Sreeti Ravi

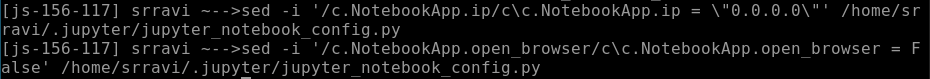
AI Fairness

Installing Anaconda

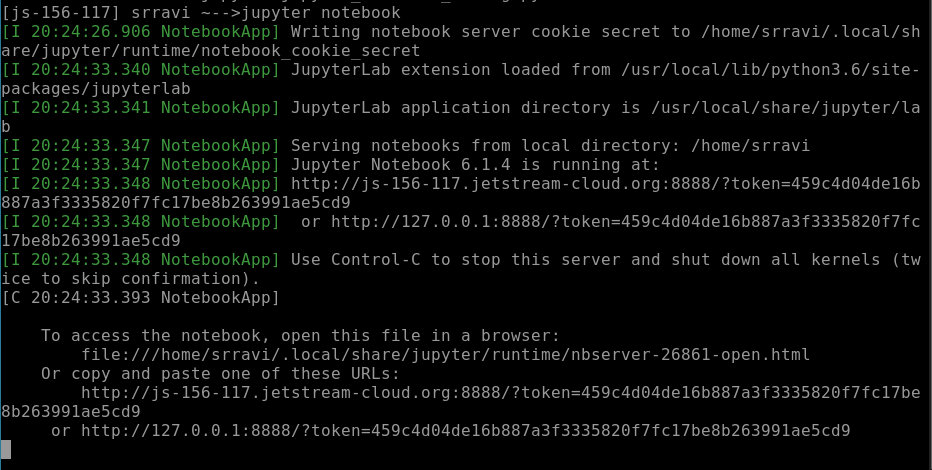


Preparing Jupyter Notebook

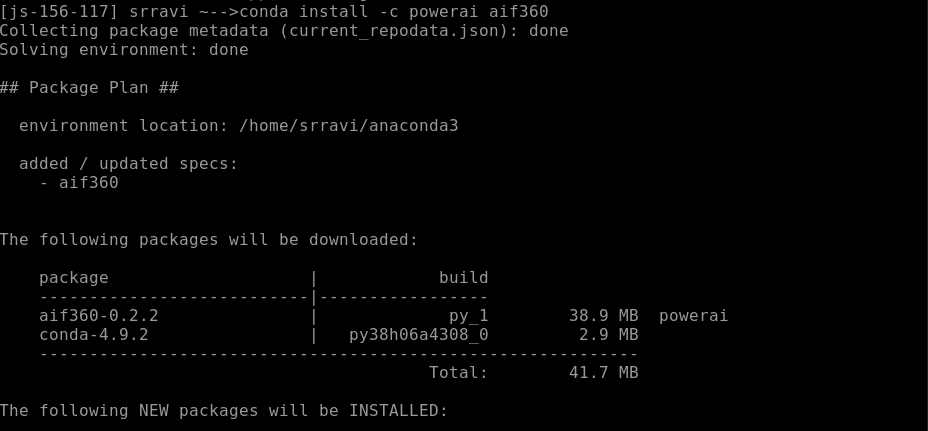


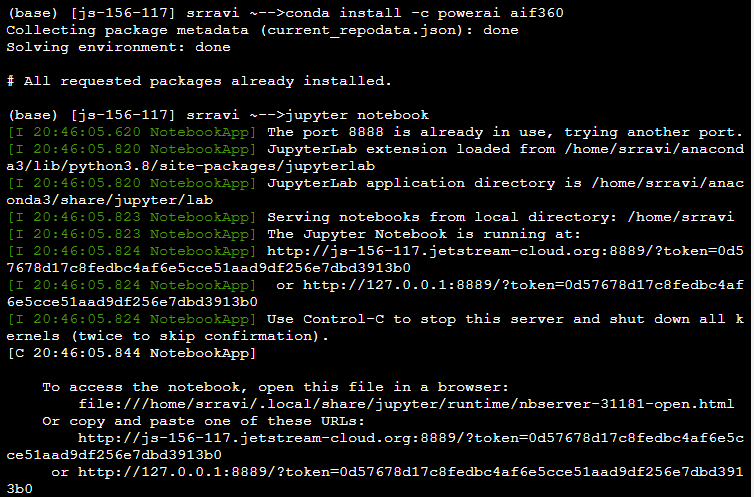


Starting Jupyter Notebook



Running Jupyter Notebook







Using a different column for detecting and mitigating bias



The learning model makes predictions based on a particular instance. In this case, the model uses the training dataset to make predictions. A machine learning algorithm finds patterns in the training set to use when a prediction for a new instance is needed. This can help address fairness metrics and bias mitigators and help find patterns that should not exist. Once biases are identified, there can be steps taken to mitigate them. This could be useful in the workplace or really any area of daily life where people are treated differently due to biases.

For this assignment, I used a Jetstream VM from a previous assignment and installed Anaconda on it. After the VM was launched, I started Jupyter Notebook. I installed the AI Fairness 360 toolkit and even ran a script for the numba package because I was getting error messages. I went through the assignment following the provided code and went through it using the column “sex”. I defined the protected attribute name as ‘sex’ and set the privileged class to ‘male’. I kept 70% of the original dataset as the train set and 30% as the test set and created two lists, one for sets of privileged groups and another for sets of unprivileged groups. To detect bias in this group, I subtracted the percentage of favorable results for the privileged groups from the unprivileged groups. The negative value I got was -0.087. This means that the privileged group, male, was getting 8.7% more positive outcomes in the training dataset. Because this is not desirable, I was supposed to mitigate the bias using the Reweighing class, but I kept getting error messages and I could not figure out what the issue was. I tried ‘conda install numba’ and it said the package already existed. I tried starting over a few times and had the same issue every time. If it was working, the Reweighing class would have transformed the dataset to be more equal in positive outcomes for both groups. After doing that, the difference in mean outcomes should have been 0.