The pet prediction problem solved based on the Random Forest-ARIMA ensemble algorithm

Pengbin Wang, Zhenyu Hu*

Jiangsu Normal University KeWen College, JiangSu, China

ABSTRACT

With the enhancement of consumer awareness, the pet industry market size has been steadily increasing, becoming a leading emerging sector. As more people treat pets as family members, the demand for pet-related products and services continues to rise, making accurate forecasting of market trends essential for industry growth. This paper aims to explore the future development trends of the pet industry and formulate sustainable planning strategies by employing flexible prediction methods based on historical data. To address the research needs, this paper first utilizes **Python**-based web scraping techniques to deeply mine relevant industry data and compile an "initial dataset". This dataset, however, contains missing values, which are filled using the mode imputation method. Subsequently, the built-in find function in **MATLAB** is applied to traverse the dataset files and ensure that there are no outliers present. After processing the data, correlations between various indicators are examined, and relevant features are extracted for further analysis. To predict the number of pet cats and dogs in China over the next three years, a **Random Forest-Multiple Linear Regression-ARIMA** integrated model is constructed. The results from this model show that the pet **cat** population will be approximately 65 million, 57 million, and 90 million in the coming years, while the pet **dog** population will be around 52 million, 54 million, and 52 million. These predictions provide valuable insights into the future of the pet industry, helping businesses and policymakers plan and strategize for sustainable growth.

Keywords: Python; Random Forest; XGBoost; ARIMA; Pet cats and dogs

1 INTRODUCTION

With the rapid progress of the global economy and rising per capita income, people's consumption concepts have gradually shifted, with increasing attention to emotional and mental needs. As an emerging industry, the pet sector has begun to grow. The Civil Code of the People's Republic of China, which came into effect in 2021, explicitly stipulates liability for damages caused by animals under the section on tort liability, emphasizing pet owners' responsibility for their pets' behavior and reflecting the protection of pet rights. In recent years, driven by sustained economic growth and the widespread adoption of the "pet companionship" concept, China's pet industry has demonstrated strong market potential [1].

However, the progress of China's pet industry also faces challenges, including intensified market competition, demographic changes affecting demand, and the lack of comprehensive policies and standards. Therefore, analyzing the current fiscal environment and market demand to explore the future progress trends of China's pet industry and formulating sustainable industrial strategies is of significant importance.

2 RELEATED WORK AND ASSUMPTION

The study analyzed changes in the Chinese pet market from 2019 to 2023, observing an increase in pet cats and a decline in pet dogs, driven by the lower cost and urban adaptability of cats. Correlation analysis found positive links with market size and pet food expenditure, and negative links with the Gini coefficient and tax rate. Predictions using Random Forest, linear regression, and ARIMA models indicated continued growth in pet cats and stable pet dog numbers.

Therefore, this paper makes the following assumptions

(1) It is assumed that the number of pets and the market size in the dataset are significantly correlated, accurately reflecting the overall trends and changes in the pet industry.

(2) It is assumed that models such as Random Forest, Multiple Linear Regression, ARIMA, and XGBoost can effectively capture the trends and features in the data, enabling accurate predictions of future changes in the pet market.

(3) It is assumed that economic indicators (e. g. , GDP per capita, Gini coefficient, total tax rate) have a significant impact on the size of the pet market and the number of pets, making them effective predictive factors in the models.

3 MODEL ESTABLISHMENT AND SOLUTION

3.1 Data sources

This study begins with a comprehensive data collection process, utilizing a range of search engines such as Google, Bing, and GitHub to locate relevant websites and resources. In addition, specialized platforms were explored to gather detailed information on key metrics, including the pet food market size, U.S. pet household penetration rates, and total pet food export values. These sources provided a foundational dataset, which was then expanded by integrating international policy data from various European and Asian countries. This combination of data points was crucial for constructing a robust preliminary dataset, offering a holistic view of the pet food industry from both market and regulatory perspectives. The preliminary dataset was subsequently refined to ensure accuracy and relevance, serving as the basis for deeper analysis of the global pet food market dynamics.

3.2 Data cleaning

The Paper initially conducted a manual inspection of the preliminary dataset, identifying the presence of missing values. The dataset was then imported into MATLAB, where the built-in `find` function was employed to traverse the files and systematically detect missing entries. Advanced statistical methods, including the coefficient of variation, Mann-Kendall trend analysis, and normalization techniques, were applied for comprehensive data processing.

Missing values, such as those for U. S. pet food consumption and annual per-pet expenditure, were imputed using mode substitution to ensure data completeness and reliability.

Table 1: Original data set (part)

Year	Population aged 20-24	Population aged 60-64	……	Number of pet raising families in China
2019	61519	60712	……	5989
2020	None	None	……	6369
2021	77256	70755	……	9168
2022	73629	71964	……	9800
2023	None	None	……	None

Table 2: Filled data set (part)

Year	Population aged 20-24	Population aged 60-64	……	Number of pet raising families in China
2019	61519	60712	……	5989
2020	61519	60712	……	6369
2021	77256	70755	……	9168
2022	73629	71964	……	9800
2023	61519	60712	……	5989

After imputing missing values, further analysis was conducted to identify outliers, revealing no anomalies in the dataset

3.3 Correlation analysis of the collected data

This article uses Python's "**seaborn**" library to create a heatmap of the correlations between cats, dogs, and various indicators, as shown in Fig.4.

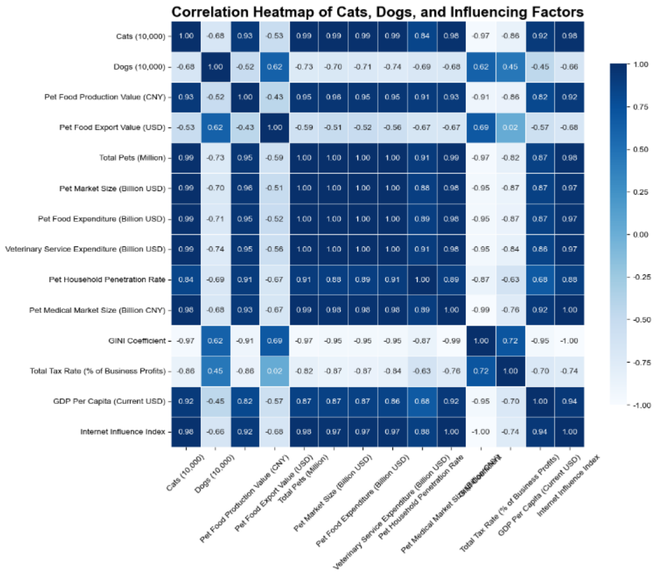

Fig.4: Correlation Heatmap of Cats, Dogs, and Factors

Fig.4, in the form of a correlation heatmap, illustrates the correlations between the numbers of cats, dogs, and their related factors. The horizontal and vertical axes represent the numbers of cats, dogs, and various associated factors. From the Fig., it can be observed that the number of cats has a strong positive correlation (>0.9) with variables such as pet market size and pet food expenditure, indicating that the number of cats is driven by the overall market scale [2]. Conversely, it has a strong negative correlation (<-0.8) with the Gini coefficient and total tax rate, suggesting that unequal income distribution and higher tax rates suppress the growth of the cat population. Pet market size shows an almost perfect positive correlation with factors such as pet food expenditure and pet medical market size, indicating that market expansion is primarily driven by spending on medical care and food.

Based on the analysis of the correlation heatmap above, the increase in pet food expenditure, rising tax rates, growth in veterinary service costs, and the increasing pet population collectively reflect the continuous expansion of the pet market size.

3.4 Data analysis

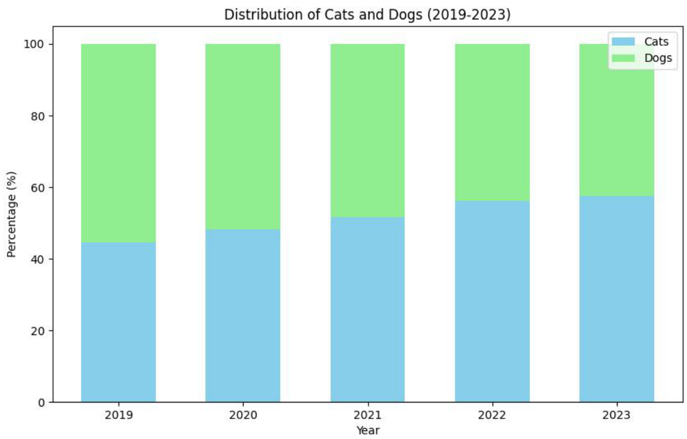

Fig.5: Number of cats and dogs stacked in the pet market

Fig.5 illustrates the dynamic distribution of cats and dogs in China's pet market from 2019 to 2023. The horizontal axis represents the years, while the vertical axis indicates the proportion of cats and dogs in the total pet population. Over time, the proportion of cats rose from 44. 5% to 57. 4%, with an annual growth rate of 10. 2%. In contrast, the proportion of dogs decreased from 55. 5% to 42. 6%, declining at an annual rate of 5. 1%. The 12. 9% rise in the proportion of cats is assigned to the economic advantages of raising cats and their better adaptability to urban living environments.

3.5 Random Forest Multiple Linear Regression - ARIMA Integrated Model

This article first uses Random Forest to extract indicators closely related to the pet industry from the dataset, as shown in the table below.

Table 3: shows the importance analysis of 6 pet indicators extracted by features

Index	Value (Cat/Dog)	Index	Value (Cat/Dog)

Total tax rate	0. 13 / 0. 18	Rural population	0. 09/ 0. 07
Total output value of pet food in China	0. 11 / 0. 12	Per capita GDP（U. S. dollar）	0. 06 / 0. 10
Gini coefficient	0. 10 / 0. 10	Food production index（2014-2016）	0. 08 / 0. 07

Based on the analysis of Table 3, the values highlight the relative significance of the pet market and its associated indicators. Notably, the total tax rate exhibits an importance score of approximately 0.18, underscoring its substantial impact on the pet market. Conversely, the food production index holds a comparatively lower importance score of 0.07, indicating a relatively modest influence. Meanwhile, the remaining indicators, with importance scores clustering around 0.09, suggest a moderate yet noteworthy contribution to the dynamics of the pet market.

(1) The establishment of the Random Forest model

Random Forest is an ensemble learning algorithm based on decision trees, commonly used for classification and regression tasks [3]. It improves the accuracy and robustness of the model by constructing multiple decision trees and determining the final prediction result through majority voting (for classification) or averaging (for regression).

1) Sample Selection.

The first step in constructing a Random Forest is sample selection. From the original training set S containing N samples, the bootstrap method is used to generate m samples, with the sample set denoted as S_m. The sample selection formula for each tree can be expressed as:

$$S_m = \{x_i \mid i \in \text{ Random Sample of Size } m \subseteq S\} \tag{1}$$

2) Feature Selection.

When splitting at each node, features are randomly selected. For example, m features are chosen ($m < M$, where M is the total number of features), and the best feature for splitting is selected based on certain criteria (such as information gain, Gini index, etc.).

$$X_f = \left\{f_i \mid i \in \text{ Random sample of Size } m \subseteq \{1, 2, \ldots, M\}\right\} \tag{2}$$

3) Construction of Decision Trees.

For each decision tree T_j, the final model for tree construction can be expressed as:

$$Y_j = T_j(X) \tag{3}$$

Where Y_j is the output of the j-th tree, and X represents the input features.

4) Final Prediction Integration.

For classification tasks: Use the majority voting method (mode) as the final prediction:

$$\hat{Y} = \text{mode}\left(Y_1, Y_2, \ldots, Y_J\right) \tag{4}$$

Where J is the total number of trees.

For regression tasks: Use the average of the predictions from all trees as the final prediction:

$$\hat{Y} = \frac{1}{J}\sum_{j=1}^{J} Y_j \tag{5}$$

(2) Establishment of Multiple Linear Regression

Multiple linear regression is a statistical method used to study the linear relationship between a dependent variable (target variable) and multiple independent variables (predictor variables). The goal of multiple linear regression is to find a set of regression coefficients that minimize the error between the predicted values and the actual values, thereby establishing a predictive model [4].

1) Basic Form of Multiple Linear Regression:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon \tag{6}$$

Where y is the dependent variable (target variable), $x_1, x_2, \dots, x_n$are the independent variables (predictor variables), β_0 is the intercept term, representing the value of the dependent variable when all independent variables are $\beta_0, \beta_1, \beta_2, \dots, \beta_n$ are the regression coefficients, indicating the impact of each independent variable on the dependent variable, and ϵ is the random error term, representing the part of the model that is not explained [5].

2) Model Matrix Form

Multiple linear regression can be expressed in matrix form as:

$$y = X\beta + \epsilon \tag{7}$$

Where $y: n \times 1$ is dependent variable vector, representing the target values of n samples. $X: n \times (p+1)$ design matrix, containing n samples, p independent variables, and a constant term.

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \tag{8}$$

Where $\beta: (p+1) \times 1$ regression coefficient vector. $\epsilon: n \times 1$ random error vector.

3) Objective Function

Multiple linear regression estimates the regression coefficients β by minimizing the Mean Squared Error (MSE):

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \frac{1}{n} \parallel y - X\beta \parallel^2 \tag{9}$$

4) Solution Using Least Squares Method

By minimizing the objective function, the optimal solution for the regression coefficients is obtained as:

$$\hat{\beta} = (X^T X)^{-1} X^T y \tag{10}$$

Where $X^T X$: Covariance matrix of the independent variable matrix. $(X^T X)^{-1}$: Inverse of the covariance matrix. $X^T y$: Inner product of the independent variable matrix and the dependent variable vector.

5) Prediction Formula

Using the regression coefficients $\hat{\beta}$, new input data can be used for prediction:

$$\hat{y} = X\hat{\beta} \tag{11}$$

(3) ARIMA Model

The ARIMA (Auto-Regressive Integrated Moving Average) model is a type of time series prediction model used to capture trends, seasonality, and random fluctuations in data [6]. It is widely applied in the fields of economics, finance, and statistics. ARIMA combines three mechanisms: Auto-Regressive (AR), Integrated (I), and Moving Average (MA). This enables the model to construct predictions for both stationary and non-stationary time series data [6].

The Three Core Components of ARIMA:

1) AR (Auto-Regressive):

Uses the past values of the time series itself to make predictions, relying on the autocorrelation of the series.

2) I (Integrated):

Transforms non-stationary series into stationary ones through differencing.

3) MA (Moving Average):

Models the random fluctuations of the series by utilizing past prediction errors to create a smoother time series [7].

AR Model (Auto-Regressive Model)

The AR model assumes that the current value is determined by a linear combination of the past p values plus a noise term:

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \epsilon_t \tag{12}$$

Where X_t: The value of the time series at time $t. c$: Constant term. $\phi_1, \phi_2, \ldots, \phi_p$: Auto-regressive coefficients. ϵ_t: Random noise (white noise).

MA Model (Moving Average Model)

The MA model assumes that the current value is a linear combination of the past q error terms:

$$X_t = c + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} \tag{13}$$

Where $\theta_1, \theta_2, \ldots, \theta_q$: Moving average coefficients. ϵ_t: White noise.

3.6 Solving ensemble model

3.6.1 Steps and Prediction Results of Random Forest Solution

(1) Data Normalization

Normalize the original data to the $[0, 1]$ range to eliminate the influence of different feature magnitudes:

$$X_{\text{normalized}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{14}$$

(2) Lagged Features Creation

Generate lagged features for the time series to capture historical information:

$$\text{lag}_i = X_{t-i}, \forall i \in \{1, 2, \ldots, \text{lag}\} \tag{15}$$

(3) Feature Matrix

The lagged feature matrix is expressed as:

$$X_{\text{lagged}} = \begin{bmatrix} X_{t-1} & X_{t-2} & \cdots & X_{t-\text{lag}} \\ X_t & X_{t-1} & \cdots & X_{t-\text{lag}+1} \\ \vdots & \vdots & \ddots & \vdots \end{bmatrix} \tag{16}$$

(4) Random Forest Feature Selection

Random forest calculates feature importance to select key features:

$$\text{Feature Importance } = \frac{\text{Reduction in Error (MSE)}}{\text{Total Features}} \tag{17}$$

The higher the importance score, the more significant the feature. Select the k most important features for further analysis.

(5) Cross-Validation

Using leave-one-out cross-validation (LOO-CV) to evaluate the generalization performance of the model:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{18}$$

Where: y_i: Actual value and $\hat{y}_i$: Predicted value.

(6) Prediction

The prediction of the random forest model is determined by averaging the predictions of multiple trees:

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} \hat{y}_t \tag{19}$$

Where: $\hat{y}_t$: Prediction of the t-th decision tree.

Future values are generated based on lagged features and model predictions:

$$X_{t+1} = f\left(X_t, X_{t-1}, \ldots, X_{t-\text{lag}}\right) \tag{20}$$

Future inputs are updated iteratively to produce step-by-step predictions.

(7) Mean Squared Error (MSE)

To measure the bias between the predicted and actual values:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{21}$$

(8) Inverse Transformation of Predictions

Transform normalized predicted values back to the original data scale:

$$X_{\text{original}} = X_{\text{normalized}} \cdot (X_{\max} - X_{\min}) + X_{\min} \tag{22}$$

(9) Time Series Extension

Extend the time series to include future years and their predicted values:

$$\begin{gathered} \text{Years Extended} = [\text{Original Years, Future Years}] \\ \text{Values Extended} = [\text{Original Values, Predicted Values}] \end{gathered} \tag{23}$$

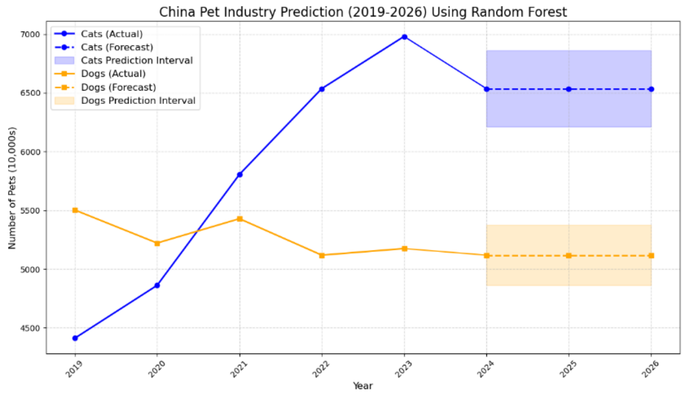

Fig.6: Cat and Dog Population Prediction by Random Forests

Fig.6 predicts the trends in the number of pet cats and dogs in China from 2024 to 2026 using the Random Forest model. The prediction results are evaluated using the Mean Squared Error (MSE), with an MSE value of approximately 0.025, indicating minimal error between the predicted and actual values. The horizontal axis represents the years, while the vertical axis shows the scale of pet cats and dogs in China. The prediction indicates that the numbers of both pet cats and dogs will stabilize from 2024 to 2026, with pet cats maintaining a level of 65 million and pet dogs around 52 million. The increase in both numbers reflects the growing demand in China's pet market [8].

3.6.2 Steps for Solving and Predicting Using Multiple Linear Regression

(1) Data Normalization

To eliminate the impact of scale differences, Min-Max normalization is used:

$$X_{\text{normalized}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{24}$$

Where: X: Original feature values. $X_{\min}$: Minimum value of the feature. $X_{\max}$: Maximum value of the feature.

(2) Calculation of Regression Coefficients β

The regression coefficients β are determined by minimizing the Mean Squared Error (MSE):

$$\beta = (X^T X)^{-1} X^T y \tag{25}$$

(3) Real-Time Prediction

The target prediction value $\hat{y}$:

$$\hat{y} = X\beta \tag{26}$$

Where: X: Input feature matrix. β: Regression coefficient vector.

(4) Mean Squared Error (MSE) Calculation

MSE is used to measure the prediction error of the model:

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{27}$$

Where: y_i: True value. $\hat{y}_i$: Predicted value. n: Number of samples.

(5) Prediction Error Adjustment

Future prediction values are adjusted using an error correction term:

$$y_{\text{future },i} = \hat{y}_{\text{future },i} + i \cdot MSE \cdot \alpha \tag{28}$$

Where: i: Prediction step (starting from 1). MSE: Model prediction error. α: Adjustment coefficient (set to 0. 02 in the code).

(6) Inverse Normalization

Predicted values in normalized form are converted back to their original scale:

$$X_{\text{original }} = X_{\text{normalized }} \cdot (X_{\max} - X_{\min}) + X_{\min} \tag{29}$$

(7) Feature Mean Calculation

The mean of normalized features is calculated to generate future prediction features:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{30}$$

(8) Time Series Extension

Extend the time series with future years:

$$\begin{aligned} \text{Years Extended } &= [\text{ Original Years, Future Years }] \\ \text{Values Extended } &= [\text{ Original Values, Predicted Values }] \end{aligned} \tag{31}$$

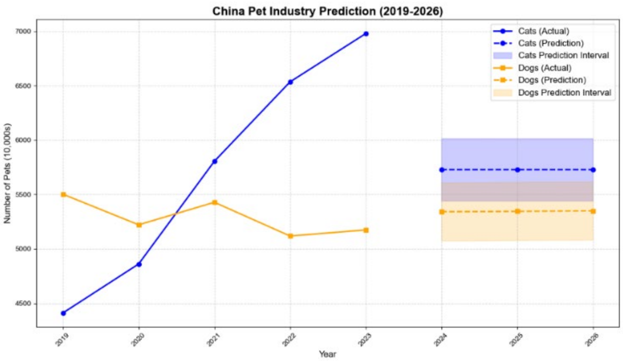

Fig.7: Predicted Cat and Dog Population Changes via Linear Regression

Fig.7, based on a multiple linear regression model, predicts the changes in the number of pet cats and dogs in China from 2024 to 2026. The prediction results are evaluated using MSE, with the MSE for domesticated cats being 0.00037 and for pet dogs being 0.48, indicating minimal errors between the actual and predicted values. Depending on the predictions, the number of pet dogs remains stable at approximately 54 million during 2024–2026. However, between 2023 and 2024, the number of pet cats drops sharply from 68.15 million to 57 million, indicating data anomalies and reflecting certain errors in the model's handling of this dataset. This suggests there is still for improvement in the predictive performance of the multiple linear regression model.

3.6.3 The ARIMA modeling process and prediction results are as follows

(1) Normalization Formula

The data is normalized to the range $[0, 1]$ using MinMaxScaler:

$$X_{\text{normalized}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{32}$$

Where X is the original feature value, $X_{\min}$ is the minimum feature value, and $X_{\max}$ is the maximum feature value [9].

(2) Differencing Formulas

Differencing is applied to stabilize the time series and eliminate trends:

First-order differencing:

$$\text{Diff}_1 = X_t - X_{t-1} \tag{33}$$

Second-order differencing:

$$\text{Diff}_2 = \text{Diff}_1(t) - \text{Diff}_1(t-1) \tag{34}$$

(3) ADF Test for Stationarity

The Augmented Dickey-Fuller (ADF) test is used to check the stationarity of the time series:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \cdots + \delta_k \Delta y_{t-k} + \epsilon_t \tag{35}$$

Where ADF Statistic: Indicates whether a unit root is present. p-value: Less than 0. 05 suggests the time series is stationary.

(4) ARIMA Model Formula

The ARIMA model consists of three components: autoregression (AR), differencing (I), and moving average (MA):

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \cdots - \theta_q \epsilon_{t-q} \tag{36}$$

Where y_t is Time series value at time $t. c$: Constant term. ϕ: Autoregression coefficients. θ: Moving average coefficients. ϵ_t: White noise.

(5) Prediction Formula

The ARIMA model predicts future values as follows:

$$\hat{y}_{t+h} = c + \sum_{i=1}^{p} \phi_i y_{t+h-i} - \sum_{j=1}^{q} \theta_j \epsilon_{t+h-j} \tag{37}$$

Where h: Prediction horizon. $\hat{y}_{t+h}$: Predicted value at future time $t + h$.

(6) Inverse Transformation Formula

The normalized prediction results are transformed back to the original scale:

$$X_{\text{original}} = X_{\text{normalized}} \times (X_{\max} - X_{\min}) + X_{\min} \tag{38}$$

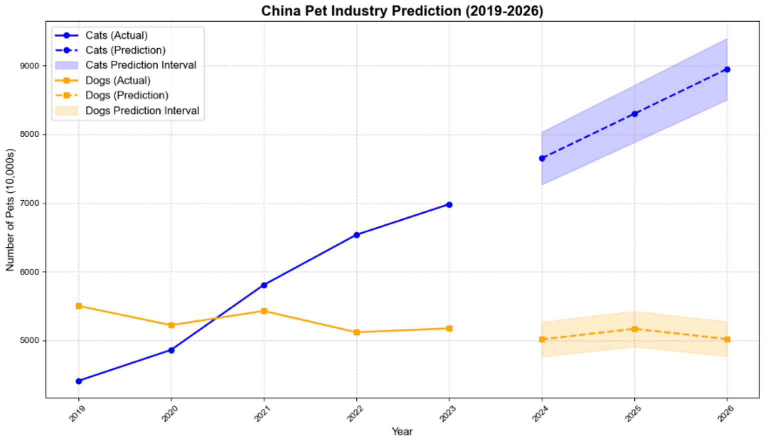

Fig.8: Line chart of predicted cat and dog populations based on ARIMA

Fig.8 uses the ARIMA model to predict the future changes in the number of pet cats and dogs in China from 2024 to 2026. According to the prediction results, the number of pet cats will increase year by year, reaching approximately 90 million by 2026, while the number of pet dogs will remain stable at around 52 million. This suggests a significant rise in the popularity of pet cats, which may be influenced by changing consumer preferences and urban living conditions. The number of pet dogs, on the other hand, reflects a more mature market. The continuous growth in both categories highlights the ongoing expansion of China's pet market, driven by increased disposable incomes and a greater focus on pet care and companionship.

Table 4: evaluates the P-value of the ARIMA model

Cats p-value (Diff)	0.42811060946796686
Dogs p-value (Diff)	0.7929541291227802

Table 4 evaluates the prediction results using p-values, where the p-value for pet cats is 0. 43 and the p-value for pet dogs is 7.93e-08, indicating that both predictions exhibit high stability.

3.7 Model detection

In statistics, the margin of error (ME) is generally related to the confidence interval and standard error. Its basic formula is expressed as:

$$ME = Z \times SE \tag{39}$$

Where: ME is Margin of Error. Z is he z-value corresponding to the selected confidence level in the standard normal distribution.

For example, for a 95% confidence level, the z-value is typically 1.96; for a 99% confidence level, the z-value is typically 2. 576. SE is Standard Error, which represents the variability of the sample mean [10].

Table 5: MSE of Random Forest and Linear Regression

Model	MSE value (Cat/Dog)
Random forest	0. 030 / 0. 021
Multiple linear regression	0. 00037 / 0. 48

The population of pet cats and dogs experienced consistent year-on-year growth from 2019 to 2023. Projections for 2024 to 2026 suggest an accelerated growth trajectory for pet cats, while the population of pet dogs is expected to plateau. This trend underscores the continued expansion of China's pet market, driven in part by the rising preference for feline companions.

4 CONCLUSION

The pet market in China is expected to continue growing, with a particularly significant increase in the number of pet cats. Data analysis and forecasting models, such as Random Forest, Multiple Linear Regression, and ARIMA, all show that the pet industry will maintain an expanding trend.

Next, the study identifies various factors driving this growth, including the rising expenditure on pet food, veterinary services, and the increasing number of pet owners. As these factors continue to evolve, the scale and influence of the pet industry will steadily expand.

Additionally, the projections for 2024 to 2026 indicate that while the pet dog population will remain stable, the pet cat population is expected to experience significant growth, reaching approximately 90 million by 2026. This trend reflects the ongoing shift in the pet market, with the focus on the growth of the cat population in the coming years.

The conclusion emphasizes the importance of these predictions for businesses and policymakers. Accurate market forecasting will help them formulate effective strategies to ensure the sustainable development of the pet industry and to address the challenges and changes that may arise in the future.

REFERENCES

[1] Qin, Y. J. (2024). A Comparative Study of Housing Price Prediction Models Based on Multiple Linear Regression and Random Forest Algorithm. Modern Information Technology, 22, 127-131. doi:10.19850/j.cnki.2096-4706.2024.22.025.

[2] He, X. F., He, H. H., Yang, L., Yu, Y., Jiang, M. F., & Zhang, T. (2024). Analysis of Tobacco Sales Influencing Factors Based on Random Forest Model. Information Technology, 11, 147-153. doi:10.13274/j.cnki.hdzj.2024.11.022.

[3] Zhang, X., & Li, L. (2024). Seasonal Electric Vehicle Charging Load Prediction Based on Random Forest. Software Engineering, 11, 11-14+37. doi:10.19644/j.cnki.issn2096-1472.2024.011.003.

[4] Lu, X. Y. (2024). A Study on the Prediction Method of VOCs Content in Coatings Based on Multiple Linear Regression Model. Popular Standardization, 20, 181-183.

[5] Sun, F. N., & Zhang, Z. J. (2024). A Model for the Influence of Logistics Demand in Inner Mongolia Based on Multiple Linear Regression Method and Analysis of Its Influencing Factors. Chinese Business Theory, 17, 99-103. doi:10.19699/j.cnki.issn2096-0298.2024.17.099.

[6] Zhang, Z. Q., & Du, J. (2024). Port Logistics Demand Prediction Analysis of Ningbo City Based on Double Exponential Smoothing and Multiple Linear Regression. Logistics Technology, 17, 78-82. doi:10.13714/j.cnki.1002-3100.2024.17.020.

[7] Qi, P. Y., Yao, X. W., Liu, Q. H., Xu, K. Q., Ren, H. F., & Xu, K. L. (2024). Prediction of Ash Melting Characteristics Temperature for Biomass and Bituminous Coal Co-firing Based on Multiple Linear Regression Model. Agricultural Engineering Journal, 15, 174-182.

[8] Cui, Y. H., Zhu, Z. H., & Li, T. (2024). Forecasting of Freight Turnover in Shijiazhuang Based on Grey Prediction-ARIMA Model. Logistics Technology, 22, 8-11+18. doi:10.13714/j.cnki.1002-3100.2024.22.002.

[9] Li, Q. J., Wang, N. L., Li, W. X., Yin, D. P., Jin, Y., Qiu, L., & Lu, Y. (2024). Study on the Prediction of Influenza-like Cases in Hainan Province Based on ARIMA Model. Journal of Xinjiang Medical University, 11, 1533-1538.

[10] Wu, C. Y. (2024). Forecasting Analysis of Per Capita Disposable Income Based on ARIMA-GM(1, Combined Model. Journal of Higher Education Science, 08, 50-56.