Machine Learning
IST 402, Week 8

Samantha Grossman, Wesley Lo, Joseph Han

Big Data vs. Machine Learning

- Big Data cannot be labeled as fair or unfair
- Machine learning can be labeled as unfair because it takes big data and uses it in an unfair way
    - ex. Machine learning technologies used in the criminal justice system are biased

Types of Harm

- Allocative Harm
  - A system allocates or withholds certain groups on opportunities or resources
  - Ex. A decision regarding who gets a loan is biased
- Representational Harm
  - A system reinforces the subordination of some groups along the lines of identity
  - Ex. Nikon’s camera mischaracterized Asian features

Machine Learning Scientists vs. Social Scientists

- Machine Learning Scientists
  - Create algorithms and use big data to input to those algorithms
  - Focused on creating technology that makes the lives of humans more convenient
- Social Scientists
  - Create theory based upon human interaction and socialization
  - Focused on finding human bias that might be present in society
- Takeaway: Machine learning scientists can benefit from social scientists by considering their bias theories when creating new machine learning technologies

COMPAS System

- COMPAS
  - Correctional Offender Management Profiling for Alternative Sanctions
- Used by US courts to assess the likelihood of a defendant becoming a recidivist
  - Recidivist - person repeating an undesirable behavior after they have either experienced negative consequences of that behavior, or have been trained to extinguish that behavior
- Pretrial Release Risk Scale
  - Measure of the potential for an individual to fail to appear and/or to commit new felonies while on release
  - Current charges
  - Pending charges
  - Prior arrest history
  - Previous pretrial failure
  - Residential stability
  - Employment status
  - Community ties
  - Substance abuse

COMPAS System Con't

- General Recidivism scale
  - Designed to predict new offenses upon release, and after the COMPAS assessment is given
    - Criminal history and associates
    - Drug involvement
    - Indications of juvenile delinquency
- Violent Recidivism scale
  - Designed to predict new violent offenses upon release, and after the COMPAS assessment is given
    - History of violence
    - History of non-compliance
    - Vocational/educational problems
    - The person's age-at-intake
    - The person's age-at-first- arrest.

Importance of COMPAS

- The United States locks up far more people than any other country, a disproportionate number of them African-American. For more than two centuries, the key decisions in the legal process have been in the hands of human beings guided by their instincts and personal biases.
- If computers could accurately predict which defendants were likely to commit new crimes, the criminal justice system could be fairer and more selective about who is incarcerated and for how long.
- Rating a defendant's risk of future crime is often done in conjunction with an evaluation of a defendant's rehabilitation needs.
- The Justice Department encourages the use of such combined assessments at every stage of the criminal justice process.

Brisha Borden vs Vernon Parter

- One possible conclusion: COMPAS is racially biased
- What is flawed in the argument in the initial part of their analysis?
  - Hidden factors
  - Machine error
  - Confirmation bias
  - Very small sample size x2
  - Examples show that African American, who were low risk but were predicted high risk (false positive)
  - Caucasian man, who were low risk but were predicted high risk (false positive)
  - We can only say this confidently if system is able to prove causation

Propublica Points

- COMPAS score correctly predicted an offender's recidivism 61 percent of the time
- COMPAS score correct in its predictions of violent recidivism 20 percent of the time
- African-American defendants were often predicted to be at a higher risk of recidivism than they actually were.
- Analysis found that African-American defendants who did not recidivate over a two-year period were nearly twice as likely to be misclassified as higher risk compared to their white counterparts.
- Even when controlling for prior crimes, future recidivism, age, and gender, African-American defendants were 45 percent more likely to be assigned higher risk scores than white defendants. In violent crime category, African-American defendants were 77 percent more likely to be assigned higher risk scores than white defendants.
- White violent recidivists were 63 percent more likely to have been misclassified as a low risk of violent recidivism, compared with African-American violent recidivists.

Conclusion

- Logistic regression model might be an accurate explanation model since it correctly mimics the predictions of the original model. It would not be faithful to what the original model computes
- Created a linear explanation model for COMPAS that depended on race, and then accused the black box COMPAS model of depending on race, conditioned on age and criminal history.
- COMPAS seems to be nonlinear, and it is entirely possible that COMPAS does not depend on race.
- ProPublica's linear model was not truly an "explanation" for COMPAS, and they should not have concluded that their explanation model uses the same important features as the black box it was approximating.