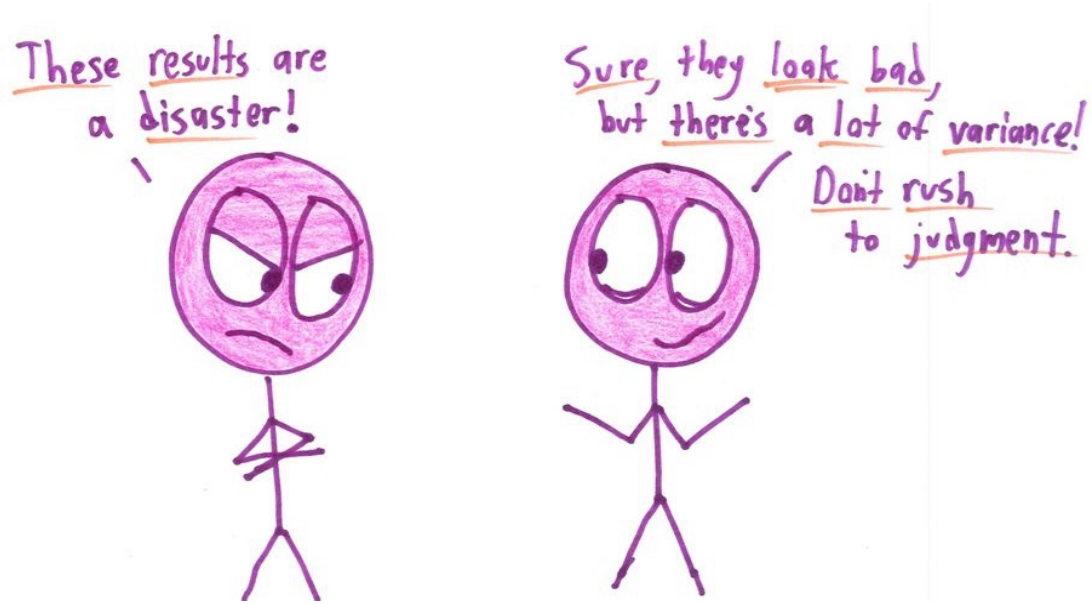
Using Triangulation to Evaluate Machine Learning Models

Andrew Fast, Ph.D.
@counterflowai
https://counterflow.ai

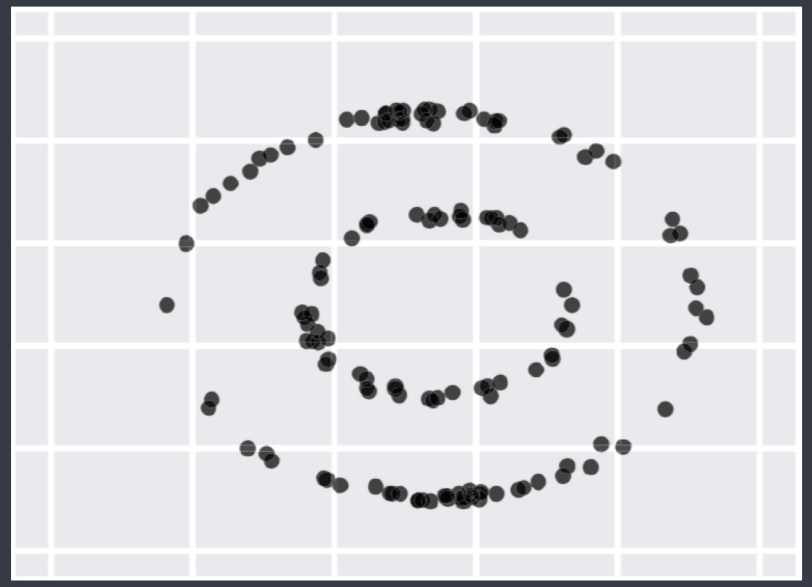
"A statistic is an imperfect witness, it tells the truth but not the whole truth."

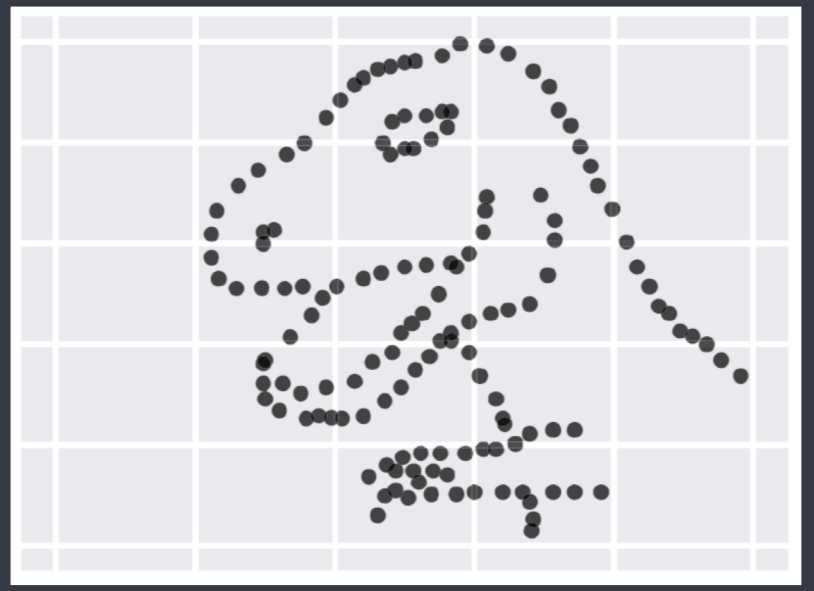
- Ben Orlin, *Math with Bad Drawings (2018)*

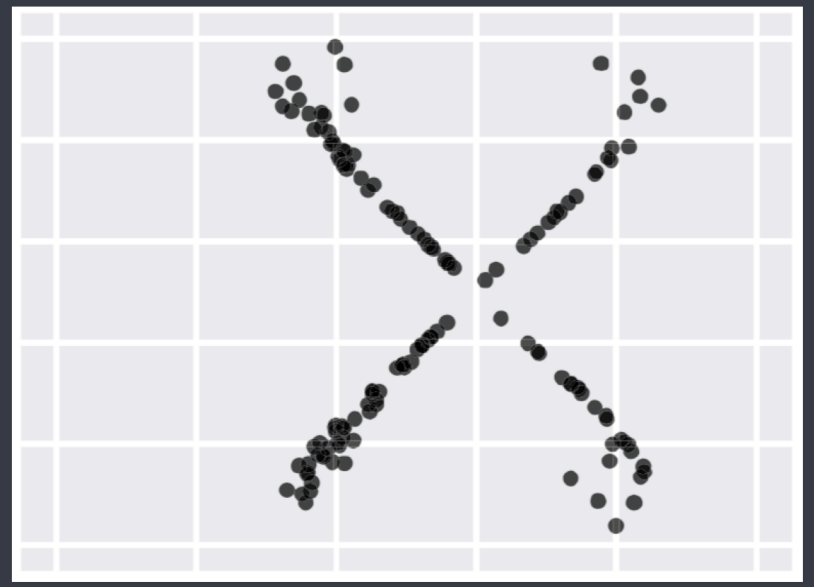


CounterFlow AI

Meet the “Datasaurus”





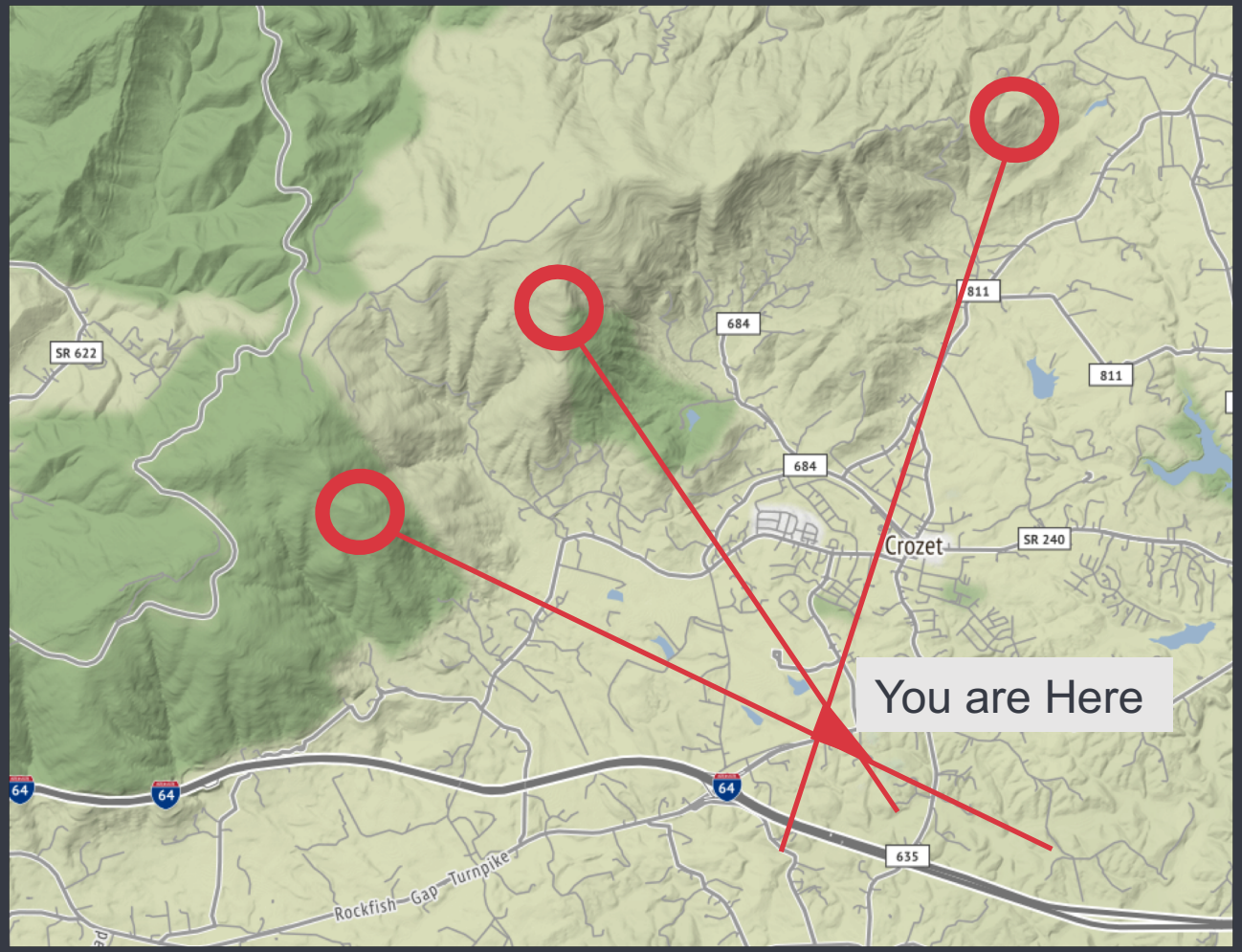


- Each of these datasets have the same basic sufficient statistics to 2 decimal places
- *$\bar{x}$ =54.26, $\bar{y}$ = 47.83, sdx = 16.76, sdy = 26.93, Pearson’s r = -0.06*

CounterFlow ai

Triangulation in Real Life

- Use relative direction to determine absolute position
- Requires views of multiple landmarks
- Precision comes from fusing information



CounterFlow ai

Triangulation Framework for ML

	Model	Metric	Data	Focus
Key Idea	Ablation	Falsifiability	No Free Lunch	Right Question
Landmarks	Default Baseline	Point Metrics	Train/Test/Eval	Streetlight Effect
	Simple Model	Relative Metrics	Cross-validation	Type III Errors
	Feature Selection	Across Thresholds	Multiple Data Sets	Counterfactuals
	Remove Structure	Visualization	Randomization	

CounterFlow ai

Triangulating the Model

Model

Key Idea

Default Baseline

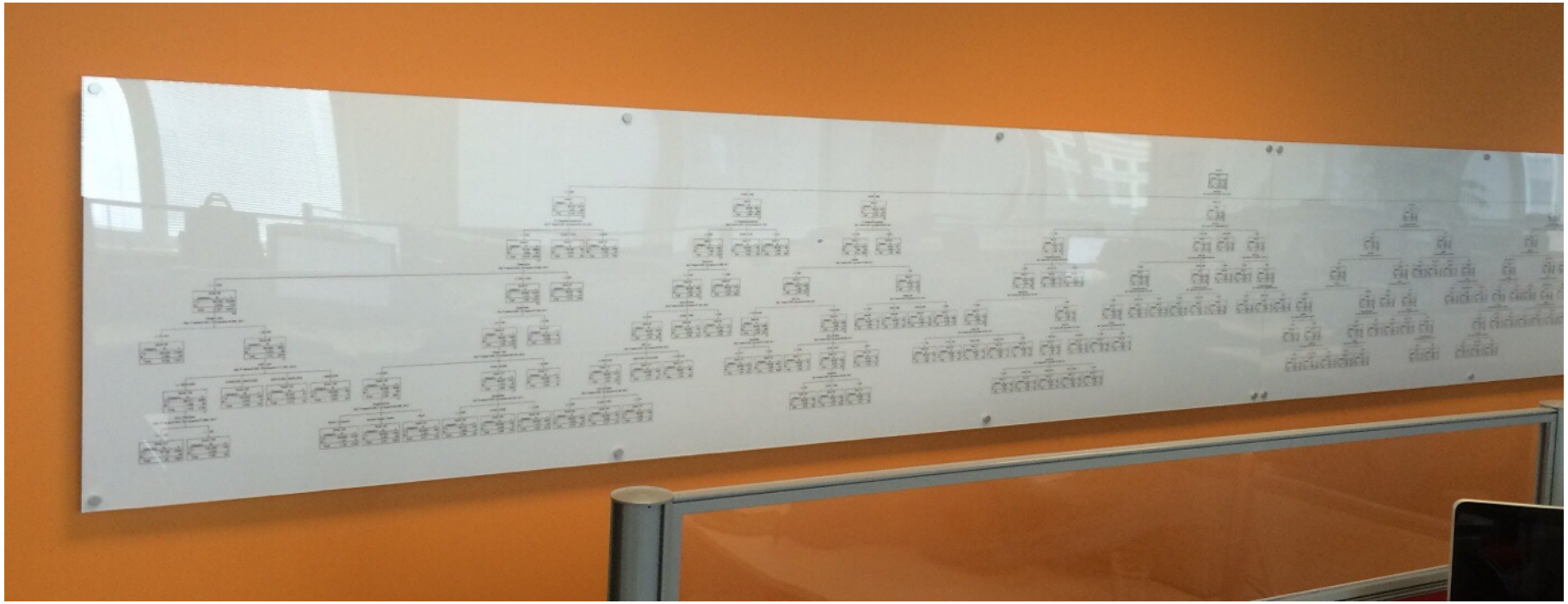
Simple Model

Feature Selection

Remove Structure

CounterFlow ai

Ablation



- *Ablation* - Removing critical parts of a model to test performance
- *Overfitting* – Model inefficiency resulting in unnecessary complexity in the model

Model

Key Idea

Default Baseline

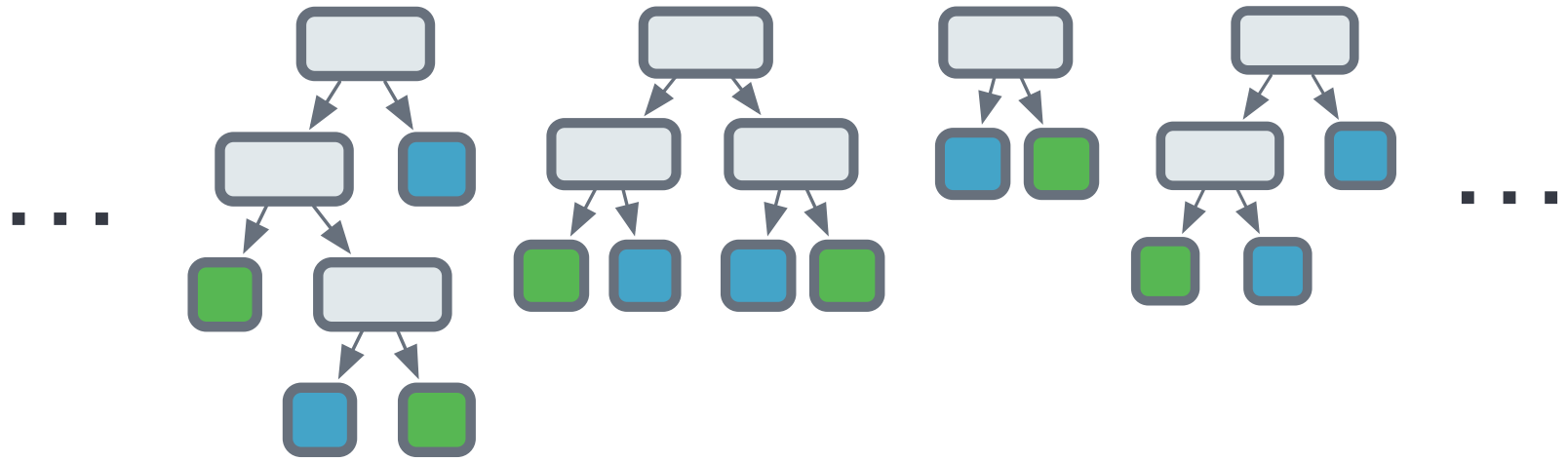
Simple Model

Feature Selection

Remove Structure

CounterFlow ai

Ablation



- *Ablation* - Removing critical parts of a model to test performance
- *Overfitting* – Model inefficiency resulting in unnecessary complexity in the model

Model

Ablation

Default Baseline

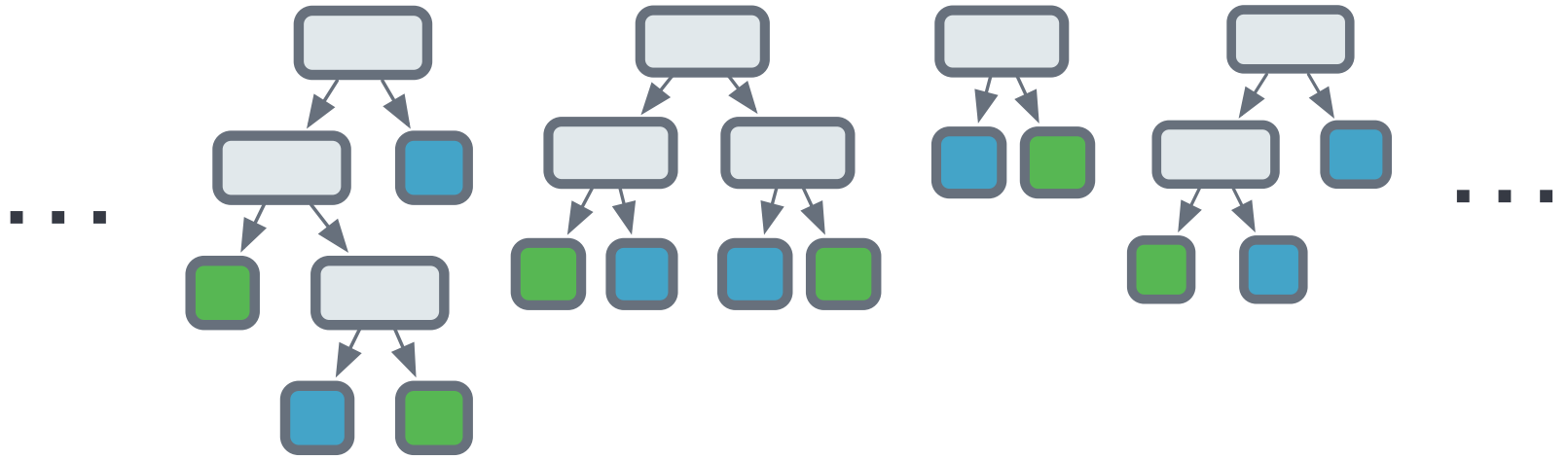
Simple Model

Feature Selection

Remove Structure

CounterFlow ai

Default Baseline (No Model)



Model

Ablation

Default Baseline

Simple Model

Feature Selection

Remove Structure

CounterFlow ai

Default Baseline (No Model)

Model

Ablation

Default Baseline

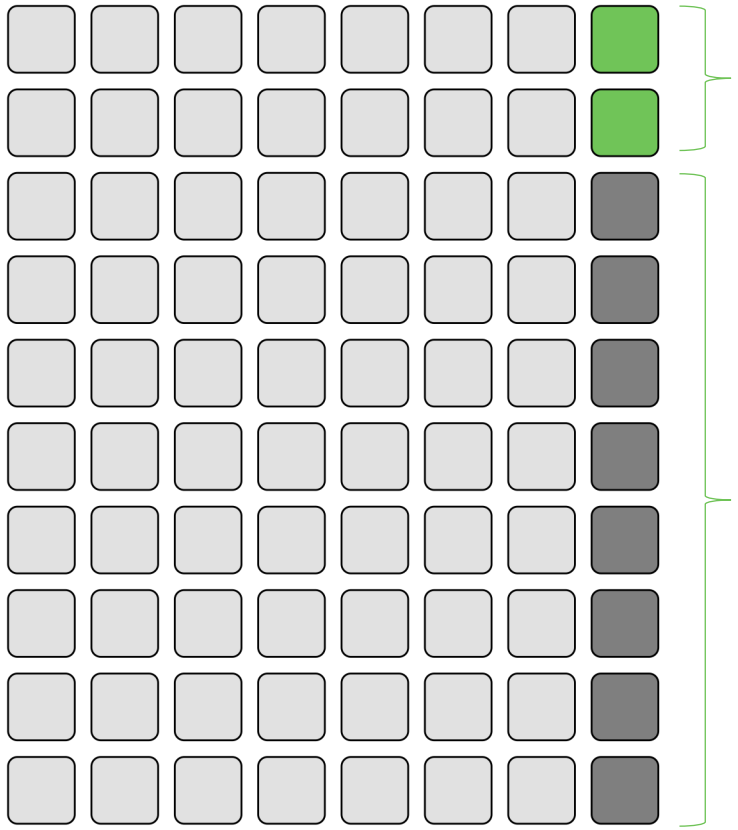
Simple Model

Feature Selection

Remove Structure

CounterFlow ai

Default Baseline (No Model)



- Look at the proportion of positives and negatives in the data
- Your model should beat predicting the majority class
- For timeseries, model should beat predicting the previous value

Model

Ablation

Default Baseline

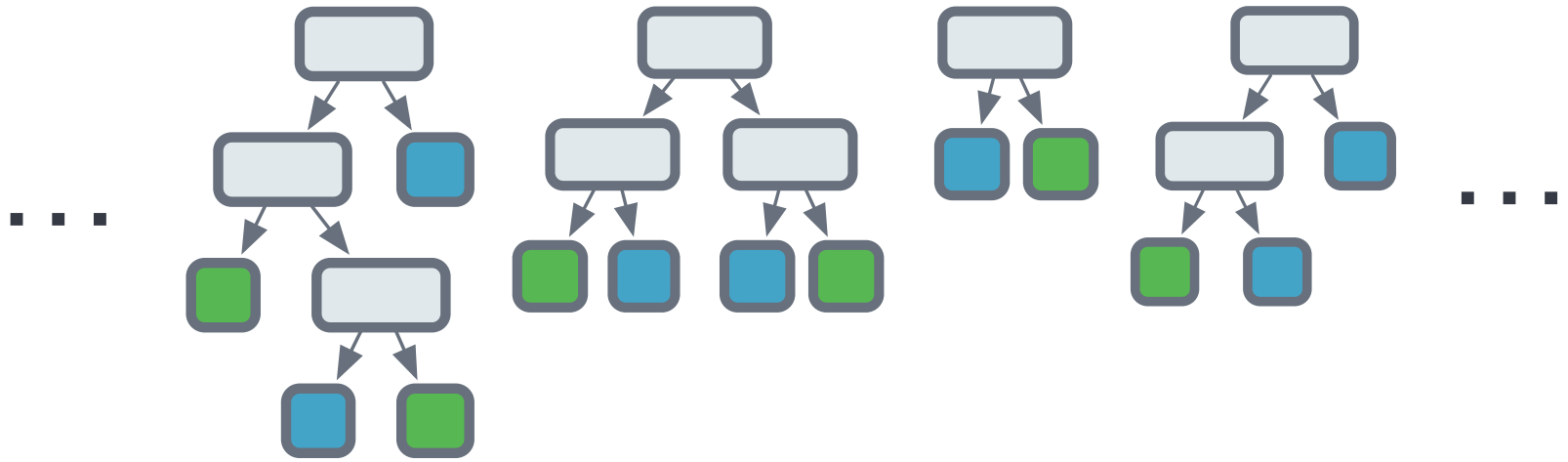
Simple Model

Feature Selection

Remove Structure

CounterFlow ai

Use a Simple Model



- Replace the complex model with a single model
- Decision Tree or Logistic Regression

Model

Ablation

Default Baseline

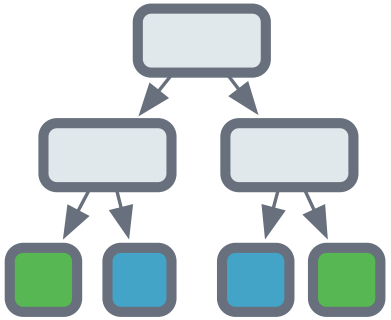
Simple Model

Feature Selection

Remove Structure

CounterFlow ai

Use a Simple Model



- Replace the complex model with a single model
- Decision Tree or Logistic Regression

Model

Ablation

Default Baseline

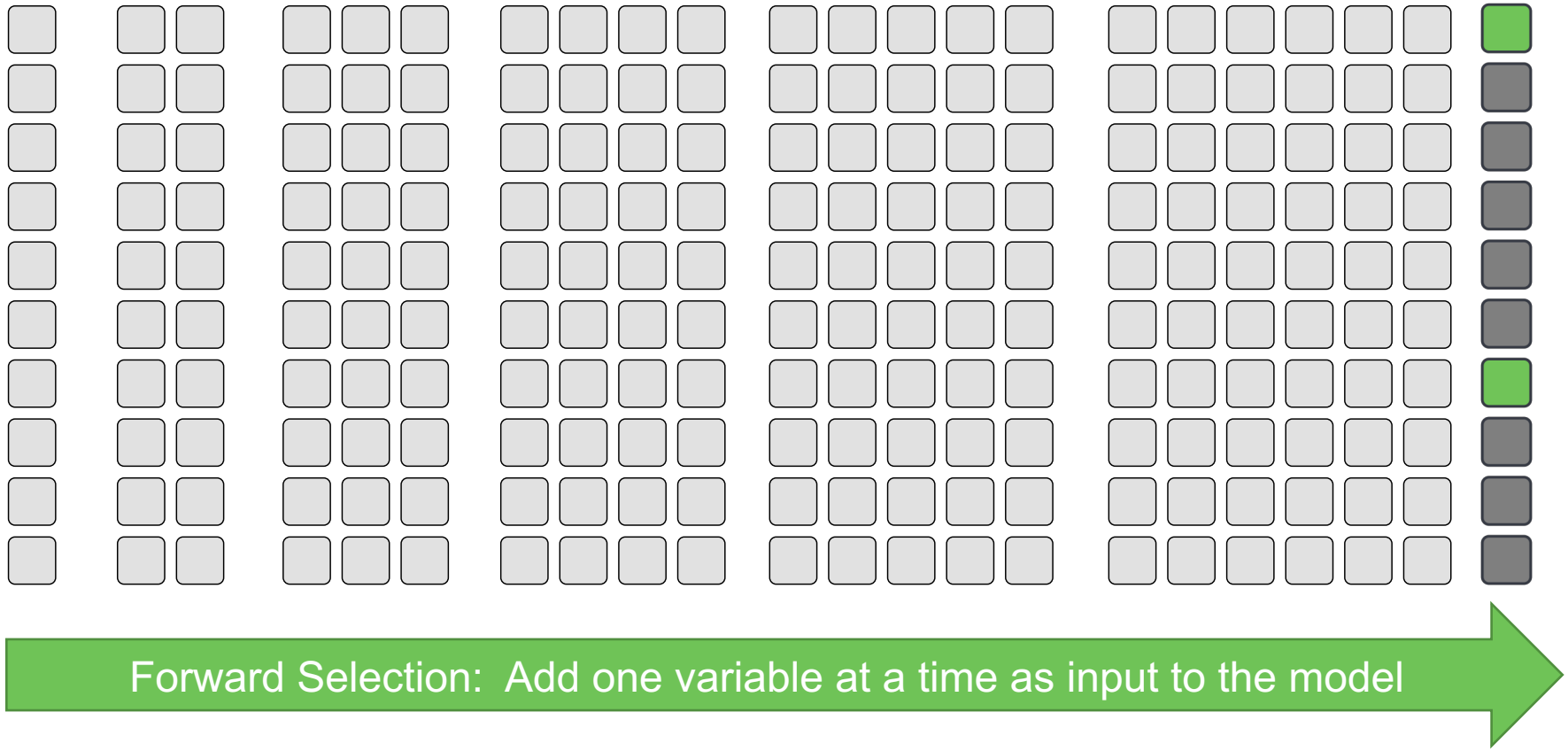
Simple Model

Feature Selection

Remove Structure

CounterFlow ai

Feature Selection



Model

Ablation

Default Baseline

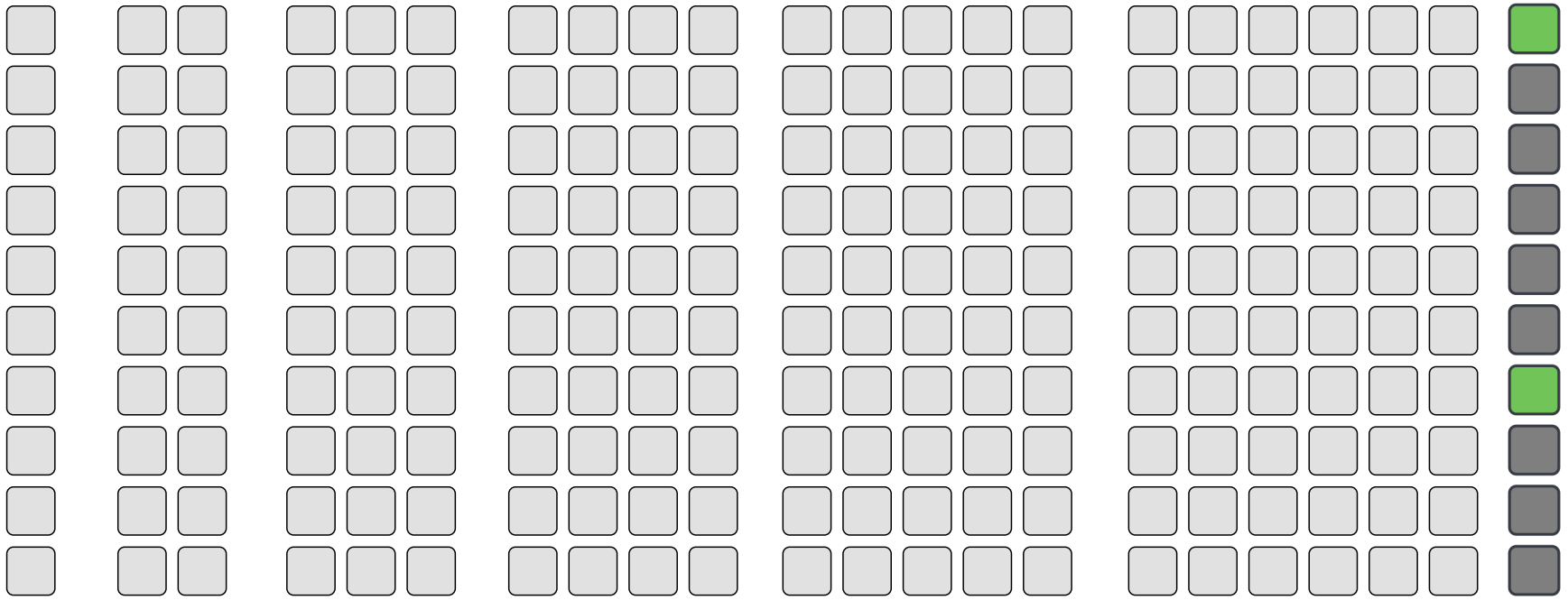
Simple Model

Feature Selection

Remove Structure

CounterFlow ai

Feature Selection



Forward Selection: Add one variable at a time as input to the model

Backwards Selection: Remove one variable at a time as input to the model

Model

Ablation

Default Baseline

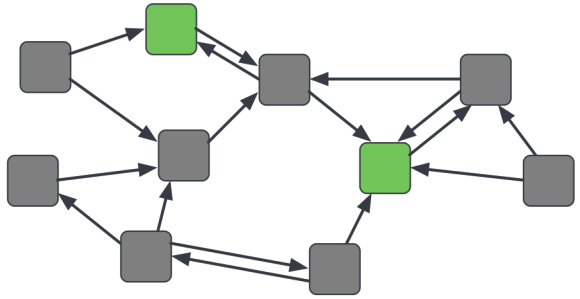
Simple Model

Feature Selection

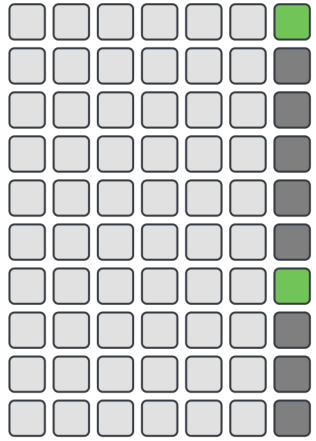
Remove Structure

CounterFlow ai

Remove Structure



- Most cyber models include either a time or graph component
- Disregard that complexity and use a table instead



Triangulating using Metrics

Metrics

Key Idea

Point Metrics

Relative Metrics

Across Thresholds

Visualization

CounterFlow ai

Falsifiability

- Falsifiable statements are able to be proven wrong
- Avoid fooling yourself, use quantitative evaluations
- Clustering is not falsifiable on its own

Metrics

Falsifiability

Point Metrics

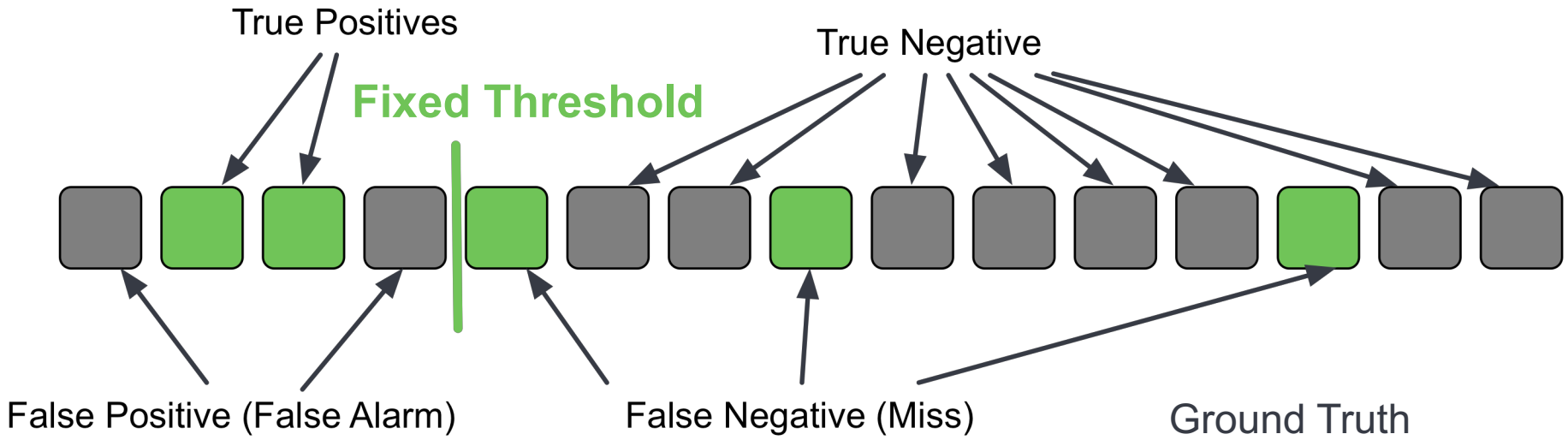
Relative Metrics

Across Thresholds

Visualization

CounterFlow ai

Accuracy, Precision, and Recall



- Point metrics are in relation to a fixed threshold

Model Prediction	Ground Truth: Bad	Ground Truth: Good
Bad	True Positive	False Positive
Good	False Negative	True Negative

Metrics

Falsifiability

Point Metrics

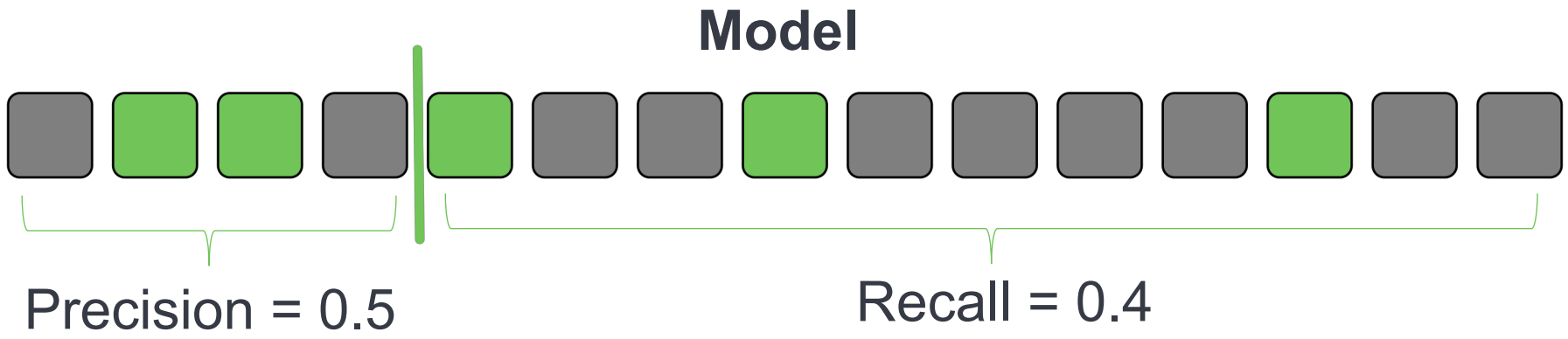
Relative Metrics

Across Thresholds

Visualization

CounterFlow ai

Example 1: F-Metric



$$F_1 = 2 * \frac{precision + recall}{(1 * precision) * recall} = 9$$

- F-Metric combines precision and recall using the harmonic mean

Metrics

Falsifiability

Point Metrics

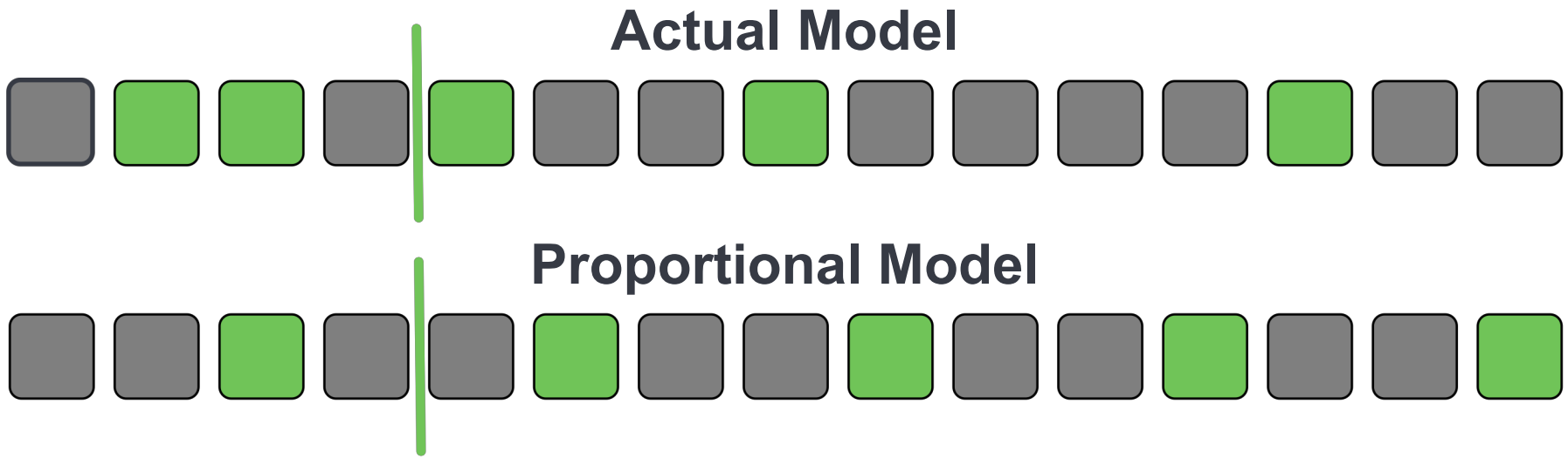
Relative Metrics

Across Thresholds

Visualization

CounterFlow ai

Example 2: Lift



$$Lift = \frac{\#\ Found\ By\ Model}{\#\ Expected\ by\ Chance} = 2x$$

Metrics

Falsifiability

Point Metrics

Relative Metrics

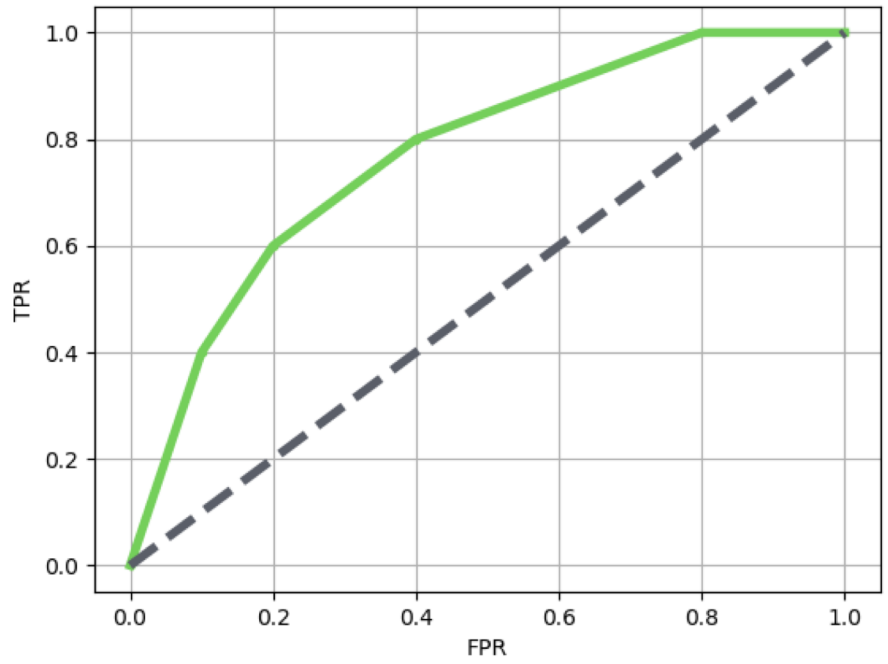
Across Thresholds

Visualization

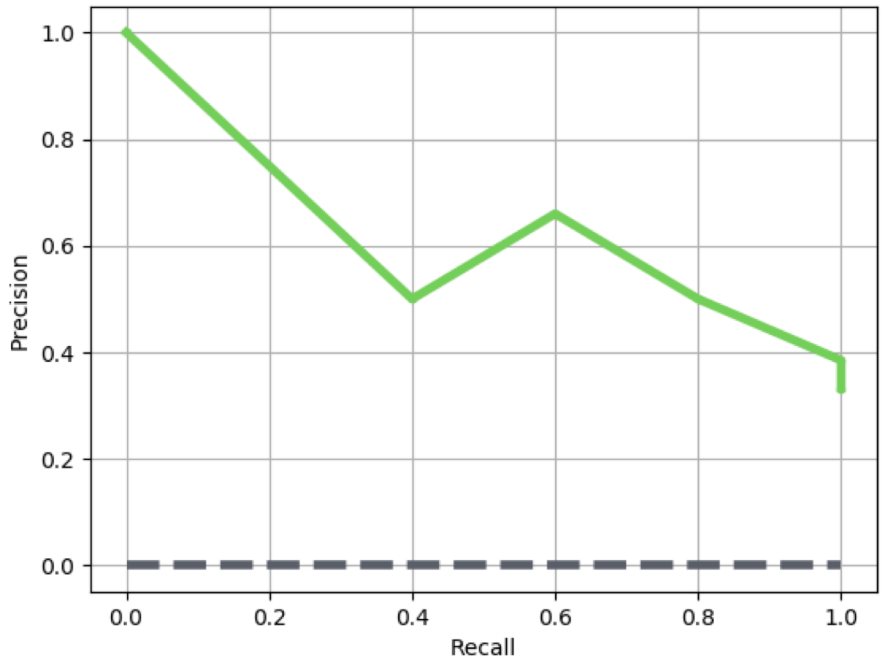
CounterFlow ai

Area Under the Curve (AUC)

ROC Curve



Precision/Recall



- Considers scores across *all* possible thresholds

Metrics

Falsifiability

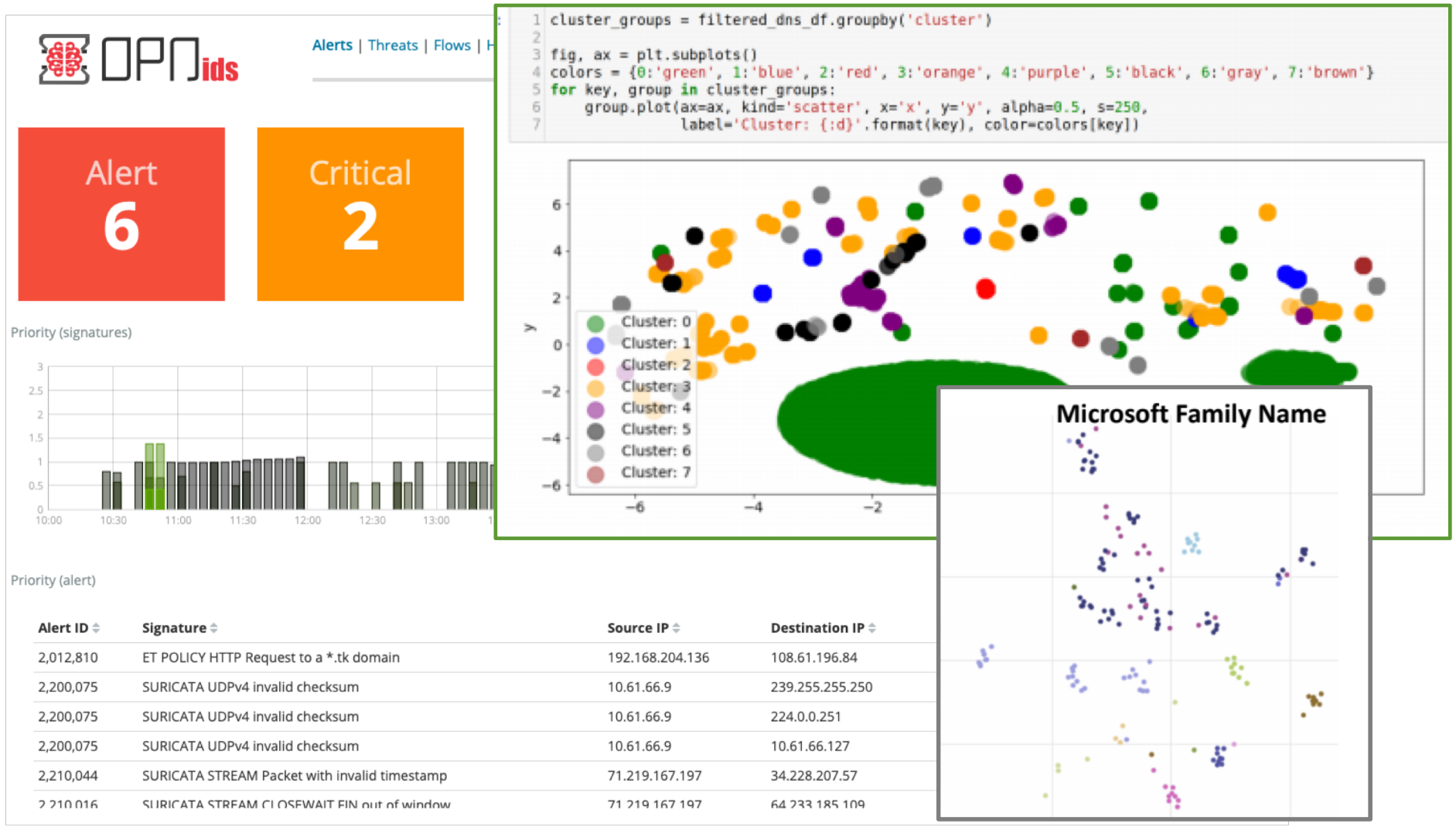
Point Metrics

Relative Metrics

Across Thresholds

Visualization

Visualization and EDA



CounterFlow ai

Triangulating with Data

CounterFlow ai

Data

Key Idea

Train | Test | Eval

Cross-Validation

Multiple Data Sets

Randomization

CounterFlow ai

There is No Free Lunch

- NFL Theorem - No algorithm performs best on every data set
- Testing on multiple data sets ensures success was not due to chance alone
- Evidence that performance will continue on unseen data

Data

No Free Lunch

Train | Test | Eval

Cross-Validation

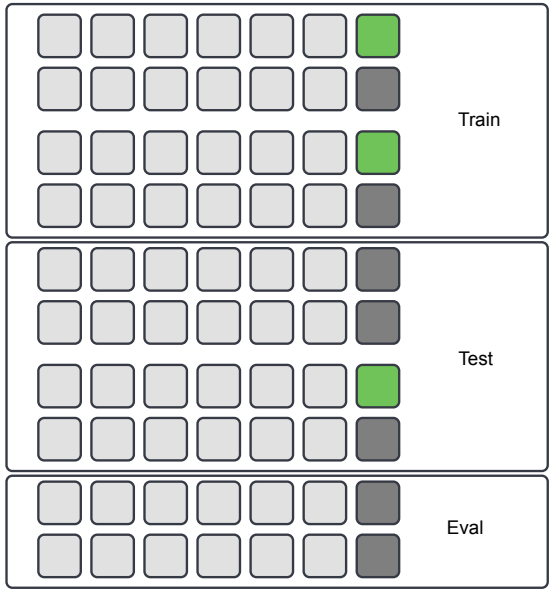
Multiple Data Sets

Randomization

CounterFlow ai

Data Triangulation

Train | Test | Eval



Data

No Free Lunch

Train | Test | Eval

Cross-Validation

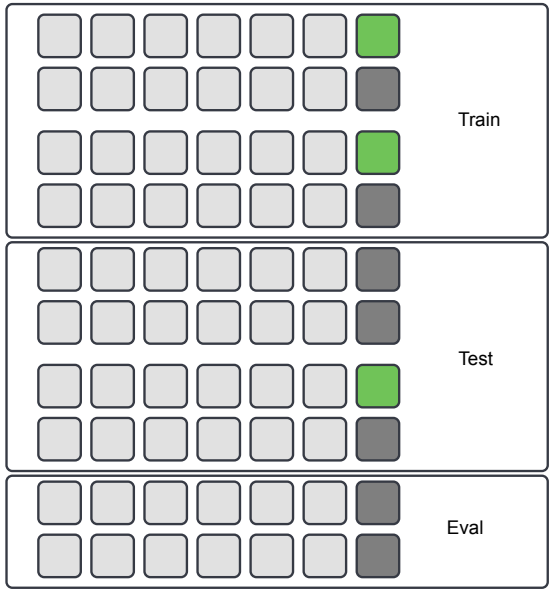
Multiple Data Sets

Randomization

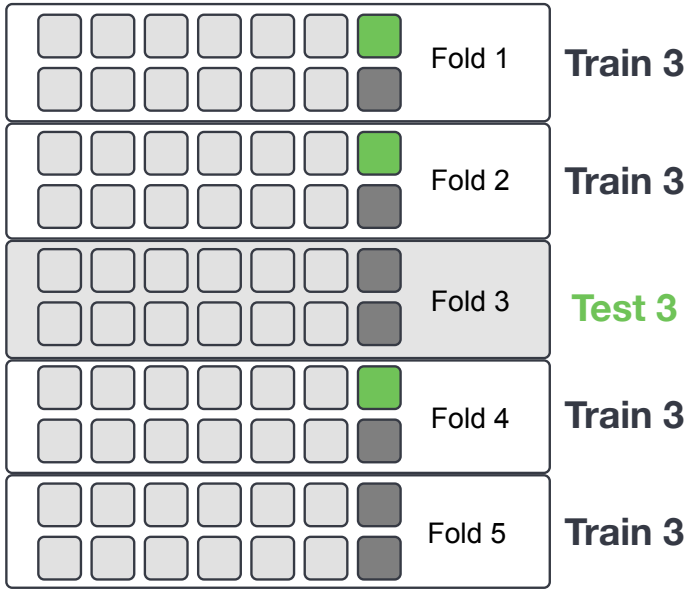
CounterFlow ai

Data Triangulation

Train | Test | Eval



Cross-Validation



Data

No Free Lunch

Train | Test | Eval

Cross-Validation

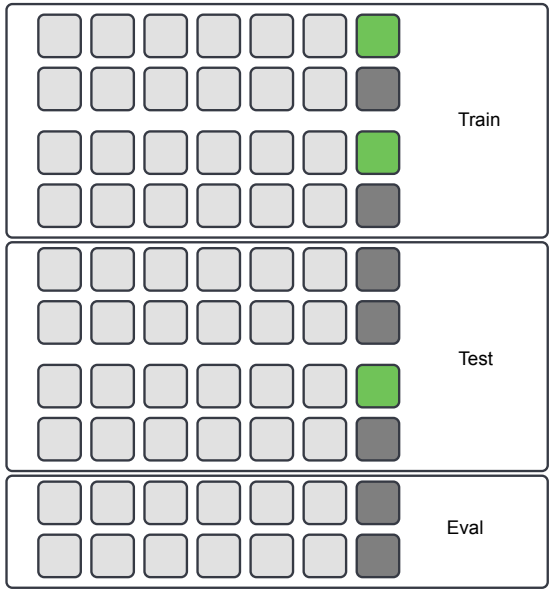
Multiple Data Sets

Randomization

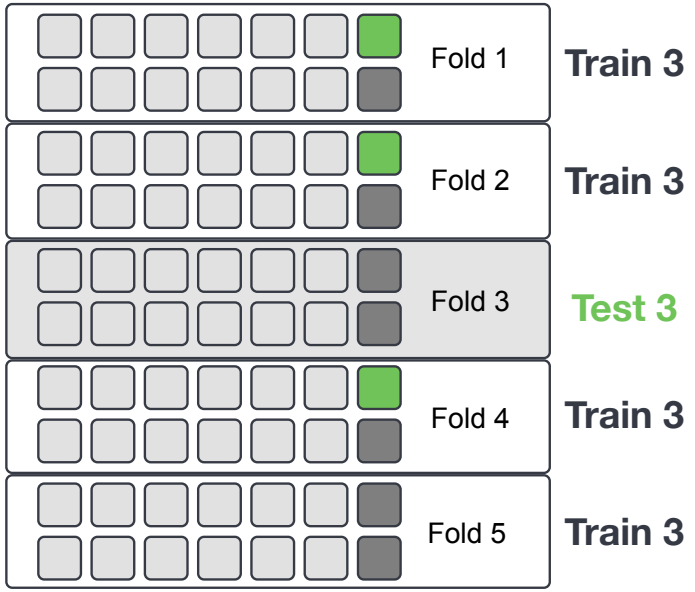
CounterFlow ai

Data Triangulation

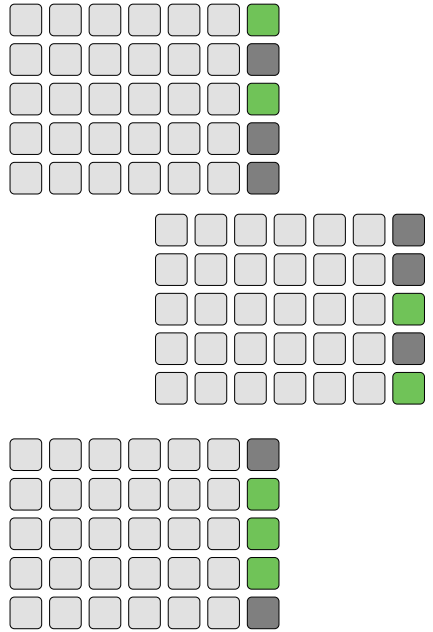
Train | Test | Eval



Cross-Validation



Multiple Data Sets



Data

No Free Lunch

Train | Test | Eval

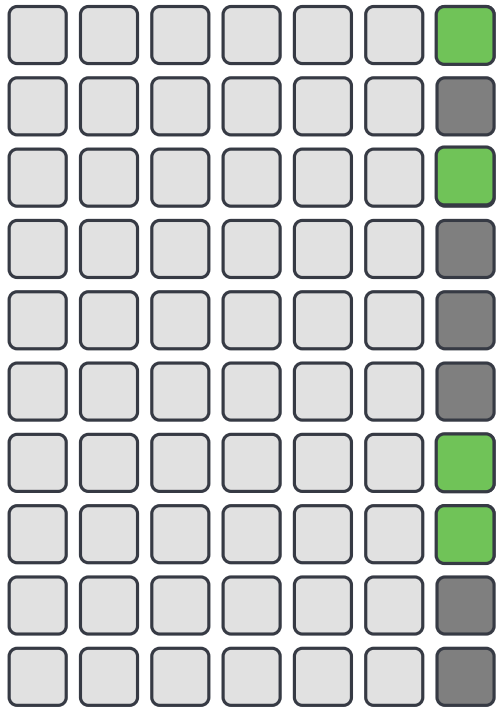
Cross-Validation

Multiple Data Sets

Randomization

CounterFlow ai

Randomization and Permutation



- Use permutations to break correlations in the data
- Repeat many times for non-parametric hypothesis testing

Data

No Free Lunch

Train | Test | Eval

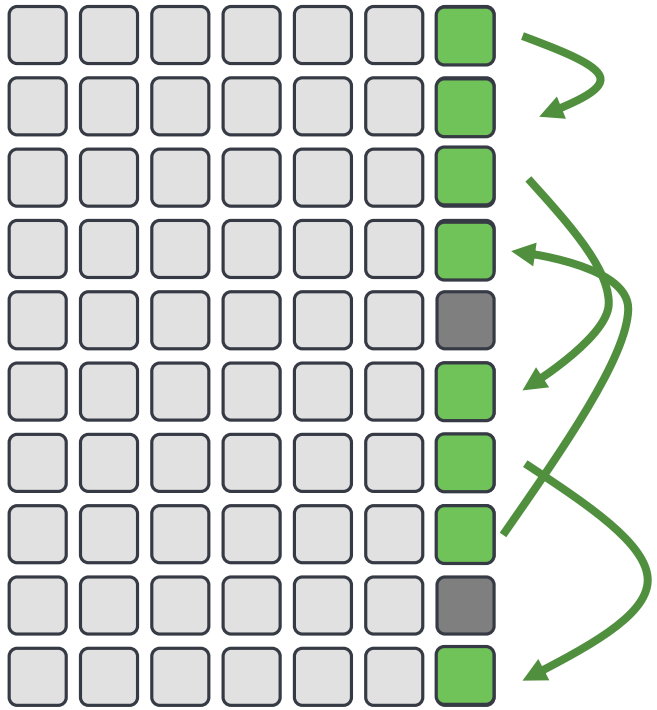
Cross-Validation

Multiple Data Sets

Randomization

CounterFlow ai

Randomization and Permutation



- Use permutations to break correlations in the data
- Repeat many times for non-parametric hypothesis testing

Data

No Free Lunch

Train | Test | Eval

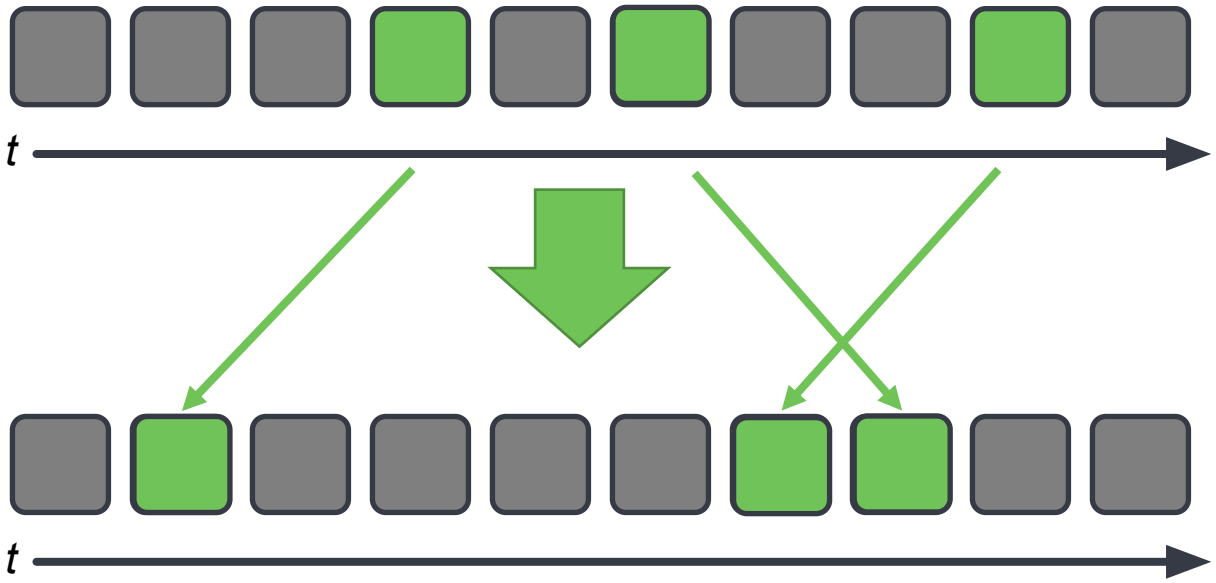
Cross-Validation

Multiple Data Sets

Randomization

CounterFlow ai

Randomization and Permutation



- Shuffle timestamps to break temporal correlations

Triangulating your Focus

Focus

Right Question

Streetlight Effect

Type III Errors

Counterfactuals

CounterFlow ai

The Streetlight Effect

Focus

Right Question

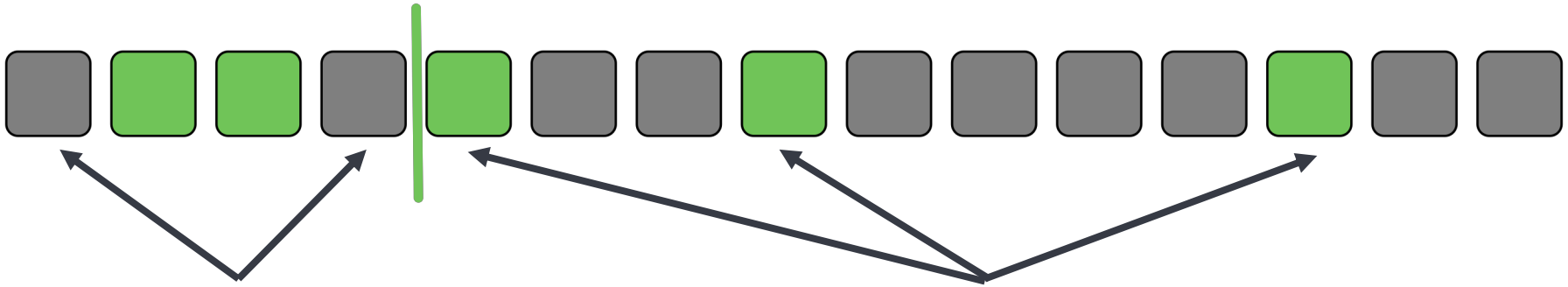
Streetlight Effect

Type III Errors

Counterfactuals

CounterFlow ai

Type III Errors



Type 1 Error: False Positives

Type 2 Error: False Negatives

Focus

Right Question

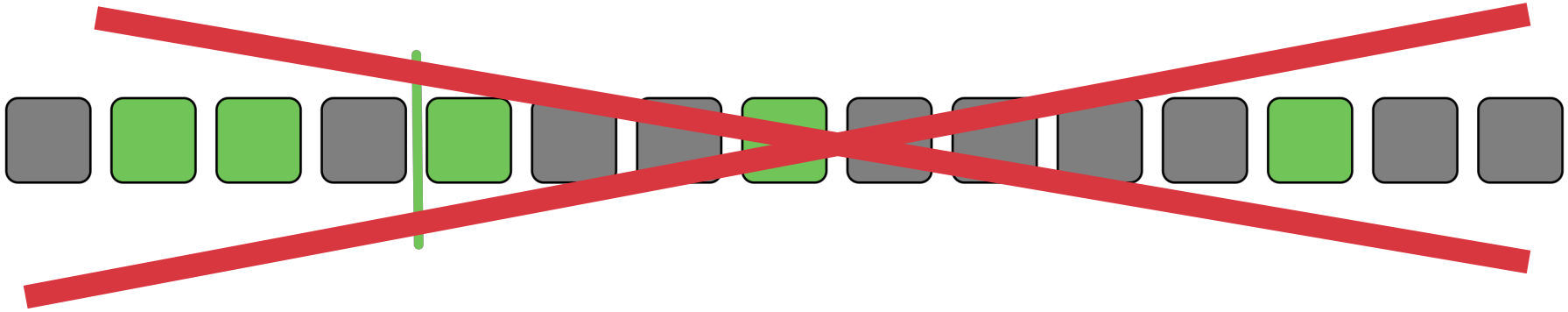
Streetlight Effect

Type III Errors

Counterfactuals

CounterFlow ai

Type III Errors



Type 3: Right Answer, Wrong Question

Focus

Right Question

Streetlight Effect

Type III Errors

Counterfactuals

Counterfactuals

Two roads diverged in a yellow wood,

And sorry I could not travel both

And be one traveler, long I stood

And looked down one as far as I could

To where it bent in the undergrowth…

- *The Road Not Taken,* Robert Frost (1916)

Triangulation Framework for ML

	Model	Metric	Data	Focus
Key Idea	*Ablation*	*Falsifiability*	*No Free Lunch*	*Right Question*
Landmarks	Default Baseline	Point Metrics	Train/Test/Eval	Streetlight Effect
	Simple Model	Relative Metrics	Cross-validation	Type III Errors
	Feature Selection	Across Thresholds	Multiple Data Sets	Counterfactuals
	Remove Structure	Visualization	Randomization	

CounterFlow ai

Questions?

CounterFlow ai

af@counterflowai.com
@counterflowai
https://counterflow.ai

Andrew Fast. Ph.D.
Chief Data Scientist

CounterFlow ai

Andrew Fast is the Chief Data Scientist and co-founder of CounterFlow AI. CounterFlow AI is building the next-generation security analytics platform enabling overwhelmed SOC teams to take a Data Science and AI approach to threat hunting. By transforming raw network traffic data into actionable insights in a streaming fashion, our products significantly reduce time to detection and response.

Previously, Dr. Fast served as the Chief Scientist at Elder Research, Inc., a leading data science consulting firm, where he helped hundreds of companies expand their data science capabilities. He is a frequent author, teacher, and invited speaker on data science topics. In 2012, he co-authored the book Practical Text Mining that was published by Elsevier and won the PROSE Award for top book in the field of Computing and Information Sciences for that year. His work on analyzing NFL coaching trees was featured on ESPN.com in 2009.

Dr. Fast earned PhD and MS degrees in Computer Science from the University of Massachusetts Amherst and a BS in Computer Science from Bethel University.

af@counterflowai.com
@counterflowai
https://counterflow.ai

References

- *Math With Bad Drawing*, Ben Orlin, 2018, Black Dog & Leventhal Press
- *Datasaurus* - http://www.thefunctionalart.com/2016/08/downloaddatasaurus-never-trust-summary.html
- *Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing*, Justin Matejka and George Fitzmaurice, CHI '17
- Anscombe, F.J. (1973). Graphs in Statistical Analysis. The American Statistician 27, 1, 17--21.
- Ablation links:
  - https://www.quora.com/In-the-context-of-deep-learning-what-is-an-ablation-study
  - https://twitter.com/fchollet/status/1012721582148550662?lang=en
  - https://nlpers.blogspot.com/2016/08/feature-or-architecture-ablation.html
- *No Free Lunch* - https://ti.arc.nasa.gov/m/profile/dhw/papers/78.pdf
- *Randomization and Permutation* - https://www.elderresearch.com/company/target-shuffling
- *Mutt & Jeff* - 1942 June 3, Florence Morning News, Mutt and Jeff Comic Strip, Page 7, Florence, South Carolina. (NewspaperArchive)
- *Type III error* - Allyn W. Kimball (1957) - https://en.wikipedia.org/wiki/Type_III_error
- *Counterfactuals* - Book of Why,Judea Pearl and Dana Mackenzie, 2018, Basic Books

CounterFlow ai