

Study on Influencing Factors of Childbearing-Aged Women's Willingness to Have Multiple Children in China*

Dandan Cheng, Jiashan Tang

Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu, China

The fertility rate of China has shown an overall decline trend, so the fully understanding of the factors affecting China's fertility intention has become the focus of the research. Based on CFPS2020 data, the study subjects were women of age between 20 and 49, who were born after the 1970s and were influenced by the family planning policy. Then, SMOTE-Catboost algorithm was used to construct a model. The results show that: (1) For the women's willingness to have multiple children, the SMOTE-Catboost algorithm is more effective than the Catboost algorithm, and the classification accuracy is improved by 8 percentage points. (2) The factors influencing women's willingness to have multiple children were ranked by social status, intergenerational care, education level and age. Women with high social status are more willing to have multiple children; All-day intergenerational care has a positive effect on women's willingness to have multiple children. And women's willingness to have multiple children declines with the rise of education level but increases with age. (3) Combined with the CFPS2018 data, the influencing factors changed over time. In 2018, the important factors related to women's willingness to have multiple children were mainly related to economic, within which the income ranked locally is the most important one. In 2020, the most important factor changed to be social status while economic became less important.

Keywords: SMOTE algorithm, Catboost algorithm, fertility intention to have multiple child, influencing factors, China

Research Background

Since 1949, China's fertility level has changed through the stages of independent fertility, intervention from family planning policy and improvement of fertility policy (Liu & Wang, 2021). From 1949 to 1978, China's total fertility rate was at a high level, with two successive fertility peaks; and from 1978 to 2013, China implemented the family planning policy to achieve state controlling of the number of births, then during 1988 and 2013, China's overall fertility level continuously declined. However, with the advent of an aging population, China's fertility policy was further adjusted and switched to an improving-fertility policy, aiming to increase the fertility rate and ease the pressure of an aging population. A short-lived rise was found in the national fertility rate with the implementation of a separate two-child policy in 2014 and 2015. However, since 2016, with the implementation of the full two-child policy, China's total fertility rate has been declining after a short-term rebound. The data from the seventh national census in 2020 showed that China's total fertility rate

* This paper has been sponsored by the research project of Jiangsu Provincial Bureau of Statistics (number: KZ0072421051).

Dandan Cheng, master student of Applied Statistics, College of Science, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu, China.

Corresponding author: Jiashan Tang, Ph.D., Professor, Dean of the College of Science, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu, China.

was only 1.3 (tr.)¹. This is below the ‘safe zone’ of fertility. The low levels of fertility have become an urgent problem for China.

Fertility intention is influenced by a variety of factors, such as psychological perceptions, family expectations, the cost of childbirth, etc. Facing the current overall decline in fertility rates, it is therefore of importance to understand the factors influencing women of childbearing age to have multiple children in China.

A Brief Overview of Historical Research

China has entered a stage of inclusive fertility since 2016 and the factors affecting women’s fertility intention have become more intricate. But ultimately, they were divided into realistic factors and cultural perception factors (Tian, Sun, Zhao, & Zhang, 2022). Realistic factors include gender, age, economic level and the level of education, while cultural perceptions refer to changes in citizens’ perceptions of fertility. Qing and Jiang (2022) found that husbands had higher fertility intention compared to wives. Gao (2018) pointed out that household economic factors had a significant negative impact on residents’ fertility intention, the higher the income was, the lower the number of fertility intention would be. With the development of the economy, the change of women’s employment and retirement concepts in China also has a significant impact on fertility intention. Sheng (2019) found that the higher the number of children women had, the lower the quality of female employment was. In addition, Zhu and Hu (2019) concluded that urban residents who held the government’s concept of old-age care and rural residents who held the concept of old-age care in which the government, the elderly and children share equally had higher fertility intention. Similar findings can be found in foreign studies. Alderotti, Vignoli, Baccini, and Matysiak (2021) specified that employment instability had a non-negligible negative effect on fertility. V. Țarcă, E. Țarcă, and Luca (2022) found that the effect of consumption was less than per capita income on female fertility in the study of 171 countries. However, findings vary somewhat due to differences in national conditions. Fox, Klüsener, and Myrskylä (2019) believed that the negative relationship between fertility and economic development had diminished in many countries, and in some countries a positive relationship had emerged.

Research on women’s fertility intention is mostly regression analysis currently, and usually focusing on the effect of a particular factor on fertility intention, and fixed data research for a certain year. In this paper, we analyze the importance of different factors influencing fertility, the effect of each factor in different ranges on women’s intention to have multiple children, and the change in the factors influencing fertility over time by means of machine learning.

Data and Research Methods

Data Sources and Characteristic Variables

The data for this study were obtained from the public data of the China Family Panel Studies (CFPS) in 2018 and 2020. And in this paper, women of childbearing age between 20 and 49 were selected as the survey subjects. A total of 6,614 samples of women of childbearing age were extracted from 28,590 individual samples published in 2020, and 960 valid samples were retained after data pre-processing. Of the publicly available 37,354 individual samples in 2018, 1,410 valid samples were retained.

¹ With data from the seventh national census in 2020 announcing that China’s total fertility rate was only 1.3. (tr. 国家统计局. 第七次全国人口普查主要数据结果新闻发布会答记者问, 2021-05-11, retrieved from http://www.stats.gov.cn/zt_18555/zdtjgz/zgrkpc/dqcrkpc/ggl/202302/t20230215_1904005.html/ (August 15, 2022).

The study's response variable is the desired number of children named fertility intention. The rest are characteristic variables (see Table 1).

Table 1

Study Variables

Variables	Definition
Response variable	
Fertility intention	0: Up to one child, 1: Two or more children
Characteristic variables	
Intergenerational care	1: None, 2: 1-2 days a week, 3: 3-4 days a week, 4: Almost every day
Account	0: Countryside, 1: Town
Age	20-49 years old
Education	1: Unschooled/Illiterate/Semiliterate, 2: Primary school, 3: Junior high school, 4: High school/Polytechnic school/Technical school/Vocational high school, 5: Bachelor's degree or above
Marital satisfaction	1-5, 1: Very dissatisfied, 5: Very satisfied
Spouse's economic contribution	1-5, 1: Very dissatisfied, 5: Very satisfied
Spouse's contribution to housework	1-5, 1: Very dissatisfied, 5: Very satisfied
Job income satisfaction	1-5, 1: Very dissatisfied, 5: Very satisfied
Job satisfaction	1-5, 1: Very dissatisfied, 5: Very satisfied
Endowment insurance	0: No, 1: Yes
Medical insurance	0: No, 1: Yes
Unemployment insurance	0: No, 1: Yes
Employment injury insurance	0: No, 1: Yes
Birth insurance	0: No, 1: Yes
Happiness	0-10, 0: Minimum, 10: Maximum
Self-life satisfaction	1-5, 1: Very dissatisfied, 5: Very satisfied
Income in local locations	1-5, 1: Very low, 5: Very high
Social status	1-5, 1: Very low, 5: Very high
Impact of maintenance	1-5, 1: Almost no effect, 5: Very strong effect

Basic Characteristics of the Study Population

Analyzing the basic characteristics of the study population, it shows a left-skewed distribution in both age structure and education level (see Table 2).

Table 2

Basic Characteristics of the Study Population

Features	Group	Frequency	Proportion (%)
Age	20-25	33	3.4
	26-30	204	21.2
	31-35	363	37.8
	36-40	236	24.6
	41-45	124	12.9
Education	No schooling/Illiterate/Semiliterate	22	2.3
	Primary School	44	4.6
	Junior high School	324	33.8
	High school/Polytechnic school/Technical school/Vocational high school	205	21.4
	Bachelor's degree or above	365	38
Account	Countryside	620	64.6
	Town	340	35.4

Research Methodology

SMOTE algorithm. The Synthetic Minority Oversampling Technique (SMOTE) algorithm is an algorithm that solves the sample imbalance problem by algorithmically synthesizing a new minority of samples, avoiding the problems of overfitting and undergeneralization of the learned model caused by traditional sampling. The principle of SMOTE sampling is as follows:

$$X_c = X + \text{rand}(0,1) \cdot (X_{\text{neighbor}} - X),$$

Where X is a sample from one of the few categories in the original data, X_{neighbor} is a nearest neighbor sample of the selected X , $\text{rand}(0,1)$ represents a random number between 0 and 1, and X_c represents a new sample of the final composite.

Catboost algorithm. Catboost is an integrated algorithm based on symmetric decision trees, which is claimed to have higher accuracy than xgboost and lightgbm. It is ideally suited to machine learning tasks with classified, heterogeneous data (Hancock & Khoshgoftaar, 2020). Its most important feature, besides the automatic processing of category features and the use of combined category features, is the use of a correction method for unbiased estimation of gradient steps in the phase of tree structure selection to mitigate prediction bias. The steps are as follows:

- Step 1: The sample set is $\{(x_k, y_k)\}_{k=1}^n$ sorted by the random sequence σ and the tree is I ;
- Step 2: For the sample x_k , initialize the model M_k ;
- Step 3: For each tree, iterate through each sample, calculating the gradient named g_i of $Loss$ for the first $k-1$ samples;
- Step 4: Use g_i and x_j of the first $k-1$ samples to construct the model;
- Step 5: For each sample x_k , correct the initialized M_k with M to obtain a segmentation model M_k ;
- Step 6: Repeat the above to obtain a segmentation model of M for each sample x . It can be seen that each M_k shares the same tree structure.

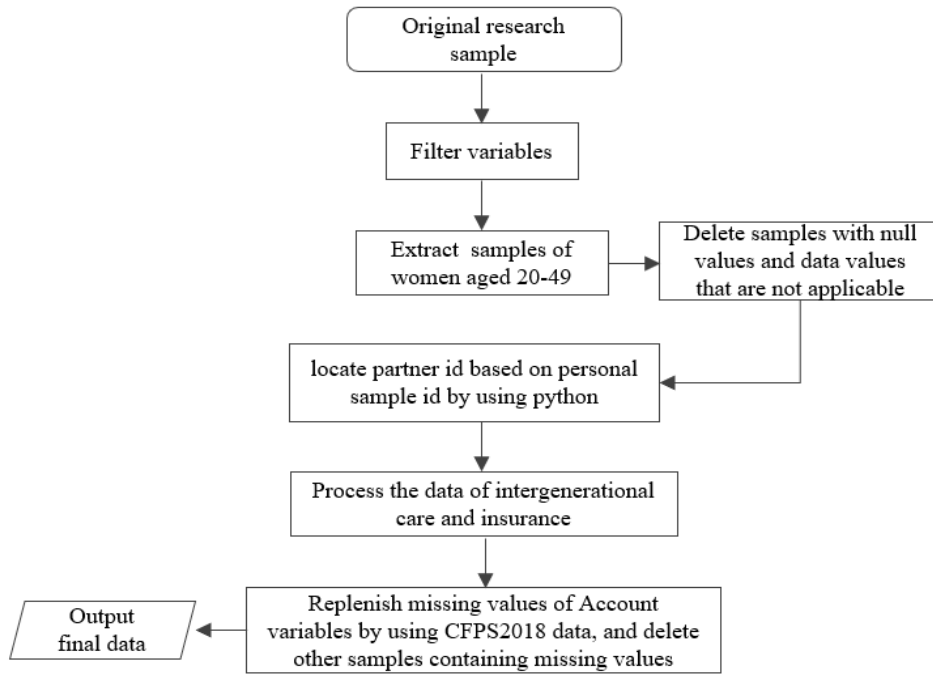


Figure 1. Data processing flow.

Data Pre-processing

Before data analysis, we need to perform data pre-processing. Firstly, screened the variables needed for the study; secondly, filtered the subjects for study and removed invalid samples; finally, integrated the intergenerational care data and insurance data and supplemented the missing data with the data from 2018 based on the characteristic that CFPS data are tracking survey data. There shows detailed data processing process (see Figure 1).

Comparison of Algorithms and Analysis of Influencing Factors

Sample Equalization and Parameter Adjustment

The ratio of the number of women in the sample who have more than one child to the number of those who have at most one child is 3:1 (see Figure 2), so this sample is a mild category imbalance sample.

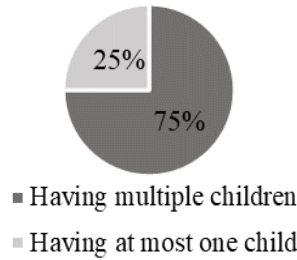


Figure 2. Distribution of fertility intention.

For the imbalanced sample, the first step is to divide the sample of childbearing-aged women into a training set and a test set in the ratio of 4:1, and the second step is that the SMOTE algorithm was applied to oversample the sample data to achieve the effect of expanding the sample. And after oversampling, the proportion of female samples in both categories is the same, and the SMOTE algorithm effectively achieves sample balance.

Based on the default parameters of the system, Bayesian optimization was used to tune the parameters and output the best parameter model. Value of optimum learning rate is 0.17, tree depth is 4, and the sampling rate of row and column are 0.84, 0.61 respectively (see Table 3).

Table 3

Table of Optimum Parameters

Parameters	Value
Learning rate	0.17
Tree depth	4
Row sampling rate	0.84
Column sampling rate	0.61

Then, the auc value of the SMOTE-Catboost algorithm model after tuning is 0.62, which is more predictive compared to the Catboost model. The precision score of the SMOTE-Catboost algorithm is 83%, which is an improvement of 8 percentage points compared to the Catboost algorithm (see Table 4).

Table 4

Model Comparison

Algorithm	auc	precision_score
Catboost	0.52	75%
SMOTE-Catboost	0.62	83%

Feature Importance Ranking and Analysis

Based on the SMOTE-Catboost algorithm model, the SHAP machine learning interpreter was used to explain the factors influencing women's intention to have multiple children.



Figure 3. Feature importance ranking.

The results find that top 10 major features are social status, intergenerational care, educational attainment, age, birth insurance, satisfaction with spouse's financial contribution, happiness, satisfaction with self-life, income in local location, and marital satisfaction. In addition, the results show that birth insurance and pension insurance are the most important insurance factors among the five types of insurance (see Figure 3).

Afterwards, a detailed analysis is conducted on the impact of these 10 main factors.

In terms of social status, women with higher social status are more likely to have multiple children, but the social status of women is concentrated in the middle of the group. In terms of intergenerational care, the frequency of all-day intergenerational care has a positive effect on women's willingness to have multiple children, while the rest of the care frequency has a negative effect on women's willingness to have multiple children (see Figure 4).

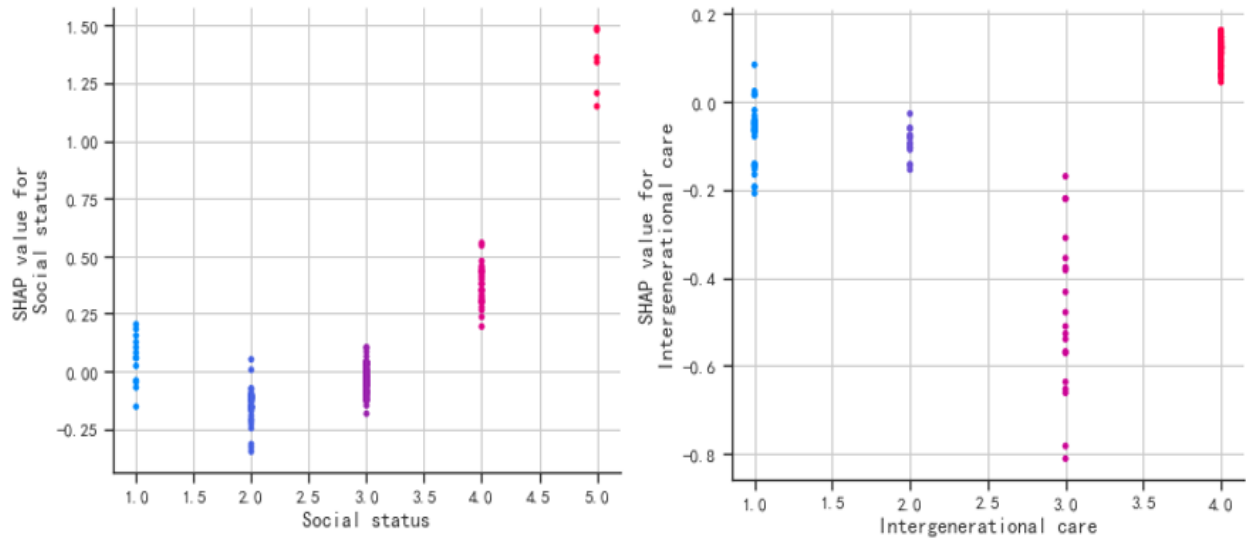


Figure 4. Analysis of influencing factors (social status, intergenerational care).

In terms of the level of education, women's willingness to have multiple children decreases with education level rises. As Wang, Xu, and Tang (2019) argued that education affected perceptions, and individuals with higher level of education identified more with the idea of 'social endowment'. In terms of age, the willingness to have multiple children increases with age. However, below the age of 33, age has an inverse effect on willingness to have multiple children, and after the age of 35, age has a positive effect on willingness to have multiple children (see Figure 5).

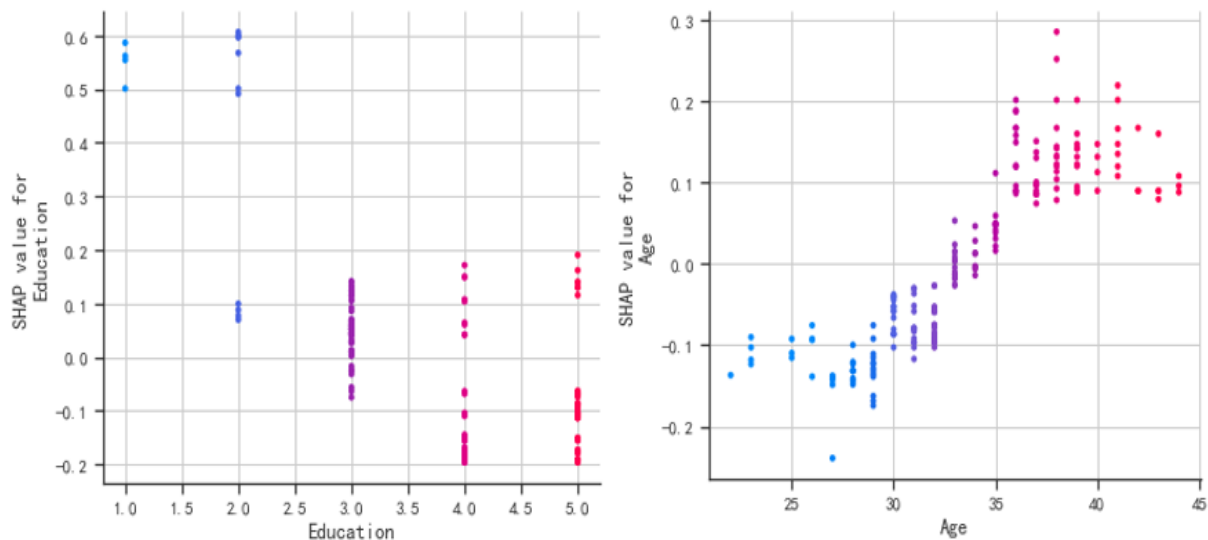


Figure 5. Analysis of influencing factors (education level, age).

In terms of birth insurance, those who pay for birth insurance are mostly reluctant to have multiple children. As Chen (2011) found that birth insurance policies can significantly reinforce wage discrimination against women, and secondly, if insurance can further improve the quality of infant and child births, according to the number of children and quality trade-off theory, families will tend to pursue the quality of children and reduce the need for quantity, then birth medical compensation will also have a negative impact on childbirth (Shen, Yang, & Zhang, 2019). In terms of satisfaction with the spouse's financial contribution, very high satisfaction increases women's willingness to have multiple children, low satisfaction increases women's willingness to have multiple children, and normal satisfaction decreases women's willingness to have multiple children. Satisfaction with spouse's economic contribution depends on spouse's income level, and those who are very satisfied tend to be high income earners, which is consistent with the direction of change in income level and fertility studied by Peng and Bi (2019) (see Figure 6).

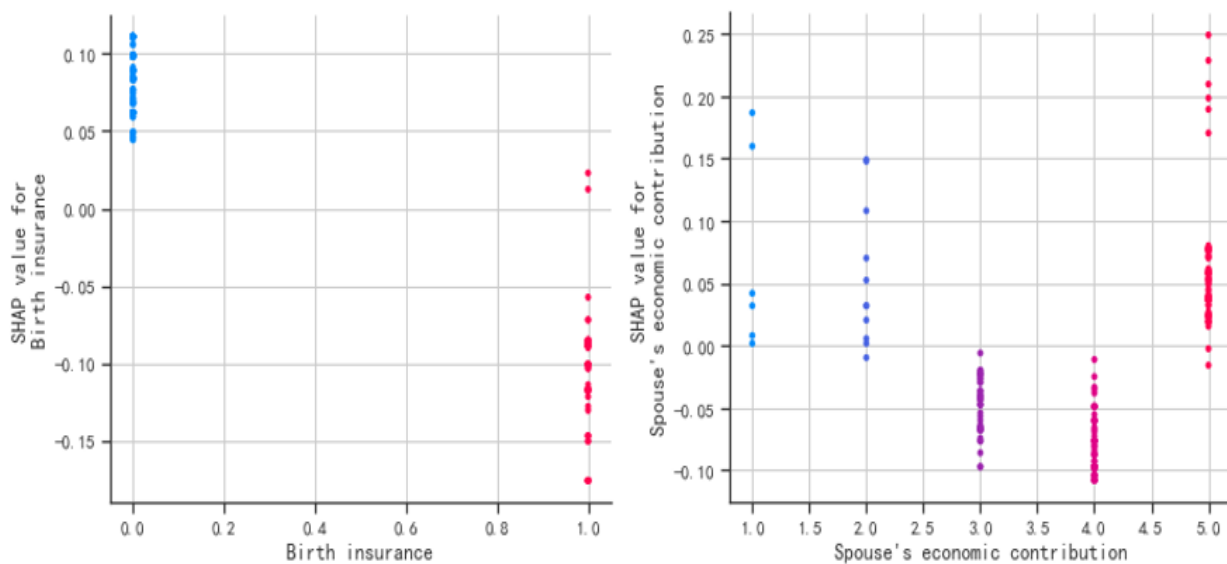


Figure 6. Analysis of influencing factors (birth insurance, spouse's economic contribution).

From the perspective of happiness, the impact of happiness on fertility intention fluctuates greatly. People with high happiness will increase their willingness to have multiple children, but women with a happiness level of 7 are more unwilling to have multiple children. When conducting research on the female population with a happiness level of 7, it was found that the social status and income of the female population with a happiness level of 7 are mostly at a moderate level in their local location, but they are in a satisfactory state regarding their marriage, self-life status, and work. Therefore, it also satisfies many working class women's unwillingness to change their balanced life state due to childbirth. From the perspective of self-life satisfaction, women who are satisfied with their lives most are more likely to have multiple children. This result is similar to the conclusion that Qiu, Hou, and Ren (2019) believed that with the higher satisfaction with life, the married population after the 1980s was more likely to have a second child (see Figure 7).

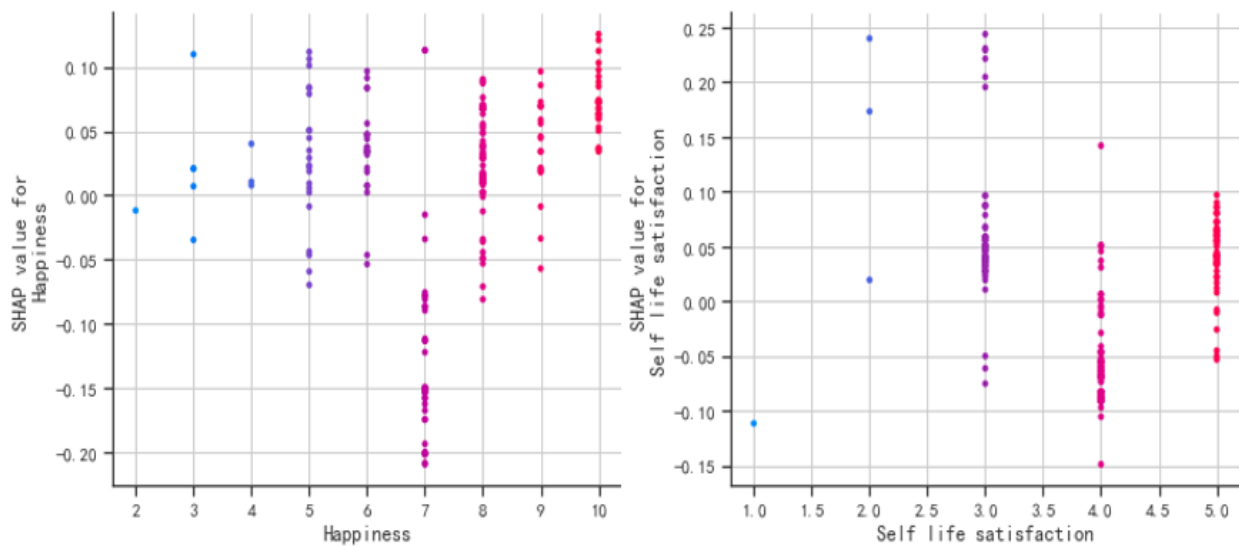


Figure 7. Analysis of influencing factors (happiness, self-life satisfaction).

In terms of the level of local income, women at the lowest and higher levels are more willing to have multiple children, while those in the middle position are more inclined to be unwilling to have multiple children. This conclusion is similar to the previous study on the impact of spouse's economic contribution satisfaction on women's fertility intention. In terms of marital satisfaction, the women's sample size of low satisfaction is relatively small, and the representativeness is relatively low. However, it can be seen that the highest marital satisfaction will increase the willingness of women to have multiple children (see Figure 8).

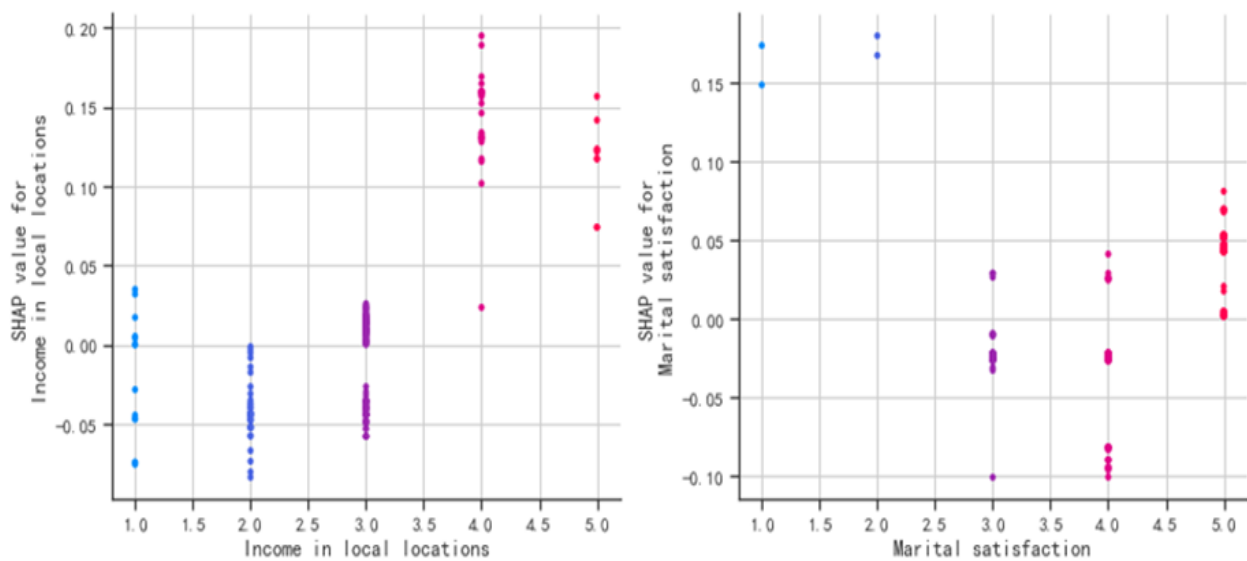


Figure 8. Analysis of influencing factors (income in local locations, marital satisfaction).

Analysis of Changes in Factors Influencing Women's Intention to Have Multiple Children

As society develops, changes in the factors influencing childbearing aged women's willingness to have multiple children are subject to further research and analysis. The result illustrates the changes in the factors influencing women of childbearing age to have multiple children in China from 2018 to 2020 (see Figure 9).

The study finds that the intention to have multiple children decreases with increasing education in 2018 and increases with age which are the same in 2020. In addition, the graph shows that the influencing factors in 2018 are mostly economic, and the most important influencing factor being the rank of income in the local area. However, in 2020, the importance of economic influencing factors was decreasing, factors reflecting women's psychological indicators such as happiness, birth satisfaction and self-life satisfaction were added to the important influencing factors of women's willingness to have multiple children, and the importance of intergenerational care increased significantly from 2018 to 2020.

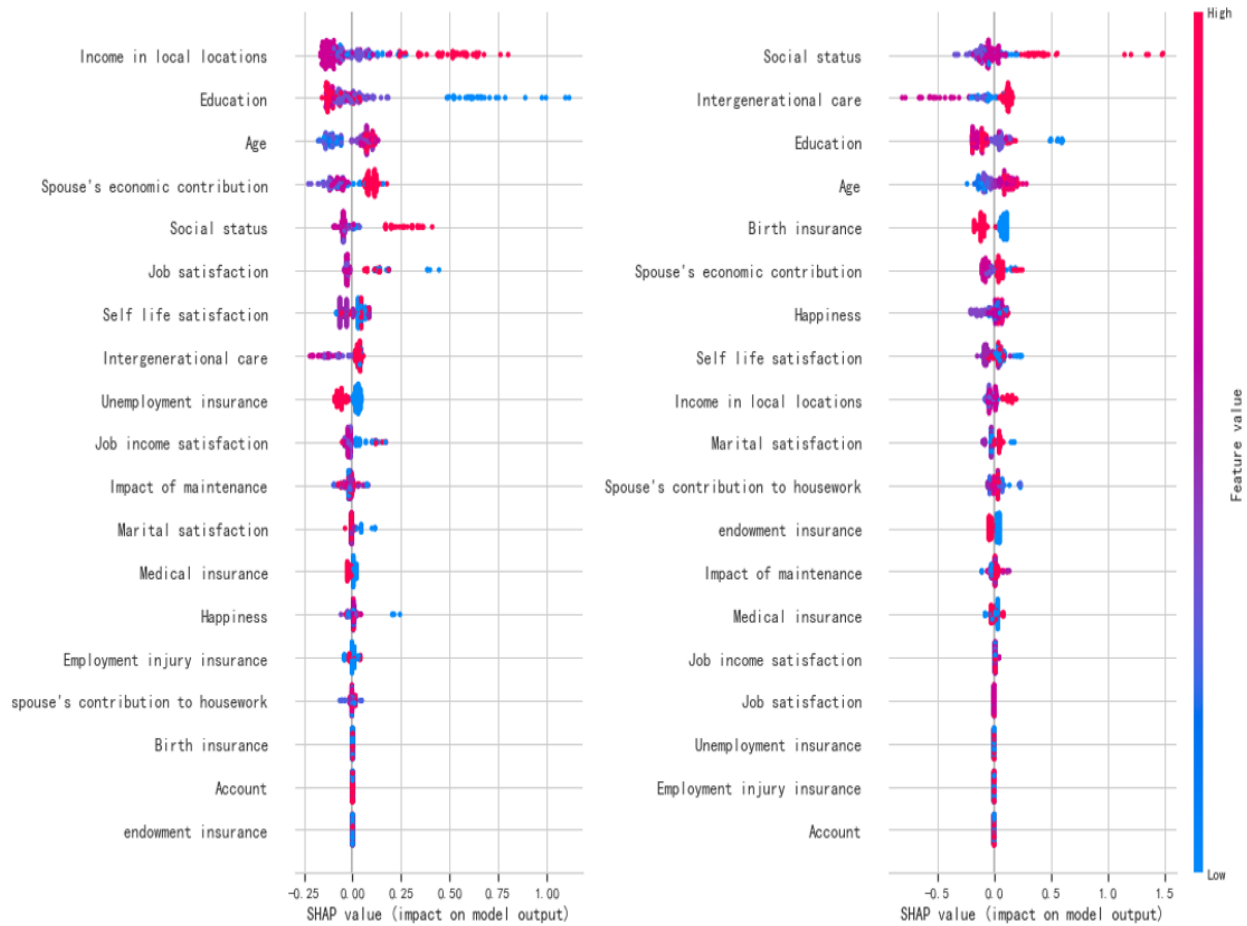


Figure 9. Rank of factors influencing intention to have multiple children in 2018 and 2020.

Conclusions

In order to further explore the factors affecting women's willingness to have multiple children, the SMOTE-Catboost algorithm was used to analyze the factors influencing women of childbearing age to have multiple children. The conclusions are obtained as follows:

Firstly, the SMOTE-Catboost algorithm is more convincing than the Catboost algorithm for modeling. The accuracy of the SMOTE-Catboost model is 83%, an improvement of 8 percentage points compared to the Catboost model.

Secondly, the ranking of importance affecting the willingness to have multiple children in 2020 are social status, intergenerational care, education, age, birth insurance, satisfaction with spouse's financial contribution,

happiness, satisfaction with self-life, income in local location and marital satisfaction. The study finds that groups with higher social status are more likely to have multiple children while intermediate groups of women are less likely to have multiple children. All-day intergenerational care has a positive effect on the willingness to have multiple children. The willingness to have multiple children decreases with education level and increases with age, and women with high and low satisfaction with their spouse's financial contribution are more likely to have multiple children, women with higher local position are the same.

Thirdly, the factors influencing the willingness of childbearing aged women to have multiple children are changing over time. The main factors influencing women's fertility in 2018 were related to economic within which the location of income in the local area was the most important factor. However, in 2020, social status was the most important one. The importance of economic factors declined and the importance of intergenerational care raised. At the same time, women of childbearing age were focusing more on psychological satisfaction, with happiness and marital satisfaction becoming important factors.

These findings also have beneficial policy implications. Firstly, based on the analysis of social status on the willingness to have multiple children, it is necessary to improve the social status of the female group. China needs to protect women's rights and interests in a more effective way, including supporting women's development and helping to break down gender barriers. Secondly, the issue of childcare is an assignable consideration, to address which, the government may enforce maternity policy of prolonging maternity leave, regulating the mature childcare market as soon as possible, and reducing the pressure on families especially on women. Finally, both families and the society should pay close attention to the psychological state of women. The government can learn from overseas experience like setting up 'respite services' for the female population or providing community health psychological services. With the above actions, the female population can maintain a good psychological state and increase the sense of well-being.

References

- Alderotti, G., Vignoli, D., Baccini, M., & Matysiak, A. (2021). Employment instability and fertility in Europe: A meta-analysis. *Demography*, 58(3), 871-900.
- Chen, L. (2011). Birth insurance, female employment, and child care—An analysis based on China's micro data. *Economist*, (7), 54-60. [陈琳. (2011). 生育保险、女性就业与儿童照料—基于中国微观数据的分析 (pp. 54-60). 经济学家, (7)].
- Fox, J., Klüsener, S., & Myrskylä, M. (2019). Is a positive relationship between fertility and economic development emerging at the sub-national regional level? Theoretical considerations and evidence from Europe. *European Journal of Population*, 35, 487-518.
- Gao, L. (2018). The influence of family economic and housing conditions on residents' fertility intentions (Master dissertation, Shandong University, Jinan). [高璐. (2018). 家庭经济与住房条件对居民生育意愿的影响. 山东大学(硕士学位论文), 济南].
- Hancock, J. T., & Khoshgoftaar, T. M. (2020). CatBoost for big data: An interdisciplinary review. *Journal of Big Data*, 7(1), 1-45.
- Liu, Z., & Wang, X.-Y. (2021). Changes in fertility: A study on the factors influencing fertility in China from 1949 to 2019. *Northwest Population*, 42(01), 107-116. [刘卓, 王学义. (2021). 生育变迁: 1949~2019年中国生育影响因素研究 (pp. 107-116). 西北人口, 42, (1)].
- Peng, J., & Bi, Z.-P. (2019). Income levels, family policies and changes in fertility: Western trends and China's challenges. *South China Population*, 34(1), 44-55. [彭皎, 毕忠鹏. (2019). 收入水平、家庭政策与生育率变化—西方的趋势和中国的挑战 (pp. 44-55). 南方人口, 34, (1)].
- Qing, S.-S., & Jiang, Y.-S. (2022). Analysis of the differences in fertility intention and influencing factors between couples. *Journal of East China Normal University (Philosophy and Social Science)*, 54(3), 140-152+187-188. [卿石松, 姜雨杉. (2022). 夫妻生育意愿及影响因素的差异分析 (pp. 140-152+187-188). 华东师范大学学报(哲学社会科学版), 54, (3)].

- Qiu, H.-Y., Hou, L.-Y., & Ren, Y.-J. (2019). The impact of childbirth experience and life satisfaction on the second child planning of the married and childbearing occupational group in the '80s generation'. *Chinese Journal of Health Education*, 35(2), 151-155. [邱红燕, 侯丽艳, 任杨洁. (2019). 生育经历、生活满意度对“80后”已婚育职业群体二胎生育计划的影响 (pp. 151-155). 中国健康教育, 35, (2)].
- Shen, Z., Yang, H.-L., & Zhang, W.-C. (2019). Can maternity insurance promote family fertility. *Finance & Economics*, (3), 52-65. [沈政, 杨华磊, 张文超. (2019). 生育保险能促进家庭生育吗? (pp. 52-65). 财经科学, (3)].
- Sheng, Y.-N. (2019). The impact of maternity policy adjustment on the quality of women's employment. *Population & Economics*, 234(3), 62-76. [盛亦男. (2019). 生育政策调整对女性就业质量的影响 (pp. 62-76). 人口与经济, 234, (3)].
- Țarcă, V., Țarcă, E., & Luca, F.-A. (2022). The impact of the main negative socio-economic factors on female fertility. *Healthcare*, 10(4), 734.
- Tian, H.-J., Sun, H.-Y., Zhao, X., & Zhang, X.-D. (2022). Low fertility intentions under the three-child policy: Realistic constraints and cultural concepts constraints. *Youth Exploration*, (4), 77-86. [田宏杰, 孙宏艳, 赵霞, 张旭东. (2022). 三孩政策下的低生育意愿: 现实束缚与文化观念制约 (pp. 77-86). 青年探索, (4)].
- Wang, Y., Xu, K., & Tang, Y. (2019). Study on the mechanism of the impact of regional education quality on residents' willingness to have a second child—Evidence from CGSS 2015 experience. *Northwest Population*, 40(4), 1-11. [王英, 徐鲲, 唐云. (2019). 地区教育质量对居民二孩生育意愿的影响机制研究——来自CGSS2015经验证据 (pp. 1-11). 西北人口, 40, (4)].
- Zhu, Q.-M., & Hu, S.-Z. (2019). Study on the fertility willingness of urban and rural population of childbearing age and its influencing factors—Based on the perspective of elderly care concept. *Journal of West Anhui University*, 35(1), 49-54+62. [朱青梅, 胡书芝. (2019). 城乡育龄人口生育意愿及其影响因素研究——基于养老观念的视角 (pp. 49-54+62). 皖西学院学报, 35, (1)].