

Fairness Explorer: An Interactive Machine Learning Tool for Bias Assessment in Recruitment

Simon Iyamu Perisanidis, Tom Wartmann, Vito Pagone, Jannis Widmer
Interactive Machine Learning: Visualization and Explainability FS 2023



visit our platform

Introduction

Unconscious biases in recruitment have become a pressing concern, as they can lead to unfair and discriminatory hiring practices.

To address this problem, our project focuses on developing an interactive machine learning web application that enables recruiters to explore and analyze their own unconscious biases during the applicant selection process.

Users and Tasks

The primary users of our web application are recruiters involved in the hiring process. They can use our platform to:

- Investigate past hiring decisions through intuitive charts
- Reflect on biases by interacting with the interface and comparing pairs of candidates
- Provide feedback on counterexamples to fine-tune the model's predictions.
- Review candidates for reconsideration

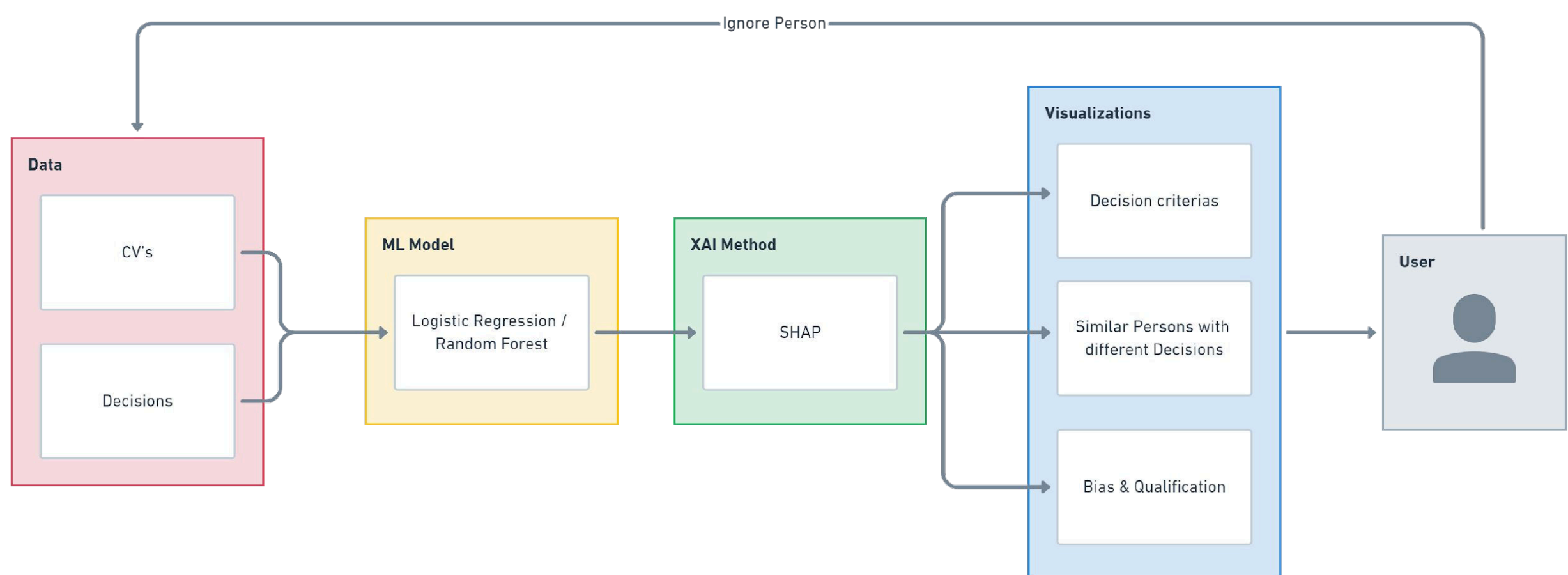
User Interface

The web application's user interface is structured as a dashboard with five panels

- The first row consists of the "Overall Fairness" panel (1) and the "Fairness by Group" (2) panel, offering a quick overview of general fairness scores and biases related to attributes like gender, nationality, and age.
- The next row includes the "Influence by Group" panel (3) and the "People to Reconsider" panel (4), allowing recruiters to explore the impact of different applicant properties and review individuals who might have been influenced by bias
- The last row features the "Fairness Explorer," (5) a dynamic visualization that clusters decision points by demographic groups, enabling interactive exploration

Machine Learning

To measure potential biases of past hiring decisions we experimented with two approaches: Logistic Regression and Random Forests together with SHAP to foster interpretability. Furthermore, to provide the user with counter examples of similar candidates, we relied on TF-IDF and cosine similarity.



Interaction

Since the model's understanding of biases might be limited, we want to foster collaborative assessment together with the user. Therefore, our platform enables users to interact with the machine learning model by choosing to "ignore" candidates they believe were treated fairly. This feedback triggers retraining of the model without the ignored candidates and recalculating the scores.

Discussion

- Recruiters must interpret the interface carefully to avoid any misinterpretations that could exacerbate the negative effects of biases.
- Future work: While this project primarily focused on designing the user interface, further research is needed to develop a robust ML model capable of accurately measuring biases.

