Data Ethics Assignment 3 - Answers

True/False & Multiple Choice (30 Marks)

Week 5

1. B. a concept of fair procedures and a concept of fair outcomes

2. B. disparate impact and disparate treatment

3. False

4. False

5. C. government salaries

6. True

7. False

9. C. when the algorithm makes perfect predictions or the target trait is equally prevalent in all relevant groups

10. False

11. E. membership in a protected class, per se, is very often irrelevant to a decision even if it is correlated with relevant factors

12. C. individual characteristics rather than group membership

13. B. some groups of people are more valuable than other groups of people

14. B. In general, we should only judge people based on traits that they can modify

Week 6

16. True

17. E. misleading messages which mislead by the deceptive or selective presentation of true information

18. False

19. D. distinguishing science from non-science

20. True

21. E. two or more people have conspired to achieve some objective, when such a belief is false and unsupported by evidence

22. D. conspiracy theories appeal to those who are paranoid

23. True

24. False

25. D. a need to feel unique

26. E. the effects of disinformation; the amount of disinformation

27. E

28. False

29. True

30. C. Introduce mechanisms to track censorship and to give those affected appropriate avenues for recourse and appeal

Short Answer (15 Marks)

Week 5:

I. (12-16 Sentences, 9 marks) The lecture mentioned two formal definitions of fairness for algorithmic decision-making: calibration and balance. We saw that these two definitions cannot be mutually satisfied, outside of highly unusual circumstances. Suppose you find yourself in a typical case (i.e., with a less than perfectly accurate algorithm, and differences between two groups with respect to the target trait. Which definition of fairness would you choose to satisfy and why, assuming no other definitions are under consideration? To answer this question, begin by clearly identifying the definition of fairness you think is best (1 sentence, 1 point). Next, describe a hypothetical case of decision-making where you believe that your preferred definition is most clearly better than the alternative (3-4 sentences, 2 points total, .5 points per element). This description should identify (i) the goal of the decision-making process (e.g., to evaluate loan applications), (ii) a target trait to be predicted (e.g., loan repayment), (iii) the input/output of the algorithm (e.g., a loan application/probability of loan repayment) and (iv) two groups who differ with respect to the target trait (e.g., Group A and Group B, men and women, etc.). Next, describe a hypothetical subject of this decision-making process (e.g., a particular loan applicant), making sure to (i) identify which group they belong to, (ii) identify whether they actually possess the target trait, and (iii) provide any ethically relevant details about their situation (3-4 sentences, 3 points, 1 point per element).

Next, explain how satisfying each definition would affect this subject (3-4 sentences, 2 points). Finally, explain why you believe that your preferred definition results in a fairer

decision for your subject (2-3 sentences, 1 points). (Note: You can pick any hypothetical case you like, so long as it is NOT the loan risk example.)

Definition of Fairness in Algorithmic Decision-Making

In a typical case of algorithmic decision-making where there are inaccuracies and noticeable differences between groups, I would choose calibration as the preferred definition of fairness. Calibration ensures that the predicted probabilities reflect the actual outcomes across all individuals, regardless of which group they belong to.

Preferred Definition of Fairness

Calibration is the approach that aligns predicted outcomes with actual results within each group, helping ensure that each person is assessed fairly based on their own data rather than assumptions about the average performance of the group they belong to.

Hypothetical Case Description

Let's consider a system used to assess college admissions applications. The goal is to predict which students will perform well academically during their first year. The algorithm takes inputs like grades, test scores, and background information, and gives a probability score for academic success. Group A includes students from private schools with strong academic resources, while Group B includes students from rural public schools with limited resources. These two groups differ in average academic preparation due to their environments.

Hypothetical Subject of Decision-Making Process

Prince is a student from Group B, the rural school group. He does possess the target trait, as he has consistently scored well in his studies and shown strong academic motivation. Ethically, it's relevant that Prince helps his family on their farm every day after school and still maintains excellent grades, which reflects his time management and dedication.

Effects of Satisfying Each Definition

If calibration is satisfied, Prince's score will reflect his actual likelihood of academic success based on his own profile, giving him a fair chance. But if balance is used, his score might be adjusted downward because of lower average performance in his group, even though he individually outperforms it. This could lead to an unfair rejection from college based solely on group-level assumptions.

Justification for Preferred Definition

Calibration leads to a fairer result in this case because it allows Prince's unique strengths to shine through without being overshadowed by group averages. It supports a merit-based decision that respects individual effort and potential.

Week 6:

II. (6-9 Sentences, 6 marks) To begin, identify a belief (real or hypothetical) that you consider to be an example of pseudoscience or conspiratorial thinking (e.g., that the Earth is flat or that the US moon landings were staged) (1 sentence, 1 point). Next, identify an unsupported auxiliary hypothesis that exemplifies "built-in subterfuge" as discussed in the lecture (1-2 sentences, 1 points). Next, identify an observation or experiment with two possible outcomes, one that counts as evidence in favor of the pseudoscientific /conspiratorial belief and one that counts as evidence against the pseudoscientific /conspiratorial belief (3-4 sentences, 1 point). Finally, explain how the unsupported auxiliary hypothesis could be used to undermine the outcome that counts against the belief but not the outcome that counts in favor of the belief (2-3 sentences, 2 points).

(Note: You can pick any hypothetical case you like, so long as it is NOT flat earth or the moon landing conspiracy.)

Example of Pseudoscience and Conspiratorial Thinking

One example of conspiratorial thinking is the belief that 5G mobile networks are harmful to human health or can spread viruses like COVID-19, even though scientific studies show no evidence for this.

Unsupported Auxiliary Hypothesis

Supporters of this belief often claim that any research showing 5G is safe is either fake or controlled by governments and telecom companies trying to protect their profits.
This creates a built-in excuse to dismiss any scientific disagreement.

Observation or Experiment with Two Possible Outcomes

We could test this belief by comparing the health of people living near 5G towers with those in areas without 5G.
If the people near 5G have more health problems, that would support the belief.
But if there's no difference — or if both groups are equally healthy — that would go against the belief.

Use of the Unsupported Auxiliary Hypothesis

Believers might ignore the negative results by saying the data was manipulated or that the researchers were bribed.
This lets them reject any evidence that doesn't fit their view, while still accepting anything that supports it.
So the belief becomes impossible to challenge, no matter what the evidence says.