Causal Inference and Uplift Modeling

PyConf Hyderabad 2025

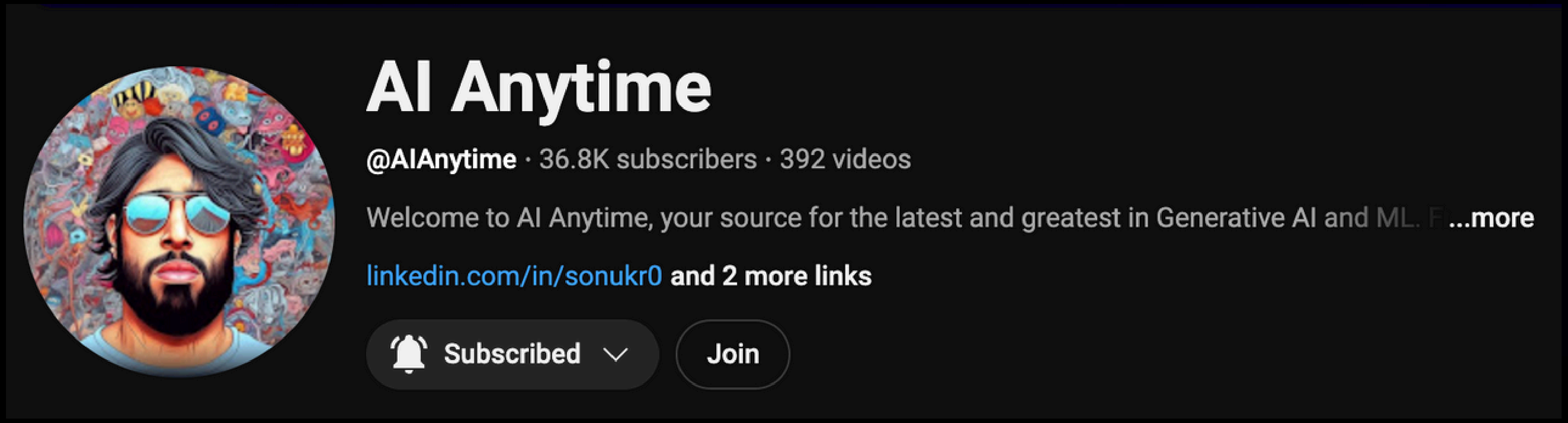
Sonu Kumar, Sporo Health / AI Anytime

Sonu Kumar

AI Engineer | Founder | YouTuber

- 7+ **years in AI**, building AI solutions & automation tools.
- **Founder of Sporo Health,** transforming clinical workflows with AI.
- **YouTube Creator** @AI Anytime, educating 40k+ AI enthusiasts globally.

SPORO HEALTH



Introduction to Causality vs. Correlation

- **Correlation** measures the association between two variables. For example, ice-cream sales and sunglasses sales might be correlated because both increase during summer.

- **Causality** asks: “Does one event cause another?”—for instance, does an advertising campaign increase sales?

Key Idea:
“Correlation is not causation.”
Even if two variables move together, it doesn’t imply one causes the other. Causal inference techniques help us tease apart true cause-effect relationships from mere correlations.

The Impact of a Marketing Intervention Scenario

Imagine you are a data scientist at a retail company. Your marketing team recently launched an online advertising campaign to boost sales. They want to know if the campaign really caused an increase in sales or if the observed uptick is due to other factors (like seasonality or trends).

Questions to Explore:

- What is the effect of the campaign (the treatment) on sales (the outcome)?
- How can we rule out alternative explanations (confounders) for the observed increase in sales?

Why Causal Inference?

Helping us answer “Did the campaign work?” and “How much did it help?”

Core Causal Inference Concepts

- Treatment (Intervention): The advertising campaign.
- Outcome: Sales revenue.
- Confounders: Variables that influence both the likelihood of receiving the treatment and the outcome (e.g., seasonality, store location, product popularity).

We worked for a large healthcare provider. A new treatment (e.g., a medication or procedure) has been introduced to reduce recovery time for patients after surgery. However, patients differ in age, baseline health status, and other factors.

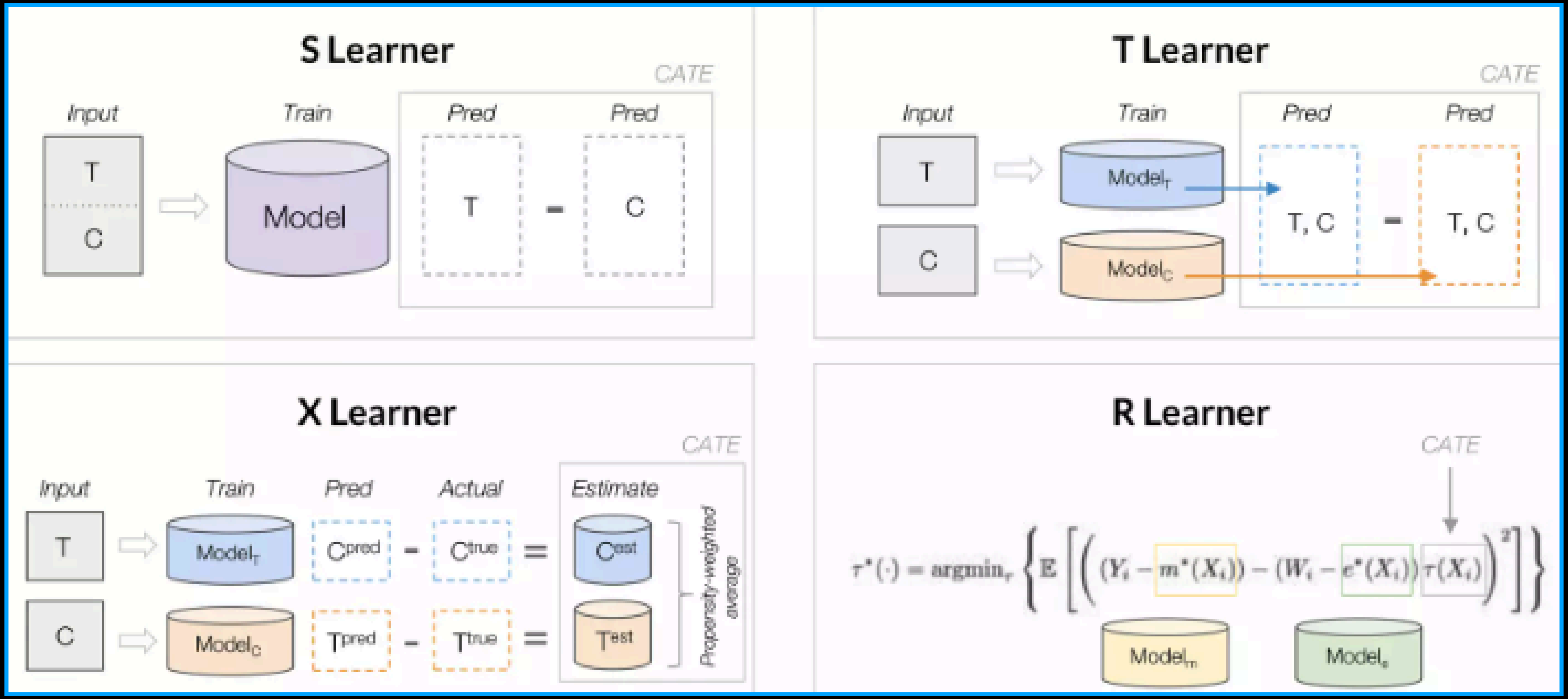
We want to estimate:

- *The overall causal effect of the treatment on recovery time.*
- *How the effect varies across subgroups (i.e., heterogeneous treatment effects).*

We will assume that:

- Treatment: A binary indicator (1 if the patient received the new treatment, 0 otherwise).
- Outcome: Recovery time (in days).
- Confounders: Age, baseline health score, and hospital type.

Family of Meta Learners



To estimate heterogeneous treatment effects (HTEs)—i.e., how the treatment effect varies across individuals.

Uplift Modeling: Explore how to use these techniques to design personalized interventions.

Python Libraries for Causation and Uplift Modeling

CausalML

py-why/dowhy

DoWhy is a Python library for causal inference that supports explicit modeling and testing of causal assumptions. DoWhy is based...

EconML