

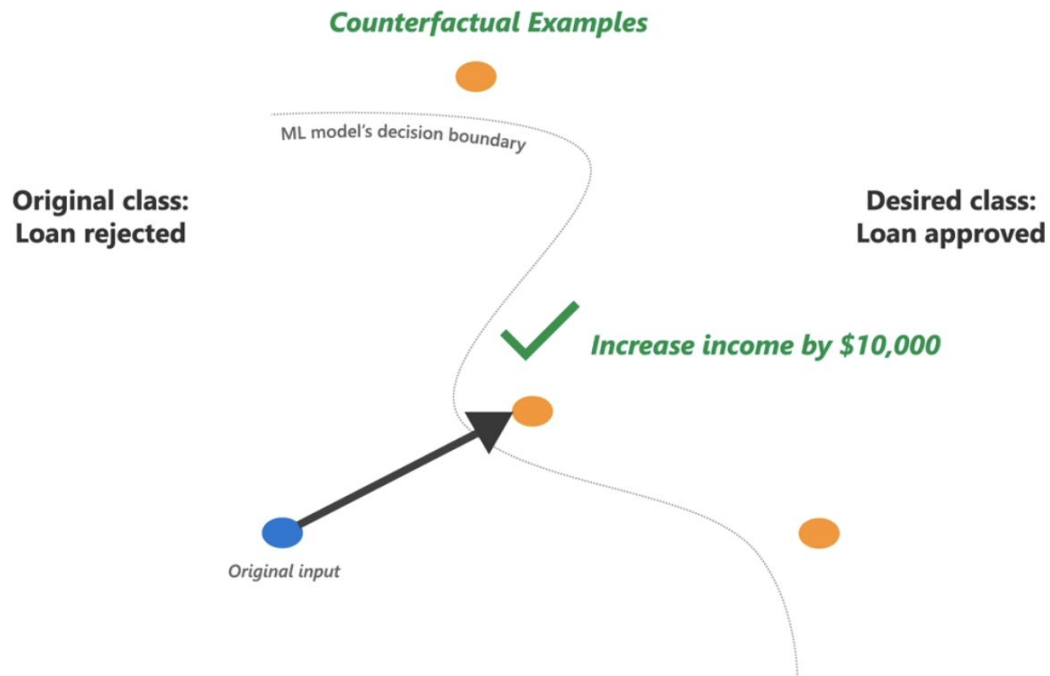


Four ways to mix causal inference and machine learning

Strengths and weaknesses

	Machine learning	Causal inference
Objective function	Maximise predictive performance subject to generalisation constraint	Identify a causal effect with observational data
Type of knowledge generated	Correlational / predictive knowledge (if I see X I am more likely to see Y)	Causal (if I do X I am more likely to see Y)
Advantages	Uses all the data Doesn't require any domain knowledge	Policy-relevant Interpretable Easier to assess generalisability (& robustness)
Weaknesses	Less useful for policy Comp & data intensive (-> unfair) Hard to interpret Brittle	Requires a lot of domain knowledge Harder to assess model performance
Example use case	Product recommender	Analysis of policy intervention

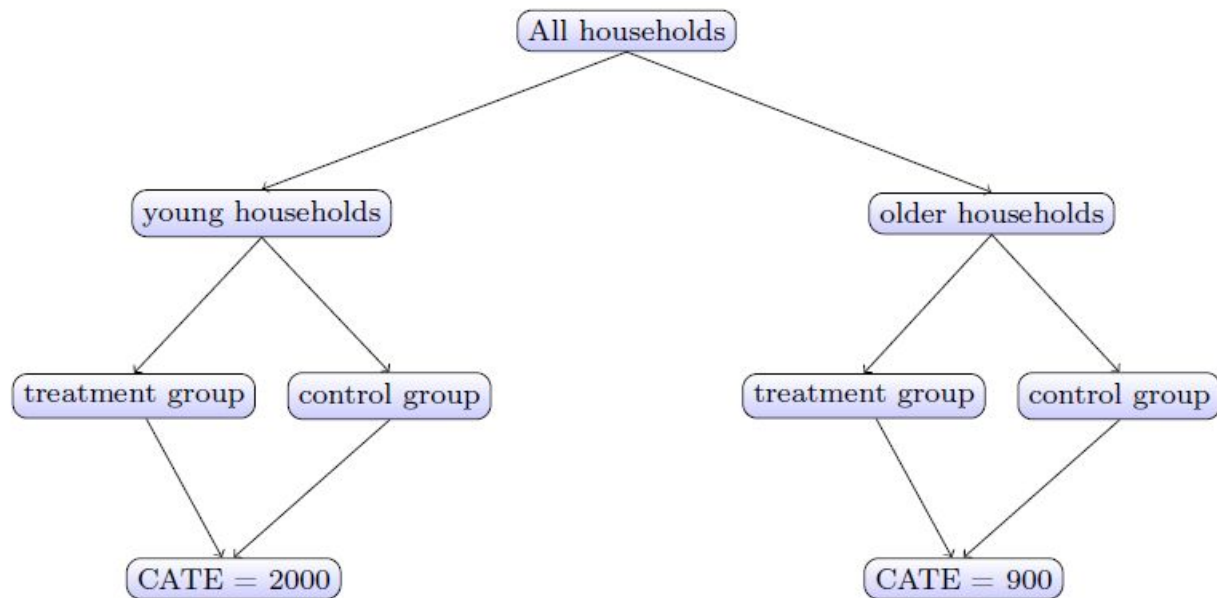
Application 1: ML enhancement via improved interpretability / fairness



Use causal methods to improve interpretability and assess algorithmic impacts on fairness [[blog](#)]



Application 2: CI enhancement via heterogeneous treatment effects



Use machine learning to identify variables that maximally explain differences in outcomes between variables

Application 2b: CI enhancement via better controls



Lukas Mergele
@LukasMergele

...

Real footage of a Phd student trying to match on observables.

[#doitlike](#) @KhoaVuUmn



Prince Charles

Male
Born in 1948
Raised in the UK
Married Twice
Lives in a castle
Wealthy and Famous



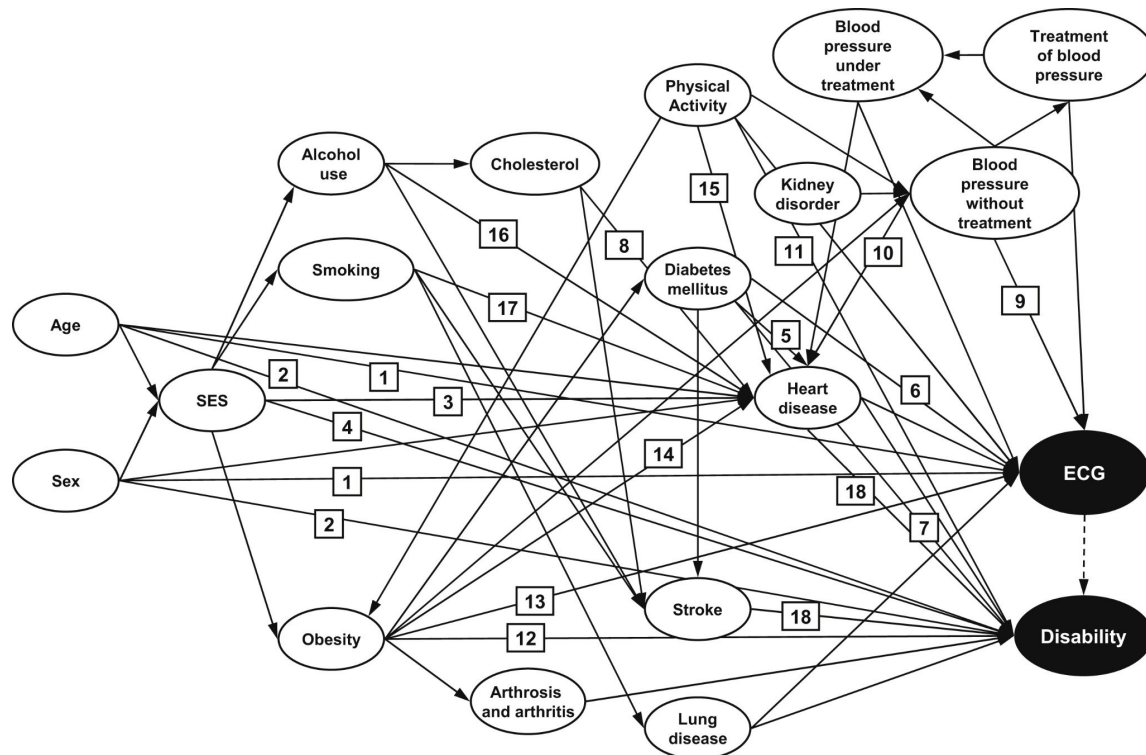
Ozzy Osbourne

Male
Born in 1948
Raised in the UK
Married Twice
Lives in a castle
Wealthy and Famous

9:51 AM · Dec 9, 2021 · Twitter Web App

Use machine learning to measure unobservables allowing better comparisons between treated & untreated

Application 3: Causal AI



Use network theory to automatically identify model specifications that make it possible to identify causal effects

Where to next?

Causal ML flavour	Use case	Observations
$ML \leftarrow CI$	Use DICE to analyse interpretability / fairness in one of our datasets (EST?) or an open benchmark dataset	
$CI \leftarrow ML$	Use EconML to analyse heterogeneous effects of an intervention on an outcome (Covid 19 early years data?)	<ul style="list-style-type: none">• Intervention doesn't have to be a policy; methodologically complex
$CI \cap ML$	Use DoWhy to analyse the causal structure of a dataset of interest (impact of research on business outcomes analysis?)	<ul style="list-style-type: none">• Need to create a causal graph• Complex data processing e.g. geocode data to enable a geographical analysis?

One suggestion

Break into two teams and select two cases to explore

- Literature
- Data
- Implement a prototype analysis
- Document everything in code and slides and present to each other on Friday PM before end of week party?