Data Management

Dr Emma Murphy

Week 9, March 2021

OLLSCOIL TEICNEOLAÍOCHTA BHAILE ÁTHA CLIATH
DUBLIN
TECHNOLOGICAL UNIVERSITY DUBLIN

Overview

- CA2
- Data Bias

CA2

- The aim of this assessment is to apply the data management concepts, methods and ethical theories that we have studied to a practical case study.
- Based on a case study you will create a data management plan and data protection impact assessment by answering the questions using templates that will be provided.

Data Management Plan

- *A data management plan (DMP) is a formal document that describes the data you expect to acquire or generate during the course of a project, how you will manage, maintain and protect it.*
- *You will be given a shortened DMP template where you will need to create data management solutions based on the case study.*

Data Protection Impact Assessment

- *The purpose of a Data Protection Impact Assessment is to determine if the concept of 'privacy by design' is adequately embedded into processes, systems or projects that will affect or bring about high risk or high-volume processing of personal data.*

- *You will be given a shortened DPIA template where you will need to answer questions based on your data management plan and details in the case study.*

CA2

- Some questions in the data management plan and impact assessment can be answered using the details in the case study. However, some aspects of the data management plan will require you to create individual data management solutions.

CA2

- Submission Deadline: Sunday 2nd May 2021

- Published on Brightspace: Monday 29 March

Bias in Data

(Based on slides by John Gilligan)

Headlines

TECHNOLOGY NEWS OCTOBER 10, 2018 / 4:12 AM / A YEAR AGO

Amazon scraps secret AI recruiting tool that showed bias against women

Jeffrey Dastin

8 MIN READ

SAN FRANCISCO (Reuters) - Amazon.com Inc's (AMZN.O) machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

> …the company realized its new system was not rating candidates for software developer jobs and other technical posts in a gender-neutral way… because Amazon's computer models were trained to vet applicants by observing patterns in resumes submitted to the company over a 10-year period. Most came from men, a reflection of male dominance across the tech industry….

https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

Headlines

Support the Guardian

Available for everyone, funded by readers

Contribute → Subscribe →

Search jobs

News | Opinion | Sport | Culture | Lifestyle

Coronavirus World UK Environment Science Global development Football Tech Business Obituaries

Facebook

‘Equity and inclusion’ teams to be assess Facebook and Instagram algorithms for bias

▲ Facebook and Instagram apps on a phone. Algorithmic discrimination has long been a concern of civil liberties watchdog. Photograph: Andre M Chang/Zuma/Rex/Shutterstock

Facebook is launching an investigation into whether its algorithms discriminate against minority ethnic groups, after internal protests forced the company to reassess the possibility that its machine learning systems could have picked up on real-world bias.

Headlines

Algorithms and bias: What lenders need to know

White & Case LLP

Global, USA | January 20 2017

Much of the software now revolutionizing the financial services industry depends on algorithms that apply artificial intelligence (AI)—and increasingly, machine learning—to automate everything from simple, rote tasks to activities requiring sophisticated judgment. These algorithms and the analyses that undergird them have become progressively more sophisticated as the pool of potentially meaningful variables within the Big Data universe continues to proliferate.

In an algorithmic system, there are three main sources of bias that could lead to biased or discriminatory outcomes: input, training and programming. Input bias could occur when the source data itself is biased because it lacks certain types of information, is not representative or reflects historical biases…. An algorithm that inadvertently disadvantages a protected class now has the potential to create expensive and embarrassing fair lending claims, as well as attendant reputational risk.

https://www.whitecase.com/publications/insight/algorithms-and-bias-what-lenders-need-know

Headlines

TECHNOLOGY

Voice Recognition Still Has Significant Race and Gender Biases

by Joan Palmiter Bajorek

May 10, 2019

Research by Dr. Tatman published by the North American Chapter of the Association for Computational Linguistics (NAACL) indicates that Google's speech recognition is 13% more accurate for men than it is for women.

https://hbr.org/2019/05/voice-recognition-still-has-significant-race-and-gender-biases

Headlines

Harvard Business Review

Coronavirus Magazine Popular Topics Podcasts Video Store The Big Idea

ANALYTICS

Hiring Algorithms Are Not Neutral

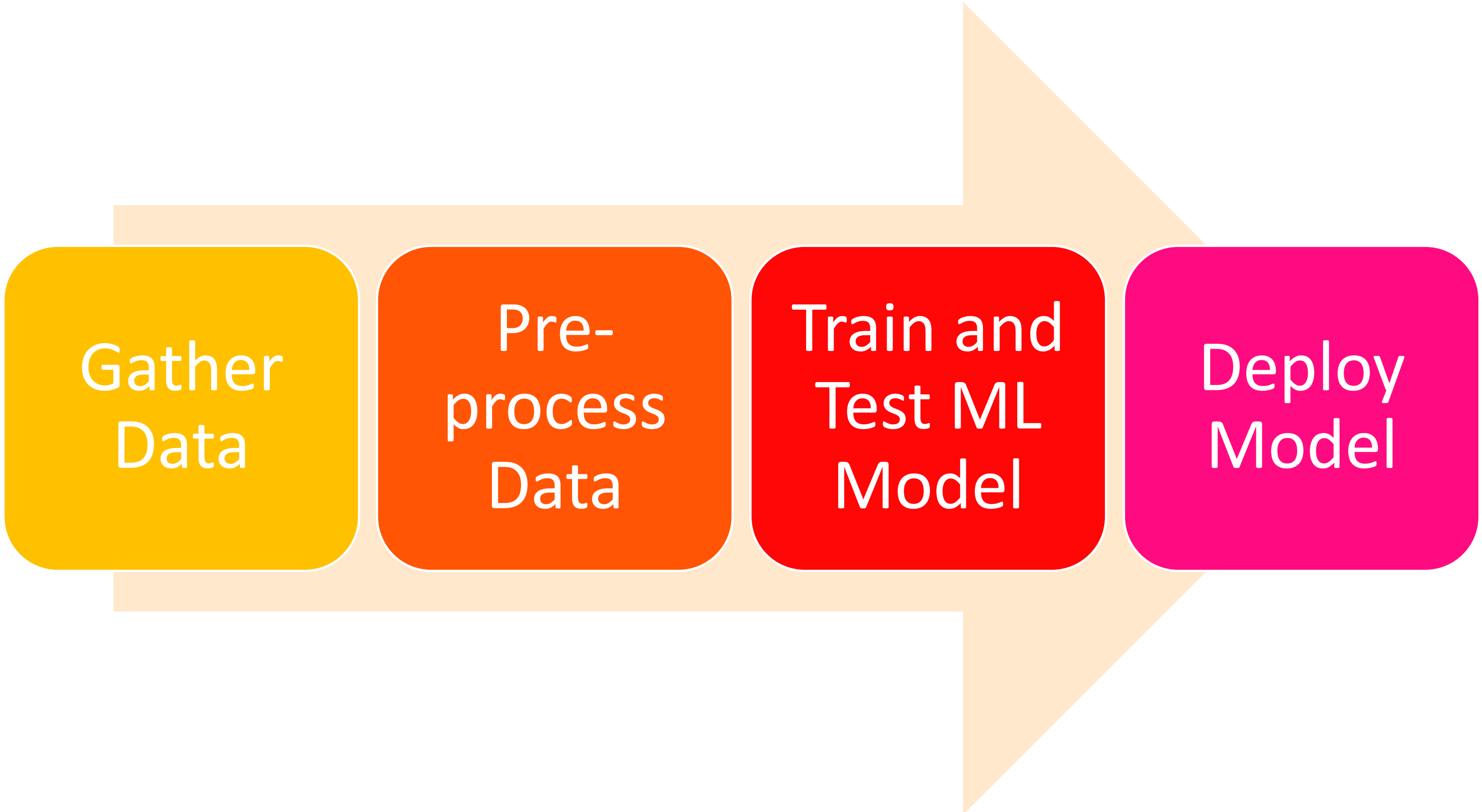
by Gideon Mann and Cathy O'Neil

December 09, 2016

When humans build algorithmic screening software, they may unintentionally determine which applicants will be selected or rejected based on outdated information — going back to a time when there were fewer women in the workforce, for example — leading to a legally and morally unacceptable result.

https://hbr.org/2016/12/hiring-algorithms-are-not-neutral

Stages in typical Machine Learning Application



Data and discrimination

- Machine learning classifiers train themselves on historical datasets
- Any problems with data become problems for the classifiers
- These problems lead to exclusion or poor representation or poor performance especially when dealing with minorities

Areas of Discrimination

Areas of concern where discrimination is possible

- Access to Credit
- Access to Employment
- Access to Higher Education
- Criminal Justice

(Obama White House Report, 2016)

Preventing Discrimination

Recommendations

1. Active Inclusion
2. Fairness
3. Right To Understanding
4. Access to Redress

(World Economic Forum, 2018)

Preventing Discrimination

- **Active Inclusion:** The development and design of ML applications must actively seek a diversity of input, especially of the norms and values of specific populations affected by the output of AI systems.

Preventing Discrimination

- **Fairness:** People involved in conceptualizing, developing, and implementing machine learning systems should consider which definition of fairness best applies to their context and application, and prioritize it in the architecture of the machine learning system and its evaluation metrics.

Preventing Discrimination

- **Right to Understanding:** Involvement of ML systems in decision-making that affects individual rights must be disclosed, and the systems must be able to provide an explanation of their decision-making that is understandable to end users and reviewable by a competent human authority. Where this is impossible and rights are at stake, leaders in the design, deployment and regulation of ML technology must question whether or not it should be used. –

Preventing Discrimination

- **Access to Redress**: Leaders, designers and developers of ML systems are responsible for identifying the potential negative human rights impacts of their systems. They must make visible avenues for redress for those affected by disparate impacts, and establish processes for the timely redress of any discriminatory outputs

Activity

Listen to the Spark podcast and consider the following questions:

What is the issue with edge or “outlier data” and machine learning for people with disabilities?

What lessons on research practice can we learn from this?

https://www.cbc.ca/radio/spark/362-machine-learning-outliers-smart-device-ownership-and-more-1.4279433/ai-s-problem-with-disability-and-diversity-1.4712521

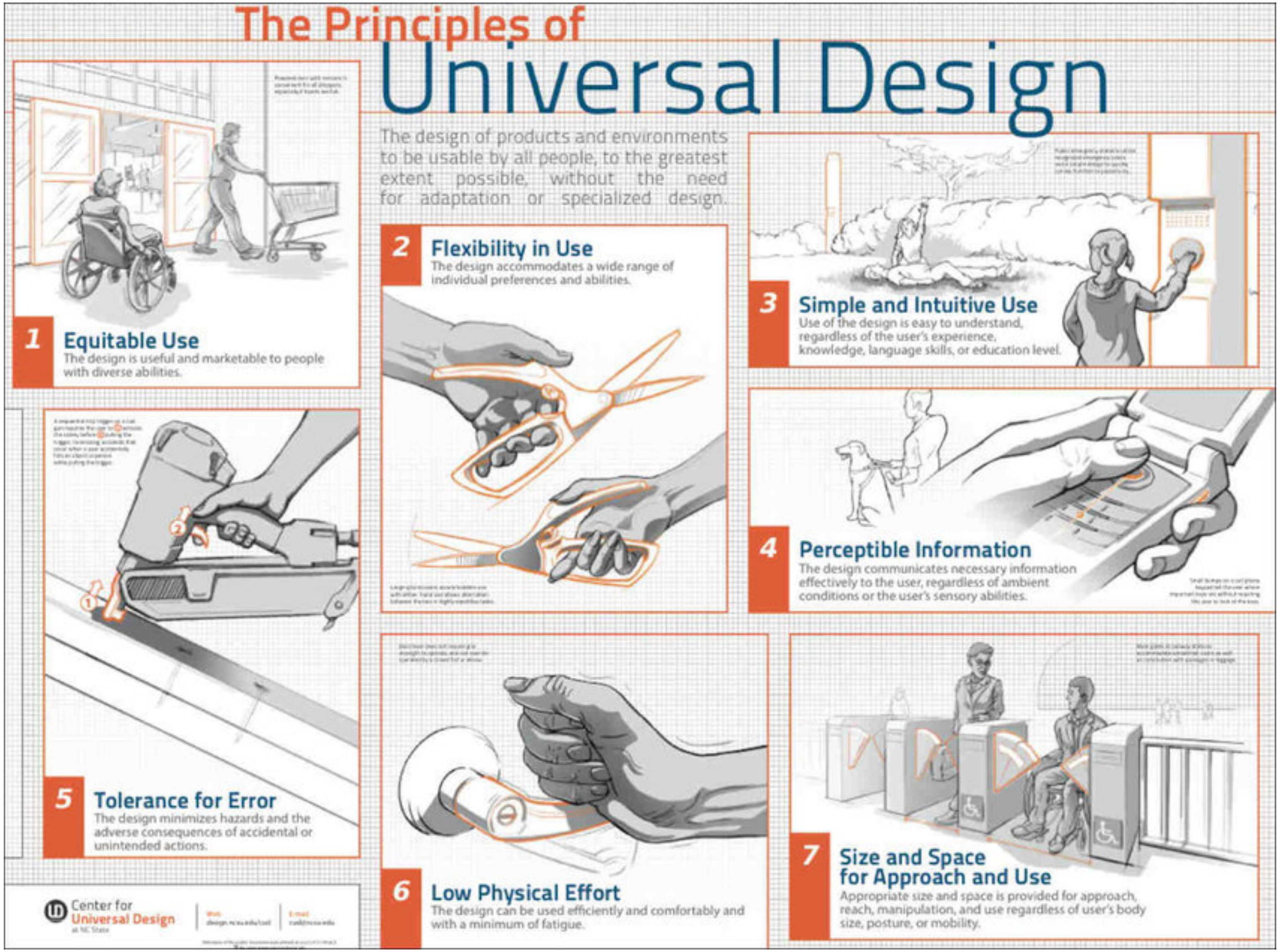
Inclusion

- The Principle of Active Inclusion
- reflects
- The First Principle of Universal Design
- (Equitable Use)

- The 7 Principles of Universal Design were developed in 1997 by a working group of architects, product designers, engineers and environmental design researchers, led by the late Ronald Mace in the North Carolina State University. The purpose of the Principles is to guide the design of environments, products and communications. According to the Center for Universal Design in NCSU, the Principles "may be applied to evaluate existing designs, guide the design process and educate both designers and consumers about the characteristics of more usable products and environments."

http://universaldesign.ie/What-is-Universal-Design/The-7-Principles/

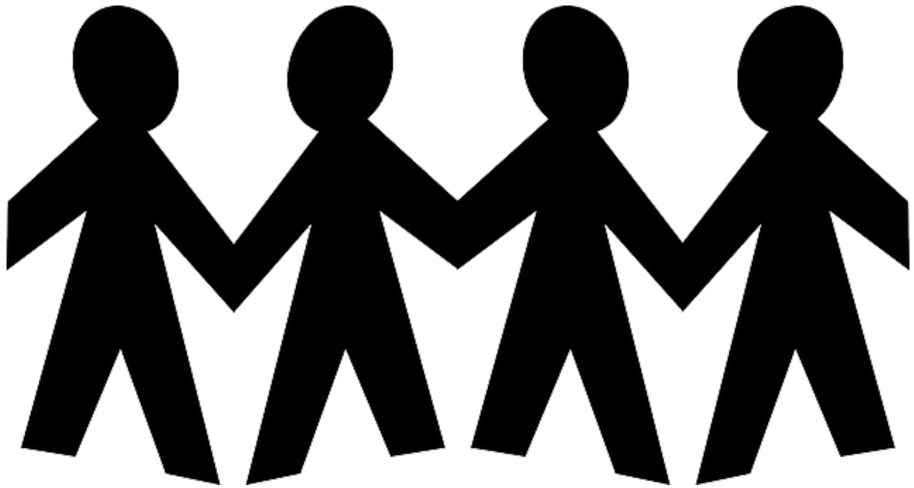
Principles of Universal Design



Principle 1: Equitable Use

- The design of a device should be useful and marketable to people with diverse abilities.
  - Provide the same means of use for all users: identical whenever possible; equivalent when not.
  - Avoid segregating or stigmatizing any users.
  - Provisions for privacy, security, and safety should be equally available to all users.
  - Make the design appealing to all users

People are not all the same



Consider a wide range of people and their different abilities

Issues for PwD

- Issues in datasets
- Under-represented or excluded from data sets
- Misrepresented by data sets
- Inherent bias in data sets
- Overwhelmed by the majority
- Probabilistic models lean towards the majority

Consequences for PwD

- Exclusion e.g from employment or credit or education
- Discrimination and loss of civil rights
- Policies made on faulty or suspect conclusions due to AI and bad data
- Technology not working for them

Research Question

- Is the problem with data with respect to people with disabilities simply a case of design by exclusion or is it exclusion by design?

(Gilligan, 2018)

Case Studies

- Automatic Speech Recognition (ASR)
- Education
- Fall Detection
- Credit Card Fraud

Case Study 1: Automatic Speech Recognition

- Study explores several new schemes to train a seq2seq (Sequence to Sequence) model (Cho et al. 2019)
- Training data based on LibriSpeech corpus which is derived from audiobooks that are part of the LibriVox project, and contains 1000 hours of speech.

http://www.danielpovey.com/files/2015_icassp_librispeech.pdf

Case Study 1: Issues

- Clean defined by automatic procedure to select higher recording quality and with accents closer to US English
- LibriVox invites readers to record in languages they speak, other than their native language, as long as they can make themselves understood in it.
- Readers from all over the world are welcome to contribute their recordings-
- No deliberate exclusion of People with Disabilities or policy to deliberately include them
- Recordings go through a process of proof listening which is a form of quality control

Case Study 1: Issues

- LibriSpeech Data Sets used in many ASR experiments because they are available
- So they are part of the development of ASR technologies
- No deliberate exclusion or inclusion of people with disability from contributing
- BUT
- Understandability requirement likely to exclude
- Dysarthic speech certainly unlikely in clean subsets
- Also challenges proof-listening methods and likely to be excluded here
- Would people with poor speech volunteer to read on LibriVOX ?

Dysarthria

- Dysarthria is a speech disorder caused by muscle weakness. It can make it hard for you to talk. People may have trouble understanding what you say.
- Causes: Stroke, Brain injury, Parkinson's disease among other conditions

Implications

- Dysarthic and other speech for example heavily accented unlikely to be a significant part of ASR datasets and therefore play limited role in developing these technologies
- Is this deliberate exclusion?
- How could the data be better processed?

Implications

- Not Deliberate Exclusion but needs better processes
- Need to actively source and curate a wider range of speech for greater inclusion

Related Work

- Project Euphonia
- Shor et al (2019) evaluate finetuning techniques to improve ASR for users with non-standard speech
- https://www.youtube.com/watch?v=OAdegPmkK-o&t=1s

Case Study 2: Education

- The goal of this study was to apply a two-stage, data-driven approach to identify associations between air pollutant exposure profiles and children's cognitive skills (Stingone et al., 2017)

- Machine Learning technique used
  - Classification and regression tree (CaRT)
- Data Set used
  - National Center of Education Statistics, the Early Childhood Longitudinal Study, Birth cohort (ECLS-B)

Training Data

- Early Childhood Longitudinal Study, Birth cohort (ECLS-B)
- Conducted by the National Center of Education Statistics
- ECLS-B is a longitudinal study of a nationally representative, random selection of children born in 2001
- Two versions
  - Public Use
  - Restricted Use

Discussion

- Exclusion depends on which files are being used public-use or restricted-use
- In this study it doesn't say which one is used
- If the public-use file is used then outlier data is not present
- This is due to legitimate privacy concerns

Implications

- While the conclusions of this study are inconclusive, interesting analysis of the learning performance of those on the margins of this data set is not part of the discussion
- People with learning difficulties may be more affected by air pollution but they are excluded from consideration if public-use data set is used

Case Study 3: Fall Detection

- The aim of this project is to create a Dataset (SisFall) which is an Activities of Daily Living and Fall Movement Data Set This then can be used with Machine Learning algorithms to predict likelihood of falls when performing Activities of daily Living
- Machine Learning technique used
  - Dataset to be used in future experiments which will decide on what ML methods to use
- Data Set used
  - SiSFall

Training Data

- The SisFall database was generated with collaboration of 38 volunteers divided into two groups: elderly people and young adults.
- The older adult group was formed by 15 participants (8 male and 7 female), and the young adults group was formed by 23 participants (11 male and 12 female).
- The older adult group is formed by retired employees of the Universidad de Antioquia and parents of current employees.
- They all were healthy and independent, and none of them presented gait problems.
- Everyday Activities included sitting, grasping, retrieving objects from the floor and walking
- Front falls, Backward Falls, falls to the side and UpSit falls were recorded

Data Gathering

- Data Gathering was done with using Microsoft Kinect v2 and acceleration samples provided by an accelerometer during the simulation of ADLs and falls.
- Standard signal pre-processing was done on the samples
- Each actor repeated each action 3 times, generating a total number of 264 sequences.

Discussion

- Automatic Falls Detection would benefit vulnerable older people and those with unsteady gait
- It could be a useful component of a tele-monitoring for independent living
- None of the potential beneficiaries were used in the study
- This is for ethics and safety reasons
- It is a feature of Fall datasets e.g. MobiFall twenty-four volunteers (22 to 42 years old) and tFall ten participants between 20 and 42 years old.

Implications

- Given these falls are simulated by agile people how representative is the data of the population likely to benefit
- In the SisFall dataset development many other factors such as case study notes and surveys were carried out to make the simulations as realistic as possible.
- Here safety is the key ethical challenge
- Is it realistic to develop Data sets such as these where there is risk?
- Serious consideration needs to be given to the design of these data gathering projects

Case Study 4 Credit Card Fraud

- This work aims to determine if it is possible to identify high risk credit card customers within the first months of the customer relationship. Using a credit card dataset consisting of customers' first 18 months of data from between January 2013 and April 2017
- Data Set used
  - SpareBank 1, credit card data from
  - over 162,000 unique customers over a period of over 3 years
- Machine Learning technique used
  - A Variety of ML methods used in a number of experiments
  - Random Forest and
  - Boosting Ensemble methods
  - Data Set used

Training Data

- SpareBank 1, a large Norwegian bank, which has provided the dataset used in this thesis.
- The dataset consists of credit card data from over 162,000 unique customers over a period of over 3 years. Over 11 million records in total.
- The dataset combines account-level data, transaction-level data, and tax assessment data for the first 18 months of each account.

Issues

- Incidents of Fraud are very low compared to non-fraud records in these datasets
- Credit Card Fraud data is typically Imbalanced
- Standard preprocessing in these cases involves redressing imbalance through techniques such as random oversampling and undersampling, SMOTE, short for Synthetic Minority Over-sampling Technique, is an oversampling method that adds synthetic data points to the dataset
- In this paper Random Oversampling was decided on after a number of tests

Discussion

- Imbalanced Data is typically balanced by these undersampling and oversampling techniques often with good reason.
- For example in the US National Health and Nutrition Examination Survey NHANES to increase the precision of estimates for certain subdomains, oversampling was carried out for Hispanic, non-Hispanic black, Asian, and white and other persons at or below 130% of the federal poverty level, and for white and other persons aged 80 and over
- The Irish Centre for Excellence in Universal Design refer to Oversampling the Tail when discussing the representation of minority populations

Implications

- Data around people with disabilities is often imbalanced.
- Balancing Techniques need to be considered for data sets
  - For example the Lawnmower of Justice (Treviranus, 2018) – "cuts off all but a small number of repeats of any given data point. This forces the learning model to level the playing field for the full spectrum of needs and pay attention to the edge or outlying data as well. It levels the hill in the normal distribution."
- One question that remains do these balancing techniques change the integrity of the datasets??

Case Study: ASR

- Exclusion Manifest
  - To Varying Degrees
- Pre-Processing
  - Yes Prooflistening and Division into Clean and Other Data Sets
- Deliberate Exclusion
  - In some parts of data set, not in others
- Implications
  - Data gathering processes unwittingly lend themselves to exclusion for people with disability

Case-study Education

- Exclusion Manifest
  - Yes
- Pre-Processing
  - Yes Public Use data Set altered to protect Privacy
- Deliberate Exclusion
  - Yes but for Privacy Reasons
- Implications

  Need to choose a data set with right representation or those on the margins will be excluded

Case Study: Fall Detection

- Exclusion Manifest
    - Yes
- Pre-Processing
    - Standard Normalization of data signals
- Deliberate Exclusion
    - Older participants with disabilities not chosen for safety reasons Young Actors and Older people in good health and mobility used
- Implications
    - It can be very difficult for good ethical reasons to acquire data and research needs to be done as to how to do this

Summary Credit Card Fraud

- Exclusion Manifest
  - No but Data naturally Imbalanced
- Pre-Processing
  - Balancing Techniques used to give bigger presence to Fraud instances
- Deliberate Exclusion
  - No Deliberate Exclusion and data balanced for greater inclusion
- Implications
  - Balancing techniques can give better representation for those with disability but does it change the integrity of the data set?

Conclusions

- Bad datasets can exclude underrepresented people from consideration
- Exclusion leads to bad design, poor policies and technology that doesn’t work
- Sometimes the exclusion is deliberate albeit for good reasons
- New datasets that are reflective of people with disabilities need to be developed or they will be left out

Remedies

- WEF Principles
- Fairness initiatives e.g. IBM Fairness 360
  - https://aif360.mybluemix.net/
  - This extensible open source toolkit can help you examine, report, and mitigate discrimination and bias in machine learning models throughout the AI application lifecycle. Containing over 70 fairness metrics and 10 state-of-the-art bias mitigation algorithms developed by the research community, it is designed to translate algorithmic research from the lab into the actual practice of domains as wide-ranging as finance, human capital management, healthcare, and education.
- Balancing Techniques e.g. SMOTE
- (Synthetic Minority Oversampling Technique)

Problems

- Fairness and Balancing done on existing flawed datasets
  - not adding new information which reflects the domain.
  - really reflecting the closed world of the dataset

Consider

- Synthetic Dataset Generation
- Around since 2008
- Many Initiatives
- Meta SIM
  - Scene Grammars
- PATE-GAN
  - Generating Synthetic Data with Differential Privacy Guarantees
- SDDL Synthetic Data Description Language
  - J. E. Hoag, Synthetic Data Generation: Theory, Techniques and Applications.

Also Consider

- Personas
- These are fictional characters, which you create based upon your research in order to represent the different user types that might use your service, product, site, or brand in a similar way.
- Used in Business, Marketing, Interaction Design, UX, Disability Studies, Case Studies, Software Development and much more
- Roles and Stereotypes in Natural Language processing
- Formal Persona Design Processes e.g. Alan Cooper 1998

Finally Steps Recommended for Companies

- 1. Identifying human rights risks linked to business operations.
  - common standards for assessing the adequacy of training data and its potential bias be established and adopted, through a multi-stakeholder approach.
- 2. Taking effective action to prevent and mitigate risks.
  - Companies should work on concrete ways to enhance company governance, establishing
- 3 Being transparent about efforts to identify, prevent, and mitigate human rights risks.
  - Companies should monitor their machine learning applications and report findings, working with certified third-party auditing bodies in ways analogous to industries such as rare mineral extraction.
  - Large multinational companies should set an example by taking the lead.
  - Results of audits should be made public, together with responses from the company.

(World Economic Forum, 2018)

References

- Executive Office of the White House (2016) Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights. Available: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf
- World Economic Forum, 2018 How to Prevent Discriminatory Outcomes in Machine Learning . Global Future Council on Human Rights 2016-2018 Available: http://www3.weforum.org/docs/WEF_40065_White_Paper_How_to_Prevent_Discriminatory_Outcomes_in_Machine_Learning.pdf
- Jaejin Cho, Shinji Watanabe, Takaaki Hori, Murali Karthick Baskar,Hirofumi Inaguma, Jesus Villalba, Najim Dehak, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) Brighton UK 2019
- Jeanette A. Stingone, Om P. Pandey, Luz Claudio, Gaurav Pandey, Environmental Pollution, Volume 230,2017,Pages 730-740, ISSN 0269-7491,
- Sucerquia, A., López, J. D., & Vargas-Bonilla, J. F. (2017) Sensors (Basel, Switzerland), 17(1), 198. doi:10.3390/s17010198
- Flood, M.H. Master of Science in Computer Science, Department of Computer Science, Norwegian University of Science and Technology July 2017. Available at https://pdfs.semanticscholar.org/5278/a2e5b96ccd8da14cec192f49217fdebdd11d.pdf
- Treviranus, Jutta (2018) Sidewalk Toronto and Why Smarter is Not Better
- https://medium.com/datadriveninvestor/sidewalk-toronto-and-why-smarter-is-not-better-b233058d01c8