

Research on the Construction of Differentiated Pricing Model for SMEs' Data Assets Based on Machine Learning in the Context of Digital Economy

Yongxia Lv

Faculty of Accounting, Zhengzhou Vocational College of Finance and Taxation, Zhengzhou 450048, Henan, China

Corresponding author: Yongxia Lv (e-mail: lyx507hu@163.com)

ABSTRACT This study focuses on the small sample size and high heterogeneity of data assets owned by small and medium-sized enterprises in the digital economy. Through questionnaire surveys and factor analysis, four core pricing factors are identified, including data characteristics, enterprise heterogeneity, market environment, and technology application. A five-stage pricing framework consisting of data augmentation, feature selection, model training, pricing output, and dynamic adjustment is then constructed. The model integrates CTGAN-based data augmentation and the XGBoost algorithm, and embeds an enterprise-heterogeneity adaptation layer to achieve differentiated pricing. Empirical results show that the proposed model performs well, with MAE of 0.072, RMSE of 0.095, and R^2 of 0.83. The pricing error rate is below 10%, significantly outperforming traditional methods and undifferentiated machine-learning benchmarks. The model also demonstrates good effectiveness and stability. This research provides operable models and tools for accurate pricing of SMEs' data assets and offers a technical reference for promoting standardized development of the data-factor market.

INDEX TERMS digital economy, smes, data assets, differentiated pricing, machine learning

I. INTRODUCTION

A. RESEARCH BACKGROUND AND SIGNIFICANCE

1) Research Background

IN the era of digital economy, the connotation and extension of production factors are undergoing profound changes as automated data and smart fabric sensor information evolve into critical strategic assets. These digital advancements redefine the traditional boundaries of manufacturing, transforming raw material inputs into intelligence-driven industrial resources. As a new type of production factor, data has been incorporated into the scope of enterprise asset management, and its value creation model is significantly different from that of traditional production factors [1]. The Opinions of the CPC Central Committee and the State Council on Establishing a Data Foundation System to Better Play the Role of Data Elements (the "Data Twenty Articles") propose establishing and improving the data property rights system and promoting innovation in data asset pricing mechanisms [2]. As an important part of the national economy, SMEs, although facing certain gaps compared with large enterprises in terms of capital scale and brand influence, can leverage data assets to explore development opportunities in specific market segments. However, the current pricing of SMEs' data assets faces several challenges, including ambiguous definitions of data

ownership, lack of standardized pricing frameworks, strong subjectivity in value assessment, and limited adaptability of traditional pricing methods to inherent characteristics such as non-consumability (i.e., data is not depleted through use) and network effects [1-3].

Meanwhile, machine learning technology has demonstrated remarkable effectiveness in price prediction, value evaluation and other fields relying on its powerful data processing and prediction capabilities [4,5]. Algorithms such as XGBoost, CatBoost, and Random Forest have been successfully applied in pricing practices in scenarios such as financial assets, providing a new path for solving the problem of differentiated pricing of data assets [6,7].

In this context, using machine learning technology to construct a differentiated pricing model for data assets adapted to textile manufacturing SMEs is the key to breaking the bottleneck of market-oriented allocation of data factors and enhancing the core competitiveness of the apparel industry. By accurately valuing the diverse data streams from automated loom operations and fabric quality inspections, this approach empowers small enterprises to capitalize on their unique industrial intelligence.

2) Theoretical Significance

(1) Enrich the theoretical system of data asset pricing and break through the adaptation limitations of traditional pricing theories to the characteristics of SMEs' data assets;

(2) Expand the application boundary of machine learning technology in the field of pricing and improve the theoretical foundation of data-driven pricing;

(3) Fill the research gap in the pricing of SMEs' data assets.

3) Practical Significance

(1) Provide guidance for SMEs to accurately evaluate the value of their data assets and avoid pricing risks;

(2) Provide standardized pricing tools for data trading platforms and improve transaction efficiency;

(3) Provide decision-making basis for the government to formulate regulatory policies for the data factor market and improve the digital support system for SMEs.

Review of Domestic and Foreign Research Status

Research on Digital Economy and Data Assets Foreign scholars have conducted in-depth research on the characteristics of data assets. Although their perspectives differ, most studies recognize that data assets possess multi-dimensional attributes, including volume, variety, velocity, value, and accuracy [8]. Compared with traditional data, big data has the advantages of large volume, diversity, and high velocity, providing enterprises with a broader data foundation [9]. In China, data assets are regarded as an important branch of information assets, and the concept of data assets was first defined based on the concept of information assets [10]. In subsequent studies, scholars generally recognize that data assets have unique attributes such as timeliness and network effects, and their pricing logic is essentially different from that of traditional assets [11,12].

Research on Data Asset Pricing Theories and Methods

Existing pricing methods are divided into two categories: traditional methods (cost method, income method, market method) and intelligent algorithms. Traditional methods have limitations such as underestimating reuse value and insufficient adaptability; intelligent algorithm-based pricing has become a hot spot. Algorithms such as XGBoost and KNN have been used to explore differentiated pricing, with the core logic of achieving accurate matching between value and price through multi-dimensional integration [13,14].

Research on the Application of Machine Learning in Pricing Machine learning realizes pricing applications through a closed loop consisting of feature learning, model training, and prediction output, and has achieved remarkable results in scenarios such as online short-term rentals and financial assets. For the small sample problem in data asset pricing, combined models of data augmentation technologies such as CTGAN (conditional tabular GAN) and machine learning have shown good adaptability [15,16].

Research on SMEs' Data Assets Existing studies have clarified the characteristics of SMEs' data assets, such as small scale, uneven quality, and distinct industry attributes

[17,18], but their value evaluation and pricing still lack targeted tools. Most studies have not deeply integrated machine learning technology, and no operable differentiated pricing scheme has been formed yet.

B. RESEARCH CONTENT

This study focuses on the core issue of differentiated pricing of SMEs' data assets and is structured as follows:

(1) Analyze the current status and core problems of SMEs' data asset pricing through surveys, and identify key influencing factors;

(2) Integrate the dimension of enterprise heterogeneity to construct a differentiated pricing model based on machine learning;

(3) Verify the effectiveness of the model through empirical tests and put forward targeted countermeasures and suggestions.

II. ANALYSIS OF THE CURRENT STATUS AND PROBLEMS OF SMES' DATA ASSET PRICING IN THE CONTEXT OF DIGITAL ECONOMY

A. SURVEY ON THE DEVELOPMENT AND PRICING STATUS OF SMES' DATA ASSETS

1) Survey Design and Data Sources

To fully grasp the current status of SMEs' data asset pricing, this study adopts a combination of questionnaire surveys and in-depth interviews.

The survey covers six industries, including manufacturing, retail, and service. A total of 326 valid questionnaires were collected, and 28 in-depth interviews were conducted.

The survey content includes the stock, type, management status, pricing practice, and existing difficulties of data assets.

2) Stock, Type and Management Status of SMEs' Data Assets

Survey results show that:

(1) The stock of data assets of SMEs is generally low, with only 35.6% of enterprises having standardized data asset libraries;

(2) Data types are mainly operational data (accounting for 42.3%) and customer data (accounting for 31.6%), while knowledge data and transaction data account for less than 20%;

(3) The management status is not optimistic. Only 28.8% of enterprises are equipped with professional data management teams, 67.2% of enterprises have problems such as incomplete data cleaning and non-standard annotation, and the data quality compliance rate (according to GB/T36325 standard) is only 41.5%. The characteristics of data assets of SMEs in different industries are significantly different: manufacturing enterprises are dominated by operational data, while retail enterprises are dominated by customer data, which also determines the heterogeneity of their pricing needs.

3) Current Pricing Practice Models of SMEs' Data Assets

The current pricing practice of SMEs' data assets is mainly based on traditional methods, and the specific model distribution is shown in Table 1.

TABLE 1. Distribution of Pricing Models for SMEs' Data Assets

Pricing model	Application proportion	Core practices	Applicable scenarios
Cost-plus pricing	45.1%	Adding 20%-50% on the basis of the full-life-cycle cost of data	Internal data asset value evaluation
Reference pricing	28.5%	Referring to the price of similar products on data exchanges and making appropriate adjustments	Data transactions with high standardization
Negotiated pricing	21.4%	Both parties determine the price through negotiation based on data usage and scale	Customized data services
Intelligent pricing	5.0%	Predicting pricing with simple algorithm models	High-value data assets of large SMEs

B. CORE PROBLEMS EXISTING IN SMEs' DATA ASSET PRICING

1) Lack of Pricing Standards and Strong Subjectivity in Value Evaluation

At present, there is no unified pricing standard for SMEs' data assets. Enterprises mostly conduct evaluations based on their own experience, leading to significant pricing differences. The survey found that:

(1) For customer data of the same type and quality, the pricing difference among different enterprises can reach 3-5 times;

(2) Some enterprises take data collection cost as the only pricing basis, ignoring data processing capabilities and application value, leading to a serious deviation between pricing and actual value;

(3) The adaptation limitations of traditional pricing theories in the field of data assets further exacerbate the subjectivity of pricing.

2) Ignoring Enterprise Heterogeneity and Insufficient Differentiated Pricing

Existing pricing practices largely adopt a one-size-fits-all model, failing to adequately account for the heterogeneous characteristics of SMEs, such as industry attributes, firm size, and data application scenarios. For instance, operational data from manufacturing SMEs and customer consumption data from service-oriented SMEs are often priced using the same logic, which does not reflect their industry-specific value. In addition, differences in data management capabilities across firms of varying sizes are frequently overlooked, resulting in the undervaluation of high-quality data assets by smaller enterprises. Effective differentiated pricing requires aligning data value with the specific characteristics of different entities; the absence of such differentiation fundamentally constrains pricing rationality.

3) Outdated Pricing Methods and Difficulty in Adapting to the Dynamic Value Characteristics of Data Assets

Traditional pricing methods are difficult to capture the dynamic value changes of data assets, such as the impact of factors such as data timeliness and network effects on value. The survey shows that:

(1) 76.3% of enterprises reflect that traditional pricing methods cannot cope with the problem of data asset value attenuation over time;

(2) Traditional methods cannot quantify the value brought by the increase in users, leading to rigid pricing. In contrast, machine learning technology has dynamic fitting and prediction capabilities, which can better match the value characteristics of data assets, but its application rate among SMEs is currently extremely low.

C. IDENTIFICATION OF INFLUENCING FACTORS FOR DIFFERENTIATED PRICING OF SMEs' DATA ASSETS

Factor analysis is used to identify the key influencing factors for differentiated pricing of SMEs' data assets. A total of 4 common factors are extracted, with a cumulative variance explanation rate of 78.6%. The specific influencing factors and weights are shown in Table 2.

TABLE 2. Results of Factor Analysis on Influencing Factors of SMEs' Data Asset Pricing

Common factor	Specific influencing factors	Weight	Influence direction
Data characteristic factor	Data quality, data quantity, timeliness, completeness	0.32	Positive
Enterprise heterogeneity factor	Industry attributes, enterprise scale, data management capabilities, compliance level	0.28	Positive (significant industry differences)
Market environment factor	Market supply and demand relations, competitive landscape, industry data scarcity	0.22	Positive (the greater the supply-demand gap, the more significant the impact)
Technology application factor	Data processing technology, algorithm adaptability, data security technology	0.18	Positive

The data characteristic factor plays a dominant role in pricing, with data quality exerting the most significant influence. Within the enterprise heterogeneity dimension, pricing coefficients for data-intensive industries such as manufacturing and finance are significantly higher than those of traditional industries. Regarding the market environment dimension, data assets in industries characterized by high scarcity (e.g., healthcare and high-end manufacturing) exhibit a clear pricing premium.

These factors provide an important foundation for constructing the model's variable system, particularly for initial variable screening and preliminary feature importance assessment, rather than replacing the adaptive feature learning mechanism of XGBoost.

III. CONSTRUCTION OF DIFFERENTIATED PRICING MODEL FOR SMES' DATA ASSETS BASED ON MACHINE LEARNING

A. MODEL CONSTRUCTION OBJECTIVES AND CORE ASSUMPTIONS

1) Model Construction Objectives

The primary objectives of the proposed model are to address the challenges of inaccurate pricing, insufficient differentiation, and limited dynamic adaptability in the valuation of SMEs' data assets, thereby achieving accurate, differentiated, and dynamically adjustable pricing.

Specifically, the model aims to (1) improve pricing accuracy by reducing the deviation between predicted and actual values to below 10%; (2) enable multi-dimensional differentiation by incorporating data characteristics and enterprise heterogeneity; and (3) enhance dynamic adjustment capabilities by allowing real-time pricing updates in response to changes in data timeliness and market conditions.

2) Core Assumptions and Constraints of the Model

Core assumptions of the model:

(1) The value of data assets has a non-linear correlation with influencing factors, which is compatible with the fitting characteristics of machine learning algorithms;

(2) The impact of enterprise heterogeneity on pricing has industry specificity, and the data asset value logic of SMEs in different industries is different;

(3) Within the scope of legal and compliant market transactions, data quality and compliance are prerequisites for pricing, and non-compliant data assets have zero legal value (the model does not cover grey/black market transactions, which involve illegal risks beyond the research scope).

Constraints:

(1) Data sources are legal and compliant, in line with the requirements of laws and regulations such as the Personal Information Protection Law and the Data Security Law;

(2) The data of model input variables is accessible and quantifiable;

(3) Pricing results need to be within the market-acceptable range to avoid extreme pricing.

B. CONSTRUCTION OF MODEL INPUT VARIABLE SYSTEM

1) Selection and Definition of Variables Based on Influencing Factors

Based on the key influencing factors identified in Chapter Stock, Type and Management Status of SMEs' Data Assets, core input variables were selected to construct a four-dimensional variable system comprising data characteristics, enterprise heterogeneity, market environment, and technology application. The core attributes of data assets (such as non-consumability and timeliness) have been integrated into the design of data characteristic variables, and enterprise heterogeneity is directly quantified through variables such as industry type and scale. The specific variable definitions and quantification methods are as follows (Table 3):

2) Variable Preprocessing

Variable preprocessing includes three steps:

(1) Data cleaning: Adopt methods such as deleting outliers (3σ principle), filling missing values (mean imputation method), and deduplication to improve data quality;

(2) Data normalization: Use the Min-Max standardization method to map all variables to the [0,1] interval to avoid the impact of dimensional differences on model training;

(3) Feature engineering: Screen important features through Lasso regression to eliminate redundant variables and reduce model computation; at the same time, construct interaction terms using polynomial features to capture the non-linear relationships between variables.

This process fully considers the requirements of machine learning algorithms for input data and lays the foundation for model training.

C. ARCHITECTURE DESIGN OF DIFFERENTIATED PRICING MODEL

1) Construction of Overall Model Framework

This model adopts a five-stage architecture consisting of data augmentation, feature selection, model training, pricing output, and dynamic adjustment.

(1) The first stage: Adopt CTGAN generative adversarial network for data augmentation to solve the problem of insufficient data samples of SMEs—this is a key design targeting the characteristic of SMEs' data assets;

(2) The second stage: Screen core features through Lasso regression and feature importance analysis;

(3) The third stage: Construct an XGBoost main model and a Random Forest auxiliary model for parallel training. The non-linear fitting ability and anti-overfitting characteristics of the XGBoost algorithm can accurately capture the complex relationships of pricing influencing factors;

(4) The fourth stage: Combine the output results of the main and auxiliary models to generate benchmark pricing and differentiated adjustment coefficients;

(5) The fifth stage: Establish a dynamic adjustment mechanism based on changes in data timeliness and market environment to update pricing results in real time and adapt to the dynamic value characteristics of data assets.

2) Embedding of Differentiated Dimensions Based on Enterprise Heterogeneity

To achieve differentiated pricing, an enterprise heterogeneity adaptation layer is embedded in the model architecture. The specific measures include:

(1) Construct segmented models according to industry types, and train dedicated pricing sub-models for different industries such as manufacturing, retail, and service industries to adapt to industry heterogeneity;

(2) Introduce interaction terms between enterprise scale and compliance level to quantify the pricing differences of enterprises of different sizes in terms of compliance;

(3) Set industry-specific pricing coefficients, and assign differentiated weights to different industries according to

TABLE 3. Definition and Quantification Methods of Input Variable System for Differentiated Pricing Model

Variable dimension	Variable name	Variable definition	Quantification method
Data characteristics	Data quality score	Comprehensive score of data completeness, accuracy, and consistency	1-10 points scale according to GB/T36325 standard
Data characteristics	Data scale	Number of valid records of data assets	Logarithmic transformation: $\ln(\text{number of records} + 1)$
Enterprise heterogeneity	Timeliness	Data update frequency and time interval	Daily update = 10, weekly update = 7, monthly update = 4, quarterly update = 1
Enterprise heterogeneity	Data type	Type of data assets	Customer data = 4, operational data = 3, knowledge data = 5, transaction data = 4 (dummy variable)
Enterprise heterogeneity	Industry type	Industry to which the enterprise belongs	Data-intensive industry = 5, general industry = 3, traditional industry = 1 (dummy variable)
Enterprise heterogeneity	Enterprise scale	Number of employees and revenue scale of the SMEs	Large SMEs = 4, medium-sized SMEs = 3, small SMEs = 2, micro-enterprise sized = 1
Market environment	Supply-demand gap	Ratio of market supply to demand for similar data	Demand/supply; the larger the ratio, the greater the gap
Market environment	Competition level	Number of suppliers of similar data assets in the market	The fewer the suppliers, the lower the competition level, and the higher the assignment (1-10 points)
Market environment	Industry prosperity	Development prosperity of the industry to which the enterprise belongs	Refer to industry prosperity index, 1-10 points scale
Technology application	Data processing technology level	Advanced nature of data cleaning, integration, and analysis technologies adopted by the enterprise	Advanced technology = 8-10, medium = 4-7, basic = 1-3
Technology application	Security technology investment	Proportion of investment in data security protection technology	Investment proportion $\times 100$ (standardized to 1-10 points)

industry data scarcity and value density to ensure that pricing results are in line with industry reality. This design directly addresses the core problem of ignoring enterprise heterogeneity identified in Chapter 2.

D. SELECTION OF CORE ALGORITHMS AND MODEL TRAINING

1) Comparison of Candidate Algorithms: Regression Algorithms, Tree Models, Neural Networks, etc.

The adaptability of four types of candidate algorithms in the scenario of SMEs' data asset pricing is compared and analyzed, and the results are as follows:

(1) Linear regression model is simple to calculate but cannot capture non-linear relationships, with a low fitting degree ($R^2 \approx 0.52$);

(2) Neural network model has strong fitting ability but requires a large number of samples and is prone to overfitting in small sample scenarios;

(3) Random Forest model has strong stability and good interpretability but limited prediction accuracy;

(4) XGBoost algorithm has the advantages of strong non-linear fitting ability, anti-overfitting, and good adaptability to small samples, and performs optimally in indicators such as RMSE and MAE. Therefore, XGBoost is selected as the main model and Random Forest as the auxiliary model.

2) Construction and Division of Training Dataset

The final dataset consists of 1,326 samples, including 326 real observations collected from SMEs and 1,000 synthetic samples generated using CTGAN. To ensure the synthetic data captures genuine market dynamics, CTGAN is constrained by industry-specific market rules and real transaction price ranges during generation, and synthetic samples that deviate from realistic market logic are filtered out.

The dataset was divided into a training set (928 samples) and a validation set (398 samples) at a ratio of 7:3. In

addition, 5-fold cross-validation was applied to enhance model generalization.

3) Model Training Process and Parameter Optimization

The model training process includes three steps: parameter initialization, iterative training, and parameter optimization:

(1) Set initial XGBoost parameters (learning rate = 0.1, max depth = 6, estimators = 100);

(2) Conduct iterative training using the training set and monitor model performance through the validation set;

(3) Optimize parameters using the grid search method, and finally determine the optimal parameter combination as: learning rate = 0.08, max depth = 5, estimators = 150, regularization parameter $\lambda = 0.1$. The optimized model achieves an R^2 of 0.89 on the training set and 0.82 on the validation set, both of which are significantly higher than those of the initial model.

E. MODEL OUTPUT AND FORMULATION OF PRICING STRATEGIES

1) Interpretation of Model Output Results: Benchmark Pricing and Differentiated Adjustment Coefficients

The model output results include two parts:

(1) Benchmark pricing: Calculated by the XGBoost main model based on core variables, reflecting the basic value of data assets;

(2) Differentiated adjustment coefficient: Calculated by the Random Forest auxiliary model based on enterprise heterogeneity and market environment variables, with a value range of [0.8, 1.5].

Final pricing = Benchmark pricing $\times$ Differentiated adjustment coefficient. Among them, the industry difference adjustment coefficient has the largest fluctuation (manufacturing industry = 1.2–1.5, traditional industry = 0.8–1.0), reflecting the significant impact of enterprise heterogeneity on pricing.

2) Formulation of Scenario-Based Differentiated Pricing Strategies

Based on the model output results, three types of scenario-based differentiated pricing strategies are formulated:

(1) High-value data asset strategy: For knowledge data, high-quality customer data, and other high-value data types, benchmark pricing is multiplied by 1.2–1.5 to reflect scarcity and application value;

(2) Standardized data asset strategy: For highly standardized data such as operational data, benchmark pricing is multiplied by 0.9–1.1 to enhance market competitiveness;

(3) Long-tail data asset strategy: For fragmented and small-scale data, benchmark pricing is multiplied by 0.8–0.9 to facilitate value realization through bulk transactions.

IV. EMPIRICAL TEST AND RESULT ANALYSIS OF THE MODEL

A. DESIGN OF EMPIRICAL TEST

Selection of Test Samples and Data Preprocessing 60 SMEs of different industries and sizes are selected as test samples, covering 6 industries including manufacturing, retail, and service industries, among which 15 are large SMEs, 18 are medium-sized SMEs, and 27 are small SMEs.

Data on 120 data assets of sample enterprises are collected, and after data cleaning, normalization and other preprocessing, a test data set is formed. At the same time, the actual transaction prices on data trading platforms are collected as real value references.

B. SETTING OF TEST INDICATORS

Four core test indicators are set:

(1) Mean Absolute Error (MAE): Measures the absolute deviation between pricing results and real value;

(2) Root Mean Square Error (RMSE): Magnifies the impact of large deviations and reflects the accuracy of the model;

(3) Coefficient of Determination (R^2): Measures the model's ability to explain pricing laws;

(4) Pricing Rationality Score: Rated by 3 experts in the field of data assets on a 1-10 points scale, taking the average value.

C. COMPARISON BENCHMARKS

Two comparison benchmarks are set:

(1) Traditional pricing method: Select the cost-plus pricing method, which is the most widely used in the survey;

(2) Undifferentiated machine learning model: XGBoost model without embedding the enterprise heterogeneity dimension.

(3) Simple non-linear model: Support Vector Machine (SVM) with RBF kernel, a classic non-linear algorithm for comparison.

By comparing with the two benchmark models, the superiority of the differentiated pricing model constructed in this study is verified.

D. PROCESS AND RESULT PRESENTATION OF EMPIRICAL TEST

The test data set is input into this model, the traditional pricing model, and the undifferentiated machine learning model respectively to obtain three types of pricing results. Compared with the real transaction prices, the test indicator results are shown in Table 4.

TABLE 4. Comparison of Test Indicators of Different Pricing Models

Test indicators	Differentiated model in this study	Undifferentiated machine learning model	SVM model (non-linear)	Traditional pricing model
MAE	0.072	0.115	0.098	0.238
RMSE	0.095	0.142	0.126	0.296
R^2	0.83	0.68	0.75	0.45
Pricing rationality score	8.6	7.2	7.8	5.3

Industry-specific test results indicate that the model performs best in data-intensive industries, such as manufacturing and finance, achieving an MAE ranging from 0.058 to 0.065.

Although performance in traditional industries is slightly lower, it remains significantly superior to the two benchmark models, demonstrating the model's adaptability across different industry contexts.

E. VERIFICATION OF MODEL EFFECTIVENESS AND STABILITY

1) Effectiveness Analysis: Matching Degree Between Pricing Results and Actual Value

Effectiveness analysis results show that the degree of alignment between the pricing results of this model and the real transaction prices reaches 89.2%, among which 76.7% of the samples have a pricing error rate of less than 10%, and only 3.3% of the samples have an error rate exceeding 20%.

SHAP value analysis indicates that data quality, industry type, and supply-demand gap are the three most influential factors, explaining 28%, 22%, and 18% of the variance, respectively. These findings are consistent with the results of influencing factor identification in Chapter 2, further verifying the effectiveness of the model.

2) Stability Analysis: Performance Under Different Data Samples and Time Dimensions

Stability analysis adopts two groups of experiments:

(1) Replacing test samples (selecting 40 new SMEs), the model test indicators are MAE = 0.078, RMSE = 0.102, R^2 = 0.80, with little difference from the original test results;

(2) Time dimension verification: After tracking the price changes of data assets over a six-month period, the degree of alignment between dynamically adjusted pricing results and actual prices still reaches 85.6%. The two groups of experiments show that the model has good stability and can adapt to changes in different samples and time.

F. DISCUSSION ON MODEL OPTIMIZATION DIRECTIONS

Although the model shows good effectiveness and stability, there is still room for optimization:

- (1) The scope of sample coverage can be further expanded to include more SMEs in central and western regions, improving the regional adaptability of the model;
- (2) Blockchain technology can be introduced to realize the traceability of data asset ownership and transaction records, enhancing the credibility of pricing;
- (3) Real-time market data can be integrated to further enhance the dynamic adjustment ability of the model and improve the response speed to market changes.

V. RESEARCH CONCLUSIONS AND COUNTERMEASURES AND SUGGESTIONS

A. MAIN RESEARCH CONCLUSIONS

This study explores the issue of differentiated pricing of data assets for manufacturing SMEs in the context of the digital economy and draws the following main conclusions regarding the valuation of fabric production intelligence and digitized design resources:

- (1) The empirical analysis confirms that pricing is primarily influenced by data characteristics, enterprise heterogeneity, market conditions, and technology application, among which data quality and industry attributes have the most significant impacts;
- (2) The current pricing of SMEs' data assets faces several challenges, including lack of standards, insufficient differentiation, and outdated methods, while traditional pricing approaches struggle to adapt to the dynamic value characteristics of data assets;
- (3) The constructed differentiated pricing model based on XGBoost has good effectiveness and stability, with significantly higher pricing accuracy than traditional pricing methods and undifferentiated machine learning models, achieving multi-dimensional differentiated pricing;
- (4) Scenario-based pricing strategies can effectively improve the targeting of pricing, and different types of data assets should adopt differentiated pricing logics.

B. COUNTERMEASURES AND SUGGESTIONS FOR DIFFERENTIATED PRICING OF SMES' DATA ASSETS

1) Suggestions for SMEs

- (1) Strengthen data asset management, establish standardized data asset libraries, improve data quality, focus on improving the whole-process management of data collection, cleaning, and annotation, and ensure data compliance;
- (2) Actively apply differentiated pricing models, select appropriate pricing strategies according to their own industry attributes and data types, adopt premium pricing for high-value data assets, and stable pricing for standardized data assets;
- (3) Strengthen data security and compliance construction, obtain relevant compliance certifications, and enhance the market recognition and pricing premium of data assets.

2) Suggestions for the Government and Industry Associations

- (1) Accelerate the formulation of pricing standards for SMEs' data assets, clarify pricing principles, core indicators, and methods, and provide reference basis for enterprises' pricing;
- (2) Strengthen policy support, encourage SMEs to transform data into recognized assets, and provide financial subsidies and technical guidance for data management and pricing model implementation;
- (3) Improve the supervision of the data factor market, standardize data transaction behaviors, establish a mechanism for defining data asset ownership and resolving disputes, and protect the rights and interests of SMEs' data assets;
- (4) Build data trading service platforms, gather SMEs' data asset resources, improve market liquidity, and provide more comparable cases for pricing.

AUTHOR CONTRIBUTIONS

Yongxia Lv designed, collected and analyzed the data, and drafted the manuscript. Yongxia Lv conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Yongxia Lv participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research was supported by, 1.The 2021 research and practice project of higher education reform in Henan Province "Internet +Accounting Factory" research and practice of Digital Intelligence education model in vocational school under the background of integration of production and education " (project number: 2021SJGLX910).

2.The 2023 research project of Henan Federation of Social Sciences" Internet +Accounting Factory "innovation research of education model in vocational school under the age of Digital Intelligence education" (project number: SKL-2023-1142).

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

- [1] J. Lin, "Innovative Data Asset Capitalization and Digital Supply Chain Finance for SMEs: A case study of Uqian Tech," SHS Web of Conferences, vol. 207, Art. no. 03011, 2024, doi: 10.1051/shsconf/202420703011.
- [2] Opinions of the CPC Central Committee and the State Council on Establishing a Data Foundation System to Better Play the Role of Data Elements, "Scientific Chinese," no. 02, pp. 60-64, 2023.

- [3] Y. Zhou, P. Li, and H. F. Zhang, "Research on the Recognition, Measurement and Accounting of SMEs' Data Assets in the Context of Digital Economy," *Fortune Today*, no. 01, pp. 97-99, 2025.
- [4] H. Korkmaz, "Prediction of Airline Ticket Price Using Machine Learning Method," *Journal of Transportation and Logistics*, vol. 9, no. 2, pp. 205-218, 2024, doi: 10.26650/jtl.2024.1486696.
- [5] Y. Qin and H. M. Kong, "Stock Price Prediction Based on Machine Learning Algorithms," *Journal of Shantou University (Humanities & Social Sciences Edition)*, vol. 41, no. 07, pp. 25-34, 94-95, 2025, doi: 10.3969/j.issn.1001-4225.2025.07.004.
- [6] C. L. Wang, "Research on Pricing Model of Consumer Finance Asset Securitization Products Based on Machine Learning," Guangzhou, China: South China University of Technology; 2024, doi: 10.27151/d.cnki.ghnlu.2024.005103.
- [7] Y. K. Xu, J. Yu, C. Jiang, J. Q. Wang, and W. Q. Zhao, "Machine Learning-Driven Power Trading Data Analysis and Strategy Optimization," *Microcomputer Applications*, vol. 41, no. 09, pp. 59-61, 2025, doi: 10.3969/j.issn.1007-757X.2025.09.016.
- [8] E. Dumbill, "3d data management: Controlling data volume, velocity and variety," *gartner*. [2026-02-06], doi: 10.1089/big.2013.1512.
- [9] H. Hassani and E. S. Silva, "Forecasting with Big Data: A Review," *Annals of Data Science*, vol. 2, no. 1, pp. 5-19, 2015, doi: 10.1007/s40745-015-0029-9.
- [10] H. Y. Wang and W. D. Chen, "Research on the Definition and Evaluation Methods of Information Assets," *Journal of Southeast University (Philosophy and Social Sciences Edition)*, no. S2, pp. 66-68, 72, 2001, doi: 10.13916/j.cnki.issn1671-511x.2001.s2.020.
- [11] X. L. Wang, A. D. Xu, and X. Y. Song, "Data Networking: Spillover Effects of Data Assets in Supply Chain Networks," *Research on Economics and Management*, vol. 46, no. 12, pp. 57-73, 2025, doi: 10.13502/j.cnki.issn10007636.2025.12.004.
- [12] G. H. Feng, Z. X. Zhang, and W. Zheng, "Core Issues of Data Assetization: Valuation Challenges, Property Rights Conflicts and Ethical Governance," *Network Security and Data Governance*, vol. 44, no. 10, pp. 16-22, 2025, doi: 10.19358/j.issn.20971788.2025.10.003.
- [13] L. Zeng, Y. Gao, and H. Hu, "Research on Retail Pricing Strategy of Supermarket Fresh Products Based on XGBoost Model," *Academic Journal of Computing & Information Science*, vol. 7, no. 10, pp. 82-88, 2024, doi: 10.25236/ajcis.2024.071012.
- [14] Y. C. Peng, X. Q. Ren, and H. S. Shi, "Design of Crowdsourcing Platform Pricing System Based on Kernel-Weighted KNN and Multi-Objective Optimization," *Software*, vol. 39, no. 06, pp. 150-154, 2018, doi: 10.3969/j.issn.1003-6970.2018.06.033.
- [15] J. X. Shen, S. N. Zhang, and F. W. Zeng, "Data Resource Pricing Method Based on Data-Augmented Small Sample Learning," *Computer Simulation*, vol. 42, no. 06, pp. 78-85, 2025, doi: 10.3969/j.issn.1006-9348.2025.06.013.
- [16] L. Xu, M. Skoularidou, A. Cuesta-Infante, and K. Veeramachaneni, "Modeling Tabular Data Using Conditional GAN," *arXiv Preprint arXiv:1907.00503*, 2019, doi: 10.48550/arXiv.1907.00503.
- [17] L. Shang, W. Huang G, and W. Xin, "Practical Needs and Promotion Paths of Digital Transformation of SMEs in China—Taking 27 SMEs in Zhuhai as Examples," *Economic and Social Development*, vol. 22, no. 03, pp. 12-19, 2024, doi: 10.16523/j.45-1319.2024.03.002.
- [18] C. L. Li, "Research on Data Asset Operation and Management Modes in the Digital Transformation of SMEs," *China Electronic Business Journal*, vol. 31, no. 17, pp. 109-111, 2025, doi: 10.3969/j.issn.1006-6675.2025.17.37.