Academic Education Publishing House

Industry Science and Engineering Vol. 3 No. 6, 2026

Zhimu Huanxi: A Custom Big Data & Machine Learning Platform to Quantify Air Pollution Impacts on Regional Livestock Economic Output

Lu Li, Jingyang Cui, Jiaqi Zhang*

School of Information Science and Technology, Baotou Teachers' College, Baotou, Inner Mongolia, China

**Corresponding Author*

Abstract: Farmers and local environmental regulators lack practical tools to quantify how ambient air pollutants drag down livestock industry revenue. To fill this practical gap, our team built a customized analysis platform named Zhimu Huanxi, using multi-source field records collected across Inner Mongolia between 2020 and 2024. The raw datasets cover local meteorological observations, routine air quality monitoring records and annual livestock production statistics. After clearing invalid records, aligning time labels and normalizing all numerical indicators, we adopted Lasso regression to screen out six dominant impact factors: heavy pollution coverage intensity, atmospheric SO_2 concentration, inhalable PM10 particle levels, regional annual average temperature, livestock breeding scale and sustained mild pollution periods. We then constructed a Linear Support Vector Regression (LSVR) model to predict yearly livestock output value losses triggered by air pollution. We reserved all 2024 records as independent test samples to test generalization capacity; the model reached an R^2 of 0.78, with an average absolute forecasting error of 152 million RMB and root mean square error equal to 198 million RMB. Residual values scattered randomly around zero without systematic deviation. The platform backend runs on lightweight Flask, while the front-end embeds ECharts interactive charts. Operators can view pollutant weight rankings, real-time output loss forecasts and multi-year historical data with one-click query. This platform delivers quantifiable evidence for breeders to design emission reduction plans and helps environmental authorities lock high-priority pollution control zones in pastoral areas.

Keywords: Machine Learning; Multi-Source Big Data; Air Pollution Economic Assessment; Livestock Output; Lasso Feature Screening; Linear Support Vector Regression

1. Introduction

Livestock breeding sectors worldwide are suffering from growing air pollution risks. The intertwined link between airborne contaminants and breeding productivity has formed a major barrier to environmentally sustainable agriculture. Large-scale livestock farms release massive volumes of harmful airborne substances including ammonia, hydrogen sulfide and suspended particles, which in turn worsen local ambient air quality. These pollutants adversely affect animal health and production performance (e.g., causing respiratory diseases, reducing meat yield and milk production) while posing potential threats to surrounding ecosystems and human health. Traditional environmental monitoring methods relying on expensive equipment and manual sampling struggle to achieve rapid, large-scale, and precise supervision. As a major source of greenhouse gas emissions, livestock farming faces challenges in quantifying its emission volumes and impact on economic output due to fragmented data, strong temporal lag, and insufficient accuracy, severely impeding informed decision-making and sustainable development strategies. Therefore, developing an intelligent system capable of integrating multi-source data, accurately assessing pollution impacts, and providing decision support holds significant theoretical importance and urgent practical value—particularly in advancing green transformation and achieving carbon peak and neutrality goals.

In response to these challenges, scholars worldwide have conducted extensive research. Machine learning methods have demonstrated

significant potential in environmental monitoring and prediction for livestock farming. For instance, Xie Qiuju et al. used a deep learning model to predict the temperature and humidity parameters in a closed pig house [1], and Liu Chunhong et al. employed the ARIMA-BP neural network to predict ammonia concentration in pig houses [2].Yang Liang et al. used the EMD-LSTM model to improve the prediction accuracy of ammonia concentration in pig houses [3].In the broader context of assessing environmental pollution's impact on agricultural economics, Park et al. conducted a comparative analysis of ammonia concentration in mechanically ventilated gestation sow houses in South Korea using a neural network model [4]; Genedy et al. proposed a physics-informed LSTM model [5]; and He et al. estimated manure emissions from different types of livestock and poultry in China [6]. However, most current studies focus solely on predicting individual pollutants or specific farming parameters, with research emphasis primarily on crops rather than livestock production. More critically, there is a widespread lack of end-to-end systems that comprehensively integrate heterogeneous data from meteorology, environmental monitoring, and livestock operations to quantify air pollution-induced economic losses. Major technical obstacles involve inconsistent data normalization rules, low accuracy when screening pollution driving factors, and inadequate visual modules for calculating economic losses. To tackle these defects, this work builds an intelligent assessment platform titled "Zhimu Huanxi". This platform combines multi-source inconsistent datasets and two machine learning tools, Lasso regression and Linear Support Vector Regression (LinearSVR). It can quantitatively evaluate and predict how air pollutants restrain livestock economic output, linking raw monitoring data to measurable calculations of industrial revenue losses.

2. Overall System Architecture and Data Preprocessing

2.1 Three-Level Architecture Design of the System

We adopted Browser/Server deployment mode for the whole platform, splitting the whole workflow into three independent modules: data collection layer, algorithm modeling layer and visual application layer. Each module bears separate responsibilities and can be upgraded or adjusted separately without interfering with other parts.

(1) Data Collection Layer: This layer acts as the system's raw data entry, connecting three separate data sources: national and local environmental monitoring stations, meteorological bureau open API interfaces, and livestock statistical reports issued by local agriculture and animal husbandry departments. We wrote Python crawler scripts and set up regular API polling tasks to fetch daily and monthly records automatically, then stored all structured data in MySQL and PostgreSQL relational databases for subsequent modeling. Cross-domain data integration follows unified processing standards to guarantee data completeness and real-time update capacity.

(2) Analysis and Modeling Layer: Developed using Python Flask framework for the backend [7], using simple route mapping to call different algorithm modules, which serves as the core computing unit of the whole platform. This layer completes full data processing pipelines, including raw data cleaning, feature extraction, machine learning training and output value prediction, covering two core algorithm units: Lasso feature filtering and Linear SVR prediction. When front-end users send query requests, the backend distributes computing tasks to corresponding model modules, calculates target results and returns feedback to the webpage in JSON format. We applied Ajax asynchronous communication to realize page-free data refresh, reducing response delay caused by page switching, and added multiple optimization steps to adapt machine learning models to web engineering deployment.

(3) Visual Application Layer: The front-end interface is coded with HTML, CSS and JavaScript, integrating open-source ECharts chart tools to display model outputs in intuitive interactive graphs. Ajax technology maintains real-time data transmission between front and backend. We customized query and chart functions for three types of users: livestock farmers, administrative managers and agricultural researchers, supporting historical record retrieval, output trend forecasting and pollutant impact weight comparison. Flask is selected as the backend framework for its low resource consumption and fast deployment speed, suitable for medium-small regional data

computing tasks. ECharts provides high customization flexibility to match multi-pollutant and multi-time dimension visualization demands. The overall architecture maintains concise logic and high operating efficiency, reserving sufficient space for function expansion in follow-up iterations.

2.2 Multi-Source Data Fusion and Preprocessing

The dataset employed in this study encompasses various types of data from Inner Mongolia Autonomous Region covering meteorological conditions, environmental monitoring, and livestock industry output value from 2020 to 2024. Environmental monitoring and output value data were sourced from statistical reports published on the official website of the Inner Mongolia Autonomous Region Statistics Bureau and monitoring reports from local environmental protection authorities; livestock industry output value data primarily originated from the *Inner Mongolia Statistical Yearbook 2024*[8]; meteorological data were obtained via APIs of free weather service platforms.

Raw records contain repeated samples, abnormal values and missing entries, which cannot be directly input into model training. We carried out four preprocessing steps in sequence:

(1) Duplicate record elimination: Scan all data tables and delete fully identical samples to avoid repeated data interfering model training stability.

(2) Abnormal value correction: Combine the 3σ outlier detection rule and local breeding industry business logic to identify unreasonable data points, such as negative pollutant concentration values and abnormally sharp output fluctuations. For indicators exceeding normal value ranges, we cross-check records from adjacent monitoring stations and local annual agricultural policy adjustments; invalid samples are only deleted after double verification.

(3) Time label unification: Different data sources adopt inconsistent collection cycles and timestamp formats, so we unify all records into monthly summary units to realize horizontal comparison between different indicator types.

(4) Missing value filling: The overall proportion of missing samples stays low; we abandon simple mean filling (which distorts original feature distribution) and adopt multiple imputation methods including K-nearest neighbor interpolation and regression-based filling. The filling process references regional temperature, breeding scale and other associated indicators of the same period to retain original data distribution characteristics.

Finally, all numerical features are standardized using the Z-score method: $z = (x - \mu) / \sigma$, where x represents the original data point, μ is the mean value of the feature, and σ is its standard deviation. This step eliminates the impact of varying scales across features, thereby enhancing performance in subsequent Lasso regression feature selection and predictive modeling.

3. Core Algorithms and Model Construction

3.1 Lasso Regression Feature Selection Based on L1 Regularization

The loss function for Lasso (Least Absolute Shrinkage and Selection Operator) regression builds upon the mean squared error (MSE) loss of linear regression by incorporating an L1 regularization term. Its mathematical expression is as follows:

$$J(\mathbf{w}, b) = \frac{1}{2n}\sum_{i=1}^{n}(y_i - (\mathbf{w}^T\mathbf{x}_i + b))^2 + \lambda\sum_{j=1}^{p}|w_j| \quad (1)$$

Here, n denotes the number of samples, y_i represents the true output value, x_i is the feature vector of the i-th sample, w is the weight vector, b is the bias term, λ is the regularization coefficient (which penalizes weights), and p is the number of features. The core mechanism of Lasso regression involves imposing absolute value penalties on feature coefficients w, thereby compressing coefficients of irrelevant or weakly correlated features to zero during iterative optimization, enabling automatic feature selection and generating a sparse feature subset.

This study employed 12 preprocessed initial features as inputs, covering dimensions such as duration of different pollution levels, concentrations of SO_2, NO_2, PM2.5, and PM10, average temperature, precipitation, and aquaculture scale. During the optimization of the regularization parameter λ, we utilized 5-fold cross-validation to determine the optimal penalty weight. Cross-validation involved dividing the dataset into five parts, using four parts alternately as the training set and one part as the validation set, to evaluate the model's performance metrics (e.g., MSE) across various λ values, ultimately selecting the value that maximizes the model's generalization ability. Throughout the screening process, we also referenced the correlation results from feature relevance heatmaps to assess feature importance,

ensuring that the selection process effectively removed redundant variables without compromising the interpretability of key influencing factors.

Following rigorous screening using Lasso regression, six core features were ultimately retained: severe pollution intensity, SO_2 concentration, inhalable particulate matter (PM10) concentration, average temperature, farming scale, and duration of moderate pollution. The screening results are shown in Figure 1. These features are statistically significant and exhibit a clear operational correlation with livestock industry output value. This sparse feature set effectively reduces the input dimensions of subsequent prediction models, helps mitigate overfitting risks, and enhances model generalization capabilities. Moreover, the retained core features provide a solid foundation for quantifying weights of pollution factors, enabling decision-makers to focus on the most critical environmental factors for governance. The design of related optimization strategies draws on algorithmic approaches that balance efficiency and screening accuracy under complex constraint scenario.

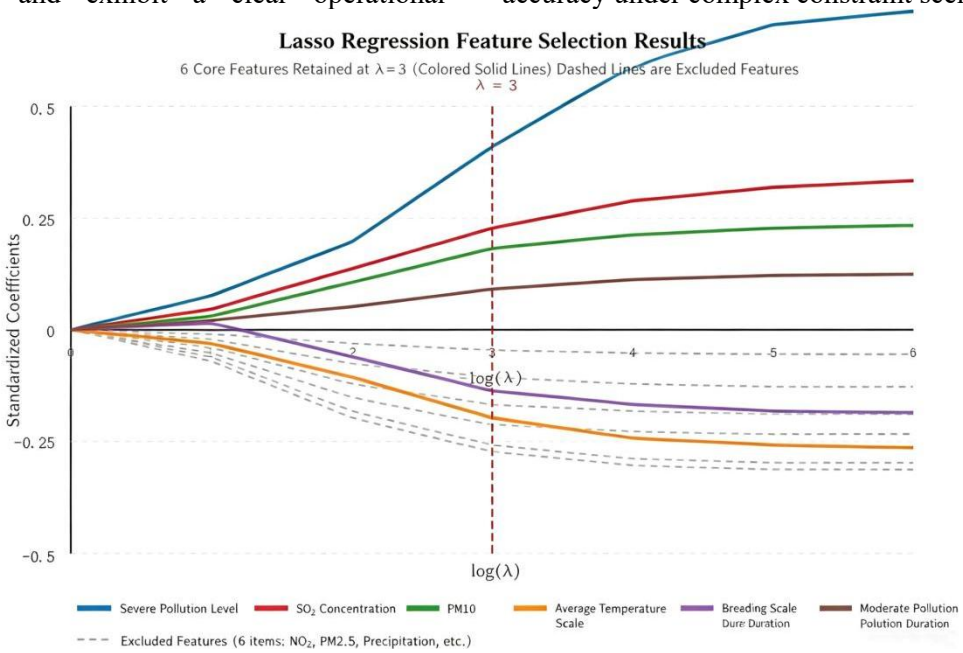

Figure 1. Results of Lasso Regression Feature Selection

3.2 Output Value Prediction Based on Support Vector Machine Linear Regression

After feature selection, this study employed a linear support vector regression (Linear SVR) model implemented using the Scikit-learn framework [9] to predict livestock industry output value. The core concept of SVR involves mapping data into a high-dimensional space and identifying an optimal hyperplane that minimizes the sum of distances from all training samples to this hyperplane, while allowing errors within a specified tolerance level ϵ.

(1) Principle: Identify the optimal hyperplane to achieve linear fitting within a tolerance ε.

The Linear SVR model aims to identify a function $f(\boldsymbol{x}) = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} + b$ such that the error between the predicted value f(xi) and the true value yi for each training sample (xi, yi) is less than or equal to a predefined tolerance level ϵ. When the error exceeds ϵ, relaxation variables θ_i and θ*_i are introduced to quantify the extent of deviation from this tolerance. The optimization objective of SVR is to minimize both model complexity and these excess errors simultaneously. The optimization problem can be formulated as:

$$\min_{\boldsymbol{w},b,\xi_i,\xi_i^*} 1/2\|\boldsymbol{w}\|_2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \quad (2)$$

$$s.t. \quad \boldsymbol{y}_i - (\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b) \le \varepsilon + \xi_i \quad (3)$$

$$(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b) - y_i \le \varepsilon + \xi_i^* \quad (4)$$

$$\xi_i, \xi_i^* \ge 0, \quad i = 1, \ldots, n \quad (5)$$

Here, $\|w\|^2$ represents the model's complexity (desirable to minimize for enhanced generalization performance), C is the regularization parameter that balances model complexity and training error (i.e., the penalty for errors exceeding the ϵ threshold), and ϵ denotes the ε-insensitive band parameter, defining the acceptable range of prediction errors the model can tolerate. The advantage of Linear SVR lies in its insensitivity to outliers, as

it only considers sample points outside the ε-insensitive band, while also demonstrating high efficiency when handling linear relationships.

(2) Parameter optimization: Adjust the regularization parameter C and tolerance ε using cross-validation.

The model's performance is highly dependent on the choice of regularization parameter C and tolerance ϵ.

C (penalty coefficient): Controls the penalty strength for errors outside the ϵ-insensitive band. A larger C value means the model will fit the training data more closely, potentially leading to overfitting; a smaller value allows greater error tolerance, which may result in underfitting.

ϵ (insensitivity bandwidth): Defines the acceptable error range for the model. Errors within this range are not included in the loss function. A smaller ϵ enables the model to fit data more precisely but makes it more susceptible to noise; a larger ϵ produces smoother predictions but may overlook valid patterns in the data.

To determine the optimal combination of C and ϵ, this study employed grid search combined with 5-fold cross-validation. Within a predefined parameter grid, the system explores all possible parameter combinations and performs cross-validation for each to evaluate the model's performance on the validation set. The parameter combination yielding the best evaluation metrics (e.g., highest R^2 or lowest MAE/RMSE) was ultimately selected.

(3) Model evaluation: Assess metrics such as R^2 score and MAE to verify the model's feasibility.

Decision Coefficient (R^2 Score): Measures the proportion of the dependent variable's variance explained by the model, ranging from 0 to 1. A higher R^2 value indicates better model fit. The calculation formula is as follows:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (6)$$

Where y_i is the true value, $\hat{y}_i$ is the predicted value, and $\bar{y}$ is the average of the true values.

Mean Absolute Error (MAE): The average absolute error between predicted and actual values. A smaller MAE indicates higher model prediction accuracy. The calculation formula is:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \quad (7)$$

Root Mean Squared Error (RMSE): It is the square root of the sum of squared errors between predicted and actual values, imposing a heavier penalty on large errors. A smaller RMSE indicates higher model prediction accuracy. The calculation formula is:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (8)$$

Through comprehensive evaluation of these indicators, the feasibility and effectiveness of the SVR model in livestock production value forecasting can be fully assessed.

3.3 Experiments and Result Analysis

This experiment utilized data from 2020 to 2023 for model training and parameter tuning, with data from 2024 serving as an independent test set to evaluate the model's generalization performance. The datasets were segmented chronologically to ensure accurate representation of the model's performance on unknown temporal data.

Visualization of feature correlation matrices: Interpret the heatmaps to illustrate the positive or negative correlations between each pollution indicator and output value.

To provide a more intuitive understanding of the relationships among the features and between these features and the target variable (livestock output value), we generated a feature correlation heatmap, as shown in Figure 2.

By examining the heat map (Figure 2), we observed a significant negative correlation between severe pollution levels, SO_2 concentrations, and inhalable particulate matter concentrations (PM10) with livestock industry output value, indicating that higher concentrations of these pollutants correlate with lower output values. For instance, the correlation coefficient between SO_2 concentration and output value is-0.65, demonstrating its strong negative impact. Conversely, breeding scale and average temperature show positive correlations with output value, suggesting that optimal conditions enhance productivity. Moderate pollution duration also exhibits a negative correlation, albeit to a lesser extent than severe pollution. These findings align with our operational intuition and existing research, further validating the effectiveness of Lasso regression feature selection.

Prediction performance evaluation: Compare the actual output value with the predicted output value curve and analyze the residual behavior.

On the test set, the Linear SVR model achieved satisfactory prediction results, and Figure 3 illustrates a comparison of its predictive performance.

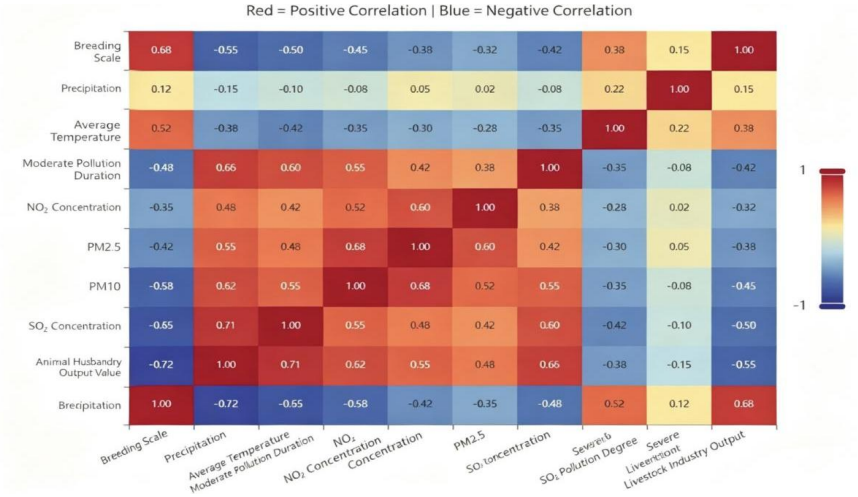

Figure 2. Heatmap of Feature Correlation

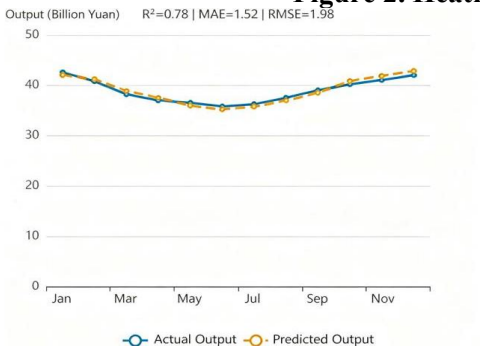

Figure 3. Comparison of Model Prediction Performance

As shown in the Figure 3, the model's predicted output value curve closely aligns with the actual output value curve in terms of trend, particularly effectively capturing directional changes during periods of significant output fluctuations. Quantitatively, the model achieves a coefficient of determination (R^2) of 0.78 on the test set, indicating it explains approximately 78% of the variability in livestock industry output values and demonstrating strong fitting capability. The ean absolute error (MAE) is 1.52 (in billion yuan), and the root mean square error (RMSE) is 1.98 (in billion yuan). These low MAE and RMSE values suggest minimal average deviation between predictions and actual values, reflecting high forecasting accuracy; Figure 4 illustrates the residual distribution plot.

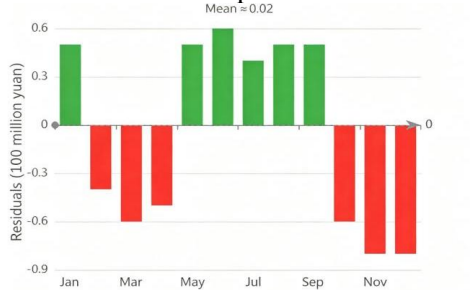

Figure 4. Residual Distribution Plot

Residual distribution is plotted in Figure 4. All residual points cluster evenly around zero without obvious upward or downward trends alongside predicted values. The average residual value is merely 0.02, nearly equal to zero. This phenomenon proves the model has no systematic overestimation or underestimation bias and delivers stable forecasting results under different output levels. All experimental evidence validates that the Zhimu Huanxi platform combining Lasso feature screening and Linear SVR can reliably quantify air pollution's economic impact on local livestock industry.

4. System Implementation and Visualization

4.1 Backend API Design

All original monitoring data are converted into standardized JSON formats and stored in the server database after preprocessing. The Flask backend is configured with dedicated routing channels to receive JSON data requests. After the backend service is launched, the front-end page sends asynchronous Ajax requests to pull processed JSON data and render visual charts on web pages. Figure 5 demonstrates the complete request-response transmission logic between front and backend.

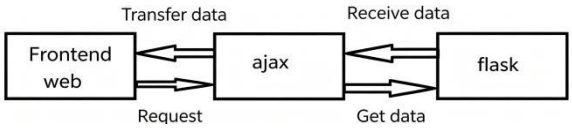

Figure 5. Backend API Request and Response Flow

4.2 Interactive Visualization Panel

The system's front-end utilizes the ECharts chart library [10] to create an intuitive, interactive visualization panel that enables users to flexibly query and analyze data, as shown in Figure 6.

Figure 6. Zhimu Huanxi Visualization Dashboard

5. Discussion

5.1 Application Value

Zhimu Huanxi transforms traditional post-hoc pollution loss statistics into forward-looking pre-warning prediction. Once meteorological authorities release heavy pollution alerts, local breeders can take targeted mitigation measures according to the platform's calculated output loss range: sealing livestock barns, upgrading air filtration ventilation systems or adding antioxidant additives to daily feed. For agricultural and environmental management departments, differentiated governance schemes can be formulated based on the varying weight of SO_2 and PM10 pollutants. Regions with high SO_2 load should prioritize low-sulfur fuel replacement and desulfurization equipment installation; areas with severe PM10 pollution need to focus on dust control and reducing straw burning activities. Administrators can also calculate annual total economic losses caused by air pollution through historical data retrieval, which provides quantitative basis for ecological compensation standard formulation.

5.2 Technical Advantages

Compared with mainstream single-function smart breeding platforms and environmental monitoring software, Zhimu Huanxi has three prominent technical strengths.

First, the multi-source heterogeneous data fusion module breaks the data silo problem existing in separated agricultural and environmental management systems. Unified timestamp alignment and standardized preprocessing guarantee consistent data quality, laying a solid foundation for subsequent joint modeling unavailable for most independent monitoring tools.

Second, the platform goes beyond simple pollutant concentration display and calculates the economic loss contribution of each pollutant factor. By combining Lasso feature screening and Linear SVR prediction, it establishes a direct quantitative link between pollutant concentration and breeding revenue loss, offering clear factor priority ranking for managers. For example, the platform verifies

SO_2 exerts stronger negative influence on output than NO_2, so pollution reduction projects should give priority to sulfur emission control. Such quantitative impact analysis capability is rarely equipped in existing visualization-only platforms.

Third, at the application level, this system adopts a hybrid delivery model combining "software platform + data services + technical consulting." This enables the system to flexibly provide either SaaS cloud services or localized private deployment solutions tailored to different user groups. Additionally, integrated technical consulting helps users better understand evaluation results and translate them into actionable decisions, thereby maximizing the system's value. This flexible and comprehensive service approach enhances both the system's market adaptability and user engagement.

5.3 Existing Limitations and Areas for Improvement

This study still has limitations in the following aspects, which require further improvement in subsequent research.

First, in the current system, Linear SVR regression prediction and time series trend prediction (e.g., ARIMA models) operate relatively independently, lacking deep integration and collaborative forecasting capabilities between the two models. Linear SVR excels at capturing causal relationships between features and targets, whereas time series models are better suited for identifying inherent trends and seasonality in data. Future research could explore developing an ensemble learning framework that combines the outputs of both models through weighted fusion techniques such as Stacking or Blending methods, thereby simultaneously capturing both the causal effects of pollution factors and the inertial trends of time series, potentially enhancing prediction accuracy and robustness.

Second, due to limitations in data availability, the modeling and validation of this study primarily relied on data from Inner Mongolia Autonomous Region covering the period 2020–2024. The geographical coverage and temporal span of the sample are relatively limited, and the model's generalization capacity under varying regional conditions (e.g., southern regions with significant climate differences), climatic variations, and livestock types (e.g., poultry, aquaculture) has not been adequately validated. Future research should expand data collaboration efforts by incorporating more diverse and extensive datasets, conduct cross-regional and cross-type model validation, and enhance the model's versatility across different environments through transfer learning or domain adaptation techniques.

Third, the current support vector machine linear regression model has inherent limitations when dealing with potentially nonlinear complex relationships. While LinearSVR performs well for linear relationships, there may be highly nonlinear interactions between livestock production value and environmental pollution. Future research could consider employing more robust nonlinear models for comparative experiments, such as random forests, gradient boosting trees (e.g., XGBoost, LightGBM), or lightweight deep neural networks. Recent studies have shown that dynamic machine learning models exhibit superior performance in ammonia emission prediction [11], and the combination of high-resolution emission inventories with machine learning methods holds significant value in environmental economic assessment [12,13].

6. Conclusion

The "Zhimu Huanxi" system integrates data streams from meteorology, environmental monitoring, and livestock production, employs Lasso regression to identify key pollution factors affecting output value, and utilizes Linear Support Vector Regression (SVR) for reliable prediction of output losses. On the 2024 test dataset, it achieved an R^2 of 0.78 and a Mean Absolute Error (MAE) of 152 million yuan. Compared to full-variable linear regression and unparameterized SVR models, this approach demonstrates significant improvements in both accuracy and stability. The system has been preliminarily deployed at livestock technology extension stations in selected counties of Inner Mongolia, providing farmers with a quantitative tool for addressing air pollution challenges. Future efforts will focus on cross-regional generalization validation, integration of nonlinear models, and enhancing early warning timeliness.

References

[1] Xie Q J, Zheng P, Bao J, et al. Temperature and Humidity Prediction Model in Closed Pig House Based on Deep Learning. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(10): 353-361.

[2] Liu C H, Yang L, Deng H, et al. Prediction of Ammonia Concentration in Pig House Based on ARIMA and BP Neural Network. China Environmental Science, 2019, 39(6): 2320-2327.

[3] Yang L, Liu C H, Guo Y C, et al. Prediction of Ammonia Concentration in Pig House Based on EMD-LSTM. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(S1): 353-360.

[4] Park J, Jo G, Jung M, Oh Y. Comparative Analysis of Neural Network Models for Predicting Ammonia Concentrations in a Mechanically Ventilated Sow Gestation Facility in Korea. Atmosphere, 2023, 14(8): 1248.

[5] Genedy R A, Chung M, Shortridge J E, et al.

A physics-informed long short-term memory (LSTM) model for estimating ammonia emissions from dairy manure during storage. Science of the Total Environment, 2024, 912: 168885.

[6] He T, Zhang W, Zhang H, Sheng J. Estimation of Manure Emissions Issued from Different Chinese Livestock Species: Potential of Future Production. Agriculture, 2023, 13(11): 2143.

[7] Li H. Flask Web Development in Action: Introduction, Advanced Techniques and Principles Analysis. Beijing: China Machine Press, 2018.

[8] Statistics Bureau of Inner Mongolia Autonomous Region. Inner Mongolia Statistical Yearbook 2024. Beijing: China Statistics Press, 2025.

[9] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 2011, 12: 2825-2830.

[10] Fan L Q, Zhang L J. Web Data Visualization: ECharts Version. Beijing: Posts & Telecom Press, 2021.

[11] Favrot A, Génermont S, Guigue V, et al. Improving ammonia emission predictions with dynamic machine learning models. EGUsphere, 2026.

[12] Lu S K, Chao L, Chen J Y, et al. Anthropogenic Ammonia Emission Inventory in Zhejiang Province from 2013 to 2020. China Environmental Science, 2022, 42(11): 1-12.

[13] Ma L, Ma W Q, Velthof G L, et al. Modeling Nutrient Flows in the Food Chain of China. Journal of Environmental Quality, 2010, 39(4): 1279-1289.