| DS 100: Principles and Techniques of Data Science | Date: June 24, 2019 |
| --- | --- |
| Discussion #1 | |
| *Name:* | |

# A Big Data Fail

Consider the 1936 federal presidential election of FDR vs. Al Landon. The magazine Literary Digest's straw poll had correctly predicted the outcome of the previous five presidential elections. Running up to the election, they polled over 10 million individuals including

- magazine subscribers

- registered automobile owners

- telephone owners

and received responses from about 2.4 million of those polled. The Literary Digest predicted Landon would win in a landslide. By contrast, George Gallup's quota sample consisted of bi-weekly surveys of 2, 000 individuals, and correctly predicted a landslide for FDR.

1. What are some potential sources of bias in each of these polling schemes?

For Literary Digest's poll, the people they surveyed were disproportionately wealthy and living in urban areas, meaning the sample wasn't truly random or representative of the total population. Additionally, only those felt strongly about the elections would respond to the survey, which again isn't random. Here, it can be mentioned that a large sample size will not lead to an accurate outcome if the data itself is biased or not sampled randomly.

For the quota sample, the issue is that once the quota had been met by the surveyor, they were allowed to also interview people of their own choosing, which meant the sample wasn't perfectly random and still had some bias associated with it

# Data-Driven Study Design: COMPAS Algorithm for Predicting Recidivism

Recidivism is the tendency of a convicted criminal to reoffend. The COMPAS (Correc- tional Offender Management Profiling for Alternative Sanctions) algorithm, developed by the company Northpointe (now equivant), predicts recidivism risk based on variables related to criminal history, drug involvement, and juvenile delinquency. It is used by US courts for the purpose of case management, to predict a defendant's risk of committing more crimes.

2. We will examine the COMPAS algorithm and, in particular, a ProPublica study pointing to racial biases associated with it (https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm). We will discuss general issues raised by the application of such algorithms, in terms of ethics, privacy, security, and governance. We will also walk through steps you might take to address questions related to the accuracy and potential racial bias of the COMPAS algorithm.

The questions are meant to be discussed with the people around you as a group and there is no right or wrong answer.

As mentioned in the question, it should be emphasised that most, if not all of these answers are subjective to some degree

(a) What is the population of interest for COMPAS?

The population of interest is present and past prisoners in the US prison system

(b) What is the imagined utility of the algorithm in contrast to a human judge?

Human beings are always biased, regardless of whether this bias is implicit/explicit and the magnitude of the bias itself. Since the algorithm has no human emotions and is based on mathematics and data, one may perceive the algorithm to be "objective" and unbiased, even though this is not necessarily true (elaborated later)

(c) What are some features or attributes that were used by COMPAS to design the algorithm? Are there features or attributes that you think should've been included or taken out?

Age, income, gender, race, zipcode or postal code of where they were born and live, offence committed

(d) How do you define "accuracy" and "racial bias"?

Here, one way of defining accuracy is the number of correct predictions made by the algorithm in the event someone is released from prison.

Racial bias can be defined as a disproportionate difference in the outcome predicted by the algorithm based on race, which may be based on the design of the algorithm itself, or due to the data being input into the algorithm itself being racially biased

(e) How does the history of criminal justice institutions inform the data used by the algorithm?

Unfortunately, there is a history of institutional and systemic racism in the United States. As such, this means historical and current data will also be racially biased, which a human judge can be conscious of and try accounting for. However, no such protection is offered in the case of the algorithm. As mentioned in the earlier questions, this input of racially biased data will result in the algorithm itself being racially biased, while still appearing to be "objective"

(f) How should data be collected or obtained to assess the accuracy of predictors like COMPAS? Would you sample at random from the population of interest?

When assessing accuracy, one should look at the data points associated with the individual in question. Additionally, one could look at other individuals in the

system with similar data points from past and present, and use this data to better inform our decision

(g)  What are some ways we can assess the accuracy of COMPAS?

Referring back to our previous definition of accuracy, we note that it does not account for predictions made when the offender in question remains in prison. We can also look at past data before the implementation of the algorithm, see what the algorithm would have predicted, and compare it with the known outcome

(h) Think about the concepts of false positives and false negatives in this scenario. What are the ramifications or costs of a false positive and/or false negative?

False positive: In actuality, the recidivism would not have taken place, but the algorithm predicts it would have

False negative: algorithm predicts no recidivism, but recidivism takes place

Here, one can discuss how the cost on an individual and societal level differs for a false positive vs. a false negative. It is undesirable if someone who was thought to be at low risk of recidivism commits a crime again, but it is far worse for a person who wouldn't commit a crime again to stay in jail because of the perception that they would.

(i) Is the COMPAS algorithm fair? For whom? According to what/whose definition of fairness?

The answer is a cumulation of the previous ones. The algorithm is far insofar as there is no individual human bias associated with it, yet the bias in the data itself is propagated in the algorithm. Additionally, the algorithm is without a sense of ethics. Is this really "fair"?