BIAS

Lecturer: Mark Gillan

BIAS DEFINED

- When bias is discussed it can mean three related, but slightly different things:

- **Statistical bias**: the estimated value of the results is systematically different from the true underlying parameter.
- **Data or sampling bias**: the available data is not representative of the population or phenomenon of study.
- **Algorithmic bias**: a computer system reflects the implicit values of the humans who created it.

CAUSES OF BIAS

Sample bias

The dataset used for model development does not represent the population the model will be used on.

Exclusion bias

Predictive variables are removed prior to the modelling process.

Measurement bias

Arises from a systematic issue with the collection of the data, often through calibration or faulty detectors.

Confirmation bias

Arises from cherry-picking data or variables that confirm a human's existing beliefs and expectations.

Stereotype bias

Cultural or gender stereotypes may lead to an uneven distribution of modelling data.

Survivorship bias

It arises from concentrating on the successful output of a process and ignoring those that don't survive.

Simpson's paradox

A statistical phenomenon that leads to incorrect inferences when trends for subgroups reverse the overall norm.

Correlation bias

It arises from inferring a correlation where it doesn't exist due to confounding variables that have not been accounted for.

AGAIN … CAUSES OF BIAS – TRY EXPLAINING

- Sample bias
- Exclusion bias
- Measurement bias
- Confirmation bias
- Stereotype bias
- Survivorship bias
- Simpson's paradox
- Correlation bias

CAUSES OF BIAS – CHECK - REMINDER

Sample bias

The dataset used for model development does not represent the population the model will be used on.

Exclusion bias

Predictive variables are removed prior to the modelling process.

Measurement bias

Arises from a systematic issue with the collection of the data, often through calibration or faulty detectors.

Confirmation bias

Arises from cherry-picking data or variables that confirm a human's existing beliefs and expectations.

Stereotype bias

Cultural or gender stereotypes may lead to an uneven distribution of modelling data.

Survivorship bias

It arises from concentrating on the successful output of a process and ignoring those that don't survive.

Simpson's paradox

A statistical phenomenon that leads to incorrect inferences when trends for subgroups reverse the overall norm.

Correlation bias

It arises from inferring a correlation where it doesn't exist due to confounding variables that have not been accounted for.

IMPACTS OF BIAS

- There are many examples of bias in machine learning algorithms that have led to a negative outcome for those for which the biased algorithms have been used upon.
- The impacts vary from perpetuating biases in recruitment processes; to an increased chance of being arrested; to requiring unnecessary surgery; to missing out on a place at university.
- This last example happened In the UK, when an algorithm was used to predict exam grades.

MITIGATING BIAS

The best way to avoid bias in data science is to have a diverse team who each bring a different perspective to the problem:

- Have multiple people with differing backgrounds involved in the project
- Review interpretations and conclusions with team members to identify gaps in reasoning
- Review the findings with subject matter experts who understand the data and how it was captured
- Include multiple data sources from as wide a range of providers as possible
- Look for simple interpretations to the findings, such as data quality issues and rule these out before moving on to more complex interpretations

Transparency in the model development approach and dataset used, alongside extensive testing and validation will also mitigate bias.

FINALLY

Finally, the ability to put a "human in the loop" for critical decisions should always be available.