CM2203 – Oliver Hancock – Portfolio 3

Task 1

I believe one of the main reasons that this classifier is producing biased results id due to the training data being used to make its classification, a form of sample bias or unintentional prejudice bias known as data heterogeneity. Specifically, when looking at the data we can see that there is only a single race of humans in the dataset, specifically darker skinned individuals. This skews the classifier as it is not trained for data that includes a broad range of skins tones like the testing data being fed for classification. This is also the case for the gender of the people in the training data as there is 75% men and only 25% women in the human category. In both cases this would be acceptable practice if the data being tested from this classifier was broadly the same as that which was being used to feed it, however this is not the case and using more accurate training data, relative to the testing data would, I believe, create a far less biased output from the program.

Task 2

This can be solved by creating guidelines and rules for how data is collected and processed before either training or testing a classifier. These all fall under the umbrella of the 'Social Responsibility of AI' [1] framework. One of the ways we could combat the bias that is present in our current algorithm is through using a much larger dataset, this would implicitly create greater variance in the overall dataset leading to a reduction in bias. However, it would also result in greater accuracy of the resulting classification, with only processing time and storage space being the downfall of this improvement, definitely a net positive for the classification. This is discussed in the paper with the example of the US ornament using only 7 known terrorists in its AI to find them, leading to horrendous results. Using large scale data leads to the solving of one of classifications biggest issues finding attributes that are predictive but uncorrelated as this which is present in our current classification in task1 due to heterogeneity in the training data. Another improvement under the social responsibility guidelines is the transparency of the algorithm itself. This involves better understanding of the training data, data collection methods and details of what the algorithm itself is trying to solve for. Obviously, in this coursework it is different as we had a task to solve but in real world use this has huge impact. One of the most interesting elements in the paper is its discussion of the *Why* when developing an AI classification. Machine learning an AI is most powerful when the output of the program solved a direct goal that is easily explainable and the steps from problem to completion of the model all follow tightly to what is being measured. All too often systems can go astray from the why of the task leading to a lessened understanding of the processes as a whole and resulting in classifications that do not solve the original problem.

P.S. Thank you Sylwia for your great lectures and for making this module much more interesting

References

[1] L. Cheng, K. R. Varshey and H. Liu, "Socially Responsible AI Algorithms:Issues, Purposes, and Challenges," *Journal of Artificial Intelligence Research* , vol. 71, 2021.