ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202209_5(9).0054

Influence of Parental Background on Gender Differences in Offspring Education: Evidence from Machine Learning

Mengfan Li[1, a]

[1]School of Economics and Management, Communication University of China, Beijing100024, China

[a]limengfan0824@163.com

Abstract

This study examines the relationship between gender differences in educational attainment of Chinese children and their parents' background, as well as their interaction. In this study, XGBoost algorithm and SHAP value with better performance and explanatory power are used to quantify the influence between variables. The model results show that women's educational attainment is more affected by their parents' background. Parents' cultural capital is more important than economic status. In addition, mother's background is of great significance to children's education, especially women's education. In order to improve the overall level of social education and reduce gender differences, more attention should be paid to mothers' economic capital, cultural capital and their relative discourse power in the family.

Keywords

Education attainment; XGBoost; SHAP; Parental background; Gender role concept.

1. Introduction

As of 2018, China's survey data show that the average years of education for women aged 18-64 years are 9.41 years, while men are 9.66 years. A large number of studies have shown that inequality between men and women in educational attainment is obvious (Bauer et al., 1992; Asadullah & Chaudhury, 2009). Although countries have issued relevant policies in recent years to narrow the gender gap in education, there is no denying that this gap is difficult to eliminate in a short time, because of the precipitation of long-term history and culture (Yang, 2014). However, gender inequality in education significantly affects public well-being (Ma and Piao, 2019) and inhibits economic growth (Klasen and Lamanna, 2019), which seriously hinders social development and cannot be ignored.

Family is the principal part of children's education investment. Many existing studies tend to explain the impact of family social status on gender differences in children's education from a macro perspective (Butcher & Case, 1994; Jacobs, 1996; Post, 2001), which to some extent ignores the role of individual micro-level cognition, choice and game within the family. Bourdieu (1970) stressed that the redistribution of social and cultural resources by individual thinking and family's 'strategy' together. In addition to the objective social resources, parents' traditional gender roles such as the division of labor between men and women, women's marriage will significantly affect their investment in offspring education (Johnson & Howard, 2008). This difference is more pronounced when there is more than one child (Blake, 1989). At the same time, due to the differences in their own attributes, parents have different inclinations to invest in their children, so we have to pay attention to the impact of the game between parents on the children's education.

Most existing studies are based on linear methods to explore the influencing factors of gender education attainment differences, but there is a certain correlation between education and

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202209_5(9).0054

income (Kslff, 2001). The establishment of linear regression model will be interfered by multicollinearity, and it is difficult to better analyze the influence of cross-related variables even through the interaction term (Fleming and Goodbody, 2019). Therefore, the limitations of research methods make it difficult to conduct in-depth research in this area and the explanatory power is weak. Machine learning can well solve the multicollinearity problem of traditional linear regression, and more intuitively discharge the importance of the influence of each variable and the joint influence of multiple factors (Bajari et al., 2015). As one of the most effective algorithms in machine learning, XGBoost algorithm has been widely used in different disciplines and application scenarios in recent years (Gumus & Kiran, 2017; Ogunleye & Wang, 2019; Jiang et al., 2019). In terms of prediction indicators, compared with the β value and p value of classical linear regression and other indicators, SHAP value has better explanatory power (Lundberg and Lee, 2017). As a feature attribution method, it assigns a specific prediction value for each feature, which is more helpful to explain the prediction results. At the same time, the powerful data visualization function provided by SHAP package also provides convenience for the display of model results.

The purpose of this paper is to use XGBoost algorithm and more explanatory SHAP value to quantify the long-standing personal capital of parents and gender role concepts, and the single and cross effects on gender education differences. The new application of machine learning method combined with sociological, psychological and economic research theories may provide better optimization enlightenment for policy makers and educators.

1.1. The Different Role of Parental Background in Children 's Education Attainment

Under the background of social change and the improvement of the overall economic level, the influence of social environment on educational attainment is becoming smaller and smaller, and the relationship between family background and educational attainment is becoming more and more closely (Zhou et al., 1998). According to the theory of family resource transformation, parents can use all kinds of advantageous resources, including economic capital and cultural capital, to influence their children 's educational opportunities and quality through transformation and reproduction (Hart, 2019). The empirical results of many studies have revealed that family background has different degrees of impact on children’s educational attainment (Bukodi & Goldthorpe, 2013; Kainuwa & Binti, 2013). However, these studies unilaterally regard the family as a whole analysis and ignore the different roles of individuals, which aspect is flawed.

Blau and Duncan (1967) conducted a pioneering study of American class structure and occupational status attainment, using path analysis to establish a ‘status attainment model’. It reveals the direct influence of father’s socioeconomic status on children’s educational attainment, but does not take into account the role of mother. When Blau and Duncan studied America, the employment rate of American women was not high. However, with the continuous development of society in recent years, the socioeconomic status of women has been significantly improved, and China is one of the countries with the highest female employment rate in the world (Zeng, 2014). In this case, it is much worse than the actual situation if only the variables of father are inserted into the model and the influence of mother is ignored (Erola et al., 2016). The main undertaker of children’s educational burden occupies the main redistribution power of children’s educational resources. The improvement of mothers’ economic and educational level makes their position in the allocation of family resources cannot be ignored (Goldscheider et al., 2015).

In addition to ignoring the importance of women, many studies have ignored the differences in preferences between men and women. Empirical evidence from some African countries shows that the assumption that parents show the same preference is not supported (McElroy, 1990;

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202209_5(9).0054

Schultz, 1999). In fact, because parents with different resources, cultural backgrounds and cognitions tend to show inconsistent subjective preferences and make different decisions, it is not appropriate to simply analyze them as a whole (Schultz, 1999). In order to better reveal the influence of family on children's education, we should separate the influence of parents for comparative analysis. Based on this, McElroy and Horney (1981) and Manser and Brown (1980) proposed a Nash bargaining model on family behavior. It regards family decision as a bargaining process within the family, separating the utility function of husband and wife. According to this model, Handa (1996) identified the important role of mother education through data studies in Jamaica.

1.2. The Background of Parents and Gender Differences in Children's Educational Attainment

The significant gender difference in children's educational attainment is an important manifestation of unequal educational opportunities between intergenerational mobility (Roksa & Potter, 2011). The neoclassical family model points out that parents have an optimal educational investment level for each child in the process of family-based optimization (Lv, 2021). Gender discrimination in the labour market inevitably leads parents to invest more resources in the education of boys in order to achieve higher returns on investment (Jayachandran, 2015). This potential theoretical logic will be affected by parents' own background. A study found that the poorer regions and families, the greater gender differences in children's access to education (Post, 2001). According to the 'family investment model', families with higher socioeconomic status with more capital can provide adequate resources for children's development, at a time when parents may take less account of gender investment returns. Parents with low socioeconomic status are under real pressure to make selective concessions in resource allocation (Bradley & Corwyn, 2002).

In addition to the limitations of family resources, parents' perception of gender roles also has a great impact. The traditional concept of gender role, in the context of China, refers to the division of labor and gender status of 'male-dominated outside and female-dominated inside' (Cheung & Halpern, 2010). In the traditional gender role concept, women should undertake more housework and obtain higher education to find better jobs, which is more important for men (Kollmayer, Schober & Spiel, 2018). People with this tendency are more likely to think that the return on investment in women's education is lower than that of men, so educational resources are more inclined to men (Robeyns, 2006). Hatlebakk (2017) believes that girls should undertake more domestic work so that boys can focus more on learning, so it has a positive impact on the number of years of education for men. Li and Tsang (2003) also found that parents' educational expectations for boys were higher than for girls, while women's enrolment rates were significantly lower than men's enrolment.

Parents have different gender, different ways of thinking, and have different resources, which makes them tend to present inconsistent cognition, so as to make different decisions on children's education investment (Baker & Stevenson, 1986). Zhang et al. (2007) found that in the allocation of resources within the family, mothers have a greater impact on girls' educational attainment. The higher the mother's educational level, economic ability and autonomy, the more fathers can undertake household chores and care for children, the smaller the boy preference (Li & Lavely, 2003). However, when more than one child in the family, limited resources will be competed, which will enlarge the original gender preference difference. Resource dilution theory explains this phenomenon (Wang & Feng, 2021). An empirical study found that the size of Indonesian siblings had a significant negative impact on children 's educational attainment, especially for children of mothers with lower educational attainment and children belonging to early birth cohorts (Feng, 2021). According to the contradictory results of previous studies on the competition between siblings in the allocation

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202209_5(9).0054

of family resources in the United States, Bauer & Gang (2001) further found that culture also affects the impact of the number of children on the educational attainment of children.

Through the above analysis, it will be found that parents who have different resources, the perception of gender roles and the number of children will significantly affect the gender differences in children's education. However, the specific impact and difference relationship have not been well proved.

1.3. Feasibility and Innovation of XGBoost algorithm and SHAP value

There are many studies on gender differences in educational attainment, but the most commonly used traditional linear regression models are analyzed (Gandhi Kingdon, 2002; Gibb, 2008;). It is based on a large number of subjective assumptions, such as 'model is linear', 'residual normal distribution' and so on, but these assumptions cannot be verified (Gelman & Hennig, 2017). It is also subjective to judge the influence by the explicitness of p value. Other traditional statistical methods are also unsatisfactory. Compared with the traditional statistical subjective model driven, the data driven of machine learning method is more persuasive (Churpek et al., 2016; Athey & Imbens, 2019).

For the purpose of this study, the XGBoost algorithm with better effect is finally selected (Song et al., 2020). Although other machine learning methods such as support vector machine (Wu, Chen & Zheng, 2011), K-nearest neighbor method (Kim & Hoi-Kyun, 2001), Bayesian network method (Ozbay & Noyan, 2006) do not need or less to assume and deal with high-dimensional nonlinear data more flexibly, they cannot explore the potential relationship between independent variables and dependent variables, which is unfavorable for the exploration of variables in this study. XGBoost is an optimized distributed gradient lifting library that inherits the advantages of statistical models and machine learning models (Wang et al., 2022). It decomposes the loss function into second-order Taylor expansion and improves the prediction accuracy. When comparing the relative importance between variables, it can output the results efficiently and accurately (Suo et al., 2019). At the same time, it has fast computing speed, flexibility to solve problems and ability to handle variables (Ma et al., 2018), which makes this method a better choice to study gender differences in educational attainments. In recent years, XGBoost algorithm has been widely used in complex disease prediction, price prediction, behavior prediction and other fields (Mo et al., 2019; Li & Zhang, 2020; Li et al., 2020), which proved its applicability to us.

Pearson correlation and other typical importance estimation methods can only determine the overall relationship, but cannot study the individual relationship. The Hapley Additive ExPlanations (SHAP) value proposed by Lundberg and Lee (2017) can solve this shortcoming, better reflect the influence of independent variables in each sample, and further understand how a single variable affects the output results. SHAP allows us to measure cross effects between variables, because of its ability to solve multicollinearity problems (Moncada-Torres et al., 2021).

1.4. Aim of the Study

In short, this study has two goals. The first is to compare the importance of different backgrounds of parents to gender differences in offspring education by XGBoost algorithm and interpretive SHAP value. The second is to explore the effect of interaction between different variables on gender differences in offspring education.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202209_5(9).0054

2. Material and Methods

2.1. Data

Due to the needs of research topics, this study is based on the survey data of the 2014 China Family Panel Studies (CFPS2014). CFPS is a national, large-scale and multidisciplinary social tracking survey project. The sample covers 25 provinces. The stratified multi-stage sampling design used in the survey enables the sample to represent about 95% of the Chinese population (Xie, 2012). The questionnaire design draws on the experience of many famous international surveys (such as PSID, CDS, HRS) and has been widely used in education, economic and social research. A large number of research results have proved the reliability and adaptability of the data.

After data cleaning and missing value elimination, 18016 samples were obtained. Taking into account the impact of social change, taking reform and opening up as the dividing point, the samples aged 40 years and below were selected as the analysis samples, and the sample size was finally determined to be 6951. Each sample includes himself and his or her parents.

2.2. Variables

2.2.1. Explained Variable

This study focuses on the influence of parents' background, gender role concept and the number of children on gender differences in educational attainment, so the explained variable is the level of educational attainment. Years of education based on time to better quantify the family's capital investment, directly as a measure of education.

2.2.2. Explanatory Variables

①Gender : male as the reference object, assigned 0, and female assigned 1. ②Parents' income level : considering that income is skewed, and because of the existence of 0 value, the natural logarithm is taken for linear transformation after adding 1 to the income data (Huang & Lin, 2009). ③Parental occupational status : translate occupational code into occupational prestige, drawing on the International Socio-Economic Index for Occupational Status (ISEI) developed by Ganzeboom et al (1992). ④Mother's level of education: compared to years of education, the highest degree can better reflect the cultural capital of parents. It is therefore used as an indicator of parental educational attainment. ⑤Gender role concept: The four variables of 'division of labor between men and women', 'women's marriage', 'women's children' and 'men's housework' which measure gender role concept in the data are merged into an index by K-means clustering.

It should be noted that this paper is a comparative study of the differences in parental influence, so the above variables are divided into father and mother.

2.2.3. Control Variables

In addition to the variables studied in this paper, some personal heterogeneity and structural factors may also affect educational attainment. Therefore, this study takes ethnic groups (Cherng et al., 2019), regions, urban (Rodríguez-Pose, 2009) and rural types (Meng, 2012) and other factors as control variables.

2.3. Analytical Strategy

For the main research variables in this paper, XGBoost is used for in-depth explanation and analysis. The core of the XGBoost algorithm selected in this study is based on the regression tree model, and hundreds or even thousands of regression tree models constructed by continuously extracting some variables are linearly combined to obtain the final model. It does not rely solely on a single model prediction, which makes the model ensemble prediction

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202209_5(9).0054

results more realistic. Its basic idea is the same as gradient lifting tree (GBDT), but it is partly optimized. The overall objective function is described by the expression of GBDT as follows.

$$L^{(t)}=\sum_{i=1}^{n} l\left(y_i, y_i^{(t-1)}+f_t(x_i)\right)+\sum_k \Omega(f_k) \tag{1}$$

Formula (1) $f_t(x_i)$ is the model of the t tree, and $\sum_k \Omega(f_k)$ is the regularization term of the model. By optimizing the loss function and regularization term, the final objective function is obtained.

$$L^{(t)}=\sum_{j=1}^{T} l\left[G_j W_j+\frac{1}{2}(H_j+\lambda)W_j^2\right]+\gamma T \tag{2}$$

This algorithm is simple, efficient and accurate. Because the data of both parents does not necessarily exist at the same time in this study, the advantage of XGBoost algorithm in dealing with missing data can also make the experimental results better.

The importance and influence direction of global variables are explored by using the performance advantages of XGBoost algorithm. Then, the models of men and women educational attainment are further fitted, and explanatory variables of different importance in the models of male and female educational attainment are found. Finally, SHAP is used to analyze the interaction of variables. Explain the complex XGBoost algorithm scientifically and explore the causes of gender differences in educational attainment.

3. Results

3.1. Descriptive Statistics

In 6951 analysis samples, there are 3328 women (47.9%), and the average years of education for women is 9.64 years, while there are 3623 men (52.1%), and the average years of education for men is 10.18 years. Gender differences in education are obvious. The regional distribution of samples is consistent with the overall situation. Basic information on parental and economic capital, cultural capital, gender identity and the number of children for men and women is provided in Tables 1 and 2.

Table 1. Descriptive statistics for the variables for man.

Variables	Father				Mother			
	Mean	SD	Skewness	Kurtosis	Mean	SD	Skewness	Kurtosis
child_num	1.68	0.58	0.96	0.58	1.69	0.62	1.74	8.62
ISEI	31.42	184.27	1.92	3.06	28.75	147.74	2.38	5.15
degree	2.59	1.19	0.20	-0.30	2.10	1.23	0.65	-0.41
income	4.08	19.14	0.34	-1.66	2.59	15.65	1.07	-0.59
gender_role	2.95	1.87	0.33	-1.24	2.80	1.81	0.63	-1.0

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202209_5(9).0054

Table 2. Descriptive statistics for the variables for woman.

Variables	Father				Mather			
	Mean	SD	Skewness	Kurtosis	Mean	SD	Skewness	Kurtosis
child_num	1.99	0.70	0.86	1.77	1.98	0.74	1.55	7.95
ISEI	33.12	195.72	1.67	2.18	29.94	142.15	1.88	2.96
degree	2.76	1.26	0.26	0.13	2.33	1.25	0.45	-0.39
income	4.61	24.24	0.20	-1.88	3.23	19.33	0.71	-1.37
gender_role	3.06	1.88	0.17	-1.31	2.91	1.99	0.36	-1.28

3.2. Results of Global Variable Analysis

k-Fold cross-validation was used to assess the performance of machine learning models. After adjusting the parameters, the XGBoost model established in this study obtains relatively excellent performance through 10-Fold cross-validation, and the training accuracy is 0.940. The experiment also compares the proposed model with the baseline model (SVR, RF, KNN), as shown in table 3. The results show that the MSE, RMSE, MAE, MAPE and other performance evaluation indexes of XGBoost are due to other models, and the accuracy of the model is high.

Table 3. Mean performance metrics and class sizes of machine learning models evaluated in this study.

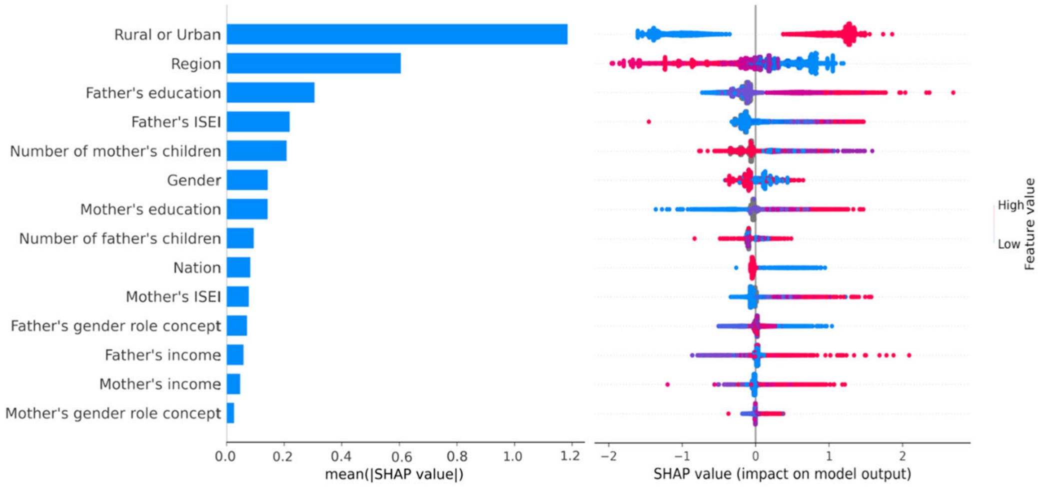
Classifier	Dataset	MSE	RMSE	MAE	MAPE	R^2
SVR	train	9.372	3.061	2.369	23.365	0.094
	10-Fold cross-validation	9.431	3.042	2.400	0.034	23.373
	test	11.753	3.428	2.545	29.232	-0.019
RF	train	4.861	2.205	1.806	16.437	0.530
	10-Fold cross-validation	8.149	2.847	2.301	0.165	20.736
	test	10.902	3.302	2.610	23.735	0.055
KNN	train	6.706	2.590	2.058	18.715	0.352
	10-Fold cross-validation	10.461	3.228	2.606	-0.069	23.585
	test	13.279	3.644	2.911	25.789	-0.152
XGBoost	train	0.625	0.790	0.388	4.363	0.940
	10-Fold cross-validation	10.088	3.166	2.508	-0.034	23.933
	test	16.237	4.029	3.193	26.143	-0.408

Figure 1 visualises the behaviour of predictors within the XGBoost model in predicting the impact of parents on educational attainment of children, using SHAP values. The redder the color is, the greater the value of the variable itself is, and the bluer the color is, the smaller the value of the variable is. Figure 1(a) shows the average magnitude of SHAP values indicating the overall importance of each predictor within the model. Fig. 1(b) summarises the model behaviour for each local prediction (each dot represents an individual prediction), hence revealing the direction of effects at different levels of each predictor.

Figure 1(a) shows that the structural characteristics include the urban and rural types of residence and the region where they live, and have the greatest impact on the educational attainment of children, which is mainly due to regional differences that affect the concept of parents and the educational resources of children (Gates & Guo, 2014). In addition, father's education level has the greatest impact on children's education. Father's professional status has

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202209_5(9).0054

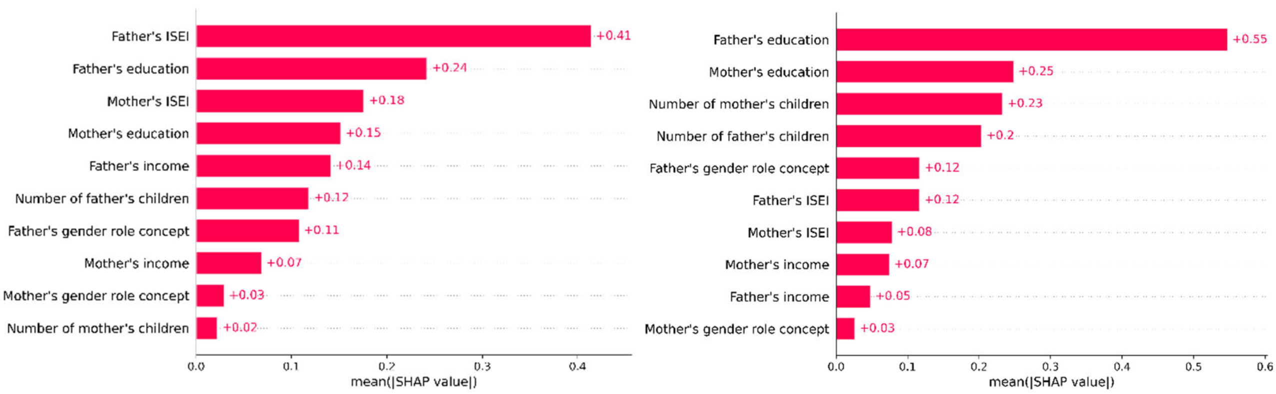
a greater impact on men's educational attainment, but it has less impact on women's educational attainment. At the same time, the number of fathers' children has a greater impact on women's educational attainment than men. As a whole, parents' income and gender roles have a relatively small impact on children's educational attainment. Combined with Figure 1(b), it can be clearly seen that parents' education level, income and occupational status have a positive impact on children's educational attainment. The number of parents' children is a negative impact on children's education, especially the mother's children.



(a) Global feature importance (b) Local explanation summary

Figure 1. A visual explanation of predictor behaviour based on the XGBoost prediction model.

In order to further compare the impact of parental educational attainment, income, occupational status, gender role attitudes and the number of children on gender differences in educational attainment of offspring, considering the removal of urban and rural types and provinces, two prediction models of educational attainment for men and women (M_men and M_women) were established to observe the explanatory variables. The results are shown in Fig. 2. In men's educational attainment model, father's occupational status has the greatest impact, followed by father's educational attainment and mother's occupational status. In contrast, the occupational status of parents has less impact on women's educational attainment. But like men, parents' educational attainment is largely affected. However, the influence of the number of children of parents with greater importance in the female education attainment model on men is not obvious in a single dimension. By comparing the two models, the SHAP value shows that the influence of parents' background on women is greater than that on men.



(a) Men (b) Women

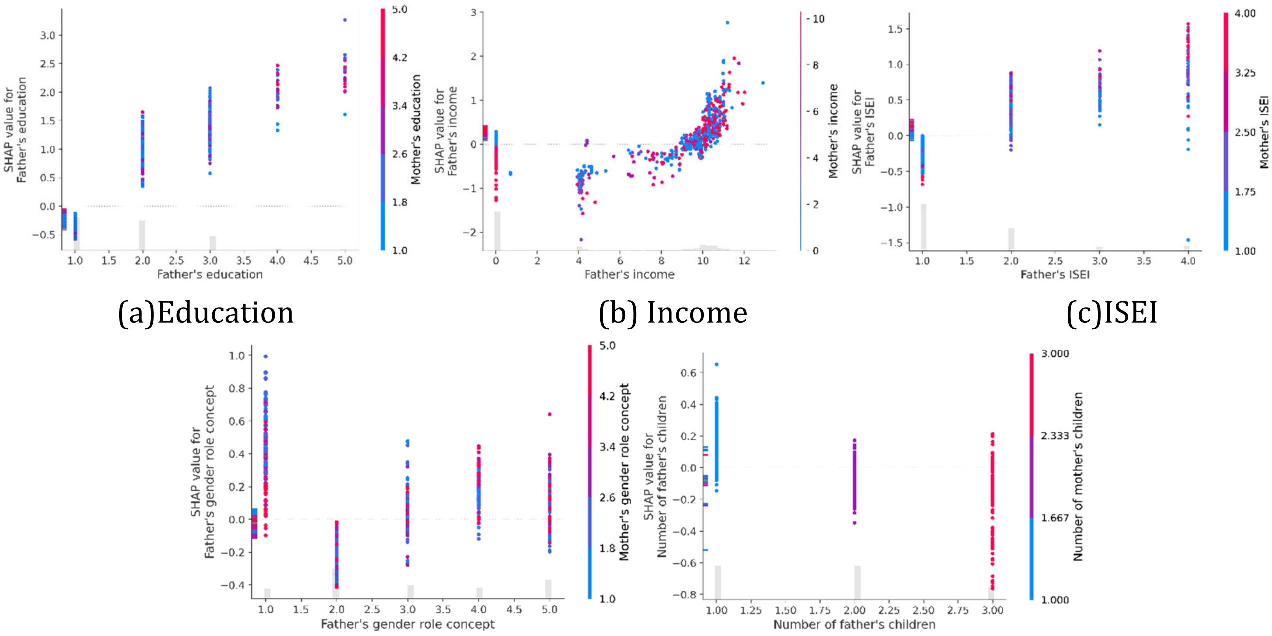
Figure 2. Gender-specific education attainment prediction model

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202209_5(9).0054

3.3. Complex Interactions Among Variables

The analysis of the prediction model obtained from the education of M_men and M_women shows that some variables have different impacts on men and women, because the XGBoost model captures the complex interaction between variables. Therefore, we analyze the SHAP interaction values to explore some predictive interactions. In the study, parents' variables are analyzed by interaction (reported in Figure 3(a)-3(e)). The diagonal of the image represents the main effect, and the upper and lower sides represent the interaction between the two variables.

Figure 3(a) shows the interaction between fathers' educational attainment and mothers' educational attainment. The educational attainment of mothers and fathers at or above the undergraduate level (4 and 5) improved the educational attainment of children. Figure 3(b) shows the interaction between mother's income and father's income. Figure 3(c) shows that when the father's occupational status is high, the mother's occupational status can also improve the educational attainment of the offspring. Figure 3(d) shows that the interaction between father's gender role concept and father's gender role concept is complex. When the father's concept of gender role is very low (value 1 represents a basic disapproval of the traditional concept of gender role), the mother's very high concept of gender role will reduce the educational attainment of children. Figure 3(e) shows that higher number of parents' children will reduce the educational attainment level of offspring.



(a)Education (b) Income (c)ISEI

(d) Gender role concept (e) Children's number

Figure 3. Interaction of parents' corresponding variables.

4. Discussion

The purpose of this study is to use a new machine learning method to analyze the gender differences in traditional education acquisition, and combine the theory and method of natural science and social science to more reliably explain the influence of parental background on gender differences in offspring 's education acquisition.

In this study, XGBoost algorithm and SHAP value were used to explore the factors that affect gender differences in educational attainment by parents' background, and better visualization

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202209_5(9).0054

was performed. The results show that in the background of parents, father's educational level, father's occupational status and mother's number of children are the most important factors affecting children's education. Father's educational level and father's professional status positively affect children's educational attainment. For father's background, the results of previous studies (Steinmayr, 2010; Davis-Kean, 2021) have been confirmed in current studies. The number of children of mothers negatively affects the educational attainment of children, which is consistent with the research results of Guo et al. (2017). They found that the number of children of parents had an impact on the equal quality of children's educational attainment, especially for children born first. In the comparative analysis of educational attainment models between men and women, it is found that the educational attainment level of women is more affected by parents' background than that of men (Gandhi Kingdon, 2002), and the educational attainment of parents and the number of children of parents have a greater impact on girls' educational attainment. It can be seen from the cross-impact between parents' educational level, income and occupational status that parents have a 'game' in their offspring's educational attainment, and that different backgrounds of parents will affect each other, and one of them will dominate. The impact of the interaction of similar variables between parents is also consistent with the concepts of McElroy and Horney (1981) and Manser and Brown (1980) 'bargain models'.

In addition, this study found that parents' cultural capital had a greater impact on children's educational attainment than economic capital. This is mainly because under the influence of the social environment, the overall economic level of China's society has been greatly improved. At the same time, China's nine-year compulsory education has borne a large part of the education cost, which reduces the educational burden of the family. Compared with economic and cultural capital, the influence of parents' gender role concept on children's educational attainment is weaker, but the interaction is more obvious. The influence of mother's background on children's educational attainment is large, not weaker than that of father (Korupp, 2002). Compared with father's background, mother's background has greater influence on women than men. Women's educational attainment is closely related to their parents' background. This gap is influenced by parents' economic status, education level and concept. Parents have more capital, and children's education is more likely to be improved. At the same time the impact on women is greater.

4.1. Implications

Whether male or female, parents' background is very important for their education. From the current research results, the gender difference in education attainment is still relatively obvious, and the influence of parents on the gender difference in education attainment of children is also different. There is still much work to be done to narrow the gender gap in Chinese education.

Men are more likely to get more investment in educational resources from their parents than women, while women may get less educational resources due to the traditional gender role concept. However, when parents' economic and cultural resources are at a high level, parents' investment in children' s education will also increase. Therefore, in order to improve the overall education level of society, policy makers must decide how to increase income and provide educational assistance to low-income families (Xue et al., 2020). In this regard, China's nine-year compulsory education has played a good effect, this policy is worth further optimization. In addition, the increase in the number of parents' children also makes family resources diluted (Wang & Feng, 2021), parents have to invest in selective education in limited resources. Taking into account the level of educational attainment of children, supported by quantity-quality (Q-Q) model of children, which was proposed by Becker and Lewis (1973), parents should plan the number of affordable children from their own resources in order to achieve quality education

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202209_5(9).0054

for children. China's one-child policy has effectively improved women's educational attainment over the past few decades (Zhang, 2017) and reduced gender disparities in access to education. Therefore, policy makers should encourage high-quality fertility and avoid over-fertility resulting in insufficient overall allocation of educational resources.

Another strategy that can be effective in improving children's educational attainment and reduce gender differences is to improve the status of mothers. The research results show the importance of mothers in the acquisition of children's education. The promotion of mother's status can effectively enhance the positive impact of father and reduce the negative impact of father. Therefore, in order to improve the overall social education level, more attention should be paid to the mother's socio-cultural and economic status and the relative discourse power in the family. It is suggested that universities should introduce relevant policies to support female students to complete their studies. Government and enterprises should encourage and support women's employment to improve women's income and professional status. Reasonable support of relevant policies may effectively improve inequality and gender differences in education.

4.2. Limitations and Future Directions

There were several major limitations in this study. First, the research data we use is not the latest data.For the sake of sample size and research credibility, this study uses CFPS data. However, limited by the matching degree between the design of the questionnaire scale and this study, the data of 2014 can only be used. Due to the early years of data, the analysis may have a certain degree of error with the current situation. One caveat that should be kept in mind is that in the CFPS survey, family members may have different concepts about gender role concept, and it may be biased to summarize gender role concept from four dimensions. Furthermore, this study did not take into account the impact of parents' marital relationship on children's educational attainment. The effects of fathers and mothers were analyzed separately when using machine learning methods for model fitting, but the effects of parents' background on children in single-parent families and restructured families may differ from those in ordinary families (Ginther & Pollak, 2004). Moreover, the introduction of the new algorithm in this study is just a preliminary exploration, and it is also worth improving to achieve better fitting results.

Due to the limitations of the data itself, the experiment has some shortcomings. However, this study innovatively uses XGBoost algorithm which has good effect in machine learning algorithm to analyze the gender difference field obtained by traditional education, which brings new explanations for the research in this field. The depth of this study is still limited, and then we can use a richer questionnaire to collect sample data, so as to further explore the mechanism of action of various factors.

5. Conclusion

This study expands the traditional research methods in the field of gender differences in educational attainment. This study finds that XGboost algorithm performs better in exploring the impact of parents' related capital on children's educational attainment and gender differences. The research results reveal the important influence of mother's background on gender differences in children's education. Governments and social organizations should take active measures to alleviate gender disparities in access to education in China. It is particularly important to pay more attention to the status of mothers and the needs of women.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202209_5(9).0054

References

[1] Asadullah, M. N., & Chaudhury, N. (2009). Reverse gender gap in schooling in Bangladesh: insights from urban and rural households. Journal of Development Studies, 45(8), 1360-1380. http://dx.doi.org/10.1080/00220380902935824

[2] Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. Annual Review of Economics, 11, 685-725. https://www.annualreviews.org/doi/pdf/10.1146/annurev-economics-080217-053433

[3] Bajari, P., Nekipelov, D., Ryan, S. P., & Yang, M. (2015). Demand estimation with machine learning and model combination (No. w20955). National Bureau of Economic Research. https://doi.org/10.3386/w20955

[4] Baker, D. P., & Stevenson, D. L. (1986). Mothers' strategies for children's school achievement: Managing the transition to high school. Sociology of education, 156-166. https://doi.org/10.2307/2112340

[5] Bauer, J., Feng, W., Riley, N. E., & Xiaohua, Z. (1992). Gender inequality in urban China: Education and employment. Modern China, 18(3), 333-370. https://doi.org/10.1177/009770049201800304

[6] Bauer, T., & Gang, I. (2001). Sibling rivalry in educational attainment: The German case. Labour, 15(2), 237-255. https://doi.org/10.1111/1467-9914.00163

[7] Becker, G. S., & Lewis, H. G. (1973). On the Interaction between the Quantity and Quality of Children. Journal of political Economy, 81(2, Part 2), S279-S288. https://doi.org/10.1086/260166

[8] Blake, J. (1989). Number of siblings and educational attainment. Science, 245(4913), 32-36. https://doi.org/10.1126/science.2740913

[9] Blau, P. M., & Duncan, O. D. (1967). The American occupational structure. https://doi.org/10.2307/2092399

[10] Bourdieu, P. (1971). Reproduction culturelle et reproduction sociale. Social Science Information, 10(2), 45-79. https://doi.org/10.1177%2F053901847101000203

[11] Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. Annual review of psychology, 53(1), 371-399. https://doi.org/10.1146/annurev.psych.53.100901.135233

[12] Bukodi, E., & Goldthorpe, J. H. (2013). Decomposing 'social origins': The effects of parents' class, status, and education on the educational attainment of their children. European sociological review, 29(5), 1024-1039. https://doi.org/10.1093/esr/jcs079

[13] Butcher, K. F., & Case, A. (1994). The effect of sibling sex composition on women's education and earnings. The Quarterly Journal of Economics, 109(3), 531-563. https://doi.org/10.2307/2118413

[14] Cherng, H. Y. S., Hannum, E., Lu, C., Kong, P. A., & Yu, X. (2019). China: Sociological perspectives on ethnicity and education: Views from chinese and english literatures. The Palgrave handbook of race and ethnic inequalities in education, 301-344. https://doi.org/10.1007/978-3-319-94724-2_8

[15] Cheung, F. M., & Halpern, D. F. (2010). Women at the top: Powerful leaders define success as work+ family in a culture of gender. American psychologist, 65(3), 182. https://psycnet.apa.org/doi/10.1037/a0017309

[16] Churpek, M. M., Yuen, T. C., Winslow, C., Meltzer, D. O., Kattan, M. W., & Edelson, D. P. (2016). Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards. Critical care medicine, 44(2), 368–374. https://doi.org/10.1097/CCM.0000000000001571

[17] Davis-Kean, P. E., Tighe, L. A., & Waters, N. E. (2021). The role of parent educational attainment in parenting and children's development. Current Directions in Psychological Science, 30(2), 186-192. https://doi.org/10.1177%2F0963721421993116

[18] Erola, J., Jalonen, S., & Lehti, H. (2016). Parental education, class and income over early life course and children's achievement. Research in Social Stratification and Mobility, 44, 33-43. https://doi.org/10.1016/j.rssm.2016.01.003

[19] Feng, N. (2021). The effect of sibling size on children's educational attainment: Evidence from Indonesia. ECNU Review of Education, 4(4), 830-856. https://doi.org/10.1177%2F2096531120921703

[20] Fleming, S. W., & Goodbody, A. G. (2019). A machine learning metasystem for robust probabilistic nonlinear regression-based forecasting of seasonal water availability in the US west. IEEE Access, 7, 119943-119964. https://doi.org/10.1109/ACCESS.2019.2936989

[21] Gates, P., & Guo, X. (2014). How British-Chinese parents support their children: a view from the regions. Educational Review, 66(2), 168-191. https://doi.org/10.1080/00131911.2013.768958

[22] Gandhi Kingdon, G. (2002). The gender gap in educational attainment in India: How much can be explained?. Journal of Development Studies, 39(2), 25-53. https://doi.org/10.1080/00220380412331322741

[23] Ganzeboom, H. B., De Graaf, P. M., & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. Social science research, 21(1), 1-56. https://doi.org/10.1016/0049-089x(92)90017-b

[24] Gelman, A., & Hennig, C. (2017). Beyond subjective and objective in statistics. Journal of the Royal Statistical Society: Series A (Statistics in Society), 180(4), 967-1033. https://doi.org/10.1111/rssa.12276

[25] Gibb, S. J., Fergusson, D. M., & Horwood, L. J. (2008). Gender differences in educational achievement to age 25. Australian Journal of Education, 52(1), 63-80. https://doi.org/10.1177%2F000494410805200105

[26] Ginther, D. K., & Pollak, R. A. (2004). Family structure and children's educational outcomes: Blended families, stylized facts, and descriptive regressions. Demography, 41(4), 671-696. https://doi.org/10.1353/dem.2004.0031

[27] Goldscheider, F., Bernhardt, E., & Lappegård, T. (2015). The gender revolution: A framework for understanding changing family and demographic behavior. Population and development review, 41(2), 207-239. https://doi.org/10.1111/j.1728-4457.2015.00045.x

[28] Gumus, M., & Kiran, M. S. (2017, October). Crude oil price forecasting using XGBoost. In 2017 International conference on computer science and engineering (UBMK) (pp. 1100-1103). IEEE. https://doi.org/10.1109/UBMK.2017.8093500

[29] Guo, R., Yi, J., & Zhang, J. (2017). Family size, birth order, and tests of the quantity–quality model. Journal of Comparative Economics, 45(2), 219-224. https://doi.org/10.1016/j.jce.2016.09.006

[30] Handa, S. (1996). Maternal education and child attainment in Jamaica: Testing the bargaining power hypothesis. Oxford Bulletin of Economics and Statistics, 58(1), 119-137. https://doi.org/10.1111/j.1468-0084.1996.mp58001006.x

[31] Hart, C. S. (2019). Education, inequality and social justice: A critical analysis applying the Sen-Bourdieu Analytical Framework. Policy Futures in Education, 17(5), 582-598. https://doi.org/10.1177%2F1478210318809758

[32] Hatlebakk, M. (2017). Son preference, number of children, education and occupational choice in rural Nepal. Review of Development Economics, 21(1), 1-20. https://doi.org/10.1111/rode.12237

[33] Huang, H. C., & Lin, S. C. (2009). Non-linear finance–growth nexus: A threshold with instrumental variable approach 1. Economics of Transition, 17(3), 439-466. https://doi.org/10.1111/j.1468-0351.2009.00360.x

[34] Jacobs, J. A. (1996). Gender inequality and higher education. Annual review of sociology, 22(1), 153-185. https://doi.org/10.1146/annurev.soc.22.1.153

[35] Jayachandran, S. (2015). The roots of gender inequality in developing countries. economics, 7(1), 63-88. https://doi.org/10.1146/annurev-economics-080614-115404

[36] Jiang, Y., Tong, G., Yin, H., & Xiong, N. (2019). A pedestrian detection method based on genetic algorithm for optimize XGBoost training parameters. IEEE Access, 7, 118310-118321. https://doi.org/10.1109/ACCESS.2019.2936454

[37] Johnson, E., & Howard, T. C. (2008). Issues of difference contributing to US education inequality. In Inequality in education (pp. 444-460). Springer, Dordrecht. https://doi.org/10.1007/978-90-481-2652-1_19

[38] Kainuwa, A., Binti, N., & Yusuf, M. (2013). Influence of socio-economic and educational background of parents on their children's education in Nigeria. International journal of scientific and research publications, 3(10), 2250-3153. https: //www.academia.edu/download/37602723/ijsrp-p2258. pdf

[39] Kalff, A. C., Kroes, M., Vles, J. S. H., Bosma, H., Feron, F. J. M., Hendriksen, J. G. M., ... & Jolles, J. (2001). Factors affecting the relation between parental education as well as occupation and problem behaviour in Dutch 5-to 6-year-old children. Social psychiatry and psychiatric epidemiology, 36(7), 324-331. https://doi.org/10.1007/s001270170036

[40] Kim, H. J., & Hoi-Kyun, C. H. O. I. (2001). A comparative analysis of incident service time on urban freeways. IATSS research, 25(1), 62-72. https://doi.org/10.1016/S0386-1112(14)60007-8

[41] Klasen, S., & Lamanna, F. (2009). The impact of gender inequality in education and employment on economic growth: new evidence for a panel of countries. Feminist economics, 15(3), 91-132. https://doi.org/10.1080/13545700902893106

[42] Kollmayer, M., Schober, B., & Spiel, C. (2018). Gender stereotypes in education: Development, consequences, and interventions. European Journal of Developmental Psychology, 15(4), 361-377. https://doi.org/10.1080/17405629.2016.1193483

[43] Korupp, S. E., Ganzeboom, H. B., & Van Der Lippe, T. (2002). Do mothers matter? A comparison of models of the influence of mothers' and fathers' educational and occupational status on children's educational attainment. Quality and Quantity, 36(1), 17-42. https://doi.org/10.1023/A:1014393223522

[44] Li, D., & Tsang, M. C. (2003). Household decisions and gender inequality in education in rural China. China: An International Journal, 1(02), 224-248. https://doi.org/10.1142/S0219747203000153

[45] Li, H., Cao, Y., Li, S., Zhao, J., & Sun, Y. (2020). XGBoost model and its application to personal credit evaluation. IEEE Intelligent Systems, 35(3), 52-61. https://doi.org/10.1109/MIS.2020.2972533

[46] Li, J., & Lavely, W. (2003). Village context, women's status, and son preference among rural Chinese women. Rural sociology, 68(1), 87-106. https://doi.org/10.1111/j.1549-0831.2003.tb00130.x

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202209_5(9).0054

[47] Li, S., & Zhang, X. (2020). Research on orthopedic auxiliary classification and prediction model based on XGBoost algorithm. Neural Computing and Applications, 32(7), 1971-1979. https://doi.org/10.1007/s00521-019-04378-4

[48] Lu, J., Chen, S., Wang, W., & Ran, B. (2012). Automatic traffic incident detection based on nFOIL. Expert Systems with Applications, 39(7), 6547-6556. https://ascelibrary.org/doi/abs/10.1061/41186(421)241

[49] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30. https: //proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper. pdf

[50] Lv, Z. (2021, October). Reviewing Gender Discrimination in K12 Education. In 2021 International Conference on Public Relations and Social Sciences (ICPRSS 2021) (pp. 1176-1180). Atlantis Press. https://dx.doi.org/10.2991/assehr.k.211020.325

[51] Ma, X., & Piao, X. (2019). The impact of intra-household bargaining power on happiness of married women: evidence from Japan. Journal of Happiness Studies, 20(6), 1775-1806. https://doi.org/10.1007/s10902-018-0023-y

[52] Ma, X., Sha, J., Wang, D., Yu, Y., Yang, Q., & Niu, X. (2018). Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning. Electronic Commerce Research and Applications, 31, 24-39. https://doi.org/10.1016/j.elerap.2018.08.002

[53] Manser, M., & Brown, M. (1980). Marriage and household decision-making: A bargaining analysis. International economic review, 31-44. https://doi.org/10.2307/2526238

[54] McElroy, M. B. (1990). The empirical content of Nash-bargained household behavior. Journal of human resources, 559-583. https://doi.org/10.2307/145667

[55] McElroy, M. B., & Horney, M. J. (1981). Nash-bargained household decisions: Toward a generalization of the theory of demand. International economic review, 333-349. https://doi.org/10.2307/2526280

[56] Meng, X. (2012). Labor market outcomes and reforms in China. Journal of Economic Perspectives, 26(4), 75-102. https://doi.org/10.1257/jep.26.4.75

[57] Mo, H., Sun, H., Liu, J., & Wei, S. (2019). Developing window behavior models for residential buildings using XGBoost algorithm. Energy and Buildings, 205, 109564. https://doi.org/10.1016/j.enbuild.2019.109564

[58] Moncada-Torres, A., van Maaren, M. C., Hendriks, M. P., Siesling, S., & Geleijnse, G. (2021). Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival. Scientific reports, 11(1), 1-13. https://doi.org/10.1038/s41598-021-86327-7

[59] Ogunleye, A., & Wang, Q. G. (2019). XGBoost model for chronic kidney disease diagnosis. IEEE/ACM transactions on computational biology and bioinformatics, 17(6), 2131-2140. https://doi.org/10.1109/TCBB.2019.2911071

[60] Ozbay, K., & Noyan, N. (2006). Estimation of incident clearance times using Bayesian Networks approach. Accident Analysis & Prevention, 38(3), 542-555. https: // doi.org/ 10.1016 /j.aap. 2005. 11.012

[61] Post, D. (2001). Region, poverty, sibship, and gender inequality in Mexican education: Will targeted welfare policy make a difference for girls?. Gender & Society, 15(3), 468-489. https://doi.org/10.1177%2F089124301015003008

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202209_5(9).0054

[62] Robeyns, I. (2006). Three models of education: Rights, capabilities and human capital. Theory and research in education, 4(1), 69-84. https://doi.org/10.1177%2F1477878506060683

[63] Rodríguez-Pose, A., & Tselios, V. (2009). Education and income inequality in the regions of the European Union. Journal of Regional Science, 49(3), 411-437. https://doi.org/10.1111/j.1467-9787.2008.00602.x

[64] Roksa, J., & Potter, D. (2011). Parenting and academic achievement: Intergenerational transmission of educational advantage. Sociology of education, 84(4), 299-321. https://doi.org/10.1177%2F0038040711417013

[65] Schultz, T. P. (1999). Women's Role in the Agricultural Household: Bargaining and Human Capital. https://www.econstor.eu/bitstream/10419/98390/1/cdp803.pdf

[66] Song, K., Yan, F., Ding, T., Gao, L., & Lu, S. (2020). A steel property optimization model based on the XGBoost algorithm and improved PSO. Computational Materials Science, 174, 109472. https://doi.org/10.1016/j.commatsci.2019.109472

[67] Steinmayr, R., Dinger, F. C., & Spinath, B. (2010). Parents' education and children's achievement: The role of personality. European Journal of Personality, 24(6), 535-550. https: // doi.org/10.1002%2Fper.755

[68] Suo, G., Song, L., Dou, Y., & Cui, Z. (2019, November). Multi-dimensional short-term load Forecasting based on XGBoost and fireworks algorithm. In 2019 18th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES) (pp. 245-248). IEEE. https://doi.org/10.1109/DCABES48411.2019.00068

[69] Wang, K., Li, M., Cheng, J., Zhou, X., & Li, G. (2022). Research on personal credit risk evaluation based on XGBoost. Procedia Computer Science, 199, 1128-1135. https://doi.org/10.1016/j.procs.2022.01.143

[70] Wang, X., & Feng, X. (2021). Family Resource Dilution in Expanded Families and the Empowerment of Married Only Daughters: Evidence from the Educational Investment in Children in Urban China. Frontiers in Psychology, 12. https://dx.doi.org/10.3389%2Ffpsyg.2021.610482

[71] Xue, Y., Xuan, X., Zhang, M., Li, M., Jiang, W., & Wang, Y. (2020). Links of family-and school-level socioeconomic status to academic achievement among Chinese middle school students: A multilevel analysis of a national study. International Journal of Educational Research, 101, 101560. https://doi.org/10.1016/j.ijer.2020.101560

[72] Yang, J., Huang, X., & Liu, X. (2014). An analysis of education inequality in China. International Journal of Educational Development, 37, 2-10. https://doi.org/10.1016/j.ijedudev.2014.03.002

[73] Zeng, B. (2014). Women's political participation in China: improved or not?. Journal of International Women's Studies, 15(1), 136-150. https://vc.bridgew.edu/cgi/viewcontent.cgi?article=1739&context=jiws

[74] Zhang, J. (2017). The evolution of China's one-child policy and its effects on family outcomes. Journal of Economic Perspectives, 31(1), 141-60. https://doi.org/10.1257/jep.31.1.141

[75] Zhang, Y., Kao, G., & Hannum, E. (2007). Do mothers in rural China practice gender equality in educational aspirations for their children?. Comparative Education Review, 51(2), 131-157. https://doi.org/10.1086/512023

[76] Zhou, X., Moen, P., & Tuma, N. B. (1998). Educational stratification in urban China: 1949-94. Sociology of education, 199-222. https://doi.org/10.2307/2673202