



Python Package Fair Detect

Group C – 2nd Term

July-2022

(PYTHON FOR DATA ANALYSIS II)



Group C – 2nd Term



**HEBA
MUSHTAHA**



**ALFONSO
FERRÁNDIZ**



**GESHAN
COLOMBATANTRI**



**MAXIMILIAN
BECKER**



**NURIA VIVERO
CABEZA**



**JULIÁN GAONA
GONZÁLEZ**

Agenda

Introduction

Fair Detect Package

Improvements done

Improvements for the user

Fair Learn Package

Conclusions



Introduction



OBJECTIVES

The main objective is to understand the package developed by alumnus **Ryan Daher** called **Fair Detect**. Subsequently, we will convert the Fair Detect function into a class and make improvements to it. Finally, we will test the Microsoft **Fairlearn** tool and analyse the differences.

CONTEXT

The use of models using Machine Learning algorithms is widespread in today's society and is becoming increasingly important. The use of these models can lead to the **emergence of biases** that perpetuate, encourage or create **discrimination** between social groups. Being able to identify and correct these biases is in most cases an hard task, because we can find ourselves in situations of **Black Boxes**. In such a situation, tools emerge to be able to detect the existence of discrimination in the models and to correct it. With the aim of solving this problem, our colleague **Ryan Daher** has developed the **Fair Detect package**.

AIM OF THIS REPORT

This report is complementary to the other materials provided and should be understood as a facilitator of the understanding of these materials in the study, but in no way a substitute for them. This report is an introduction to the work done from a high-level perspective. The low-level detail of the work done can be found in the materials enclosed in the package.



STEPS PERFORMED

Firstly, we have carried out a **detailed analysis** of the documentation and material relating to the Fair Detect package as well as the Fairlearn tool.

All classes, methods and functions have been commented using **docstrings** based on the **pep257** convention to facilitate the use of the packages.

All of the above is put together in a package that is uploaded to **GitHub** and **PyPi**.

Analysis

Improving code

Document

Fairlearn

Merging

Regarding enhancements, two classes (**Fair Detect** and **ExtendedDataFrame**) are created. Other improvements are developed to simplify the code and improve the user experience.

A study of Microsoft's **Fairlearn** tool is developed, and its **results are compared** with those obtained in FairDetect.

The FairDetect Package.



The FairDetect package is a code that has been elaborated by the IE alumni Ryan Daher.

In it, Ryan has developed a notebook called "**fairdetect_functions**" that contains a series of functions that will allow us to detect the existence of biases, as well as their analysis, in the predictions emitted by a Machine Learning model. The author's objective is to apply the three core steps to provide robust, and responsible artificial intelligence: **Identify, Understand, and Act (IUA)**, as established by the **HLEG framework**.

To do this, a dataframe is imported and a machine learning model is applied to it, obtaining **predictions**. Ryan will apply two main functions to these predictions: **identify_bias** and **understand_shap**. The **first** is a **grouping of several functions** (representation, ability, ability_plots, ability_metrics and predictive) that analyse and show the possible deviations that may occur in relation to a pre-selected sensitive variable. The **second** allows the marginal contribution of each of the variables to be isolated according to the category taken by the sensitive variable,, applying the **Shapley Value**, (a game theory solution).

Improvements done

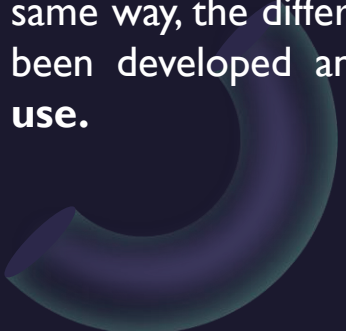


THE FAIRDETECT CLASS

Following the steps provide in the material, we have modified the `fairdetect_functions` script. A **Fairdetect class** has been created and initialized. The functions of the notebook have been transformed into methods part of this class, simplifying the use of the package.

DOCUMENTING THE CODE

Both classes, as well as all their methods, have been documented using **docstrings** that follow the conventions set out in **pep257**, as well as the best practices followed by the community. In the same way, the different notebooks that integrate the package have been developed and commented in order to **facilitate their use**.



THE EXTENDED DATAFRAME CLASS

A new class has been created, **ExtendedDataFrame class**. This class extends the pandas dataframe functionalities. It allows the user, in a simple way, to perform a statistical and visual analysis of the data to be worked with, to prepare the data for the model, as well as to detect potential sensitive variables, prior to applying the model.

Improvements in the user experience

DICTIONARY

We have added a dictionary mapping the following metrics: True Positive, True Negative, False Positive, False Negative, Selection rate, and Accuracy Score. Others can be added also. This makes it easier to edit in case the business requires a different metric assessment.

VARIABLE SIMPLIFICATION

We have created one variable called “metric_frame” instead of using different four variables for the metrics i.e. TP, TN, FP, and FN. Hence, the query can support a variety of metrics and plot as many as needed in the metric_frame dictionary, without adding complexity.

P-VALUE DICTIONARY

We have implemented a dictionary that stores the metric used in the notebook, in this case storing the chi-square for the p-values of TP, TN, FP and FN. This will be highly useful afterward to compare the confidence level and alpha.

CHART PLOTTING

We have improved the code in a way that we are able to plot the chart using the by_group from the metric_frame, to make it more generic. Hence the user can adapt the metrics as needed.

TRY & EXCEPT

We have included a try and except statement which will deliver an error message if the sensitive variable in the data frame is not an integer.

LOOP SIMPLIFICATION

We have simplified the sens_rep and labl_rep loops into just one loop. As well we have simplified the value_counts().sum() function into count() achieving the same result, as can be seen in our project.

Improvements in the user experience



CONFIDENCE LEVEL LOOP

We have implemented a loop that maps the p-values values in order to compare them and decide to Accept or Reject the hypothesis. This way we have reduced the code from 40 to 9 lines.

MESSAGES TO THE USER

We have improved the messages delivered to the user making them easier to understand.

Note: All the improvements are documented in the class while keeping Ryan's code for comparison.

Fairlearn is an open-source by Microsoft, community-driven project to help data scientists improve fairness of AI systems.

[Fairlearn.org](https://fairlearn.org)

The logo features a large blue circle containing a white icon of three horizontal bars of increasing length, followed by the word "Fairlearn" in white text. The background is dark blue with decorative elements: a small 3D cube on the right and a blurred sphere at the bottom left.

≡ Fairlearn

Conclusions

As a conclusion, Ryan's is a very detailed well-structured model to assess disparity in a dataset. However, Microsoft developed a flexible robust pipeline to assess and mitigate unfairness in AI models. I would recommend enhancement on Rayan based on solving fairness problems in machine learning models and also expanding to resolve other types of models as regression.

Note: Our comparison is highly detailed on the Jupyter notebook.






Our packages

FAIR DETECT PROJECT


<https://pypi.org/project/fairdetectteamc/0.1/>

EXTENDED DATAFRAME

<https://test.pypi.org/project/extendeddataframeteamc/0.2/>


 julian.gaona ▼

fairdetectteamc 0.1

`pip install fairdetectteamc==0.1` 

✓ Latest version

Released: less than 20 seconds ago

 julian.gaona ▼

extendeddataframeteamc 0.2

`pip install -i https://test.pypi.org/simple/extendeddataframeteamc==0.2`

✓ Latest version

Released: less than a minute ago

Thank You

Group C – 2nd Term

