In this module, you will:

* Observe the effects of unfair bias for a real world healthcare use case
* Apply bias mitigation algorithms to reduce bias in the above use case using the AI Fairness 360 toolkit

To be successful in this module, prior knowledge is recommended in:

* Python
* Module 1 : The big picture of Trustworthy AI and Algorithmic Fairness
* Module 2 : Mitigating Bias using AI Fairness 360

The Medical Expenditure Panel Survey (MEPS) is a set of large-scale surveys of families and individuals, their medical providers, and employers across the United States. MEPS is the most complete source of data on the cost and use of health care and health insurance coverage. This data is discussed in the Medical Care Management use case lesson of this course.

Click <https://meps.ahrq.gov/mepsweb/> link to open resource.

This extensible open source toolkit can help you examine, report, and mitigate discrimination and bias in machine learning models throughout the AI application lifecycle. We invite you to use and improve it.

Learn how to put this toolkit to work for your application or industry problem by trying some prebuilt tutorials. Browse through the state-of-the-art bias mitigation algorithms that can address bias throughout AI systems. Are individuals treated similarly? Are privileged and unprivileged groups treated similarly? Find out by using metrics like these that measure individual and group fairness. There are more than 70 metrics in the GitHub repository already. Add new metrics to the repository and use the Slack channel to let the community know about them.

Click <https://aif360.mybluemix.net/> link to open resource.

This project is about explaining what machine learning classifiers (or models) are doing. At the moment, we support explaining individual predictions for text classifiers or classifiers that act on tables (numpy arrays of numerical or categorical data) or images, with a package called lime (short for local interpretable model-agnostic explanations). Lime is based on the work presented in [this paper](https://arxiv.org/abs/1602.04938).

Our plan is to add more packages that help users understand and interact meaningfully with machine learning.

Lime is able to explain any black box classifier, with two or more classes. All we require is that the classifier implements a function that takes in raw text or a numpy array and outputs a probability for each class. Support for scikit-learn classifiers is built-in.

Click <https://github.com/marcotcr/lime> link to open resource.

We highly recommend that you **use Watson Studio** for running the tutorial notebooks as 'AI Fairness 360' opensource toolkit has a lot of dependencies and its easier to use this toolkit on Watson studio rather than any other platform like Jupyter notebook ([website](https://github.com/Trusted-AI/AIF360/tree/master/examples)). **Detail instructions and step-by-step guidance is provided on how to open and run these notebooks using Watson Studio**

|  |  |
| --- | --- |
|  | U,{ef5cad08-9157-481e-8a05-b7b43178f69c}{75},3.125,3.125 |

|  |  |  |
| --- | --- | --- |
|  |  | U,{ef5cad08-9157-481e-8a05-b7b43178f69c}{98},3.125,3.125[Running Course Notebooks in Watson Studio on the IBM Cloud.pdf](https://learn.ibm.com/pluginfile.php/1303760/mod_folder/content/0/Running%20Course%20Notebooks%20in%20Watson%20Studio%20on%20the%20IBM%20Cloud.pdf?forcedownload=1) |

|  |  |  |
| --- | --- | --- |
|  |  | U,{ef5cad08-9157-481e-8a05-b7b43178f69c}{121},3.125,3.125[tutorial\_medical\_expenditure.ipynb](https://learn.ibm.com/pluginfile.php/1303760/mod_folder/content/0/tutorial_medical_expenditure.ipynb?forcedownload=1) |

|  |  |  |
| --- | --- | --- |
|  |  | U,{ef5cad08-9157-481e-8a05-b7b43178f69c}{144},3.125,3.125[tutorial\_setup.ipynb](https://learn.ibm.com/pluginfile.php/1303760/mod_folder/content/0/tutorial_setup.ipynb?forcedownload=1) |

Section 3 and 4 of the “tutorial\_medical\_expenditure.ipynb” notebook discusses Random Forest (RF) classifier as an alternative classifier that can be used for predicting healthcare utilization, with and without bias mitigation. Complete these additional exercises and observe and record the results, then compare your results to the provided solutions document.

|  |  |
| --- | --- |
|  |  |

|  |  |  |
| --- | --- | --- |
|  |  | U,{ef5cad08-9157-481e-8a05-b7b43178f69c}{238},3.125,3.125[Helpful exercise - Questions.docx](https://learn.ibm.com/pluginfile.php/1304182/mod_folder/content/0/Helpful%20exercise%20-%20Questions.docx?forcedownload=1) |

|  |  |  |
| --- | --- | --- |
|  |  | U,{27cff971-03e7-4b4c-8f5a-7ddacdad040c}{6},3.125,3.125[Refer this document only after solving Helpful Exercise - Questions.docx](https://learn.ibm.com/pluginfile.php/1304182/mod_folder/content/0/Refer%20this%20document%20only%20after%20solving%20Helpful%20Exercise%20-%20Questions.docx?forcedownload=1) |

#### Article on Racial Bias in Health Algorithms

Health systems rely on commercial prediction algorithms to identify and help patients with complex health needs. We show that a widely used algorithm, typical of this industry-wide approach and affecting millions of patients, exhibits significant racial bias: At a given risk score, Black patients are considerably sicker than White patients, as evidenced by signs of uncontrolled illnesses. Remedying this disparity would increase the percentage of Black patients receiving additional help from 17.7 to 46.5%. The bias arises because the algorithm predicts health care costs rather than illness, but unequal access to care means that we spend less money caring for Black patients than for White patients. Thus, despite health care cost appearing to be an effective proxy for health by some measures of predictive accuracy, large racial biases arise. We suggest that the choice of convenient, seemingly effective proxies for ground truth can be an important source of algorithmic bias in many contexts.

* Link to article - Dissecting racial bias in an algorithm used to manage the health of populations: <https://science.sciencemag.org/content/366/6464/447>