algoritma
data science education center

CAUSAL INFERENCE FOR BUSINESS DECISIONS: UNDERSTANDING CAUSE & EFFECT

Date: 11 - 14 November 2024

algoritma
learn data science by building

Instructor at Algoritma Data Science School Since 2021

Python, SQL, R
Machine Learning, Data Analytics

Computer Science, Universitas Indonesia

Wulan

hwulanayu

Hamam Wulan Ayu

Teaching Assistant

algoritma
learn data science by building

Victor

VictorNugraha

Victor nugraha

Visualization & Dashboarding

Irfan

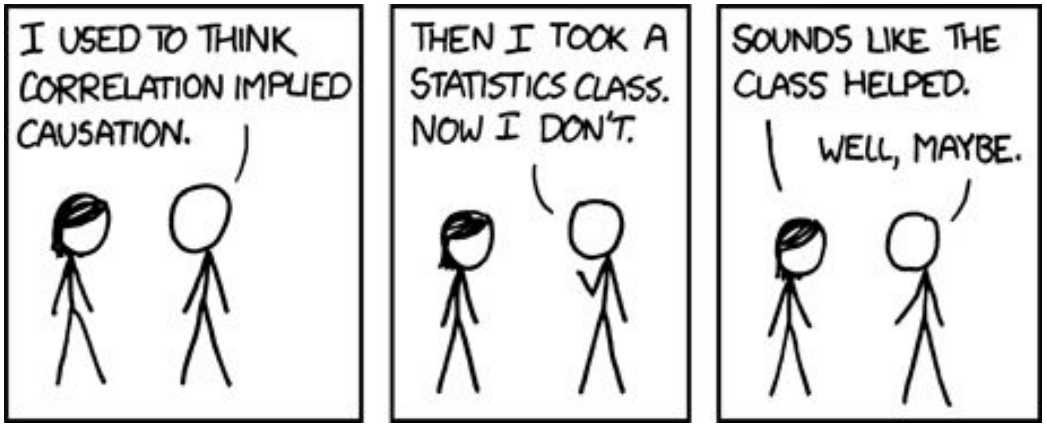
IrfanChairurrachman

Irfan Chairur Rachman

Machine Learning, Data Engineering

Introduction

What is Causality?



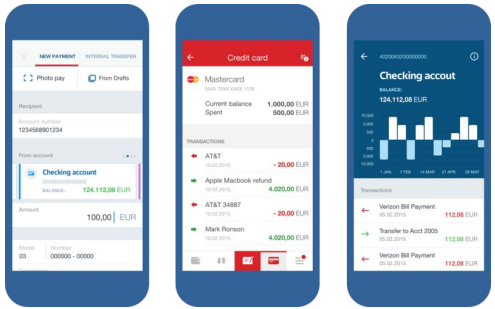
Causality

the study of how things influence one other, the relationship between cause and effect.

Motivation:

- **We rely on causation to make it happen**
- **A/B Testing doesn't always works**

Example problem:



[Mobile app - transaction]

[Tarif - cc usage]

[bundling product insurance]

Causality vs Correlation

(+) Kausalitas dan (+) Korelasi: penjualan jaket meningkat ketika suhu juga meningkat

(+) Kausalitas dan (-) Korelasi: TIngkat pertumbuhan tanaman dengan banyak air yang diberikan

(-) Kausalitas dan (+) Korelasi: tingkat konsumsi margarin dengan tingkat perceraian di Maine

(-) Kausalitas dan (-) Korelasi: tinggi badan dan warna mata

Another spurious correlation: https://www.tylervigen.com/spurious-correlations

Causality

Causal Inference

the process of determining whether an observed association truly reflects a cause-and-effect relationship.

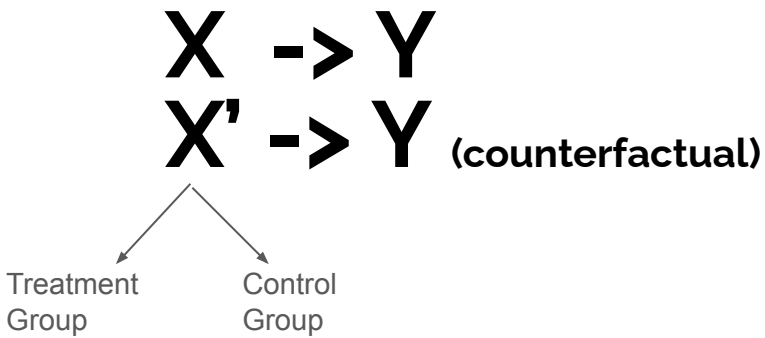
Causal Inference

X -> Y
X' -> Y (counterfactual)

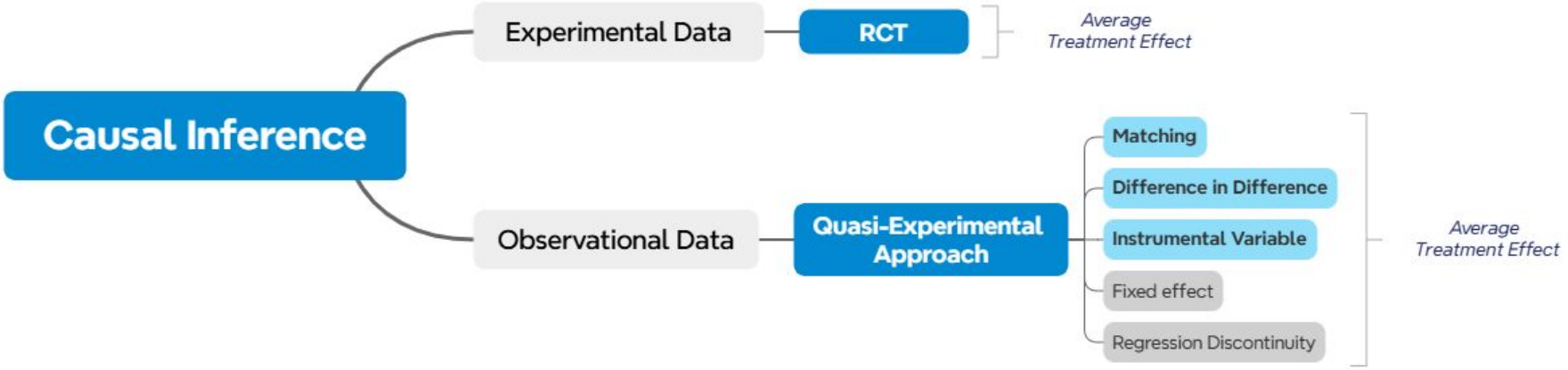
RCT

a.ka Randomized Control Trial

Randomized Controlled Trial



Mindmap


Causal Inference
Experimental Data
RCT
Average Treatment Effect
Observational Data
Quasi-Experimental Approach
Matching
Difference in Difference
Instrumental Variable
Fixed effect
Regression Discontinuity
Average Treatment Effect

algoritma
learn data science by building

Introduction for Programming in Python

algoritma
data science education center

Tools Analogy

Data Analyst

Painter

ANACONDA

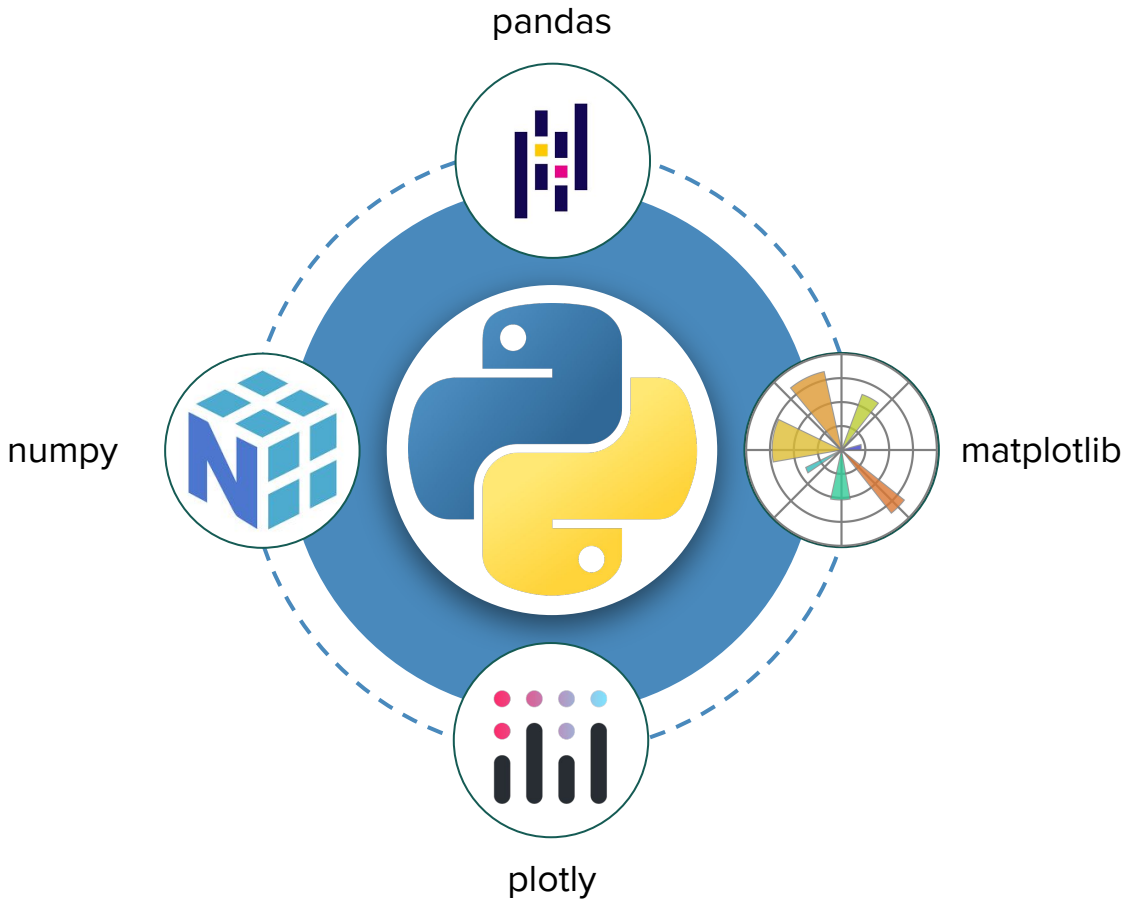
Studio

python™

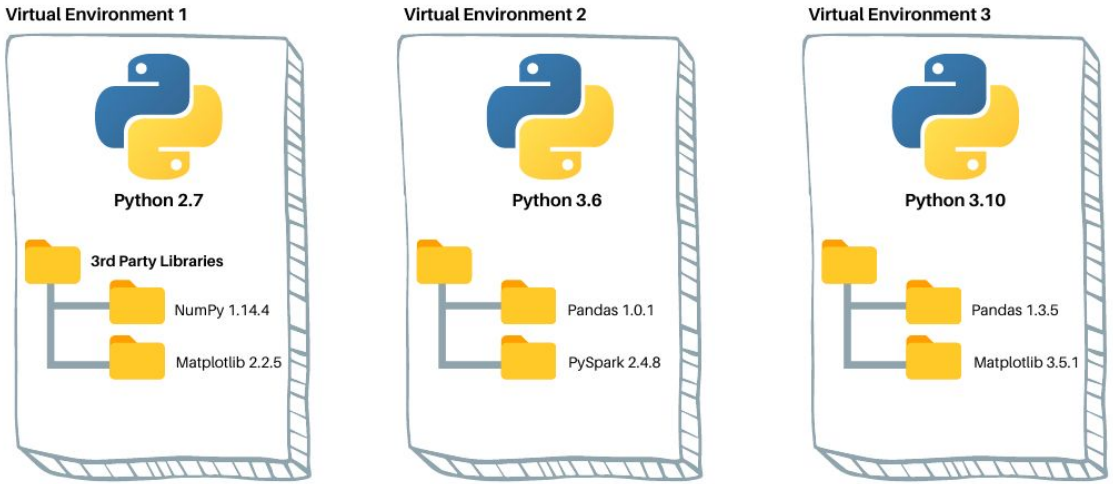
Paint and Brush

Canvas

Package / Library



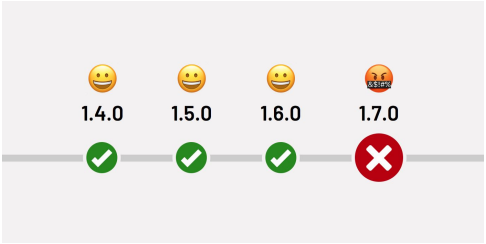
Virtual Environment



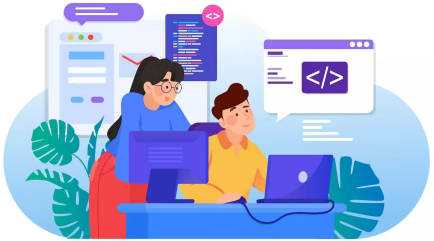
Project 1 *Project 2* *Project 3*

Why do we need Python environments?

You might ask: shouldn't I just install the latest Python version?



Isolate package versions to avoid breaking changes



Sharing virtual environment to enable project collaboration

Publishing or deploying an application requires setting up an environment

Package / Library

Chef + *Raw Materials* + *Kitchen Set* = *New Food* + *Recipe*

You *Data* *Library* *Product* *Notebook*

Visual Studio Code

Open VS Code

How to Create a Virtual Environment

1. Open VS Code
2. Open **terminal in VS Code**
3. Create new virtual environment with:
   conda create --name dss_causality python=3.8
4. Activate the new virtual environment:
   conda activate dss_causality
5. Download requirements.txt from installation post in google classroom
6. Install required libraries
   pip install - r requirements.txt

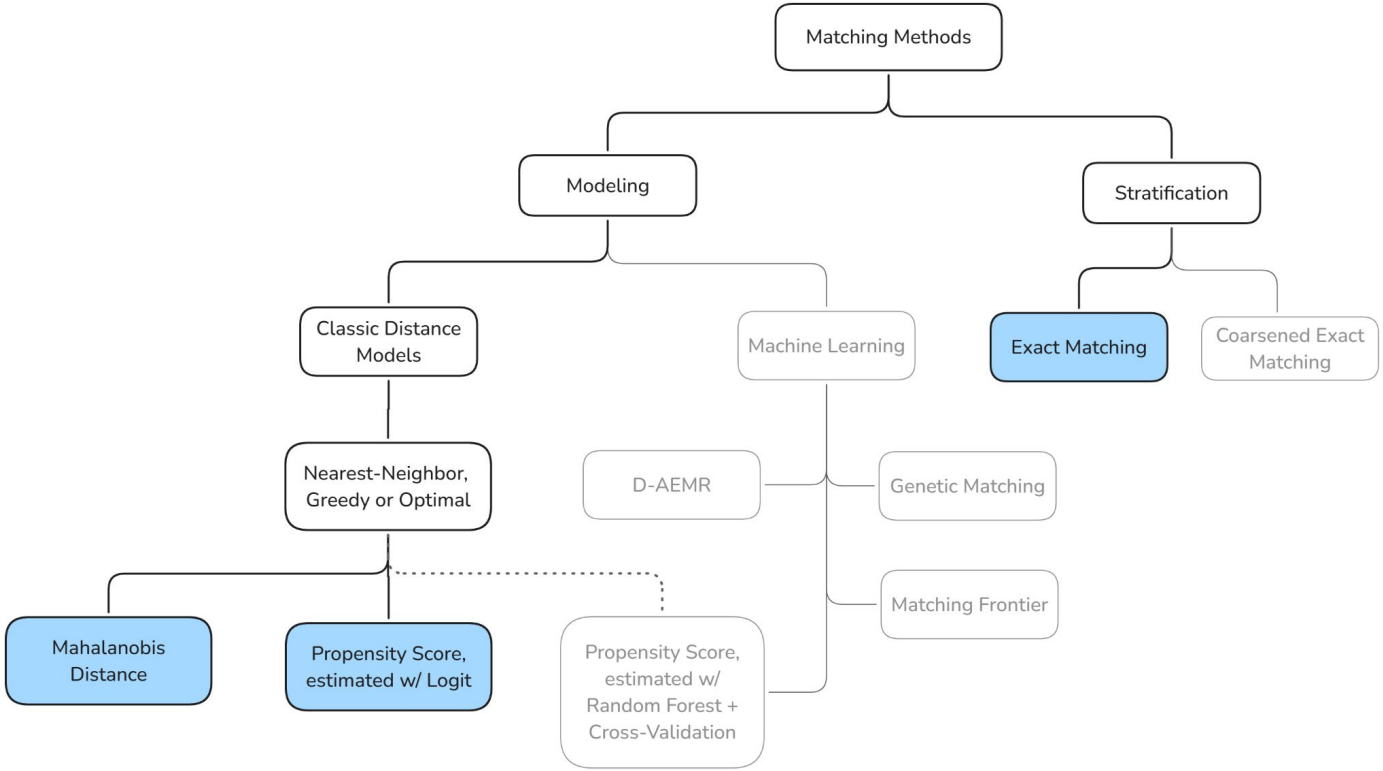
- `conda env list`: check environment list

Workflow Causal Inference untuk data observasional

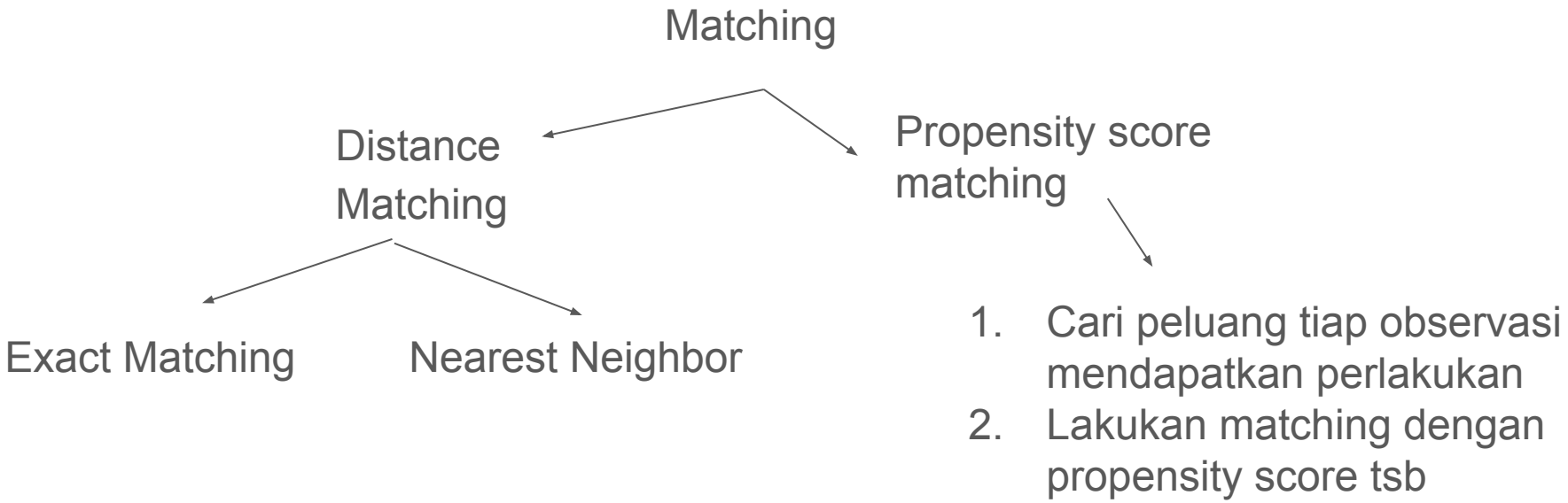
1. Definisikan tujuan dan asumsi
    a. Tentukan treatment, outcome, confounder
2. Membuat diagram causal
3. Preprocessing data
    a. Missing value, nilai outlier, dll
4. Pilih metode causal inference yang cocok dengan kasus dan karakteristik datanya
    a. Regresi
    b. Distance matching
    c. Propensity matching
    d. DiD
    e. Instrumental Variabel
    f. ……
5. Evaluasi dan Asumsi
    a. Statistically significantl
6. Estimasi Efek Kausal
7. Kesimpulan

algoritma
learn data science by building

Matching

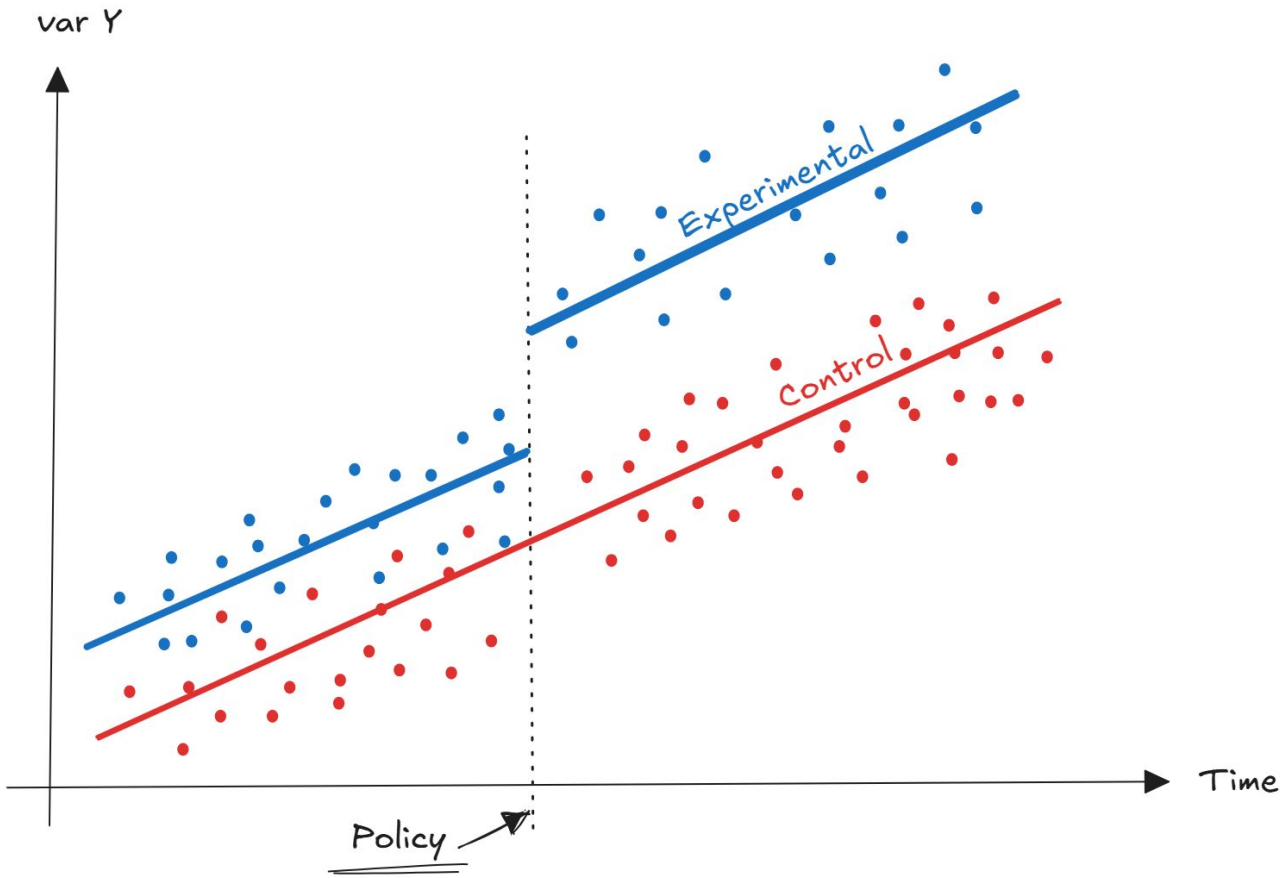


Matching Tree Methods

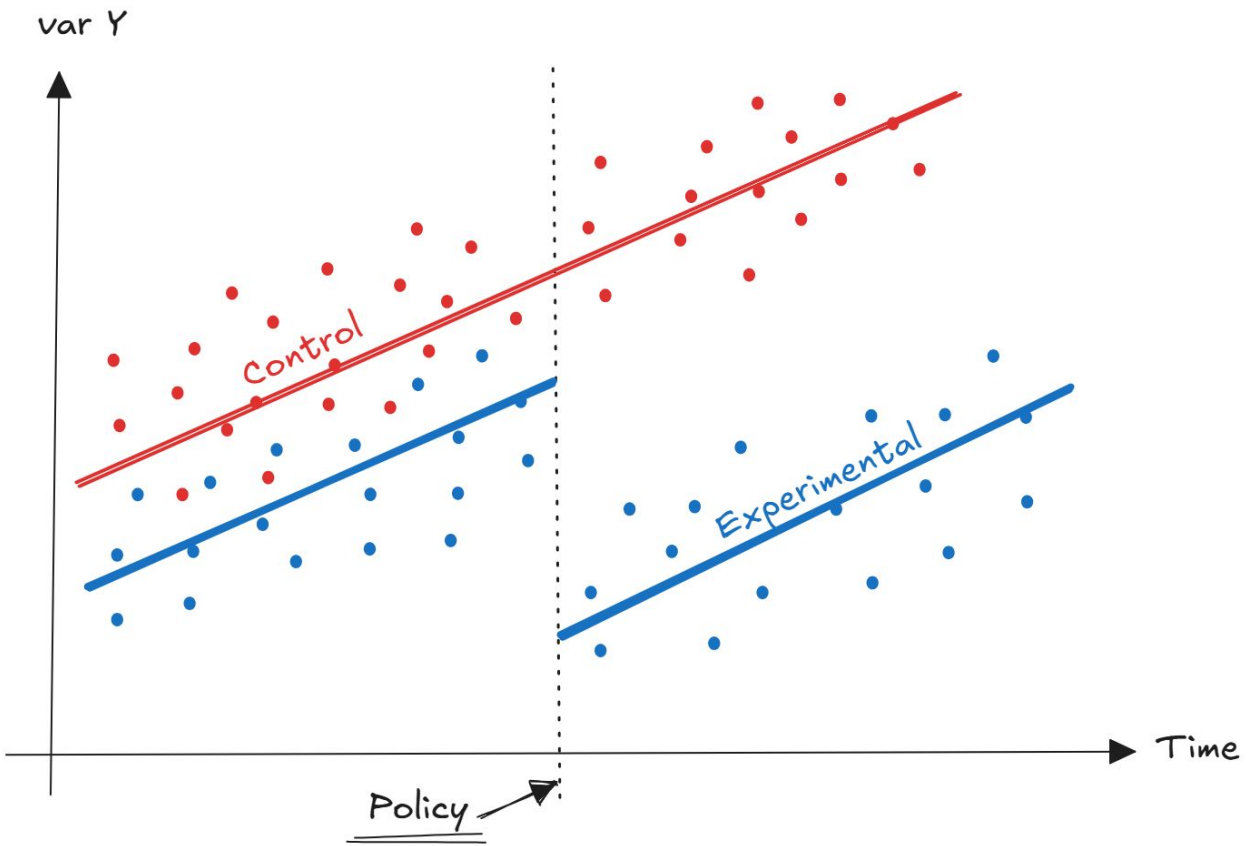

Matching
Distance Matching
Propensity score matching
Exact Matching
Nearest Neighbor
1. Cari peluang tiap observasi mendapatkan perlakukan
2. Lakukan matching dengan propensity score tsb

Differences in Differences

DiD

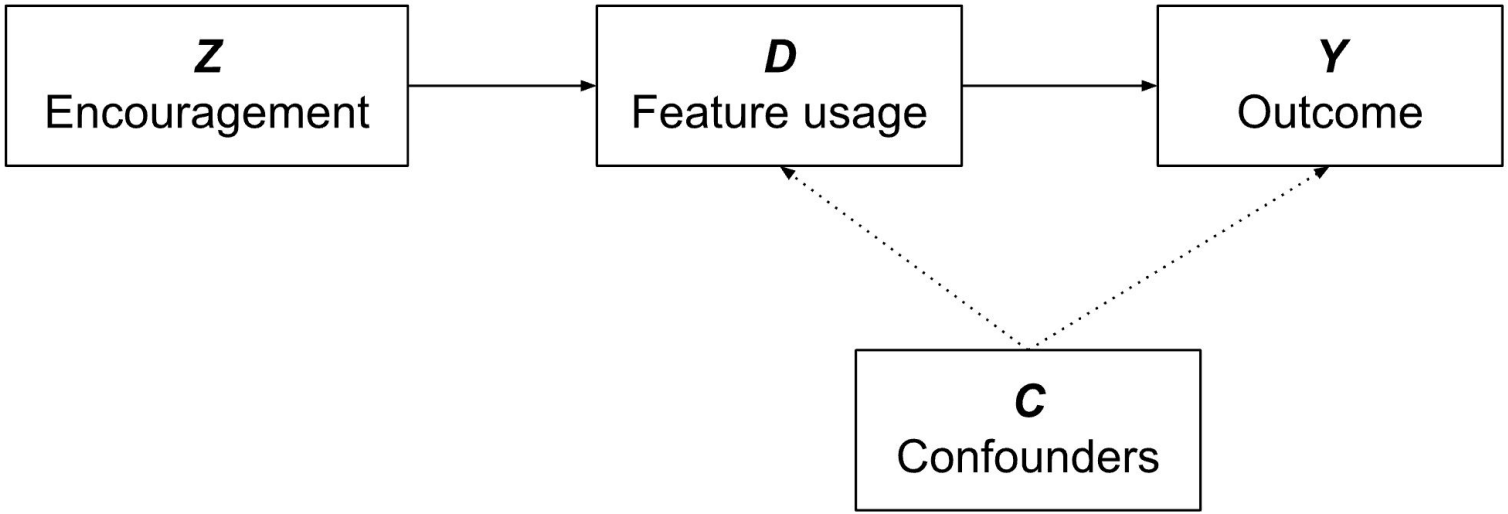


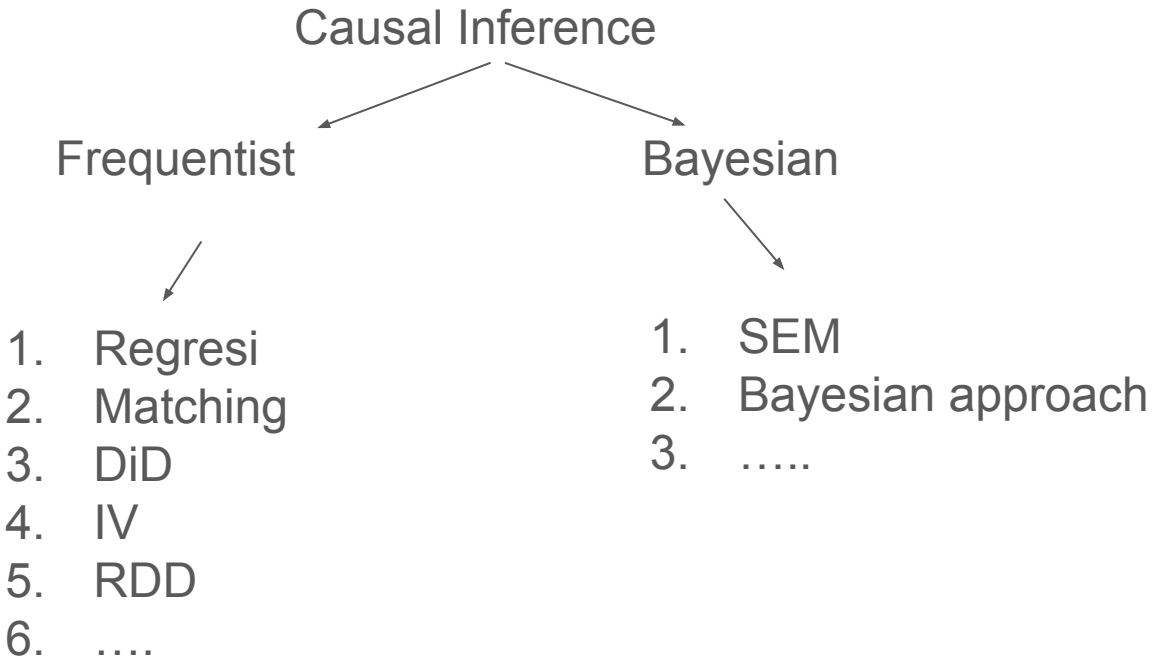
DiD



Instrumental Variable

Spotify Case




Causal Inference
Frequentist
Bayesian
1. Regresi
2. Matching
3. DiD
4. IV
5. RDD
6. ….
1. SEM
2. Bayesian approach
3. …..