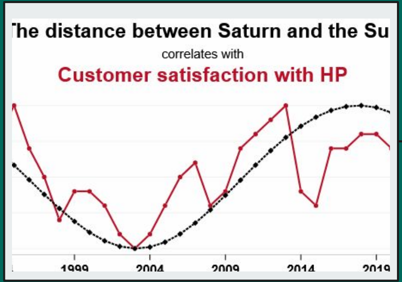
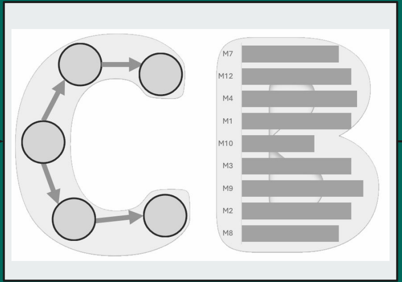
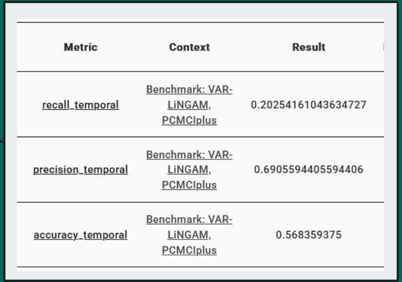
CausalBench: Causal Learning Research Streamlined Understanding Causality



Causality



CausalBench



Metric	Context	Result
recall_temporal	Benchmark: VAR-LiNGAM, PCMCIplus	0.20254161043634727
precision_temporal	Benchmark: VAR-LiNGAM, PCMCIplus	0.6905594405594406
accuracy_temporal	Benchmark: VAR-LiNGAM, PCMCIplus	0.568359375

Benchmarking

Ahmet Kapkiç
Pratanu Mandal
Abhinav Gorantla
Dr. Kasim S. Candan

tutorial.causalbench.org

August 3–7, 2025
KDD2025

What is Causality?

Causality is the relationship between an action, event, or state (the cause) and a resulting event or state (the effect).

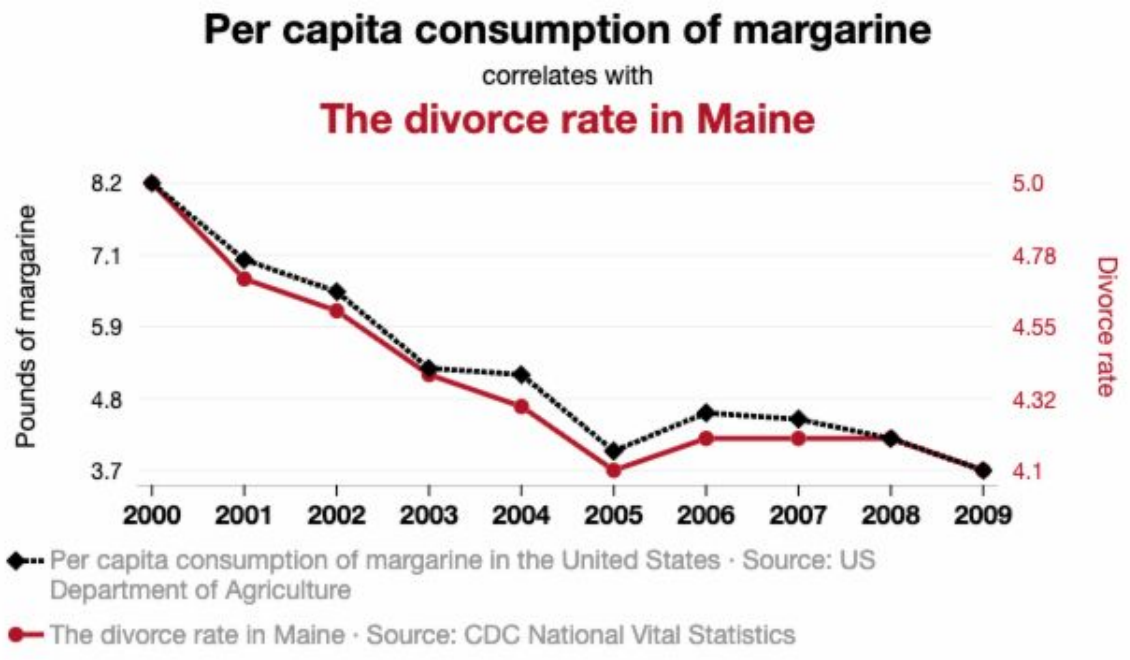
It seeks to answer the question: “Why?”

For example:

- Is rainfall the cause of river flooding?
- What’s the effect of a drug on a patient’s health
- How temperature affects crop yields?

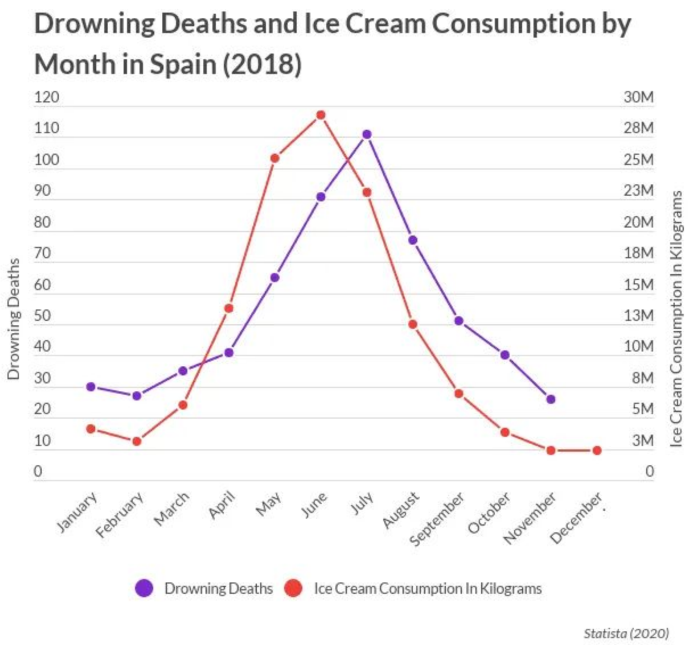
Hook: What does causality produce?

Causality: Correlation is not Causation



https://tylervigen.com/spurious/correlation/5920_per-capita-consumption-of-margarine_correlates-with_the-divorce-rate-in-maine

Ice cream causes drowning?



Ice cream causes drowning?

- A common cause: summer
- Summer is a confounding factor.
- Observational data can lead to wrong conclusions.

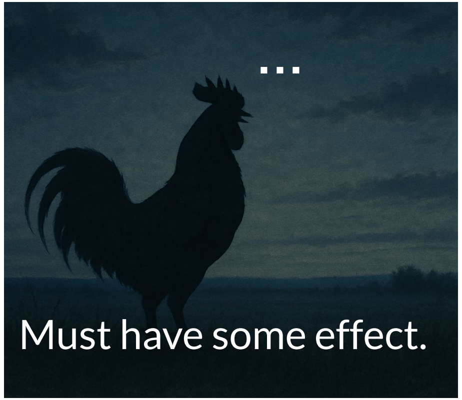
- But how can we learn causality from data?

Causality: How can we learn causality

*Post hoc ergo propter hoc

Intervention on potential cause

Observing without potential cause



Causality: the Simpson Paradox

**Same data, different understanding*

Scenario 1: "Strong"/"Weak" as Precondition

- If *Strong* and *Weak* reflect **baseline susceptibility**.
- **Use grouped data** to examine heterogeneous effects across susceptibility levels.

Scenario 2: "Strong"/"Weak" as Post-Treatment Outcomes

- If *Strong* and *Weak* reflect **response to the vaccine**, they occur after treatment.
- Use **aggregated data** to estimate the overall causal effect.

	Full Population N = 52		
	Success	***Failure***	***Success Rate***
Treatment	20	20	50%
Control	6	6	50%
	Weak, N = 20		
Treatment	8	5	38%
Control	4	3	43%
	Strong, N = 32		
Treatment	12	15	56%
Control	2	3	60%

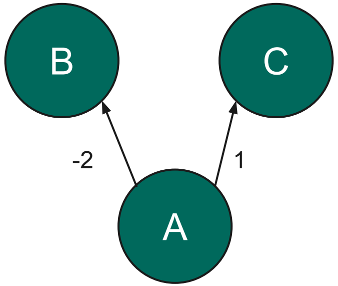
Table: Treatment is the vaccine, Success means not getting the disease. "Strong" and "Weak" refer to patient health conditions.

Causality: Causal Models

1. DAG
2. Potential Outcome
3. SCM or functional causal models (FCM)

Hook: How to explain a cause-effect knowledge to a computer?

Representation: Causal Knowledge - Models



How do we represent causal knowledge?

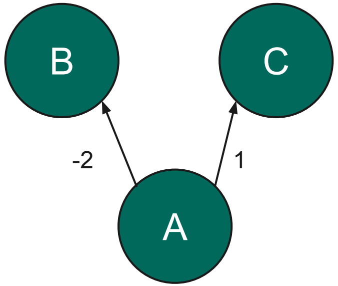
- Directional Acyclic Graphs (DAGs)
  - Variates in the Causal Model are represented as nodes in the DAG.
  - Causal Effects are represented by a directed edge from the node that *causes* the effect to the node that is affected by the cause.
- Structural Causal Matrices (SCM)[1]
  - Causal Model is represented as a set of equations.
  - A variate in a causal model is represented as a function of its parents and external, unobserved noise.
  - The causal model represented as a DAG on this slide can be represented as an SCM in the following manner:
    - B = 2*A + Ub
    - C = 1*A + Uc
    - Where, Ub and Uc are external unobserved noise on variates B and C.

[1] Pearl, J. (2009). *Causality*. Cambridge university press.

Representation: Causal Knowledge - Models

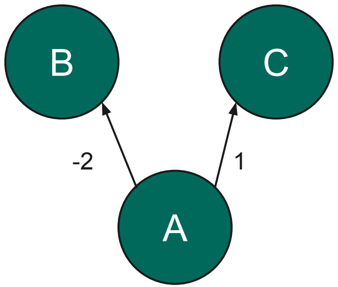
How do we represent a DAG or an SCM in a way a computer can understand?

- Adjacency matrices is one of the common ways a DAG and (linear) SCMs can be encoded.
- The DAG used as an example here can be represented as an adjacency matrix in the following manner:



$$\begin{bmatrix} 0 & -2 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Representation: Causal Knowledge - Models



- Structural Causal Matrices (SCM)[1]
    - Causal Model is represented as a set of equations.
    - A variate in a causal model is represented as a function of its parents and external, unobserved noise.
    - The causal model represented as a DAG on this slide can be represented as an SCM in the following manner:
        - B = 2*A + Ub
        - C = 1*A + Uc
        - Where, Ub and Uc are external unobserved noise on variates B and C.

[1] Pearl, J. (2009). *Causality*. Cambridge university press.

What Is Temporal Causal Discovery?

- Extracting cause–effect links from time-ordered data
- Causes must precede their effects

Core Idea:

Learn a directed graph of lagged edges where past values of X help predicting current Y

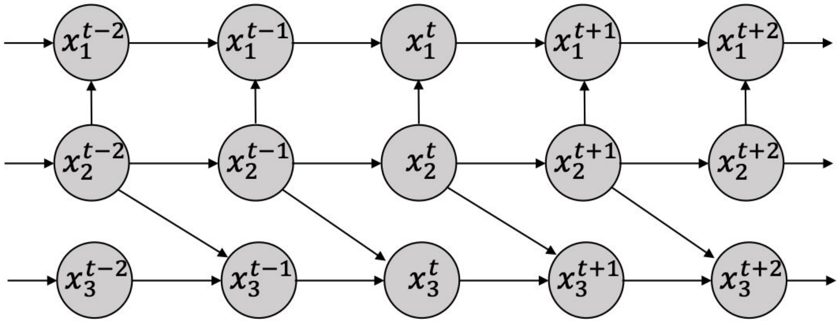
$$X_{t-T} \longrightarrow Y_t$$

Key Difference from Static Causality:

Captures time delays and feedback loops

Not just instantaneous associations

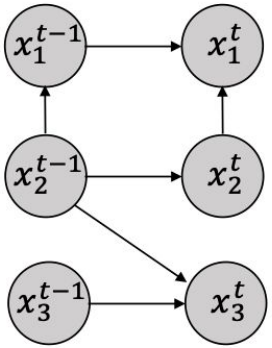
Three Types of Temporal Causal Graphs



(a) Full time causal graph

- Complete graph of dynamic system
- Usually difficult to discover due to the single observation for each series at each time point

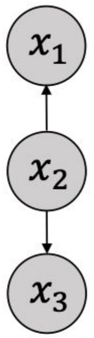
Three Types of Temporal Causal Graphs



(b) Window causal graph

- Assumes time-homogenous causal structure
- The size of window causal graph represents the maximum lag in the full-time causal graph

Three Types of Temporal Causal Graphs



(c) Summary causal graph

- Each individual time-series variable is merged into a single node to create the summary causal graph.
- Represents the causal relations among the time series without displaying any time lags.

Models: Granger Causality (VAR)

- Assumption of linear time-series dynamics
- A time series X Granger-causes Y if past values of X provide unique, significant information about future values of Y
- Learns predictive causality, not necessarily true mechanistic causation.

Hyperparameters

- Maximum lag order: Defines how many past steps will be considered
- Significance level: Threshold to decide causal vs. non-causal

Models: PCMCI[+]

- Constraint-Based method
- Combines linear or nonlinear conditional independence tests to discover the window causal graph
- Initializes by constructing a partially connected graph, where all pair of nodes (x_i^{t-k}, x_j^t) are directed as $x_i^{t-k} \rightarrow x_j^t$ if $k > 0$
- Removes all unnecessary edges based on conditional independence

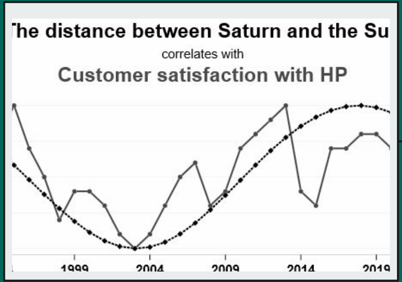
Hyperparameters

Models: Dynamic Bayesian Networks

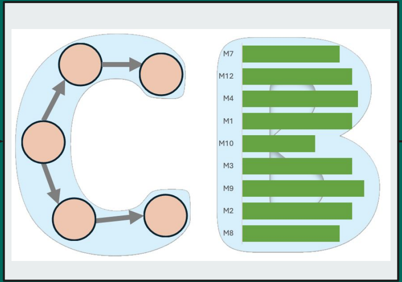
- Score-Based method
- Probabilistically scores multiple models and output the most probable one

Questions?

CausalBench: Causal Learning Research Streamlined
Understanding CausalBench



Causality



CausalBench

Metric	Context	Result
recall_temporal	Benchmark: VAR-LiNGAM, PCMCIplus	0.20254161043634727
precision_temporal	Benchmark: VAR-LiNGAM, PCMCIplus	0.6905594405594406
accuracy_temporal	Benchmark: VAR-LiNGAM, PCMCIplus	0.568359375

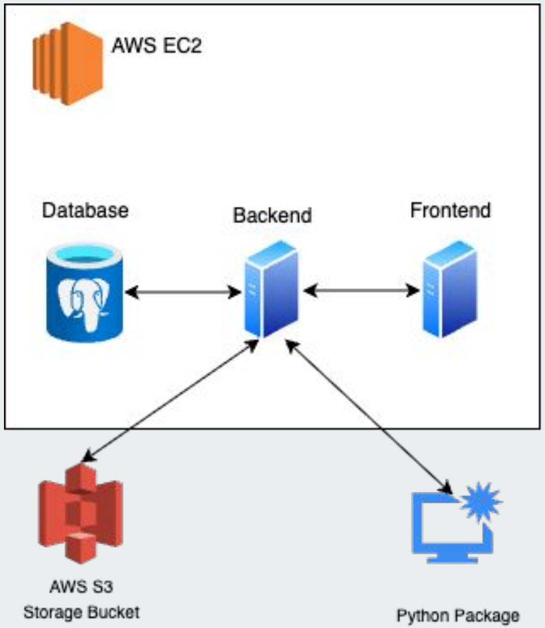
Benchmarking

Ahmet Kapkiç
Pratanu Mandal
Abhinav Gorantla
Dr. Kasim S. Candan

tutorial.causalbench.org

August 3-7, 2025
KDD2025

Introduction: CausalBench



- CausalBench is a benchmarking platform for Causal Learning research.
- The goals of CausalBench are:
  - Facilitating collaboration between researchers in sharing their datasets, models and metrics.
  - Enabling reproducibility of experiments.
- CausalBench allows researchers to share and publish their experimental setups and results.

Introduction: CausalBench

- Transparent experiment results. All public experiment results are published to Zenodo.
- Causal Analysis in **CausalBench** - Analyzes experiment results to determine the causal relationship between experiment parameters and results.
- Causal Recommendation in **CausalBench** - Causal Analysis results are used to recommend new experiment settings.

Components of CausalBench

docs.causalbench.org

- CausalBench contains three components:
  - Python Package
  - The Web Server: Backend + Frontend
- The python package handles the process of benchmarking.
- The web backend receives the results from the python package and publishes it to zenodo.
- The web frontend provides users with a GUI to browse benchmark runs, datasets, models, metrics and contexts already published to causalbench.org.

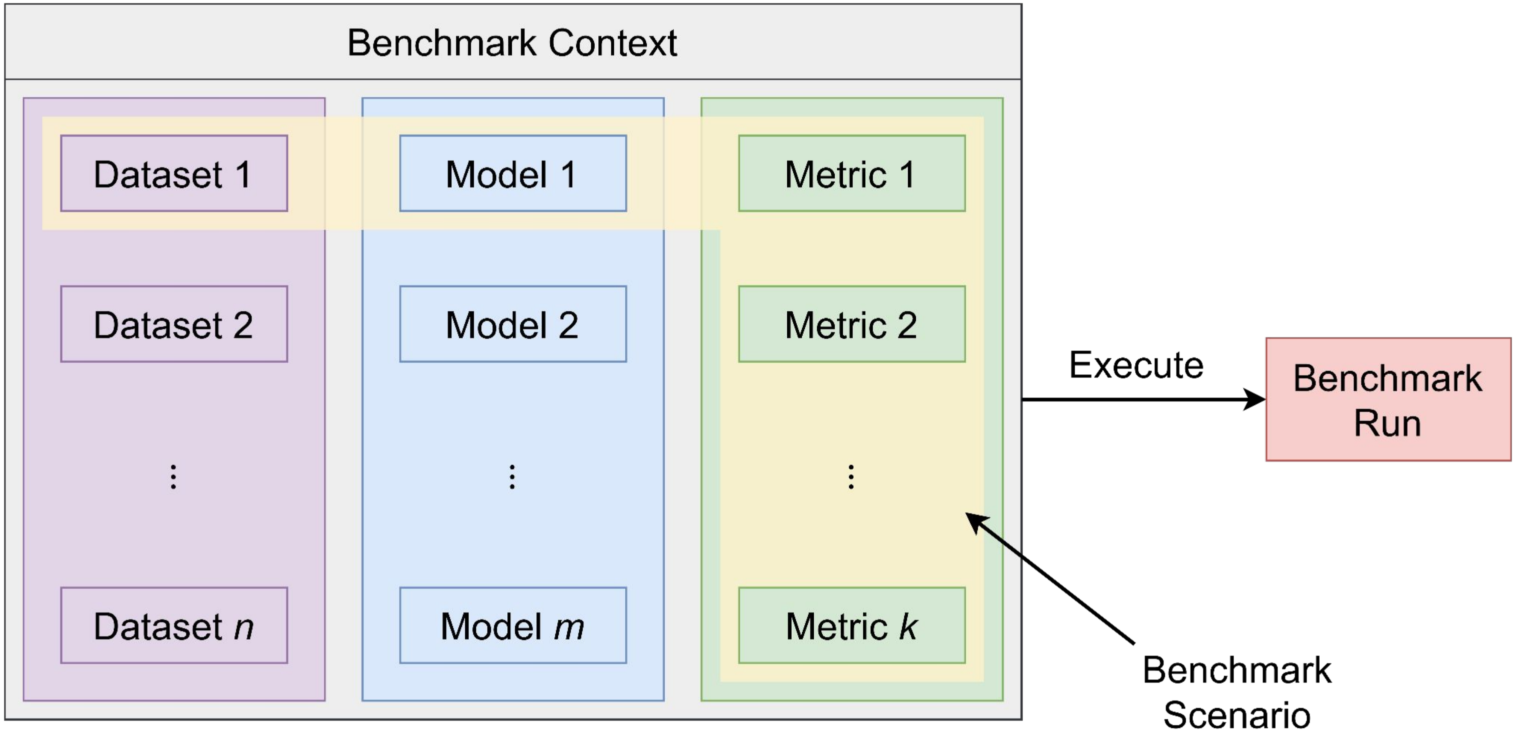
Importance of Proper Benchmarking

- It is important to properly document the experiment and hardware setup when presenting results in a technical/research document.
- Documenting the software and hardware setup makes it easier for the readers of the technical/research document to interpret the results.
- Proper documentation of experimental setup also ensures reproducibility of experiments.

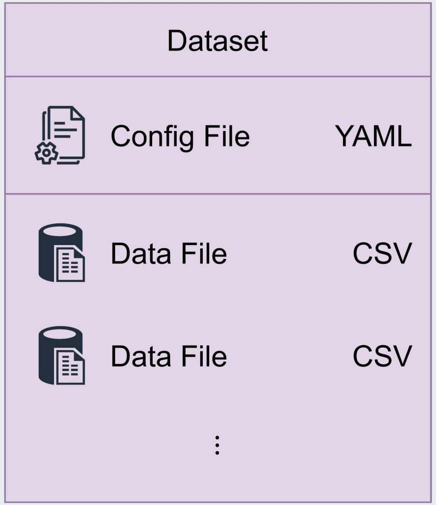
Importance of Profiling

- CausalBench automates this process of documenting the hardware and software configuration.
- Such precise benchmarking is crucial also because it is difficult to compare results from different researchers without knowing the complete context in which the experiment was performed.
- Profiling experiments gives additional insights to the researcher on how their algorithm impacts hardware usage (like memory and disk).

CausalBench: Modules



CausalBench: Modules (Dataset)



Config file:

- Metadata - name, description, and URL
- Names and metadata of data files
- Specifies structure
  - Number of rows
  - columns – number, name, data type, description
  - Index – time, space, etc.

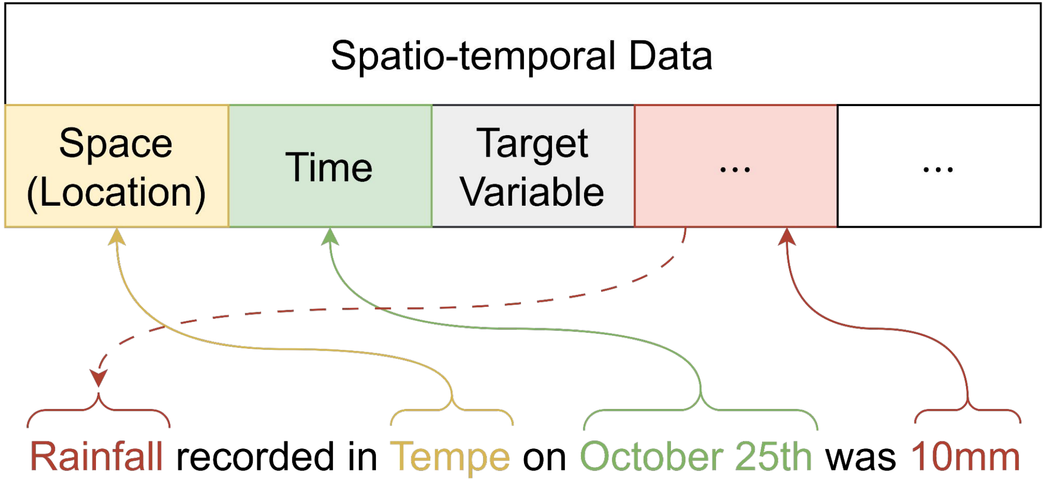
Data file:

- Tabular data
- Data formats
  - Spatio-temporal Data
  - Spatio-temporal Graph

CausalBench: Modules (Dataset) – Data Formats

Spatio-temporal Data:

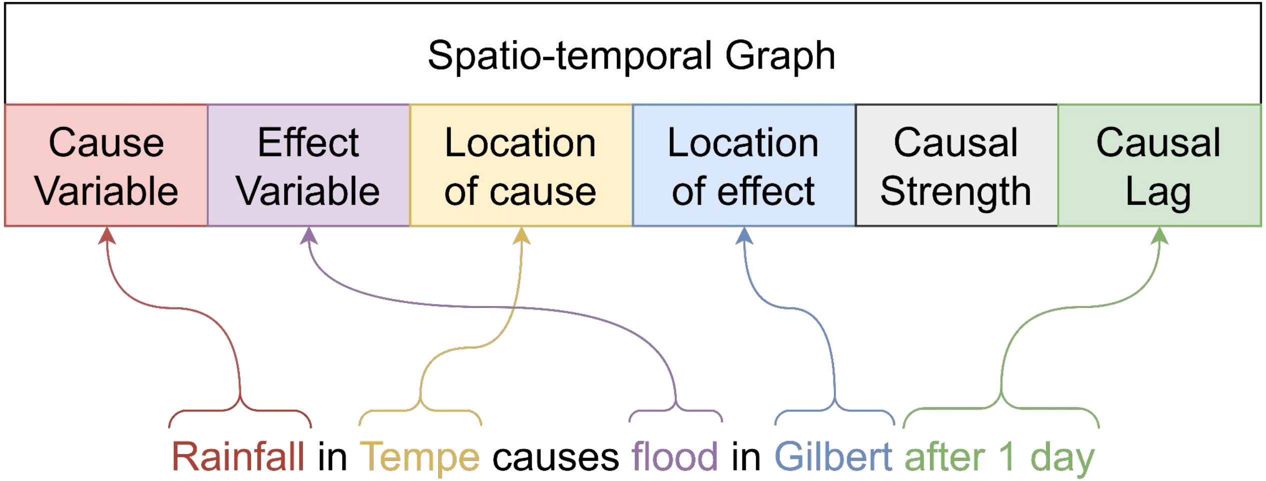
- Tabular format



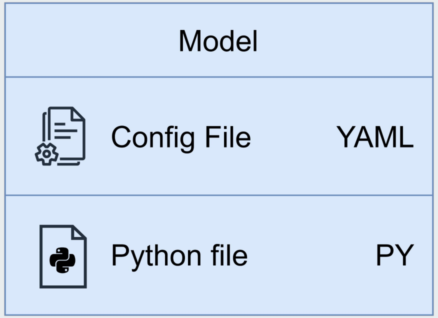
CausalBench: Modules (Dataset) – Data Formats

Spatio-temporal Graph:

- Tabular format



CausalBench: Modules (Model)



Config file:

- Metadata – name, description, and URL
- Name and metadata of python file
- Task – Causal Discovery, Causal Inference, etc.
- Hyperparameters – data type, description, and default value

Python file:

- Function to take accept inputs and provide outputs
    - Comply with function signature specified by task

CausalBench: Modules (Metric)

Metric	
Config File	YAML
Python file	PY

Config file:

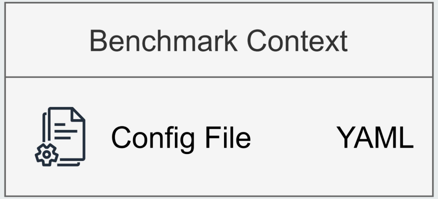
- Metadata – name, description, and URL
- Name and metadata of python file
- Task – Causal Discovery, Causal Inference, etc.
- Hyperparameters – data type, description, and default value

Python file:

- Function to take accept inputs and provide outputs
    - Comply with function signature specified by task

Same structure as model

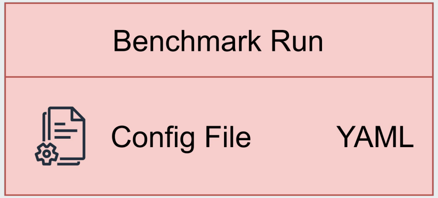
CausalBench: Modules (Benchmark Context)



Config file:

- Metadata – name, description, and URL
- Task – Causal Discovery, Causal Inference, etc.
- Datasets
  - Dataset IDs and versions
  - Data files to task mapping
- Models
  - Model IDs and versions
  - Model hyperparameters (if not using default values)
- Metrics
  - Model IDs and versions
  - Metric hyperparameters (if not using default values)

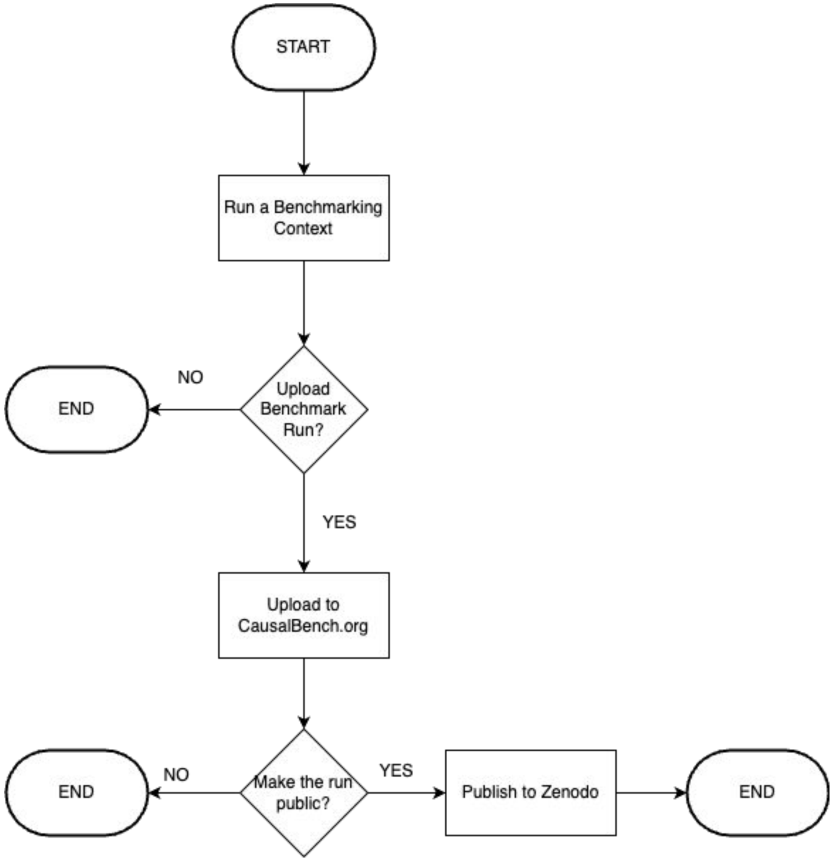
CausalBench: Modules (Benchmark Run)



Config file:

- Reference to Benchmark Context
  - Benchmark Context ID and version
- Hardware Information – Operating System, CPU, GPU, Memory, Disk
- Scenarios
  - Consists of 1 dataset, 1 model, and multiple metrics
  - Dataset
    - ID and version
  - Model and Metrics
    - IDs and versions
    - Output
    - Profiling information – Execution time, CPU used, GPU used, etc.

CausalBench Benchmarking Flow

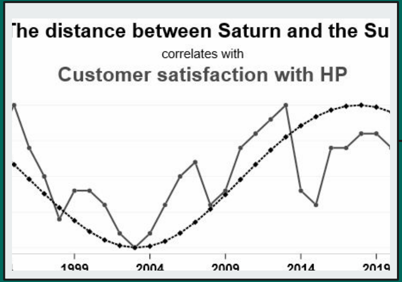


CausalBench Package Setup

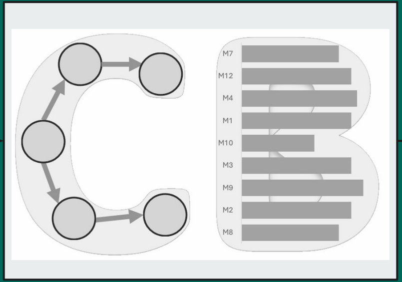
- Pre Requisites: Python (>= 3.10) and pip.
- Install CausalBench python package using "`pip install causalbench-asu`".
- To use the **CausalBench** package, you need a **CausalBench** account.
- On first use, **CausalBench** package will prompt user to input their credentials.

```
WARNING:root:Failed to initialize 'pynvml' library: NVML Shared Library Not Found
WARNING:root:Failed to import 'pyadl' library: name 'struct_AdapterInfo' is not defined
Credentials required
Email: user@gmail.com
Password:
```

CausalBench: Causal Learning Research Streamlined Hands-On Benchmarking



Causality



CausalBench

Metric	Context	Result
recall_temporal	Benchmark: VAR-LiNGAM, PCMCIplus	0.20254161043634727
precision_temporal	Benchmark: VAR-LiNGAM, PCMCIplus	0.6905594405594406
accuracy_temporal	Benchmark: VAR-LiNGAM, PCMCIplus	0.568359375

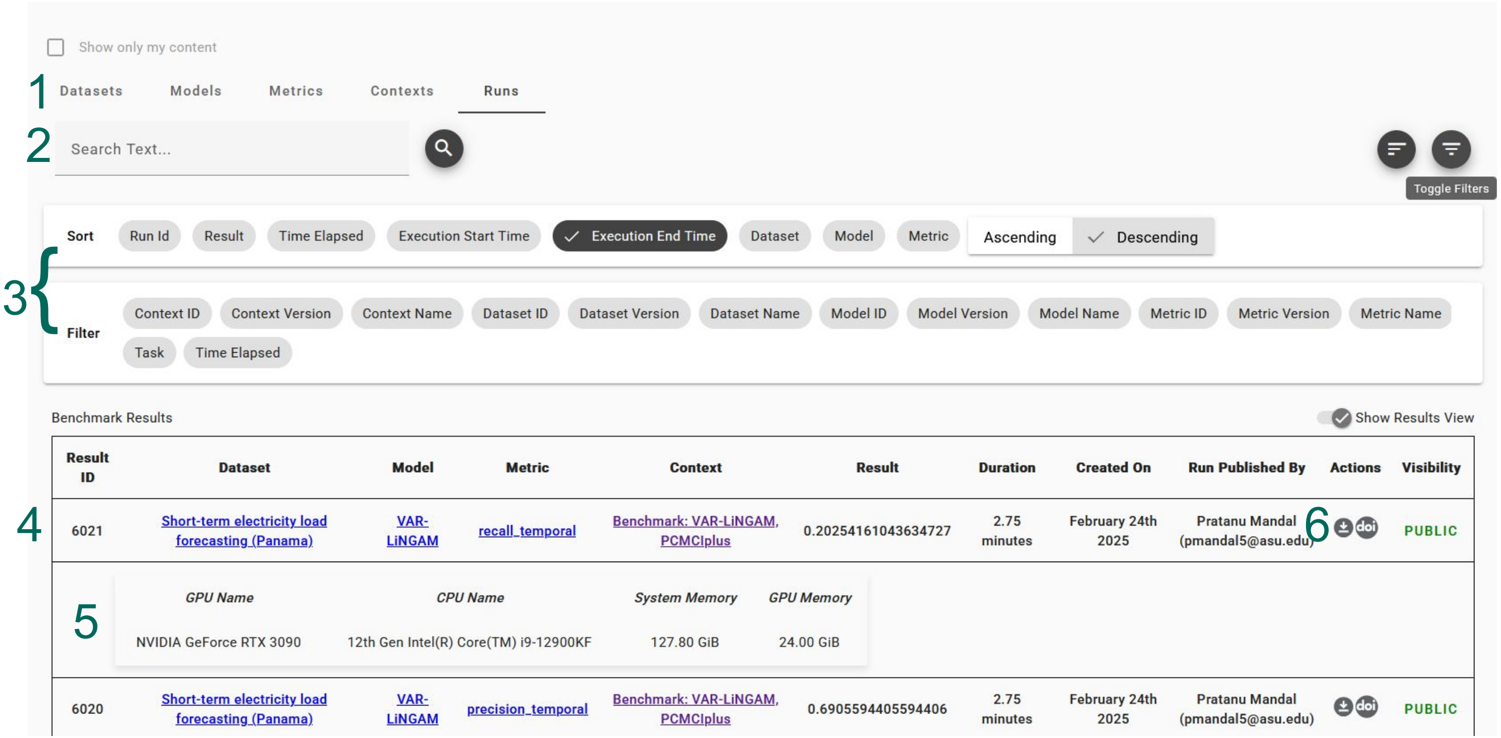
Benchmarking

Ahmet Kapkiç
Pratanu Mandal
Abhinav Gorantla
Dr. Kasim S. Candan

tutorial.causalbench.org

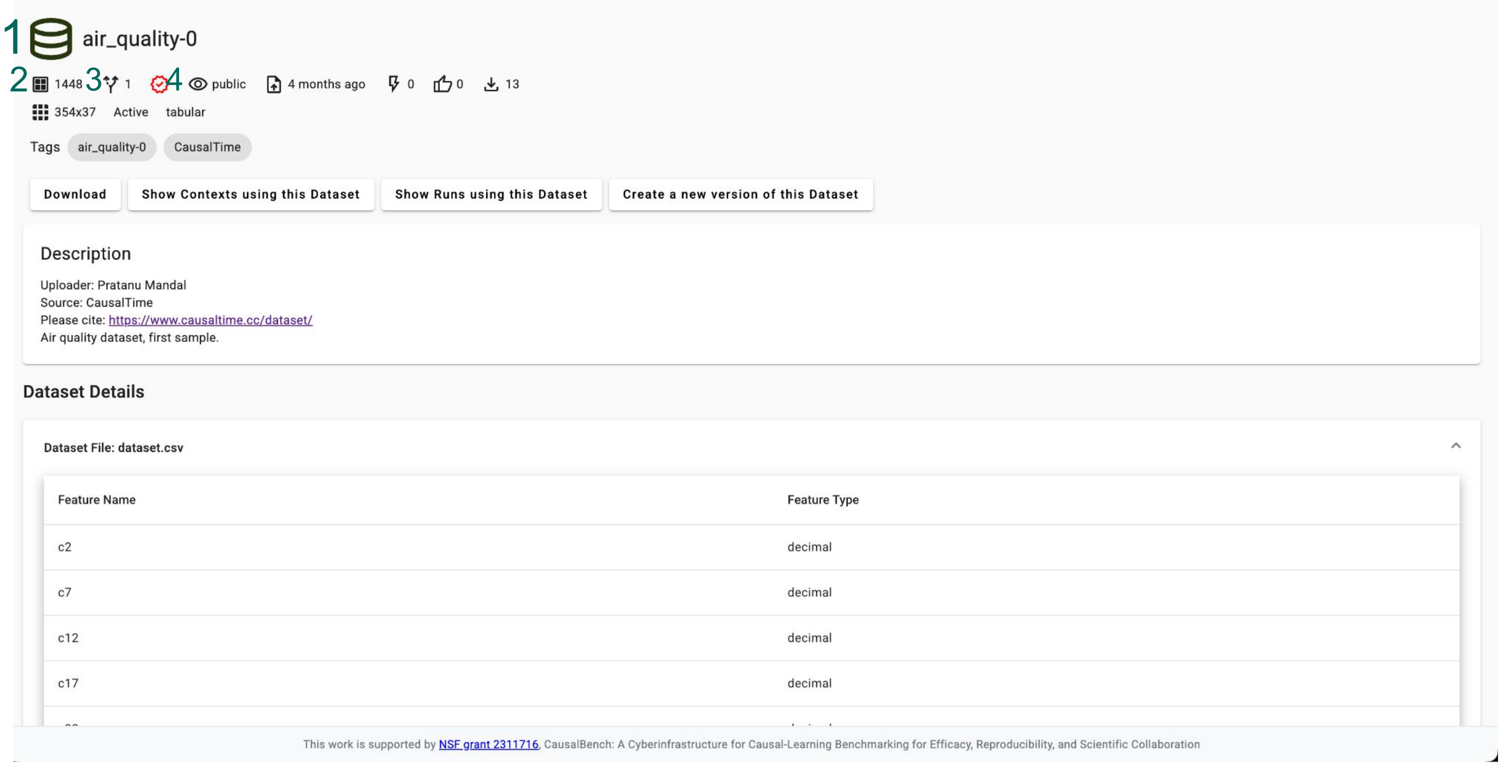
August 3–7, 2025
KDD2025

CausalBench Hands-On: Browsing Repositories



1. Repository selector
2. Search Function
3. Filter/Sort
4. Detail overview
5. On-demand details
6. Download/Cite

CausalBench Hands-On: Dataset Detail Page



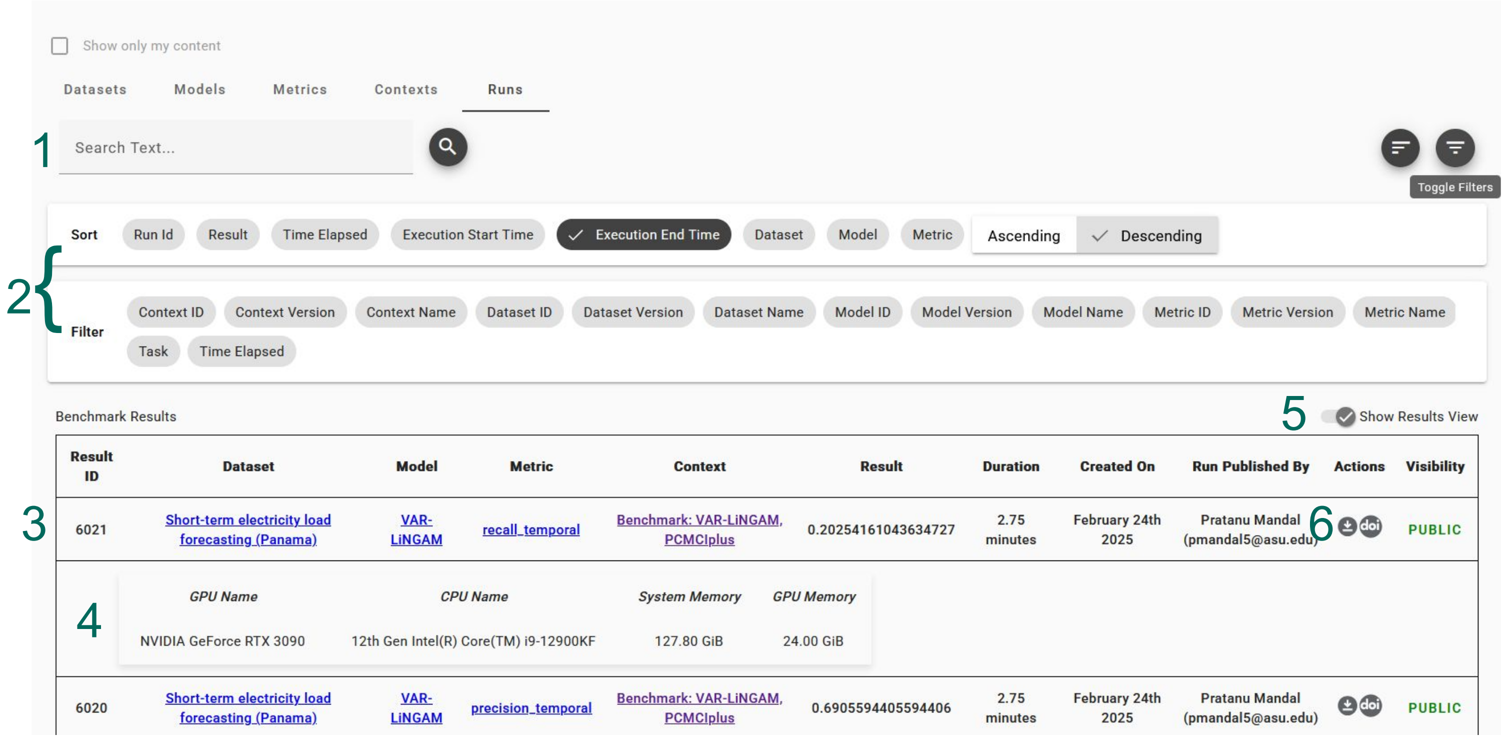
1. Dataset Name
2. Dataset ID
3. Dataset Version
4. Dataset Visibility

CausalBench Hands-On: Building a Context

```
dataset1: Dataset = Dataset(module_id=1, version=1)
```

```
context1: Context = Context.create(module_id=10,
                                  name='Context1',
                                  description='Test static context',
                                  task='discovery.static',
                                  datasets=[(dataset1, {'data': 'file1', 'ground_truth': 'file2'})],
                                  models=[model1, model2],
                                  metrics=[metric1, metric2])
```

CausalBench Hands-On: Browsing Benchmarks



1. Search function
2. Filter/Sort
3. Result overview
4. On-demand details
5. Changing result view
6. Download/Cite

Thank you!

Any questions?

CausalBench

(Product)

KDD Tutorial

(Usage)

Docs/Github

(Contribution)

Further questions? Feedbacks? Want to use CausalBench? support@causalbench.org / akapkic@asu.edu