Algorithmic Fairness and Bias Notes

Sunday, July 9, 2023 23:39

- How can we use algorithmic fairness to mitigate bias
- Use Al and ml to transform a current biased data set into a more objective solution
- Representational harm: when Al and ml system amplifies or reflects negative stereotypes about a particular group.
- Opportunity denial: when Al and ml system negatively impacts individuals access to opportunities resources, and overall quality of life
- Disproportionate failure - experience of interacting with system is disproportionately failing for particular groups
- is the model doing good things or bad things to people
  - If your model is sending people to jail, maybe, be better to have more false positives than false negatives
  - If your model is handing out loans maybe better to have more false negatives than false positives
- False positives might be better than false negatives
  - False positive: some thing that doesn't need to be gets blurred
  - Bad for surveillance applications
- False negative: something that needs to be blurred is not blurred
  - Can be a bummer for protecting against identity tleff
- False negatives: spam email that is not caught, so gets delivered to your inbox
  - Annoying at minimum
- False positive: email flagged as spam is removed from your inbox
  - If it’s a highly sensitive/ time-critical email, could be disastrous
- False positive - test result incorrectly indicates the presence of a condition
- False negative- the test result incorrectly indirectly indicates the absence of a condition wen it is actually present
- Accuracy rate: measurement is used to multiply it with the results, rewarding the model for providing high confidence values for its correct assessments
- Accuracy refers to the closeness of a measured value to a standard or known value
- Precision refers to the closeness of two or more measurements to each other