Data Shapley:

Equitable Valuation of Data for Machine Learning

Prepared by Tian Xiao

Overview

Motivation

Cooperative Games
& Shapley Value

Data Shapley

Application

Collaborative Machine Learning

- **Data** is the fuel powering machine learning.
- Where does data come from?

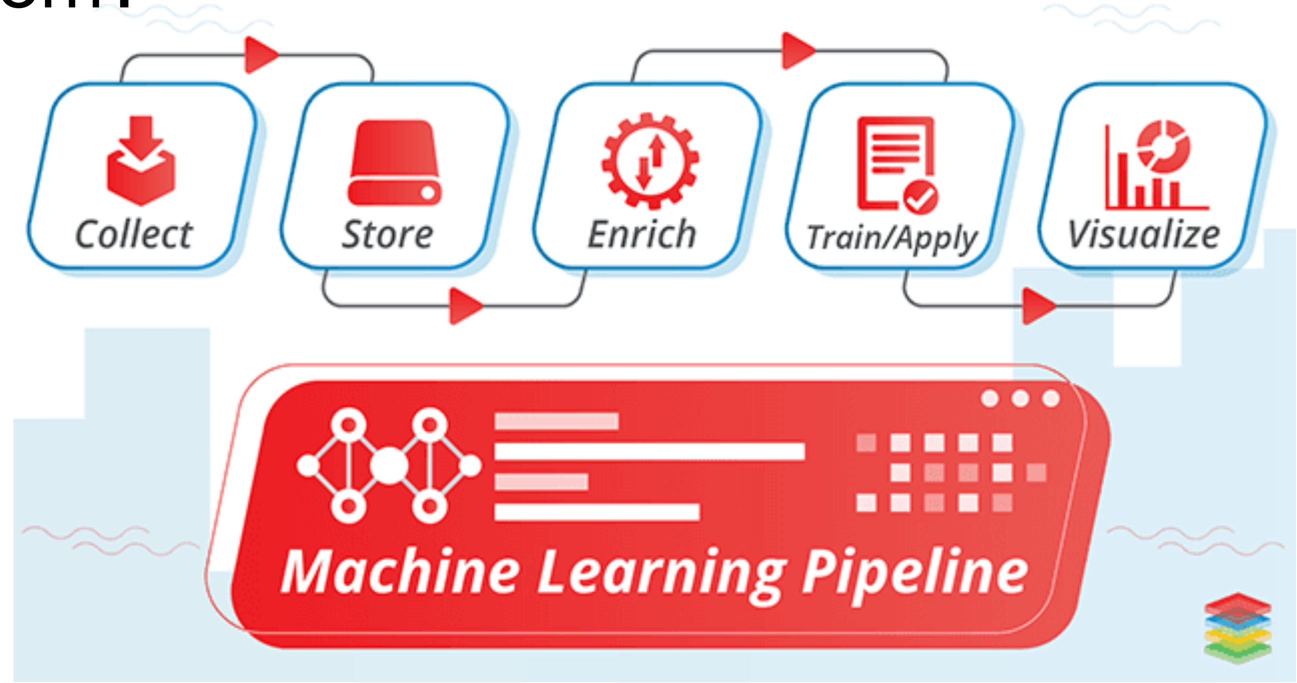
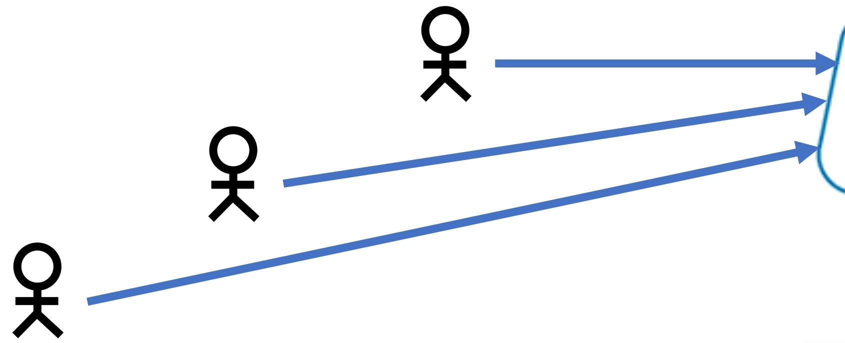


Figure: Machine Learning Pipeline *(Gill, 2022)*.

Collaborative Machine Learning

- **Data** is the fuel powering machine learning.
- Where does data come from?



Individuals

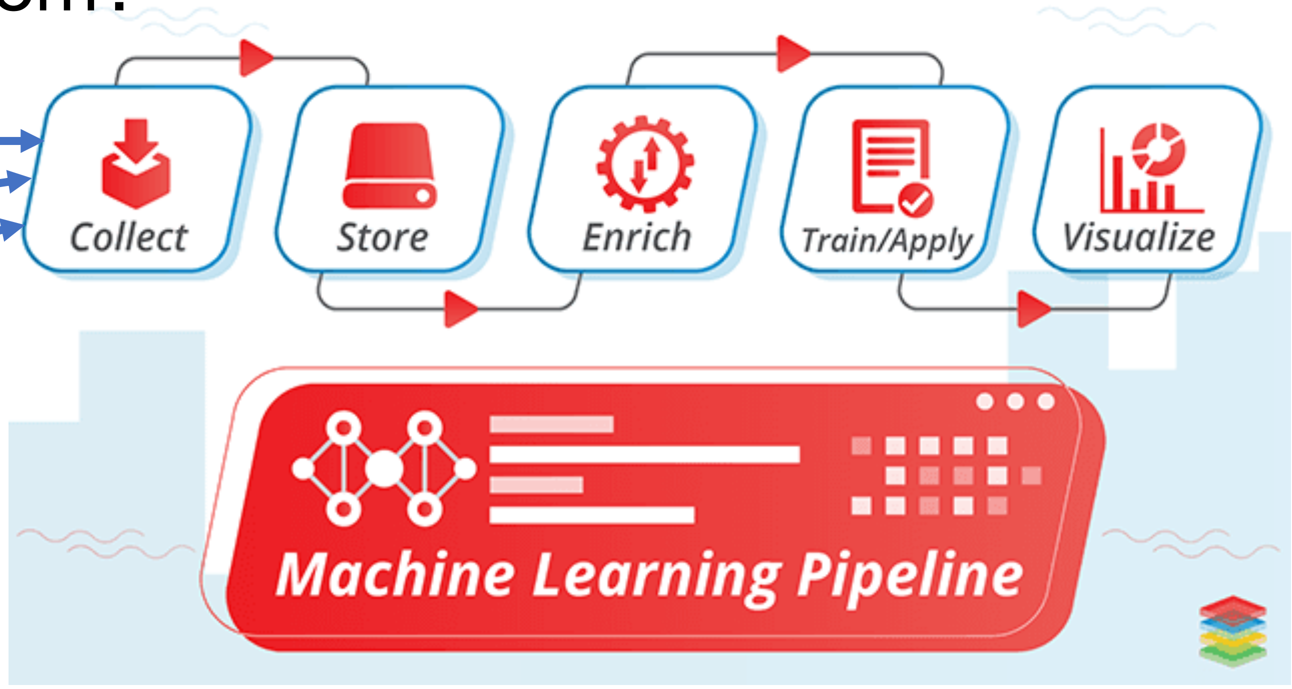


Figure: Machine Learning Pipeline *(Gill, 2022)*.

General Data Protection Regulation

- **Data** are properties. Properties are not **free** for use.

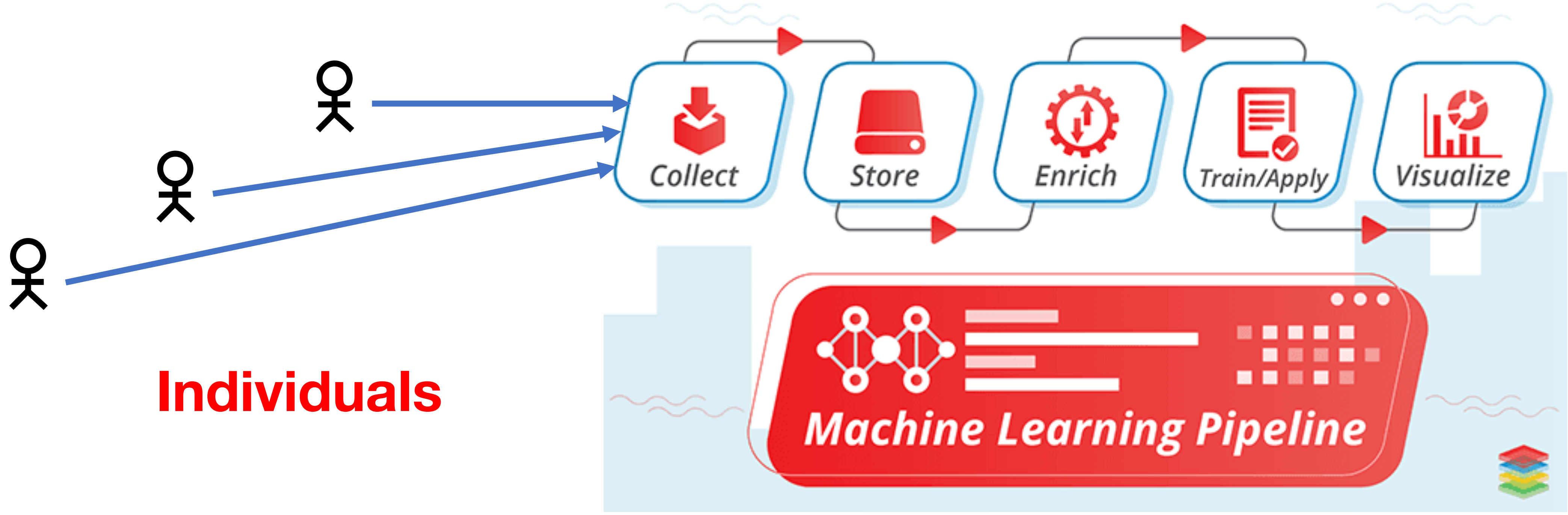


Figure: Machine Learning Pipeline *(Gill, 2022)*.

Data Valuation

- Need to assign a value to each individual's data so that everyone is fairly compensated.

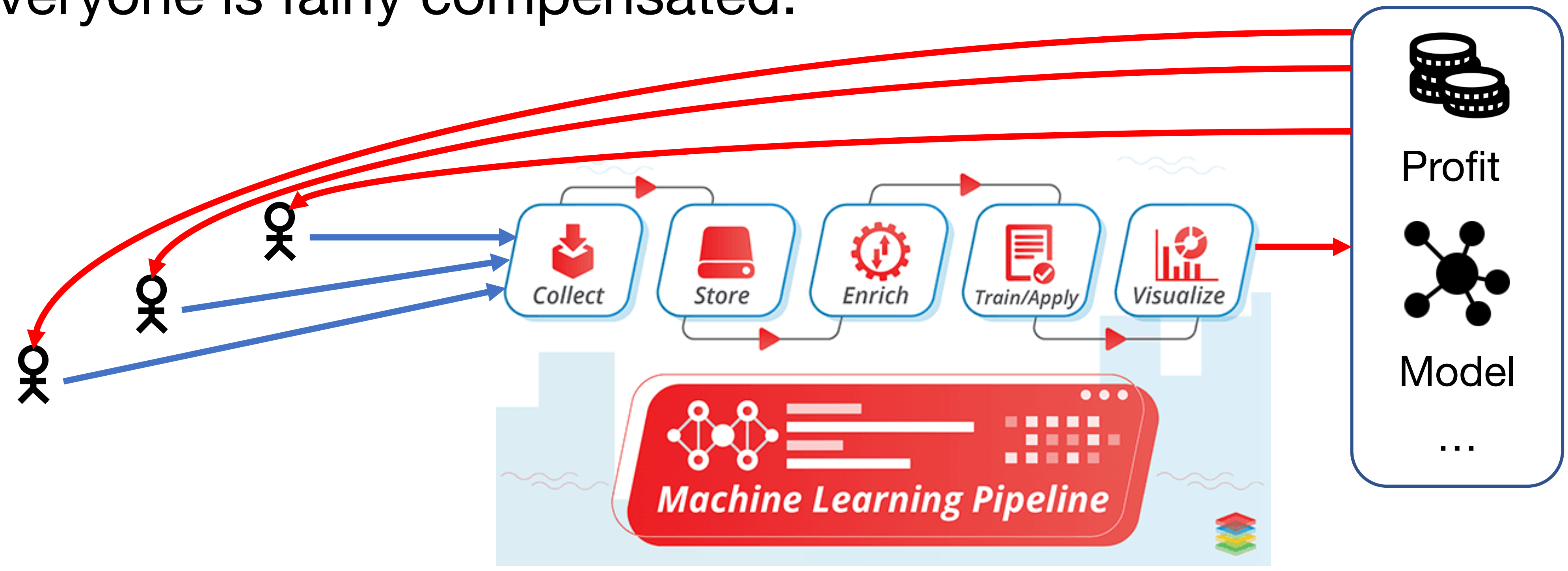
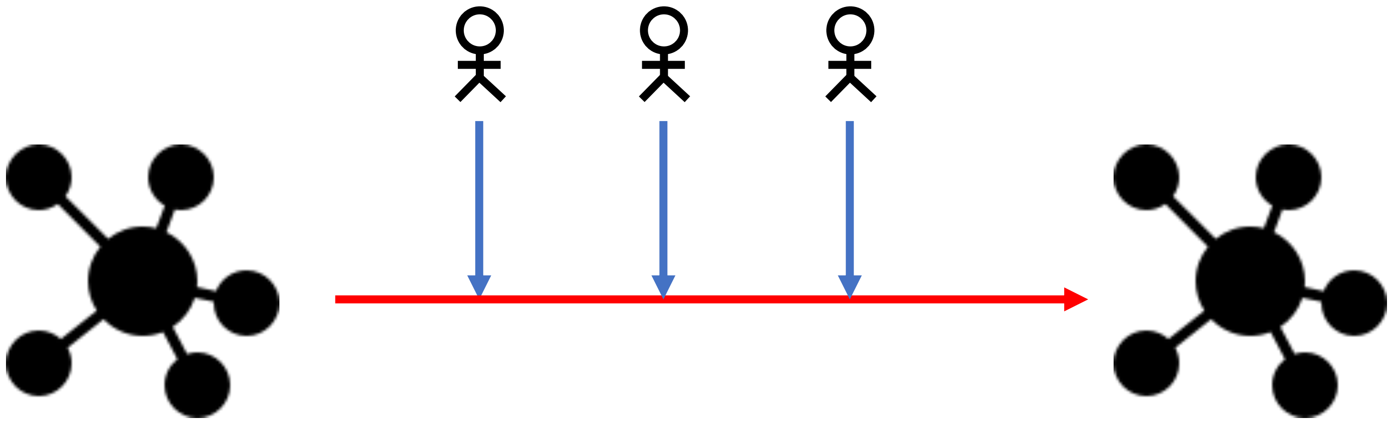


Figure: Machine Learning Pipeline *(Gill, 2022)*.

A cooperative game!

- Through cooperation, we obtain a **better** model than without cooperation.



Initial model

Better model

Evaluation metrics

- Accuracy
- MSE
- F1 score
- Information gain
- …

Game Theory

Traditional

- Players are **rational** and **selfish**.
- “Prisoner's Dilemma”: Both prisoners will eventually choose to **defect** because whatever the other prisoner choose, to defect gives the better outcome.

Figure: Prisoner’s Dilemma *(Forsythe, 2012)*.

Game Theory

Traditional

- Players are **rational** and **selfish**.
- “Prisoner's Dilemma”: Both prisoners will eventually choose to **defect** because whatever the other prisoner choose, to defect gives the better outcome.

But this is not the best outcome!

Figure: Prisoner’s Dilemma *(Forsythe, 2012)*.

Game Theory

Traditional

- Players are **rational** and **selfish**.
- "Prisoner's Dilemma": Both prisoners will eventually choose to **defect** because whatever the other prisoner choose, to defect gives the better outcome.

Cooperative

- Players have **common interests**, **information exchange** and **compulsory contract**.
- Both prisoners should **not** defect to gain mutual benefits.

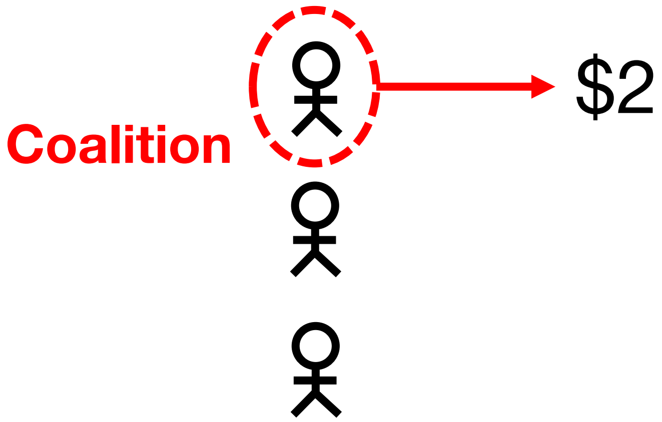
Cooperative Games

- A **game** is uniquely defined by a set function

$$V: 2^N \to \mathbb{R}$$

aka Value Function

where N represents the set of players in the game.



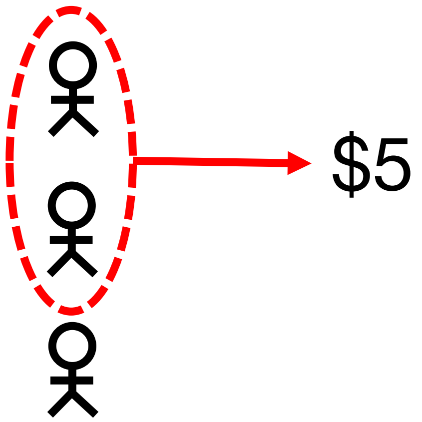
Cooperative Games

- A **game** is uniquely defined by a set function

$$V: 2^N \to \mathbb{R}$$

aka Value Function

where N represents the set of players in the game.



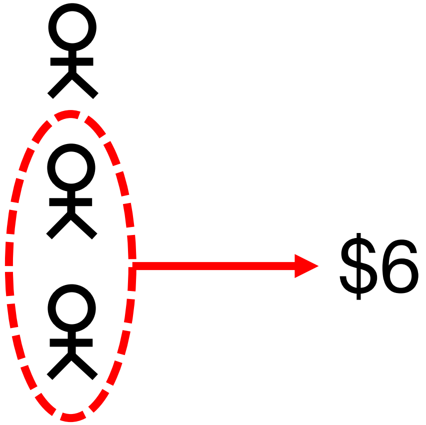
Cooperative Games

- A **game** is uniquely defined by a set function

$$V: 2^N \to \mathbb{R}$$

aka Value Function

where N represents the set of players in the game.



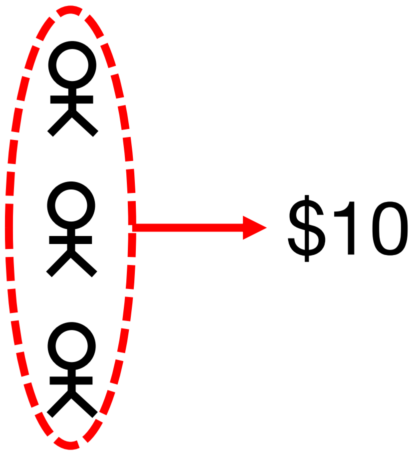
Cooperative Games

- A **game** is uniquely defined by a set function

$$V: 2^N \to \mathbb{R}$$

aka Value Function

where N represents the set of players in the game.

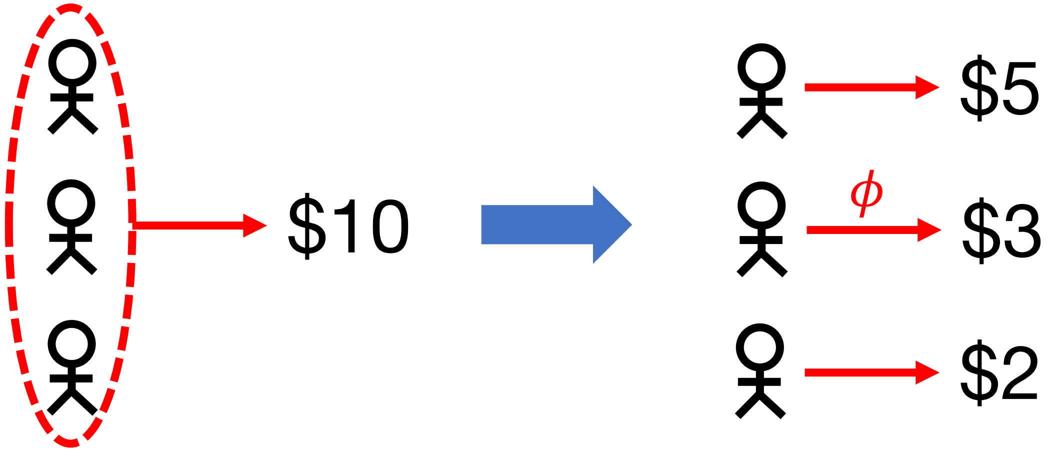


Contribution Function

- To measure the contribution of each player, we define

$$\phi_V : N \to \mathbb{R}$$

where N represents the set of players in the game.

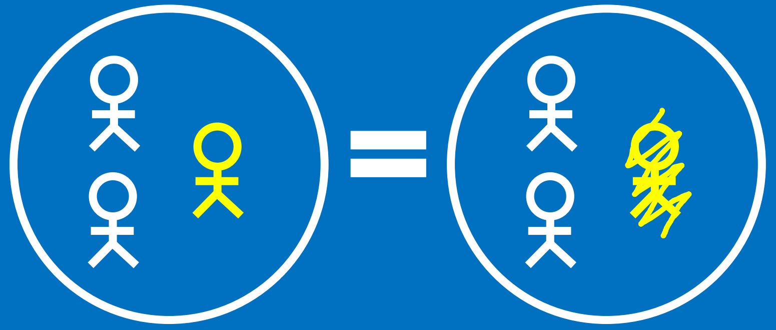


Fair Measure of Contribution

Analogy: Measure the value of a new colleague in the workplace.

Null Player

When player i joins any existing work group, he does not add value to that group.



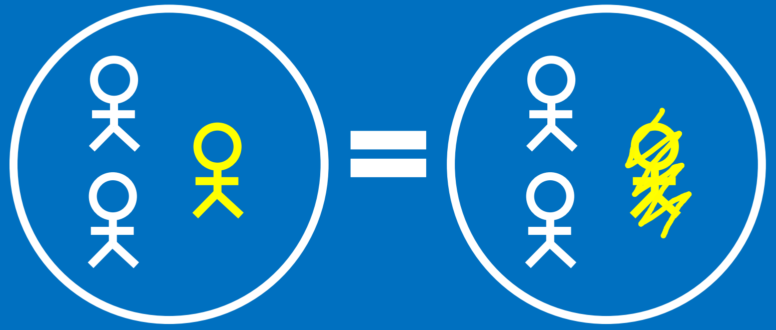
$$\forall S \subseteq N \backslash \{i\}\ [V(S) = V(S + \{i\})]$$
$$\Rightarrow \boldsymbol{\phi(i) = 0}$$

Fair Measure of Contribution

Analogy: Measure the value of a new colleague in the workplace.

Null Player

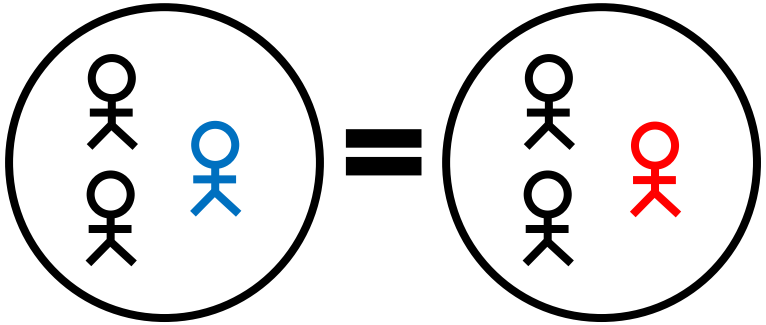
When player i joins any existing work group, he does not add value to that group.



$$\forall S \subseteq N \backslash \{i\}\ [V(S) = V(S + \{i\})] \Rightarrow \boldsymbol{\phi(i) = 0}$$

Symmetry

When player i and j join any existing work group, they add the same value to that group.



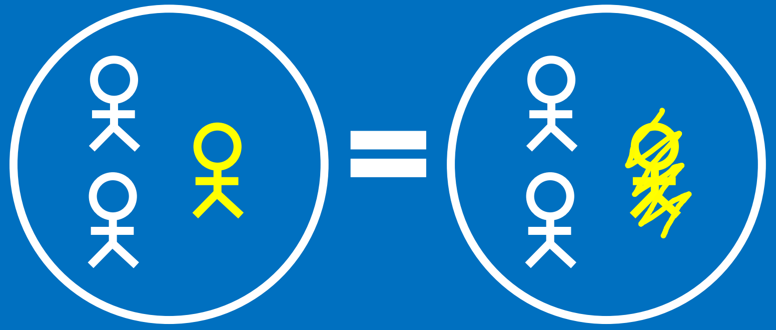
$$\forall S \subseteq N \backslash \{i, j\}\ [V(S + \{i\}) = V(S + \{j\})] \Rightarrow \boldsymbol{\phi(i) = \phi(j)}$$

Fair Measure of Contribution

Analogy: Measure the value of a new colleague in the workplace.

Null Player

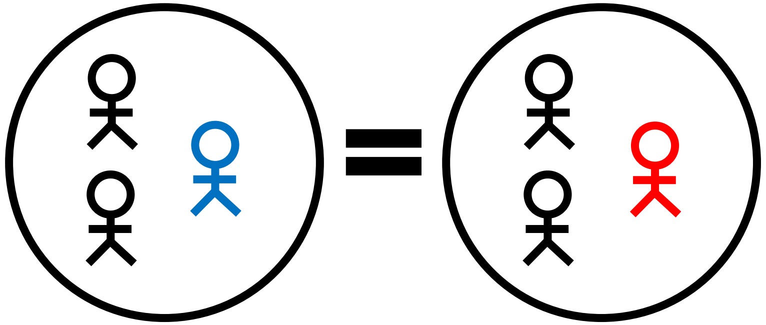
When player i joins any existing work group, he does not add value to that group.



$$\forall S \subseteq N \backslash \{i\}\ [V(S) = V(S + \{i\})] \Rightarrow \boldsymbol{\phi(i) = 0}$$

Symmetry

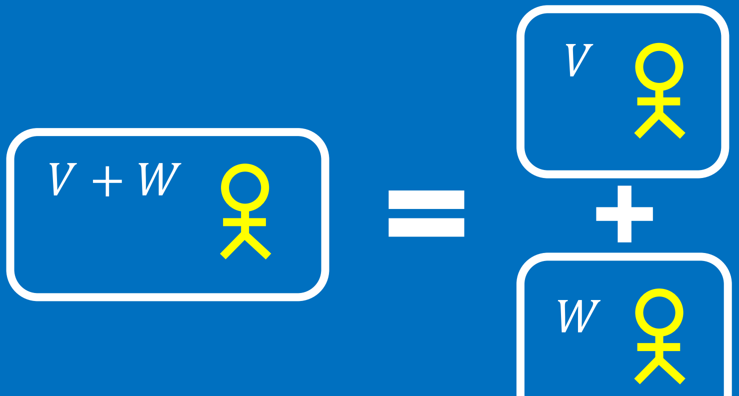
When player i and j join any existing work group, they add the same value to that group.



$$\forall S \subseteq N \backslash \{i, j\}\ [V(S + \{i\}) = V(S + \{j\})] \Rightarrow \boldsymbol{\phi(i) = \phi(j)}$$

Linearity

We have two scores V and W for each work group. We take the combined score as $V + W$.



$$\forall i \in N$$
$$[\boldsymbol{\phi_V(i) + \phi_W(i) = \phi_{V+W}(i)}]$$

Shapley Value

- Shapley found such a value:

$$\phi(i) = \frac{1}{|N|} \sum_{S \subseteq N \setminus \{i\}} \frac{V(S + \{i\}) - V(S)}{\binom{n-1}{|S|}}$$

  - Besides Null Player, Symmetry and Linearity, the Shapley value is special such that it is the only one that satisfies **Efficiency**:

$$\sum_{i \in N} \phi(i) = V(N)$$

Figure: Lloyd S. Shapley *(Moreno et al., 2018)*.

Shapley Value

Marginal contribution

- Shapley found such a value:

$$\phi(i) = \frac{1}{|N|} \sum_{S \subseteq N \setminus \{i\}} \frac{V(S + \{i\}) - V(S)}{\binom{n-1}{|S|}}$$

  - Besides Null Player, Symmetry and Linearity, the Shapley value is special such that it is the only one that satisfies **Efficiency**:

$$\sum_{i \in N} \phi(i) = V(N)$$

Figure: Lloyd S. Shapley *(Moreno et al., 2018)*.

Data Shapley

Performance metrics/Information gain

$$\phi(i) = \boldsymbol{C} \sum_{S \subseteq N \setminus \{i\}} \frac{V(S + \{i\}) - V(S)}{\binom{n-1}{|S|}}$$

- S is every subset of N, leading to **very high computational cost** (in machine learning, we usually have millions of data!).

Truncated Monte Carlo (TMC) Shapley

- General idea I: Take a random permutation of data and calculating the marginal contribution in a **rolling** basis.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

$$V(\{1\}) - V(\emptyset)$$

Truncated Monte Carlo (TMC) Shapley

- General idea I: Take a random permutation of data and calculating the marginal contribution in a **rolling** basis.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

$$V(\{1,2\}) - V(\{1\})$$

Truncated Monte Carlo (TMC) Shapley

- General idea I: Take a random permutation of data and calculating the marginal contribution in a **rolling** basis.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

$$V(\{1,2,3\}) - V(\{1,2\})$$

Truncated Monte Carlo (TMC) Shapley

- General idea I: Take a random permutation of data and calculating the marginal contribution in a **rolling** basis.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

$$V(\{1,2,3,4\}) - V(\{1,2,3\})$$

Truncated Monte Carlo (TMC) Shapley

- General idea I: Take a random permutation of data and calculating the marginal contribution in a **rolling** basis.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

$$\boxed{V(\{1,2,3,4,5\}) - V(\{1,2,3,4\})}$$

Truncated Monte Carlo (TMC) Shapley

- General idea I: Take a random permutation of data and calculating the marginal contribution in a **rolling** basis.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

$$V(\{1,2,3,4,5,6\}) - V(\{1,2,3,4,5\})$$

Truncated Monte Carlo (TMC) Shapley

- General idea I: Take a random permutation of data and calculating the marginal contribution in a **rolling** basis.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

$$\boxed{V(\{1,2,3,4,5,6\}) - V(\{1,2,3,4,5\})} < \epsilon$$

Truncated Monte Carlo (TMC) Shapley

- General idea II: When the marginal contribution becomes very small, mark all the remaining contribution as 0.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

$$\boxed{V(\{1,2,3,4,5,6\}) - V(\{1,2,3,4,5\}) = 0}$$

Truncated Monte Carlo (TMC) Shapley

- General idea II: When the marginal contribution becomes very small, mark all the remaining contribution as 0.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

$$\boxed{V(\{1,2,3,4,5,6,7\}) - V(\{1,2,3,4,5,6\}) = 0}$$

Truncated Monte Carlo (TMC) Shapley

- General idea II: When the marginal contribution becomes very small, mark all the remaining contribution as 0.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

$$V(\{1,2,3,4,5,6,7,8\}) - V(\{1,2,3,4,5,6,7\}) = 0$$

Truncated Monte Carlo (TMC) Shapley

- General idea II: When the marginal contribution becomes very small, mark all the remaining contribution as 0.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

$$V(\{1,2,3,4,5,6,7,8,9\}) - V(\{1,2,3,4,5,6,7,8\}) = 0$$

Truncated Monte Carlo (TMC) Shapley

- General idea II: When the marginal contribution becomes very small, mark all the remaining contribution as 0.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

$$V(\{1,2,3,4,5,6,7,8,9,10\}) - V(\{1,2,3,4,5,6,7,8,9\}) = 0$$

Application: Low Quality Data

Mislabelled data has
low (even –ve) Data Shapley value!

Label: Sunflower
Value = -0.00484

True Label: Daisy

Label: Daisy
Value = -0.00395

True Label: Rose

Label: Sunflower
Value = -0.00456

True Label: Dandelion

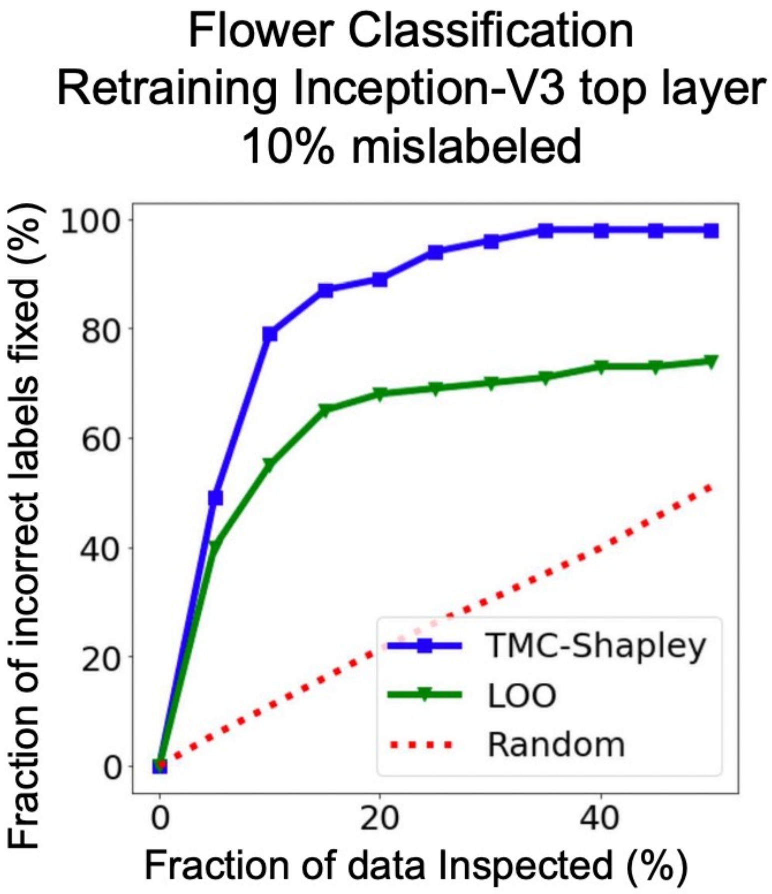
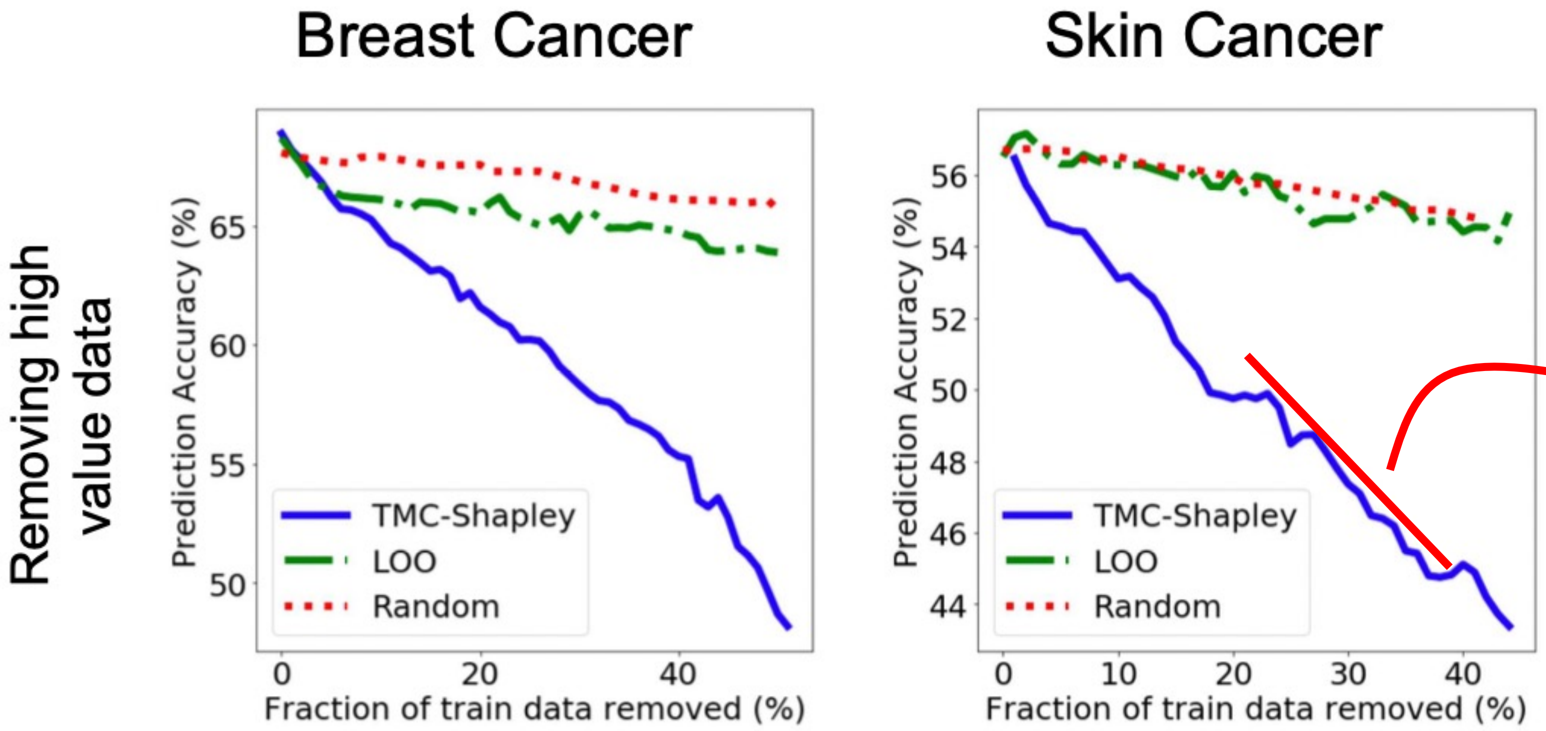


Figure: Identifying mislabelled data and correcting them *(Ghorbani & Zou, 2018)*.

Application: Differentiate Data Sources

- "All data sources are not created equal."



Largest gradient → More reflective of the true value

Figure: Change of prediction accuracy as high value data are removed gradually *(Ghorbani & Zou, 2018)*.

Application: Adapt to New Data

1. Use performance metrics on **target data** as value function.
2. Remove –ve value data.
3. Use value of data as **weight** when training them.

Source to Target	Prediction Task	Trained Model	Original Performance (%)	Adapted Performance (%)
Google to HAM1000	Skin Lesion Classification	Retraining Inception-V3 top layer	29.6	37.8
CSU to PP	Disease Coding	Retraining DeepTag top layer	87.5	90.1
LFW+ to PPB	Gender Detection	Retraining Inception-V3 top layer	84.1	91.5
MNIST to UPS	Digit Recognition	Multinomial Logistic Regression	30.8	39.1
Email to SMS	Spam Detection	Naive Bayes	68.4	86.4

Figure: Original performance vs Data Shapley Adapted Performance on different prediction tasks *(Ghorbani & Zou, 2018)*.

Alternatives to Data Shapley

- **Cook's Distance** in Linear Regression
- **Leverage** and **Influence**

These quantities does not satisfy **Null Player**, **Symmetry** and **Linearity**!

References

Forsythe, G. (2012, December 4). *Prisoner's Dilemma*. Flickr. https://www.flickr.com/photos/gforsythe/8245423564

Ghorbani, A., & Zou, J. (2019, May). Data Shapley: Equitable Valuation of Data for Machine Learning. In *International Conference on Machine Learning* (pp. 2242-2251). PMLR.

Ghorbani, A., Kim, M., & Zou, J. (2020, November). A distributional framework for data valuation. In International Conference on Machine Learning (pp. 3535-3544). PMLR.

Gill, N. S. (2022, August 19). *Machine Learning Pipeline Deployment and Architecture*. Xenonstack. https://www.xenonstack.com/blog/machine-learning-pipeline

Jia, R., Sun, X., Xu, J., Zhang, C., Li, B., & Song, D. (2019). An empirical and comparative analysis of data valuation with scalable algorithms.

Koh, P. W., & Liang, P. (2017, July). Understanding black-box predictions via influence functions. In International conference on machine learning (pp. 1885-1894). PMLR.

Kwon, Y., & Zou, J. (2021). Beta Shapley: a unified and noise-reduced data valuation framework for machine learning. arXiv preprint arXiv:2110.14049.

Moreno, V., Ramírez M. E., Oliva C. D. L., & Moreno E. (2018, May 21). *Biografía de Lloyd S. Shapley*. Busca Biografías. https://www.buscabiografias.com/biografia/verDetalle/9903/Lloyd%20S.%20Shapley

Appendix: Use C instead of $\frac{1}{|N|}$

$$\phi(i) = \boldsymbol{C} \sum_{S \subseteq N \setminus \{i\}} \frac{V(S + \{i\}) - V(S)}{\binom{n-1}{|S|}}$$

- In data valuation, the **Efficiency** axiom is not that useful.
- C can be any arbitrary constant representing the scale since it does not affect the relative weight between data points.

Appendix: Limitation of Data Shapley

- Still expensive in **time**!
- Data Shapley gives each cardinality a **uniform weight** ($\frac{1}{|N|}$). This is actually **suboptimal**!
- The value of each data depends on our **chosen dataset.**
- The Efficiency axiom is **not** important in ML setting ☺!

Appendix: Leave-one-out (LOO) Value

$$LOO(i) = V(N) - V(N\{i\})$$

This is actually the marginal contribution to the grand coalition without i!

- Leave-one-out value is much **easier to compute** than the Shapley value, and it is **robust to clone**.
- However, it does not satisfy **linearity**.

Appendix: Variants of Data Shapley

$$\phi(i) = \frac{1}{|N|} \sum_{S \subseteq N \setminus \{i\}} \frac{\text{marginal contribution of } i}{\binom{n-1}{|S|}}$$

- **Banzhaf index:** $\frac{1}{2^{|N|-1}} \sum_{S \subseteq N \setminus \{i\}} \text{marginal contribution of } i$
- **Beta Shapley:** $\frac{1}{|N|} \sum_{S \subseteq N \setminus \{i\}} w \cdot \frac{\text{marginal contribution of } i}{\binom{n-1}{|S|}}$, where $w \sim Beta(\alpha, \beta)$.
- **$\mathfrak{D}$-Shapley:** $\mathbb{E}_{D^{|N|}}(\phi(i))$