Fairness in Context and Philosophy

Martin Mose Bentzen (DTU) and Jens Ulrik Hansen (RUC)

2019-08-30

The end note yesterday

- Fairness is super hard!

Where are we now?

- Fairness in Machine Learning
    - A first thought
        - Fairness is about treating people from different sensitive groups similarly
        - So let us just leave out information about membership of that sensitive group
    - A second thought
        - Technically we can define fairness of an algorithm in several different inconsistent ways
        - Just leaving out the variable representing the sensitive group might not be enough or a good idea at all
        - What fairness is, might also depend on the application and the stakeholder perspective
        - … fairness is super hard!
    - Why and when should we aim for fairness in first place?

Sneak peek at the conclusions

- Technical work on fairness is important
- In many cases, there is no one right solution, so we need to apply critical thinking
- While, you as a machine learning engineer have a responsibility to aim for fairness, you are not alone in this responsibility!

Outline

- Technical fairness definitions
- Fairness in context
- Ethics principles, fairness and philosophy
- How to be fair

Fairness in context

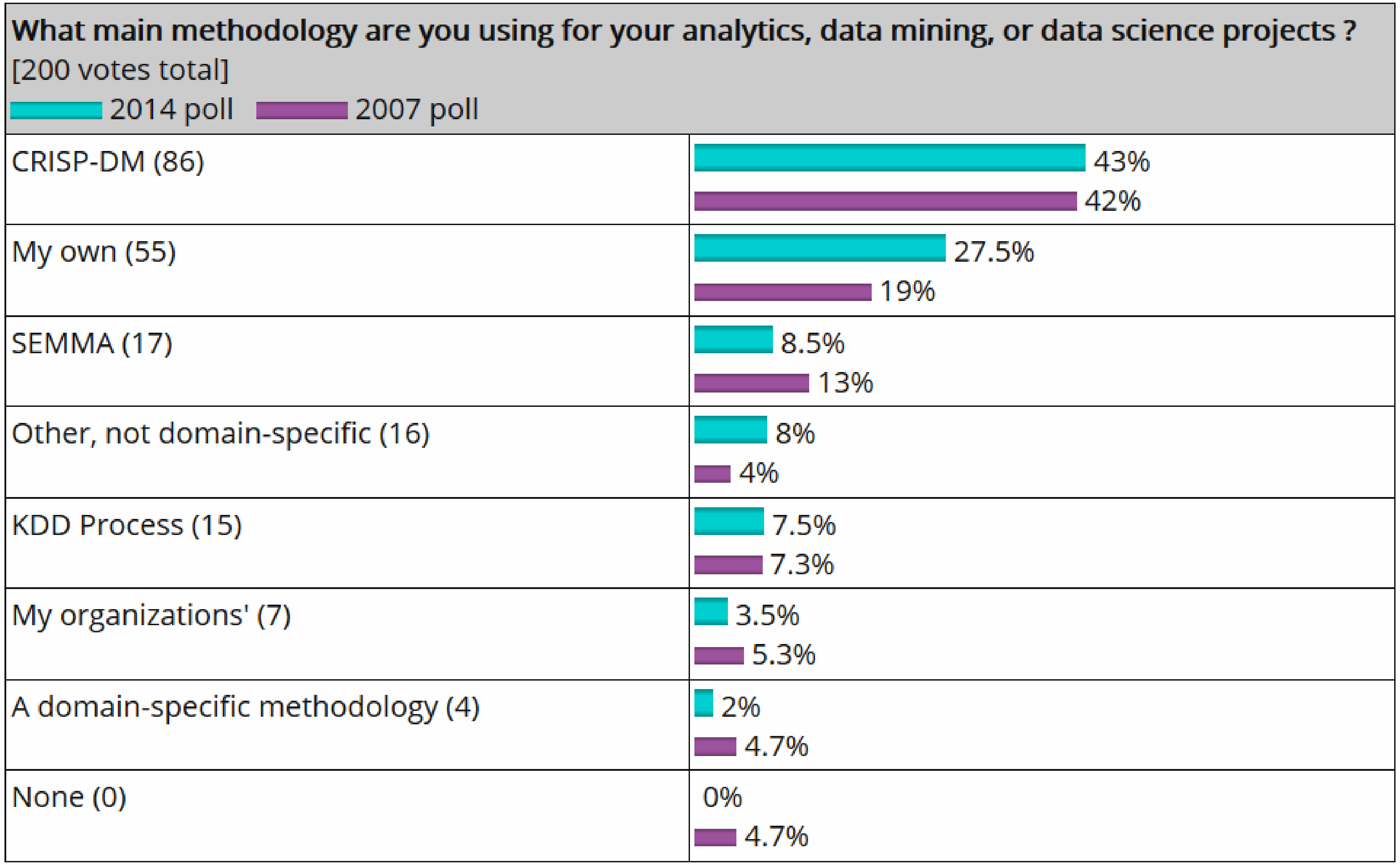
- Different contexts or application might lead to different notions of fairness being important
  - Deciding on how to distribute a scarce resource (the preventive medicine) or an abundant resource (the cat video treatment on facebook)
- Different stakeholders might have different views on what is fair
  - I will feel a travel ban unfair if it is not very certain I have the disease, while all the people that potentially could get infected will feel it is fair. Will it be fair to the airline companies?
- Difference between individual fairness and what is fair from a societal point of view
  - I might want doctors to be distribute such that I have the same access to one as elsewhere in the country, but from a societal point of view, it might be more fair to distribute the doctors according to where there are more sick people
- There can be other ethical aspects to take into account than just fairness
  - From a societal perspective it might be desirable to save as many lives as possible

Clarifying Group fairness

- Fairness from a societal point of view
- Fairness from the viewpoint of a particular group of people

The Data Science Process

- Different process models
  - https://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html



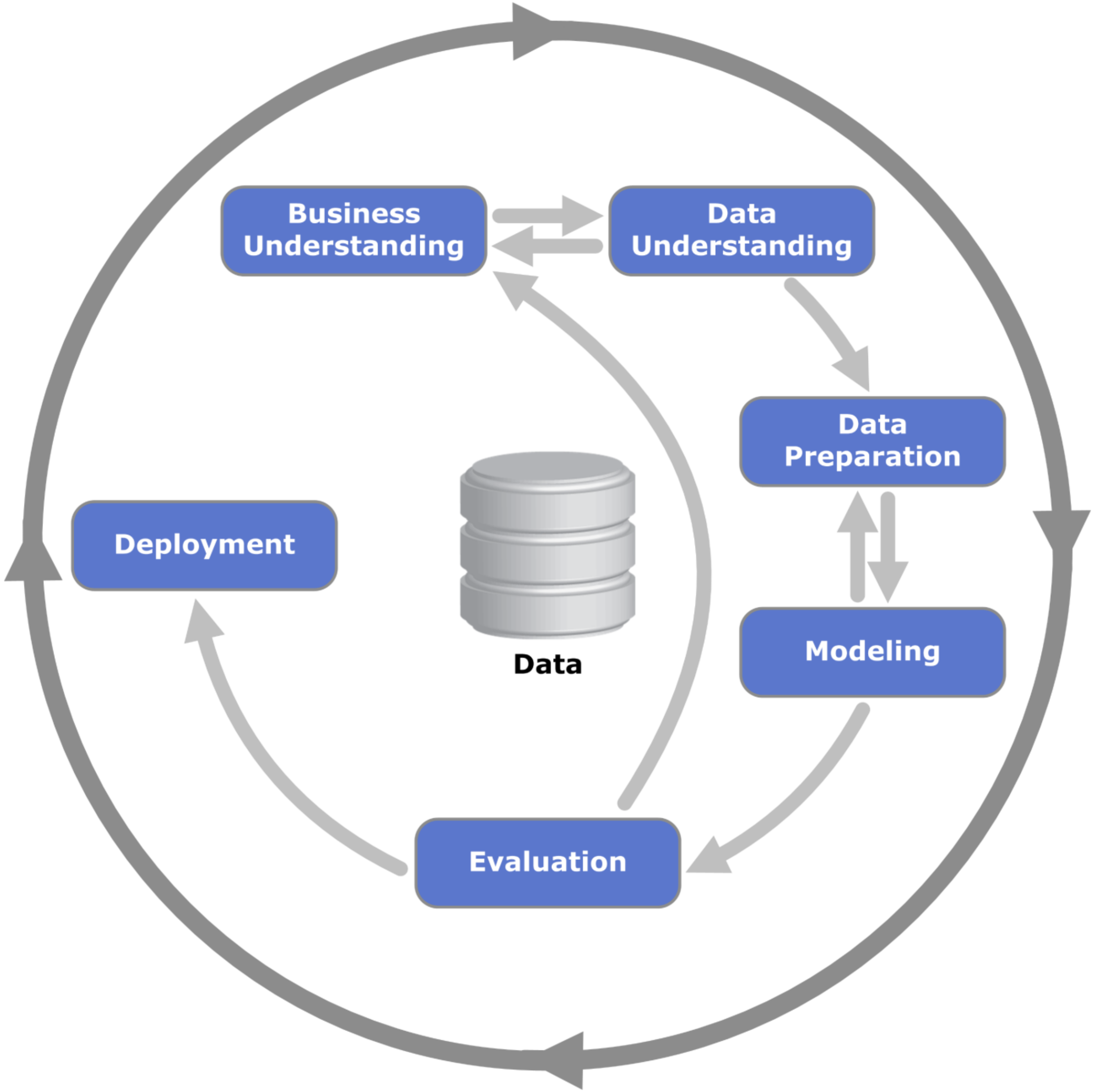
What main methodology are you using for your analytics, data mining, or data science projects ?
[200 votes total]

🟦 2014 poll 🟪 2007 poll

	2014 poll	2007 poll
CRISP-DM (86)	43%	42%
My own (55)	27.5%	19%
SEMMA (17)	8.5%	13%
Other, not domain-specific (16)	8%	4%
KDD Process (15)	7.5%	7.3%
My organizations' (7)	3.5%	5.3%
A domain-specific methodology (4)	2%	4.7%
None (0)	0%	4.7%

CRISP-DM

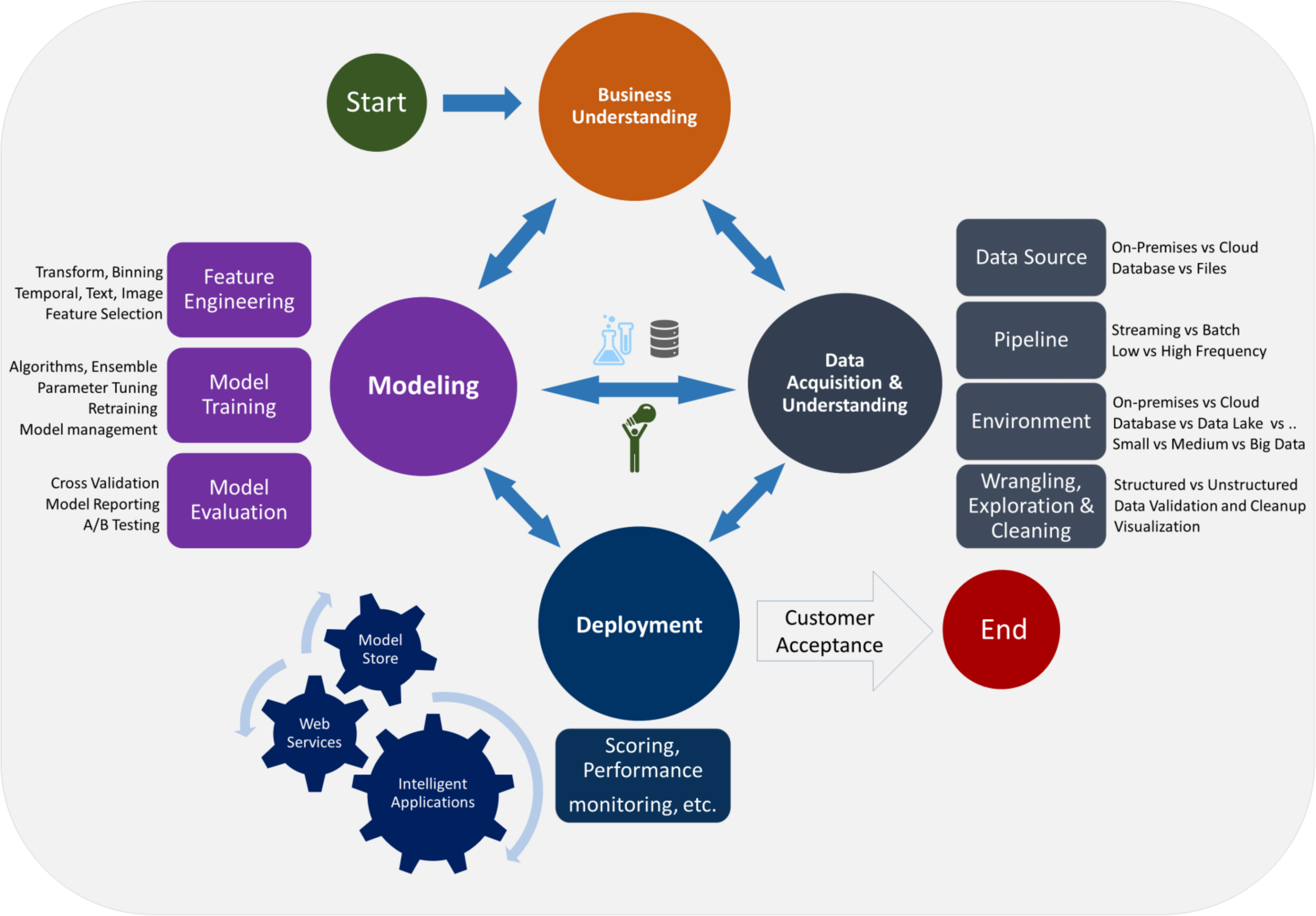
- ***Cross-industry standard process for data mining***
  - https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining



TDSP

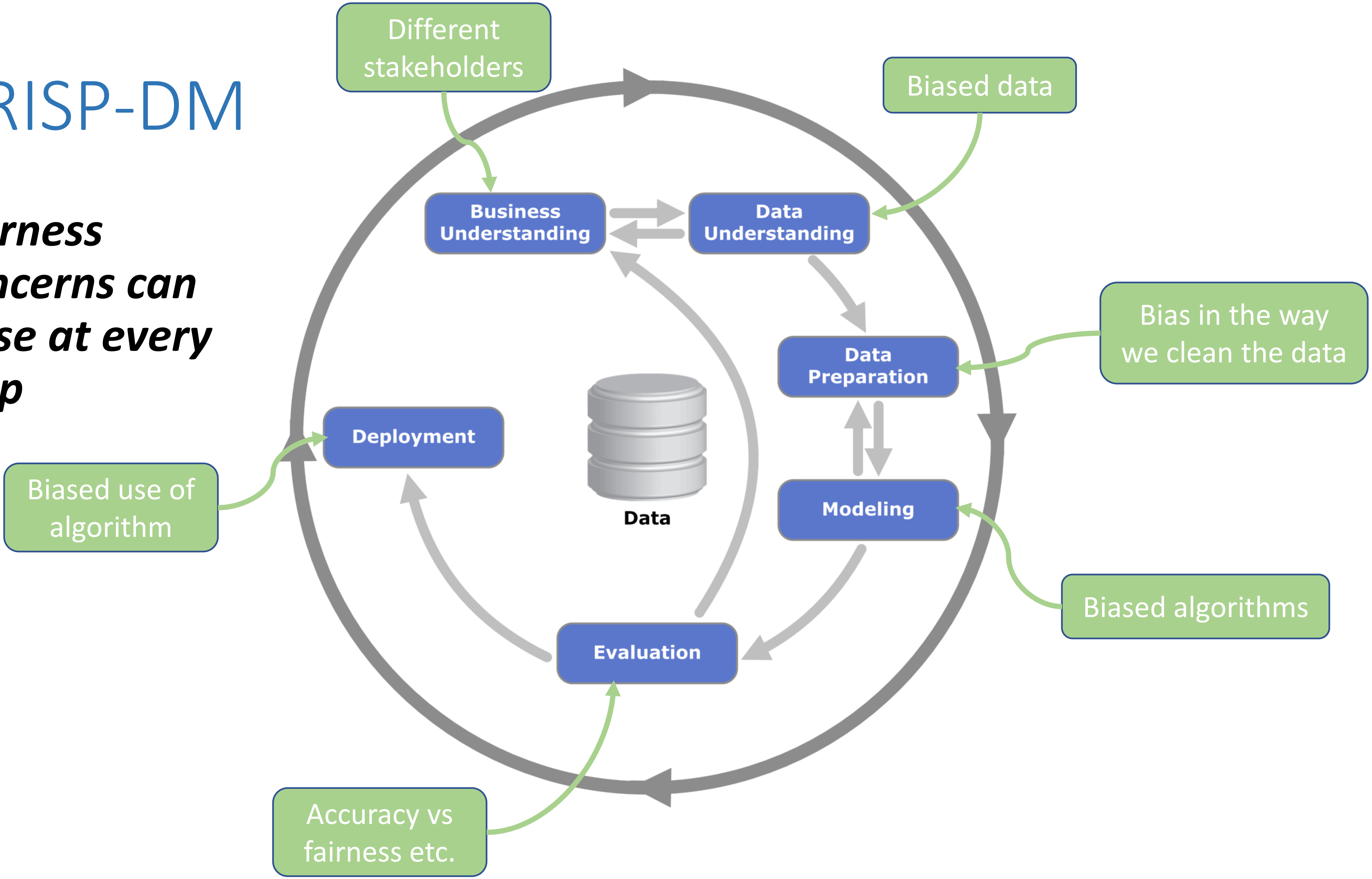
- ***Team Data Science Process***
  - Microsoft
  - https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/overview

Data Science Lifecycle



CRISP-DM

Fairness concerns can arise at every step



Ethics principles, fairness and philosophy

- Can Philosophy help?
  - Philosophy is about questioning things, rarely about coming up with answers
  - However, Philosophy can provide a useful conceptual framework for taking the discussion of fairness in Machine Learning
  - (Note, philosophers often talks about justice instead of fairness. They are very similar concepts, but not necessarily the same…)

Disclaimer: ethics vs law

- While, the law might have paragraphs that tries to define and ensure fairness, the law will never be complete
- One can act in accordance with the law, but in an unethical way
- One can act in an ethical way and still break the law
- So we cannot rely on the law to provide us with the right definition of fairness, neither should we expect that we can come up with a definition of fairness that can be implemented in law and cover all future cases whatsoever

Ethics

- Ethics is about how we should act
- Ethics is usually divided into metaethics, normative ethics, and applied ethics
    - **Metaethics:** What is ethics? Where do ethical principles come from? … etc.
        - We do not need to go here… - Philosophers only!
    - **Normative ethics:** Which actions are right? Which actions are wrong? … etc.
        - Three major approaches: Consequentialism, Deontology/duty ethics, and Virtue ethics
        - Adhering to a normative ethics can guide us in what to do and thus it seems relevant to Fairness in Machine Learning
    - **Applied ethics:** Ethical consideration about concrete cases such as abortion, gun control, …etc.
        - Fairness in Machine Learning could be considered applied ethics
        - In a sense what you did yesterday doing the exercises

Virtue ethics

- We should act as the virtuous person would do
  - Problem: How do we recognize the virtuous person?
- It is not as much about abiding to ethical rules/principles or assessing consequences of actions, but more about building moral character based on virtue
  - Such values could be wisdom, courage, justice, generosity, self-respect, sincerity, … etc.
- Fairness could be considered a virtue
  - In doing so, virtue ethics will tell us to be fair
  - While, virtue ethics cannot be dismissed completely, it currently less obvious how it should feed into applied ethics such as fairness in Machine Learning

Deontology/duty ethics

- It is about fundamental principles of obligation, such as “don´t commit murder”, irrespective of the consequences
  - The consequences of an action alone is not enough to determine whether it is right or wrong, in general
  - Rule based
- Examples
  - Rights theory, I have certain fundaments rights such as not being harmed by you, right to freedom of speech, etc.. The rights of one person implies duties of another person
  - Immanuel Kant’s ***categorical imperative***: *“Act in such a way, that whoever is treated as a means through your action (positively or negatively and including yourself), must also be treated as an end of your action.”*

Consequentialism

- The rightness of an action boils down to the consequences of that action through a cost-benefit analysis (weighing positive and negative consequences)
- Examples
  - Utilitarianism: an action is morally right if the consequences of that action are more favorable than unfavorable
  - Maximizing expected utility: Chose the action that maximizes the total expected utility (Utilitarianism with uncertainty)
  - Pareto efficiency/optimality: Chose the action such that it is impossible to take an alternative action to make some better off, without also making someone worse off
- Problem: How do we actually compute the utilities?

How to be fair…

- **Think critically about**
  - What is your own point of view (or you boss)?
  - What other relevant point of views could there be?
  - What legal compliance issues are relevant?
  - What ethical principle will you be arguing from?
  - Are fairness a relevant concern?
  - If yes, what definitions of fairness are relevant?
  - How can you implement the relevant fairness definitions?
  - How will you ensure that that fairness continue downstream – your fair algorithm is not used in an unfair way?
  - How do you monitor fairness after deployment?
  - How to justify the choices you have made?

Who has the responsibility of fairness?

- Some responsibility of fairness lies with the Data Scientists/Machine Learning Engineers
  - Discover biases in datasets
  - Attempt to be fair in data preprocessing
  - Attempt to develop fair algorithms and asses them on more than just accuracy
- However, implementation, purpose, stakeholders and the context in general influences the fairness of Machine Learning Systems
  - Thus, other people jointly bares the responsibility of fairness
  - You should not accept to take on the responsibility of fairness alone!

The next buzzword

- Fairness in Machine Learning, Explainable AI (XAI)
- **Responsible AI** (First four hits on Google yesterday was: Google, PwC, Accenture, and McKinsey…)
    - **Google:** *“The development of AI is creating new opportunities to improve the lives of people around the world, from business to healthcare to education. It is also raising new questions about the best way to build fairness, interpretability, privacy, and security into these systems”*
    - **PwC:** *“With great potential comes great risk. Are your algorithms making decisions that align with your values? […] How is your brand affected if you can’t explain how AI systems work? It’s critical to anticipate problems and future-proof your systems so that you can fully realise AI’s potential. It’s a responsibility that falls to all of us — board members, CEOs, business unit heads, and AI specialists alike.”*
    - **Accenture:** *“However, with great power comes great responsibility. Specifically, AI raises concerns on many fronts due to its potentially disruptive impact. These fears include workforce displacement, loss of privacy, potential biases in decision-making and lack of control over automated systems and robots. While these issues are significant, they are also addressable with the right planning, oversight, and governance.”*
    - **McKinsey:** *“Deploying AI requires careful management to prevent unintentional but significant damage, not only to brand reputation but, more important, to workers, individuals, and society as a whole.”*