Ethics in AI, Machine Learning, and Data Science

February 21, 2021

Preliminaries

This assignments' focus is on ethical and societal aspects of AI, Data Science, and Machine Learning. The aim of this assignment is to gauge your understanding of data science methods, in the context of applications that may involve sensitive private data and with societal implications, and their ethical pitfalls. There will be particular focus on your ability to reflect on the methods inherent limitations and potential misuse.

The assignment (10 pts(=points)) is comprised of three parts:

1. Reading material and multiple choice questions (3.5 pts)
2. Case 1: Wearable health-technologies and private health insurance. (3.5 pts)
3. Case 2: AI-enabled human personality prediction (3.0 pts)

Part 1 is **10** multiple choice questions with **5 choices for each**. Each question has **atleast** one correct choice. Each question can give up to 0.35 pts and down to 0 pts.

Parts 2 and **3** are composed of multiple essay questions where the answers should be made:

1. As succinctly as possible (for example, 1 paragraph, 5-6 sentences)
2. Using clear and logical argumentation, based on technical, statistical, and other scholarly merits. Use these observations and arguments to discuss ethical problems.

Overly long answers with lengthy discussions of irrelevant aspects will be deducted a maximum of -0.2 pts per question.

We are aware that some of the topics covered in this assignment may invoke valid emotional reactions. However, the point of this assignment is to make arguments on a technical and scientific basis, and to illustrate how poorly designed statistical and data science models can serve malicious ends – either intentionally or unintentionally. The best way to combat these ethical pitfalls is by systematic and scholarly argumentation.

When building arguments keep in mind the on the aims outline above and the learning outcomes from the course canvas page.

Part 2: Wearable health-technologies and private health insurance

A vibrant start-up – Vivaksa – is developing an app for health and stress monitoring. The team is three young caucasian men in their late 20s early 30s, with degrees in Machine Learning, Software Engineering, and Business. The product aims to collect physical activity data and use this to predict health and stress states.

The business model of Vivaksa is to sell the data-analytics software, predicting the health status and stress of costumers in an online manner to private health insurance companies. The company leverages the broad use among the public of smart and fitness watches which include motion and health tracking sensors.

Vivaksa intends that the insights provided by their app will help health insurance providers to fairly tailor insurance premiums to customers dependent on their health statuses. Vivaksa wants to deploy its product ethically and responsibly, so they ask health insurance companies to allow their customers to opt-out of using their data for price adjustments.

1. At Vivaksa, they have collected training data generated by Vivaksa employees using the employees personal Apple Watches over one full month. Their big dataset covers physically active and inactive times of the day, along with daily blood and saliva tests indicating levels of important biological markers for health and stress. The machine learning model uses the sensory data from smart watches and health trackers – any combination of input from photoelectric pulse wave sensors, barometers, accelerometers, gyroscopes, and orientation sensors – to train a machine learning model to predict the biological markers of health and stress.
   (1.5 pts) How can the training pipeline affect the predictions of Vivaksa's health status prediction algorithm when deployed on the general public?
   (1 pt) Can Vivaksa improve their predictions? if so how? and if not, why?

2. A health insurance company has agreed to use Vivaksa app in a pilot program offering customers with good health predictions a lower health insurance premium. To motivate their customers to opt-in on the program, the health insurance company increases the premium for customers who opt-out of the program. A core motivation for the health insurance company to adopt this technology is to increase fairness. Costumers who are healthier and better at maintaining an active and healthy lifestyle have fewer sick-days, fewer hospitalization, and lower costs for medicine and healthcare for the insurance company.
   (1 pt) Can the roll-out of this pilot program undermine the insurance company's own value of fairness? – explain why or why not.

Part 3: AI-enabled human personality prediction

The new start-up – Selfie2Personality – is developing a mobile phone app that let users give active feedback to an AI system that applies visual filters to their smartphones front-facing ("selfie") camera. They allow their users to use the app for free and to share customized 'personality-matched' filtered selfies of themselves on social media.

The founders of Selfie2Personality were inspired by two scientific papers (see reading material) that claim to be able to provide accurate predictions of criminality and trustworthiness based on photo evidence alone. The Selfie2Personality app augments these approaches to allow for the integration of the information about filter settings. The idea of Selfie2Personality is to generate fingerprints of the users' personality traits, including preferences to certain products, political stance, education and income level, trustworthiness, and criminality. Selfie2Personality claim to be able to generate such fingerprints by using photo and filter settings of users combined with their engagement with posts and products on social media (likes or comments) and news articles they share. This information is valuable for several industrial and public partners including mortgage lending companies or law enforcement.

1. (1pt) What are the limitations of Selfie2Personalitys technology - if any?
2. Following protests in a city that caused significant damages to public and private property law-enforcement is looking for suspects. Analysing social media posts prior to the protests, the law-enforcement identify a pattern of certain political leanings, however without being able to pin-point specific suspects as to who caused the damages. To narrow their possibilities, law-enforcement reach out to Selfie2Personality who agree to cooporate, as they see it as their civic duty to help bring justice in society.
   (2pt) Selfie2Personality share the social media accounts and a psychological profile of users from the city metropolitan area who fit the political stance (as predicted from their app) that law-enforcement found to be sympathetic to the protests. Are there any ethical and privacy concerns with this collaboration and how may they manifest?