EAS 504 Applications of Data Science - Industrial Overview

Assignment 4

Name - Aboli Rawanhate

UB Number - 50374341

Q1. Discuss with 2-3 examples some ethical, legal and privacy issues that you might need to consider in designing a data science application.

The lecture presented by Jonathan Manes was focused on Law & Ethics in Data Science. In that he mentioned several examples of ethical, legal and privacy issues that we need to consider in designing a data science application.

When the line between right and wrong blurs in the design of a data science application, ethical difficulties arise. It's critical to consider ethical guidelines such as whether the data is valid for us to use, how the data was collected and how biased it is, whether we're in any way jeopardizing the privacy of anyone, and whether the person whose data is being used is aware of it and has consented. Also models should not be biased for a group of people.

Following things we need to consider while designing a data science application:

1. ML model should not discriminate
2. Companies will face legal consequences if their algorithms are biased
3. Model should provide fairness and transparency
4. It should protect person's privacy
5. It should keep their data secured
6. It should notify user if data breaches happen

Examples :

1. Compucredit companies reduce credit limit of users if they spend on marital counseling, purchase at bars/clubs, tire repair or tire shop without knowing them. This comes under fairness and transparency issues as the company is deducting money without their permission. They have the right to know how the system works. Also they should have the opportunity to be heard and request for reconsideration.
2. COMPAS (Criminal Risk Scoring Tool) was used to score criminals. Two criminals were scored incorrectly as one with two robberies who have already been sentenced for 5 years , also shoplifting was scored as 3 out of 10. However, one juvenile who kidnapped a child from the street was scored as 8. Algorithm was highly biased on age. The younger the criminal, the higher the score they will have. The algorithm was discriminating and they have right to get explaination for the decision.
3. Facebook was showing housing advertisements only to particular groups of people based on some factors. It was taking opportunity from other people to be aware of the

housing options available. Facebook was sued for allegedly allowing housing discrimination.

Q2. How can algorithms be potentially discriminatory - illustrate using some of the examples referenced in the talk

Discriminating against a group of people because of their qualities has the potential to be discriminatory. Discrimination on the basis of caste, color, or creed can occur in humans. Algorithms can potentially discriminate with following events:

1. The discriminations can be included into the definition of class variables, training set collections, or labels based on samples or selection bias as those are defined by humans.
2. Even if color, gender, age, and sexual orientation are not included in the dataset, ML/statistical techniques unwittingly incorporate their effect.
3. Oversampling and undersampling of data of particular groups can lead to bias in the algorithm.
4. It's possible that the data we have doesn't match what we're attempting to measure.
5. Because they are substantially correlated with their characteristics, features in our dataset might be 'proxies' for protected qualities.

Examples :

1. The procedure of screening resumes was automated. A program was created to automate the selection of applicants. But that didn't work out. The algorithm apparently became prejudiced towards female candidates after the corporation trained it on ten years of its own recruiting data. The term "women," as in women's sports, would trigger the system to rate candidates lower specifically. The developers tried to address the problem, but the algorithm was still not up to scratch, thus the project was canceled.
2. Many job ads on Facebook illegally exclude women. It threatens the opportunity they should be getting.
3. Users' credit limits are reduced by Compucredit if they spend on marital therapy, purchases at bars/clubs, tire repair, or tire store. Users were unaware of this and they were discriminated against based on their purchases.

Q3. Discuss data privacy issues in the context of the Facebook-Cambridge Analytica example.

When the problem of Facebook-Cambridge Analytica was brought to light, it received a lot of attention.

The integrity of the data and the privacy of its users are the most important aspects of any data-related application. The entire incident centered on Facebook's betrayal of this trust. This controversy occurred during the 2016 presidential election in the United States. Until recently, such scandals have caused us issues.

As a result, Cambridge Analytica's Alexander Kogan conducted a user survey. This was a polling data firm, and the poll was about the elections at the time. This survey gathered information about a person's personal characteristics. However, because it was logged in through Facebook, it also received additional information such as demographics, age, likes, and all other Facebook data.

Not only that, but the data of persons in the person's buddy list was also gathered. Despite the fact that only 2,70,000 persons responded to the poll, the data from 87,000,000 people was utilized. Cambridge Analytica claimed that they had not violated Facebook's privacy policies because no passwords were used or shared, Facebook acknowledged that the users' trust had been violated because the company used the user's profile information, network of friends, and pages or articles liked without their consent. This was then utilized to target users for political purposes. So, despite the fact that the vast majority of respondents were unaware of the survey, their information was compromised and managed by a third party.

This event has affected:

- Privacy of users - Without the approval of millions of Facebook users, personal profile information such as name, birthday, city, and connections was stolen and used.
- Security of user - Security of user is hampered on both individual and national level.
- Trust - The campaign results were suspect since each person's perception was swayed by enticing commercials.

Q4. Describe in the context of data collection, storage and use, some safeguards that are necessary to be in compliance with US privacy laws.

There are regulations that control the privacy of specific records or sectors in each domain, which are based on the fair information practice concept and are enforced by the Federal Trade Commission. Some industry-specific privacy legislation include:

- HIPAA (Health/Medical Records)
- FERPA (Educational Records)
- FCRA, FACTA (Credit Reports/Consumer Reports)
- RFPA, GLBA (Financial/Banking Records)
- COPPA (Children's Online Information)
- VPPA (Video Rentals)
- Privacy Act (Government Records)

Above regulations come from a distinct domain, with differing restrictions on what may and cannot be done with data in specific situations. Generally, these privacy laws cover the following issues:

- Notice - Letting user know which data is being collected and how it will be used
- Consent - Provide user choices about secondary uses of information - opt-out or opt-in
- Access/Correction - The user has access to his or her data and may correct any inaccuracies
- Integrity/Security - obligation to keep information accurate and secure against breach
- Enforcement - Effective mechanisms should be there to enforce these rights.

It is crucial to figure out which of these regulations apply while working for a certain firm in a specific industry. Google has filtered search engine with applicable country or state laws with domains such as - 'google.fr' and unfiltered general domain as - 'google.com'

Q5. Discuss what additional safeguards might be necessary to be in compliance with the EU GDPR requirements

The EU method is different from the United States method, it is intended to include all data handling, including data processors and data controllers. The General Data Privacy Regulation governs any personal data processing of a person (GDPR).

Additional safeguards necessary to be in compliance with EU GDPR requirements are:

- To fulfill contractual obligations with a data subject, or for tasks at the request of a data subject who is in the process of entering into a contract.
- To comply with a data controller's legal obligations
- To protect the vital interests of a data subject or another individual
- To perform a task in the public interest or in official authority
- All uses of information are forbidden unless it is specifically permitted.
- General consent isn't enough. Need affirmative, opt-in. Can't bundle broad consent as a condition of access.
- Rights to access information and to correct it.
- Right to know who data was shared with.
- Requirement to "make it as easy to withdraw consent as it is to give it"
- Right to Data Portability.
- Right to Data Erasure ("Right to be Forgotten")

Q6. MCQ:

1. A
2. D
3. C
4. B
5. A