sustainability

MDPI

Article

Digital Inclusive Finance and Family Wealth: Evidence from LightGBM Approach

Ying Liu [1,2,3], Haoran Zhao [1], Jieguang Sun [2,*] and Yahui Tang [2]

1 School of Management Science and Information Engineering, Jilin University of Finance and Economics, Changchun 130117, China
2 Key Laboratory of Financial Technology of Jilin Province, Changchun 130117, China
3 Business Big Data Research Center of Jilin Province, Changchun 130117, China
* Correspondence: 401065@jlufe.edu.cn

Abstract: With the rapid development of digital technology in China, Digital Inclusive Finance, which uses digital financial services to promote financial inclusion, is developing rapidly. This paper uses the Peking University Digital Financial Inclusion index of China and China Family Panel Studies (CFPS) data to construct a predictive model using the LightGBM machine learning algorithm to study whether Digital Inclusive Finance can predict household wealth and analyze the characteristics of strong predictive ability for household wealth. They found that: (1) the introduction of the Digital Financial Inclusion index can improve the prediction performance of the household wealth model; (2) financial literacy and age characteristics are the key characteristics of household wealth accumulation; (3) the coverage and depth of Digital Inclusive Finance has a significant effect on family wealth accumulation, but the degree of digitization acts as a disincentive factor. This paper not only uses machine learning methods to do research on Digital Inclusive Finance and family wealth from a more comprehensive perspective, but also provides effective theoretical support for the key factors that enhance family wealth.

Keywords: family wealth; machine learning; Digital Inclusive Finance

check for updates

Citation: Liu, Y.; Zhao, H.; Sun, J.; Tang, Y. Digital Inclusive Finance and Family Wealth: Evidence from LightGBM Approach. *Sustainability* **2022**, *14*, 15363. https://doi.org/10.3390/su142215363

Academic Editors: Albert Y.S. Lam and Yanhui Geng

Received: 25 September 2022
Accepted: 14 November 2022
Published: 18 November 2022

1. Introduction

In recent years, the popularity of Digital Inclusive Finance has largely improved the efficiency of financial services, as well as promoted innovation, consumption, and national income, which has had a huge impact on socioeconomic effects. China established inclusive finance as a national strategy for the first time after the State Council promulgated the Plan for Promoting the Development of Inclusive Finance (2016–2020). Inclusive finance refers to the provision of appropriate and effective financial services at affordable costs for all segments and groups of society in need. After the approval of the G20 Advanced Principles for Digital Inclusive Finance at the G20 Hangzhou Summit in 2016, China began to focus on promoting the integration of inclusive finance and digital technology, specifically financial products and services such as payments, credit, insurance, securities, and savings, through digital tools for transactions. Since then, China has entered the era of digital economy. Digital Inclusive Finance adapts to the scenario of traditional financial inclusive finance where there is little information, scattered distribution, small customer size, and lack of collateral. Digital Inclusive Finance has social and economic effects, such as improving financial service efficiency, optimizing resource allocation, and promoting innovation, consumption, and income growth, because it can solve the problems of information asymmetry, high cost, and high risk [1]. China has made remarkable achievements in economic and social development so far, and the national family wealth and comprehensive national capacity have increased significantly.

The family is an important demand side in the financial market, and it is an important element of social aggregation. Family wealth, as the material basis for a better life for the people, plays an important role in maintaining the stability of social and economic structure and alleviating the imbalance of development [2–4]. Therefore, it is important to study the effects of the impact of Digital Inclusive Finance on family wealth accumulation and to meet the demand for financial accessibility of families in all social classes and regions.

At present, most of the current literature on family wealth and Digital Inclusive Finance uses explanatory methods, which use statistical models to infer the causal relationship between variables based on hypothesis testing. These methods manually assume a specific functional form of the model beforehand, such as exponential relationships, linear relationships, and so on. However, many models that fit well have no specific functional form set in advance. In contrast, predictive models using machine learning algorithms start from the data itself, and gradually train the model to fit the data best by iterating many times according to the principle of constantly finding the optimal solution, so that more complex relationships between variables can be uncovered and patterns between the data can be explained, thus helping to advance the relevant theories. The machine learning model itself belongs to the black box model, but scholars improve the interpretability of machine learning algorithms by studying alternative variables, which improves the prediction effect of machine learning, and also supplements the theoretical category of interpretative methods [5].

In view of this, this paper designs a household wealth prediction indicator system, based on a large number of papers related to Digital Inclusive Finance and household wealth, using 38,459 household panel data and 77,697 individual panel data from 31 provinces and municipalities directly under the central government of China from 2014 to 2018. Furthermore, the LightGBM machine learning algorithm is used to verify whether Digital Inclusive Finance can predict household wealth. To analyze the relationship with household wealth in multiple dimensions to improve the model interpretation, the Accumulated Local Effects Plot algorithm is proposed to further explore the main factors affecting family wealth accumulation in three dimensions: breadth of coverage, depth of use, and degree of digitization. The remainder of the paper is organized as follows: (1) a review of the literature related to Digital Inclusive Finance, family wealth, and machine learning algorithms applied to the fields of finance and management, as well as providing a theoretical basis for the design of model indicator systems in Section 2; (2) introduce the data sources and the specific indicator system constructed in Section 3; (3) show the specific algorithm used, the experimental procedure, and the experimental results in Section 4; (4) analysis and discussion for experimental results in Section 5; and (5) finally, conclusions.

2. Review of the Literature

2.1. Factors Affecting Family Wealth

Regarding the analysis of factors affecting family wealth accumulation, scholars have conducted more adequate research, mainly including two perspectives of family characteristics and family member characteristics. In studying the household asset portfolio in China, Lei and Zhou [6] found that the health factors of urban residents have an important impact on guiding the distribution of family wealth. This was later confirmed in the findings of Xie [7], while this study found a significant effect of education and knowledge literacy, and whether or not self-employment had a differential impact on household wealth. However, the process in the study regarding the impact of knowledge literacy on household wealth has not been shown. Wu et al. [4] also argue that knowledge literacy can promote household wealth accumulation, as shown by the fact that households with higher financial literacy allocate more assets to financial assets, especially risky financial assets. Furthermore, this study also argues that household size has different degrees of positive effects on household wealth accumulation. There are also many examples in the literature that analyze this characteristic of household location and confirm that urban–rural differences in households and geographical differences lead to different asset allocation structures of households, which have an impact on household wealth [8–10] In the literature that has been studied from the perspective of family member's characteristics, some scholars have found that individual cognition [11,12] and individual education [13,14] allow individuals to understand financial instruments more comprehensively and allocate assets more rationally. It has also been found that mental health [15] individual self-control [16], childhood family structure [17], and gender [18] all have important effects on the accumulation of household wealth. From the above studies, it is clear that previous studies on household wealth have focused on certain characteristics to explore their relationship with household wealth, and there are few studies that have analyzed them under a comprehensive framework. In addition, the previous literature mostly measures and studies households and family members separately, while the impact indicator systems of households and family members are different, so there is a need to analyze them together. In this paper, based on the above-mentioned literature, we conduct the construction of a specific indicator system for the impact factors of household wealth, so as to prepare the premise for the experiments of the modeling method afterwards.

2.2. Digital Inclusive Finance and Family Wealth

Dedicated to providing financial services to groups excluded from the financial system, financial inclusion has grown to some extent worldwide. There is a range of literature that demonstrates the positive effects of financial inclusion on promoting social status equity [19], access to banking services [20], alleviating distributional inequalities [21], and optimizing social behavior outcomes [22]. Along with the advancement of digital technology, Digital Inclusive Finance has emerged, and its development breaks through time and space constraints and reduces the dependence of traditional transaction models on physical space. Currently, the Digital Inclusive Finance development index has risen from 40 in 2011 to 341.22 in 2020. Digital Financial Inclusion has enlivened the economy and quickly penetrated all aspects of household life [23,24], influencing household economic behavior and becoming a new factor affecting household wealth growth [25]. More and more scholars are exploring the impact of digital inclusion on countries (regions) from a macro perspective, for example, by looking at rural areas as an entry point. Song empirically analyzed the income gap between urban and rural residents in 31 Chinese provinces using the Theil index, and found that Digital Inclusive Finance can significantly reduce the urban–rural income gap [26]. Using multidimensional poverty as an entry perspective, Zhang et al. demonstrated that financial inclusion improves the viability of rural families, and thereby improves the quality of poverty reduction by reducing the likelihood of future poverty among rural households [27]. Based on the perspective of urban–rural differences, Zhang and Wu argued that Digital Financial Inclusion can narrow the gap between urban and rural incomes, improve the transmission mechanism of income distribution, and stimulate the entrepreneurial dynamics of rural residents, especially for small and micro or labor-intensive entrepreneurial activities [28]. Moreover, some scholars have also studied the impact of Digital Inclusive Finance on income from the perspectives of entrepreneurship [29], farmers' income increase [30], distribution patterns [31], and human capital [32]. In addition to macro analysis, some studies focused on micro households to explore the impact of Digital Inclusive Finance development on household financial asset allocation [4], entrepreneurship [33–35], residential consumption [27], household risk [36], credit [37], and other effects. As well as the positive effects of digital inclusion, some scholars argue that digital inclusion can have an impact on exacerbating the gap between rich and poor. Hu found that the lack of digital tools for low-income households made it difficult to participate in the process of Digital Inclusive Finance, which, in turn, suppressed low-income household income and increased relative poverty [38]. Direct studies use household net worth as the explanatory variable and discuss less analysis of factors of Digital Inclusive Finance for family wealth growth [3]. The above studies use explanatory models to test the causal relationship between variables and prove the positive effect of Digital Inclusive Finance development on promoting family wealth accumulation; more commonly used are structural equation and linear regression models [25,34,39,40]. On the one hand, explanatory models require specific functional assumptions and restrictions on the relationship between data and variables, but most real-world problems have difficulty assuming functional form among variables [41]; on the other hand, explanatory models are prone to overfitting in order to promote the perfect fitting of sample data when dealing with high-dimensional data, resulting in their poor extrapolation prediction [42] In the context of Digital Inclusive Finance, the characteristics that affect household wealth are complex and multidimensional, and the characteristics themselves can be relational. This is difficult to describe using traditional linear fitting models. Therefore, it is necessary to explore new and effective methods to deeply explore the factors of family wealth accumulation.

2.3. The Application of Machine Learning on Family Wealth Issues

Machine learning is an important area of artificial intelligence that stems from statistical model fitting. Machine learning uses inference and sample learning to derive appropriate theory from data, especially for solving problems such as "noisy" patterns and large data sets. It plays an increasingly important role in the analysis of large samples, multi-vectors, and uncertain data [43,44]. The main features of machine learning methods include: (1) machine learning can get the results that fit the data best by continuously learning and looping to optimize the target problem; (2) the fitted model can help us explore more statistical relationships between data system characteristics and variables, and can find more complex patterns in the data [45]; (3) machine learning uses a range of methods, such as regularization and pruning, to better solve the overfitting problem [46], which is difficult to achieve with explanatory models. Some scholars have now used machine learning methods to deal with certain economic-type problems, such as stock return prediction [47,48], the effect of executive characteristics on firm performance [49,50], commercial bank liquidity monitoring [51], the issue of changing organizational boundaries of firms and the selection of collaborative innovation partners [52,53], etc. However, there is less research literature on the analysis of factors influencing family wealth accumulation using machine learning methods.

Combing the above literature, we find that, firstly, there is a large amount of literature examining the impact of Digital Inclusive Finance on asset allocation and narrowing the wealth gap, but less literature examining the impact of Digital Inclusive Finance on household income and household wealth. Secondly, the studies focus more on the problem itself, few focus on the modeling methods, and lack of systematic and quantitative findings from the predictive capability. Under this background, this paper studies the impact of Digital Inclusive Finance on household wealth accumulation based on the LightGBM model. The main contribution of this study is that it can better identify the regular pattern and main factors that affect the accumulation of family wealth. At the same time, the proposed model also enriches the measurement methods of family wealth models.

3. Description of Data Sources and Variables

3.1. Data Source

The first part of the data is the "Digital Inclusive Finance Index of Peking University", which is jointly compiled by Peking University Digital Finance Research Center and Ant Group Research Institute, and the data set contains 33 specific indicators in 3 dimensions. The second part of the data is the China Family Panel Studies (CFPS) of the China Social Science Survey Center of Peking University, which has been conducted every two years since 2010, and the data covers three levels of individuals, families, and communities. Based on the need of model setting and data availability, individual and household data of three periods in 2014, 2016, and 2018 were selected, extracted, and combined, and finally, a total of 77,697 sample data were obtained for the three periods.

3.2. Variable Definition

From the existing literature, the definition of household wealth is unclear. In this paper, we refer to Hurst and Lusardi [54] Gai et al. [55] (2013), Gan et al. [56], and other scholars' methods of defining household wealth and the characteristics of the database. Finally, family net worth is used as a proxy variable for household wealth.

Digital Inclusive Finance consists of three dimensions: the breadth of coverage of digital finance, the depth of use of digital finance, and the degree of digitization of inclusive finance. The breadth of coverage of digital finance is expressed in the provision of adequate digital financial services. The depth of use of digital finance refers to the effective demand for digital financial services. The degree of digitization of inclusive finance is the convenience, low cost, and creditability of financial services. If Digital Inclusive Finance can contribute to household wealth growth, then breadth of reach, depth of use, and digitalization can also contribute to household wealth growth.

We filtered the rest of the core variables in this paper on the basis of the literature in Section 2 and summarized them into two main categories. One is the family-level characteristics, which include family size, urban–rural classification, whether working in agriculture, whether self-employed, and the amounts of books in the family collection. The other is the characteristics of family member-level, which can be divided into objective characteristics and subjective characteristics according to the ease of observation. Objective characteristics include: age, marital status, gender, health status, education level, whether accessing the Internet, and other variables that can be easily measured objectively. Subjective characteristics include: satisfaction with life, satisfaction with the environment, perception of the wealth gap, perception of employment, perception of medical care, perception of education, perception of housing, perception of social security, and other variables that are difficult to measure directly.

3.3. Descriptive Statistics

The descriptive statistics of family-level characteristics and family member-level characteristics are shown in Table 1, which includes the sample size, mean, and standard deviation of each characteristic. From the results of the description of family-level characteristics: the average net assets of household are CNY 586,323.3; the average size of family is 3–4 persons; 50.24% of respondents live in urban areas; 48.06% of families are engaged in agricultural work; 9.40% of families are engaged in the self-employed private sector; the average number of books in the household collection is 68; and the average financial inclusion index of each region is 217.68. From the results of the description of family member-level characteristics, in terms of objective characteristics: the age span is disparate, the highest is 102 years old, the lowest is 16 years old, and the average age is 46; the percentage of men was 49.27 and the percentage of women was 50.73, thus the gender ratio is almost equal; the average education is between junior high school and high school; the percentage of respondents using the Internet was 40.93; 79.19% of respondents are married; and the average health level of the respondents was relatively healthy. In terms of subjective characteristics: the average view of respondents on various social issues is more positive, and views on life satisfaction, environmental issues, wealth gap, employment issues, medical issues, education issues, housing issues, and social security issues are all above the average level and more positive and optimistic.

Table 1. Variable description statistics.

Variable	Variable Definition	Sample Size	Average Value	Standard Deviation
	family-level			
NetAsset	Family net worth	38,459	586,323.3	1,570,973
FamilySize	Family size	38,459	3.647209	1.869107
AgBagd	Whether work in agriculture (Yes = 1; No = 0)	38,459	0.4806417	0.4996316
InPrivate	Whether the family is engaged in self-employed private (Yes = 1; No = 0)	38,459	0.0940482	0.2918996
InCity	Whether the household is located in a town (yes = 1; no = 0)	38,459	0.5024312	0.5000006
BookSum	Household book collection	38,459	67.38932	333.8275
Index*	Digital Financial Inclusion index	38,459	217.6776	52.54776
coverage_breadth*	Breadth of coverage index	38,459	188.4211	56.74687
usage_depth*	Depth of use index	38,459	215.0571	68.72808
digitization_level*	Degree of digitization index	38,459	319.0469	65.27698
	Family member-level			
NetAsset	Family net worth	77,697	597,015.2	1,526,159
Age	Age of family member	77,697	46.28097	16.47327
Gender	Gender (male = 1; female = 0)	77,697	0.4927089	0.4999501
InNet	Internet access or not (Yes = 1; No = 0)	77,697	0.409308	0.4917093
Marry	Whether married	77,697	1.46868	1.112772
Health	Health status (0–4)	77,697	1.997889	1.227294
Education	Academic qualifications (No need to study = 0; elementary school = 1; middle school = 2; high school = 3; college = 4; high school = 5; bachelor's degree = 6; master's degree = 7; doctorate = 8)	77,697	2.626742	1.50079
LifePoint	Satisfaction with life (1–5)	77,697	3.810919	1.027157
EvPoint	Satisfaction with the environment (0–10)	77,697	6.694274	2.672565
PgPoint	Views on the gap between the rich and the poor (0–10)	77,697	7.005058	2.402166
EmpPoint	Opinions on employment (0–10)	77,697	6.469658	2.4133
EduPoint	Views on education (0–10)	77,697	6.247938	2.681608
MedcPoint	Opinions on medical care (0–10)	77,697	6.317966	2.618888
HousPoint	Opinions on housing (0–10)	77,697	6.121897	2.710084
SocisePoint	Opinions on social security (0–10)	77,697	5.959805	2.636506
Index*	Digital Inclusive Finance index	77,697	212.9693	52.87561
coverage_breadth*	Breadth of coverage index	77,697	183.3167	57.07364
usage_depth*	Depth of use index	77,697	210.6925	67.90457
digitization_level*	Degree of digitization index	77,697	341.9367	66.21179

4. Research Methods and Model Construction

4.1. Research Methods

4.1.1. LightGBM

LightGBM (Light Gradient Boosting Machine) is an open source framework proposed by Microsoft in 2017 that combines the ideas of integrated learning and decision trees, and it is an improved gradient boosting algorithm based on the GBDT (Gradient Boosting Decision Tree) algorithm. GBDT is an iterative decision tree algorithm where each learning is based on the residuals of the previous training. LightGBM adopts the Histogram-based Algorithm, Gradient-based One-Side Sampling (GOSS), and Exclusive Feature Bundling (EFB) technology to reduce the running time of the algorithm complexity. When LightGBM processes high-latitude and massive data, it occupies less memory, and has higher computing speed and prediction accuracy [57]. The specific modeling process is as follows: firstly, the initial value of the model is set to 0, as in Equation (1):

$$\rho_0 = 0 \tag{1}$$

ρ_0 denotes the initial value, followed by the calculation of the optimal split point among all features according to the Histogram-based Algorithm, at this point, the first tree is obtained, as in Equation (2):

$$\rho_1 = \rho_0 + \psi_1 \tag{2}$$

ρ_1 represents the prediction model of the first round, and ψ_1 denotes the first tree. Based on the first tree, we calculate the tree that minimizes the objective function, which is called the second tree. Put the second tree behind the first tree and become the prediction model for the second round. The above steps are called an iterative process. Repeat the above iteration until the last tree, as in Equation (3):

$$\rho_0 = 0\rho_1 = \rho_0 + \psi_1\rho_2 = \rho_1 + \psi_2 \ldots \rho_i = \rho_{i-1} + \psi_i \tag{3}$$

4.1.2. Accumulated Local Effects Plot (ALE)

Interpretability is summarized as the degree to which humans can understand the reasons for model decisions [58] In current research, most machine learning methods tend to increase the complexity of the model in order to increase its own prediction accuracy, leading to a decrease in the interpretability of the model [59]. In this paper, we try to use the Accumulated Local Effects Plot algorithm [60] to deeply analyze the explanatory relationship between Digital Financial Inclusion and family wealth on the LightGBM model.

The ALE algorithm describes the impact mode of a single feature variable on the prediction results, that is, controlling other features as observed values, and modifying the target features to change the model fitting results. The core of the observed impact model is calculating the prediction bias. First, the value of the target feature is divided into multiple intervals according to quantiles to form a grid. Then, the effects of all target features in an interval on a single instance are summed, and the average effect in the interval is calculated at the same time. Finally, accumulating the average effect of all grid intervals to find the impact result. Explain Equation (4) as:

$$\hat{\tilde{f}}_{j,ALE}(x) = \sum_{k=1}^{k_j(x)} \frac{1}{n_j(k)} \sum_{i:x_j^{(i)} \epsilon N_j(k)} \left[f\left(z_{k,j}, x_{\setminus j}^{(i)}\right) - f\left(z_{k,j}, x_{\setminus j}^{(i)}\right)\right] \tag{4}$$

where, x is the target feature, j is the number of features, k is the number of intervals divided, $n_j(k)$ is the total amount of data in the kth interval, $Z_{k,j}$ is the upper and lower bound taken for the kth interval of the jth feature, and $i : x_j^{(i)} \epsilon N_j(k)$ is at the ith sample point of the kth interval. Centralize the effects, that is, let each effect subtract the average effect. When the feature effect is positive, indicating that the feature has a positive effect on the predicted outcome, the expression is shown in Equation (5):

$$\tilde{f}_{j,ALE}(x) = \hat{\tilde{f}}_{j,ALE}(x) - \frac{1}{n} \sum_{i=1}^{n} \hat{\tilde{f}}_{j,ALE}\left(x_j^{(i)}\right) \tag{5}$$

4.2. Model Construction

4.2.1. Indicator System

In this paper, the family net worth (NetAsset) is set as the predicted variable, and the model is defined as a family model (Fa_Md) and a family member model (Fn_Md) according to the family characteristics and family member characteristics, and these two models are collectively called the Basic Model. In addition, the Digital Inclusive Finance index and its sub-indices are introduced in the Basic Model, including: breadth of coverage index, depth of use index, and degree of digitization index. The model, after the introduction of the Digital Inclusive Financial Index and its sub-indices, is defined as the Family Digital Inclusive Model (FaDi_Md) and the Family Member Digital Inclusive Model (FnDi_Md), and the two models are collectively called the Digital Inclusive Model. The specific model composition, predictor variables, and predicted variable composition are shown in Table 2.

Table 2. Model indicator system.

Predicted Variables	Model		Predictive Variables
NetAssets	Basic Model	Family Model (Fa_Md)	FamilySize, AgBagd, InPrivate, InCity, BookSum
		Family Member Model (Fn_Md)	Age, Gender, InNet, Marry, Health, Education, LifePoint, EvPoint, PgPoint, EmpPoint, EduPoints, MedcPoint, HousPoint, SocisePoint
	Digital Inclusive Model	Family Digital Inclusive Model (FaDi_Md)	FamilySize, AgBagd, InPrivate, InCity, BookSum, Index*(coverage_breadth*, usage_depth*, digitization_level*)
		Family Member Digital Inclusive Model (FnDi_Md)	Age, Gender, InNet, Marry, Health, Education, LifePoint, EvPoint, PgPoint, EmpPoint, EduPoint, MedcPoint, HousPoint, SocisePoint, Index* (coverage_breadth*, usage_depth*, digitization_level*)

Note: The Digital Inclusion Index* is synthesized from the coverage_breadth*, the usage_depth*, and the digitization_level*.

4.2.2. Algorithm Implementation

Step 1 randomly splits the data set of each period of the family model, with part of the samples as the training data set and part of the samples as the test data set. Step 2 constructs the LightGBM model, and the whole process of the experiment is done by using single-period data training and next-period data rolling fit. In the learning process, we should adjust the learning rate, the number of iterations, the depth of the single-base regression tree, and other parameters of the model. Step 3 calculates the goodness of fit R^2 and the mean absolute error MAE to test the model performance. Step 4 calculates the ranking index using the relative importance built into LightGBM to obtain the relative importance features that affect family wealth accumulation. Step 5 introduces the cumulative local effect algorithm to further analyze the impact of the Digital Financial Inclusion sub-index on family wealth prediction and marginal effect. The specific algorithm implementation steps are shown below.

Step 1: Data splitting

The data set is split proportionally according to the experimental requirements to form a training data set and a test data set.

Step 2: Build the LightGBM model

① Build the model

Using the model indicator system in Table 2 and following the process of Equations (1)–(3) to construct four models, respectively, as follows:

$$\rho_I = \sum_{I=1}^{I} \psi_i(X, \theta_i) \tag{6}$$

where, ρ_I is the established LightGBM model. ψ_i is the ith tree, I is the total number of trees, X is the predictor variable dataset, and θ_i is the set of learning parameters for the ith tree.

② Objective function

Since the problem in this paper belongs to the regression problem, the model internal loss function chosen is MSE: $T(\rho_{obs}, \rho_i) = (\rho_{obs} - \rho_i)^2$, and the objective function of the model can be obtained as:

$$\Gamma_i = T(\rho_{obs}, \rho_i) + \Omega(f_i) \tag{7}$$

The objective function Γ_i contains two components, the first term is the loss function T, in which ρ_{obs} are the control observations; the second term Ω is the regularization term.

③ Regular items

The regular items are used to control the complexity of the model to avoid overfitting problems, and two penalty functions are added to the regular items. The specific canonical terms are as follows.

$$\Omega(f_i) = \lambda \sum_{j=1}^{J} |w_j| + \frac{1}{2}\alpha \sum_{j=1}^{J} w_j^2 \quad (8)$$

In which w_j is the number of leaf scores of the model split, w_j^2 is the L2 parametric number of leaf node scores, λ, α are coefficients, and the specific values can be adjusted in the process of use, and the default values are used in this paper.

④ Parameter adjustment

According to the experimental comparison results, the learning rate is set to 0.1, the number of iterations is 6000, and the depth of a single regression tree is 7.

⑤ Rolling verification

A single period training data set is used for training and prediction, and then the next period prediction data set is used for validation, rolling backwards in sequence by year for fitting.

Step 3: Evaluation metrics

Evaluation metrics: one is the goodness of fit R^2, and the other is the mean absolute error MAE. The goodness of fit R^2 is calculated as:

$$R^2 = 1 - \frac{\sum_{i=0}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=0}^{n}(y_i - \overline{y})^2} = 1 - \frac{RSS}{TSS} \quad (9)$$

The mean absolute error MAE is calculated as:

$$MAE = \frac{1}{n}\sum_{l=1}^{n}|y_i - \hat{y}_i| \quad (10)$$

y_i is the true value of the test set, $\overline{y}$ and $\hat{y}_i$ are the mean value of the previous sample and the prediction value trained using the previous sample, respectively. RSS and TSS are the sum of squared residuals and total squared deviations between samples. R^2 takes the value range [0, 1], and the model's prediction ability is better when R^2 is closer to 1. MAE takes the value range [0, +∞], and the smaller the MAE, the better the prediction ability of the model.

Step 4: Relative importance ranking

In the splitting process of the decision tree in Equations (1)–(3), the individual feature weights and the number of times selected scores are counted, where the larger the feature is to the splitting point improvement performance measure (closer to the root node), the larger the weight is, the more decision trees are selected and the higher the score is. Finally, the results of a feature in all decision trees are averaged after weighted summation to obtain the importance score, and all features are ranked and processed accordingly.

Step 5: Accumulated local effects plot

The Digital Inclusive Finance index, the breadth of coverage index, the depth of use index, and the degree of digitization index are brought into Equations (4) and (5) as target features to measure the specific impact effects of the respective features with respect to the predicted features.

4.3. Empirical Research

4.3.1. A Comparison of Approaches to Digital Inclusive Finance for Household Wealth Building

The objectives of the experiments in this section include: (1) comparing the performance of machine learning methods and comparing the proposed model with the ordinary least squares (OLS) method; (2) verifying whether the introduction of the Digital Financial Inclusion index into the index system can improve the performance of household wealth prediction. The experimental data years span from 2014 to 2018, and the experiments are fitted in a rolling fashion, with the following specific experimental procedures: firstly, the data set is divided into two phases, 2014–2016 and 2016–2018; each year of data is split according to 7.5:2.5, with the first 75% data set training the model and the last 25% data set validating the model. For example, the model was trained using 75% of the 2014 dataset, 25% of the 2016 dataset was used to validate the model, and then the model was trained again using 75% of the 2016 dataset to predict the next period, and so on. The experimental results are shown in Table 3, and it is easy to find that the LightGBM method outperforms the OLS method for both the Base Model (Fa_Md, Fn_Md) and the Digital Inclusive Model (FaDi_Md, FnDi_Md). Taking the Family Member Digital Inclusive Model (FnDi_Md) as an example, for the 2014–2016 and 2016–2018 data for both periods, the R^2 fitted using the LightGBM method is 20.72% and 19.96%, respectively, while the R^2 results of the OLS method are 11.83% and 14.10%, which are 8.89% and 5.86% higher, respectively. Similarly, for the MAE evaluation index, LightGBM obtains 27.82% and 36.08%, while the OLS results are 39.02% and 39.76%, which are 11.2% and 3.68% lower, respectively. The above analysis shows that the LightGBM method performs better than the traditional statistical model in fitting the data, so it is feasible to use the LightGBM machine learning method to explore the relationship between digital financial inclusion and household wealth.

Table 3. Fitting results.

Evaluation Methods	Crossover Year	Evaluation Indicators	Model Fitting Results				Degree of Promotion	
			Basic Model		Digital Inclusive Model			
			Fa_Md	Fn_Md	FaDi_Md	FnDi_Md	FaDi_Md–Fa_Md	FnDi_Md–Fn_Md
OLS	14–16	R^2 (%)	1.82	3.51	4.33	11.83	2.51	8.32
		MAE (%)	37.40	38.37	39.09	39.02	1.69	0.65
	16–18	R^2 (%)	7.01	1.76	18.19	14.10	11.18	12.34
		MAE (%)	39.16	45.13	35.73	39.76	−3.43	−5.37
LightGBM	14–16	R^2 (%)	4.33	5.82	11.36	20.72	7.03	14.9
		MAE (%)	31.15	34.19	26.92	27.82	−4.23	−6.37
	16–18	R^2 (%)	8.49	3.83	27.34	19.96	18.85	16.13
		MAE (%)	38.41	42.34	31.81	36.08	−6.6	−6.26

This part analyzes whether the Digital Inclusion Model can improve the predictive power of the model for household wealth to a greater extent, using the Base Model as a control group. According to the fitting results of LightGBM in Table 3, it can be observed that for the two periods of data from 2014 to 2016 and 2016 to 2018, the R^2 of the Family Digital Inclusive Model (FaDi_Md) is improved by 7.03% and 18.85%, respectively, compared to the Family Model (Fa_Md) at the family level. Similarly, at the family member level, the Family Member Digital Inclusive Model (FnDi_Md) compared to the Family Member Model (Fn_Md) increased R^2 by 14.9% and 16.13%. The results of the Digital Inclusion Model outperformed the Base Model both at the family level and at the family member level, and the results suggest that the introduction of the Digital Financial Inclusion index can significantly improve the predictive power of family wealth.

4.3.2. Analysis of the Key Features of Digital Financial Inclusion for Family Wealth Growth

The experiments in the previous section compare and analyze the fitting effects of the Base Model and the Digital Inclusion Model, verifying that the Digital Financial Inclusion index can improve the predictive power of household wealth, but the extent of its influential role in the prediction of results is not known. The purpose of the experiments in this section is to compare the relative importance of features in the Family Digital Inclusive Model (FaDi_Md) and the Family Member Digital Inclusive Model (FnDi_Md). Using the proposed model step 4 method, the scores obtained from feature selection during regression tree splitting were weighted and averaged and ranked, and the results are shown in Table 4. The experimental results found that for the Family Digital Inclusive Model (FaDi_Md), the family book collection (BookSum, 31.45%), Digital Financial Inclusion index (Index*, 24.95%), and family size (FamilySize, 24.20%) were ranked highly, while for the Family Member Digital Inclusive Model (FnDi_Md), the feature age (Age, 24.92%), education (Education, 21.89%), and Digital Financial Inclusion index (Index*, 10.05%) were more important. The results show that the Digital Financial Inclusion index (Index*) occupies an important position in both models because the technology platform of digital finance has a long-tail effect and low marginal cost, which can alleviate the information asymmetry phenomenon, reduce investment risk, enhance family investment participation, and thus promote the increase of household income and wealth growth. The family book collection (BookSum) reflects the overall knowledge and cultural literacy of the family, and the characteristic education is the education level of individual family members, which indicates that individual knowledge literacy and education level have a great role in promoting family wealth management, and family members with a higher literacy level are more conducive to the creation and increase of family wealth. The characteristic family size (FamilySize) indicates the number of family members. FamilySize indicates the number of family members who create wealth in the family. The age variable reflects the social experience, wealth accumulation experience, and wealth cognitive differences of individual family members. It means that the more people who create family wealth, the more conducive to family wealth accumulation. The elderly, with rich social experience as the leaders of family capital decision making, pay more attention to wealth accumulation.

Table 4. Feature importance order.

	Family Digital Inclusive Model (FaDi_Md)		Family Member Digital Inclusive Model (FnDi_Md)	
	Variables	**Feature Importance (%)**	**Variables**	**Feature Importance (%)**
1	BookSum	31.45	Age	24.92
2	Index*	24.95	Education	21.89
3	FamilySize	24.20	Index*	10.05
4	InCity	7.35	LifePoint	5.93
5	AgBagd	6.58	HousPoint	5.41
6	InPrivate	5.47	Health	5.34
7			EvPoint	3.99
8			InNet	3.98
9			MedcPoint	3.70
10			EmpPoint	3.53
11			EduPoint	3.13
12			PgPoint	2.64
13			Marry	2.55
14			SocisePoint	2.52
15			Gender	0.42

4.3.3. Multidimensional Analysis of the Effect of Digital Inclusive Finance on Family Wealth Growth

The above two experiments confirm that Digital Inclusive Finance is an important factor influencing household wealth accumulation from the perspective of algorithm and relative importance of features, respectively. Since the Digital Financial Inclusion index is synthesized from three sub-indices: breadth of coverage, depth of use, and degree of digitization, where breadth of coverage is reflected by sufficient digital financial products and financial services, depth of use is the effective demand for digital financial services, and degree of digitization is expressed as the flexibility, low cost, and creditability of financial services [3]. This section of the experiment uses the ALE to further explore the explanatory relationship between digital financial inclusion and household wealth in different dimensions.

The ALE algorithm acts on the household digital inclusion model (FaDi_Md) and the household member digital inclusion model (FnDi_Md), respectively, to produce a graph of the Digital Financial Inclusion index and its sub-indices in relation to household wealth (as shown in Figures 1–4). Figures 1a–c and 2a–c are the analysis graphs of the Family Digital Inclusive Model and the Family Member Digital Inclusion Model analysis diagram for 2014–2016. Figures 3a–c and 4a–c are the analysis diagram of the Family Digital Inclusive Model and the Family Member Digital inclusive Model for 2016–2018, respectively. The results of the four graphs show that as the Digital Financial Inclusion index, depth of use, and breadth of coverage increase, household wealth shows an increasing trend. The relationship between the Digital Financial Inclusion index, depth of use, breadth of coverage, and household wealth shows a more significant positive correlation. The results indicate that the development of Digital Inclusive Finance is conducive to increase the overall wealth of households, and that the breadth of coverage and depth of use elements of Digital Inclusive Finance are the main reasons for the positive correlation in the results. This may be due to the fact that digital inclusion offers a wide range of financial products and services through information technology, which lowers the threshold for households to participate in financial activities, thereby boosting household economic dynamics and increasing the total value of household wealth.

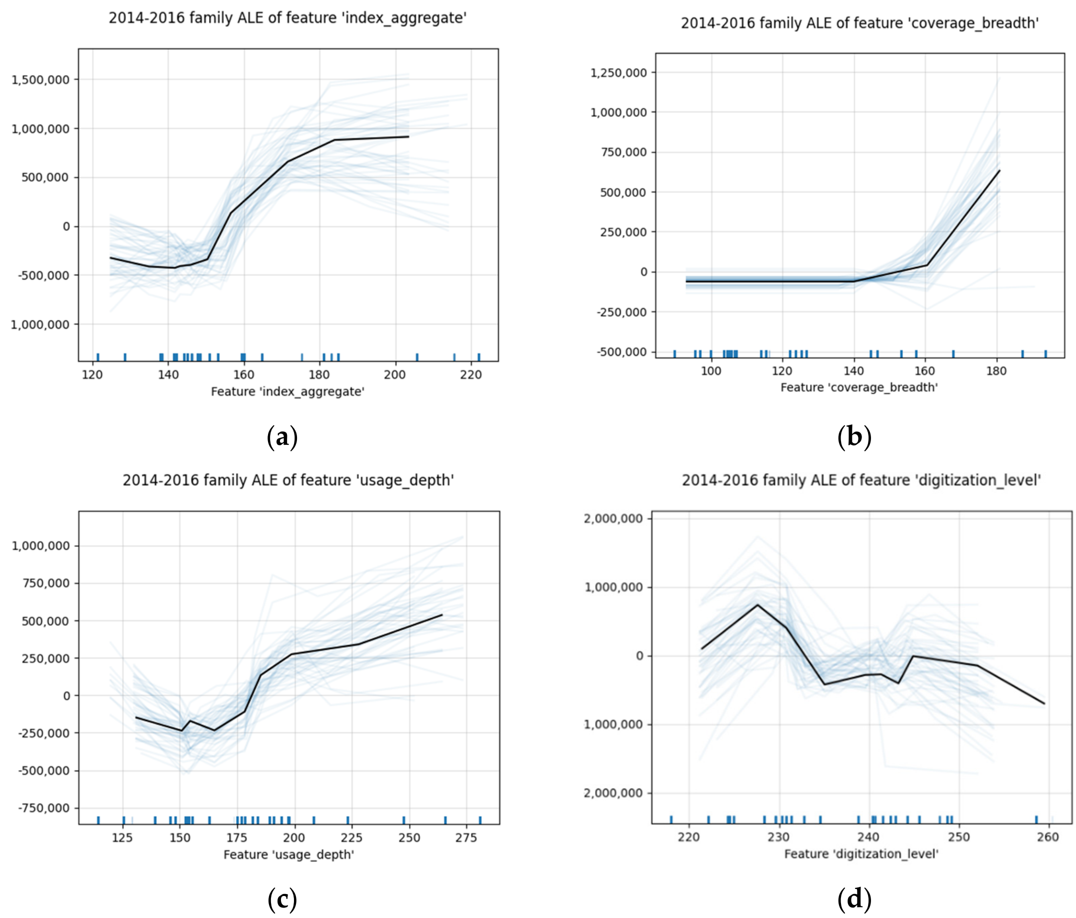

Figure 1. ALE analysis of Family Digital Inclusive Model (2014–2016). (**a**–**d**) represent the relationship between digital inclusion index, breadth of coverage index, depth of use index, and digitalization index and household wealth, respectively.

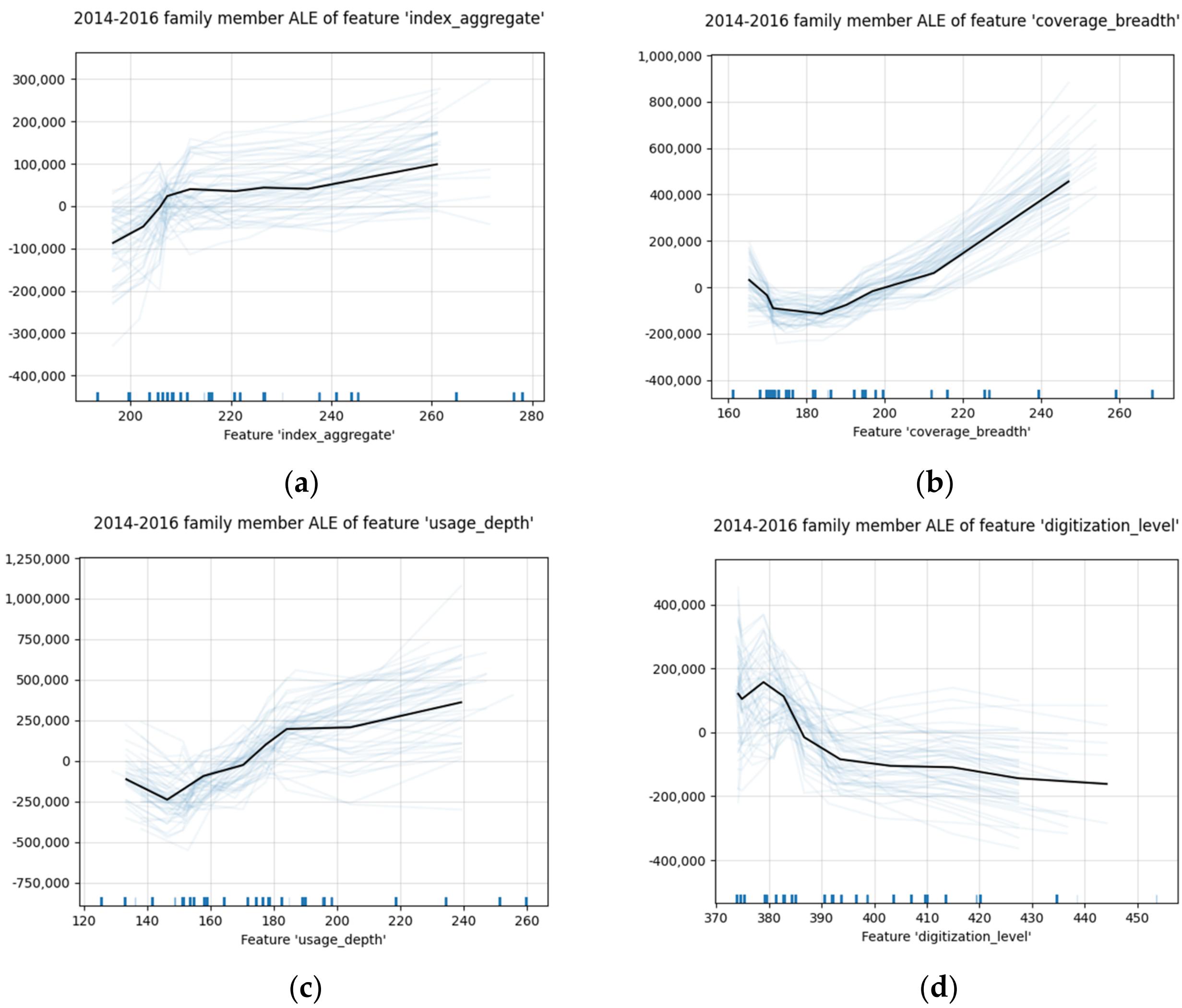

Figure 2. ALE analysis of Family Member Digital Inclusive Model (2014–2016). (**a**–**d**) represent the relationship between digital inclusion index, breadth of coverage index, depth of use index, and digitalization index and household wealth, respectively.

On the other hand, Figures 1d and 2d show that there is no obvious linear relationship between the degree of digitization and household wealth, indicating that the degree of digitization has less effect on family wealth and even has a suppressive effect within a certain range, while Figures 3d and 4d show the negative correlation between the degree of digitization and household wealth. This phenomenon indicates that there is a "digital divide" in Digital Inclusive Finance. The main reasons: (1) digital financial services rely excessively on Internet connection, which makes them unable to provide services to people who do not have digital devices and have low access to digital technology; (2) due to the different education levels of households, there is a cognitive bias towards the products and services provided by financial institutions, which leads to financial exclusion; (3) network data vulnerability and network security problems also reduce the trust of customers in digital financial platforms. As a result of the "digital divide", the spatial environment for information access, participation and information sharing is hindered, and some groups are unable to successfully participate in economic and social activities, lacking opportunities and being isolated, thus aggravating the polarization between rich and poor and hindering the growth of household wealth.

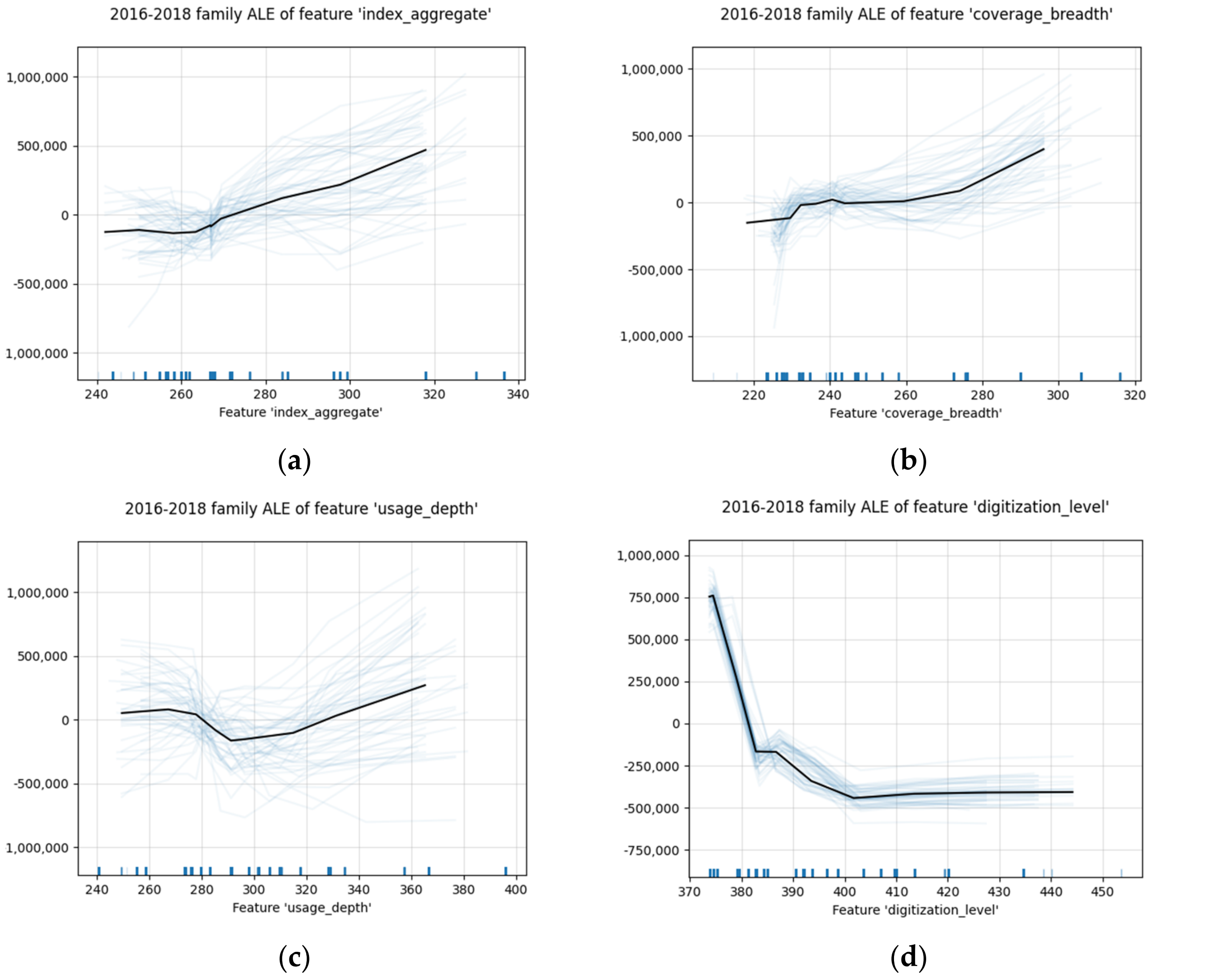

Figure 3. ALE analysis of Family Digital Inclusive Model (2016–2018). (**a**–**d**) represent the relationship between digital inclusion index, breadth of coverage index, depth of use index, and digitalization index and household wealth, respectively.

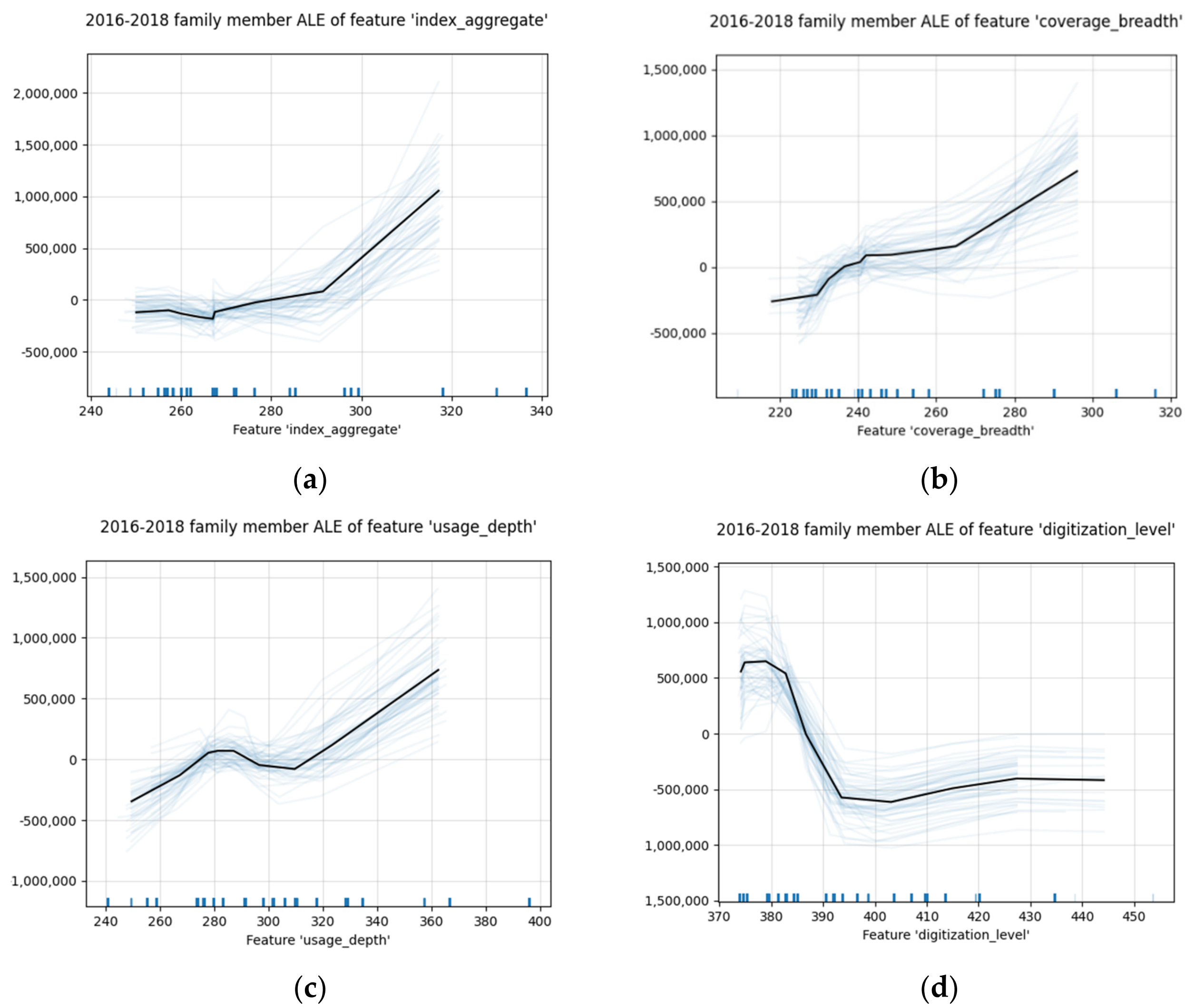

Figure 4. ALE analysis of Family Member Digital Inclusive Model (2016–2018). (**a**–**d**) represent the relationship between digital inclusion index, breadth of coverage index, depth of use index, and digitalization index and household wealth, respectively.

5. Discussion

5.1. Comparison Based on Machine Learning Methods and Suggestion

Table 3 shows the fitting results of the LightGBM method and the OLS method, which was commonly used in the previous literature for the data set. Based on the goodness of fit and mean absolute error results, we find that LightGBM outperforms the OLS method, which indicates that the effect of choosing LightGBM to improve the traditional linear model is very significant. Therefore, in future related research, we can try to use more machine learning algorithms to process the data to supplement the theoretical scope of the traditional linear model.

5.2. Key Characteristics Affecting Family Wealth and Suggestion

The results on the ranking of the importance of characteristics in Table 4 show that cognition and literacy play an important influence on household wealth, both at the level of the household as a whole and at the level of household members. However, the mechanism of cognition and literacy in influencing household wealth is usually manifested in the rational understanding of financial instruments and rational allocation of financial assets. Therefore, despite the rapid development of digital financial inclusion today, the relevant departments should still make great efforts to popularize financial literacy among the public, so that the ability of family wealth management and allocation can be enhanced and the overall level of family wealth can be improved.

5.3. Multidimensional Analysis for Digital Inclusive Finance and Suggestion

According to Figures 1–4, we find that the breadth of coverage and depth of use of Digital Inclusive Finance has a significant contribution to household wealth, and the higher the level of development, the more pronounced the promotion effect. Therefore, relevant authorities should continue to boost the coverage area and financial business functions of Digital Inclusive Finance, so that digital technology and inclusive finance can be more deeply integrated. Furthermore, it is important to implement the policy of financial inclusion and promote digital transformation so that more customer groups can participate in financial activities. Finally, we will enhance the public's awareness and acceptance of digital finance and their knowledge, so that more people can enjoy the benefits of digital financial inclusion, and thus increase their income.

Unfortunately, Digital Inclusive Finance has a significant disincentive effect on household wealth in terms of the degree of digitization due to a possible digital divide. Digital Inclusive Finance is overly reliant on internet connectivity and is unable to reach individuals without mobile phones or digital devices. Today, a large number of rural households in China still suffer from barriers to internet access, and many face a greater degree of digital exclusion. The digital divide has a dampening effect on household wealth accumulation, making the contribution of digital inclusion to household wealth weaker. Alleviating the 'digital divide' and achieving the goal of equalization are important elements in the development of Digital Inclusive Finance. In promoting the implementation of digital inclusion, the relevant authorities should consider the balanced characteristics of the hardware, as well as the universal rules of software design. It is also important to ensure data security and avoid information leakage.

6. Conclusions

Digital finance is one of the most influential forms in the development of inclusive finance, and also an important driving force for the development of inclusive finance. In the context of digital penetration, it is important to deeply explore the key factors affecting household wealth accumulation to enhance household property income and achieve wealth preservation and appreciation. Based on this, this paper uses data from the Peking University Digital Financial Inclusion index and China Household Tracking Survey to construct a more comprehensive household wealth prediction index system at both household and family member levels. The LightGBM machine learning method has been used to verify the relationship between Digital Inclusive Finance and household wealth, and a cumulative local effect algorithm is also introduced to multi-dimensionally analyze the impact of a digital inclusive sub-index on the impact of family wealth accumulation. It was found that the machine learning model could avoid the disadvantages of traditional statistical models, because it does not rely on samples to assume a specific functional form, and calculates the degree of influence of variables on target values based on node storage gain values. A machine learning model is more suitable for analyzing the nonlinear and interactive relationships between variables, identifying the contribution of variables, and then mining the data for more complex and reliable patterns. Secondly, the fitting effect of the model adding Digital Inclusive Finance is significantly improved, which indicates Digital Inclusive Finance has a positive impact on the prediction of household wealth; in addition, there is a strong positive relationship between Digital Inclusive Finance and individual knowledge literacy, family size, and member structure. Families with a strong understanding and rich life experience can better promote wealth creation and accumulation. Thirdly, the analysis of the effect of Digital Inclusive Finance on the growth of family wealth finds that the Digital Financial Inclusion index, the breadth of coverage, and the depth of use sub-indices all show a significant positive relationship with family wealth, while the degree of digitization has a negative impact, indicating that there is a "digital divide" phenomenon.

The shortcomings of this paper are that, on the one hand, the degree of development of digital financial inclusion is changing rapidly. There may be some deviations in the description of the real situation due to the constraints of data availability. On the other hand, this paper only uses one machine learning model method, and the arguments for robustness testing are not sufficient. The comparison of multiple models can be incorporated into the study in the future.

Author Contributions: Conceptualization, Y.L. and H.Z.; methodology, Y.L. and H.Z.; software, Y.L. and H.Z.; validation, Y.L. and H.Z.; formal analysis, Y.L. and J.S.; investigation, Y.L. and H.Z.; resources, Y.L. and H.Z.; data curation, H.Z.; writing—original draft preparation, H.Z.; writing—review and editing, Y.L. and H.Z.; visualization, H.Z.; supervision, J.S. and Y.T.; project administration, J.S. and Y.T.; funding acquisition, Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Social Science Foundation of China, grant number 20BTJ062.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The [family wealth related data e.g., "TotalAssect"] data that support the findings of this study are available from ["China Family Panel Studies (CFPS)"], [hyperlink to data source, "http://www.isss.pku.edu.cn/cfps/download" (accessed on 25 February 2021)]. The [index of Digital Inclusive Finance e.g., "coverage_breadth"] data that support the findings of this study are available from ["Digital Inclusive Finance Index of Peking University"], [hyperlink to data source, "https://idf.pku.edu.cn/zsbz/index.htm" (accessed on 18 February 2021)].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Huang, Y.P.; Huang, Z. China's Digital Finance Development: Now and the Future. *China Econ. Q.* **2018**, *17*, 1489–1502. [CrossRef]
2. Yin, Z.C.; Peng, C.Y.; Angela, C.L. The Development and Impact of Financial Inclusion of Household in China. *J. Manag. World* **2019**, *35*, 74–87. [CrossRef]
3. Zhou, T.Y.; Chen, M.X. Digital Penetration, Inclusive Finance and Household Wealth Growth. *J. Financ. Econ.* **2021**, *47*, 33–47. [CrossRef]
4. Wu, Y.; Peng, C.Y.; Yin, Z.C. Financial Literacy, Wealth Accumulation and Household Asset Allocation. *Mod. Econ. Sci.* **2016**, *38*, 19–29.
5. Greenwell Brandon, M.; Boehmke, B.C.; McCarthy, A.J.A. Simple and Effective Model-Based Variable Importance Measure, statistics. *arXiv* **2018**, arXiv:1805.04755.
6. Lei, X.Y.; Zhou, Y.G. On the Portfolio Choice of Chinese Households:Health and Risk Preference. *J. Financ. Res.* **2010**, *1*, 31–45.
7. Xie, M.B. Household Wealth, Education, and Age Equalization of Wealth—A Study of Household Net Worth Determinants Based on CHFS. *Southeast Acad. Res.* **2017**, *5*, 201–219. [CrossRef]
8. Wang, C.; Yao, L.; Chai, S.J. The Influence of Age Structure on Family Asset Allocation and Its Regional Difference. *Stud. Int. Financ.* **2017**, *2*, 76–86. [CrossRef]
9. Lu, X.M.; Zhao, S.; Luo, R.H. Will Regional Financial Development Promote Rational Household Investment?—From the Perspective of Household Asset Portfolios Diversification. *Res. Econ. Manag.* **2019**, *40*, 60–87. [CrossRef]
10. Yu, N.; Wang, Y. Can Digital Inclusive Finance Narrow the Chinese Urban–Rural Income Gap? The Perspective of the Regional Urban–Rural Income Structure. *Sustainability* **2021**, *13*, 6427. [CrossRef]
11. Smith, J.P.; McArdle, J.J.; Willis, R. Financial Decision Making and Cognition in a Family Context. *Econ. J.* **2010**, *120*, F363–F380. [CrossRef] [PubMed]
12. Meng, Y.J. Cognitive ability and household asset selection. *Econ. Res. J.* **2014**, *49*, 132–142.
13. Behrman, J.R.; Mitchell, O.S.; Soo, C.K.; Bravo, D. How Financial Literacy Affects Household Wealth Accumulation. *Am. Econ. Rev.* **2012**, *102*, 300–304. [CrossRef]
14. Hu, Z. Financial Literacy and Household Wealth Accumulation:Based on Micro-data from Urban Household in China. *J. Zhongnan Univ. Econ.* **2018**, *4*, 110–117. [CrossRef]
15. Johnston, D.; Kassenboehmer, S.C.; Shields, M.A. Financial decision-making in the household: Exploring the importance of survey respondent, health, cognitive ability and personality. *J. Econ. Behav. Organ.* **2016**, *132*, 42–61. [CrossRef]

16. Biljanovska, N.; Palligkinis, S. Control thyself: Self-control failure and household wealth. *J. Bank. Financ.* **2018**, *92*, 280–294. [CrossRef]
17. Bernardi, F.; Boertien, D.; Geven, K. Childhood Family Structure and the Accumulation of Wealth Across the Life Course. *J. Marriage Fam.* **2018**, *81*, 230–247. [CrossRef]
18. Chu, C.Y.C.; Kan, K.; Lin, J.C. Variations of wealth resemblance by family relationship types in modern Chinese families. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 6548–6553. [CrossRef]
19. Stein, L.; Yannelis, C. Financial Inclusion, Human Capital, and Wealth Accumulation: Evidence from the Freedman's Savings Bank. *Rev. Financial Stud.* **2020**, *33*, 5333–5377. [CrossRef]
20. Célerier, C.; Matray, A. Bank-Branch Supply, Financial Inclusion, and Wealth Accumulation. *Rev. Financ. Stud.* **2019**, *32*, 4767–4809. [CrossRef]
21. von Fintel, D.; Orthofer, A. Wealth inequality and financial inclusion: Evidence from South African tax and survey records. *Econ. Model.* **2020**, *91*, 568–578. [CrossRef]
22. Duvendack, M.; Mader, P. Impact of financial inclusion in low- and middle-income countries: A systematic review of reviews. *Campbell Syst. Rev.* **2019**, *15*, 594–629. [CrossRef]
23. Chen, L. From Fintech to Finlife: The case of Fintech Development in China. *China Econ. J.* **2016**, *9*, 225–239. [CrossRef]
24. Zhang, Q.; Posso, A. Thinking Inside the Box: A Closer Look at Financial Inclusion and Household Income. *J. Dev. Stud.* **2017**, *55*, 1616–1631. [CrossRef]
25. Li, J.; Wu, Y.; Xiao, J.J. The impact of digital finance on household consumption: Evidence from China. *Econ. Model.* **2019**, *86*, 317–326. [CrossRef]
26. Song, X.L. Empirical Analysis of Digital Inclusive Finance Bridging the Urban-rural Residents' Income Gap. *Financ. Econ.* **2017**, *6*, 14–25.
27. Zhang, D.H.; Yin, Z.C.; Sui, Y.B. Can Financial Inclusion Improve the Quality of Poverty Reduction:Analysis Based on Multidimensional Poverty. *South China J. Econ.* **2020**, *10*, 56–75. [CrossRef]
28. Zhang, B.Q.; Wu, W.T. Digital Inclusive Finance, Entrepreneurship and Income Distribution:An Empirical Research from the Perspective of Urban-rural Differences in China. *Chin. Rev. Financ. Stud.* **2021**, *13*, 31–44.
29. Ji, X.; Wang, K.; Xu, H.; Li, M. Has Digital Financial Inclusion Narrowed the Urban-Rural Income Gap: The Role of Entrepreneurship in China. *Sustainability* **2021**, *13*, 8292. [CrossRef]
30. Li, Y.; Wang, M.; Liao, G.; Wang, J. Spatial Spillover Effect and Threshold Effect of Digital Financial Inclusion on Farmers' Income Growth—Based on Provincial Data of China. *Sustainability* **2022**, *14*, 1838. [CrossRef]
31. Zhao, H.; Zheng, X.; Yang, L. Does Digital Inclusive Finance Narrow the Urban-Rural Income Gap through Primary Distribution and Redistribution? *Sustainability* **2022**, *14*, 2120. [CrossRef]
32. Song, J.; Zhou, H.; Gao, Y.; Guan, Y. Digital Inclusive Finance, Human Capital and Inclusive Green Development—Evidence from China. *Sustainability* **2022**, *14*, 9922. [CrossRef]
33. Xie, X.L.; Zhang, H.X.; Guo, F. Can Digital Finance Promote Entrepreneurship?—Evidence from China. *China Econ. Q.* **2018**, *17*, 1157–1580. [CrossRef]
34. Zhang, X.; Wan, G.H.; Zhang, J.J.; He, Z.Y. Digital Economy, Financial Inclusion, and Inclusive Growth. *Econ. Res. J.* **2019**, *54*, 71–86. [CrossRef]
35. Jia, L.; Tan, W.; Abu, M. Financial Literacy, Family Wealth and Family Entrepreneurial Decision-Making. *Southwest Financ.* **2021**, *1*, 83–96.
36. Sun, Y.; Zhang, T.; Wang, X.; Li, D. Status, Problems and Prospects of Digital Inclusive Finance in China. *Quant. Technol. Econ.* **2021**, *38*, 43–59. [CrossRef]
37. Fu, Q.Z.; Huang, Y.P. The Heterogeneous Impact of Digital Finance on Rural Financial Needs—Evidence from China Household Finance Survey and Peking University Digital Inclusive Finance Index. *J. Financ. Res.* **2018**, *11*, 68–84.
38. Hu, L.; Yao, S.Q.; Yang, C.Y.; Ji, L.H. Is digital financial inclusion good for alleviating relative poverty? *J. Financ. Econ.* **2021**, *47*, 93–107. [CrossRef]
39. Yu, C.; Jia, N.; Li, W.; Wu, R. Digital inclusive finance and rural consumption structure—Evidence from Peking University digital inclusive financial index and China household finance survey. *China Agric. Econ. Rev.* **2021**, *14*, 165–183. [CrossRef]
40. Ye, Y.; Pu, Y.; Xiong, A. The impact of digital finance on household participation in risky financial markets: Evidence-based study from China. *PLoS ONE* **2022**, *17*, e0265606. [CrossRef]
41. Hong, Y.; Wang, S. Big Data, Machine Learning and Statistics: Challenges and Opportunities. *China J. Econom.* **2021**, *1*, 17–35.
42. Araújo, A. Polynomial regression with reduced over-fitting—The PALS technique. *Measurement* **2018**, *124*, 515–521. [CrossRef]
43. Adner, R.; Puranam, P.; Zhu, F. What Is Different About Digital Strategy? From Quantitative to Qualitative Change. *Strat. Sci.* **2019**, *4*, 253–261. [CrossRef]
44. Nambisan, S.; Lyytinen, K.; Majchrzak, A.; Song, M. Digital Innovation Management: Reinventing Innovation Management Research in a Digital World. *MIS Q.* **2017**, *41*, 223–238. [CrossRef]
45. Athey, S.; Imbens, G.W. The State of Applied Econometrics: Causality and Policy Evaluation. *J. Econ. Perspect.* **2017**, *31*, 3–32. [CrossRef]
46. Chernozhukov, V.; Chetverikov, D.; Demirer, M.; Duflo, E.; Hansen, C.; Newey, W.; Robins, J. Double/debiased machine learning for treatment and structural parameters. *Econ. J.* **2018**, *21*, C1–C68. [CrossRef]

47. Sezer, O.B.; Ozbayoglu, A.M. Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach. *Appl. Soft Comput.* **2018**, *70*, 525–538. [CrossRef]
48. Hu, W.X.; Su, Z.X.; Yang, L. Research on the Prediction Effect of Investor Attention to the Concept Index on the Re-turn of Concept Index and the Investment Strategy of Concept Index Based on the Random Forest Algorithm. *Front. Sci. Technol. Eng. Manag.* **2021**, *40*, 60–66.
49. Malhotra, S.; Reus, T.H.; Zhu, P.; Roelofsen, E.M. The Acquisitive Nature of Extraverted CEOs. *Adm. Sci. Q.* **2017**, *63*, 370–408. [CrossRef]
50. Lu, Y.; Zhang, Y.Q.; Li, H.; Zhao, H.Y. Managerial individual characteristics and corporate performance: Evidence from a machine learning approach. *J. Manag. Sci. China* **2020**, *23*, 120–140.
51. Wen, B.H.; Gou, S.D.; Zhang, B.R. Evaluation of Liquidity Monitoring Methods of China's Commercial Banks in the Process of Leverage Adjustment: Comparison of Algorithms Based on Deep Feedforward Networks. *Front. Sci. Technol. Eng. Manag.* **2019**, *38*, 59–66.
52. Chen, D.M.; Wang, L.Z.; Chen, A.N. Digitalization and Strategic Management Theory:Review, Challenges and Pro-spects. *J. Manag. World* **2020**, *36*, 220–236. [CrossRef]
53. Li, J.; Zhu, X.Q.; Wang, L.X. Partners Selection in the Cooperative Innovation of Cluster Enterprises Based on Knowledge-associated Evaluation Index System and Improved ELMAN Neural Network Model. *Front. Sci. Technol. Eng. Manag.* **2020**, *39*, 23–29.
54. Hurst, E.; Lusardi, A. Liquidity Constraints, Household Wealth, and Entrepreneurship. *J. Politi Econ.* **2004**, *112*, 319–347. [CrossRef]
55. Gai, Q.E.; Zhu, X.; Shi, X.H. The Heterogeneous Effect of Wealth on Entrepreneurship: An Empirical Analysis Based on Farmers in Three Provinces. *J. Financ. Econ.* **2013**, *39*, 134–144. [CrossRef]
56. Gan, Y.; Zhu, J.; Liu, C.Y. The study on Impacts of Financial Constraints towards Household Entrepreneurship and its Differences between Urban and Rural Areas. *Shanghai J. Econ.* **2015**, *9*, 15–23. [CrossRef]
57. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3146–3154. [CrossRef]
58. Miller, T. Explanation in artificial intelligence: Insights from the social sciences. *Artif. Intell.* **2018**, *267*, 1–38. [CrossRef]
59. Murdoch, W.J.; Singh, C.; Kumbier, K.; Abbasi-Asl, R.; Yu, B. Definitions, methods, and applications in interpretable machine learning. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 22071–22080. [CrossRef]
60. Apley, D.W.; Zhu, J. Visualizing the effects of predictor variables in black box supervised learning models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2020**, *82*, 1059–1086. [CrossRef]