

# Fairness using AIF360 Lab

## Introduction

Machine learning models are increasingly used to inform high stakes decisions about people. Although machine learning, by its very nature, is always a form of statistical discrimination, the discrimination becomes objectionable when it places certain privileged groups at systematic advantage and certain unprivileged groups at systematic disadvantage. Biases in training data, due to either prejudice in labels or under-/over-sampling, yields models with unwanted bias.

AI Fairness 360 is an open source toolkit developed by IBM Research, that can help you examine, report, and mitigate discrimination and bias in machine learning models throughout the AI application lifecycle. The goal of this lab is to introduce its bias detection functionalities and help mitigate the bias in the machine learning model using Adversarial Debiasing technique.

For more information see links below:

AIF360 Demo: <https://aif360.mybluemix.net>

AIF360 GitHub: <https://github.com/IBM/AIF360>

AIF360 API Docs: <https://aif360.readthedocs.io/en/latest/>

## Objectives

In this notebook you will utilize AIF360 to detect and mitigate bias on Compas dataset which is used to assess the likelihood that a criminal defendant will reoffend.

Upon completing this lab you will learn:

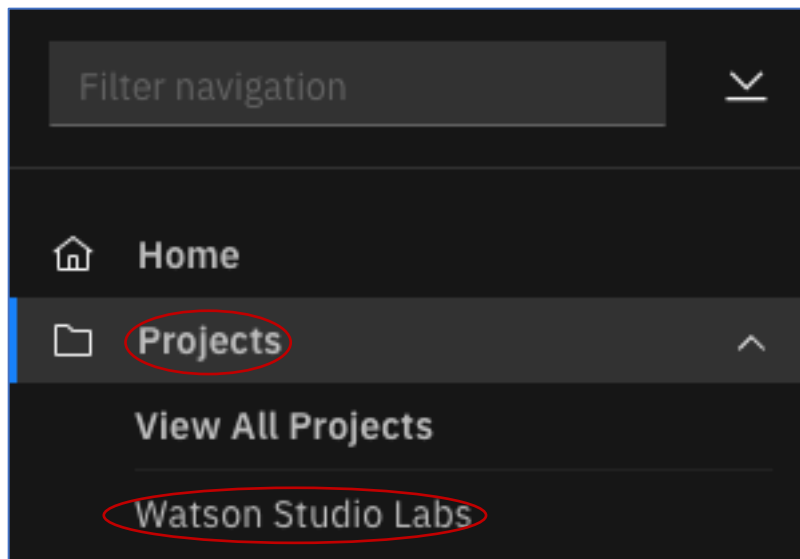
1. How to load datasets from the toolkit package
2. Check the dataset for bias
3. Mitigate existing bias by using Adversarial Debiasing technique
4. Train on both original and corrected dataset and compare results

## Lab Steps

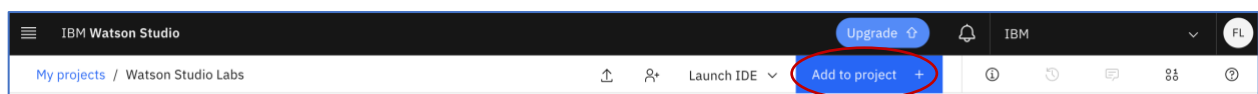
### Step 1 - Create a Jupyter Notebook

1. Click on the hamburger icon , then click on **Projects**, and then **Watson Studio Labs** (or whatever you named the project)

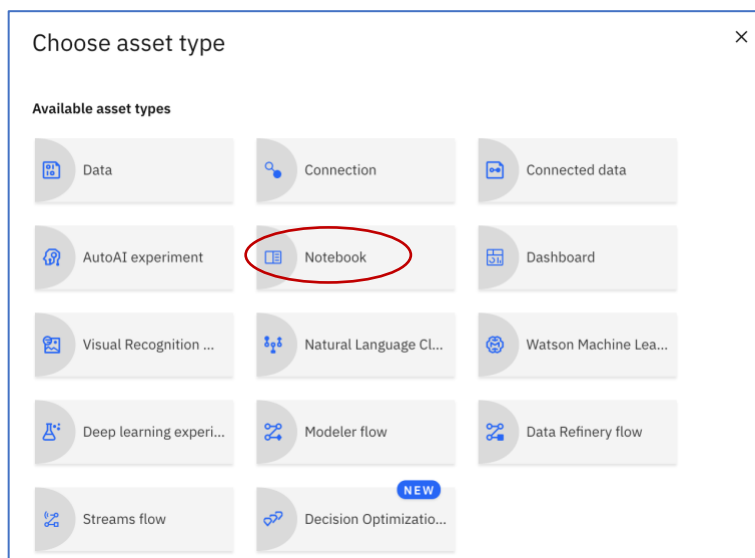




2. We are now going to create a notebook in our project. This notebook will be created from a url that points to the AIF notebook in the github repository. Click the **Add to project** link.



3. Click on **Notebook**



- Click on **From URL** under **New Notebook**, enter **AIF Demo** for the **Name**, optionally enter a **Description**, leave the default for the **runtime**, and cut and paste the following url into the **Notebook URL** field.

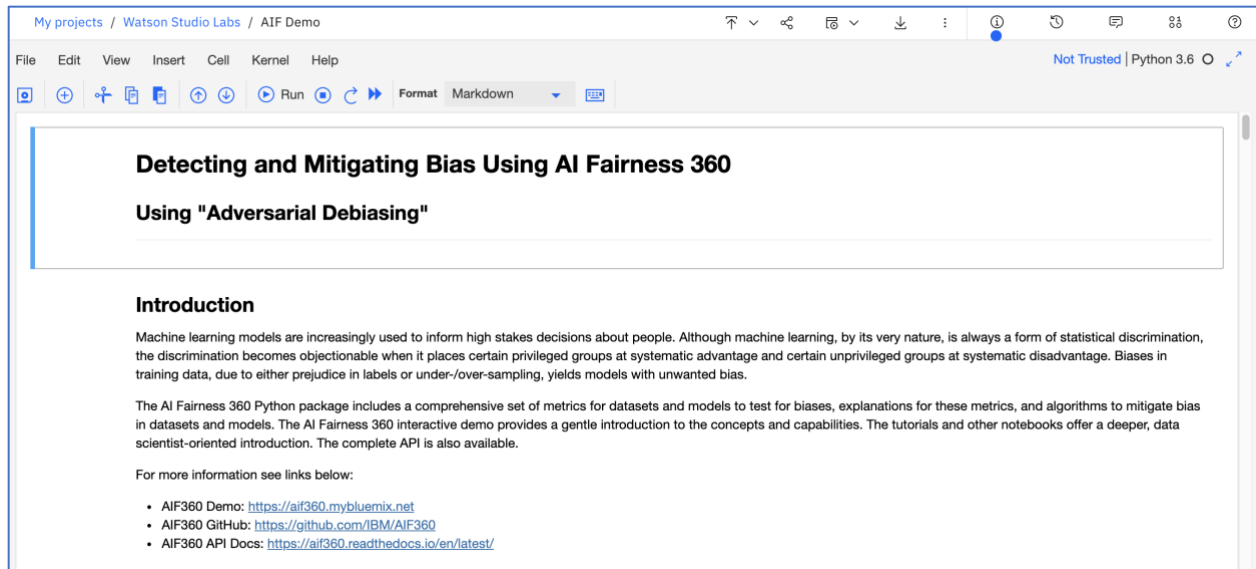
<https://github.com/bleonardb3/Think2020/blob/master/Lab-1/AIF%20Demo.ipynb>

Click **Create**.

The screenshot shows the 'New notebook' interface with the following elements:

- Tab Selection:** Three tabs are visible: 'Blank', 'From file', and 'From URL' (which is selected and circled in red).
- Name Field:** A text input field containing 'AIF Demo', circled in red.
- Description (optional) Field:** A text area with the placeholder 'Type your description here'.
- Select runtime:** A dropdown menu showing 'Default Python 3.6 XS (2 vCPU 8 GB RAM)'.
- Notebook URL:** A text input field containing the URL 'https://github.com/bleonardb3/Think2020/blob/master/Lab-1/AIF%20Demo.ipynb', which is circled in red.
- Buttons:** At the bottom right, there are two buttons: 'Cancel' and 'Create'. The 'Create' button is highlighted with a red circle.

- Place the cursor in the first documentation cell.



6. Execute the code cells in the notebook. For those unfamiliar with Jupyter notebooks, read below.

A Jupyter notebook consists of a series of cells. These cells are of 2 types (1) documentation cells containing markdown, and (2) code cells (denoted by a bracket on the left of the cell) where you write Python code, R, or Scala code depending on the type of notebook. Code cells can be run by putting the cursor in the code cell and pressing **<Shift><Enter>** on the keyboard. Alternatively, you can execute the cells by clicking on **Run icon** on the menu bar that will run the current cell (where the cursor is located) and then select the cell below. In this way, repeatedly clicking on **Run** executes all the cells in the notebook. When a code cell is executed the brackets on the left change to an asterisk **'\*** to indicate the code cell is executing. When completed, a sequence number appears. The output, if any, is displayed below the code cell.