Assignment 5

Find 2 examples of where an application in data science went wrong (excluding the examples on the ethics in data science podcasts shown at Discussing Model Behaviour datascienceethics.com), from an ethical standpoint. Describe what happened, and how the problem could have been avoided (roughly 1000 words). The assignment is optional and can be used to replace a lower mark on one of the first three assignments. The assignment has to be your original work. Submit a pdf of your essay.

Data Science is a part of our lives in more ways than we realize today. Whether it is technological advancements or recruitments and admissions, data science model affects every aspect of our society – admissions to school, who to hire or fire based on their information, work schedules, granting a loan, meeting new people online, ads shown on websites, news and social media directed to a user ‘s account. How does a data science model even get the training data? We often hear about ‘Facebook` trying to steal our data or rumors like ‘government planting devices using vaccines’, but which of these methods are actually accurate and possible? GPS, search history, social media posts most engaged with and cameras everywhere are some examples of methods that contribute towards data collection. There could be different reasons for an organization to collect data – it can be for testing and experimenting to improve their services; or it could be for using the data against the user in a meticulous way. ‘Consent’ to data collection is required and many people can deny giving that consent, but sometimes organizations can collect data and test it without us knowing – this is called as ‘A/B Testing’. These are few of the factors that affect application of data science and end up going in a wrong direction.

Scenario 1: Algorithm Mishap

Any system’s operation and method of procedure can be defined and altered by a few lines of code – “Algorithm”. An algorithm defines a product and its functioning. Any mistakes in the algorithm can completely change the way the product operates. Developing an algorithm requires detailed information on various factors affecting the final product features. This means that if we don’t have enough data collected or if we miss any factor, then it can result in an algorithm mishap.

One such case example includes a big company we have all heard of – Staples. Staples is one of the biggest Retail Companies whose primary sales revolve around office supplies. There are many big companies similar to Staples such as Office Depot and Office Max that are considered as Staples’ biggest competitors.

In order to compete with their competitors, mainly Office Max in this scenario, Staples decided to alter prices for a certain neighborhood. Staples decided to lower prices and have some ‘sale-off’ for their products in the neighborhood closer to Office Max – they settled on this tactic thinking customers will be drawn to their store because of the lower prices. However, during this change in

algorithm, they forgot to consider one factor – the type of neighborhood. Turns out, Office Max is usually set up in well-off neighborhood areas. Conclusively, Staples offered cheaper prices to a rich neighborhood. This attracted a lot of hate for Staples from its other customers therefore leading to losing more of its customers. Staples later clarified that they had not intended for this to happen, it was a simple ignorant mistake.

A simple data science application went wrong for Staples and their business due to lack of understanding the model properly.

Scenario 1: Data exploitation

With the rising data science applications, there is an increasing fear among people of being exposed with all their personal information in front of the whole world. A simple teenage mistake can be the hindrance in their career path later on in life. Similar thing happened in the 2016 American election campaign.

Some Trump consultants from a Data Analytics Firm Cambridge Analytica decided to harvest private information from more than 50 million Facebook users to develop new techniques in order to support Donald Trump's 2016 election campaign. Upon investigation it was learned that the Trump campaign had hired the firm and paid them more than $6.2 million.

Even though there were criminal and civil enforcements actions pursued for the case, however, the millions of people whose data was harvested without their knowing could not receive their privacy back. The data harvesting was not done to harm the users but to use their interests and target their personality in a way to encourage them towards voting for Trump. This can still be considered as manipulating the election.

After all investigations and prosecutions, they said that they destroyed all the harvested data. However, later on Facebook received reports that "not all data was destroyed". This is another scary scenario of data science application gone wrong.

References:

Case Study 1: https://techreport.com/news/24105/wsj-staples-others-adjust-prices-based-on-location/

Case Study 2: https://www.reuters.com/article/us-facebook-cambridge-analytica-idUSKCN1GT02Y