Assignment_6

March 3, 2021

0.1 Assignment 6: Ethics in AI, Machine Learning, and Data Science

Group 6:

Name	Contribution
1. Himanshu Chuphal (guschuhi@student.gu.se)	7 H
2. Claudio Aguilar Aguilar(claagu@student.chalmers.se)	7 H

---

1 Part 2: Wearable health-technologies and private health insurance

1.1) How can the training pipeline affect the predictions of Vivaksa's health status prediction algorithm when deployed on the general public?

[Answer] We believe the training pipeline will affect the predictions of Vivaksa's health status prediction algorithm for the following mentioned reasons::

1. *Small training dataset* : The team is ONLY three young caucasian men in their late 20s early 30s. This indicates probably small sample size and not much diversity in the dataset. This would definitely impact the concerned predictions algorithm, which is being trained on a very small and specific dataset. This in not good as the solution/app is intended to be deployeded on the general public.
2. *Missing diversity in the training set* : If the algorithm is not being trained with diverse dataset including factors like different age groups, ethnicity, genders sets, area, education background, food habits etc, the algorithm can make misleading predictions and thereby an inaccurate service to the general public.
3. *Inaccurate Prediction* : Since the service involves to help people's future insurance, it is very curucial to include related dataset for training the algorithm before it can be deployed for the general public.

1.2) Can Vivaksa improve their predictions? if so how? and if not, why?

[Answer] Yes, by addressing the issues mentioned in 1.1), we believe Vivaksa can indeed improve the predication to a great extent. For instance, including large sample dataset and including diverse dataset for training intended for general public.

The other factors that can probably help to improve the prediction can be decreasing the risks of outliers, training the algorithm for longer duration and probably adding more options to the general public related to their health status on daily basis, which can be easily predicted using the diverse datasets and history of medical records of each user. Also, as accuracy is very imprortant in such services, more efforts can be made in testing and training the prediction algorithm.

2.) Can the roll-out of this pilot program undermine the insurance company's own value of fairness? – explain why or why not

[**Answer**] Yes, we believe the roll-out of the program, which is only based on health perspective of just three young caucasian men in their late 20s early 30s and lacking diverse datasets, would NOT be seen as fair by customers. Considering that the core motivation for the health insurance company to adopt this technology is to increase fairness but only comparing to health data of 3 people to decide what is healthy and deciding the premium, which would kind of not fair for all age groups or different background. Also, not everyone uses the gear application actively or correctly, what happens to such customers in terms of their health status. If some users show their health data or the status wrong, let's say more healthier than they actually are, that would lower their premium but is NOT fair for people using the applicatin in the right spirit but end up paying more premium than to those who are cheating on the application. The insurance company needs to seriously consider all these aspects before deciding on thresholds on insurance premium and also clearly define the health status for all age groups by asking for the used technology/algorithm to adapt to diverse data sets to get better predictions for all groups.

2 Part 3: AI-enabled human personality prediction

1)What are the limitations of Selfie2Personalitys technology - if any?

[**Answer**] The limitations of Selfie2Personalitys technology are:

1. Seems like Selfie2Personalitys technology work is still under development and with varying image qualities, we are not sure how reliable will be the predictions from an under development algorithm.

2. Bias toward training sample, the predication of the criminal look will very much inclined toward the training sample used and might give different result if a different dataset was used.

3. Since the app provide predictions of criminality, it is possible that some of the application users can adapt so that they are perceived in a different way. The app needs to be transparent as to how the predictions are made by the used technology. One limitation is regarding the study done by Xiaolin Wu and Xi Zhang where the results indicated that people who smirked a bit(usually own taken pictures) tended to be non-criminal in comparison to people that made no facial expression(for instance a mugshot or drivers license id) which tended to be criminal. The limitation here is that the technology can be manipulated by society to make themselves less prone to be "criminal" by either taking photos that the technlogy will scan as non-criminal.

4. Another limitation is that if the company states openly that the technology base there information on activity in social media then you can also manipulate the way you use social media(for instance not posting nor liking anything).

2)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?

[**Answer**]

Yes

First of the marketing of the app where the company states that it is an app for sharing customized 'personality-matched' filtered selfies on social media which people tend to believe it is but has no clue that their personal data is being shared/sold.

The so called 'Selfie2Personality' filter app is intented to be some filter based fun application for the users, but seems like the collected date is then sold/shared to several industrial and public partners including mortgage lending companies or law enforcement.

In this example the data is being shared to law enforcement but who knows, next time it could be sold to another interested clients.

There is no information regarding the age limit of using the app, which can be a concern depending on the laws of that particular country. Different countries in Europe have different age of consent in relation to GDPR (source), where in some countries the age limit is none and in other from the ages 13-16 years. This means that in some cases the app cannot collect data from children without consent of their parents.

One concern could be that the app recognizes many people from a certain race as likely being violent which would be reflected in the psychological profile and could lead to incorrect suspicions and create social inequalities.