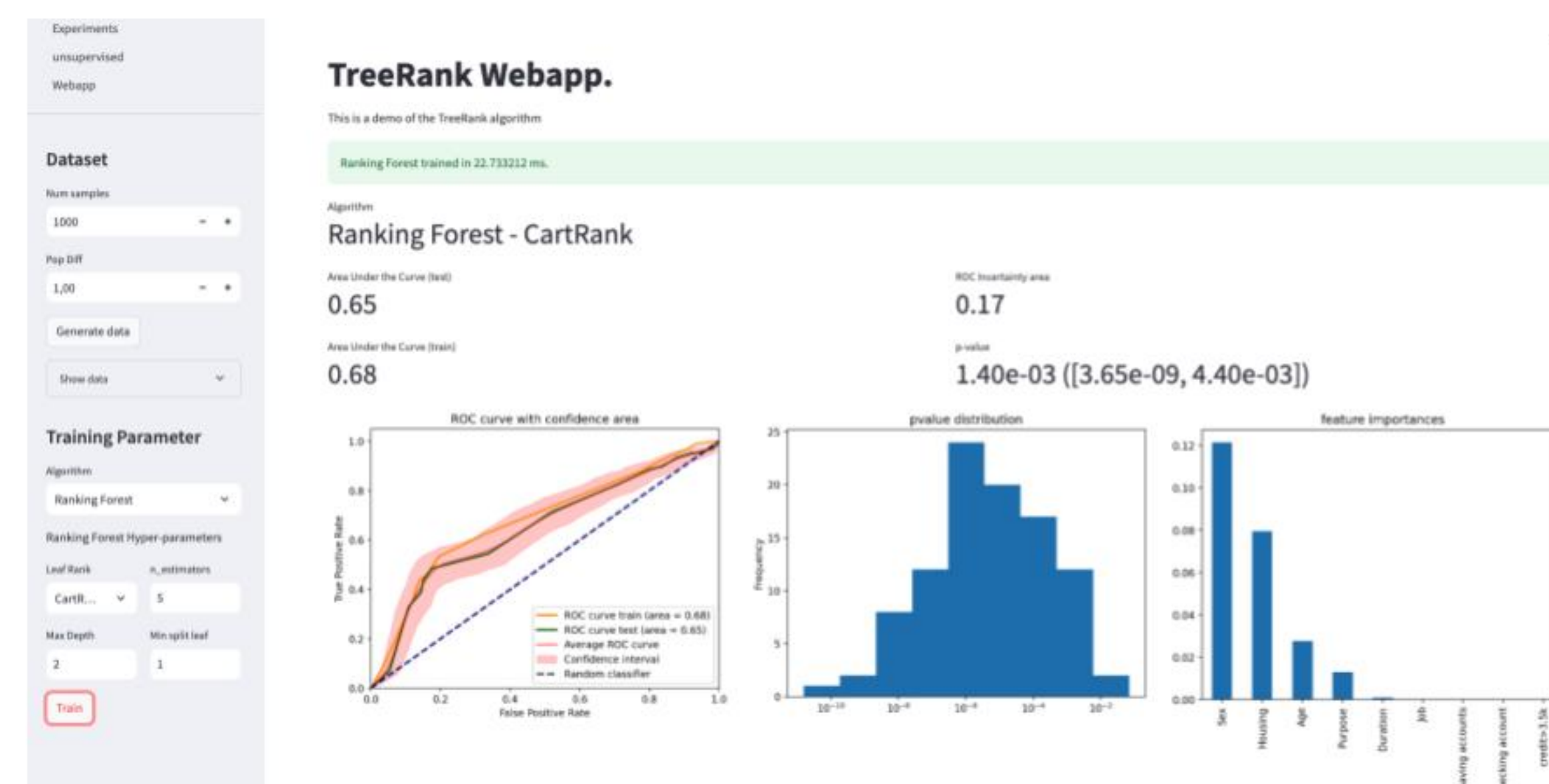


Fig. 2: The Generic ARTIST Interface.

Key Advantages

- **A very simple GUI** to assess the presence of biases in your data for **non-experts** (default setting).
- A GUI **customizable on demand** tailored to your needs by our team.
- A **python library** to access to a complete set of functionalities in expert mode
- **A fast, scalable and frugal** technology
- **Quantifies** the presence of bias in the data (p -value)
- Incorporates **statistical uncertainty** (confidence interval/tube)
- Natively Interpretable (**visualization** of results)
- **Provides explanations**: 'Which variables are responsible for the bias'?
- Allows **bias correction, database enrichment**

Alternatives?

- **Human expertise**, costly and non-exhaustive: project massive data in 2D, 3D and find differences by hand
- Use **rigid metrics** (i.e. not learned by an algorithm) to quantify distribution deviations? Not operational: no table giving p -values (unless data are e.g. Gaussian), no explainability of observed deviations
- **Ignore biases in data and accept the consequences!** Compromised predictive performance, regulation, distrust, ...

Additional Functionalities

- (un)supervised anomaly detection
- **Similarity scoring**
- **Fairness assessment**
- **Explainable segmentation** (e.g. risk analysis)

How to work with us

We are a winning project in the 2023-25 IP Paris maturation competition. (level TRL 3-4 software), reaching 5-6 TRL levels and the commercial start-up phase by EoY '25. In the meanwhile, to start a pilot or project :

- **Directly**, with licence from Telecom Paris and at first the start ups implied in the development (Polygon.AI, AI-vidence), then the forecoming dedicated structure.
- **Indirectly**, with : either the partner start ups (Polygon.AI, AI-vidence), or your favorite consulting partners

References

- [1] S. Cléménçon and N. Vayatis. "Tree-based ranking methods". In: *IEEE Transactions on Information Theory* 55.9 (2009), pp. 4316–4336.
- [2] Stephan Cléménçon, Marine Depecker, and Nicolas Vayatis. "Ranking Forests". In: *Journal of Machine Learning Research* 14.2 (2013), pp. 39–73. URL: <http://jmlr.org/papers/v14/clemencon13a.html>.
- [3] Stephan Cléménçon, Myrto Limnios, and Nicolas Vayatis. *A Bipartite Ranking Approach to the Two-Sample Problem*. 2025. URL: <https://arxiv.org/abs/2302.03592>.
- [4] Stephan Cléménçon et al. "Learning from Biased Data: A Semi-Parametric Approach". In: *Proceedings of the 38th International Conference on Machine Learning*. Ed. by Marina Meila and Tong Zhang. Vol. 139. Proceedings of Machine Learning Research. PMLR, 18–24 Jul 2021, pp. 803–812. URL: <https://proceedings.mlr.press/v139/bertail21a.html>.
- [5] *Patent US 9235805 B2: Method for providing with a score an object, and decision-support system.*
- [6] *Patent WO 2012032118 A2 Méthode d'octroi d'un score à un objet, et système d'aide à la décision.*