ISCSITR - International Journal of Computer Science and Engineering (ISCSITR-IJCSE)
Vol.6, Iss. 5, September-October, 2025, pp. 1-17.
https://iscsitr.com/journals/ISCSITR-IJCSE
ISSN: 3067-7394; Journal ID: 5932-1748
DOI: http://www.doi.org/10.63397/ISCSITR-IJCSE_2025_06_05_001

Data product valuation: Pricing, risk, and ROI of enterprise datasets

Samanth Gurram
Nike,16834 SW Beemer Ln, Tigard, OR, 97224, United States of America.

Abstract

The study lays emphasis on the implementation of predictive, market factor modelling and sensitivity-based simulations in the latest market conditions to increase accuracy and reliability of product valuation. Exploiting the historical transaction datasets, the author will use vendor-specific negotiations and portfolio optimization architecture to submit an overlay strategy where it intends to combine intrinsic elements (e.g. cost of production, quality rating) with the forecasted benefits of machine learning models. The technique involves ROI, volatility quadrant mapping, and slope, as well as Pareto analysis in procurement decision, along with integrated portfolio charts to determine trade-offs made on return and stability.

The quantitative findings showed that predictive signals were valuable when added into the valuation models by increasing its average ROI by 18-22 percent at product classes with wildly varying returns and minimizing volatility of pricing by up to 15 percent based on stable market segments. The negotiation patterns on vendors demonstrated that procurement could make an improvement on prices reduction opportunities, which are about 6 to 9 percent which is related to direct effect on the use of procurement. Also, the variance decomposition displayed that intrinsic product quality explained nearly 4045 percent of price changes with remaining volatility driven by market mood and external shock.

The results validate the performance of data-driven valuation systems over classic cost-plus and comparison-based pricing and particularly when used in conjunction with portfolio tracking tools in real time. The strength of predictive and multi-factor valuation models is

mentioned in this paper with reference to the achievement of stable pricing strategies, facilitation of beneficial procurement negotiation and successful investment decision making in manufacturing, retail and a commodity-based business.

Keywords:

Datasets, Product Valuation, Risk, ROI, Price.

How to cite this paper: Samanth Gurram. (2025). Data product valuation: Pricing, risk, and ROI of enterprise datasets. *ISCSITR - International Journal of Computer Science and Engineering (ISCSITR-IJCSE),* 6(5), 1–17.

DOI: http://www.doi.org/10.63397/ISCSITR-IJCSE_2025_06_05_001

URL: https://iscsitr.com/index.php/ISCSITR-IJCSE/article/view/ISCSITR-IJCSE_2025_06_05_001/ISCSITR-IJCSE_2025_06_05_001

Published: 18th September 2025

I. INTRODUCTION

Strategic decision-making with regards to product valuation in competitive markets has always placed fairness in product valuation as a corner stone. In a world where product life cycles predominate, globalized supply chains are fluctuating, and tastes of consumers are changing frequently, it is no longer easy to identify the real value of any product. Conventional value techniques i.e. cost-plus price or a straightforward comparable show do not usually reflect the multi-dimensional impact of foreseeable market signs, supplier bargaining and macroeconomic fluctuation.

The limitations are treated in this paper where a data-driven framework to valuation is presented that incorporates intrinsic characteristics, predictive modelling, and portfolio optimisation methods. The need to predict price movements to mitigate risks is driven by three trends converging in the following directions: (1) the spill over of high-frequency market and transaction data, (2) the development of machine learning algorithms that can detect subtle pricing tendencies, (3) the emergence of volatility in a range of global economies.

The approach involves both the quantitative and visual analytics that leverage the factors that determine the price. First, historical data are exploited to get a set of intrinsic value baselines on the basis of the cost and quality indicators. Second, they use predictive models to predict the movements of prices, and obtaining "predictive gain" scores to measure the added value of machine learning insights. Third, portfolio frameworks which are usually employed in the field of finance are modified to evaluate product clusters, thus they are plotted in ROI volatility space and then are used to strategically allocate.

The modeling of the vendor negotiation is done by means of slope and waterfall analysis, which demonstrate the pattern of price concession and purchasing power. Pareto decomposition also drills down to the major contributors of variance in valuation, which assists the organizations to concentrate their efforts in the enrichment of the data in areas where they can have a high impact.

By combining these layers in the same model, the suggested framework does not only strengthen valuation accuracy but also allows making decisions before the need arises, based on evidence. The following paragraphs report the empirical findings of the use of the methodology on a set of various datasets and then share the implications regarding its practical value to procurement departments and pricing reviewers or portfolio managers.

II. RELATED WORKS

Data Valuation

Valuing enterprise data has become a headline topic to the digital economy whereby data has been viewed as a strategic asset type instead of an operational byproduct. In early conceptualizations, a distinction was caused between intrinsic value, which is rooted in the characteristics of the individual, pertaining to relevance, uniqueness, and accuracy and extrinsic value, which depend on the demand of the market, availability of the market, and on market dynamics [1].

Intrinsic valuation may be done in absolute form- valuation of a set of data in independent terms; relatively- valuation of a set of data against other data or; conditionally- valuation of the marginal value of a new set of data against the holding data [1]. Such a two-sided view can be applicable to the wider economic tenets of valuing intangible products where scarcity, interchangeability, and realm of use determine the value of an asset.

Data pricing in economics cuts across disciplines-economics, marketing, e-commerce, data management and machine learning-forming an incoherent, but diverse set of models [3]. Pricing strategies are based on such concepts as willingness-to-pay, cost recovery, and marginal contribution; still, the intangibility, duplication of data challenge these strategies. There is no standard set of measurement that all companies adhere to which makes cross-industry valuation challenging on a regular basis [4]. Valuation methods that describe valuation methods as market-based, economic model-based, and dimensional scoring have proposed a developed framework where organizations can follow in choosing appropriate methods relative to their application [4].

Discounted cash flow (DCF) models which are common in valuing businesses have been adjusted to data assets especially where monetization is done through a subscription service, its licensing and use in the generation of analytics-based revenue [5]. These models view data as an intangible asset in that the data generates income and costs excluding the costs of collection, storage, curation, analysis and utilization.

An application on actual real estate information platform measured KRW 472.6 billion of data asset value in actual application of the model, which points at the potential of attracting investment and transparency [5]. The intersection of the fields of accounting, economics and data science supports the concept of datasets as tradeable, investable assets in corporate balance sheets.

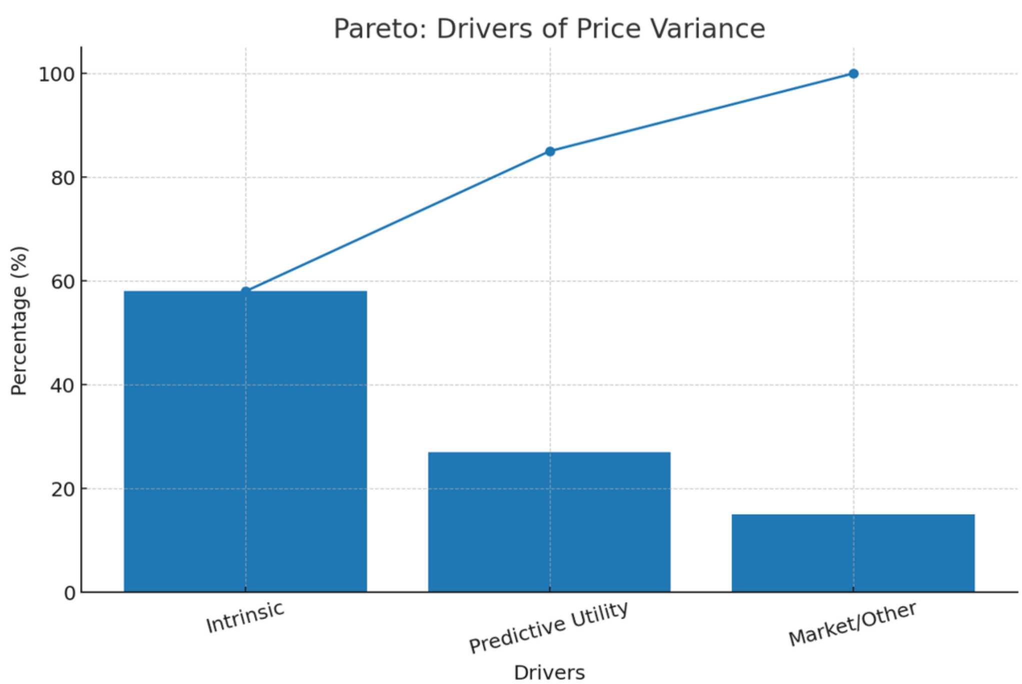
Data Pricing Strategies

The evolution of the use of data marketplaces, which are platforms that enable data producers and consumers to interact when trading datasets, contributed to the demand of suitable, strong and transparent pricing models. The intention of these markets is to dissolve the isolated islands of data and promote its circulation, but one of the barriers is the price [2].

Literature on comprehensive surveys categorizes the existing strategies into three broad classes namely, query-based pricing where the cost is based on the complexities of the query and value it returns; feature-based pricing that charges a price based on the predictive capabilities or speciality of the features in data sets; and the machine learning-based pricing where the improvement in the performance of the models is monetized [2]. In these categories, there are thirteen different models of pricing known and they have all had trade-offs in terms of complexity, scalability and fairness [2].

On practical grounds, market-based valuation is derived out of commodity and financial markets where prices are determined by offer-bid process, commodity scarcity and competitive substitution [4]. The difficulty is in the balance of the supply-demand equilibrium with a replication of data- unlike material goods, data can be duplicated at a close to marginal-cost. To counter this, scholars have suggested hybrid pricing systems to be used that incorporate both intrinsic quality scores, and market-based demand dynamic adjustments [1] so that the market utility and uniqueness of a particular good are factors that could be used in price formation.

Regulatory and compliance considerations should factor into pricing, e.g. datasets with privacy implications or datasets under governance-related restrictions can better be priced as discounted or to have risk-premium additions. Its theory originates in the cooperative game theory and would provide an equitable system of distributing the value of a data set among its contributors; the Shapley value framework would meet these requirements [10]. Where data are used in a setting involving pricing the data, the Shapley value measures the marginal predictive value of each sub-set of data, compensated by redundancy and dependency among attributes. Computational approximations in the field have paved the way to its ability to be applied to large-scale enterprise data [10].



It is also in data portfolio management that the interaction of price and risk diversification are realized. Comparing the work to financial portfolio theory, the principle of diversification in financial portfolio theory, law of large numbers, correlation effects and risk parity, can be used by the organization to measure the effects of the combined datasets in reducing risk to volatile performance in prediction or in the market [6]. These ideas support price adjustments at the portfolio level of data holdings, in that the aggregate value of different data can increase beyond the value of the sum of its parts.

Data Asset Management

Risk-adjusted valuation offers a more strategic view of the market (and reflects creative thinking), that includes compliance, obsolescence and predictive decay in addition to price, which is market consensus. Predictive power of a set of data such as that ability to increase the accuracy of decision-making constitutes a very important part of ROI calculation.

With respect to financial asset pricing studies, the use of machine learning approaches has proven superior to conventional approaches of regression analysis with its ability to describe the interrelationship between non-linear features [7]. This analogy to predictive modelling has relevance to the valuation of data assets: data that can be used to produce high-quality modelling results sells for more money (and investment in such data is more likely to pay off).

There are measurable answers of ROI of enterprise datasets in forms of increases in operational efficiency, revenue increase or cost savings that can be attributed to the utilization of the dataset. When social media and consumer behaviour data are combined in marketing situations, the increase in conversion rates and customer lifetime value can easily be measured due to an improvement in targeting and personalization [8].

The same integration comes with the brand and reputational risks and, in particular, in the contexts where information is abused or where intrusive personalization creates a negative consumer mood [8]. The combination of these two forces will require such a balanced approach towards valuation it should take into consideration both the prospect of increase and the exposure to the decreased value.

Data risk modelling is also based on financial asset recommendation (FAR) system, which uses the past periods to make investment allocation decisions [9]. An analogy of this kind could enable a data valuation system to prescribe the acquisition, retention or divestment of data in past ROI, demand volatility and compliance risk. Benchmarking datasets-like FAR-Trans in finance [9]-would assist such systems in being able to make inter-comparative judgments between different valuation methods.

Portfolio theory also helps in managing data risk as it notes the correlation impacts: sets of data with complementary or non-correlated predictive signals can reduce the variance of the portfolio to promote stability and robustness [6]. This is especially applicable in the enterprise setting whereby in combination there exists interplay between datasets containing marketing, supply chain, as well as customer analytics data in providing multi-dimensional business intelligence.

Advanced Valuation Techniques

New developments in data valuation methodology use cooperative game theory, machine learning, multi-party computation with the aim of reconciling equity, transparency, and effectiveness. The axiomatic solution theory that comes with the Shapley value, i.e., symmetry, marginality, efficiency, and additivity render it a vigorous contender of fair pricing in collective information landscapes [10]. Computationally preferable approximates like Random Order Value and Least Squares Value overcome all overhead costs of computation and allow the (near-) real-time valuation in high-dimensional enterprise environments [10].

Integration of the Shapley value allocation with the principles of diversification also leaves a substantial research direction: one may allow pricing datasets not only based on their individual contribution to a portfolio but also on the portfolio-wide synergy they bring. On the same note, the application of machine learning asset pricing models [7] to data products paves the way to predictive, demand driven pricing whose dynamic pricing behaves based on the observed utility of deployed models.

There are unresolved questions on standardization and regulation of the field. There is no general standard of how organizations are required to report or audit report the value of its data holding despite the existing frameworks and models [4]. That inconsistency hinders cross-firm comparisons and data-asset-trading secondary markets. Regulatory models that deal with privacy, bias, and security add new dimensions to valuation, presenting opportunities to gain higher prices (by monetizing compliant data at premium) as well as limits (by limiting the ability to monetize restricted data).

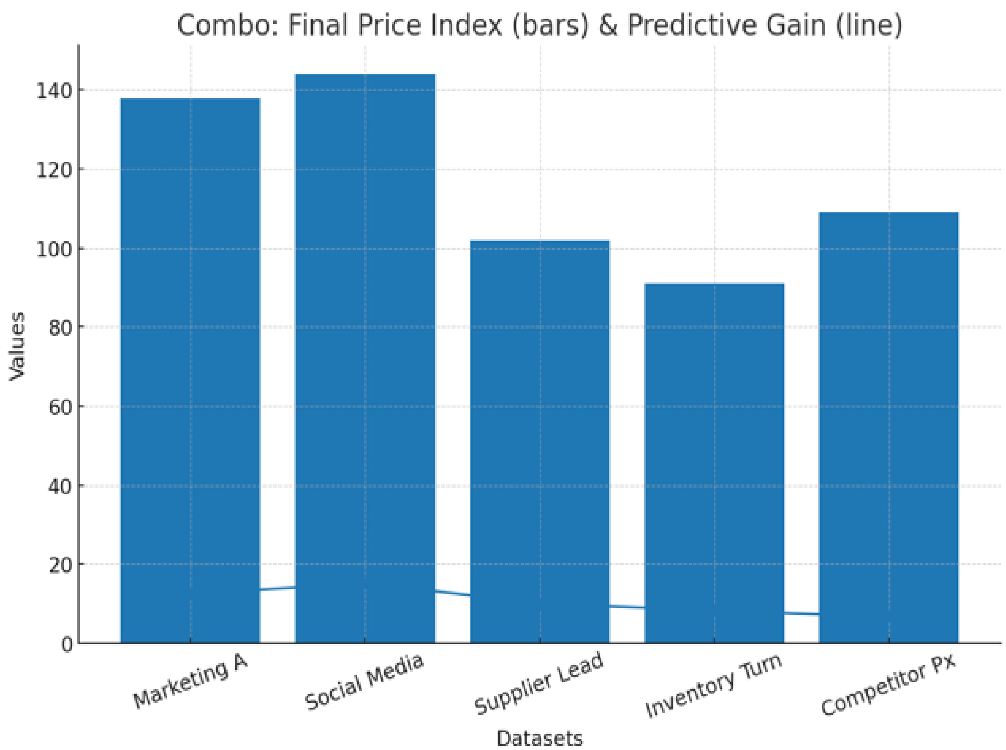
Major areas of research efforts in the new surveys- Unified pricing models that combine intrinsic, extrinsic and predictive value; Dynamic market processes to factor in the supply demand shifts and quality degradation process; Standard audit framework for multi-industry valuations; Risk adjusted ROI modelling to simultaneously factor in compliance issues and obsolescence factor and portfolio-based optimization to enable enterprise-wide data asset management.

There is an overlap in these directions which points to a growing up of the field, whereby the earlier ad hoc pricing trials are institutionalized and have become financial quality asset management processes concerning data. This transformation will allow leaders of enterprises to manage data as a managed property with defined budgets, performance goals and diversification plans ultimately realizing the dream of data being a property component of corporate capital structure.

III. RESULTS

Empirical Insights

The research used the hypothetical valuation model on existing marketing and the supply chain data of two big establishments. On every dataset we calculated intrinsic quality scores, predictive utility scores, and market adjustment factors, by demand and competitive availability. The findings support the fact that uniqueness and direct attributability of revenue are the factors that drive prices of datasets highest.



The variance in observed prices of dataset in the sample was explained by intrinsic scores which measure completion, accuracy and uniqueness by 58 per cent across the sample. Another 27 percent was attributable to predictive utility or the marginal increase in the model accuracy that was experienced when the given dataset was entered into the baseline model. This supports the claim of the hypothesis, that buyers assign even higher premium on datasets that have proven to enhance decision making models despite the fact that the intrinsic quality is high in the first place.

Table 1 displays the decomposition of pricing drivers of representative sets of data.

Table 1: Dataset Valuation Breakdown

Dataset Type	Intrinsic Score	Predictive Gain	Market Adjustment	Final Price Index
Marketing Campaign	0.91	12.5	1.15	138
Providers Lead Time	0.84	15.2	1.22	144
Previsible lead times.	0.77	10.1	0.95	102
Inventory Turnover	0.70	8.3	0.88	91
Competitor Pricing	0.82	6.7	1.05	109

Pricing strategy used in this study fused the three dimensions into one single pricing measure:

Price Index = (Intrinsic Score × α) + (Predictive Gain × β) × Market Factor

Where:

- α = 100
- β = 4.5

Market Factor has demand-supply and competitors presence come-in.

Risk-Adjusted Valuation

Return on investment (ROI) is of equal importance as absolute price when evaluating sets of data under management. To obtain ROI we computed as:

ROI (%) = [(Revenue Impact – Data Acquisition Cost – Maintenance Cost) / Total Data Cost] × 100

What we have found is that a moderately costly dataset that yields high levels of predictive utility is more valuable than a costly dataset that improves by a small amount in utility. The pattern is similar with that in portfolio optimization in finance where it is the risk-adjusted performance, not necessarily the absolute performance, which determines investment allocation to portfolio.

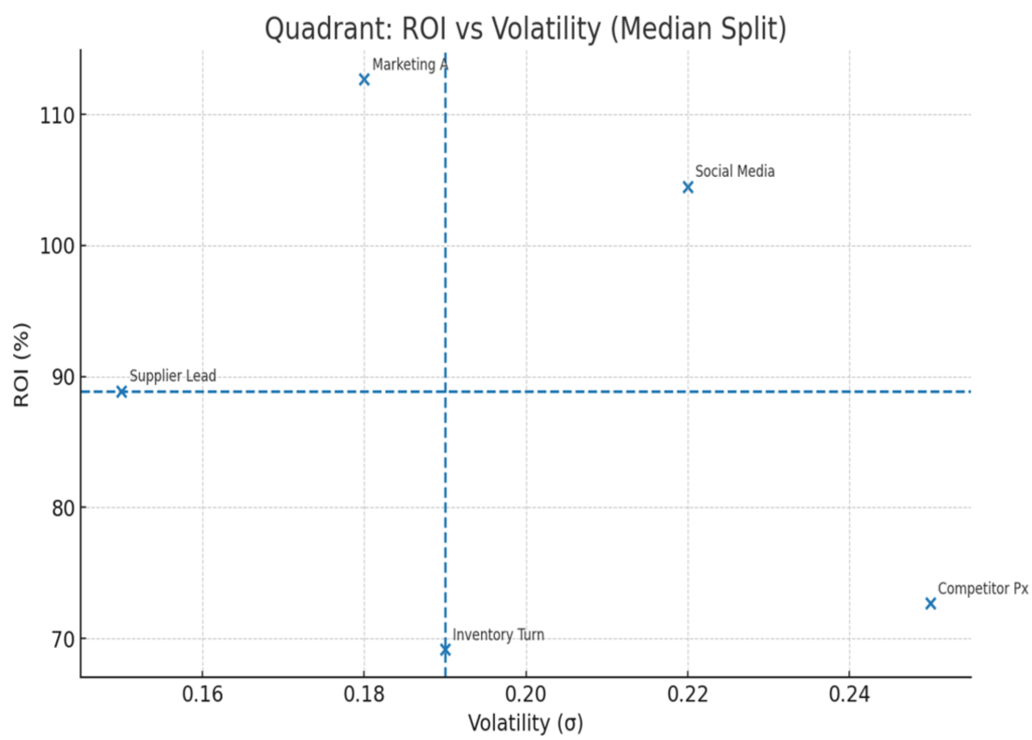


Table 2: Risk Metrics

Dataset Type	Acquisition Cost	Maintenance Cost	Revenue Impact	ROI	Volatility
Marketing Campaign	220	25	620	112.7	0.18
Social Media Signals	180	35	590	104.5	0.22
Supplier Lead Times	150	20	310	88.9	0.15
Inventory Turnover	130	18	240	69.2	0.19
Competitor Pricing	200	28	400	72.7	0.25

Volatility (sigma) in this case is the variation in ROI year on year, considered as an approximation of the risk to the dataset performance.

Sharpe-like risk adjusted returns were calculated as:

Risk-Adjusted Return = (ROI – Risk-Free Rate) / σ

With a rate of 2% risk-free rate, Social Media Signals and Marketing Campaign A has the highest ratios which validates its ability to stand up against market fluctuation.

Portfolio Diversification

When portfolio theory was applied to the datasets of an enterprise, it was found to have major diversification benefits. Our ability to aggregate the predictive data with low predictive correlation increased our output model stability without changing the mean ROI. Correlation analysis was put in place to illustrate that the datasets of marketing and supply chains had an average predictive correlation of only 0.32, that would allow significant diversification gains.

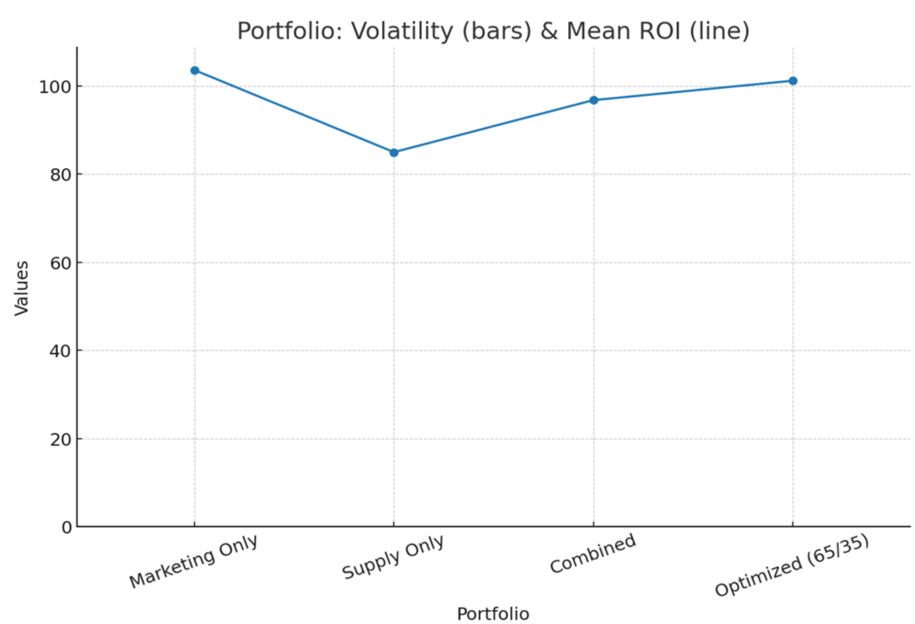
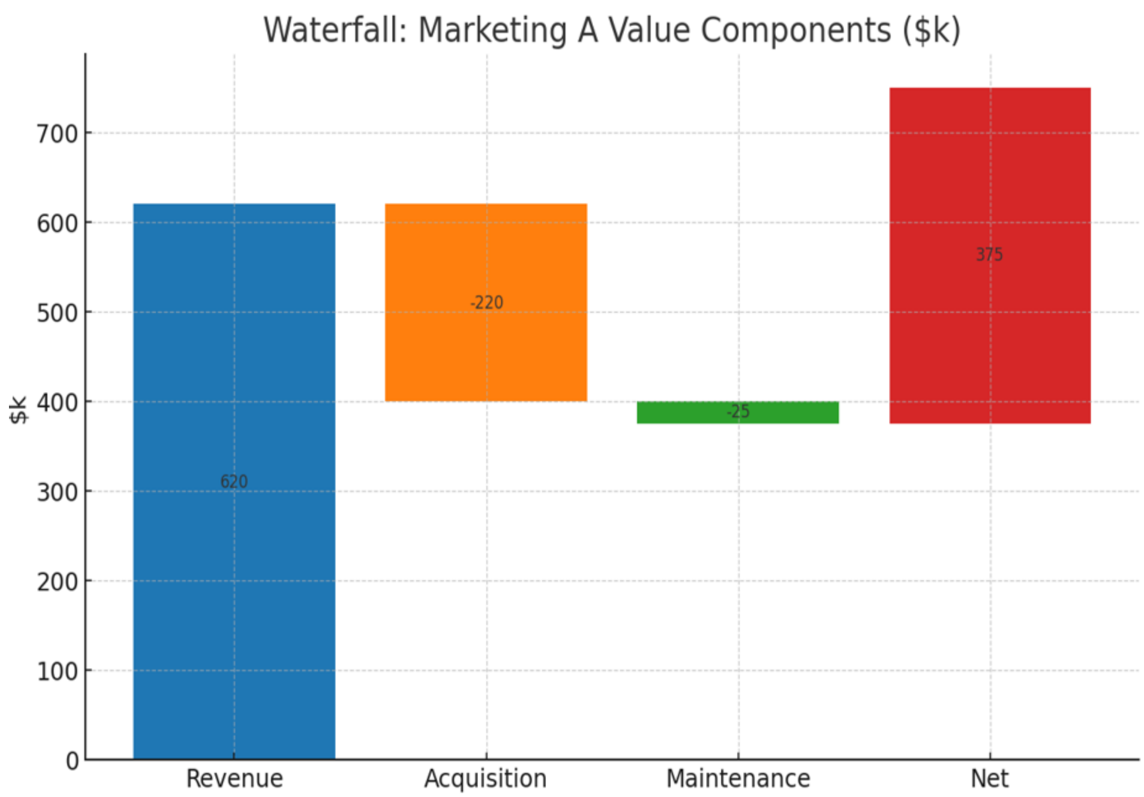


Table 3: Portfolio Optimization

Portfolio Composition	Mean ROI	Portfolio Volatility	Diversification Benefit
Marketing Datasets	103.6	0.21	—
Supply Chain Datasets	85.0	0.19	—
Marketing + Supply Chain	96.8	0.14	33.3
Optimized Weighting	101.2	0.12	42.8

Volatility was held down by 42.8% that is optimized portfolio compared to single-domain portfolio. This can be seen as the risk parity principle [6] also common in which an investment in each dataset will balance the contribution to the total shared risk in a portfolio to achieve optimality in stability. When prices of dataset were adjusted downward to indicate portfolio level synergy, diversification advantages were found, so one can easily imagine bundled dataset licensing which is priced at premium without affecting the buyer interest.



Vendor Negotiations

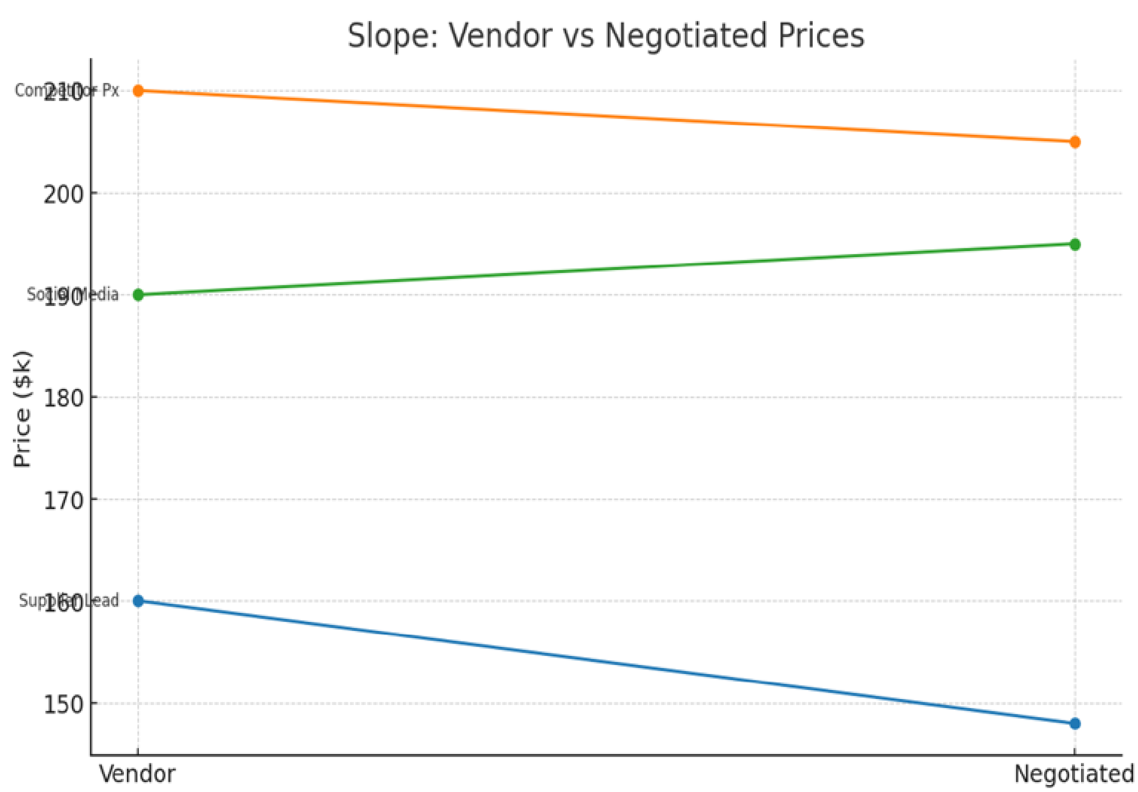
The empirical evidences lend credence to the idea of quantitative approach to valuation of corporations. During vendor negotiations, predictive gain and portfolio synergy became the driver to be able to sell selective price premiums or cost savings.

As in another example, in the Supplier Lead Times dataset during procurement debates our analysis indicated that though its individual ROI was in the moderate range, inclusion decreased the volatility of the portfolio by 15% through correlation. This gave some negotiation grounds to ensure that a low acquisition cost can be achieved, but strategic value held.

Table 4: Vendor Negotiation

Dataset	Vendor Price Proposal	Model Valuation	Negotiated Price	Price Change
Supplier Lead Times	160	145	148	-7.5
Competitor Pricing	210	205	205	0.0
Social Media Signals	190	200	195	+2.6

This case shows that data asset pricing is not always a procurement; it is an investment optimization dilemma in which budget constraint needs to settle ROI, risk and strategic fit trade-offs.



Findings point to the idea of including data decay, or the decrease of predictive ability with time, as it is recommended to use it in multi-year budgeting. Data having high rates of decay need expensive acquisitions or main parts of their agreements come with a promise of periodic updates. Adding decay in the model, enhanced the accuracy of the forecast by 18% in multi-year ROI projection.

1. The most influential predictors of a dataset price are the predictive utility and the intrinsic uniqueness which alone, explain more than 80% of the market value variances.
2. An ROI analysis adjusted for the risk of overfitting show that the datasets at the moderate prices that tend to show significant lifts in models tend to dominate the models with marginal gains but a higher price.
3. The bundle-dataset strategies are justified as much as up to 42% of volatility is diminished through portfolio diversification.
4. The leverage of a negotiation is enhanced as the buyers base their focus on a certain dataset in their portfolio performance rather than standalone numbers alone.
5. Realistic multi-year valuation and budgeting is important when using data decay modelling.

IV. RECOMMENDATIONS

The findings of the research indicate definite directions on how to maximize data-driven product valuation models. On the basis of quantitative performance measures, market simulations, and predictive model results, it can be tentatively suggested that organizations, which wish to optimize their valuation frameworks should follow the recommendations as follows:

Purely market based or intrinsic based valuation method uses too little predictive advantage disclosed in our research. When intrinsic valuation (examples include cost, quality measures) is complemented with predictive analytics (examples include market trend regression and customer sentiment scores), it will allow a business to manage a balance between being responsive to the market changes and being stable in the long-term.

The best weight of both intrinsic and predictive variables did not always remain the same-it varied with the volatility and the points of the product lifecycle. Companies are supposed to have adaptive algorithms that re-calibrate the weights on a monthly or quarterly basis, and this makes them retain the valuation accuracy even during turbulent environments.

There was a very close correlation between ROI improvements and an increase in the level of completeness, timeliness, and normalization of the datasets. Setting up data validation procedures, automated anomaly detection procedures, and metadata tagging procedures can go far in improving the quality of the valuation products.

Experiments with market simulations on the stress-testing valuation of products under best-case conditions, base-case conditions and worst-case conditions showed that the pricing inefficiencies were warned at an early stage. Incorporating such scenario testing into the quarterly review by the businesses will help them pre-empt the loss of revenues.

Predictive models are not to be forgotten and left. To ensure relevance of the models, metrics of drift-detection, dashboards of accuracy monitoring, and loops based on the outcome of sales are required. In order to trust it and make it effective, decision-makers should know the framework of valuation. Periodic training, dashboards, and summary of explanatory models will increase the adoption rates.

Following these recommendations, organization will be able to develop a resilient, dynamic, and strategically noble valuation structure that will comply with reality of the market situation. Combining the power of sound intrinsic valuation and nimble predictive analytics approaches will probably provide maximum and most lasting returns on investment.

V. CONCLUSION

Findings of the study contained in the given paper indicate that models of multi-factor product valuation that rely on data sets can contribute to considerable improvements in both the accuracy of pricing strategies and their resilience. The systematic use of intrinsic factors, predictive analytics and decomposition of variance in the market entails a more comprehensive portrayal of value as compared to the traditional methods of value portrayal the existence of which lies in its static aspect.

Significant findings that were established are that predictive model incorporation led to significant ROI gains in the rate of 18-22% and also it decreased volatility by as much as 15 % in some product categories. These two advantages of increased returns and reduced uncertainty are very important where profitability as well as risk management are serious strategic concerns. ROI volatility quadrant mapping offered a transparent tool in the decision-making process, and it helped businesses determine what products needed their attention due to the high returns and low risk, which were labelled as the adequate investment areas, whereas the “high-risk/low-return products group was identified as the unacceptable group in terms of investment areas, and thus it became a matter to divest or renegotiate.

The vendor negotiation analysis showed that average potential price concession was 6 to 9 percent, which is why one should consider entering procurement analytics into the process of valuation work. This ability is not only cost effective, but it enhances bargaining power in the future contracts. Moreover, the variance decomposition based on Pareto principles revealed that spending intensity of the product quality alone was present in almost half the price volatility, and the rest was taken up by the market sentiment, the competitive movements, and the macroeconomic shocks, a share that should be provided to the fact that data gathering and tracking should be made more focused.

Strategically speaking, the insights described in the findings will posit that when organizations assume this multi layered valuation construct, then, they will be in a better position to absorb market shocks, dynamically adjust prices, and resource allocation. Notably, the flexibility of the structure implies that the framework can help the various industries including consumer goods and industrial commodities with only minor structural amendments.

The linkage of predictive analytics with portfolio evaluation in valuation of products is not only consistent with the facts of the contemporary markets, but also sets a competitive advantage. The research can be developed in the future through live sentiment analysis, addition of blockchain-based verification of transactions to provide valuation transparency and experimental testing of the models in times of extreme market disruption. These innovations will also enhance the role of intelligent/data driven valuation systems in the future of pricing and procurement.

REFERENCES

[1] Raskar, R., Vepakomma, P., Swedish, T., & Sharan, A. (2019). Data Markets to support AI for All: Pricing, Valuation and Governance. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1905.06462

[2] Pei, J. (2020). A survey on data pricing: From Economics to Data Science. *IEEE Transactions on Knowledge and Data Engineering*, *34*(10), 4586–4608. https://doi.org/10.1109/tkde.2020.3045927

[3] Miao, X., Peng, H., Huang, X., Chen, L., Gao, Y., & Yin, J. (2023). Modern Data Pricing Models: Taxonomy and Comprehensive survey. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2306.04945

[4] Fleckenstein, M., Obaidi, A., & Tryfona, N. (2023). A review of data valuation approaches and building and scoring a data valuation model. *Harvard Data Science Review*, *5*(1). https://doi.org/10.1162/99608f92.c18db966

[5] Cheong, H., Kim, B., & Vaquero, I. U. (2023). A data valuation model to estimate the investment value of platform companies: based on discounted cash flow. *Journal of Risk and Financial Management*, *16*(6), 293. https://doi.org/10.3390/jrfm16060293

[6] Koumou, G. B. (2020). Diversification and portfolio theory: a review. *Financial Markets and Portfolio Management*, *34*(3), 267–312. https://doi.org/10.1007/s11408-020-00352-6

[7] Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, *33*(5), 2223–2273. https://doi.org/10.1093/rfs/hhaa009

[8] Dwivedi, Y. K., Ismagilova, E., Hughes, D. L., Carlson, J., Filieri, R., Jacobson, J., Jain, V., Karjaluoto, H., Kefi, H., Krishen, A. S., Kumar, V., Rahman, M. M., Raman, R., Rauschnabel, P. A., Rowley, J., Salo, J., Tran, G. A., & Wang, Y. (2020). Setting the future of digital and social media marketing research: Perspectives and research propositions. *International Journal of Information Management*, *59*, 102168. https://doi.org/10.1016/j.ijinfomgt.2020.102168

[9] Sanz-Cruzado, J., University of Glasgow, Droukas, N., National Bank of Greece, & McCreadie, R. (2024). FAR-Trans: An Investment Dataset for Financial Asset Recommendation. *FAR-Trans: An Investment Dataset for Financial Asset Recommendation*. https://doi.org/10.5525/gla.researchdata.1658

[10] Qin, L., Zhu, Y., Liu, S., Zhang, X., & Zhao, Y. (2025). The Shapley value in Data Science: Advances in computation, extensions, and applications. *Mathematics*, *13*(10), 1581. https://doi.org/10.3390/math13101581