****

**Shapley Values: Unlocking Fair Attribution in Adtech**

**Introduction**

In the fast-evolving world of adtech, accurately attributing value across data sources, models, and campaign strategies is a persistent challenge. As campaigns grow more sophisticated—incorporating seed data, audience analytics, and machine learning models—the question of *who* or *what* deserves credit for performance becomes both more complex and more important. Enter the Shapley value: a mathematically rigorous approach borrowed from cooperative game theory, now finding new relevance in advertising attribution and machine learning interpretability.

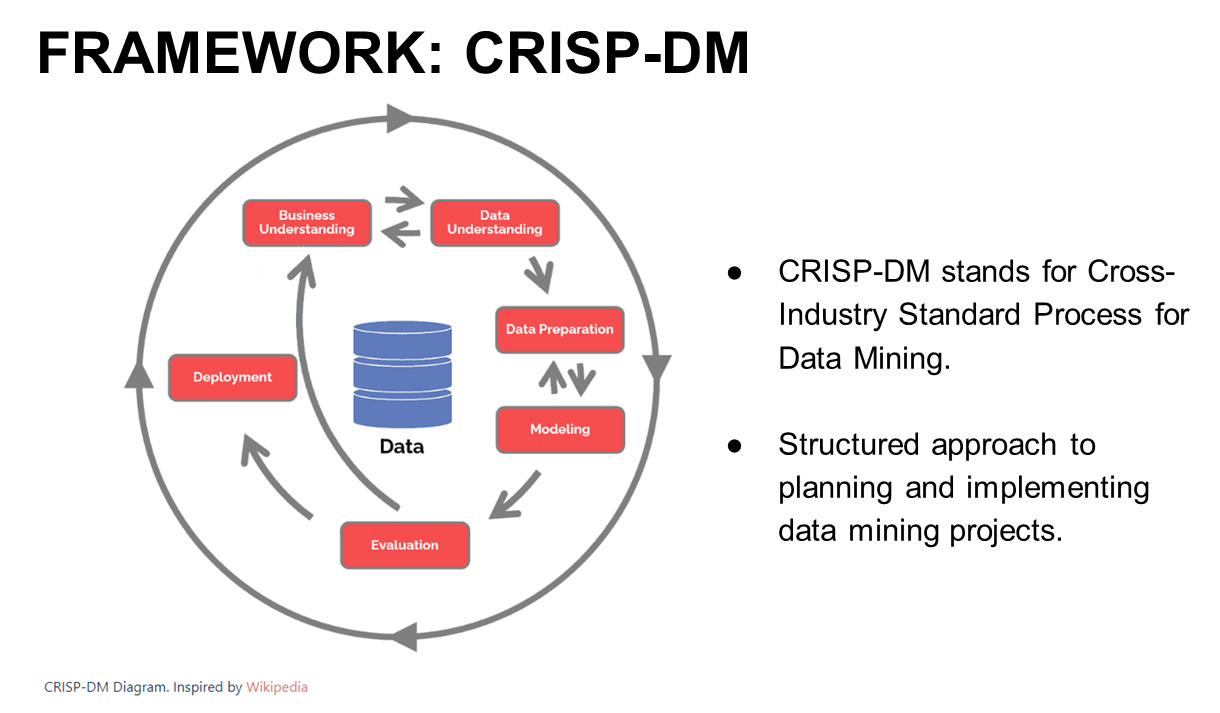
**What Are Shapley Values?**

Shapley values provide a principled way to distribute gains among a coalition of players, ensuring each participant receives a "fair" share of the total value generated by their cooperation. In the adtech context, these "players" might be data sources, audience segments, or predictive models, and the "gain" is typically campaign performance—often measured as revenue per ad spend.

Formally, for a set N*N* of n*n* players and a characteristic function v*v* mapping subsets of players to real numbers, the Shapley value for player i*i* is:

φi(v)=1n∑S⊆N∖{i}(n−1∣S∣)−1(v(S∪{i})−v(S))*φi*(*v*)=*n*1*S*⊆*N*∖{*i*}∑(∣*S*∣*n*−1)−1(*v*(*S*∪{*i*})−*v*(*S*))

This formula sums the marginal contributions player i*i* makes to every possible subset S*S* of the coalition, normalizing to ensure fairness.



**Why Shapley Values Matter in Adtech**

**Fair Attribution Across Data and Models**

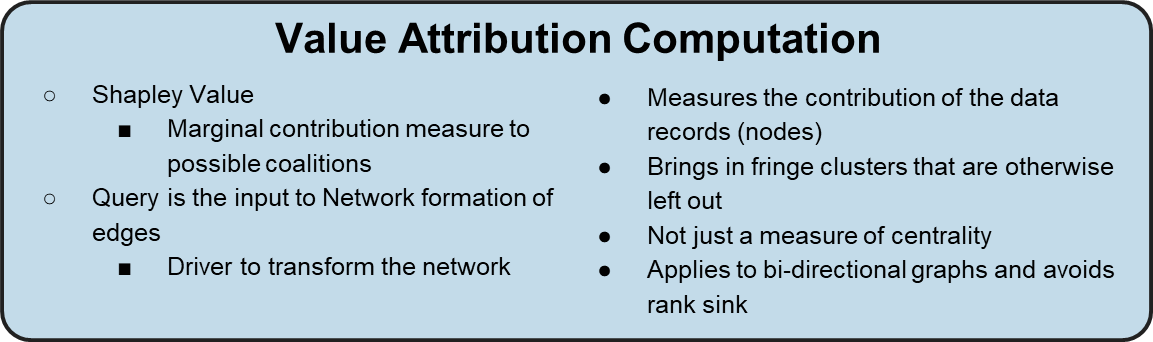
Consider a campaign using three key components:

* **Seed Data (S)**
* **3rd Party Data Provider Data (A)**
* **Logistic Regression Model (LR)**

Each of these can contribute to different audience segments, but none can drive meaningful results alone. The challenge is to fairly distribute the campaign's total value (e.g., revenue/ad spend) among these contributors.

**Properties That Make Shapley Values Unique:**

* **Efficiency:** The total value is fully distributed—no more, no less.
* **Symmetry:** Equivalent contributors receive equal credit.
* **Linearity:** Contributions add up logically across combined scenarios.
* **Null Player:** Contributors that add no value receive none.



**Applying Shapley Values to Campaign Attribution**

Let's see how this works in practice with three audience segments:

| **Segment ID** | **Description** | **Sources Used** |
| --- | --- | --- |
| PR-BRAND-001 | Lookalike segment | S + A + LR |
| PR-BRAND-002 | Lookalike/historic | S + A, S + A + LR |
| PR-BRAND-003 | Historic SGT buyers | S + A |

The value for each subset (e.g., S+A, S+A+LR) is calculated as total revenue divided by total ad spend across relevant segments.

* V(S+A)=Rev003+Rev002Spend003+Spend002*V*(*S*+*A*)=*Spend*003+*Spend*002*Rev*003+*Rev*002
* V(S+A+LR)=Rev003+Rev002+Rev001Spend003+Spend002+Spend001*V*(*S*+*A*+*LR*)=*Spend*003+*Spend*002+*Spend*001*Rev*003+*Rev*002+*Rev*0011

**Shapley Value Calculations:**

* **For LR:** 13(V(S+A+LR)−V(S+A))31(*V*(*S*+*A*+*LR*)−*V*(*S*+*A*))
* **For S and A:** 16V(S+A)+13V(S+A+LR)61*V*(*S*+*A*)+31*V*(*S*+*A*+*LR*)

This ensures that each contributor's marginal impact is considered across all possible combinations, yielding a fair and transparent attribution1.

**Addressing Attribution Challenges**

One insight from applying Shapley values is that foundational data sources (like S and A) often receive significant credit—because they're prerequisites for advanced modeling. However, this can sometimes overstate their unique value, especially if alternative data sources could be substituted without loss of performance.

**Alternative Attribution Strategies:**

* **Full Credit to LR:** Assign all value from lookalike segments to the LR model, reflecting its central role in those results.
* **Split Credit:** Divide lookalike segment value between LR and the data sources (e.g., 50/50 split), acknowledging both the model's and the data's contributions.

These variations can be modeled by adjusting the value functions and Shapley formulas accordingly, allowing for flexible and context-sensitive attribution.

**Shapley Values in Machine Learning Interpretability**

Shapley values also underpin leading methods for interpreting machine learning models, such as SHAP. Here, each feature is treated as a "player," and the Shapley value quantifies its contribution to a specific prediction. This approach unifies disparate feature attribution methods and ensures explanations are locally accurate, consistent, and fair.

**Summary**

Shapley values offer a robust, mathematically grounded framework for attribution in adtech. Whether distributing credit among data sources, models, or campaign strategies, they ensure transparency and fairness—qualities essential for optimizing spend and fostering trust among stakeholders. As adtech continues to integrate more complex data and modeling approaches, Shapley-based attribution is poised to become a cornerstone of effective campaign measurement and optimization as well as more efficient workflow.

About Precise.ai

Precise is a leader in data verification and secure collaboration, specializing in Proof of Data technology to ensure transparency, trust, and compliance. Through blockchain-backed proof methodologies, Precise provides data owners with an immutable record of accuracy, lineage, and responsible data use. In addition to data verification, Precise enables privacy-preserving data collaboration through federated machine learning, allowing organizations to extract insights and drive value while keeping sensitive data protected and decentralized. By establishing verifiable data integrity and privacy-first data activation, Precise empowers businesses to confidently leverage their data while maintaining compliance and accountability.  Contact us at [info@precise.ai](mailto:info@precise.ai).