Causal ML

Estimating Heterogeneity in Treatment Effects

@ravibapna

AB Testing

- Gold standard for causal inference
- Inference (using standard statistical inference methods such as T-tests and regression models) gives us the **average treatment effect (ATE)**
  - IS the treatment group (version A) significantly different, **on average**, than the control group (version B)
- But averages can mask extreme outcomes --- may have extreme positive effects and extreme negative effects in different subgroups which on average may wash out!
  - There may be no ATE, but say ‘older black men who run 5 miles per week’ may respond very positively to a drug!
- If trying multiple variants of an Ad say or a Call to Action, maybe different sub-groups respond differently to different versions (even though there is not ATE between them)

Challenges with Heterogeneous Treatment Effect (HTE) Estimation

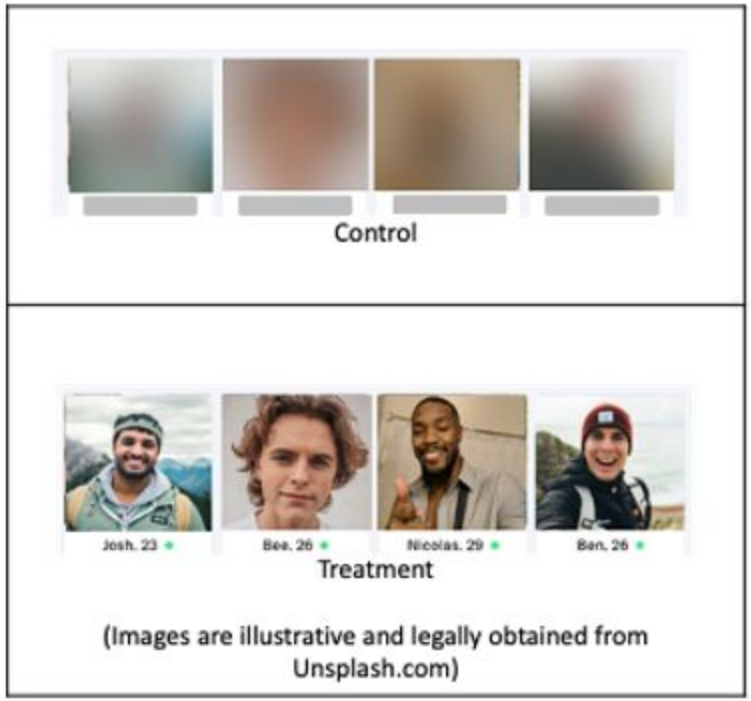
- Potential outcomes framework - in any AB test we only observe a given user in treatment or control, not both
  - If user 7 is in control group she cannot be in treatment group
  - If user 978 is in treatment group she cannot be in control group
  - Thus, we cannot estimate y(treatment) - y(control) for individual users!
- We don’t know ex ante which subgroup (combination of user characteristics) has high or low treatment effects

Solution - Use ML to Overcome Potential Outcomes Challenge

1. We can generate best in class predictive models for treatment and control groups, separately → data generation process for treatment group and control group
2. Say user 7 was in control group. We can apply the predictive model learnt from the treatment group to predict the potential outcome for user 7 if she was in the treatment group!
    a. We can now estimate y_hat(treatment) - y(control) for user 7
3. Say user 978 was in treatment group. We can apply the predictive model learnt from the control group to predict the potential outcome for user 978 if she was in the controlgroup!
    a. We can now estimate y(treatment) - y_hat(control) for user 978
4. We can estimate y(treament) - y(control) for all users!

CAUSAL FOREST ALGORITHM DOES THIS

Case -- Who Likes You Feature in Online Dating (Bapna et al 2022)



Do different subgroups of users benefit (or get hurt) from this feature differently?