USAID
FROM THE AMERICAN PEOPLE

Machine Learning

USAID MENA Advanced MEL Workshop

Session Objectives

By the end of this session, you should be able to..

- Understand basic concepts of machine learning
- Recognize the distinction between supervised and unsupervised learning
- Recognize the most common machine learning algorithms and how they are applied in development settings

Level Set

How does USAID do machine learning?

- Managing Machine Learning Projects in International Development
- Making AI Work for International Development

What is Machine Learning?

- Machine learning is a collection of algorithms that learn from data
- Machine learning attempts to learn the data so well that it can tell you what to expect with new data you haven't seen yet
- Machine learning algorithms can be broadly classified into unsupervised or supervised

How Machine Learning is Different

What are the defining characteristics of machine learning?

- Prediction vs. causation
- Bias vs. variance
- Over- vs. under-fitting

Prediction vs. Causation

- Descriptive or exploratory analysis helps us understand our data
- Causal inference helps us understand the relationships between variables of interest
- Machine learning tries to learn our data so well that it can predict future data!

Bias vs. Variance

- The causal inference tradition seeks to eliminate bias first, then looks to improve efficiency by reducing variance
- Machine learning doesn't care about bias
- In fact, sometimes machine learning can actually use bias to reduce uncertainty

Underfitting vs. Overfitting

- Standard analysis will take the data as given, and seek to understand it
- More advanced analysis may try to *explain* the given data
- Machine learning algorithms are evaluated according to how well they work on data they haven't seen yet
  - Run algorithms on training data, evaluate performance on test data

Unsupervised and Supervised Learning

Unsupervised Learning

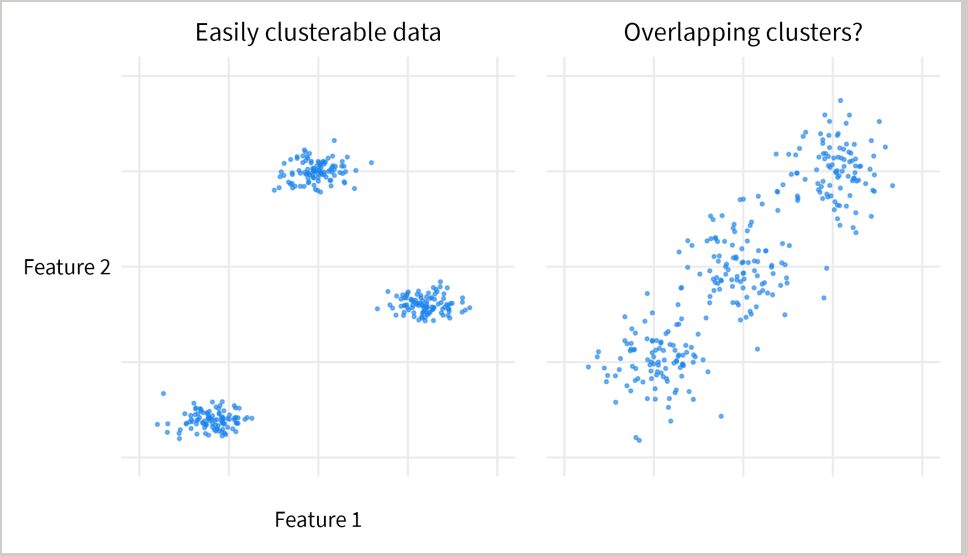
- We have a set of variables that we suspect or have reason to believe are related to each other in some way
- Is there any hidden structure in the data that standard analysis would not find?
- The task is to find patterns that we can then turn into new variables that capture the patterns

Algorithms for Unsupervised Learning

- Clustering based on similarity
  - K-means, Latent Class Analysis, Density-Based Spatial Clustering
- Dimension reduction
  - Principal Component Analysis, Factor Analysis, Item Response Theory, Expectation Maximization

Clusters in Data

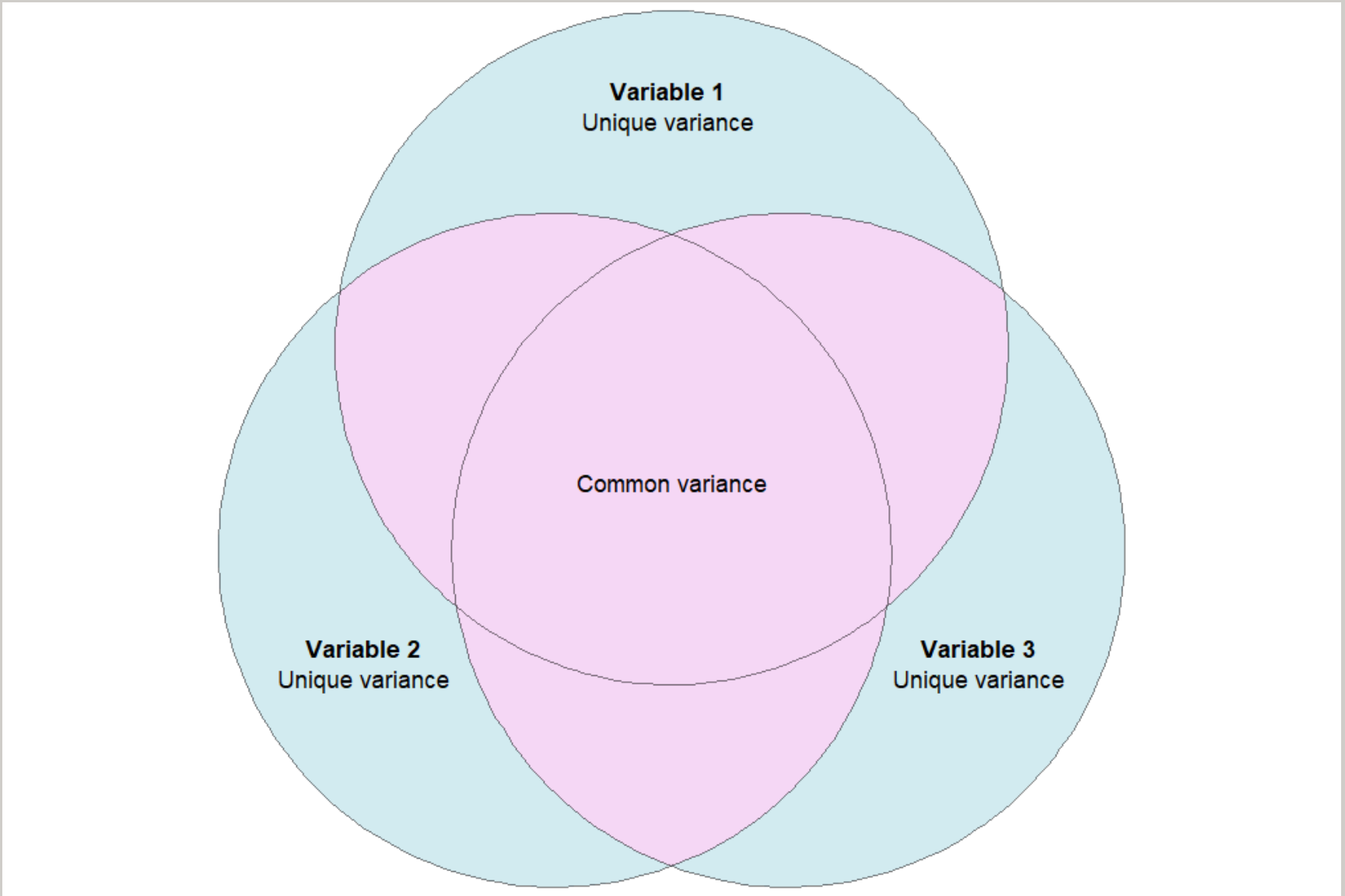
- Clustering algorithms minimize variance between data points, or use a distance metric
- Clustering is difficult when there is overlap between clusters



Correlations in Data

- Principal components analysis (PCA) attempts to capture as much of the total variance as possible in a single or few variables
- Factor analysis (FA) attempts to capture common variance among a set of variables

Total and Common Variance

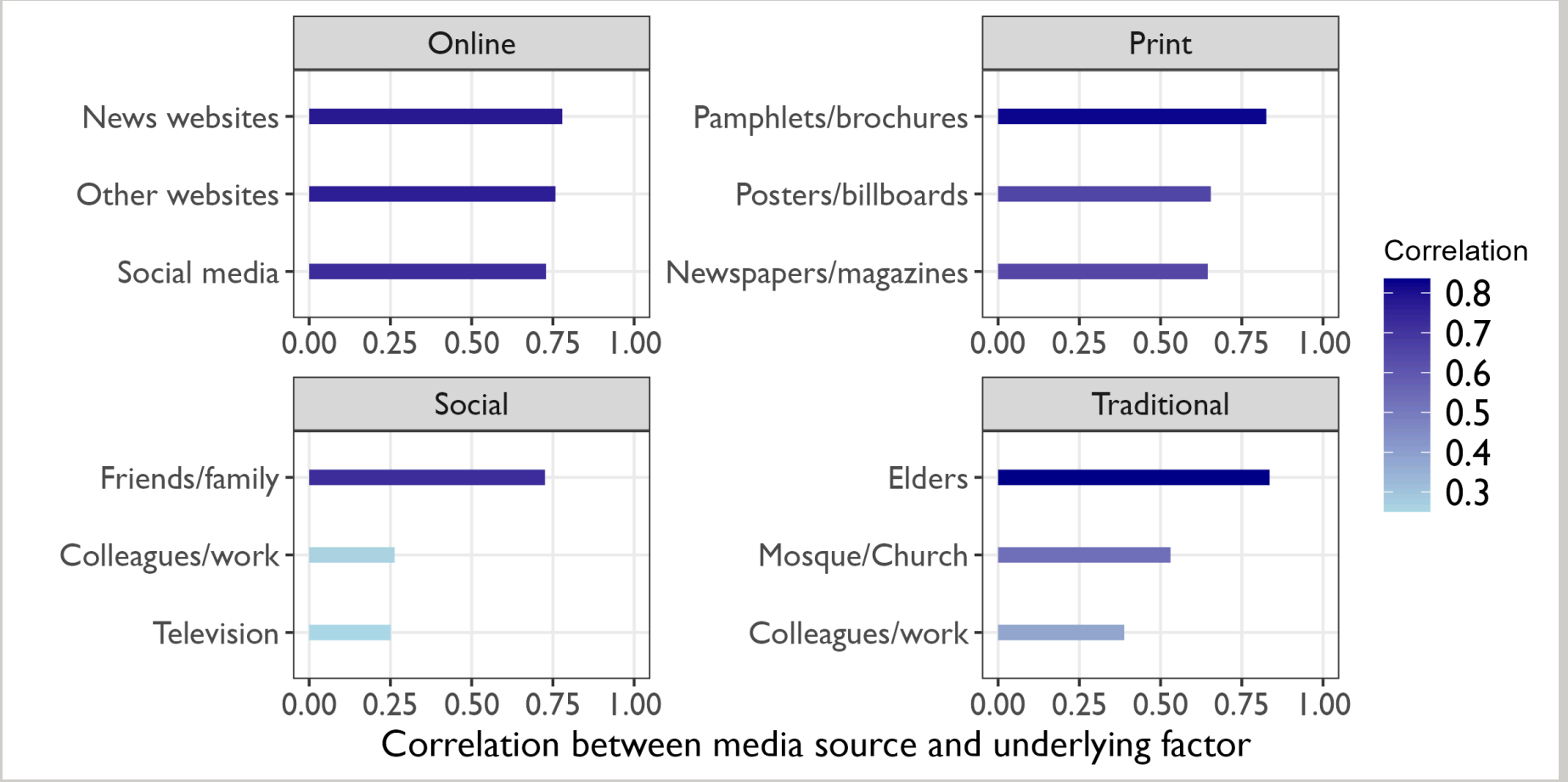


Total Variance of Household Assets

Household item	Mean of household ownership	Standard deviation	Factor Weight
Table	3.496	3.359	0.097
Fan	1.495	1.502	0.095
Mobile	1.945	1.265	0.091
Bed	2.623	1.245	0.091
Cabinet	1.656	1.543	0.090
SofaSet	0.175	0.436	0.086
Refrigerator	0.155	0.387	0.082
Dressing Table	0.191	0.459	0.081
Charger Light	0.553	0.823	0.080
Trunk	0.673	1.222	0.077

Common Variance of Media Profiles

What media or other sources do you turn to for news and information?

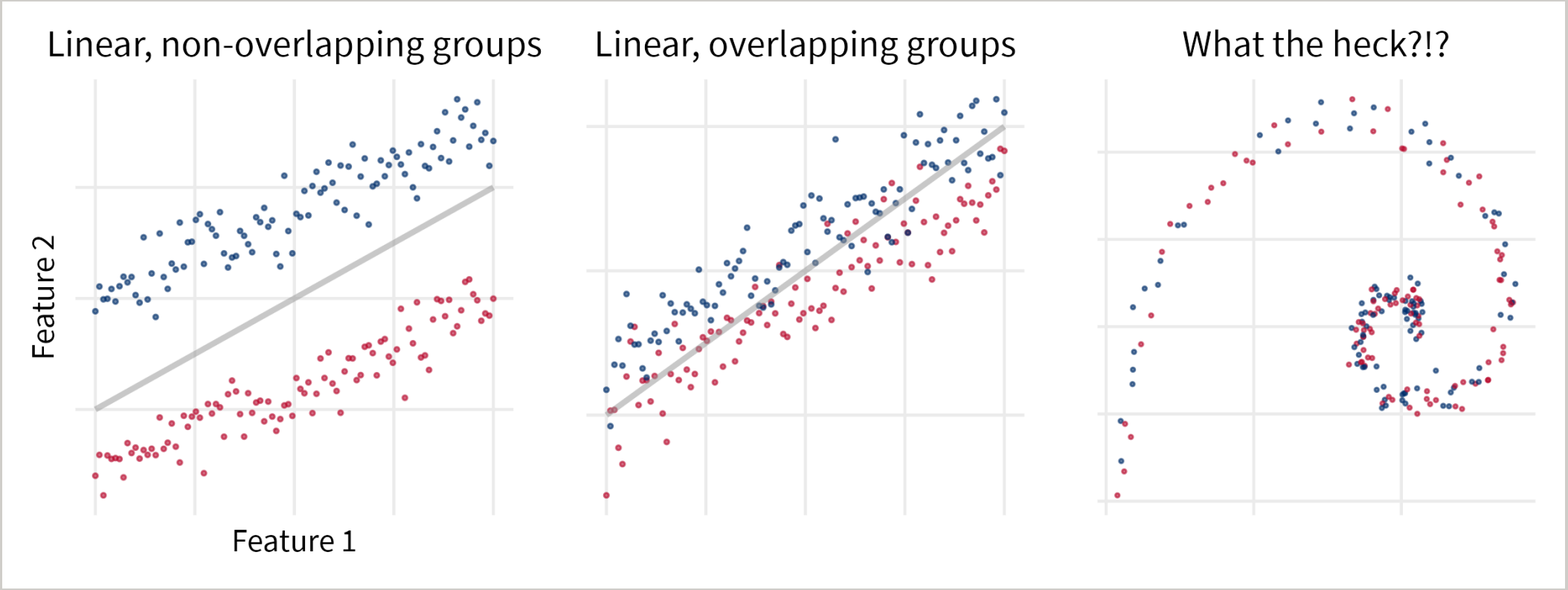


Supervised Learning

- Now we have data that we know is related to some target variable
- How well can we use the data to hit the target?
- The task is now prediction/classification

Using Features to Predict a Target

Previously we just had features. Now we have features that predict a target.



Interested in that spiral thingy? See https://www.mathnasium.com/blog/golden-ratio-in-nature

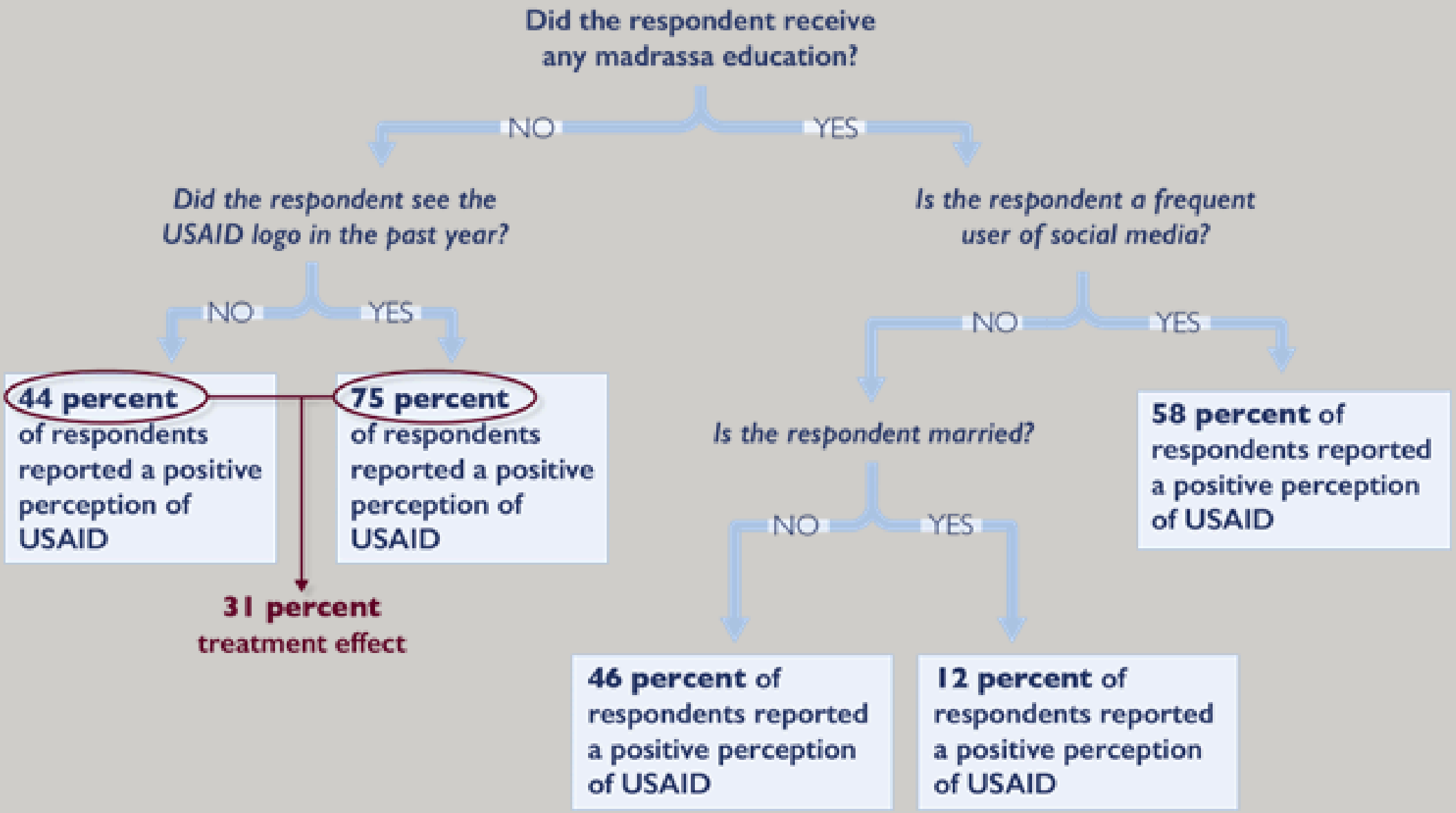
Algorithms for Supervised Learning

- Conditional Inference Trees/Random Forests: Break data into branches for classification and regression tasks
- Gradient Descent: Minimizes error by iterating through partial derivatives
- Neural Networks: Identifies ‘hidden layers’ in complex data to conduct deep learning tasks

Supervised learning - Conditional Inference Tree

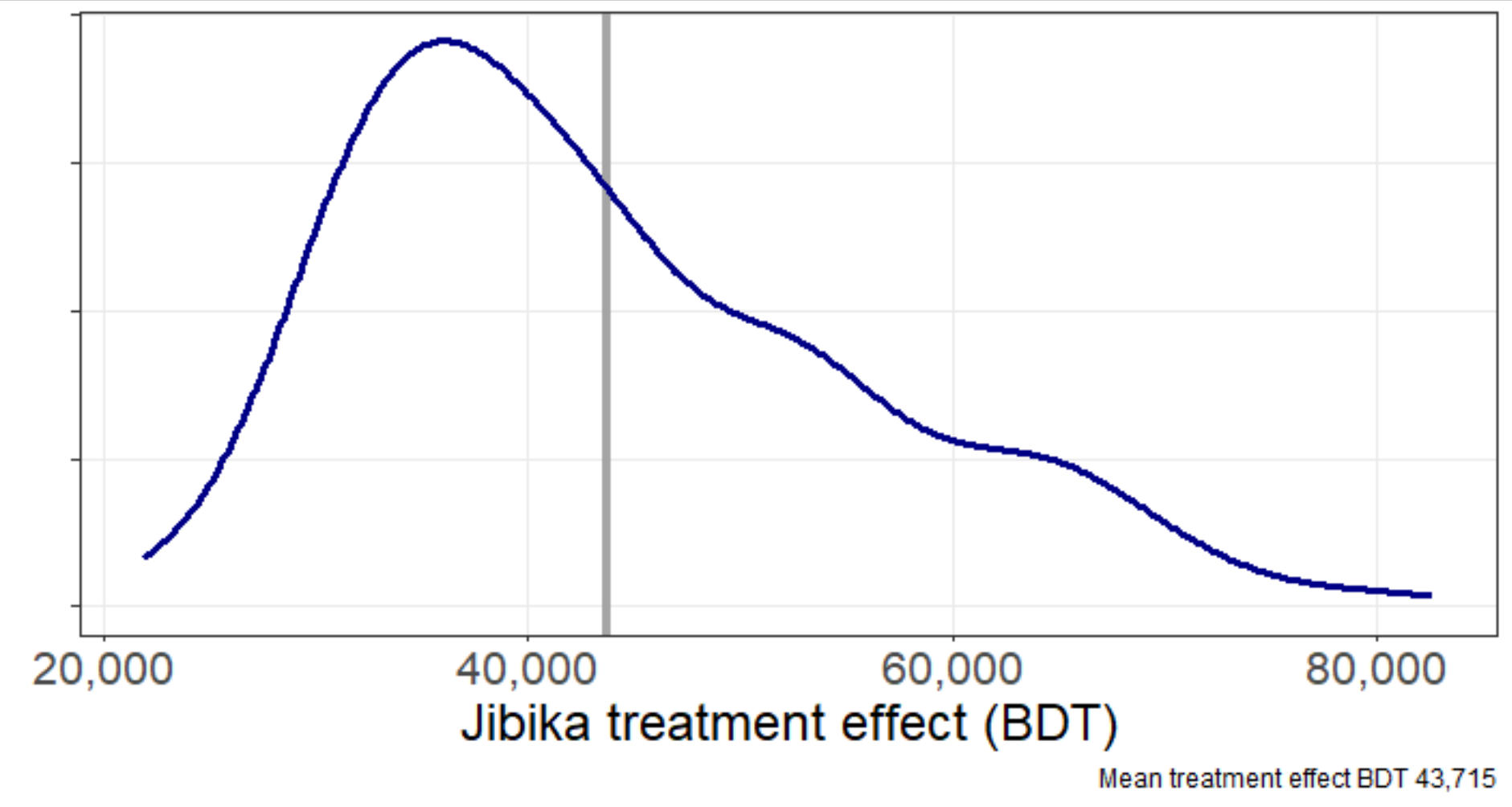
Conditional inference trees crawl through your data, split everything into different pathways (branches), and retain the branches that capture the most variation

- Advantage: identifies interesting sub-groups to explore further
- Disadvantage: Overfits your data!


Did the respondent receive any madrassa education?
NO
YES
Did the respondent see the USAID logo in the past year?
NO
YES
44 percent of respondents reported a positive perception of USAID
75 percent of respondents reported a positive perception of USAID
31 percent treatment effect
Is the respondent a frequent user of social media?
NO
YES
58 percent of respondents reported a positive perception of USAID
Is the respondent married?
NO
YES
46 percent of respondents reported a positive perception of USAID
12 percent of respondents reported a positive perception of USAID

Supervised learning - Random Forest

Random forests makes predictions using a large number of trees

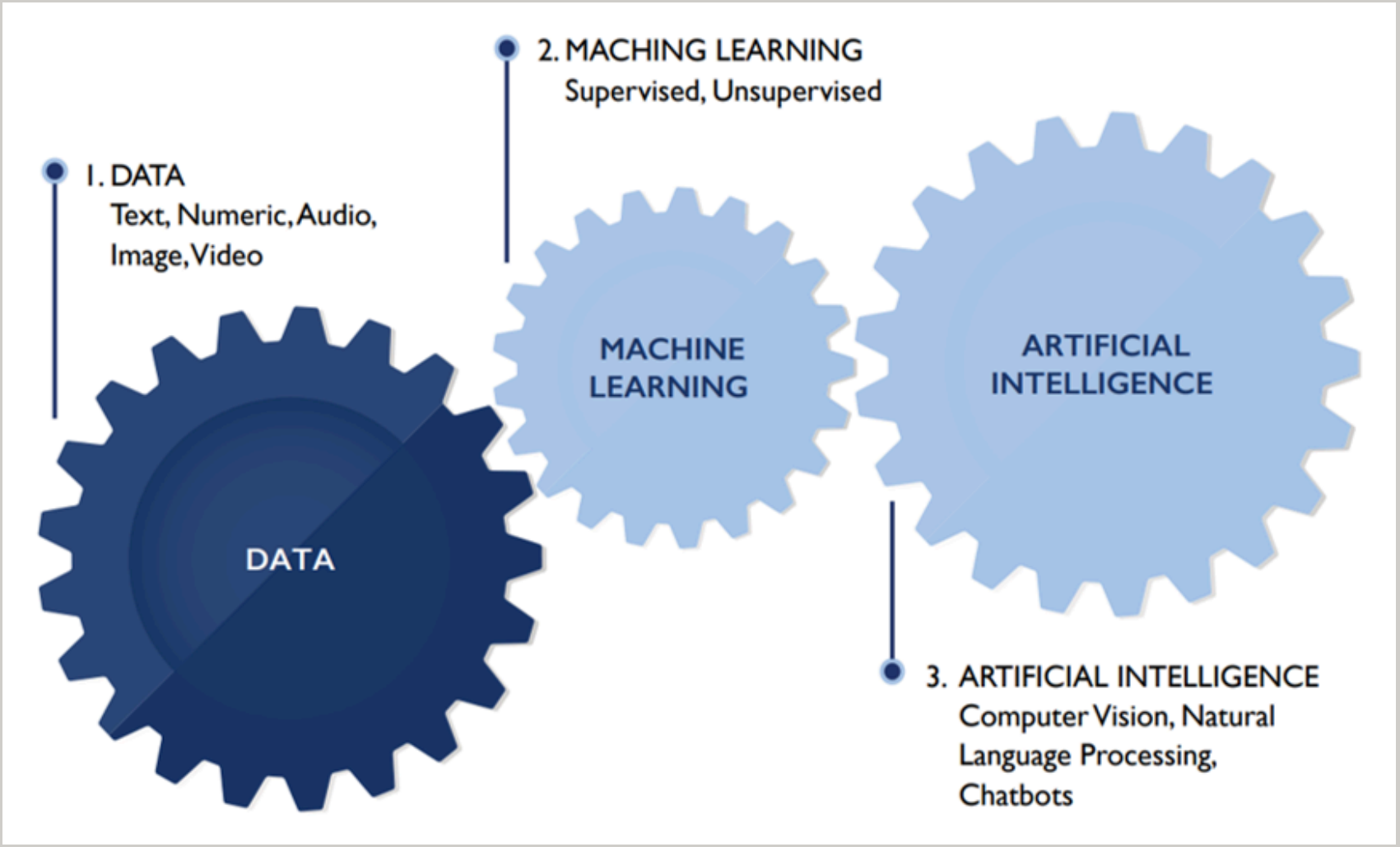


Heterogeneous treatment effects on household income

From Machine Learning to AGI

- We've mentioned how machine learning is about the use of algorithms to learn from data
- What happens if the algorithms learn from data, and then use that learning to improve themselves and run again?
- Artificial General Intelligence (AGI) refers to algorithms that seem to us like thinking and reasoning human intelligences

Learn from Your Data, Then Learn from What You Learn



Welcoming Our New Machine Learning Overlords

- Currently, the closest we have to AGI are the Large Language Models that power ChatGPT and other chatbots
- "In the next couple of decades, we will be able to do things that would have seemed like magic to our grandparents." [Sam Altman, 9-23-24]
- Stay tuned for next session on Large Language Models!

Thank you!