Model Fairness & Transparency

Description:

Image you are a data scientist in a company, you train a model that has a great performance. When you want to push the model to production for decision making, are you able to show your user this model is reliable and make your users to trust your model? The performance matrix cannot tell you and your users that your model is trustable and reliable. For example, will you believe that you have cancer just because a medical model which has 85% accuracy tells you so ? Aren't you curious about any further explanation or do you wonder if the model has bias and makes mistakes ? To make a good model, model fairness and transparency are essential. Currently, they are major concerns in the real application in every industry. You, as a data scientist, not only want to build a good performance model but also want to spend effort in solving concerns around the model and make end users trust your model.

There are two topics:

Model fairness: The model can do good only if it can provide fair insights. When your model has bias on certain features, can you find it ? Can you remove the bias from the model ? Another important issue is that if you don't solve the bias in the model and the model keeps collecting bias data, your model will become more bias and eventually go further against your original goal and cause unpredictable damage.

Model Transparency: How can you explain your model prediction ? You may think we have feature importance for tree models. Do you really think that can explain the output well to the end user ? Put your feet into your users' shoes, are you curious about why the model predict a 1 or 0 for a certain case? For each individual prediction, what is the supportive evidence ?

There are a couple of tools you can research: IBM AI Fairness 360, FairML, Lime, SHAP and Google What-If Tool.

In this project, you can use any public dataset and choose a machine learning model. Improve the model by improving its fairness and add model prediction explanation. You can try out the tools mentioned here or explore your own tool or method.

Model Fairness & Transparency are essential in terms of model adoption. Many startups are working on this topic. You will be stood out as a data scientist or can be an entrepreneur once you successfully complete this project.

Dataset:

Public machine learning dataset

Contact Person:

yuyi.zhou@technicalsafetybc.ca

Contributor of the Project Idea:

Yuyi Zhou

Data Scientist

Technical Safety BC

Who We Are and What We Do

Technical Safety BC is an independent, self-funded organization that oversees the safe installation and operation of technical systems and equipment across the province. In addition to issuing permits, licences and certificates, we work with industry to reduce safety risks through assessment, education and outreach, enforcement, and research.

Our Vision

Safe technical systems. Everywhere.