UxV-DPN: Utility-vs-Value Data Pricing and Negotiation Mechanism in Machine Learning Data Marketplace

Hajar Baghcheband
Faculty of Engineering, University of Porto
Porto, Portugal
Artificial Intelligence & Computer Science Laboratory- LIACC
Porto, Portugal
h.baghcheband@fe.up.pt

Carlos Soares
Faculty of Engineering, University of Porto
Porto, Portugal
Artificial Intelligence & Computer Science Laboratory- LIACC
Porto, Portugal
Fraunhofer AICOS Portugal
Porto, Portugal
csoares@fe.up.pt

Luis Paulo Reis
Faculty of Engineering, University of Porto
Porto, Portugal
Artificial Intelligence & Computer Science Laboratory- LIACC
Porto, Portugal
lpreis@fe.up.pt

ABSTRACT

As data becomes a central asset in the AI-driven economy, data marketplaces have emerged to support efficient, utility-aware exchanges between data providers and consumers. However, traditional models such as fixed pricing or auctions often fail to account for evolving budgets, asymmetric incentives, and predictive utility, resulting in suboptimal transactions. This paper introduces UxV-DPN (Utility-vs-Value Data Pricing and Negotiation), a novel negotiation-based pricing protocol designed to align the buyer's learning utility with the seller's value expectations through dynamic, utility-informed pricing. We evaluate UxV-DPN against baseline strategies, including Simple negotiation, Single-Agent learning (centralised), and No-Exchange, across multiple datasets in a distributed learning environment. Experimental results indicate that both data exchanges improve predictive performance over No-Exchange, while UxV-DPN reduces buyer costs compared to the Simple negotiation strategy, offering over 80% of its accuracy at nearly half the price. This demonstrates that UxV-DPN delivers a more cost-effective negotiation than Simple. These findings underscore the significance of structured negotiation strategies in enhancing economic efficiency and model accuracy in distributed data exchange.

PVLDB Workshop Format:
Hajar Baghcheband, Carlos Soares, and Luis Paulo Reis. UxV-DPN: Utility-vs-Value Data Pricing and Negotiation Mechanism in Machine Learning Data Marketplace. VLDB 2025 Workshop: 3rd Data EConomy Workshop (DEC).

1 INTRODUCTION

Data Economy (DE) has become a crucial component of modern digital infrastructure, with data recognised not only as a technological resource but also as an essential economic asset. Harnessing its potential and aligning producer and consumer incentives requires organised systems for data exchange, valuation, and monetisation. Consequently, organisations need efficient, reliable market-based mechanisms for data sharing and trading.

Data marketplaces have gained traction as platforms that facilitate the discovery, exchange, and monetisation of data assets between different entities [1, 15, 21]. These platforms function as intermediaries between data providers and consumers [10], facilitating transactions while ensuring and providing data valuation [11, 24], pricing [19, 28], and, in some cases, analytics services and privacy-preserving mechanisms [14, 16]. Data marketplaces provide structured and scalable frameworks for data exchange, facilitating access for organisations to a diverse array of high-quality datasets without the necessity of customised data-sharing agreements. By reducing transaction costs and enabling dynamic pricing and licensing, these platforms are reshaping data-driven innovation in sectors such as finance, healthcare, mobility, healthcare, and smart cities [2, 19, 25]. A variety of data marketplace architectures have emerged, including centralised platforms like AWS Data Exchange, decentralised protocols such as Ocean Protocol, and domain-specific exchanges tailored to industry needs [2].

Several methods exist for data valuation and monetisation, ranging from market-based pricing models to algorithmic approaches that estimate value based on utility or metadata. However, these methods often face limitations in accurately capturing a dataset's true worth, particularly in dynamic and decentralised environments [2, 19]. Current data marketplaces face pricing asymmetries, weak metadata and version control, regulatory constraints, and poor interoperability, hindering scalability and trust. In this paper, we propose a negotiation-based approach that aims to align buyer utility with seller valuation, addressing both economic efficiency and fairness in Machine Learning Data Marketplaces (MLDM).

This paper is structured as follows: Section 2 reviews Data Marketplaces for Machine Learning, focusing on pricing and negotiation. Section **??** introduces the UxV-DPN protocol. Section 4 presents the empirical evaluation, including setup and results. Finally, Section 5 concludes the paper and suggests future work.

2 DATA MARKETPLACES FOR ML

Artificial Intelligence relies heavily on large amounts of data. Yet in many real-world scenarios, data is inherently distributed across autonomous agents, edge devices, or institutional boundaries [14,

This work is licensed under the Creative Commons BY-NC-ND 4.0 International License. Visit https://creativecommons.org/licenses/by-nc-nd/4.0/ to view a copy of this license. For any use beyond those covered by this license, obtain permission by emailing info@vldb.org. Copyright is held by the owner/author(s). Publication rights licensed to the VLDB Endowment.

25]. Each agent may collect its own localised data stream, such as sensor readings, user transactions, or domain-specific logs. This decentralised collection process means that each agent often lacks the volume or diversity of data required to train a robust and high-performance model. This fragmentation creates a barrier to model performance and scalability.

Centralising these disparate datasets is often impractical. Privacy regulations, bandwidth limitations, and a lack of trust among data owners can prevent raw data from being pooled in a single repository [23]. Decentralised approaches like federated learning attempt to address these concerns by allowing agents to share model updates instead of raw data. Although this approach addresses some of the limitations above, it does not account for the real costs of data acquisition, preprocessing, or labeling. In practice, data represents a tangible economic asset, one that requires incentives and protections if owners are to participate in collaborative learning.

Data marketplaces for machine learning embrace this economic perspective [8, 18, 22]. By viewing data as a scarce, valuable commodity, these marketplaces enable agents to buy and sell datasets based on measurable utility. In such a setting, each dataset is priced according to the utility it brings to a downstream learning task (e.g., how much adding a new batch of examples boosts accuracy). In fact, data trading becomes a utility-driven transaction rather than purely data exchange.

An example of one such marketplace is the *Machine Learning Data Marketplace (MLDM)* framework [3, 4]. MLDM is a decentralised framework that facilitates data exchange among autonomous learning agents. It enables data trading through integrated modules for acquisition, valuation, monetisation, negotiation, and exchange (see Figure 1). Each participant acts as a data prosumer, producing (selling) and consuming (buying) data. Participants train supervised models on local datasets and evaluate their data using the Gain-Data-Shapley-Value (GDSV) [5], which determines the contribution of each sample to model performance. These evaluations inform pricing and negotiation, allowing high-impact data to command higher prices. After negotiations, participants assess the data exchange's impact on performance improvements (details in Section 2.1).

2.1 Algorithmic Framework for ML data Marketplace

The MLDM is a decentralised society of intelligent agents that act both as data producers and consumers (i.e., prosumers), aiming to collaboratively enhance their predictive performance through data exchange (see Algorithm 1). Let $A = \{A_1, A_2, ..., A_n\}$ denote the set of all agents, where each agent $A_i \in A$ develops a supervised learning model $\mathcal{M}_{i,t}$ at time t trained on its local dataset, $D_{i,t} = \left\{X_{i,t}, y_{i,t}\right\}$, where $\mathcal{M}_{i,t} : X_{i,t} \to y_{i,t}$. Model performance is estimated at each iteration using an appropriate methodology (e.g., hold-out) and measure (e.g. accuracy) as $\mathcal{PR}_{i,t}$.

At each time step t, seller A_j offer subsets of its data for trade, denoted as $TrB_{j,t}$, for potential trade with multiple buyers. Each buyer. The seller A_j calculates the value of $TrB_{j,t}$ using a data valuation function:

$$\mathcal{DV}(TRrB_{j,t}, \mathcal{M}_{j,t}, \mathcal{PR}_{i,t}) \to \phi_{j,t} \tag{1}$$

Algorithm 1 MLDM Algorithmic Framework

```
t ← 0
Buyer A_i, Seller A_j
while t ≤ T do
    All agents train their models (M_{i,t}, PR_{i,t})&(M_{j,t}, PR_{j,t})
    Seller value its data DV(TRrB_{j,t}, M_{j,t}, PR_{i,t}) → φ_{j,t}
    Data Pricing PF(φ_{j,t}) → p^S_{j,t}, p^B_{i,t}
    Negotiation between A_i, A_j: NF(p^S_{j,t}, p^B_{i,t}) → p^{i,j}_t
    Exchange traded data: EF(TrB_{j,t}, X_{i,t}, p^{i,j}_t) → X'_{i,t}, Bg'_{i,t}
    Calculate Gain: GF(TrB_{j,t}, p^{i,j}_t, M_{i,t}, PR_{i,t}) → g_{i,t}, g_{j,t}
end while
```

which evaluates the potential impact of the traded data on model performance. Based on this valuation, the seller sets price (an economic value) using a pricing function: $\mathcal{PF}(\phi_{j,t}) \to p^S_{j,t}$(See equation 4).Then, the buyer assesses the The buyer, A_i, proposes its desired price for traded data $\mathcal{PF}(\phi_{j,t}) \to p^B_{i,t}$ (See equation 6). An iterative negotiation protocol $\mathcal{NF}(p^S_{j,t}, p^B_{i,t})$ is initiated to resolve differences between the seller's offer and the buyer's proposal (section 3). If the negotiation takes place, it may result in a mutually agreed-upon price, denoted as $p^{i,j}_t$. While the negotiation process has the potential to be iterative until an agreement is reached, this paper considers only a single iteration.

If an agreement is reached, the buyer pays from its budget, $Bg'_{i,t} = Bg_{i,t} - p^{i,j}_t$, then incorporates the acquired data into its training set, $X'_{i,t} = X_{i,t} \cup TrB_{j,t}$. This process is mathematically represented as follows:

$$\mathcal{EF}(TrB_{j,t}, X_{i,t}, p^{i,j}_t) \to X'_{i,t}, Bg'_{i,t} \tag{2}$$

and retrains its model accordingly, resulting in $\mathcal{M}_{i,t+1}$. Finally, both parties evaluate the gain from the transaction through a gain function:

$$\mathcal{GF}(TrB_{j,t}, p^{i,j}_t, M_{i,t}, \mathcal{PR}_{i,t}) \to g_{i,t}, g_{j,t} \tag{3}$$

which reflects the change in utility for both the buyer and seller as a result of the data trade, taking into account the trade-off between economic cost and predictive improvement.

2.2 Data Pricing and Negotiation (DPN) in ML Data Marketplace

Data pricing is the process of assigning economic value to datasets, enabling their exchange in digital marketplaces. Existing approaches span multiple dimensions [19]:

- Versioning: Creating different versions of data products tailored to customer needs, linking price to perceived value. Examples include varying data granularity, privacy levels, or delivery delays. [6, 20].
- Arbitrage-Free Pricing: Ensuring pricing models prevent exploitation of price differences across markets or channels. Techniques include query-based pricing and entropy-based pricing [17, 27].
- Revenue Maximisation: Designing pricing strategies to attract maximum customers while optimising revenue, such

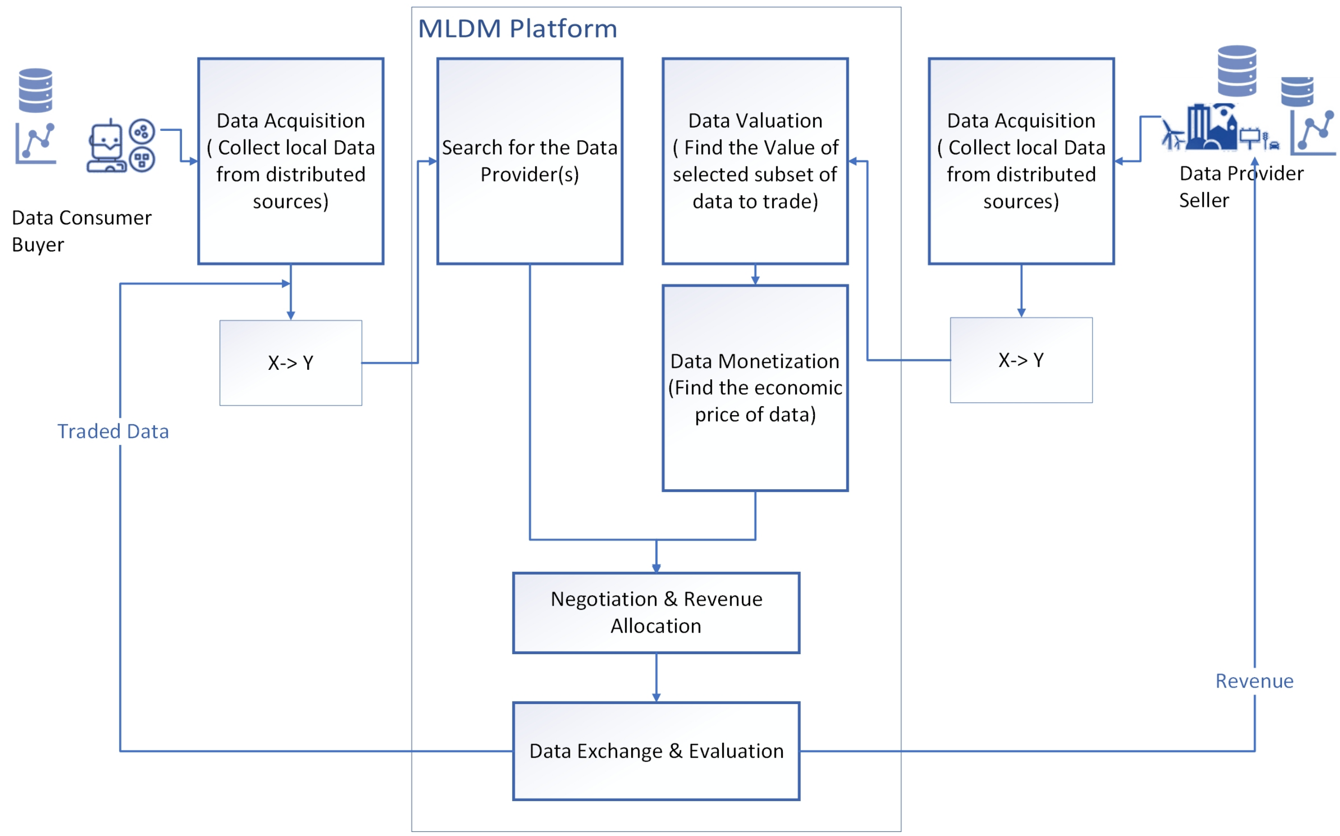

Figure 1: The Architecture of the MLDM Platform

as bundling, subscription models, and dynamic pricing [7, 9].

- Fair and Truthful Pricing: Using mechanisms like Shapley fairness to ensure equitable revenue distribution among sellers and incentivise buyers to reveal true valuations [1, 12, 13].

While these approaches offer foundational principles for economic transactions, their direct application in MLDM presents new challenges. In MLDM, data is not only an economic commodity but also a model performance enhancer. As such, pricing mechanisms must account for the *predictive utility* of data from the buyer's perspective and the *value and revenue potential* from the seller's side. Moreover, the iterative, agent-based nature of decentralised MLDM calls for negotiation-aware pricing strategies that adapt to local constraints, such as budgets, model needs, and dynamic market conditions. These challenges motivate the development of the UxV-DPN protocol, introduced in the following subsection.

Algorithm 2 Data Pricing and Negotiation Mechanism

Seller price : $\mathcal{PF}(\mu_{j,t}, \phi_{j,t}) \rightarrow p^S_{j,t}$
Buyer desired price: $\mathcal{PF}(b_{i,t}, \phi_{j,t}) \rightarrow p^B_{i,t}$
Buyer Decision-making:
if $p^B_{i,t} \geq p^S_{j,t}$ **then**
  Accept negotiation
else if $\frac{|p^S_{j,t} - p^B_{i,t}|}{p^B_{i,t}} > \epsilon_i$ **then**
  Reject (overpricing)
else
  Propose new price $p^B_{i,t}$
end if
Seller Decision-making:
if $\frac{|p^S_{j,t} - p^B_{i,t}|}{p_{i,t}} < \epsilon_j$ **then**
  Accept negotiation
else
  reject (underpricing)
end if

3 UXV-DPN: UTILITY-VS-VALUE DATA PRICING AND NEGOTIATION

We propose **UxV-DPN**, a negotiation protocol that enables prosumer agents to engage in balanced, utility-driven exchanges. By dynamically aligning value and utility through adaptive bargaining, UxV-DPN aims to support equitable transactions that balance buyer utility and seller value in decentralised data exchanges. UxV-DPN will be evaluated within the context of MLDM. However, it can serve as a general negotiation strategy in data marketplaces for ML.

Each transaction in this context represents an economic exchange between two rational but asymmetric agents:

- The **seller**, motivated by monetising their data asset, seeks to maximise **Value**, defined as the potential revenue derived from offering high-quality data.
- The **buyer**, driven by model performance improvement but constrained by its budget, evaluates the dataset's **Utility**, referring to the expected gain in model accuracy for a corresponding cost.

This perspective highlights the need for a pricing mechanism that simultaneously attempts to find high-quality data and maximise predictive utility.

The proposed mechanism is illustrated in the negotiation mechanism diagram Figure 2 and formalised in Algorithm 2. The process begins with the seller, agent A_j, setting a price for a traded set ($TrB_{j,t}$) at time t based on its data valuation method. The price function is expressed as:

$$\mathcal{PF}(\mu_{j,t}, \phi_{j,t}) \rightarrow p^S_{j,t} \tag{4}$$

where $\phi_{j,t}$ is value of data at time t and $\mu_{j,t}$ represents willing-to-sell. This willing-to-sell increases when the budget decreases, defined by the equation $\mu_{j,t} = 0.001 + \frac{Bg_{j,0} - Bg_{j,t}}{Bg_{j,0}}$. For the purposes of this paper, we simplify $\mathcal{PF}$ to be calculated as follows:

$$p^S_{j,t} = \mu_{j,t} * \phi_{j,t} \tag{5}$$

This linear formulation captures a basic proportional relationship between a seller's motivation and the intrinsic value of the data. It allows for transparent parameterisation and controlled experimental evaluation. However, we acknowledge that more complex formulations, such as non-linear functions (e.g., logarithmic or exponential) or dynamic pricing models, may more accurately reflect real-world seller behaviour under competitive or strategic pressure.

Once the price has been established, the buyer determines their desired price using the following formula:

$$\mathcal{PF}(b_{i,t}, \phi_{j,t}) \rightarrow p^B_{i,t} \tag{6}$$

where $\phi_{j,t}$ represents the value of the traded data belonging to agent A_j at time t, and $b_{i,t}$ signifies the public value that reflects how eager buyer A_i is to acquire a new dataset [1]. This is defined as:

$$b_{i,t} = \hat{\mu}_{i,t} * \mathcal{PR}_{i,t} * \frac{Bg_{i,t}}{Bg_{i,0}} \tag{7}$$

Here, $\hat{\mu}_{i,t}$ indicates the buyer's willingness to pay, which quantifies how much buyer A_i values a marginal improvement in learning performance (i.e., the price they are prepared to pay for a unit increase in performance). Based on these parameters, the desired price for a new dataset is then calculated as follows:

$$p^B_{i,t} = b_{i,t} * \phi_{j,t} \tag{8}$$

Based on both seller's price ($p^S_{j,t}$) and buyer's desired price ($p^B_{i,t}$), the buyer evaluates the offer and must make a decision: accept, reject, or propose a counter-offer. If the buyer accepts the proposed price by seller ($p^B_{i,t} \geq p^S_{j,t}$), the negotiation concludes successfully.

If a buyer outright rejects a seller's proposed price without engaging in any negotiation, it indicates that the proposed price significantly diverges from the buyer's expectations or valuation. This occurs when the relative difference between the buyer's internal valuation $p^B_{i,t}$, and the seller's offered price $p^S_{j,t}$ exceeds a predefined acceptance threshold ϵ_i. This relationship can be expressed mathematically as:

$$\frac{|p^S_{j,t} - p^B_{i,t}|}{p^B_{i,t}} > \epsilon_i \tag{9}$$

Here, ϵ_i denotes the acceptable margin of price deviation from the buyer's perspective. It encodes the tolerance level that the buyer is willing to accept before deciding that the offer is too far from their valuation to warrant a counteroffer. If this threshold is passed, the transaction is immediately terminated. In this paper, ϵ is fixed across all agents for consistency. In our experiments, we set $\epsilon = 0.25$, allowing buyers to accept offers within ±25% of their internal valuation. This threshold was empirically chosen to balance negotiation flexibility with economic discipline. If this threshold is exceeded, the transaction is terminated, discouraging extreme price proposals and preserving the integrity of the negotiation protocol. This mechanism serves to discourage unreasonably high or low initial price settings and maintains offers within a range that is mutually acceptable for negotiation.

If the buyer makes a counter-offer price as $p^B_{i,t}$, the negotiation enters a new round. Upon receiving this counter-offer, the seller assesses the revised price and has three possible actions: accept the offer, reject it, or propose a new counter-offer. This decision is formed by the relative difference between buyer's offer $p^B_{i,t}$ and the seller's original price $p^S_{j,t}$. If the proposed counter-offer is sufficiently close to the seller's expectation, formally, if

$$\frac{|p^S_{j,t} - p^B_{i,t}|}{p^S_{j,t}} < \epsilon_j \tag{10}$$

where ϵ_j is the seller's tolerance threshold, the seller accepts the offer. This leads to a successful negotiation. Conversely, if the revised offer is outside this acceptable range, the seller perceives it as underpricing and subsequently rejects the proposal. This results in a failed negotiation.

While this framework naturally supports iterative bargaining with multiple counter-offers between buyer and seller, this paper focuses on a simplified one-iteration negotiation to emphasise the core incentive and trust mechanisms. This simplified process is illustrated in 2, which depicts the sequential flow of decisions in the data pricing algorithm.

This iterative negotiation process encourages rational behaviour in a competitive but cooperative setting, where neither the buyer nor the seller is assumed to be malicious. Instead, trust and incentive alignment are promoted through a structured sequence of offers, decisions, and outcomes that balance negotiation freedom with transactional discipline.

The main goal of this method is to prevent overpricing or underpricing of data assets, providing that prices better reflect the true utility and value of datasets in machine learning tasks.

4 EMPIRICAL VALIDATION

This section evaluates the effectiveness of the proposed negotiation strategy within MLDM framework. Specifically, we analyse how the negotiation mechanism, based on the dynamic pricing

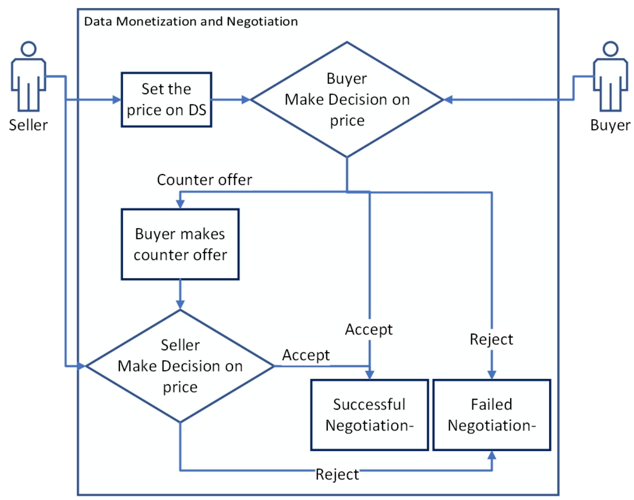

Figure 2: The Process of Data Pricing ($\mathcal{PF}$) and Negotiation ($\mathcal{NF}$) in the ML data marketplace

and counter-offer protocol, improves the average predictive performance of agents operating under the same learning algorithm. The central hypothesis is that UxV negotiation allows agents to acquire high-quality data at economically viable prices, thereby enhancing overall model accuracy.

4.1 Experimental Setup

The MLDM framework consists of five autonomous agents, each independently training a model using the K-Nearest Neighbors (KNN) classification algorithm. These settings were chosen for their simplicity and consistency across datasets, enabling a focused evaluation of the effects of data exchange and negotiation strategies.

The key elements of the setup are as follows:

- **Datasets:** The MLDM framework was evaluated by choosing a random set of 48 public classification data sets from the OpenML platform [26] with different sizes and properties to simulate diversity.
- **Budget:** Each agent starts with a fixed initial budget of $1000(Bg_{i,0} = 1000)$, which is used to acquire data from peers during negotiations.
- **Simulation Details:** All experiments were run for 10 iterations and repeated 10 times to ensure statistical reliability.
- **Global Evaluation:** Besides the local training and test data, a separate test set is left out for global evaluation, it is called the validation set (TsG). Agents estimate the predictive performance of their models on a it, indicated as $\mathcal{GPR}_{i,t}$.

Based on global evaluation, the primary evaluation metric is Total Global Performance ($\mathcal{GPR}^t$), defined as the average predictive accuracy of all agents $A_i, i = \{1, ..., n\}$ across all datasets $D_j, j = \{1, ..., N\}$ at time step t:

$$\mathcal{GPR}_{tot} = \frac{1}{n * N} \sum_{D_j=1}^{N} \sum_{i=1}^{n} \mathcal{GPR}_{i,t}^{D_j} \tag{11}$$

where $\mathcal{PR}_{i,t}^{D_j}$ stands for the performance of learning model of agent A_i at time t, on data set D_j.

To benchmark the negotiation strategy, we define three comparison scenarios:

- **No Exchange (NE)**: Agents train on their initial local data only; no trading occurs.
- **Single Agent (SA)**: A centralised environment with only one agent accessing the entire dataset.
- **Simple Negotiation** $MLDM_{\phi}$ **(SimpleNeg)**: This is a fixed-price strategy where buyer agents accept negotiations if the budget is sufficient. Otherwise, they request a reduced size of the traded set. If the budget is exhausted, they do not enter into negotiations [5].

To assess the impact of data exchange on the performance of machine learning models, we compare MLDM with its No-Exchange counterpart (NE). The Global Evaluation metrics for the no-exchange and exchange versions are denoted as $\mathcal{GPR}_{tot}^{ne}$ and $\mathcal{GPR}_{tot}$, respectively. The Accuracy Gain Function is defined as:

$$\mathcal{AG}_{tot} = \mathcal{GPR}_{tot}^{MLDM} - \mathcal{GPR}_{tot}^{NE} \tag{12}$$

This metric helps us measure which data exchange contributes to a performance improvement in machine learning models. This formulation captures buyer utility as a function of marginal learning gain; a formal utility model and analytical guarantees (e.g., convergence, fairness) are left for future work.

4.2 Results

This section evaluates the core research question: Does the proposed negotiation strategy (UxV-DPN) improve predictive performance for agents in MLDM, considering the seller's value and the buyer's utility? To address this, we compare the effectiveness of UxV-DPN against SimpleNeg and two baselines: NE and SA. Our results indicate that only after a couple of iterations, the exchange of data achieves similar levels of predictive performance as the single agent. However, UxV-DPN enables agents to achieve a better predictive performance compared to NE, while reducing the cost compared to Simple.

Figure 3 presents the average predictive performance of agents at different iterations. As expected, the NE baseline results in the lowest accuracy, highlighting the limitations of isolated learning without data sharing. In contrast, all exchange-based strategies demonstrate significant performance improvements, especially during the initial phases of data acquisition when the data volume is low, which affects the training of high-performance models. The SimpleNeg strategy achieves the highest average performance when agents have access to the complete dataset (100%), followed closely by UxV-DPN and Single Agent.

Figure 4 further explores the trade-off between the total payment made by buyers and the accuracy gain, $\mathcal{AG}_{tot}$, obtained through negotiation. The UxV-DPN strategy shows a systematic advantage in cost-effectiveness. Agents using this strategy achieve an approximate 5% increase in model accuracy compared to the NE, while incurring substantially lower payments than those using the Simple strategy. Comparing to UxV-DPN, the Simple strategy only achieves about a 1% increase in accuracy (a total gain of roughly 6%) but at nearly double the cost. This shows that UxV-DPN provides a

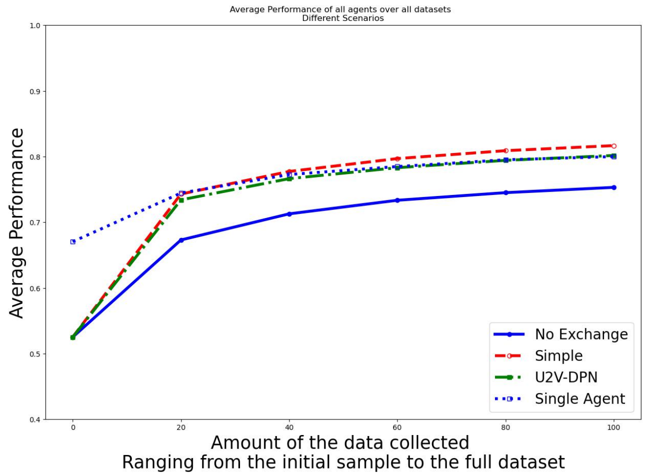

Figure 3: Average learning performance across agents under different data exchange scenarios

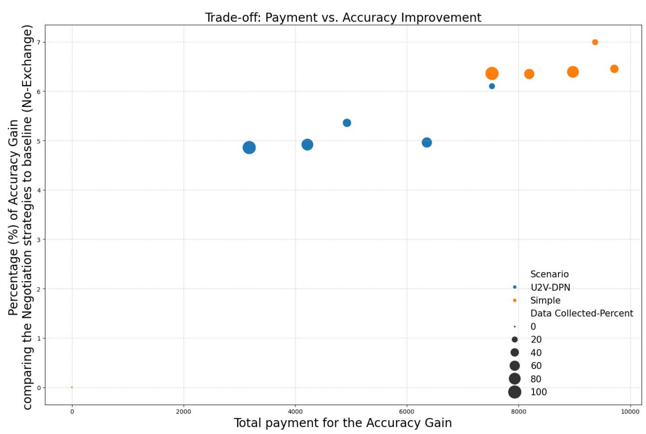

Figure 4: Trade-off between total payment and accuracy gain ($\mathcal{AG}_{tot}$) across negotiation strategies.

strategy considering the value and revenue for sellers, as well as the cost and utility for buyers.

Overall, the UxV-DPN strategy provides a balance between utility and cost in MLDM scenarios. Although the Simple negotiation achieves marginally higher accuracy, it incurs a greater cost. The UxV-DPN strategy displays that dynamic data pricing and negotiation can lead to efficient and equitable data trading, making it a practical option for real-world data marketplaces where optimising both predictive benefits and cost sensitivity is crucial.

5 CONCLUSION

The primary objective of this work is to evaluate the effectiveness of the proposed negotiation protocol, **Utility-vs-Value Data Pricing and Negotiation (UxV-DPN)**, in the context of Data Marketplaces. The primary objective of UxV-DPN is to provide cost-effective data exchanges that enhance the predictive performance of the buyer's learning model while maintaining the seller's value and revenue. To address this question, we conduct experiments through MLDM across multiple datasets using a society of five learning agents, all trained using a classification algorithm (KNN). In this setting, agents often act as prosumers – participating both as data buyers and sellers. Through comparative experiments across other scenarios (NE, SA, SimpleNeg), we indicated that the proposed UxV-DPN protocol achieves a balance between predictive improvement and economic cost. While the SimpleNeg strategy delivers slightly higher accuracy, it costs more (nearly double on average), thereby reducing overall cost-effectiveness. In contrast, UxV-DPN delivers over 80% of the performance while maintaining significantly lower payment levels. It provides a more efficient trade-off between cost and learning benefits, improving predictive accuracy by approximately 5% relative to a NE baseline. UxV-DPN converges toward stable economic equilibrium, making it particularly well-suited for iterative, real-world data exchanges, where budget constraints and learning demands continually evolve.

Overall, the results illustrate the core principle behind UxV-DPN: aligning utility for buyers with value for sellers promotes sustainable and scalable Machine Learning Data Marketplaces. Future research will evaluate UxV-DPN in more diverse settings, including heterogeneous agents with different budgets and learning models, as well as larger-scale environments with imbalanced data ownership. These directions will help assess the protocol's robustness, fairness, and scalability in more realistic marketplace conditions. We also aim to formalise the theoretical properties of UxV-DPN, including convergence behaviour and fairness guarantees, and to refine the modelling of buyer utility beyond performance-based heuristics.

ACKNOWLEDGMENTS

This work was financially supported by: UID/00027 of the LIACC - Artificial Intelligence and Computer Science Laboratory - funded by Fundação para a Ciência e a Tecnologia, I.P./ MCTES through the national funds. Additional funding was provided by Agenda "Center for Responsible AI", nr. C645008882-00000055, investment project nr. 62, financed by the Recovery and Resilience Plan (PRR) and by European Union - NextGeneration EU.

REFERENCES

[1] Anish Agarwal, Munther Dahleh, and Tuhin Sarkar. 2019. A marketplace for data: An algorithmic solution. *ACM EC 2019 - Proceedings of the 2019 ACM Conference on Economics and Computation* (2019), 701–726.
[2] Santiago Andrés Azcoitia and Nikolaos Laoutaris. 2022. A Survey of Data Marketplaces and Their Business Models. *ACM SIGMOD Record* 51, 3 (2022), 18–29. https://doi.org/10.1145/3572751.3572755 arXiv:2201.04561
[3] H. Baghcheband, C. Soares, and L. Reis. 2022. Machine Learning Data Markets: Trading Data using a Multi-Agent System. In *2022 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*. IEEE Computer Society, Los Alamitos, CA, USA, 450–457.
[4] H. Baghcheband, C. Soares, and L. Reis. 2024. MLDM: Machine Learning Data Market Based on Multi-Agent Systems. *IEEE Internet Computing* 01 (may 2024), 1–7.
[5] Hajar Baghcheband, Carlos Soares, and Luis Paulo Reis. 2024. Shapley-Based Data Valuation Method for the Machine Learning Data Markets (MLDM). In *Foundations of Intelligent Systems*. Springer Nature Switzerland, Cham, 170–177.
[6] Dirk Bergemann, Alessandro Bonatti, and Alex Smolin. 2018. The Design and Price of Information. *American Economic Review* 108, 1 (January 2018), 1–48. https://doi.org/10.1257/aer.20161079
[7] Shuchi Chawla, Shaleen Deep, Paraschos Koutris, and Yifeng Teng. 2019. Revenue Maximization for Query Pricing. arXiv:1909.00845 [cs.DB] https://arxiv.org/abs/1909.00845

[8] Lingjiao Chen, Paraschos Koutris, and Arun Kumar. 2018. Model-based Pricing for Machine Learning in a Data Marketplace. *ArXiv* abs/1805.11450 (2018). https://api.semanticscholar.org/CorpusID:195259536

[9] Lingjiao Chen, Paraschos Koutris, and Arun Kumar. 2019. Towards Model-based Pricing for Machine Learning in a Data Marketplace. In *Proceedings of the 2019 International Conference on Management of Data* (Amsterdam, Netherlands) *(SIGMOD '19)*. Association for Computing Machinery, New York, NY, USA, 1535–1552. https://doi.org/10.1145/3299869.3300078

[10] Raul Castro Fernandez, Pranav Subramaniam, and Michael J. Franklin. 2020. Data market platforms: trading data assets to solve data problems. *Proc. VLDB Endow.* 13, 12 (July 2020), 1933–1947. https://doi.org/10.14778/3407790.3407800

[11] Mike Fleckenstein, Ali Obaidi, and Nektaria Tryfona. 2023. Data Valuation: Use Cases, Desiderata, and Approaches. In *Proceedings of the Second ACM Data Economy Workshop* (Seattle, WA, USA) *(DEC '23)*. Association for Computing Machinery, New York, NY, USA, 48–52. https://doi.org/10.1145/3600046.3600054

[12] Ruoxi Jia, David Dao, Boxin Wang, Frances Ann Hubis, Nezihe Merve Gurel, Bo Li, Ce Zhang, Costas Spanos, and Dawn Song. 2019. Efficient task-specific data valuation for nearest neighbor algorithms. *Proc. VLDB Endow.* 12, 11 (July 2019), 1610–1623. https://doi.org/10.14778/3342263.3342637

[13] Ruoxi Jia, David Dao, Boxin Wang, Frances Ann Hubis, Nick Hynes, Nezihe Merve Gürel, Bo Li, Ce Zhang, Dawn Song, and Costas J. Spanos. 2019. Towards Efficient Data Valuation Based on the Shapley Value. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics (Proceedings of Machine Learning Research)*, Kamalika Chaudhuri and Masashi Sugiyama (Eds.), Vol. 89. PMLR, 1167–1176. https://proceedings.mlr.press/v89/jia19a.html

[14] Kangsoo Jung, Junkyu Lee, Kunyoung Park, and Seog Park. 2019. PRIVATA: Differentially Private Data Market Framework using Negotiation-based Pricing Mechanism. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management* (Beijing, China) *(CIKM '19)*. Association for Computing Machinery, New York, NY, USA, 2897–2900. https://doi.org/10.1145/3357384.3357855

[15] Aida Manzano Kharman, Christian Jursitzky, Quan Zhou, Pietro Ferraro, Jakub Marecek, Pierre Pinson, and Robert Shorten. 2022. On the Design of Decentralised Data Markets. *arXiv preprint arXiv:2206.06299* (2022), 1–26. arXiv:2206.06299

[16] Karl Koch, Stephan Krenn, Tilen Marc, Stefan More, and Sebastian Ramacher. 2022. KRAKEN: A Privacy-Preserving Data Market for Authentic Data. In *DE 2022 - Proceedings of the 1st International Workshop on Data Economy, Part of CoNEXT 2022 (DE 2022 - Proceedings of the 1st International Workshop on Data Economy, Part of CoNEXT 2022)*. Association of Computing Machinery, United States, 15–20. https://doi.org/10.1145/3565011.3569057

[17] Bing-Rong Lin and Daniel Kifer. 2014. On arbitrage-free pricing for general data queries. *Proc. VLDB Endow.* 7, 9 (May 2014), 757–768. https://doi.org/10.14778/2732939.2732948

[18] Jinfei Liu, Jian Lou, Junxu Liu, Li Xiong, Jian Pei, and Jimeng Sun. 2021. Dealer: An End-to-End Model Marketplace with Differential Privacy. *Proc. VLDB Endow.* 14 (2021), 957–969. https://api.semanticscholar.org/CorpusID:243742143

[19] Jian Pei. 2022. A Survey on Data Pricing: From Economics to Data Science. *IEEE Transactions on Knowledge and Data Engineering* 34, 10 (Oct. 2022), 4586–4608. https://doi.org/10.1109/tkde.2020.3045927

[20] Carl Shapiro and Hal Varian. 2008. *Information Rules: A Strategic Guide to The Network Economy.* Vol. 30. Harvard Business School Press. https://doi.org/10.2307/1183273

[21] Florian Stahl, Fabian Schomm, and Gottfried Vossen. 2014. Data marketplaces: An emerging species. *Frontiers in Artificial Intelligence and Applications* Databases, August 2013 (2014), 145–158.

[22] Peng Sun, Xu Chen, Guocheng Liao, and Jianwei Huang. 2022. A Profit-Maximizing Model Marketplace with Differentially Private Federated Learning. *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications* (2022), 1439–1448. https://api.semanticscholar.org/CorpusID:249900418

[23] Peng Sun, Guocheng Liao, Xu Chen, and Jianwei Huang. 2024. A Socially Optimal Data Marketplace With Differentially Private Federated Learning. *IEEE/ACM Trans. Netw.* 32, 3 (Jan. 2024), 2221–2236. https://doi.org/10.1109/TNET.2024.3351864

[24] Zhihua Tian, Jian Liu, Jingyu Li, Xinle Cao, Ruoxi Jia, Jun Kong, Mengdi Liu, and Kui Ren. 2023. Private Data Valuation and Fair Payment in Data Marketplaces. arXiv:2210.08723 [cs.CR]

[25] Xuezhen Tu, Kun Zhu, Nguyen Cong Luong, Dusit Niyato, Yang Zhang, and Juan Li. 2021. Incentive Mechanisms for Federated Learning: From Economic and Game Theoretic Perspective. *CoRR* abs/2111.11850 (2021).

[26] J. Vanschoren, J. van Rijn, B. Bischl, and L. Torgo. 2013. OpenML: Networked Science in Machine Learning. *SIGKDD Explorations* 15, 2 (2013), 49–60. http://doi.acm.org/10.1145/2641190.2641198

[27] Chaolun Xia and S. Muthukrishnan. 2018. Arbitrage-free Pricing in User-based Markets. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems* (Stockholm, Sweden) *(AAMAS '18)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 327–335.

[28] Yiding Zhu, Hongwei Zhang, Jiayao Zhang, Jinfei Liu, and Kui Ren. 2024. DataPrice: An Interactive System for Pricing Datasets in Data Marketplaces. *Proc. VLDB Endow.* 17, 12 (Aug. 2024), 4433–4436. https://doi.org/10.14778/3685800.3685893