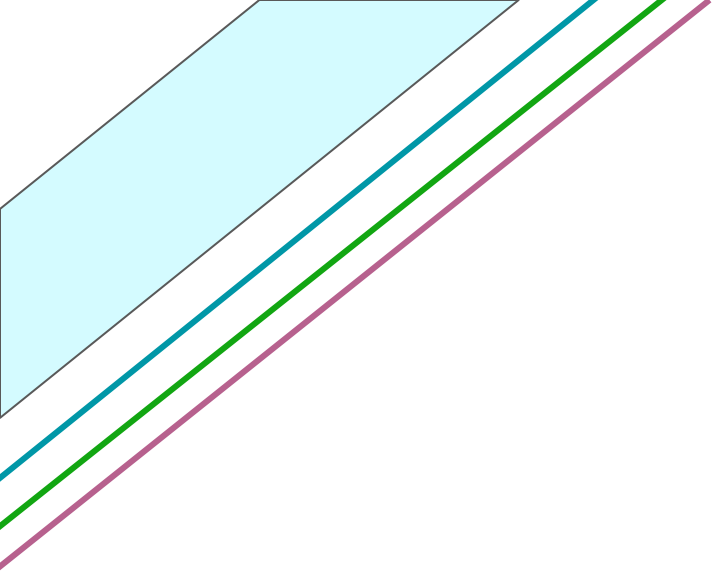


A Programmer's Guide To Inclusive AI

By Marie Tessier

Data

Data Collection - - - - Data Pre-Processing

Collection

1. ### Define the data requirements.

   This will prevent ambiguity during data collection.. Identify what biases may arise in the data. Identifying the possible biases within the targeted data subjects will help to prevent them.

2. ### Work with a subject-matter expert.

   An expert will understand what biases may emerge within the data and what groups may need extra attention.

3. ### Only collect data from data subjects who have given informed consent.

   Ensuring the data is collected ethically with consent and transparency helps mitigate bias and prejudice.

Document the process used to collect the data.

Documentation will help current and future developers catch possible unfairness in the data collection process.

Pre-Processing

1. **Identify any possible sensitive variables and their proxy variables.**
   Once identified, mark are sensitive or proxy variables. Do not use these variables to directly train the algorithm with.
2. **Within these sensitive variables check if there are any that are underrepresented or overrepresented.**
3. **Identify is the time frame of the data is current enough to provide an accurate representation.**
   When using historical data, such as employment data, it can hold past structural inequalities.

Sensitive Variables	Possible Proxies
Gender	➢ Education Level ➢ Income
Race	➢ Zipcode ➢ Socioeconomic status ➢ Criminal record
Disabilities	➢ Personality Test ➢ Level of Education

In- Processing Algorithm

Sensitive Variables - - - - Test Accuracy and Fairness

“Protect” sensitive variables and proxy variables

Sensitive & Proxy Variables

Just omitting sensitive variables may not always be the solution, this is because proxy variables will perpetuate the unfairness and decrease the accuracy of the overall algorithm. Although temporarily omitting sensitive variables can be useful when transforming the data can help to identify the proxy variables. Another approach is blinding to also help identify problematic variables.

Using fair regression will help minimize the loss of the function of actual and predicted values. At the same time this method will help guarantee fairness. This method will help apply multiple metrics to be used for general regression models. With fair regression, you can use propensity modeling and additional constraints to mitigate biases in linear regression while at the same time have the mean target variable between groups approach zero. This approach is suggested and can also be applied in post-processing.

Sensitive Variables - - - - Test Accuracy and Fairness

Pre-Processing

Test the accuracy of the classification algorithm to the confusion matrix

The matrix will help to distinguish underlying differences between groups and highlight biases.

		Actual	
		Positive	**Negative**
Predicted	**Negative**	True Positive	False Positive
	Positive	False Negative	True Negative

Confusion Matrix

The Confusion Matrix is used to calculate the true positive rate and false positive rate which is donated by the following equation:

$$TPR = \frac{TP}{TP+FN}$$

$$FPR = \frac{FP}{TP+FN}$$

This true positive rate can be used to compare the true positive rates for different groups. When the TPR is the same across different groups, the algorithm is considered fair under equal opportunity.

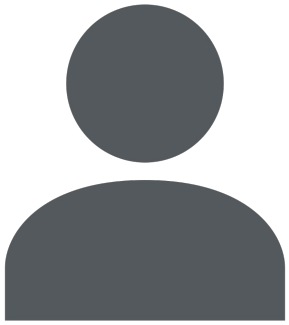
Accuracy

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

The accuracy can be compared between all groups alongside the FPR and TPR to find the most fair and accurate model.

Users

Post-Processing



1. **Apply transformations to the model output.**
   To be used for improving prediction fairness. This is a flexible phase and does not have to release the whole pipeline process.
2. **Accessible to a vast representation of users.**
   Make sure the program is usable for minority or underrepresented groups without discrimination.
3. **Provide transparency to users.**
   The user should easily understand the overview on how and why the ML model is making decisions that are affecting them.
4. **Test the model to ensure it is not making bias or prejudice decisions.**
   This could be done during a testing period with a diverse group.

End Notes

- Return to the training dataset on a regular interval to ensure that the ML model is performing as predicted and not learning new biases.

- Some Projects exist to help provide assistance in analyzing the fairness of ML models.