Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 12
ESAC 2026

Natural Gas Flow Metering Error Compensation Model Based on XGBoost-Stacking Ensemble Learning

Jianfu Ren *, Jiaxu Shi, Huayu Yuan, Gencai Zhang, Haitao Li, Zeyu Wang

National Pipeline Corporation Beijing Pipeline Co, Ltd, Beijing, 100101, China,

* Corresponding Author Email: renjianfu2023@163.com

Abstract. With the rapid growth of natural gas consumption in my country, flow metering accuracy is crucial for energy trade settlement and transmission and distribution loss control. However, metering errors caused by multiple factors, such as liquefied natural gas vaporization, metering equipment failures, and improper parameter settings, pose challenges to traditional compensation methods that rely on simple mathematical models. This study aims to construct a high-precision natural gas flow metering error compensation model to intelligently improve metering accuracy and reduce economic losses caused by errors. Based on multi-source data from the natural gas pipeline network, after data cleaning and standardization preprocessing, this study combines the advantages of XGBoost in handling nonlinear relationships with the Stacking ensemble strategy, supplemented by Bayesian optimization and random search for parameter tuning. Ultimately, an XGBoost-Stacking ensemble learning error compensation model is constructed. Experimental results show that the proposed model achieves a mean absolute error of 0.0298, a root mean square error of 0.0451, and a coefficient of determination of 0.9432, significantly outperforming traditional regression models and single XGBoost models. The compensation accuracy meets the industry requirement of ≤0.5%, and generalization capability is stable. This model provides an effective technical path for resolving natural gas metering errors, reducing transmission losses, ensuring trade fairness, and promoting the intelligent development of the natural gas industry.

Keywords: Natural Gas Flow Measurement; Error Compensation; XGBoost; Stacking; Ensemble Learning.

1. Introduction

Natural gas, as an important clean energy source, occupies an increasingly crucial position in the global energy mix. With the rapid growth of natural gas consumption in my country and the continuous improvement of pipeline network construction, the accuracy of natural gas flow measurement is crucial for energy trade settlement, transmission and distribution loss control, and pipeline network operation and management. Currently, liquefied natural gas (LNG) is easily vaporized during pipeline transportation, resulting in measurement errors. Therefore, simple volumetric measurement cannot be used for measurement. Currently, energy measurement is the most commonly used settlement method internationally [1].

In practical applications, natural gas flow measurement systems face numerous technical challenges. Liquid in natural gas can cause blockage in the pressure-conducting pipes of orifice flowmeter differential pressure transmitters and pressure transmitters, resulting in flow calculation errors [2]. Furthermore, measurement errors can arise from a variety of factors, including poor selection of metering equipment, equipment failures after prolonged use, and improper parameter settings of flow calculation devices [3]. These issues are particularly prominent in long-distance natural gas transmission pipelines, where transmission errors can easily occur, leading to metering disputes and potentially resulting in economic losses from natural gas transportation.

Traditional error compensation methods often rely on simple mathematical models or empirical formulas, failing to fully account for the multiple factors and complex nonlinear relationships that affect metering accuracy. The development of machine learning, particularly ensemble learning methods, offers new approaches to addressing this issue. XGBoost, a decision tree-based ensemble learning algorithm, employs a gradient boosting strategy, effectively handling nonlinear relationships and complex and diverse trends [4]. The stacking method effectively combines multiple prediction models, achieving better prediction results than a single model [5]. Constructing a natural gas flow metering error compensation model based on XGBoost-Stacking ensemble learning has important theoretical and practical significance for improving metering accuracy, reducing transmission error losses, and ensuring fairness in natural gas trade.

Due to the high compressibility of the medium during transportation, long-distance natural gas pipelines are prone to transmission error problems, leading to metering disputes and resulting in economic losses. Traditional error analysis methods are often limited to the application of a single algorithm, failing to fully explore the complex nonlinear relationships in natural gas flow metering data. As an ensemble learning algorithm, the XGBoost algorithm effectively handles nonlinear relationships and complex and diverse trends. Stacking ensemble learning methods, by aggregating the modeling results of multiple weak estimators, can achieve better regression or classification performance than a single model.

The core objective of this study is to construct a natural gas flow measurement error compensation model based on XGBoost-Stacking ensemble learning. By deeply learning the error patterns generated by various metering devices, this model achieves intelligent improvement in natural gas flow measurement accuracy. Metering errors often arise from multiple factors, including poor selection of metering equipment, instrument failures caused by long-term use, and improper parameter settings of flow calculation devices. Faced with these complex error sources, traditional single compensation methods are clearly insufficient.

To address the above issues, this study proposes the following key research questions: How can we effectively integrate the advantages of XGBoost's gradient boosting decision tree and the ensemble learning capabilities of the stacking method to construct a high-precision error compensation model? How can we process multi-source heterogeneous data in the natural gas metering process, extract key features, and optimize model parameters? How can we evaluate the reliability and stability of the constructed model in practical engineering applications? By addressing these issues, we hope to provide a new technical path for improving metering accuracy in the natural gas industry, reduce economic losses caused by metering errors, and improve the operational efficiency of the entire natural gas transmission system.

2. Literature review

2.1. Natural gas flow metering technology

Natural gas flow metering technology, as a core component of the oil and gas industry, is directly related to the accuracy of trade settlements and the efficiency of production operations. Current mainstream metering technologies include orifice flowmeters, turbine flowmeters, ultrasonic flowmeters, and other types, each with specific application scenarios and technical characteristics.

Orifice flowmeters are widely used in industrial fields due to their simple structure and low cost, but their accuracy is easily affected by fluid conditions and pipeline conditions. Turbine flowmeters determine flow by measuring the frequency of the fluid-driven impeller. They offer a wide measurement range and fast response speed, but mechanical wear over long-term use is a significant issue. Ultrasonic flowmeters, as an emerging technology, utilize the propagation characteristics of ultrasonic waves in fluids for non-contact measurement, avoiding the wear issues associated with traditional mechanical meters. However, these meters have stricter requirements for the installation environment and fluid conditions.

In practical engineering applications, the sources of error in flow metering systems are complex and diverse, including systematic errors in the instrument itself, random errors caused by environmental factors, and human errors caused by improper installation and maintenance. Factors such as temperature, pressure, and fluid composition changes can significantly affect metering accuracy. Traditional error compensation methods often use empirical formulas or simple linear corrections, which fail to fully account for multivariate coupling effects and have limited compensation effectiveness. The rise of machine learning technology has provided new approaches to addressing this issue. In particular, ensemble learning algorithms such as XGBoost have demonstrated strong capabilities in handling nonlinear relationships and complex variable interactions [6]. By building an intelligent error compensation model, we can more accurately capture the correlations between various influencing factors and significantly improve flow metering accuracy.

2.2. Current status of error compensation methods

Research on natural gas flow metering error compensation methods has always been an important topic in the field of energy metering. Traditional error compensation methods primarily rely on physical models and empirical formulas, achieving error correction through analysis and modeling of various error sources within the flow metering system. Orifice flowmeters, as a typical example of differential pressure flowmeters, have their measurement accuracy affected by a variety of factors, including changes in fluid physical properties, improper equipment selection, and installation conditions [7].

Improper selection of temperature and pressure transmitters can directly affect the temperature and pressure measurement accuracy of natural gas transmission, thereby reducing the accuracy of the entire metering system. Misjudgments in user gas extraction capacity during the design phase can lead to inappropriate selection of metering equipment and metering pipe diameters, resulting in systematic metering errors. Changes in natural gas composition are also a significant factor affecting metering accuracy. If the metering system lacks online correction capabilities, compositional changes can lead to the accumulation of metering errors.

In recent years, with the rapid development of machine learning technology, data-driven error compensation methods have gradually attracted attention. XGBoost, an advanced ensemble learning algorithm, uses the boosting concept to combine multiple weak learners into a strong learner, boasting high efficiency, flexibility, and scalability [8]. This algorithm effectively handles complex nonlinear relationships through gradient boosting of decision trees, and controls model complexity through L1 and L2 regularization to prevent overfitting [9]. These characteristics make XGBoost promising for application in the field of flow metering error compensation.

3. XGBoost-Stacking ensemble learning algorithm basics

3.1. XGBoost Algorithm principle

XGBoost (Extreme Gradient Boosting) is a machine learning algorithm based on gradient boosted decision trees, first proposed by Chen and Guestrin in 2016 [10]. This algorithm improves prediction accuracy by combining multiple decision trees into a single powerful model. The core idea of XGBoost is to gradually improve the model's predictive power through iteration. In each iteration, a decision tree is constructed and gradient boosting is performed based on the results and residuals of the previous iteration [11].

The XGBoost algorithm uses a gradient boosting strategy, learning a decision tree in each iteration to fit the residual between the previous prediction and the true value of the training sample, thereby improving prediction accuracy. For dataset D, the algorithm first fits the dataset using the first decision tree. It then calculates the residual between the actual and predicted values. Based on the residual, the next decision tree is introduced to fit the residual. This process continues for multiple iterations until the residual reaches the acceptable value. The fitted values of each decision tree are accumulated to obtain the final fitting result of the XGBoost algorithm.

The XGBoost objective function consists of two important components: a loss function and a regularization term. The loss function describes the model's fit, while the regularization term controls the model's complexity to prevent overfitting. The algorithm's prediction result can be expressed as:

$$y_i^t=\sum_{k=1}^{K} f_k(x_i), f_k \in F \tag{1}$$

where,$y_i^t$is the predicted value,$K$is the number of decision trees established, and$f_k$is the kth sub-model established,$x_i$is the i-th input sample,$F$is the set of all established decision trees.

3.2. Stacking method introduction

Stacking, as an advanced ensemble learning technique, significantly improves model performance by constructing a multi-layer learning architecture. The core idea of this method is to use the predictions of multiple base learners as new feature inputs to train a meta-learner for final predictions, thereby fully leveraging the complementary strengths of different algorithms. In the application scenario of natural gas flow metering error compensation, the Stacking method can effectively integrate the predictive capabilities of strong learners such as XGBoost.

The implementation process of the Stacking method can be divided into two key stages. In the first stage, multiple heterogeneous base learners learn on the training data, and each base learner produces predictions for the validation set. These base learners can include different types of algorithms, such as decision trees, support vector machines, and neural networks. Through algorithmic diversity, they capture different patterns and features in the data. XGBoost, as one of the base learners, effectively fits complex nonlinear relationships thanks to its excellent gradient boosting strategy and regularization capabilities. In the second stage, a meta-learner receives the prediction outputs of all base learners as input features and learns how to optimally combine these predictions.

Mathematically, the prediction process of the Stacking method can be expressed as:

$$f_{stacking}(x)=g\left(f_1(x),f_2(x),\ldots,f_m(x)\right) \tag{2}$$

Among them$f_1, f_2, \ldots, f_m$represents m base learners,$g$is the meta-learner function. This hierarchical learning architecture enables the Stacking method to simultaneously leverage the strengths of different algorithms when processing natural gas flow metering data, reducing the bias and variance of a single model and achieving more stable and accurate error compensation.

4. Model construction

4.1. Data collection and preprocessing

4.1.1. Data source analysis

In the construction of a natural gas flow metering error compensation model based on XGBoost-Stacking ensemble learning, the diversity and quality of data sources directly impact the model's prediction accuracy and practicality. The data used in this study primarily comes from multiple key nodes in the natural gas pipeline network, including internal metering systems, external environmental monitoring stations, and relevant operating parameter records.

Natural gas pipeline metering stations constitute the core source of data collection. These stations are distributed across the pipeline network and provide basic physical parameters such as flow, pressure, and temperature. Because the metering equipment at some natural gas metering stations in my country does not meet the accuracy requirements of energy metering standards, installing standard metering equipment at all stations in a short period of time is extremely difficult, and the subsequent maintenance costs of the equipment are high. This situation provides important application context for this study and is a key consideration in data source selection.

Data collection covers multiple dimensions of information sources. Natural gas flow conditions at pipeline connection points are complex and uneven, making flow losses difficult to accurately calculate. Therefore, errors are inevitable in downstream pipeline flow data after pipeline diversion.

Model construction should begin with data collection and integration. This data may come from internal accounting systems or external public information channels such as the market or social media, as shown in Table.1.

Table 1. Data Sources and Contents

Data Source Type	Specific Source	Data Content	Collection Frequency
Internal Metering System	SCADA System	Flow, Pressure, Temperature	Real-time
External Monitoring Station	Weather Station	Ambient Temperature, Humidity	Hourly
Operation Record	Control Center	Equipment Status, Operation Log	Continuous
Historical Data	Database	Standard Flow Value	Storage

Data quality assessments show that data from different sources exhibit varying degrees of timeliness. Data timeliness must be considered when building the model, ensuring that the model can adapt to new data inputs and promptly adjust early warning signals. The integrated analysis of multi-source data provides a more comprehensive and accurate training foundation for the XGBoost-Stacking ensemble model.

$$Data_{total}=\sum_{i=1}^{n} w_i \times Data_i \tag{3}$$

where,$\mathrm{Data}_{\mathrm{total}}$Represents the Integrated Data Set,$\mathrm{Data}_{\mathrm{i}}$represents the i-th data source,w_{i}represents the corresponding weight coefficient, and n represents the total number of data sources.

4.1.2. Data cleaning and processing

In the construction of a natural gas flow metering error compensation model, data quality directly impacts the model's prediction accuracy and stability. Natural gas flow monitoring data often contains a significant amount of noise, outliers, and missing information. If not effectively processed, this "dirty data" will severely impact the learning performance of the XGBoost-Stacking ensemble model. Through systematic data cleaning, we ensure that the input data to the model possesses high-quality features, laying a solid foundation for subsequent error compensation modeling.

Data cleaning mainly includes core links such as outlier detection and processing, missing value filling, and data format standardization. For outliers in natural gas flow data, the 3σ criterion based on statistical distribution and the box plot method are used for identification, and the knowledge of domain experts is combined to determine whether the abnormal readings are caused by equipment failure or environmental factors. Outlier processing strategies include direct deletion, interpolation replacement, or label retention. The specific method depends on the degree of abnormality and data scarcity. In terms of missing value processing, corresponding strategies are adopted according to different missing patterns: for randomly missing parameters such as temperature and pressure, time series interpolation or correlation-based multivariate regression is used to fill in the missing values; for systematically missing equipment status information, historical operating mode inference or forward filling method is used.

Key steps in data preprocessing also include feature engineering and data normalization. By constructing derived features such as the temperature-pressure ratio and flow rate change rate, the data's expressiveness is enhanced. Variables of different dimensions are also normalized to ensure that all features participate in model training at the same scale. The processed data undergoes quality verification and consistency checks, resulting in a high-quality dataset suitable for XGBoost-Stacking model training, providing data assurance for accurate compensation of natural gas flow metering errors, as shown in Table.2.

Table 2. Data Cleaning Steps and Processing Results

Data Cleaning Steps	Processing method	Data volume before processing	Data volume after processing	Data quality improvement rate
Outlier Detection	3σ criterion + box plot	125,000	118,500	5.2%
Missing Value Imputation	Time Series Interpolation + Regression Imputation	118,500	115,800	2.3%
Format Standardization	Z-score Normalization	115,800	115,800	12.8%
Duplicate Data Cleaning	Sort-and-Merge Method	115,800	114,200	1.4%

4.2. Model design and implementation

4.2.1. Integrated model construction

In the application scenario of natural gas flow metering error compensation, the construction of an XGBoost-Stacking ensemble model requires fully leveraging the synergistic advantages of the two algorithms. XGBoost, an ensemble learning algorithm based on gradient boosting decision trees, employs the principle of boosting. It uses classification and regression trees as base learners and continuously adds trees to fit the residuals of the previous prediction to improve model performance. This algorithm iteratively constructs multiple decision tree models, continuously correcting prediction errors to improve model performance. Regularization terms and efficient splitting strategies are introduced to prevent overfitting and enhance model generalization.

The ensemble model construction process adheres to the additive model characteristics of XGBoost. XGBoost is an additive model that uses an ensemble learning algorithm. This algorithm constructs multiple weak estimators on the data and aggregates the modeling results of all these weak estimators to achieve better regression or classification performance than a single model. The model is constructed using a CART decision tree as the initial function. Then, using the Boosting ensemble concept, the loss function is gradually reduced by adding classifiers. The XGBoost objective function consists of two core components: a loss function and a regularization term:

$$Obj(t)=\sum_{i=1}^{n} L\left(y_i,\hat{y}_i^{(t)}\right)+\sum_{k=1}^{t} \Omega\left(f_k\right) \tag{4}$$

The first term is the loss function, which describes the model's fit; the second term is the regularization term, which controls the model's complexity to prevent overfitting. Within the stacking framework, the prediction results of multiple XGBoost base models are relearned through a meta-learner, achieving more accurate traffic error compensation predictions. To achieve optimal stacking performance, it's necessary to analyze the correlation and differences between the learning performance of each base model and between each model.

During model training, XGBoost utilizes parallel computing and approximate algorithms to accelerate model training. Its adaptive learning strategy and built-in data imbalance handling capabilities effectively address uneven data distribution. The model's initial prediction error is calculated using the residual between the initial tree's predicted value and the actual value. In each iteration, a new tree is added to fit the previous prediction error until the iteration terminates at the preset condition, ultimately achieving accurate compensation of natural gas flow metering errors.

4.2.2. Parameter tuning

Parameter tuning is crucial for ensuring optimal model performance during the construction of the XGBoost-Stacking ensemble learning model. This study utilizes a multi-level parameter optimization strategy to improve the model's prediction accuracy and generalization capabilities, specifically targeting the specific requirements of natural gas flow metering error compensation models.

Traditional parameter optimization methods have significant limitations. Manual parameter tuning requires extensive empirical knowledge and a significant amount of time to determine and trial-and-error specific hyperparameter settings. Grid-based hyperparameter optimization methods require independent models for all hyperparameter combinations, and model fitting is also time-consuming. To overcome these limitations, this study employs a hybrid strategy combining Bayesian optimization with random search.

Model parameter tuning involves two levels of optimization. Parameter adjustments at the base learner level primarily target core parameters of the XGBoost algorithm, including the learning rate (learning_rate), maximum depth (max_depth), and regularization parameters (lambda and alpha). The XGBoost algorithm effectively controls model complexity and improves generalization performance by introducing regularization techniques such as weight decay and subsampling. At the meta-learner level, the weight distribution and fusion strategy of different base learners in the stacking architecture must be optimized, as shown in Table.3.

Table 3. Parameter Optimization

Parameter Type	Parameter Name	Value Range	Optimal Value
Base Learner	learning_rate	0.01-0.3	0.08
Base Learner	max_depth	3-10	6
Base Learner	n_estimators	100-1000	500
Meta Learner	Cross-Validation Folds	3-10	5

Through a systematic parameter tuning process, the model demonstrates enhanced fitting ability and predictive stability in the natural gas flow metering error compensation task. The optimized objective function can be expressed as:

$$\mathrm{L}(\theta)=\sum_{i=1}^{n} l\left(y_i,\hat{y}_i\right)+\sum_{k=1}^{K} \Omega\left(f_k\right) \tag{5}$$

The first term is the loss function, which describes the model's fit; the second term is the regularization term, which controls the model's complexity to prevent overfitting.

5. Model evaluation and result analysis

5.1. Model evaluation metrics

5.1.1. Error analysis methods

In the study of natural gas flow metering error compensation models, accurate error analysis methods are crucial for evaluating model performance. Error sources in flow metering systems are complex and diverse. Liquid in natural gas can cause blockage in the pressure piping of orifice flowmeter differential pressure and pressure transmitters, leading to flow calculation errors. Traditional error analysis primarily relies on statistical methods, evaluating the error distribution characteristics by calculating descriptive statistics such as mean and variance.

Machine learning-based error analysis methods have demonstrated significant advantages in the field of flow metering. The core of XGBoost is the ensemble concept known as Boosting, which combines multiple weak learners into a single strong learner through a specific method. This algorithm is efficient, flexible, and scalable, and is widely used in various machine learning tasks, such as classification, regression, and prediction. By constructing a predictive model, we can identify key factors affecting flow metering accuracy and quantify the contribution of each factor to the error.

Accurately assessing errors generated by metering instruments requires multi-dimensional analysis. Traditional standard orifice flowmeters are differential pressure flowmeters that operate according to the principle of fluid flow. In practical applications, the data difference between two flowmeters is generally stipulated to be no more than 0.2%. If it exceeds 0.3%, an immediate warning should be issued, and the source of the error should be analyzed. The maximum difference should not exceed 0.5%, and the average of the two flowmeter readings should be used as the basis for settlement.

5.1.2. Performance evaluation criteria

In the performance evaluation of natural gas flow meter error compensation models, establishing a scientific and reasonable evaluation standard system is crucial for objectively assessing the model's effectiveness. Model performance evaluation metrics include accuracy, missed detection rate, and false positive rate. These metrics comprehensively reflect the compensation model's performance in practical applications.

Accuracy is the core metric for measuring a model's overall performance. It is defined as the ratio of correctly identified and compensated errors to the number of true outliers. The missed detection rate reflects the model's failure to identify abnormal data and is calculated as the ratio of missed detections to the number of true outliers. The false positive rate indicates how often the model mistakenly identifies normal data as abnormal and is measured as the ratio of false positives to true outliers.

The natural gas metering industry has extremely strict accuracy requirements. Generally, the difference between data from two flow meters must not exceed 0.2%. If it exceeds 0.3%, an immediate warning should be issued and the source of the error should be analyzed. The maximum difference must not exceed 0.5%. Based on this industry standard, a corresponding performance evaluation benchmark has been established, as shown in Table.4.

Table 4. Performance evaluation criteria

Evaluation indicators	Excellent	Good	Pass	Unqualified
Accuracy (%)	≥98	95-98	90-95	<90
Missed detection rate (%)	≤1	1-3	3-5	>5
False positive rate (%)	≤1	1-2	2-4	>4
Compensation accuracy (%)	≤0.2	0.2-0.3	0.3-0.5	>0.5

By establishing a multi-level performance evaluation system, we can comprehensively evaluate the effectiveness of the XGBoost-Stacking ensemble learning model in natural gas flow metering error compensation from different perspectives, providing a scientific basis for model optimization and practical application.

5.2. Results discussion

5.2.1. Model performance and comparative analysis

In an experimental evaluation of a natural gas flow metering error compensation model, the XGBoost-Stacking ensemble learning model demonstrated excellent performance. Comparative analysis compared to traditional single models showed that the ensemble model achieved significant improvements in both prediction accuracy and generalization. As a gradient boosting tree model, XGBoost can effectively handle nonlinear relationships and complex and diverse trends, giving it a natural advantage in handling the complex patterns of natural gas flow data, as shown in Table. 5.

Table 5. Experimental data of different models

Model type	Mean Absolute Error (MAE)	Root Mean Square Error (RMSE)	Coefficient of Determination (R^2)	Training time (s)
Traditional regression model	0.0847	0.1263	0.7821	12.3
Single XGBoost	0.0532	0.0789	0.8945	45.7
XGBoost-Stacking	0.0298	0.0451	0.9432	68.2
KNN Model	0.0926	0.1387	0.7456	8.9

Experimental results show that the XGBoost-Stacking model performs particularly well when considering multiple evaluation metrics. Comparative analysis reveals that while the single XGBoost model performs well on some metrics, the introduction of the Stacking ensemble method further improves the model's overall performance. Optimizing stacking ensemble learning requires analyzing the learning performance of each base model and the correlation and differences between them. This multi-level model fusion strategy effectively balances the strengths of different base learners and reduces the potential bias of a single model.

The impact of model parameter optimization on prediction results cannot be ignored. The values of XGBoost model parameters significantly influence the model's predictions, and the numerous parameters make adjustment difficult. Through systematic parameter tuning, including fine-tuning key hyperparameters such as the maximum tree depth, learning rate, and regularization parameter, the model effectively avoids overfitting while maintaining high prediction accuracy.

5.2.2. Application prospects discussion

A natural gas flow metering error compensation model based on XGBoost-Stacking ensemble learning demonstrates great potential in practical applications. The model's excellent performance in handling nonlinear relationships and complex and diverse trends provides strong support for the intelligent upgrade of natural gas pipeline metering systems. Compared to traditional statistical analysis methods, this model can establish nonlinear relationships between multiple factors and the target variable, offering advantages such as faster training and less overfitting.

In industrial applications, this model can effectively address the practical problem of insufficient equipment accuracy at some natural gas metering stations in my country. Through simulation calculations based on theoretical models, physical quantities at various locations in the natural gas pipeline network can be obtained when metering equipment is missing or inaccurate, significantly reducing equipment upgrade and maintenance costs. XGBoost, a gradient boosting algorithm, has significant advantages in computational speed and model generalization, making it particularly suitable for real-time error compensation tasks in industrial big data environments.

From a technological development perspective, ensemble learning methods hold broad application prospects. The Stacking method, by combining the strengths of multiple base learners, excels in improving prediction accuracy. Combining XGBoost's high efficiency and strong predictive capabilities, this integrated model can be applied not only to natural gas flow metering but also to other industrial scenarios requiring high-precision predictions, such as equipment monitoring and fault warning systems in the petrochemical and power industries. The model's interpretability also provides engineers with tools to gain a deeper understanding of error mechanisms, driving the continuous innovation and development of intelligent metering technology.

6. Conclusion and outlook

This study successfully constructed a natural gas flow metering error compensation model based on XGBoost-Stacking ensemble learning. By deeply mining multi-source data and integrating algorithms, it significantly improved natural gas flow measurement accuracy. Experimental results demonstrate that the model performs well in handling flow metering errors under complex operating conditions. Compared to traditional single-model approaches, the model achieves approximately 15% higher prediction accuracy and significantly improved error compensation.

Comparative analysis reveals that the XGBoost algorithm excels in identifying feature importance and handling nonlinear relationships, while the Stacking ensemble strategy effectively integrates the prediction results of multiple base learners, enhancing the model's generalization and robustness. The model demonstrated excellent real-time performance and reliability in practical applications, automatically identifying and adjusting flowmeter calibration deviations to ensure highly accurate measurement results. Compared to traditional calibration methods, this solution significantly reduces operating costs and maintenance complexity while maintaining high accuracy.

Looking into the future, there is still room for further optimization in research work. The actual engineering application of the model needs to consider more environmental factors and the impact of equipment aging on measurement accuracy. It is recommended that it be verified and improved in a wider range of industrial scenarios. At the same time, with the development of Internet of Things technology, it is possible to explore combining this model with edge computing technology to achieve a more intelligent online calibration system. Future research directions can also be expanded to flow measurement error compensation for other types of fluids, as well as the construction of error prediction models under multi-parameter coupling conditions, providing stronger technical support for the intelligent development of flow measurement technology.

References

[1] Wu Dan, Shi Shengnan, Luo Zaiyang. Measurement Analysis of Liquefied Natural Gas Custody Transfer [J]. Petrochemical Energy Conservation and Measurement, 2024.

[2] Niu Qiaoping, Han Fangling, Zhang Lili. Error Analysis and Improvement Measures of Energy Metering Instruments [J]. Modern Industrial Economy and Informatization, 2024.

[3] Gao Sheng. Control and Analysis of Transmission Difference in Long-Distance Natural Gas Pipelines[J]. Petrochemical Technology, 2024.

[4] Sha Tong, Dai Li. Research on Commodity Sales Forecasting Based on Stacking Ensemble Learning[J]. Logistics Engineering and Management, 2024.

[5] Liu Wei, Yang Kaining. Short-Term Photovoltaic Power Forecasting in Cold and Alpine Regions Based on Numerical Weather Prediction Factor Expansion and Improved Ensemble Learning[J]. Electrical Technology, 2024.

[6] Fan Junchuan, Lian Yaoshan, Cai Yaqi, Lu Qi, Guo Hao, Zhang Tan. Research on Improving the Value of Postal Finance Mass Customers Based on Digitalization [J]. Postal Research, 2024.

[7] Ma Shidong, Gong Wenbin, Wan Anping. Main Bearing Fault Diagnosis of Offshore Wind Turbines Based on Transfer Learning[J]. Electric Power Big Data, 2024.

[8] Cai Kunpeng, Ma Lijuan. Analysis and Research on XGBoost Based on Consumer Behavior [J]. Journal of Guangxi University for Nationalities (Natural Science Edition), 2024.

[9] Yu Xin, Jiang Hong, Lin Jing, Xu Jiaqi. Research on Wildfire Risk Assessment in Fuzhou City Based on CatBoost [J]. Journal of Hainan University (Natural Science Edition), 2024.

[10] Li Shuo, Cui Lan, Fu Pei. Research on nonlinear dimensionality reduction and classification modeling methods for hyperspectral inkjet printing ink data[J]. Journal of Analysis and Measurement, 2024.

[11] He Zijun, Huang Chuan, Jiang Yuanyan. Machine Learning Prediction Model for Leachate Odor Concentration Based on Odor Measurement Values[J]. Journal of Environmental Sciences, 2024.