DeRDaVa: Deletion-Robust Data Valuation for Machine Learning

Xiao Tian[1,2], Rachael Hwee Ling Sim[1], Jue Fan[1,2], Bryan Kian Hsiang Low[1]

[1] Department of Computer Science, National University of Singapore
[2] Department of Mathematics, National University of Singapore
{xiao.tian, rachael.sim, jue.fan}@u.nus.edu, lowkh@comp.nus.edu.sg

Abstract

Data valuation is concerned with determining a fair valuation of data from data sources to compensate them or to identify training examples that are the most or least useful for predictions. With the rising interest in personal data ownership and data protection regulations, model owners will likely have to fulfil more data deletion requests. This raises issues that have not been addressed by existing works: *Are the data valuation scores still fair with deletions? Must the scores be expensively recomputed?* The answer is no. To avoid recomputations, we propose using our data valuation framework DeRDaVa upfront for valuing each data source's contribution to preserving robust model performance after anticipated data deletions. DeRDaVa can be efficiently approximated and will assign higher values to data that are more useful or less likely to be deleted. We further generalize DeRDaVa to Risk-DeRDaVa to cater to risk-averse/seeking model owners who are concerned with the worst/best-cases model utility. We also empirically demonstrate the practicality of our solutions.

1 Introduction

Data is essential to building machine learning (ML) models with high predictive performance and model utility. Model owners source for data directly from their customers or from collaborators in collaborative machine learning (Nguyen et al. 2022a). For example, a credit card company can train an accurate ML model to predict the probability of default based on consumers' income and payment history data (Tsai and Chen 2010). As another example, a healthcare firm can train an ML model to predict the progression of diabetes based on patients' health data (TCFOD 2019). As the quality of data contributed by multiple data sources may vary widely, several works (Fan et al. 2022; Tay et al. 2022; Xu et al. 2021) have recognized the importance of *data valuation* to help model owners compensate data sources fairly, or identify data that are most or least useful for predictions.

Data valuation studies how much data is worth and proposes how rewards associated with the ML model can be **fairly** allocated to each data source (Sim, Xu, and Low 2022). Several existing data valuation techniques (Ghorbani and Zou 2019; Jia et al. 2019; Yang et al. 2019) have adopted

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Setting

$v(S\cap D')$ / Staying set D'	$v(\{\bigstar,\blacksquare\}\cap D')$	$v(\{\bigstar\}\cap D')$	$v(\{\blacksquare\}\cap D')$
$\{\bigstar,\blacksquare\}$ (with prob .7)	$=v(\{\bigstar,\blacksquare\})$ $=.8$	$=v(\{\bigstar\})=.5$	$=v(\{\blacksquare\})=.5^{\blacksquare}$
$\{\bigstar\}$ (with prob .3)	$=v(\{\bigstar\})=.5$	$=v(\{\bigstar\})=.5$	$=v(\emptyset)=0$

Valuation

	$\bigstar$	$\blacksquare$
Data Shapley Risk-seeking ($\alpha=.7$)	$\frac{.8-.5^{\blacksquare}}{2}+\frac{.5}{2}=\underline{.4}$	$\frac{.8-.5}{2}+\frac{.5^{\blacksquare}}{2}=\underline{.4}$
DeRDaVa	$\underline{.4}\times.7+\underline{.5}\times.3=\underline{.43}$	$\underline{.4}\times.7+\underline{0}\times.3=\underline{.28}$
Risk-averse ($\alpha=.3$)	$=\underline{.5}$	$=\underline{0}$

Figure 1: Comparison of Data Shapley vs. the deletion-robustified DeRDaVa and Risk-DeRDaVa scores with Shapley prior for a game with 2 *interchangeable* sources. ★ and ■ stay with probability 1 and .7 respectively. ★ and ■ have equal Data Shapley score but ★ has higher DeRDaVa and Risk-averse DeRDaVa scores. This is because Data Shapley (Eq. (1)) considers only the initial support set {★, ■} while DeRDaVa (Eq. (8)) and Risk-averse DeRDaVa (Eq. (10)) also consider the worst-case support set {★}. Further explanation is included in App. A.1.

the use of *semivalues* from cooperative game theory. Recent works (Sim et al. 2020; Tay et al. 2022; Zhang, Wu, and Pan 2021) have also developed various reward allocation schemes based on semivalues. The core intuition behind semivalues is that a data source should be fairly valued relative to other data sources (i.e., based on the averaged model utility improvement it contributes to each sub-coalition of other data sources). Moreover, these works have justified the use of semivalues by fairness axioms such as *Interchangeability* — assigning the same valuation score to two data sources d_i and d_j with the same model utility improvement to every sub-coalition (e.g., ★ and ■ in Fig. 1).

Data deletion refers to the deletion of **data and their impact** from trained ML models. Due to the recent introduction of data protection legislation, such as General Data Protection Regulation (GDPR) in the European Union and California Consumer Privacy Act (CCPA) in the United States, data deletions are expected to occur more frequently. These laws legally assert that data are properties of their owners and data owners have the *right to be forgotten* (Shastri, Wasserman, and Chidambaram 2019). This mandates model owners to **delete the data and their impact** from trained models *without undue delay* (Magdziarczyk 2019) upon request or after a stipulated time (Ong 2018). While these regulations have led to intensive research on *machine unlearning* (Bourtoule et al. 2021; Chen et al. 2021; Sekhari et al. 2021) to efficiently remove the impact of deleted data from trained models, to our knowledge, no work has considered the impact of data deletions on data valuation.

It is our view that data deletions challenge existing semivalue-based data valuation techniques. The existing (pre-deletion) semivalue may not preserve the fairness axioms after data deletions. For example, if data sources d_i and d_j only make the same model utility improvement to every sub-coalitions after deletions, they may have been assigned different pre-deletion semivalues (which violates the Interchangeability axiom) (see App. A). At first glance, one may think of recomputing the semivalue after every deletion to address the challenge. However, the recomputations are computationally expensive and data owners (and legislators) may not tolerate the uncertainty and fluctuations in valuation (e.g., having to return monetary compensation). Thus, we propose new *deletion-robustified fairness axioms* and a more proactive approach that anticipates and accounts for data deletions: the model owner should assign higher **deletion-robust data valuation (DeRDaVa)** scores upfront to data sources with higher probability of staying (and thus contribute to *deletion-robustness*, i.e., robustly preserving model performance after deletions). We show that DeRDaVa **satisfies our deletion-robustified fairness axioms** (Sec. 3.1) and can be efficiently approximated (Sec. 3.3). Lastly, to cater to model owners who are only concerned with the expectation of **worst-cases model utility** (i.e., a *risk-averse* model owner) or best-cases model utility (i.e., a *risk-seeking* model owner), we generalize DeRDaVa to **Risk-DeRDaVa** (Sec. 3.4). In Sec. 4, we empirically demonstrate and compare the behaviour of DeRDaVa with semivalues after data deletions on real-world datasets.

2 Background and Related Works

2.1 Semivalue-Based Data Valuation

Machine learning can be viewed as a *cooperative game* among multiple data sources in order to gain the highest model performance, where each data source can be a single data point or a smaller dataset. Semivalue (Dubey, Neyman, and Weber 1981) is a concept from cooperative game theory (CGT) that measures the contribution of each data source in such a cooperative game. To formalize this, we define such a cooperative game as a pair $\langle D, v\rangle$, where the *support set* $D = \bigcup_{i=1}^{n}\{d_i\}$ represents the set of n data sources d_i indexed by i in the collaboration and the *model utility function* $v : \mathcal{P}(D) \to \mathbb{R}$ maps each coalition of data sources in the power set of D to its utility. Specifically, the utility $v(S)$ of a coalition S of data sources can represent the prediction performance (e.g., validation accuracy) of model trained with data from S. For example, $v(\{d_0, d_1, d_2\}) = 0.9$ may represent that we use data from the three data sources d_0, d_1 and d_2 to train a model and obtain an accuracy of 0.9 when evaluated on a validation set.

Let G^n represent the set of all cooperative games with respective support set D sized n. To measure the contribution of each data source, we want to find an n-sources *data valuation function* $\phi^n : G^n \to \mathbb{R}^n$ that assigns each data source d_i a real-valued valuation score $\phi_i^n(\langle D, v\rangle)$ abbreviated as $\phi_i^n(v)$. To ensure the **fairness** of data valuation functions, a common approach in CGT is *axiomatization*, where a list of axioms is provided to be fulfilled by data valuation functions. There are four important axioms that are commonly agreed to be fair (Covert, Lundberg, and Lee 2021; Ridaoui, Grabisch, and Labreuche 2018; Sim, Xu, and Low 2022): **Linearity**, **Dummy Player**, **Interchangeability** and **Monotonicity** (refer to App. A for their respective definition and connection to fairness). Here, we define *semivalue*, a unique form of data valuation functions that fulfils the four axioms concurrently below[1]:

Definition 1. [Semivalue (Dubey, Neyman, and Weber 1981)] *An n-sources data valuation function $\phi^n : G^n \to \mathbb{R}^n$ is called a **semivalue** if the valuation score $\phi_i^n(v)$ assigned to any data source $d_i \in D$ satisfies*

$$\phi_i^n(v) = \sum_{S\subseteq D\setminus\{d_i\}} w_{|S|} \cdot \mathrm{MaC}_v(d_i|S)/\binom{n-1}{|S|}, \qquad (1)$$

*where $w_{|S|} \geq 0$ is a **weighting term** associated with all coalitions S of size $s = |S|$, satisfying $\sum_{s=0}^{n-1} w_s = 1$; $\mathrm{MaC}_v(d_i|S) = v(S \cup \{d_i\}) - v(S)$ is the **marginal contribution** of data source d_i to coalition S under model utility function v, representing the additional model utility brought by d_i to S measured by v.*

Interpretation. Semivalues can be interpreted as a weighted sum of marginal contribution of d_i to each coalition S. Moreover, since a support set with n data sources has $\binom{n-1}{s}$ coalitions sized s excluding data source d_i, Eq. (1) can be reinterpreted as the expectation of the average marginal contribution of d_i to coalitions sized $0, 1, \cdots, n-1$ over some categorical distribution $W_{\mathbf{s}}$, where $W_{\mathbf{s}}(\mathbf{s} = s) = w_s$. This offers model owners freedom to place more focus on larger or smaller coalitions. For example, Leave-One-Out (LOO) only considers d_i's marginal contribution to the largest possible coalition excluding d_i and sets $w_{n-1} = 1$ and other $w_s = 0$. Beta Shapley (Kwon and Zou 2022) sets $W_{\mathbf{s}}$ to be a beta-binomial distribution with two parameters α and β such that model owners can put more weights on smaller coalitions by setting a larger α and on larger coalitions by setting a larger β.

2.2 Data Deletion and Machine Unlearning

Due to new data protection regulations, model owners must delete data sources' data from their datasets and erase their impact from their trained models upon request. Machine unlearning works (Nguyen et al. 2022b) have studied how to erase data effectively and efficiently and proposed model-agnostic methods (such as decremental learning (Ginart et al. 2019) and differential privacy (Gupta et al. 2021)),

[1] In Sec. 3.1, we will explain how the axioms should be robustified with anticipated data deletions.

model-intrinsic methods (for linear models (Izzo et al. 2021) and Bayesian models (Nguyen, Low, and Jaillet 2020)), and data-driven methods (such as data partitioning (Bourtoule et al. 2021) and data augmentation (Huang et al. 2021)). Our work complements machine unlearning: model owners foresee future data source deletions (and changes in model performance), and thus require a new data valuation approach to value a data source based on its contribution to model performance both before and after anticipated data deletions.

In our problem setting, the main challenge is **how we can adapt and extend the (C1) fairness axioms and (C2) concept of semivalues to cases where data deletion occurs**. Assume that the model owner and data sources in the collaboration decide to use the data valuation function ϕ^n for valuation when there is no deletion. Our solution should be derived from this jointly-agreed semivalue ϕ^n and satisfy some deletion-robustified fairness axioms[2].

3 Methodology

In Sec. 3.1, we define a *random cooperative game* to model the situation where data sources may be deleted and the corresponding deletion-robustified fairness axioms to address (C1). In Sec. 3.2, we describe the *null-player-out* consistency property to extend the jointly-agreed semivalue ϕ^n to the sub-games after data deletions to address (C2). In Sec. 3.3, we define our deletion-robust data valuation technique, DeRDaVa, based on our solutions to (C1) and (C2) and discuss its efficient approximation via sampling. Lastly, in Sec. 3.4, we describe a generalization of DeRDaVa for risk-averse or risk-seeking model owners.

3.1 Random Cooperative Game and Deletion-Robustified Fairness Axioms

Let D_n denote the initial set of n data sources before data deletions. In Sec. 2, we model the problem as a cooperative game $\langle D_n, v\rangle$ with data valuation function ϕ^n. When data deletion occurs, the support set D_n shrinks to a smaller set $D' \subseteq D_n$ but the same model utility function v still maps any subset of the new support set D' to its utility. In this section, we further consider a *random cooperative game* $\langle \mathbf{D}, v\rangle$ to account for deletions. The *random staying set* $\mathbf{D}$ is a subset of D_n and follows some probability distribution $P_{\mathbf{D}}$ (e.g., in Fig. 1, $P_{\mathbf{D}}$ is a categorical distribution with parameters .7 and .3).

In practice, the model owner sets $P_{\mathbf{D}}$ by

- estimating the *independent* probability $\Pr[\mathbb{I}_{d_i} = 1]$ each data source d_i stays in the collaboration from their surveys/histories, where $\mathbb{I}_{d_i}$ is an indicator variable which evaluates to 1 when d_i stays (not delete) and 0 otherwise;
- weighing the emphasis of having only $|D'|$ data sources staying out of D_n (e.g., if the model owner intends to recompute the valuation scores when there are ≥ 30 deletions, it should place higher probability on larger D' with $|D'| > n - 30$);
- using the normalized "reputation" score of data source d_i or subset D' (i.e., how unlikely d_i or D' is malicious and deletion-worthy in upcoming data audits)[3].

[2] App. G.1 discusses why DeRDaVa is superior to the simpler alternative of multiplying the pre-deletion semivalue score of each data source d_i with its staying probability.

Instead of expensively recomputing semivalues every time a deletion happens, we seek a deletion-robust data valuation function τ that acts on the random cooperative game $\langle \mathbf{D}, v\rangle$. The function assigns each data source d_i a fair valuation score $\tau_i(v)$ that accounts for anticipated deletions once/upfront. To ensure the fairness of τ, we examine and "robustify" each of the previously mentioned axioms Linearity, Dummy Player, Interchangeability and Monotonicity with minimal changes such that the robustified versions are desirable in our problem setting.

The Linearity axiom is a very important requirement for any cooperative game and valuation scheme (Shapley 1953) since it provides a way to formally analyze games with linear algebra. Moreover, it ensures that if the marginal contribution of data source d_i to each coalition S doubles, then the valuation score assigned to d_i shall also double; if a data source brings zero marginal contribution to all coalitions, then its valuation score shall be zero. This property is clearly still desirable in our problem setting:

Axiom 1. [Robust Linearity] *Given a random support set $\mathbf{D} \subseteq D_n$ and any two model utility functions v and w, a fair data valuation function τ shall satisfy*

$$\forall d_i \in D_n \quad [\tau_i(v) + \tau_i(w) = \tau_i(v + w)]. \tag{2}$$

The Dummy Player axiom defines a specific type of data source called *dummy player*, whose marginal contribution is always equal to its own utility (i.e., $\forall S \subseteq D_n \setminus \{d_i\} \quad [\mathrm{MaC}_v(d_i|S) = v(\{d_i\})]$). The axiom states that the valuation score assigned to a dummy player shall be equal to its own utility since its marginal contribution is equal to its own utility in all cases. However, in our problem setting, consider two dummy players $\mathtt{DP}_i$ and $\mathtt{DP}_j$ with equal own utility, where $\mathtt{DP}_i$ always stays in the collaboration while $\mathtt{DP}_j$ stays or deletes with a 50-50 chance. Although their contributions to pre-deletion model performance are the same, $\mathtt{DP}_i$ contributes more to model performance after anticipated data deletions (i.e., deletion-robustness) than $\mathtt{DP}_j$. Therefore, the original Dummy Player axiom is no longer desirable. Instead, our data valuation function should only reward a dummy player for cases where it stays in $\mathbf{D}$:

Axiom 2. [Robust Dummy Player] *A data source* $\mathtt{DP}$ *is called a* ***dummy player*** *if its marginal contribution is always equal to its own utility. For any dummy player* $\mathtt{DP}$*, a fair data valuation function τ shall satisfy*

$$\tau_{\mathtt{DP}}(v) = \mathbb{E}_{\mathbf{D}\sim P_{\mathbf{D}}}\left[v(\{\mathtt{DP}\}) \cdot \mathbb{I}[\mathtt{DP} \in \mathbf{D}]\right], \tag{3}$$

where $\mathbb{I}[\mathtt{DP} \in \mathbf{D}]$ is an indicator variable that equates to 1 when $\mathtt{DP}$ *is present in $\mathbf{D}$ and vice versa.*

The Interchangeability axiom defines that two data sources are *interchangeable* if their marginal contributions to any coalition S are always equal. It states that two interchangeable data sources shall receive the same valuation score. However, in a random cooperative game, two interchangeable data sources that have different probabilities

[3] App. G.2 further discusses how to set $P_{\mathbf{D}}$.

of staying contribute differently to deletion-robustness, and their valuation scores should not be equal (e.g., ★ and ■ in Fig. 1). Therefore, we add a further constraint to this axiom:

Axiom 3. [Robust Interchangeability] *Two data sources d_i and d_j are said to be **robustly interchangeable** ($d_i \cong d_j$) if their marginal contribution to any coalition $S \subseteq D_n$ is always equal and their probability of staying alongside others sources $D' \subseteq D_n \setminus \{d_i, d_j\}$ are also equal. The valuation scores assigned to any two robustly interchangeable data sources shall be equal:*

$$d_i \cong d_j \Rightarrow \tau_i(v) = \tau_j(v). \tag{4}$$

Finally, the Monotonicity axiom states that if every data source makes a non-negative marginal contribution to every coalition (i.e., model utility function v is monotone increasing), then their valuation scores shall be non-negative. This is still valid in a random cooperative game because if v is monotone increasing, then the marginal contribution of any data source is at least 0 even if it quits the collaboration. Therefore, we keep the original version of this axiom:

Axiom 4. [Robust Monotonicity] *If model utility function v is monotone increasing, then the valuation score assigned to any data source shall be non-negative:*

$$\begin{aligned} &\forall S, T \subseteq D_n \quad [S \subseteq T \Rightarrow v(S) \leq v(T)] \\ &\Rightarrow \forall d_i \in D_n \quad [\tau_i(v) \geq 0]. \end{aligned} \tag{5}$$

With the formalization of Axioms 1 to 4, we proceed to find a solution that satisfies these axioms.

3.2 Null-Player-Out (NPO) Consistency and Extension

After data deletions, the number of data sources will be $< n$. Thus, we can no longer directly apply the n-sources data valuation function ϕ^n. Instead, we need to derive a sequence of semivalues $\Phi = \langle \phi^k : k = 1, 2, \cdots, n \rangle$ to value every game with support set sized k ranging from 1 to n. We address this challenge by considering a post-deletion model utility function ν which maps each coalition of data sources to the model utility of the staying subset of D', e.g., $\nu(D_n) = v(D')$. In the cooperative game $\langle D_n, \nu \rangle$, any deleted data source is a *null player* (van den Brink 2007):

Definition 2. [Null player] *A data source in a cooperative game $\langle D, \nu \rangle$ is said to be a **null player** ($\texttt{NP}$) if its marginal contribution to every coalition S is always equal to zero, i.e.,*

$$\forall S \subseteq D \setminus \{\texttt{NP}\} \quad [\mathrm{MaC}_\nu(\texttt{NP}|S) = 0]. \tag{6}$$

A null player $\texttt{NP}$, e.g., an empty data source, should be assigned a valuation score $\phi^n_{\texttt{NP}}(\nu)$ of zero.

Consider the case where some data sources have quit D_n and only a subset $D' \subset D_n$ stays as the support set. Intuitively, the value assigned by ϕ^n to an undeleted data source d_i in the cooperative game $\langle D_n, \nu \rangle$ (using the post-deletion model utility function) should be the same as its value assigned by $\phi^{|D'|}$ in the cooperative game $\langle D', v \rangle$ (as though $D_n \setminus D'$ never joined the collaboration) (‡). This condition requires us to select the sequence of semivalues $\Phi = \langle \phi^k : k = 1, \cdots, n-1 \rangle$ to be *NPO-consistent*:

Definition 3. [NPO-consistency] *Consider a set of data sources $D_n = \bigcup_{i=1}^n \{d_i\}$ and a natural number $k \leq n$ such that only data sources in $D_k = \{d_i : 1 \leq i \leq k\}$ are **non-null players** in the cooperative game $\langle D_n, \nu \rangle$. Two semivalues $\phi^n : G^n \to \mathbb{R}^n$ and $\phi^k : G^k \to \mathbb{R}^k$ are **NPO-consistent** if the presence of null players (e.g., empty data sources) do not affect the values of the non-null players (who contribute valid datasets):*

$$\forall d_i \in D_k \quad [\phi^n_i(\nu) = \phi^k_i(\nu)] \tag{7}$$

holds for all such D_n and D_k with fixed n and k. Moreover, a sequence of semivalues Φ is NPO-consistent if every pair of semivalues in Φ is NPO-consistent.

Note that the null players $D_n \setminus D_k$ in $\langle D_n, \nu \rangle$ correspond to deleted data sources. As data sources in D_k stays undeleted, $v(S) = \nu(S)$ for all $S \subseteq D_k$ and the values $\phi^k_i(\nu) = \phi^k_i(v)$. The NPO-consistent property guarantees that $\phi^n_i(\nu)$ equals $\phi^k_i(v)$, thus, achieving (‡).

We then construct a sequence of semivalues $\Phi = \langle \phi^k : k = 1, 2, \cdots, n \rangle$ that is NPO-consistent with the following NPO-extension process:

Theorem 1. [NPO-extension] *Every semivalue $\phi^n : G^n \to \mathbb{R}^n$ can be **uniquely** extended to a sequence of semivalues $\Phi = \langle \phi^k : k = 1, 2, \cdots, n \rangle$ that is NPO-consistent through the following unified **NPO-extension** process:*

1. *From the weighting term w_s in Definition 1, calculate the quantity $w^n_s = w_s / \binom{n-1}{s}$, which is sometimes referred to as **weighting coefficient** (Carreras and Freixas 2000).*
2. *Calculate the weighting coefficients w^{n-1}_s of ϕ^{n-1} using the formula $w^{n-1}_s = w^n_s + w^n_{s+1}$. We can therefore construct ϕ^{n-1} by setting the weighting term w_s in ϕ^{n-1} to be $w^{n-1}_s \cdot \binom{n-2}{s}$ for each $s = 0, 1, \cdots, n-2$.*
3. *Repeat Steps 1 and 2 until we have constructed every semivalue in Φ.*

Intuition behind NPO-extension. Consider the cooperative game $\langle D, \nu \rangle$ with $|D| = n$ with one null player $\texttt{NP}$. The marginal contribution of any non-null data source $d_i \neq \texttt{NP}$ to any coalition S without the null player (i.e., $\texttt{NP} \notin S$) is always equal to its marginal contribution to $S \cup \{\texttt{NP}\}$ (i.e., $\mathrm{MaC}_\nu(d_i|S) = \mathrm{MaC}_\nu(d_i|S \cup \{\texttt{NP}\})$. For ϕ^{n-1} to be NPO-consistent with ϕ^n, the total weights on $\mathrm{MaC}_\nu(d_i|S)$ must be equal in ϕ^{n-1} and ϕ^n. Hence, the weighting coefficient w^{n-1}_s must equal the sum of the weighting coefficients of the above two marginal contribution terms ($w^n_s + w^n_{s+1}$) (Domenech, Giménez, and Puente 2016). In App. B, we formally prove the validity and uniqueness of the result in Theorem 1 and the NPO-consistent property of Φ defined using common semivalues such as Data Shapley (Ghorbani and Zou 2019; Jia et al. 2019), Beta Shapley (Kwon and Zou 2022) and Data Banzhaf (Wang and Jia 2023).

3.3 DeRDaVa and Its Efficient Approximation

Let $\Phi = \{\phi^k : k = 1, 2, \cdots n\}$ be the sequence of semivalues derived from ϕ^n using NPO-extension. Each data source's contribution to model performance and deletion-robustness can be regarded as an aggregate of its contribution (measured by $\phi^{|D'|} \in \Phi$) to every possible staying set $D' \subseteq D_n$. Hence, we take the expectation of valuation scores $\phi_i^{|\mathbf{D}|}(v)$ over the probability distribution $P_{\mathbf{D}}$ and regard d_i as making 0 contribution when $d_i \notin D'$. This leads to the formal definition of DeRDaVa:

Definition 4. [DeRDaVa] *Let $\langle \mathbf{D}, v\rangle$ be a random cooperative game with random support set $\mathbf{D}$ over some probability distribution $P_{\mathbf{D}}$ where the maximal support set D_n contains n data sources. Suppose also that $\phi^n : G^n \to \mathbb{R}^n$ is the jointly-agreed semivalue and $\Phi = \{\phi^k : k = 1, 2, \cdots n\}$ is the sequence of semivalues derived from ϕ^n using NPO-extension. The* ***DeRDaVa score with ϕ^n prior*** *assigned to data source d_i, $\tau_i(v)$, is given by*

$$\begin{aligned}\tau_i(v) &= \mathbb{E}_{\mathbf{D}\sim P_{\mathbf{D}}}\left[\mathbb{I}[d_i \in \mathbf{D}] \cdot \phi_i^{|\mathbf{D}|}(v)\right] \\ &= \sum_{D' \subseteq D_n} \Bigg(P_{\mathbf{D}}(\mathbf{D} = D') \cdot \mathbb{I}[d_i \in D'] \cdot \\ &\qquad \sum_{S \subseteq D' \setminus \{d_i\}} w_{|S|}^{|D'|} \cdot \mathrm{MaC}_v(d_i|S) \Bigg), \end{aligned} \tag{8}$$

where $\mathbb{I}[d_i \in \mathbf{D}]$ is an indicator variable that equates to 1 only when d_i stays present, $\phi^{|\mathbf{D}|} \in \Phi$ is the semivalue from NPO-extension, and $w_{|S|}^{|D'|}$ is the weighting coefficient to coalition S in $\phi^{|D'|}$ defined in Theorem 1.

The fairness and uniqueness of DeRDaVa are guaranteed with the following theorem:

Theorem 2. [Fairness and uniqueness of DeRDaVa] *Given a random cooperative game $\langle \mathbf{D}, v\rangle$ with the same notations $P_{\mathbf{D}}$, D_n, $n = |D_n|$ and $\phi^n : G^n \to \mathbb{R}^n$ as in Definition 4, the DeRDaVa function τ defined in Definition 4 uniquely satisfies Axioms 1, 2, 3 and 4 defined in Sec. 3.1.*

In App. C, we prove the (F) fairness and (U) uniqueness of DeRDaVa. (F) involves verifying that DeRDaVa satisfies our four robustified axioms. Let $V(\cdot)$ denote the *random utility function* that maps each coalition $S \subseteq D_n$ to the random utility after deletions, i.e., $V(S) = v(S \cap \mathbf{D})$. (U) involves identifying the *static dual* game $\langle D_n, \mathbb{E}[V(\cdot)]\rangle$ to our random cooperative game $\langle \mathbf{D}, v\rangle$ and proving that the original axioms of semivalues for the static dual game are equivalent to the four robustified axioms for any random cooperative game. The uniqueness of DeRDaVa then follows from the uniqueness of semivalues.

From Eq. (8), we need to consider every possible pair $\langle S, D'\rangle$ of subset S and staying set D' such that $S \cup \{d_i\} \subseteq D' \subseteq D_n$. Each data source $d_j \neq d_i$ has exactly three *states*: (i) It is in neither S nor D' (State 0); (ii) It is in D' but not in S (State 1); (iii) It is in both S and D' (State 2). Thus, the total number of unique state combinations or pairs $\langle S, D'\rangle$ needed to exactly compute DeRDaVa scores is $O(3^n)$. Exact computation is intractable when the number of data sources is large. Thus, model owners should efficiently approximate DeRDaVa scores based on Monte-Carlo sampling and additionally use our **012-MCMC algorithm** when it is hard to estimate/sample from $P_{\mathbf{D}}$.

Monte-Carlo Sampling DeRDaVa scores can be alternatively viewed as the expectation of marginal contribution $\mathrm{MaC}_v(d_i|S)$ over some distribution of staying set D' (i.e., $P_{\mathbf{D}}$) and coalition S. Therefore, it is natural to use Monte-Carlo sampling when it is tractable to sample from $P_{\mathbf{D}}$ directly. For example, for the special case where data sources decide to stay/delete independently, we sample whether each data source stays to determine staying set D', the size s of coalition S (using the weighting coefficients) and lastly, s data sources out of D'. In App. D.1, we prove that the number of samples needed to approximate DeRDaVa with (δ, ϵ)-error is $O(\frac{2r^2 n}{\epsilon^2} \log \frac{2n}{\delta})$, where r is the range of model uility function v.

012-MCMC When direct Monte-Carlo sampling is hard, we propose to repeatedly sample the state for each source in $D_n \setminus \{d_i\}$ from the uniform distribution over $\{0, 1, 2\}$. For the t-th sample, we construct $S_{(t)} = \{d_j : \mathtt{state}(d_j) = 2\}$ and $D'_{(t)} = \{d_j : \mathtt{state}(d_j) \neq 0\} \cup \{d_i\}$. Thus, we enforce $S_{(t)} \cup \{d_i\} \subseteq D'_{(t)}$. The DeRDaVa score of source d_i, $\tau_i(v)$, is then approximated by importance sampling and taking the average of T samples:

$$\frac{1}{T}\sum_{t=1}^{T}\left(\frac{P_{\mathbf{D}}\big(\mathbf{D}=D'_{(t)}\big)}{1/3^{(n-1)}} \cdot w_{|S_{(t)}|}^{|D'_{(t)}|} \cdot \mathrm{MaC}_v(d_i|S_{(t)})\right) . \tag{9}$$

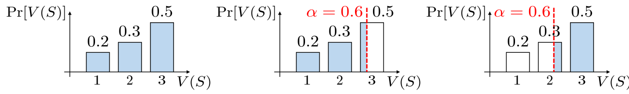
By using importance sampling, we avert computing $P_{\mathbf{D}}$ for every subset. Instead, we use the known probability $P_{\mathbf{D}}\left(\mathbf{D} = D'_{(t)}\right)$ for each sampled staying set $D'_{(t)}$.

In practice, we construct M parallel Markov chains of samples and use the Gelman-Rubin statistic (Gelman and Rubin 1992) to assess the convergence of the approximation. The threshold for the Gelman-Rubin statistic is usually set around 1.1 (Vats and Knudson 2021), but we set it to ≤ 1.005 for higher approximation precision. The time complexity of our 012-MCMC algorithm depends on the number of samples generated which is significantly smaller than $O(3^n)$. The justification and pseudocode for 012-MCMC algorithm are included in App. D.2.

3.4 Risk-DeRDaVa: A Variant for Different Risk Attitudes

In Sec. 3.3, the DeRDaVa scores are equivalent to the values assigned by ϕ^n on the static dual game $\langle D_n, \mathbb{E}[V(\cdot)]\rangle$. The model owner considers each data source's marginal contribution to the **expected** utility of each coalition S. The model owner is *risk-neutral* and indifferent between (R1) a constant random utility function $V(\cdot)$ or (R2) a varying $V(\cdot)$ with a worse worst-case (with equal expected values).

In practice, model owners may strictly prefer R1 or R2. *Risk-averse* model owners would prefer R1 with a higher worst-case model utility and thus highly value data sources that have higher marginal contributions to the worst-case model utility. In contrast, *risk-seeking* model owners would prefer R2 with a higher best-case model utility and thus value those data sources that have higher marginal contributions to the best-case model utility. In Fig. 2, we illustrate and contrast how the risk-neutral, risk-averse and



(a) Risk-neutral. (b) Risk-averse. (c) Risk-seeking.

Figure 2: Model owners with different risk attitudes will map the random utility function $V(S)$ evaluated at coalition S to a deterministic value differently. The risk-neutral owner (2a) takes expectation (blue) over all possible utilities. A risk-averse (2b)/ risk-seeking (2c) owner takes expectation over the lower/worst 0.6-tail and upper/best 0.6-tail respectively.

risk-seeking model owners will transform the random utility function $V(\cdot)$ to a static dual game.

To quantify risks, we use discrete Conditional Value-at-Risk (CVaR) (Uryasev et al. 2010) to define Coalitional Conditional Value-at-Risk (C-CVaR) in our problem setting. There are two types of C-CVaR, namely *Risk-Averse C-CVaR* (C-CVaR$^-$) and *Risk-Seeking C-CVaR* (C-CVaR$^+$), which corresponds to the expectation within the lower tail (Fig. 2b) and upper tail (Fig. 2c) respectively. For example, consider the $V(S)$ given in Fig. 2b. The C-CVaR$^-$ at level $\alpha = 0.6$ is the expectation of $V(S)$ in the blue shaded region (i.e., the lower 60% tail): C-CVaR$^-_{0.6}[V(S)] = \frac{0.2}{0.6} \times 1 + \frac{0.3}{0.6} \times 2 + \frac{0.1}{0.6} \times 3 = \frac{11}{6}$. A formal definition of C-CVaR$^-$ and C-CVaR$^+$ is included in App. E. Therefore, to provide a suitable solution for risk-averse/seeking model owners, we consider the static prior game $\langle D_n, \text{C-CVaR}^{\mp}_\alpha[V(\cdot)]\rangle$ whose data valuation function is ϕ^n and define Risk-DeRDaVa:

Definition 5. [Risk-DeRDaVa] *Given a random cooperative game $\langle \mathbf{D}, v\rangle$ with the same notations $P_{\mathbf{D}}$, D_n, $n = |D_n|$ and $\phi^n : G^n \to \mathbb{R}^n$ as in Definition 4, first define for any coalition $S \subseteq D_n$ the **random utility function** $V(S) = v(S \cap \mathbf{D})$. Let $v_{\text{risk}}(S) = \text{C-CVaR}^{\mp}_\alpha[V(S)]$. The **Risk-DeRDaVa score with ϕ^n prior at level α** for risk averse/seeking model owners is defined as*

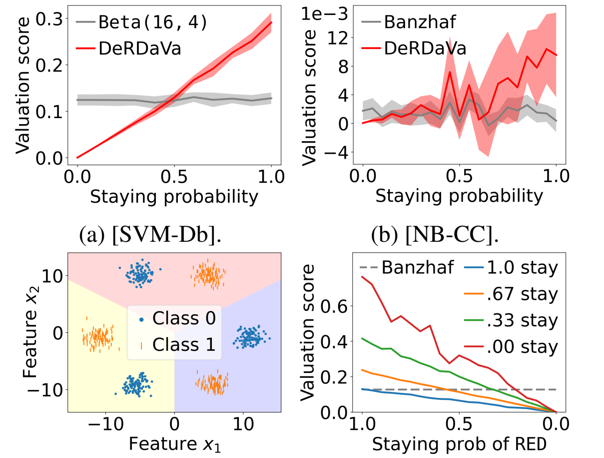
$$\rho_i(v) = \sum_{S \subseteq D_n \setminus \{d_i\}} w_{|S|} \cdot \text{MaC}_{v_{\text{risk}}}(d_i|S) / \binom{n-1}{s}, \quad (10)$$

where $w_{|S|}$ is the weighting term associated with all coalitions S of size $s = |S|$ given by ϕ^n.

Note that $\alpha = 1$ recovers the DeRDaVa scores. In practice, it is more common for model owners to be risk-averse, so we default Risk-DeRDaVa to refer to the risk-averse version. As C-CVaR is non-additive (see App. E), we approximate Risk-DeRDaVa scores by sampling S and using the Monte-Carlo CVaR algorithm (Hong, Hu, and Liu 2014).

4 Experiments

Our experiments use the following [model-dataset] combinations: [NB-CC] Naive Bayes trained on Credit Card (Yeh and Lien 2009), [NB-Db] Naive Bayes trained on Diabetes (Carrion, Dustin 2022), [NB-Wd] Naive Bayes trained on Wind (Vanschoren, Joaquin 2014), [SVM-Db] Support Vector Machine trained on Diabetes, and [LR-Pm] Logistic Regression trained on Phoneme (Grin, Leo 2022). More experimental details are included in App. F.



(a) [SVM-Db]. (b) [NB-CC]. (c) Synthetic dataset. (d) Data similarity.

Figure 3: DeRDaVa accounts for data deletions. (3a) (11 data sources) and (3b) (21 data sources) show the effect of staying probability on DeRDaVa scores with Beta Shapley and Data Banzhaf prior; (3c) and (3d) show when DeRDaVa score of a redundant data source exceeds its Banzhaf score.

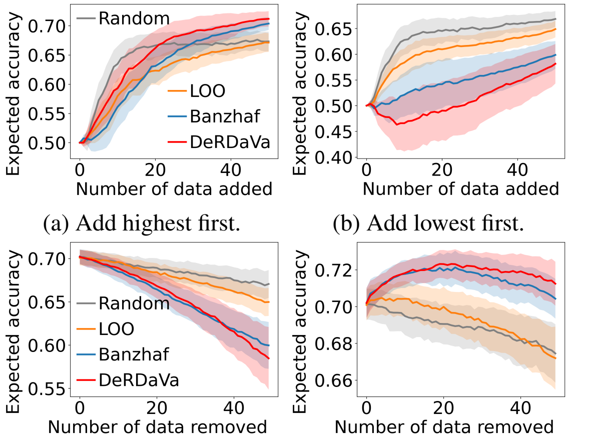
4.1 Measure of Contribution to Model Performance and Deletion-Robustness

DeRDaVa is designed to measure each data source's contribution to both model performance and deletion-robustness. By creating data sources with different contributions and comparing their DeRDaVa scores, we can verify the empirical behaviour of DeRDaVa. We analyse the three main factors that affect the contribution of data sources (**staying probability**, **data similarity** and **data quality**) below.

Staying Probability We repeat 50 runs of creating data sources with equal number of randomly sampled training examples, assigning different independent staying probabilities and computing their semivalue and corresponding DeRDaVa scores. From Fig. 3a and 3b, we observe that data sources with higher staying probability receive higher DeRDaVa scores as they contribute more to model performance after anticipated deletions.

Data Similarity We create a synthetic dataset (Fig. 3c) with 4 data sources. The yellow and blue regions are exclusively owned by 2 different data sources while the red region is co-owned by 2 data sources `RED` and `REDD`. Thus, `RED` and `REDD` data are highly similar. The model utility function is the accuracy of the trained k-Nearest Neighbours model. In Fig. 3d, we observe that the `RED` is assigned a higher DeRDaVa score (plotted as solid lines) than Banzhaf score (plotted as a dashed line) when its staying probability is high and when other data sources do not stay with certainty. This aligns with our intuition that deletion-robust data valuation should favour `RED`, despite its redundancy in the presence of `REDD`, when `RED` is more likely to stay than others.

Data Quality Data sources with poor data quality (e.g., with high noise level) make a low contribution to model performance regardless of data deletions. Similar to semivalues,



(a) Add highest first. (b) Add lowest first.

(c) Remove highest first. (d) Remove lowest first.

Figure 4: Point addition and removal experiments. All experiments are run using [NB-Wd], 100 data sources and Data Banzhaf prior.

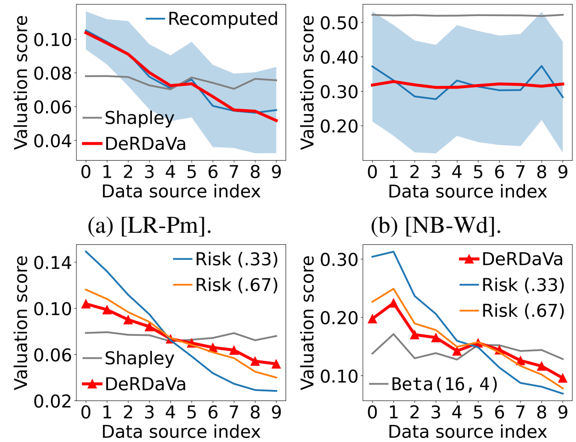
DeRDaVa is also capable of reflecting data quality and thus can be applied to identify noisy data (see App. F).

4.2 Point Addition and Removal

We perform point addition and removal experiments which are often used in evaluation of data valuation techniques (Ghorbani and Zou 2019; Kwon and Zou 2022) with an adaption to our setting — we measure the **expected** model performance after data deletion instead. When data with the highest scores are added first (Fig. 4a), `Random` shows a rapid increase in expected model performance at the beginning as the training curve has not plateaued and almost every added point contributes a lot. However, after more additions, DeRDaVa with Banzhaf prior surpasses all others as its selected points contribute to preserving high model accuracy after anticipated deletions. When data with the lowest scores are added first (Fig. 4b), DeRDaVa's expected accuracy drops since these data are harmful to both model performance and deletion-robustness. When data with the highest scores are removed first (Fig. 4c), DeRDaVa exhibits a rapid decrease in expected model performance. This is because data sources that contribute more to deletion-robustness are removed. When data with the lowest scores are removed first (Fig. 4d), DeRDaVa demonstrates a rapid increase in expected model performance at the beginning and the slowest decrease later. This is because data which contributes to preserving higher model accuracy after anticipated deletions tend to have higher DeRDaVa scores and are not removed.

4.3 Reflection of Long-Term Contribution

Next, we simulate data deletions and recompute the semivalue scores to see how the contribution of a data source changes as data deletion occurs. We then compare these recomputed scores with the pre-deletion semivalue scores and DeRDaVa scores to investigate which represent the long-term contribution better. Fig. 5a and 5b show that the average of the distribution of recomputed valuation scores is almost the same as the DeRDaVa scores but deviate significantly from the pre-deletion scores. Moreover, the recomputed semivalues can vary widely (see shaded region) with different deletion outcomes. This aligns with our motivation to avert uncertainty and fluctuations in the valuation by efficiently computing DeRDaVa scores upfront.



(a) [LR-Pm]. (b) [NB-Wd].

(c) [LR-Pm]. (d) [NB-Db].

Figure 5: When data sources stay with independent (5a) and dependent (5b) probabilities, the recomputed semivalue scores of 10 data sources always converge to DeRDaVa scores and deviate from pre-deletion scores; (5c) and (5d) compare Risk-DeRDaVa with DeRDaVa and semivalues.

4.4 Empirical Behaviours of Risk-DeRDaVa

In this section, we observe the empirical behaviours of Risk-DeRDaVa and investigate how the valuation scores change as we use C-CVaR$^-$ at different levels α, which reflects model owners with different risk attitudes. We assign a predetermined independent staying probability to each data source, where data sources with smaller indices have higher staying probability. As shown in Fig. 5c and 5d, Risk-DeRDaVa (risk-averse) assigns even higher scores to data sources with high staying probability and penalizes data sources that are likely to delete harder.

5 Conclusion and Discussion

In this paper, we propose a deletion-robust data valuation technique DeRDaVa and an efficient approximation algorithm to improve its practicality. We also introduce Risk-DeRDaVa for model owners with different risk attitudes. We have shown both theoretically and empirically that our proposed solutions have more desirable properties than existing works when data deletion occurs. Future work can consider other possible applications (e.g., heuristics of active learning) and address the limitations and negative social impacts raised in App. G.3 such as approximating DeRDaVa scores more efficiently with guarantees, estimating the staying probabilities $P_{\mathbf{D}}$ more accurately and preventing intentional misreporting of staying probabilities $P_{\mathbf{D}}$ or data.

Acknowledgments

This research/project is supported by the National Research Foundation Singapore and DSO National Laboratories under the AI Singapore Programme (AISG Award No: AISG2-RP-2020-018). The author would like to extend special thanks to Dr. Henry Wai Kit Chia for his valuable comments and suggestions on this research.

References

Bourtoule, L.; Chandrasekaran, V.; Choquette-Choo, C. A.; Jia, H.; Travers, A.; Zhang, B.; Lie, D.; and Papernot, N. 2021. Machine Unlearning. In *Proc. IEEE Symposium on Security and Privacy (SP)*, 141–159.

Carreras, F.; and Freixas, J. 2000. A Note on Regular Semivalues. *International Game Theory Review*, 2(04): 345–352.

Carrion, Dustin. 2022. Pima Indians Diabetes. https://www.openml.org/search?type=data&status=active&id=43582. Accessed: 2022-10-30.

Chen, M.; Zhang, Z.; Wang, T.; Backes, M.; Humbert, M.; and Zhang, Y. 2021. When Machine Unlearning Jeopardizes Privacy. In *Proc. ACM SIGSAC Conference on Computer and Communications Security*, 896–911.

Covert, I. C.; Lundberg, S.; and Lee, S.-I. 2021. Explaining by Removing: A Unified Framework for Model Explanation. *Journal of Machine Learning Research*, 22(1): 9477–9566.

Domenech, M.; Giménez, J. M.; and Puente, M. A. 2016. Some Properties for Probabilistic and Multinomial (Probabilistic) Values on Cooperative Games. *Optimization*, 65(7): 1377–1395.

Dubey, P.; Neyman, A.; and Weber, R. J. 1981. Value Theory Without Efficiency. *Mathematics of Operations Research*, 6(1): 122–128.

Fan, Z.; Fang, H.; Zhou, Z.; Pei, J.; Friedlander, M. P.; Liu, C.; and Zhang, Y. 2022. Improving Fairness for Data Valuation in Horizontal Federated Learning. In *Proc. IEEE International Conference on Data Engineering (ICDE)*, 2440–2453.

Gelman, A.; and Rubin, D. B. 1992. Inference From Iterative Simulation Using Multiple Sequences. *Statistical Science*, 457–472.

Ghorbani, A.; and Zou, J. 2019. Data Shapley: Equitable Valuation of Data for Machine Learning. In *Proc. ICML*, 2242–2251.

Ginart, A.; Guan, M.; Valiant, G.; and Zou, J. Y. 2019. Making AI Forget You: Data Deletion in Machine Learning. In *Proc. NeurIPS*, 3513–3526.

Grin, Leo. 2022. Phoneme. https://www.openml.org/search?type=data&status=active&id=43973. Accessed: 2022-10-30.

Gupta, V.; Jung, C.; Neel, S.; Roth, A.; Sharifi-Malvajerdi, S.; and Waites, C. 2021. Adaptive Machine Unlearning. In *Proc. NeurIPS*, 16319–16330.

Hong, L. J.; Hu, Z.; and Liu, G. 2014. Monte Carlo Methods for Value-at-Risk and Conditional Value-at-Risk: A Review. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 24(4): 1–37.

Huang, H.; Ma, X.; Erfani, S. M.; Bailey, J.; and Wang, Y. 2021. Unlearnable Examples: Making Personal Data Unexploitable. In *Proc. ICLR*.

Izzo, Z.; Smart, M. A.; Chaudhuri, K.; and Zou, J. 2021. Approximate Data Deletion From Machine Learning models. In *Proc. AISTATS*, 2008–2016.

Jia, R.; Dao, D.; Wang, B.; Hubis, F. A.; Hynes, N.; Gürel, N. M.; Li, B.; Zhang, C.; Song, D.; and Spanos, C. J. 2019. Towards Efficient Data Valuation Based on the Shapley Value. In *Proc. AISTATS*, 1167–1176.

Kwon, Y.; and Zou, J. 2022. Beta Shapley: A Unified and Noise-Reduced Data Valuation Framework for Machine Learning. In *Proc. AISTATS*, 8780–8802.

Magdziarczyk, M. 2019. Right to Be Forgotten in Light of Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons With Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC. In *6th International Multidisciplinary Scientific Conference on Social Sciences and Art*, 177–184.

Nguyen, D. C.; Pham, Q.-V.; Pathirana, P. N.; Ding, M.; Seneviratne, A.; Lin, Z.; Dobre, O.; and Hwang, W.-J. 2022a. Federated Learning for Smart Healthcare: A Survey. *ACM Computing Surveys (CSUR)*, 55(3): 1–37.

Nguyen, Q. P.; Low, B. K. H.; and Jaillet, P. 2020. Variational Bayesian Unlearning. In *Proc. NeurIPS*, 16025–16036.

Nguyen, T. T.; Huynh, T. T.; Nguyen, P. L.; Liew, A. W.-C.; Yin, H.; and Nguyen, Q. V. H. 2022b. A Survey of Machine Unlearning. arXiv:2209.02299.

Ong, E.-I. 2018. Data Protection in the Internet: National Rapporteur (Singapore). In *Congress of the International Academy of Comparative Law 20th IACL*.

Ridaoui, M.; Grabisch, M.; and Labreuche, C. 2018. An Axiomatisation of the Banzhaf Value and Interaction Index for Multichoice Games. In *Proc. MDAI*, 143–155. Springer.

Sekhari, A.; Acharya, J.; Kamath, G.; and Suresh, A. T. 2021. Remember What You Want to Forget: Algorithms for Machine Unlearning. In *Proc. NeurIPS*, 18075–18086.

Shapley, L. S. 1953. A Value for n-person Games. *Contributions to the Theory of Games*, 2: 307–317.

Shastri, S.; Wasserman, M.; and Chidambaram, V. 2019. The Seven Sins of Personal-Data Processing Systems Under GDPR. In *11th USENIX Workshop on Hot Topics in Cloud Computing (HotCloud 19)*.

Sim, R. H. L.; Xu, X.; and Low, B. K. H. 2022. Data Valuation in Machine Learning: "ingredients", strategies, and open challenges. In *Proc. IJCAI*.

Sim, R. H. L.; Zhang, Y.; Chan, M. C.; and Low, B. K. H. 2020. Collaborative Machine Learning With Incentive-Aware Model Rewards. In *Proc. ICML*, 8927–8936. PMLR.

Tay, S. S.; Xu, X.; Foo, C. S.; and Low, B. K. H. 2022. Incentivizing Collaboration in Machine Learning via Synthetic Data Rewards. In *Proc. AAAI*, volume 36, 9448–9456.

TCFOD. 2019. Sharing and Utilizing Health Data for AI Applications. https://www.hhs.gov/sites/default/files/sharing-and-utilizing-health-data-for-ai-applications.pdf. Accessed: 2022-12-02.

Tsai, C.-F.; and Chen, M.-L. 2010. Credit Rating by Hybrid Machine Learning Techniques. *Applied Soft Computing*, 10(2): 374–380.

Uryasev, S.; Sarykalin, S.; Serraino, G.; and Kalinchenko, K. 2010. VaR vs CVaR in Risk Management and Optimization. In *Proc. CARISMA*.

van den Brink, R. 2007. Null or Nullifying Players: The Difference Between the Shapley Value and Equal Division Solutions. *Journal of Economic Theory*, 136(1): 767–775.

Vanschoren, Joaquin. 2014. Wind. https://www.openml.org/search?type=data&sort=runs&id=847&status=active. Accessed: 2022-10-30.

Vats, D.; and Knudson, C. 2021. Revisiting the Gelman–Rubin Diagnostic. *Statistical Science*, 36(4): 518–529.

Wang, J. T.; and Jia, R. 2023. Data Banzhaf: A Robust Data Valuation Framework for Machine Learning. In *Proc. AISTATS*, 6388–6421.

Xu, X.; Lyu, L.; Ma, X.; Miao, C.; Foo, C. S.; and Low, B. K. H. 2021. Gradient Driven Rewards to Guarantee Fairness in Collaborative Machine Learning. In *Proc. NeurIPS*, 16104–16117.

Yang, Q.; Liu, Y.; Chen, T.; and Tong, Y. 2019. Federated Machine Learning: Concept and Applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2): 1–19.

Yeh, I.-C.; and Lien, C.-h. 2009. The Comparisons of Data Mining Techniques for the Predictive Accuracy of Probability of Default of Credit Card Clients. *Expert systems with applications*, 36(2): 2473–2480.

Zhang, J.; Wu, Y.; and Pan, R. 2021. Incentive Mechanism for Horizontal Federated Learning Based on Reputation and Reverse Auction. In *Proc. ACM Web Conference*, 947–956.