# EAS 504: Applications of Data Science – Industrial Overview – Spring 2023

*-Lecture by Jonathan Manes*

**Name: Tananki Ranga Sai Saran Rohit**

**UB ID: 50441793**

**UB Email: rangasai@buffalo.edu**

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

In today's world , huge amounts of data is getting continuously generated from various other sources like different social media platforms, browsers, plug-ins, extensions, mobile phones, applications etc… Organizations must be collecting data comes with responsibilities, including following the federal and regional laws, upholding ethical standards, and respecting privacy of their customers and clients.

**Social Media Applications:**

Social media application is one the best example to understand some ethical, legal and privacy issues ,someone needs to consider in designing a data science application. To begin with, what we see in your Facebook feed or Instagram feed or Twitter, it's not just no longer just like the latest thing someone posted, it's the algorithm of social media apps what they wanted to show us in our feed.it also decides on whose posts you see, for how long and calculate your interests and these algorithm gets on refining itself, and based on these the adds will appear on your Instagram feed . As discussed in the video , they're ultimately tuning their system to maximize the amount of time you spend paying attention to their app. Earlier Facebook was targeted many times by users regarding data censoring, data privacy concerns, data breach , user-specific ad targeting . Recently, Facebook had implemented various security measures to protect user data, such as two-factor authentication, end-to-end encryption, and data encryption.

**Government Services and public administration:**

It is important to give people clear information about the data used in forecasting and the algorithms used to make decisions. must ensure that the data used to make the prediction is not biased against certain groups or demographics.few Government immigration officials ask users to submit their social media accounts so they can analyze them and search for information on potential travelers. In china, government is trying to implement a new credit based system for entire behavior analysis of individual (similar to cibil score ) by collecting huge chunks of data. This might be huge obstacle for citizens over privacy rights ,which is so bad . Similarly as discussed in the video ,the entire road trip data will be collected by apps like street bump, which may be misused later.it is important to ensure that the application complies with all applicable data protection rules. This includes obtaining consent from individuals before collecting data, restricting access to personal data and ensuring that data is not shared with third parties without consent.

**Medicine & health care:**

Data science applications must be designed in a way that ensures fairness and avoids bias and also be accurate. This includes ensuring that the data used in the application is representative and unbiased, as well as ensuring that the algorithms used are fair and working effective. As mentioned in the lecture, IBM Watson have prescribed wrong medicines to the patient . The doctor termed this product as a piece of shit.In this field the algorithms and products must be working so effectively, because any small mistakes could cost a life.

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

**Discrimination:**

Treating people worse based on the group to which they belong too,for instance based on sex,race, disability status ,age , nationality, religion, genetic information . Few domains are covered by anti discrimination laws such as employment , housing ,public accommodations , banking , education, etc…Decisions have a disproportional effect on basis of protected characterstic , even if criteria looks

neutral,where criteria are not closely related to a legitimate business necessity. In New Haven, Connecticut, they had an example to follow, or rather advertise. Firemen liked high positions. Lieutenant, captain. And they did a test, a promotional test, for people who wanted to get that promotion. And they designed the test to make it fairer and that they hope it will produce results that don't benefit whites or blacks. Basically, blacks were excluded from the fire service. was a very club affair and they tried to design a test that was fair. So they run the test and everybody who was part of the test that this would recommend promotion except for two people are white. here,everyone of 70 people are white and no black passed the test ,the new heaven university clearly admitted that they were wrong. But if you try to benefit people based on the fact that they've historically been discriminated against and they're seen as criminals, we look at that very carefully, and that's kind of a conversation between our society and the law, the idea, as a colorblind society, that the best way in society in that to advance on the stage is simply never to consider race. Because different types of target variables can have different effects on protected groups, organizations must be very careful at this stage to avoid coding bias. When defining target variables, those target variables and class labels must be selected that are closely related to business needs. Mainly oversampling or undersampling of particular groups must be taken care when data collection and labeling data involves an exercise of judgment,can reflect biases. modeling a machine learning algorithm with Enron Mail data is bad idea because the data in this situation is not suitable for modeling machine learning learning algorithms used to read natural language. There are limitations on system and maybe those won't create like illegal bias or illegal discrimination.

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

Analytica political ad targeting scandal that was revealed after the 2016 election. The scandal involved the collection and misuse of personal data from millions of Facebook users without their consent, by a third-party app developed by Cambridge Analytica. In 2014, if you had taken a quiz online, you probably shared your personal data and your friends personal data with the company that worked for President Trump's 2016 campaign. Cambridge Analytica partnered with a UK based academic, Alexander Kogan, who was using Facebook data for research purposes. Quizzes were sent to around 300,000 Americans which looked innocuous. Political consultancy Cambridge Analytica used a personality quiz app

to collect personal information from Facebook users, including their likes, interests and friends. The program used this information to create psychographic profiles of users that were used for targeted political advertising during the 2016 US presidential campaign. One of the major issues in this scandal was the lack of transparency and user control over their personal data. Facebook allowed third-party developers to access users' data without users consent .The data collected by Cambridge Analytica was used to target political advertising to American citizens based on their data, which could have influenced their voting behavior. In this way misuse of users personal data for political purposes without their knowledge happened. This had led to increased awareness of privacy issues and the need for stronger regulatory initiatives to ensure that companies respect users' privacy rights.

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

In United States, various privacy laws govern the collection, storage, and use of personal information. Compliance with these laws requires the implementation of security measures to protect the privacy of individuals . The regulations that control the privacy of data records in particular industries in various sectors are HIPAA for Health and medical records, FERPA for educational records, FCRA and FACTA for Credit reports , RFPA for financial and banking records , COPPA for Children’s online information , VPPA for Video Rentals and privacy act for government records. Organizations must perform fair information practice principals such as notice (informing user about data being collected), consent, access to the data, integrity ( obligation to keep information accurate and secure against data breaches ), enforcement. They basically define certain uses that are allowed, and they'll typically say that no other use of data is permitted . Few States have additional data protection laws on top of these like Massachusetts has a law that says that that doesn't allow retailers to collect your zip code when you make a purchase.So,that it prevents certain kind of targeted advertising or the collection of information about customers.TransUnion , which is one of the three big credit reporting companies which collects all information about one’s credit history and provide that information to banks and others who might extend credit to individual ,and the Fair Credit Reporting Act says that they can't use that information for any reason except specific enumerated purposes and it lays out the specific purposes

that they can that that these credit reporting agents can use that your consumer reporting information for and marketing isn't one of them. TransUnion shouldn't be in the business of monetizing information,that it collects this sort of sensitive information that it collects about about its customers. So they got sued by the FTC. Organizations must secure and protect user data from data breaches. Effective procedures must be put in place to implement the aforementioned concepts in practice.

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

The way that the EU law is written, it covers both what are known as data processors and data controllers.here, both data processor is like the company or data controller, the entity that's actually sort of holding the data and using it are covered by the EU Data Privacy legislation.Analyzing the data controller is the company that is directing how the data speaks.Any processing information that pertains to person is governed by the GDPR like health , education. The GDPR gives individuals broader rights in relation to their personal data, including the right to access, correct, delete and object to their processing. Organizations must implement processes to promote these rights and respond to data subjects' requests in a timely manner. EU GDPR provides rights to access information and to correct it , right to know who the data is shared with, right to data portability , right to erasure. So if you've consented to your information being disclosed under the GDPR, you're supposed to be able to withdraw that consent just as easily.For example, when you sign up for phone service, they only need basic information to provide the service but not all the other information just for marketing .They can't make it a condition of providing the service that you also agree to allow them to sell your information to marketers which has to be separate, but not bundled ,it should be more user consent way.The GDPR restricts the transfer of personal data outside the EU to countries that do not provide an adequate level of data protection. Organizations must implement safeguards such as standard contractual clauses or mandatory corporate rules to adequately protect personal data transferred outside the EU. Obtaining consent from individuals before collecting data, restricting access to personal data and ensuring that data is not shared with third parties without consent and also erasing the data whenever needed provides more control over their personal information and feel more secure.

Q1: Based on the lecture, which of the following statements is CORRECT

A. Discrimination is defined and identified as disparate treatment and disparate impact.
B. Algorithms are inherently objective and unbiased.
C. Algorithms won't encode bias inadvertently or intentionally when defining target variables and class labels.
D. Loose definition of target variables can help prevent producing discriminatory results.

## ANS:A

Q2: Based on the lecture, select all the CORRECT statements about potential discrimination resulting from the use of Big Data

1. Labeling data involves an exercise of human judgement, which can sometimes reflect biases
2. Bias can occur when oversampling or undersampling specific groups during data collection
3. Treating people worse based on the group to which they belong, is discrimination
4. ML/Statistical processes may inadvertently introduce influence of characteristics like race, sexual orientation, gender even though they aren't specifically encoded in the data set

A. 1,2,3 B. 1,2,4 C. 2,3,4 D. All of them

## ANS:D

Q3: Based on the lecture, which of these is an INCORRECT statement

A. Pregnancy, disability status, genetic information are all commonly prohibited grounds for discrimination
B. Housing, Banking, Education, Government Entitles are all covered by anti-discrimination laws
C. Raw data without artificial modification will not result in discrimination
D. Available data that may not correspond to what you are actually trying to measure is a problem of data collection that can cause data discrimination

## ANS:C

Q4. Based on the lecture, select the CORRECT statement about Laws Relevant to Algorithmic Fairness

A. Opportunity to engage the decisionmaker directly is out of the range of individual fairness requirements
B. FTC, FCRA, GDPR are all laws that are relevant to algorithmic fairness
C. General Data Privacy Regulation (GDPR) only affects businesses that are headquartered in European Union
D. The right to request an explanation of a business decision is not a customer right protected by GDPR

## ANS:B

Q5. Based on the lecture, select all the CORRECT statements about personal privacy/data protection

1. Informing user of what is being collected and how it will be used is a good practice of Fair Information Practice Principles

2. Keeping data safe from unauthorized access is a good practice in Data Security
3. Careful, thoughtful design of systems, paying attention to data sources are all good practices of personal privacy / data protection
4. Asking for general consent for all uses of information is sufficient under GDPR.

A. 1,2,3 B. 1,2,4 C. 2,3,4 D. All of them

**ANS:A**