

Research on China's Birth Rate based on BP Neural Network

Jinyu You, Hongzhen Chen, Cheng Chen and Hanchun Yang *

School of Mathematics and Statistics, Yunnan University, Kunming, China

* Corresponding Author Email: ma.hyang@163.com

Abstract. In terms of the impact of population age structure on birth rate, some studies have shown that compared to the reference category of respondents aged 30-34, young respondents are less likely to not have children, while older recipients have a more common unwillingness to have children; Women of childbearing age consider their own physical condition and reproductive safety factors. The older their parents are, the lower their willingness to have children, and after having one child, they may not be willing to have a second child; Moreover, if there are more elderly people in the family and the social pension security system and system are not perfect, the willingness of the family to have more children will decrease.

Keywords: Birth Rate; Backward Stepwise Regression Model; BP Neural Network.

1. Background and Significance of the Study

1.1. Background

Since the establishment of the People's Republic of China, the total population of our country has shifted from rapid development to stable development. After a long and tortuous development, China's population development can be roughly divided into four stages: high-speed growth stage (1949-1970), regulated growth stage (1971-1980), growth rate recovery stage (1981-1990), and stable growth stage (1991-2022).

The low birth rate also means a decrease in population dividends, as well as the prominent issue of aging. The problem of "getting old before getting rich" has become increasingly severe, and the problem of low birth rate has become a huge challenge faced by China's social development. Therefore, addressing this issue is urgent.

1.2. Significance

Our team has studied the changes in China's birth rate during different periods and birth policies, summarized their patterns through models, and provided relevant suggestions to relevant government departments in response to the current unsatisfactory birth rate level in China. Ultimately, the goal of maintaining the birth rate at a satisfactory level is to enjoy the demographic dividend without causing a crisis to the sustainable development of society.

Analyzing the influencing factors behind low birth rates through mathematical models is beneficial for the country to adjust and deploy relevant aspects, and has great practical and social significance: from an economic perspective, solving the problem of low birth rates can improve labor supply and facilitate social wealth creation; From the perspective of political science, solving the problem of low birth rate can increase social stability and be beneficial for the long-term development of the country and nation; From the perspective of social security, addressing the issue of low birth rate can reduce inter-generational support conflicts and contribute to the stability of the social security system. Therefore, our research is of great significance.

2. Quantitative Analysis of Factors Influencing Birth Rate based on Backward Stepwise Regression Model

This study collected data on influencing factors related to China's birth rate, and incorporated a large number of factors into the model, which may lead to model complexity and overfitting. Therefore, a backward stepwise regression method was used to first delete some noise and similar variables, thereby reducing the number of input variables for modeling and simplifying the process of model establishment and solution.

2.1. Model Introduction

The basic idea of the backward stepwise regression analysis method is to automatically select the most important variable from a large number of available variables and establish a prediction or interpretation model for regression analysis. The basic idea is to put all variables into the model, and then try to remove a certain variable to see if there is a significant change in the entire model after removal (F-test). If there is no significant change, remove it, and if there is, keep it until all factors that have a significant change in the model are left. The independent variables will be removed in descending order of contribution.

After gradual regression, the explanatory variables retained in the model are both important and do not have severe multicollinearity.

2.2. Data Preparation

Based on actual life and online literature, we collected data from 2010 to 2020, including birth rate, gross national income, average selling price of commercial housing, per capita GDP index, population structure (proportion of population aged 15-64), urbanization level, Male to female ratio, marriage rate, divorce rate, and number of women of childbearing age aged 20-34.

In order to facilitate data analysis, we have processed the gross national income and per capita gross domestic product. Based on the data from 1978, it is recorded as 100, and the data from other years are calculated in proportion to it to obtain the Gross National Income Index and Per Capita Gross Domestic Product Index.

Table 1. A stepwise regression model for influencing factors of birth rate

Independent variable	Dependent variable
Marriage rate (X_1)	Birth rate (Y)
Divorce rate (X_2)	
Male to female ratio (X_3)	
Urbanization level (X_4)	
Gross National Income Index (X_5)	
Average selling price of commercial housing (X_6)	
Number of women of childbearing age aged 20-34 (X_7)	
Per capita GDP index (X_8)	
Population structure (proportion of population aged 15-64) (X_9)	

2.3. Model Building

For the dependent variable Y , there are p independent variables X_1, X_2, \dots, X_p ($p = 9$). Let these p independent variables and the dependent variable Y establish a multiple regression model

$$Y = \beta_0 + \sum_{i=1}^p \beta_i X_i + \varepsilon. \quad (1)$$

Perform F-test on all independent variables X_i and the corresponding regression coefficient β_i , taking the minimum value

$$F_{k_1}^1 = \min \{F_1^1, F_2^1, \dots, F_p^1\}, \quad (2)$$

If $F_{k_1}^1 > F_\alpha(1, n-p-1)$, then the equation is already optimal; If $F_{k_1}^1 \leq F_\alpha(1, n-p-1)$, we need to remove the independent variables and proceed to the next step.

Take the remaining variables as independent variables X_1, X_2, \dots, X_{p-1} and continue to establish a multiple regression equation as above. Perform F-test on all independent variables X_i and corresponding regression coefficient β_i in the equation, taking the minimum value

$$F_{k_2}^2 = \min \{F_1^2, F_2^2, \dots, F_{p-1}^2\}, \quad (3)$$

If $F_{k_2}^2 > F_\alpha(1, n-(p-1)-1)$, then the equation is already optimal; If $F_{k_2}^2 \leq F_\alpha(1, n-(p-1)-1)$, you need to remove the independent variables X_{k_2} and proceed to the next step.

Continue iterating until the regression coefficient test values of each variable are greater than the critical value, meaning that there are no variables in the equation that can be removed. At this point, the regression equation is the optimal regression equation, and the remaining independent variables in the equation are the key variables we want to obtain.

2.4. Model Calculation Results

After iterative calculation, the obtained model results are shown in the following table 2 and figure 1

Table 2. Backward stepwise regression model result table

	Unstandardized Coefficients		Standardized Coefficient	t	P	R ²	Adjust R ²	F
	B	standard error	Beta					
constant	-201.01	39.035	0	-5.149	0.002***	0.851	0.868	F=17.374, P=0.002***
Per capita GDP index	-0.023	0.009	-5.51	-2.448	0.050**			
Population structure (proportion of population aged 15-64)	1.661	0.354	1.748	4.689	0.003***			
Urbanization level	2.465	0.836	6.482	2.95	0.026**			
Male to female ratio	-0.016	0.007	-0.285	-2.316	0.060*			
Dependent variable: Birth rate								

Note: ***, **, * represent significance levels of 1%, 5%, and 10% respectively

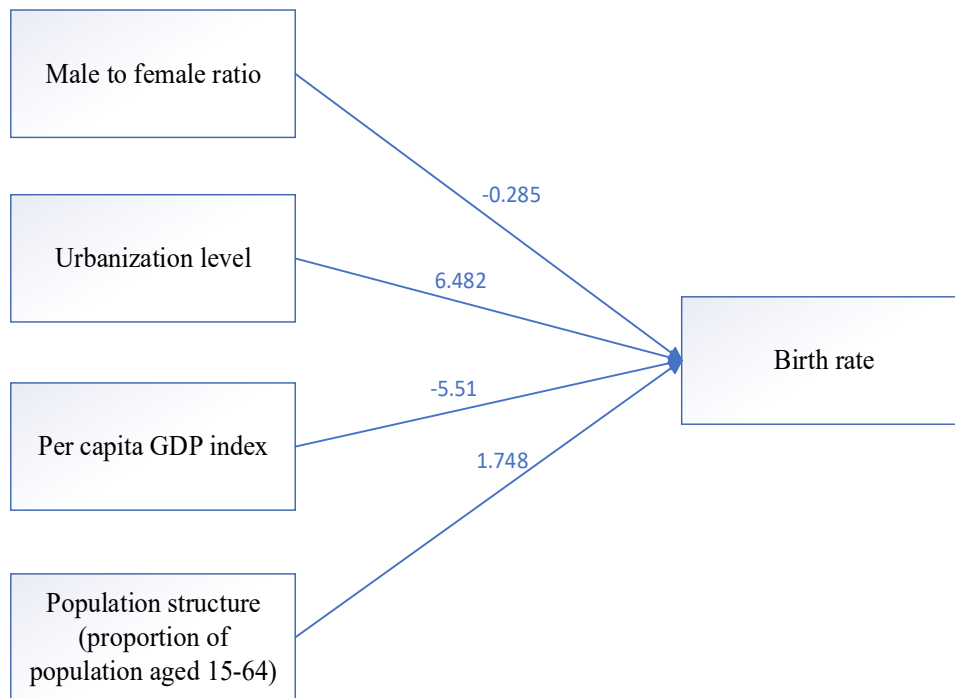


Fig 1. Model Path Map

2.5. Analysis of Measurement Results

As shown in Table 2, $R^2 = 0.851$, it is already very close to 1, indicating that the model has a good fitting effect. From the analysis of the results of the F-test, it can be concluded that the significance P-value is 0.002^{***} , indicating significant levels, rejecting the null hypothesis of a regression coefficient of 0. The overall significance of the model.

From Figure 3-1, it can be seen that the male to female ratio and per capita GDP index have a negative impact on the birth rate, meaning that as these two values increase, the birth rate will decrease accordingly. The level of urbanization has a significant impact on the birth rate, but it has a positive impact on the birth rate, which may differ from the intuitive representation of the data. Therefore, this model has some problems and the prediction effect may not be ideal.

However, this model provides us with effective assistance in selecting key influencing factors from a multitude of dependent variables, and in establishing new and appropriate models in the future.

3. Qualitative Research based on Literature Research

3.1. The Impact of Male to Female Ratio on Birth Rate

Due to the reproductive responsibility of women, the size of women of childbearing age is one of the important factors affecting the birth rate [1]. The high sex ratio at birth not only means that the number of live births of female babies has been artificially reduced, but also inevitably affects the number of future women of childbearing age. Against the backdrop of declining birth rates, the scale of women of childbearing age has significantly decreased, and the imbalance in the gender ratio of the birth population undoubtedly accelerates the shrinkage of the number of women of childbearing age in the future, further reducing the birth rate.

3.2. The Impact of Urbanization Level on Birth Rate

The direct result of the improvement of urbanization level is the increase of migrant population. The floating population in the context of urbanization is mainly composed of rural to urban migration. Due to the differences in fertility policies and levels between rural and urban areas in China, it can be

preliminarily inferred that the fertility level of rural to urban migrants is showing a downward trend compared to their source areas, which also leads to a decrease in the overall birth rate.

The theory of separation regards migrant populations as a type of population in a state of dissociation. The migration population is separated from the rural population, and the influence and control of rural social environment and customs are weakened. However, they have not yet integrated into urban life, lack dependence on the city, and are not deeply influenced by the behavior and beliefs of the urban population. Therefore, they are in a state of freedom and intermediation, and their birth rate is usually higher than that of permanent immigrants and local populations in urban destinations, but lower than that of rural populations [2].

3.3. The Impact of Per Capita GDP Index on Birth Rate

As family income increases, raising children gradually becomes an "uneconomical" behavior. Leibenstein (1957) and Becker (1960) used consumer behavior theory to conduct an economic analysis of raising children. They believed that children are a special consumer product, and that reproductive behavior is a response of consumers to their children's needs [3].

At the same time, modernization factors have led to delayed childbirth and fewer children.

3.4. The Impact of Population Structure on Birth Rate

In terms of the impact of population age structure on birth rate, some studies have shown that compared to the reference category of respondents aged 30-34, young respondents are less likely to not have children, while older recipients have a more common unwillingness to have children; Women of childbearing age consider their own physical condition and reproductive safety factors. The older their parents are, the lower their willingness to have children, and after having one child, they may not be willing to have a second child; Moreover, if there are more elderly people in the family and the social pension security system and system are not perfect, the willingness of the family to have more children will decrease.

4. BP Neural Network Prediction Model

4.1. Model Introduction

The BP neural network consists of an input layer, a hidden layer, and an output layer. The input layer is used to input one or more variables, while the output layer outputs the desired variables for the user, which can be one or more (only one in this paper, i.e. birth rate). The hidden layer serves as a bridge between the input and output layers, used to transmit and feedback information. The input layer is connected to the hidden layer, and the hidden layer is connected to the output layer with weights. The hidden layer can be multiple layers, and each hidden layer can have multiple nodes, depending on the actual situation. The learning and training of BP neural network includes two parts, namely, the forward propagation error of the signal and the backward feedback [4].

4.2. Data Preparation

In the previous analysis, we identified four factors that have a significant impact on the birth rate: per capita gross domestic product, gender ratio, population structure (15-64 year old population ratio), and urbanization level. Considering that China's policies, economy, and other factors have been relatively similar in recent years, we selected data from 2010 to 2021. However, considering that only 12 years of national data are selected, it is not sufficient to establish a BP neural network prediction model, Therefore, we collected data from multiple provinces such as Beijing, Tianjin, Guangxi, Fujian, Henan, Gansu, and Shaanxi for research to prevent data overfitting.

4.3. Model Building

For the establishment of the BP neural network prediction model, our input layer has m variables ($m = 4$) that form vectors $X = (X_1, X_2, \dots, X_m)^T$. The output layer has a variable Y , which is the birth rate we need to predict. Let the hidden layer have n layers, with the number of nodes in the first layer being p_1 , the node vectors forming the i -th layer being $H = (h_{i1}, h_{i2}, \dots, h_{ip_i})^T$, and the connection weight between the j -th neuron in the input layer and the k -th neuron in the first hidden layer being $w_{jh_{1pk}}$. For the neurons in the hidden layer, the input is

$$I_{h_{1pk}} = \sum_{j=1}^m w_{jh_{1pk}}, \quad (4)$$

after activating the data through the activation function, the output is

$$O_{h_{1pk}} = f(I_{h_{1pk}}), \quad (5)$$

the activation function is

$$f(x) = \frac{1}{1 + e^{-x}}. \quad (6)$$

Similarly, by performing the same operation between hidden layers and between hidden layers, as well as between hidden layers and output layers, we can obtain the predicted data, which is the process of signal forward propagation. The following figure2 shows the structure of the neural network.

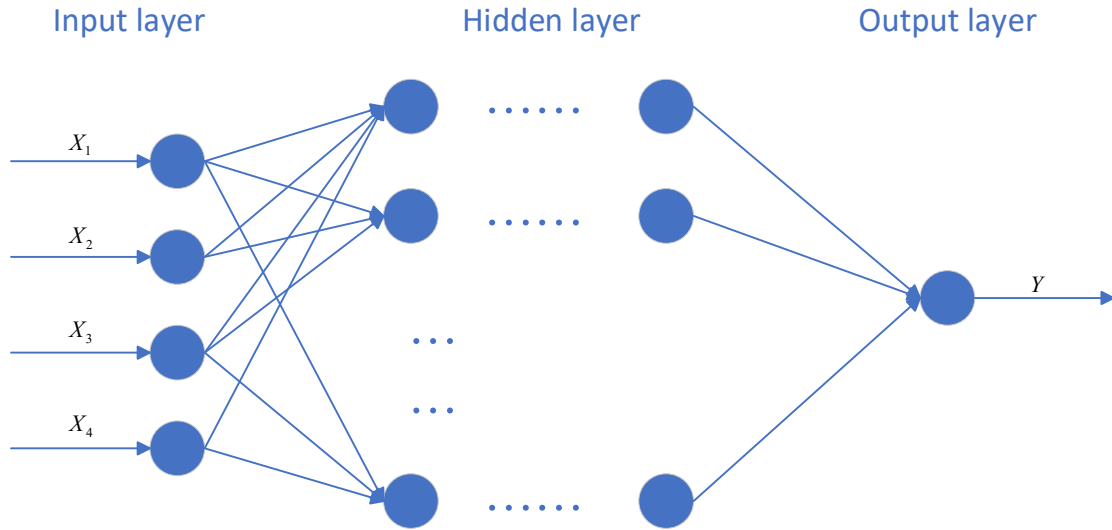


Fig 2. Neural network structure diagram

For the error feedback process, the BP neural network has a criterion to measure the performance of the model, which is the energy function. The smaller the value of the energy function, the better the performance of the model. The energy function is

$$E = \frac{1}{2} \sum_{i=1}^{p_n} (Y - O_{h_{np_i}})^2. \quad (7)$$

When constructing the model, neural networks continuously adjust the weights to make the energy function values smaller and smaller, and the simulation results closer to the actual situation.

4.4. Model Evaluation and Analysis

4.4.1. Model Training

After 18 iterations, the BP neural network gradually tends to the optimal iteration result, with error loss falling below 0.01. The number of iterations and MSE of the network are shown in Figure 1, and the number of decision trees and error loss are shown in Figures 3 and 4:

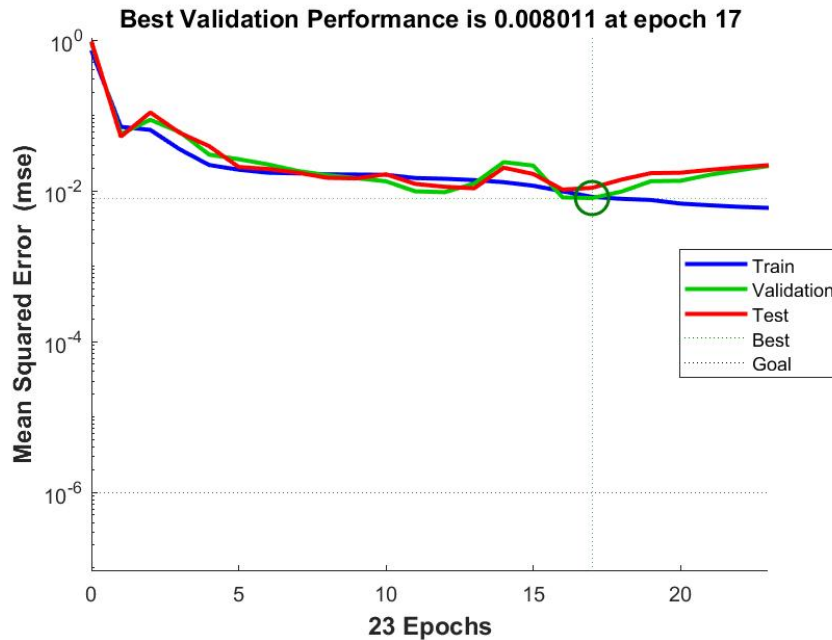


Fig 3. The number of iterations in relation to MSE

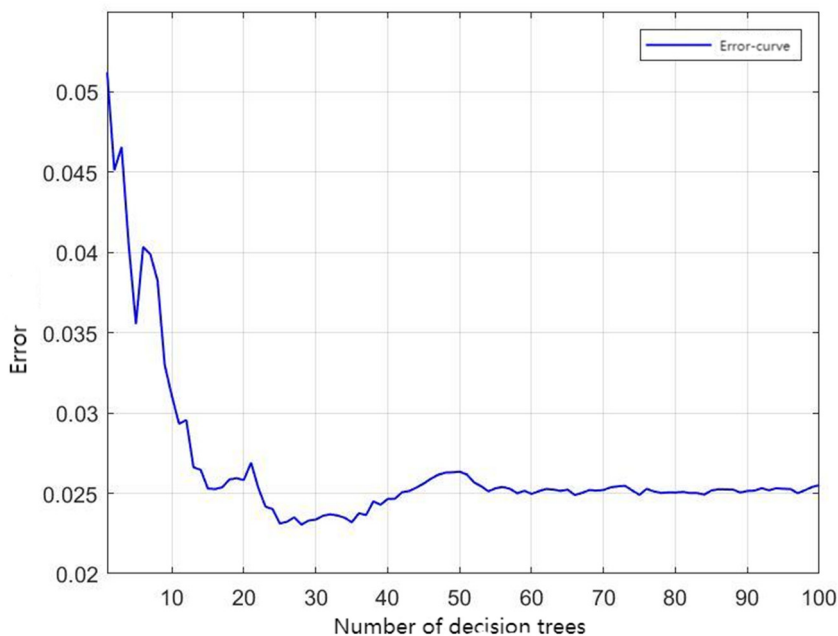


Fig 4. The number of decision trees as a function of error

4.4.2. Model Evaluation

The model evaluation uses the constructed test samples to test the trained model. The comparison between the actual and predicted values is shown in Figure 6-4. From the figure, it can be seen that the trend of the actual and predicted values basically coincides, enabling accurate prediction of the test data.

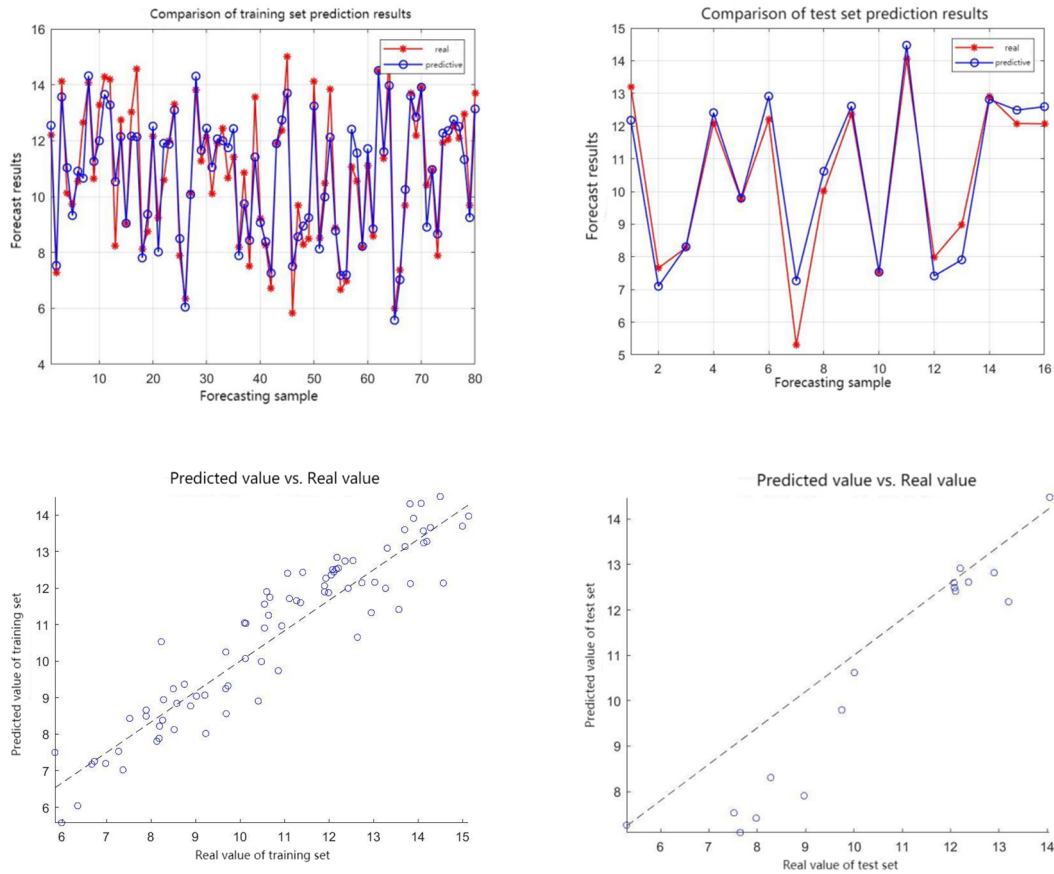


Fig 5. Prediction Results of Model Training Set and Test Set

According to the prediction results, it can be seen that the model has a high prediction accuracy, and the relevant evaluations are shown in Table 3.

Table 3. Evaluation Index Table

Evaluating indicator	Value
R ²	0.915
MAE	0.53557
MBE	0.1244
RMSE	0.72108

Compared to the $R^2=0.851$ obtained by the multiple linear regression model and the $R^2=0.915$ obtained by the BP neural network, it can be concluded that the established BP neural network model is true and effective, with a high degree of fitting and high accuracy of prediction results.

5. Conclusion and Recommendations

5.1. Conclusion

The factors that contribute significantly to the birth rate obtained from the BP neural network model in Chapter 3 are shown in the following figure:

Among them, features 1 to 4 are the per capita GDP index, population structure (proportion of population aged 15-64), urbanization level, and male to female ratio, respectively. Therefore, it can be known that the biggest factor affecting the birth rate is population structure, followed by per capita GDP and urbanization level, reflecting the significant impact of the economy on the birth rate.

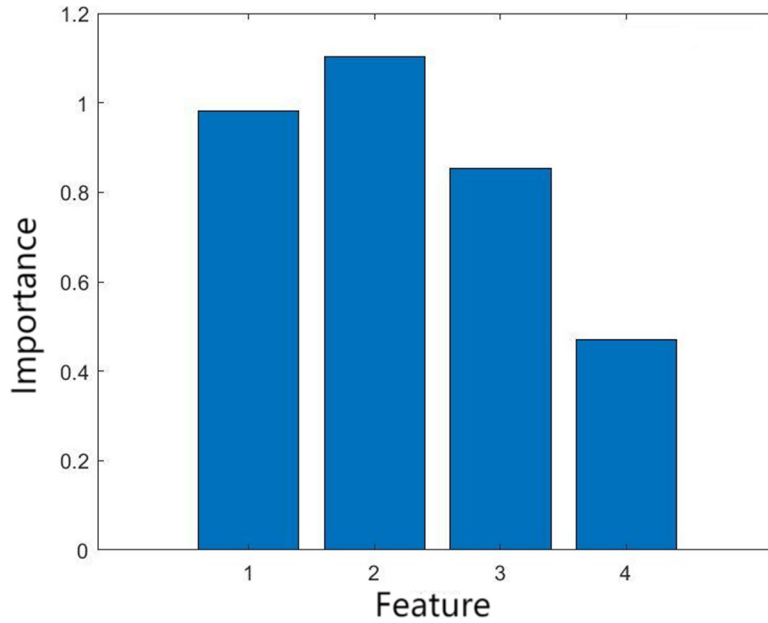


Fig 6. Feature importance

Use the BP neural network model to predict the birth rate in the next few years, as shown in the table below.

Table 4. Birth rate prediction table

year	Unit: %
2023	5.86
2024	5.21
2025	4.60

From the above table, it can be seen that if effective measures are not taken, the birth rate will further decrease, which is unfavorable for China's future development. Therefore, it is necessary to take effective measures as soon as possible to promote the increase of the birth rate.

5.2. Recommendation

5.2.1. Economic Policy and Support

The government can implement various incentive measures, such as providing childcare allowances, tax incentives, maternity leave, and paternity leave, to reduce the economic burden of raising children for families. Consider baby friendly facilities in urban planning, such as nurseries, parks, and children's medical centers, to provide a better environment for children's growth. Ensure that young families can obtain suitable housing and reduce the cost of raising children through policies and measures.

5.2.2. Population Structure and Social Support

Establish a social network for sharing parenting experiences, allowing parents to exchange experiences and solve problems, thereby reducing their anxiety. Provide flexible working hours and remote work options for parents to better balance work and family responsibilities.

5.2.3. Gender Equality and Cultural Concepts

Gender equality promotion: Through promotion and education, promote the concept of gender equality and encourage men to take on more responsibilities in childcare. Change the traditional

concept of family role allocation and encourage men and women to participate together in childcare and household affairs.

5.2.4. Health Security and Medical Support

Ensure high-quality and easily accessible medical services during pregnancy and postpartum to enhance parents' confidence. Provide comprehensive education on childbirth and parenting, helping expectant parents understand relevant knowledge and skills.

References

- [1] Chen Wei, Li Min. Long-term population consequences of high sex ratio at birth in China [J]. *Population and Development*, 2010,16(04):33-37+52.
- [2] Hu Pengcheng. Study on the impact of population migration on birth rate in the process of urbanization [D]. North China electric power university (Beijing), 2020. The DOI: 10.27140/, dc nki. Ghbbu. 2020.000902.
- [3] Becker, G. *Demographic and Economic Change in Develop ed Countries* (Princeton: Prin ceton University Press, 1960).
- [4] Xu Hong. Prediction of employment situation of college graduates in China based on BP neural network [J]. *Electronic Technology and Software Engineering*,2020(15):197-198.