MULTI INVESTMENT ATTRIBUTION

CAUSAL EFFECTS OF MULTI-LEVEL TREATMENT OF INTERVENTIONS USING OBSERVATIONAL DATA

PROJECT ID: IN04
MENTOR'S NAME: Narayan Sarkar (IBM)

PROJECT TEAM MEMBERS:

Arkaprabha Pal
Ranjitha T M
Sandhiya S
Subhadeep Rakshit

TABLE OF CONTENTS

PROBLEM STATEMENT:

In a competitive business environment, companies must allocate resources across multiple customer interventions (such as tech support and discount programs) to maximize customer engagement and revenue. However, the effectiveness of these interventions can vary depending on customer characteristics (e.g., company size), and implementing multiple interventions simultaneously may have compounding or diminishing effects on revenue.

This project aims to quantify the causal effect of each intervention (tech support and discount) on customer revenue using observational data, acknowledging the potential influence of confounding factors.

WORK APPROACH:

1) Practiced and presented a presentation to explain causal effects with a structured approach.
2) Demonstrated the application of DoWhy and EconML libraries, showcasing examples to illustrate each library.
3) For DoWhy, presented an example of Customer Loyalty, and for EconML, covered Multi-Investment Attribution.
4) Selected Multi-Investment Attribution as our main project focus based on insights from the examples.
5) Finding the Confounding Variable, Outcomes, and treatment methods for the chosen scenario.
6) Defined the treatment effect function, including true coefficients and intercepts, tailored to each treatment method.
7) Selected and trained an EconML model using generic helper models for efficient computation.
8) Calculated the treatment effect for each method, ensuring accurate evaluation.
9) Choosing the most substantial strategy that yields a higher outcome based on obtained estimates for each treatment method.

WORK PRODUCTS AND DELIVERABLES:

Dataset Used:
https://msalicedatapublic.z5.web.core.windows.net/datasets/ROI/multi_attribution_sample.csv

Libraries Used: EconML, Numpy, Pandas, XGBoost

EconML Documentation: https://econml.azurewebsites.net/

Platform Used: Google Colab

(**Colab Link:** https://colab.research.google.com/drive/11oi2p8ASbR7kkp-0AJQIopUQzJvnjnXK?usp=sharing)

GitHub Repository: https://github.com/Sandhiya-2003/Multi-Investment-Attribution

USER MANUAL:

1) Open Google Colab and select File > Open notebook.
2) Go to the GitHub tab, enter the repository URL or search for the repository by name, and select the desired notebook.
3) Click Open Notebook to load it in Colab, and then click Run All (Runtime > Run all) to execute the code.

TECHNICAL MANUAL:

Datasets: The data contains ~2,000 customers and is comprised of:

- **Customer features:** details about the industry, size, revenue, and technology profile of each customer.
- **Interventions:** information about which incentive was given to a customer.
- **Outcome:** the amount of product the customer bought in the year after the incentives were given.

Feature Name	Type	Details
Global Flag	W	whether the customer has global offices
Major Flag	W	whether the customer is a large consumer in their industry (as opposed to SMC - Small Medium Corporation - or SMB - Small Medium Business)
SMC Flag	W	whether the customer is a Small Medium Corporation (SMC, as opposed to major and SMB)
Commercial Flag	W	whether the customer's business is commercial (as opposed to public secor)
IT Spend	W	$ spent on IT-related purchases
Employee Count	W	number of employees
PC Count	W	number of PCs used by the customer
Size	X	customer's size given by their yearly total revenue
Tech Support	T	whether the customer received tech support (binary)
Discount	T	whether the customer was given a discount (binary)
Revenue	Y	$ Revenue from customer given by the amount of software purchased

Software Libraries Used:

- EconML: Provides causal inference methods tailored for estimating treatment effects from observational data in complex, multi-level intervention scenarios.
- Numpy: Enables efficient numerical operations and array manipulation, essential for handling large datasets in causal effect modeling.
- Pandas: Facilitates data loading, preprocessing, and manipulation, crucial for managing multi-investment data and treatment variables.
- XGBoost: Implements gradient boosting for treatment effect estimation, helping identify impactful investment strategies and causal relationships.

Functions Created/Used:

a) **Treatment Effect Function:** Computes the treatment effect as a function of customer size to evaluate intervention impact levels. This expression produces a vector where:
   - The first component represents a treatment effect calculated as:

$$\mathbf{5000} + \frac{\mathbf{2}}{\mathbf{100}}\boldsymbol{X}$$

   - The second component represents a treatment effect calculated as:

$$\frac{\mathbf{5}}{\mathbf{100}}\boldsymbol{X}$$

➢ The expression for the treatment effect function is as follows:

$$\textbf{TE(X)} = \begin{bmatrix} 5000 + \frac{2}{100}X \\ \frac{5}{100}X \end{bmatrix}$$

where X is the customer size.

b) LinearDRLearner Function:

Trains a causal model using LinearDRLearner(model_regression, model_propensity, random_state) to estimate treatment effects while controlling for confounding variables.

- model_regression = XGBRegressor
- model_propensity = XGBClassifier

c) XGBRegressor Function:

Trains a gradient-boosting model for regression tasks to predict outcomes based on covariates.

d) XGBClassifier Function:

Trains a gradient-boosting model for classification tasks to estimate treatment assignment probabilities across multiple interventions.

Installing EconML guide:

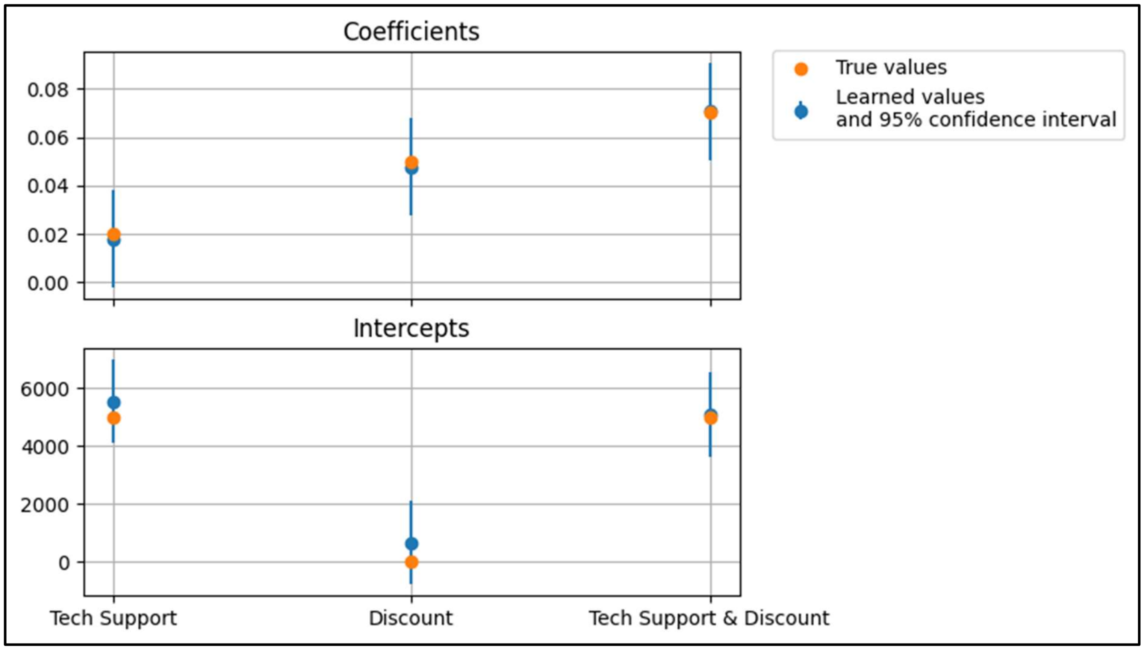
To install the EconML library in Google Colab, run the following command in a code cell:

```
!pip install econml
```

This command will install the latest version of EconML, allowing you to utilize its features for causal inference and treatment effect estimation in your analysis.

RESULTS:

Comparison of Estimates vs True Estimates (True Coefficients and Intercepts):



Investment and their coefficient and CATE intercept results:

Investment	Coefficient Results		CATE intercept Results	
	Coefficient (Size)	p-value	CATE intercept	p-value
Tech Support	0.018	0.083	5529.309	0.0
Discount	0.048	0.0	647.327	0.38
Tech Support and Discount	**0.071**	**0.0**	**5081.263**	**0.0**

CONCLUSION:

With the highest Size coefficient (0.071) and significant intercept (5081.263) - the most effective investment is Tech Support and Discount (Combined).