Ethical thinking in Data Science

Vandana P. Janeja
University of Maryland, Baltimore County

Supported by Hrabowski Innovation Fund

PI: V. P. Janeja

- Data Science is sending out the next generation [Data] decision maker, [Data] fire fighters. Any data scientist has to be aware of ethical considerations while using data science to derive knowledge and implement it into decision making.

- **Ethical thinking** must be a critical aspect of any data science project.
- Intertwining commerce and data analytics means use of analyses or availability of records affects communities.
- These considerations should be brought to the forefront, at every point in the data life cycle
  - from data cleaning, selection, mining, thresholding and pattern evaluation, for the students who are the next generation of decision makers.

Principles of Ethical thinking in Data Science

- Ethics is not equal to Privacy. EDS includes privacy, social responsibility, decision making and evaluation of impact in an ethical framework.
- EDS can also include releasing of data not just hiding it. More importantly it includes releasing data responsibly with checks and balances in place.
- EDS means considering all of us and representing all of us in data.
- EDS considers every data scientist touching the data as a data steward, this includes data collectors, data users, data re-users, data re-combiners.
- Ethical context is also heavily influenced and interpreted through a lens of other types of context such as space, time, activity.

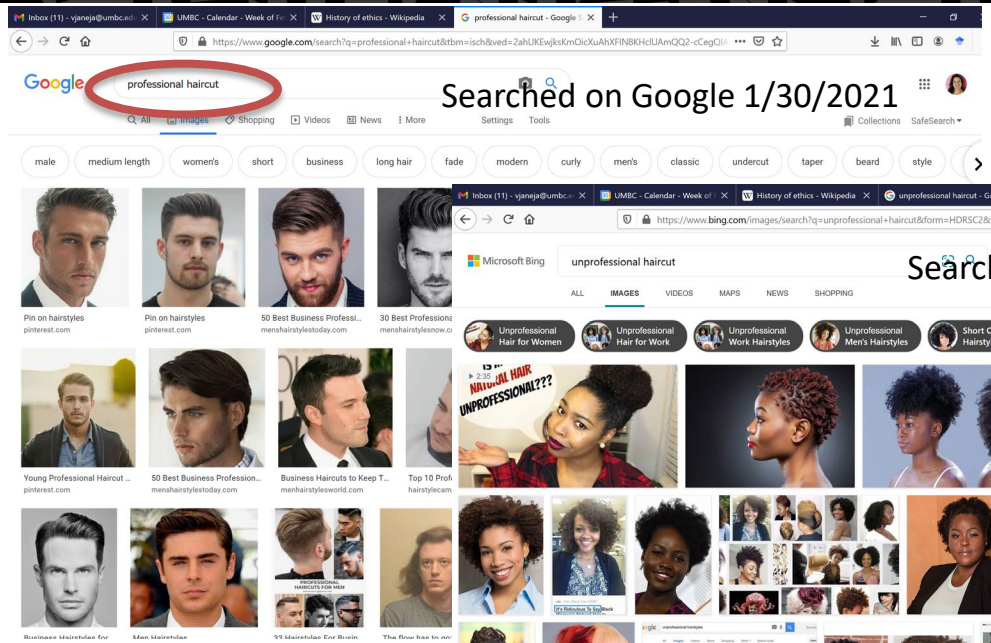
Ethics

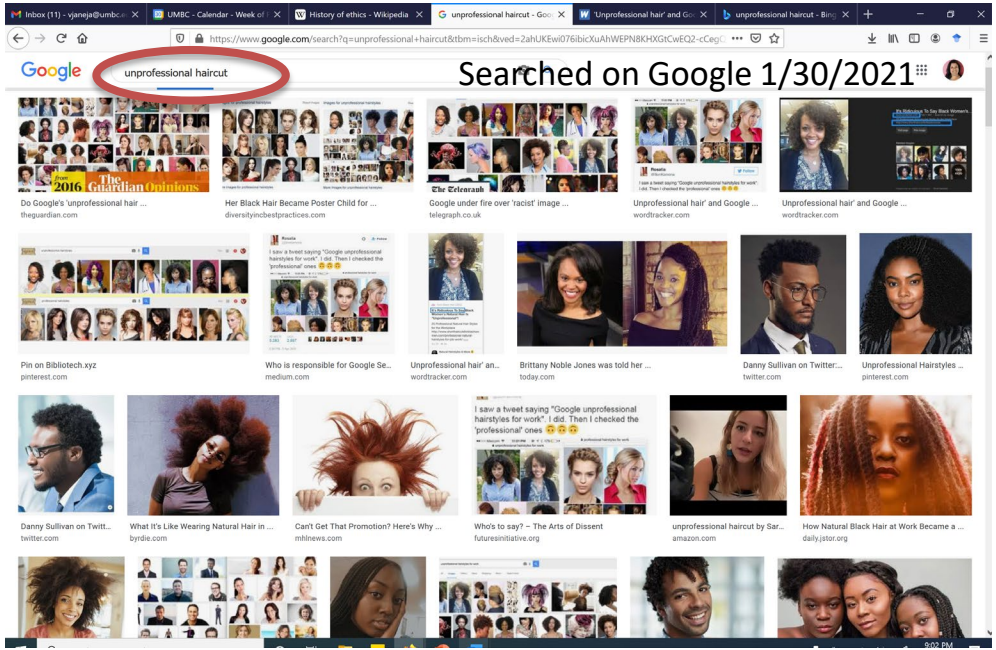
- Ethics is a complex part of our fabric
- Ethics has been taught since humans started teaching
- Ethics and its implications get morphed with the time we live in
- Ethics takes on different magnitude with the context
- Our ways and behavior evolve as we better understand the ethical implications in context and in the time we live in

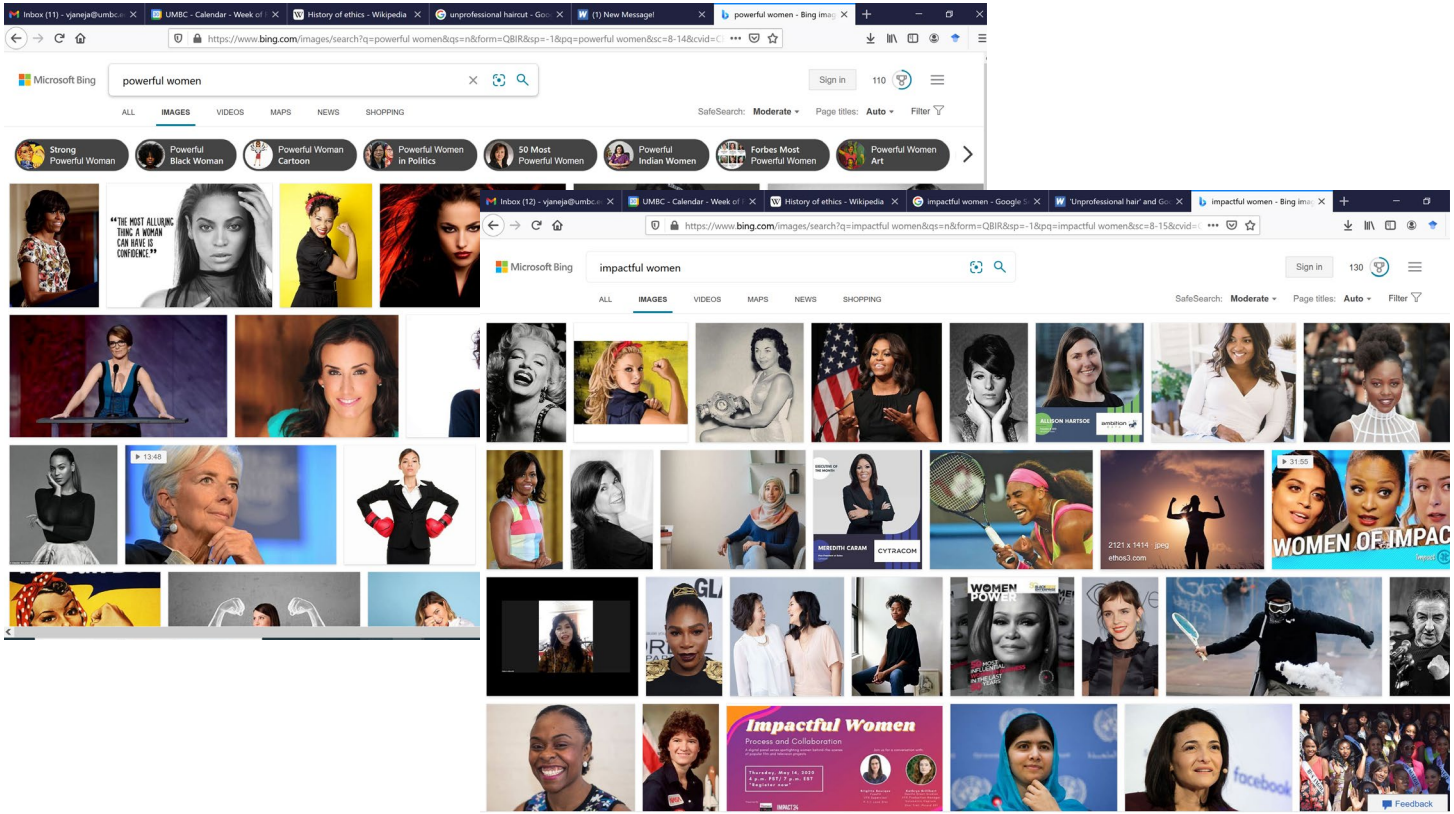
Lets do some Web search

Case in point

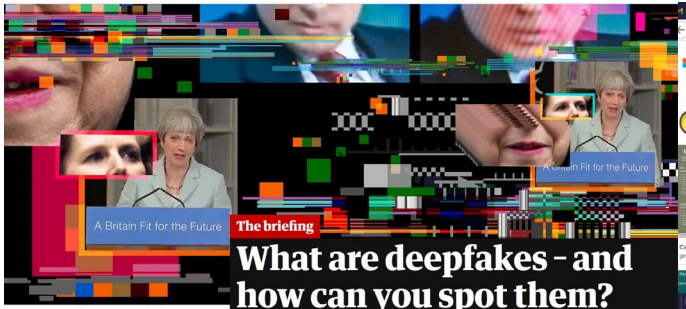
https://www.theguardian.com/technology/2016/apr/08/does-google-unprofessional-hair-results-prove-algorithms-racist-


professional haircut
Searched on Google 1/30/2021
male
medium length
women's
short
business
long hair
fade
modern
curly
men's
classic
undercut
taper
beard
style
unprofessional haircut
Searched on BING on 1/30/2021
Unprofessional Hair for Women
Unprofessional Hair for Work
Unprofessional Work Hairstyles
Unprofessional Men's Hairstyles
Short Curly Afro Hairstyles
Unprofessional Men's Haircut
Barber Shop Haircuts
Natural Curly Bob Hairstyles
IS MY NATURAL HAIR UNPROFESSIONAL???
Toronto Principal Tells Eighth-Grader Her Hair Is Office 'Too Poofy' and 'Unprofessional'

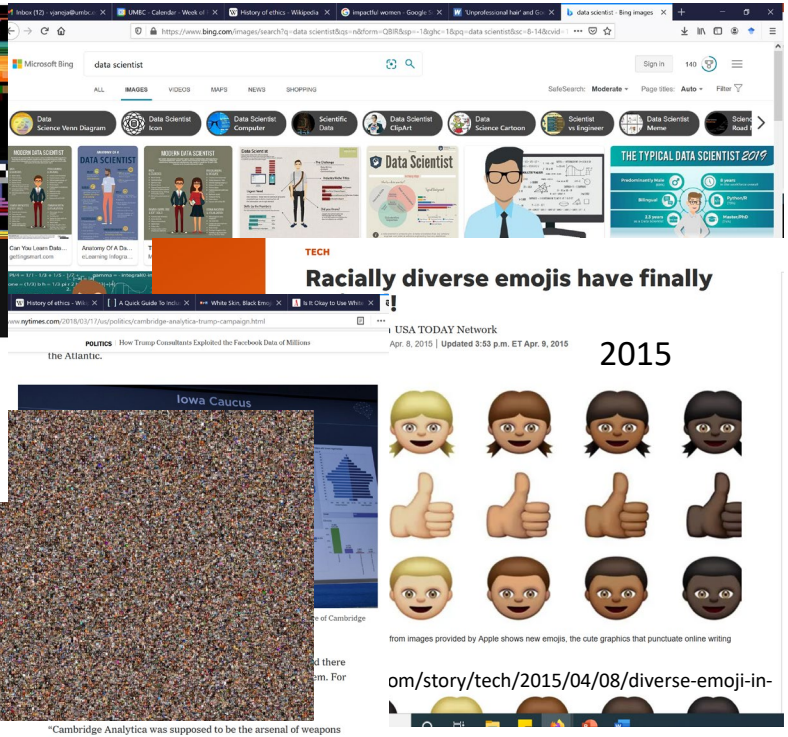

unprofessional haircut
Searched on Google 1/30/2021


powerful women
Microsoft Bing
Sign in
ALL
IMAGES
VIDEOS
MAPS
NEWS
SHOPPING
Strong Powerful Woman
Powerful Black Woman
Powerful Woman Cartoon
Powerful Women in Politics
50 Most Powerful Women
Powerful Indian Women
Forbes Most Powerful Women
Powerful Women Art
"THE MOST ALLURING THING A WOMAN CAN HAVE IS CONFIDENCE."
impactful women
Impactful Women
Process and Collaboration
WOMEN POWER
WOMEN OF IMPAC
facebook
Feedback

A discussion continues…



https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them



technology/facial-recognition-photo-tool.html

Here's a Way to Learn if Facial Recognition Systems Used Your Photos

An online tool targets only a small slice of what's out there, but may open some eyes to how widely artificial intelligence research fed on personal images.

analytica-trump-campaign.html

om/story/tech/2015/04/08/diverse-emoji-in-

Ethics

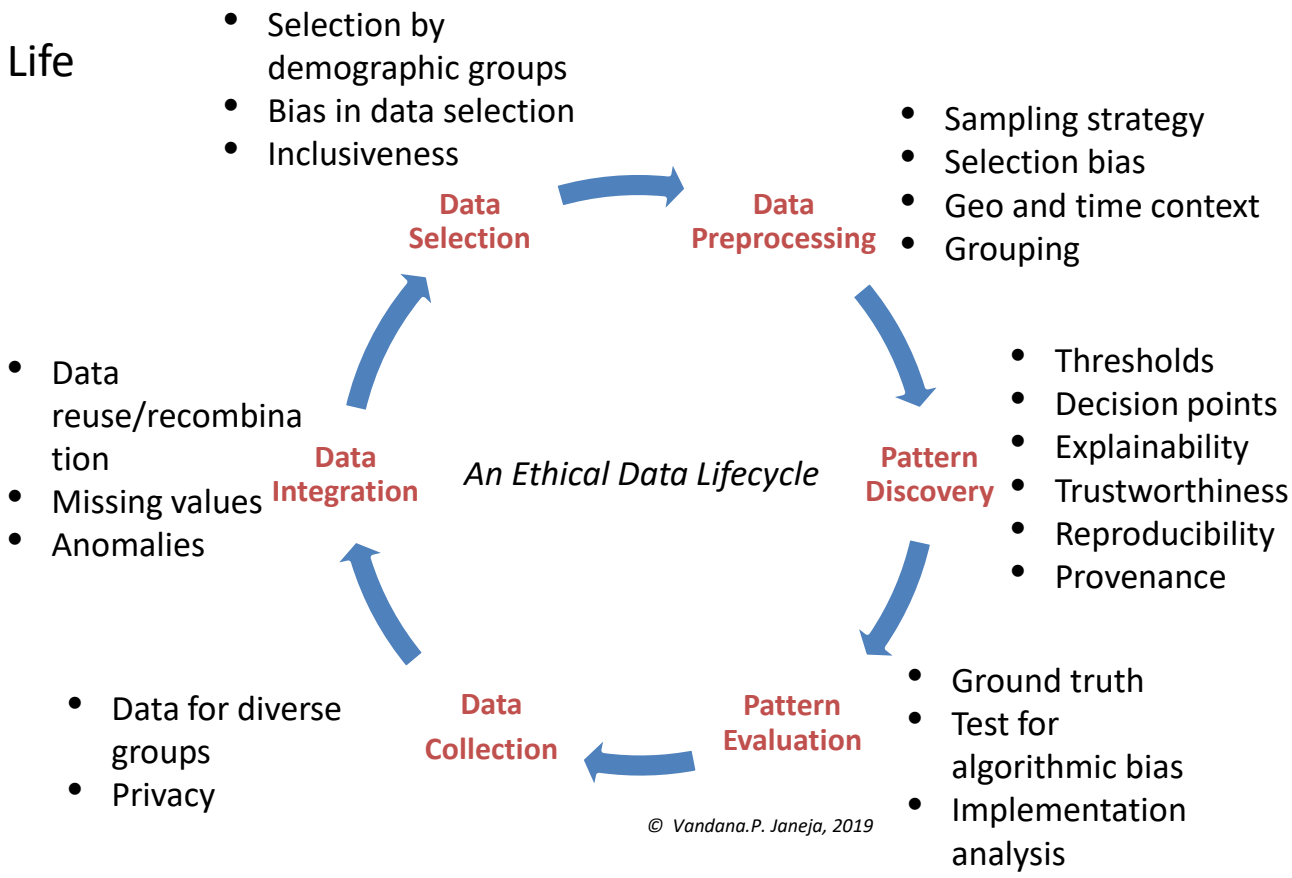
- Ethics is a complex part of our fabric
- Ethics has been taught since humans started teaching
- Ethics and its implications get morphed with the time we live in
- Ethics takes on different magnitude with the context
- Our ways and behavior evolve as we better understand the ethical implications in context and in the time we live in

Everyone needs to address ethics – in today's time and context – especially data scientists and AI experts

Data Life cycle

- Embedding ethical decision making in the data life cycle
- New use of records brings new problems, or difficulty in recognizing existing knowledge
- Data Omission
- Collaboration: learning with what research has demonstrated in other fields (e.g., selection bias)

Ethical Data Life cycle



Case Study Discussion

- **Anomaly Detection: Under the [data] Hood in Smart Cars**
  - Quader, F., & Janeja, V. (2019, June). Anomaly Detection: Under the [data] hood in Smart Cars. In *2019 IEEE International Conference on Smart Computing (SMARTCOMP)* (pp. 126-131). IEEE.
- **Subspace Discovery for Disease Management: A Case Study in Metabolic Syndrome**
  - Namayanja, J., & Janeja, V. P. (2011). Subspace discovery for disease management: A case study in metabolic syndrome. *International Journal of Computational Models and Algorithms in Medicine (IJCMAM)*, *2*(1), 38-59.
- **Contact tracing COVID**
  - **Student research**

Ethical Considerations in Data and Algorithms

Training and Test Distribution in Classification

Data

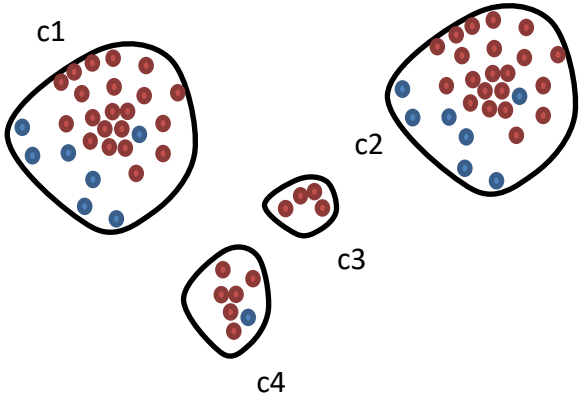
Class	Percent
B	20
W	70
H	10

Sampling Distribution

Class	Percent
B	20
W	70
H	10

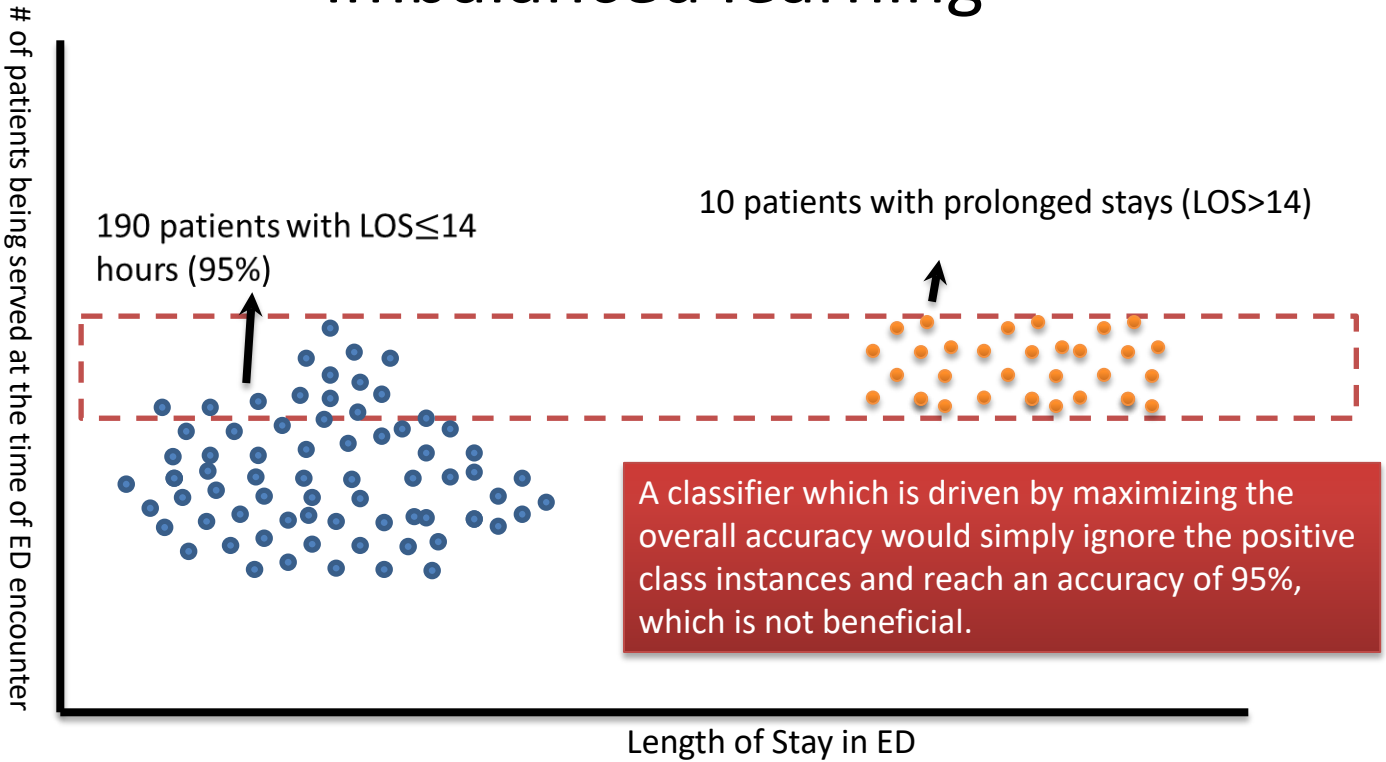
Class	Percent
B	33
W	33
H	33

Clustering and Anomaly Detection



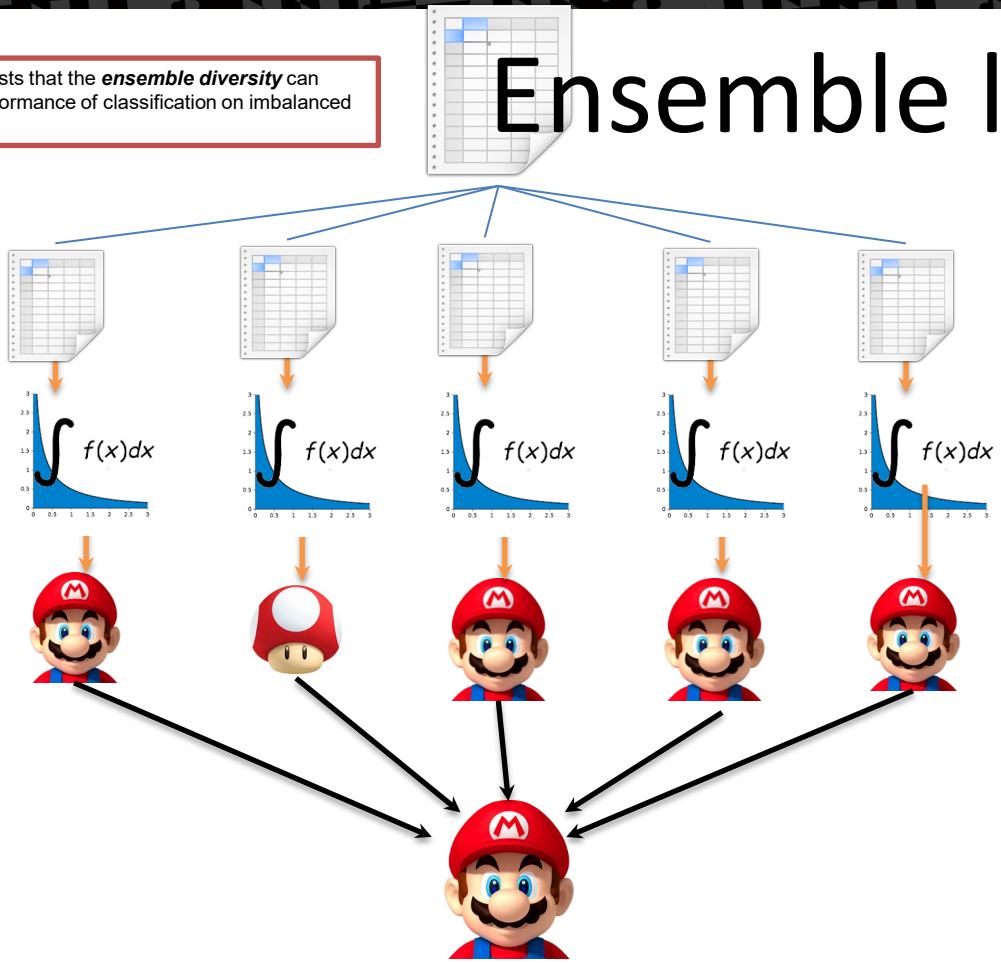
Cluster C3 and C4, Are they anomalies?
Are they minorities?
Should they be rejected or accommodated in the data?

Imbalanced learning



Ensemble learning

Literature suggests that the ***ensemble diversity*** can improve the performance of classification on imbalanced data [23], [24].



Is AI/Data Science more good than bad?

- Deep fakes
  - Mark Zukerberg
  - Nancy Pelosi

- Google: detecting breast cancer

Big Data --- Big Debate

- Pro / Cons
  - Summary of case study
  - Pros and Cons
  - Ethical questions (answered/unanswered)
  - Gap
  - What would you do different?