Societal Computing Activity 1
Your new job at ScoreCo

Overview

Divide into your groups for this exercise. Read the Introduction, and then **pick exactly one** case study.

Spend the first 5-10 minutes conversing **face-to-face** and **reviewing any related material** from class or this document. Then, create a **new Google Doc** with the questions to answer copy/pasted in and **your names at the top**. **Share this document with the instructors**. **You will submit this document to UBLearns at the end of class to be graded**

Spend the next 30-35 minutes collaboratively addressing the worksheet questions. Your instructor will let you know when to move ahead to the various parts of the worksheet.

Finally, in the last ten minutes of the class period, your instructor and TAs will lead a group discussion where each group will share some of the information from their worksheets.

Aside: if you find this activity interesting and engaging and want to learn more, check out this keynote presentation by the inestimable Kate Crawford.

Introduction

Your team just landed your first job working for ScoreCo (because you were an epic CSE199 team, they hired you as a group). The company discovered that they could use a new machine learning algorithm originally intended to score different Tupperware containers in terms of their marketability to also **score people** according to a bunch of different tasks.

ScoreCo is considered a trendy new start-up and is backed by a bunch of fancy-pants Venture Capital firms. You get paid a lot of money, and they let you bring your dog (or cat, or lemur) to work. Nice.

You're excited to start on ScoreCo's machine learning team, where you will be responsible for building a new **evaluation** pipeline. Your first task is to create a report that will summarize the potential of applying the algorithm to one of three new domains (all based on real-life settings) that the CEO is considering, and for which you now have data.

Case Study

The three areas ScoreCo is thinking of moving into are as follows. **Pick exactly one** to answer the questions below:

1. **Credit Scoring** - A credit score is a score that provides financial institutions with a measure of your likelihood of being severely past due on a loan they provide you. The higher the credit score, the more likely one is to get approved for all kinds of loans, leases, and credit cards. Paying back loans on time and keeping a low balance on credit cards are ways to increase/keep your score high. Other things, such as late payments, might decrease your score. ScoreCo thinks they can use behavioral data they have bought recently from an *online ad company* to create a new, better credit-scoring algorithm. ScoreCo plans to use features that include 1) the average price of your purchases and 2) where you live (based on IP), and will evaluate the model based on other data the ad company has provided on your current credit card debt for one of your credit cards, if you have one.

2. **Social Credit Scoring** - A social credit score is a score that depicts one's value within the community. It is meant to provide incentives to do "good" for the community, such as volunteering to plant trees as well as is quick to penalize its people for causing harm or doing "bad," like committing a crime. ScoreCo thinks they can use a combination of behavioral data, and data from secret surveillance cameras on public streets, to create a new, better social credit-scoring algorithm. ScoreCo plans to use features that include 1) how often you jaywalk (as seen on the cameras) and 2) the websites you visit, and will evaluate the model based on their ability to predict whether or not you commit a violent crime.

3. **Truth Scoring** - A trustworthiness score is a score that defines how likely it is that the content you post online is truthful. It is meant to ensure that recommendation algorithms, e.g. those from Facebook and Youtube that we talked about last week, are less likely to recommend content from untrustworthy users. ScoreCo plans to use data that was shared from Facebook through a data-sharing agreement to create a new way of identifying (un)trustworthy users. ScoreCo plans to use features that include 1) your age, since older people are more likely to share fake news and 2) how often you share content from a list of pre-defined "fake news" websites. They will evaluate the model based on how likely you are to post content in the future that is eventually rebuked by the popular fact-checking websites Snopes.

Worksheet

- **(5 minutes)** Let us first look at this from ScoreCo's point of view.
    - **What other features might be useful for this task?** Provide 3-5 other potentially predictive signals that could be pulled out of the data you're given (feel free to imagine what might be in that data)
    - **What benefits are there to having a machine learning algorithm that could score people in this way, as opposed to having people do it manually?**
- **(15 minutes)** Cool. Now, let's look at this through a Societal Computing lens
    - (Bad) Question (*note: this phrase "Bad Question" is just a reminder of what we covered in Lecture! Don't get bogged down here)*
        - **Who are the people that will be impacted by the implementation of this algorithm?** Name one group of people (e.g. "impoverished people"), and explain how they will be impacted.
        - **Do you think this algorithm will have a positive or negative impact on these people? Why?**
    - (Bad) Data (*same note as above!)*
        - **Name one potential reason why collecting this data might be invasive to people's privacy**
        - **Identify one way in which the data you will use might be biased**
    - (Bad) Evaluation
        - **What is one specific evaluation metric (e.g. F1 score) that you learned about in the data mining module that this algorithm could be evaluated? Explain what that metric would capture**
        - **What is one thing you learned about in this week's lectures that could allow you to evaluate your model beyond these standard metrics like F1, Precision, Accuracy and Recall?**
    - Drawing parallels
        - **Relate the task that you have selected for this activity to one that you have learned about in class. To do so, you can do one of the following, although you are welcome to go beyond these:**
            - Discuss why the questions asked are bad in similar ways
            - Discuss how the models could be biased in similar ways

        - Discuss how the models could have similar negative impacts for certain groups of people.
- **(10 minutes)** Finally, a little more imagination. Let's say you actually begin to analyze the data, and find the following (**focus only on your case study**):
    1. **Credit scoring** The credit score algorithm illustrates a bias towards people of color (POC). Specifically, you find that POC are 33% more likely to receive predatory loan offers, loan offers that lead to a higher delinquency and default rates, compared to white people.
    2. **Social credit scoring** The social score algorithm suggests that those who speak English as a second language (ESL) are not only less likely to be seen as "social butterflies", but are also less likely to get jobs because of their lack of social skills. Specifically, those who confirm that they speak a different language at home, or those who self-describe themselves as having a foreign accent are 40% less likely to get hired over someone who speaks english as their first language, even if they have a PhD from an accredited university in the U.S.
    3. **Truth Scoring** Women are considered less trustworthy than men by the algorithm, even when exactly the same on all other measured characteristics

  Now, let's assume that you know that the boss has already claimed to investors that the ScoreCo algorithm is "fair by any measure we are aware of." Based on this, answer the following questions:
    - **Is this a problem?** If so, why? If not, why not?
    - **How would you approach this situation with your CEO, who believes that fairness is not an important metric to be concerned with?**
    - **Let's now assume that you've told your boss about this issue, and that your boss tells you "we just don't use those measures here," and that if you provided additional information to anyone else, you "could face consequences". What would you do then?**

Grading

Notes:

- You are required to turn on your video while on Zoom. If you do not turn on your video, you will receive a maximum recitation score of 1 point for the day.
- If you arrive at your recitation more than five minutes late, you will receive a maximum recitation score of 1 point for the day.
- If you do not attend the recitation, you are not eligible for any points.
- You must submit your group's response document to UBLearns

Outside of these notes, the grading criteria for this exercise is as follows:

Participation:

- 0 pts: No attendance/participation
- 1 pt: Present, but arrived late or did not participate fully
- 1.5 pts: Participated, the group interacted face-to-face

Results:

- 1 pt: Group answers are correct but obvious or incomplete
- 1.5 pts: Each worksheet question is addressed with at least some detail, and several worksheet questions address concerns not directly drawn from course materials