Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 11
TPCEE 2025

Research on Influencing Factors of Elderly Care Model Selection for the Silver - Haired Population Based on Traditional Statistical Analysis and Machine Learning Models

Zhaohe Gu [1, *], Kaiyi Lei [2], Mengyuan Xu [1]

[1] School of Materials Science and Engineering, Yanshan University, Qinghuangdao, China, 066004

[2] School of International Education, Hebei Finance University, Baoding, China, 071000

* Corresponding Author Email: 19312542985@163.com

Abstract. To address global population aging, this study innovatively integrates statistics and machine learning, analyzing multidimensional data, including economic status, health conditions, family support, and social environment. Logistic regression is employed to elucidate the influencing mechanisms of elderly care model choices, while decision tree and random forest models significantly boost the prediction accuracy to 85%. Key findings indicate that financial capabilities and family support are the core determinants, highlighting machine learning's superiority in identifying crucial influencing factors. This research provides a scientific foundation for optimizing the allocation of elderly care service resources and designing services. Emphasizing the potential of interdisciplinary research, it aims to fundamentally transform gerontology and create a supportive aging future. The developed model can be tailored to specific regional needs, enhancing its applicability and effectiveness, and it also promotes cross - cultural exchanges of elderly care experiences, thus contributing to building a society that enables the elderly to age with dignity and receive adequate support.

Keywords: Silver haired population; Elderly care model; Traditional statistical analysis; Machine learning.

1. Introduction

Global population aging has drawn extensive attention. With the improvement of medical standards and the enhancement of living quality, the proportion of the elderly population is increasing. This phenomenon has a profound impact on the social and economic structure and also poses higher requirements for elderly care services. The elderly care needs of the silver-haired population are becoming increasingly diverse.

The choice of elderly care in the silver economy has become a research hotspot. Zhang Wenjuan et al. [1] found that healthy elderly tend to choose home - based care, while disabled elderly prefer institutional care. Wu Haixia et al. [2] pointed out that parents with only one child have a stronger willingness for institutional care. Yan Shiqi et al. [3] emphasized that the ability of nursing staff plays a key role in the quality of elderly care services.

Deng Xin et al. [4] mentioned that there is a lack of cases of integrating medical and elderly care in rural elderly care. Yang Xinyue et al. [5]'s research in Zhenjiang has enriched regional elderly care data. MODEL O N A H U [6] explored the willingness to participate in mutual aid elderly care through "time bank". Zheng Xiuyun et al. [7] studied retirement willingness, expanding research perspectives.

Zhu Hailong et al. [8] found that the concept of "independent elderly care" has weakened the traditional model. Hai Jian et al. [9] looked forward to the future trends of elderly care. Jia Yan et al. [10] pointed out that there is a lack of long - term evaluation of innovative models. Chen Weitao [11] clarified the connotation of different elderly care models. Feng Tieying et al. [12] analyzed the practical path of integrating medical and elderly care. Existing studies have shortcomings, which point out the direction for future research.

Content from this work may be used under the terms of CC BY-NC 4.0 licence (https://creativecommons.org/licenses/by-nc/4.0/).
Published by Warwick Evans Publishing.

At present, in the research on elderly care choices, the influencing factors have been explored from multiple dimensions and various models have been involved. However, there are still problems such as a lack of rural cases, insufficient quantitative research on conceptual changes, and a lack of evaluation of innovative models. Our machine learning model performs excellently in classification tasks with an accuracy of 85%, marking a significant breakthrough in understanding complex elderly needs; By analyzing the importance of features and identifying key influencing factors, unprecedented insights have been provided for policy makers and service providers.

2. Research Methods

2.1. Machine Learning Models

At present, there has been exploration of the influencing factors of elderly care mode selection, which lays the variable foundation for constructing predictive models. The introduction of traditional statistical analysis and machine learning provides methodological breakthroughs for accurately predicting elderly care mode selection. The former is good at explaining causal relationships between variables, promoting theoretical construction, while the latter is good at handling large-scale data and revealing hidden patterns, making it suitable for prediction and classification tasks. Combining the two can offset the limitations of using a single method and produce richer and more accurate research results. This study integrates these two methods to explore the factors that affect the elderly care mode selection of the silver haired population, providing a scientific basis for optimizing the design of elderly care services and resource allocation. Meanwhile, the application of this comprehensive approach paves the way for future related research and promotes the development of gerontology and social welfare policies.

2.1.1. Decision Tree Model

The decision tree model is a supervised learning algorithm based on a tree structure. It performs attribute testing and partitioning on data through nodes and branches, starting from the root node and recursively constructing based on criteria such as information gain until the stopping condition is met. It can be used for classification, regression, decision support, and data exploration, and can intuitively display data patterns and decision logic, playing an important role in multiple fields. Decision tree model flowchart is shown in Figure 1.

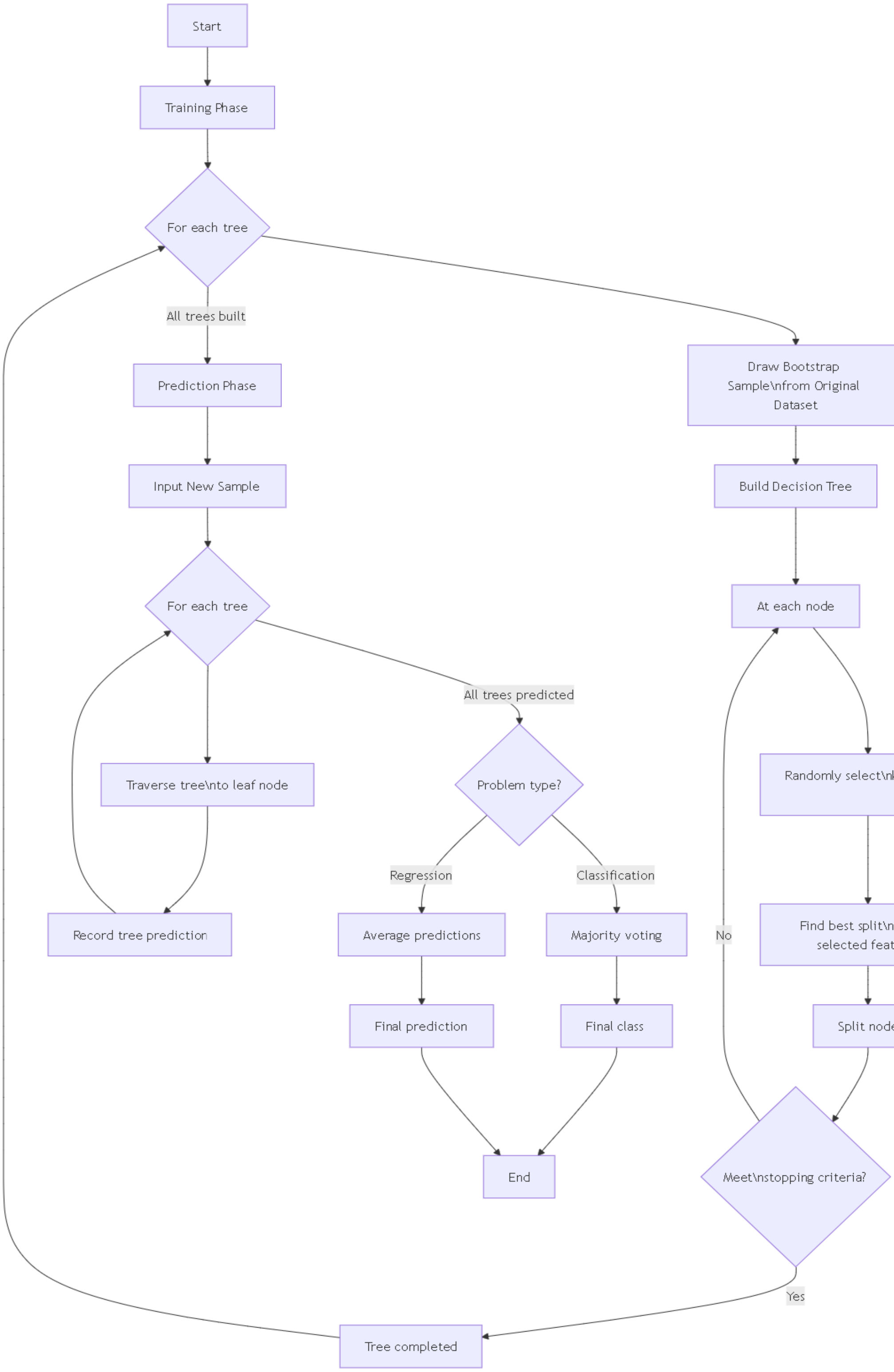

Figure 1. Decision tree model flowchart

A decision tree is a tree-based prediction model with three core components:

Decision trees operate through a hierarchy of nodes: a root containing all training data, internal nodes making feature-based decisions and branching using criteria like Gini or information gain, and leaf nodes providing predictions (class labels for classification, mean values for regression). The tree is built via recursive splitting, selecting the feature split that maximizes child node purity, until stopping criteria are met (pure nodes, max depth reached, insufficient samples). New samples are classified by traversing the tree based on feature judgments until a leaf node is reached, and the prediction associated with that leaf is output.

Taking the Gini index as an example, its calculation formula is:

$$Gini(D) = 1 - \sum_{i=1}^{n} p \tag{1}$$

where D is the dataset of the current node, and pi is the proportion of the i-th class sample in dataset D.

2.1.2. Random Forest

Random forest is an ensemble learning method composed of multiple decision trees, which improves the stability and accuracy of the model through ensemble learning. It randomly extracts samples from the dataset and constructs multiple decision trees, and the final result is voted based on the predicted results of all trees. Flow chart of Random Forest Model is shown in Figure 2.

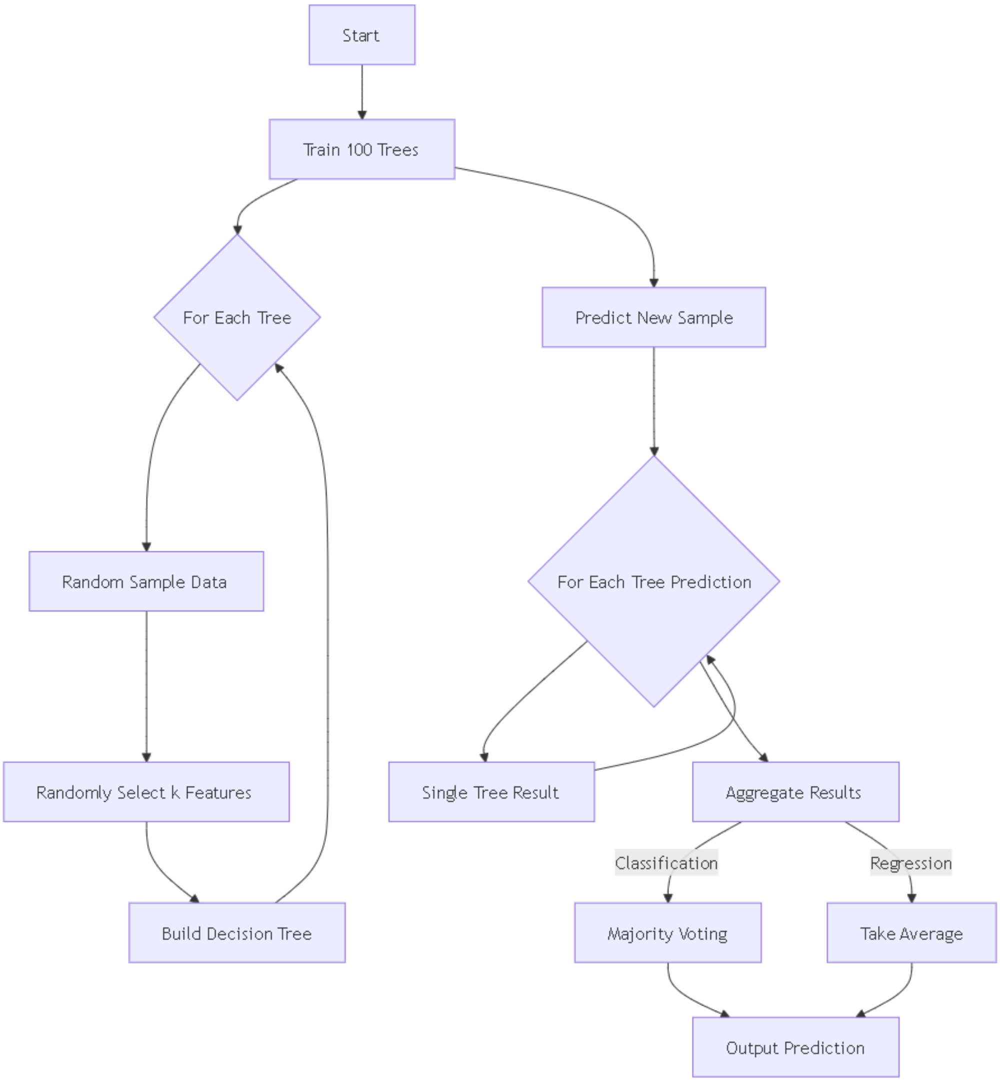

Figure 2. Flow chart of Random Forest Model

The key principles and mechanisms of random forest are as follows:

Random sampling: Multiple bootstrap samples are generated from the original data through replacement sampling, and each sample is used to train the decision tree. Each sample contains approximately 63.2% of the original data, with the remaining 36.8% being out of bag data that can be used for model validation.

Feature random selection: When splitting decision tree nodes, randomly select (where d is the total number of features) to form a subset, and choose the optimal feature split to improve model diversity and anti-interference ability.

Integration strategy: Classification tasks use voting to determine categories; The regression task calculates the output by averaging the predicted values of all decision trees.

The random forest model not only provides highly accurate classification results, but also allows us to identify which factors are most critical in determining the choice of elderly care mode through feature importance scores. Feature importance is calculated based on the contribution of each feature to the model's prediction accuracy, with higher scores indicating that the feature is more important. For example, in this study, economic status was found to be the most significant factor, followed by family support, health status, and social environment score.

2.2. Traditional Statistical Analysis Methods

In order to explore how various factors affect the choice of elderly care mode, this study adopted the traditional statistical analysis method of logistic regression. The logistic regression model is particularly suitable for binary or multi class problems, and can effectively evaluate the impact of independent variables (such as economic status, health status, etc. mentioned above) on the dependent variable (i.e. the choice of elderly care mode). By conducting logistic regression analysis on the sample data, not only can we determine which factors have a significant impact ($p<0.05$), but we can also quantify the strength and direction of each factor's impact, providing a theoretical basis for subsequent policy recommendations.

$$\log it(P(Y=1)) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n \qquad (2)$$

where β_0 is the intercept term, $\beta_1, \beta_2, \ldots, \beta_n$ are the coefficients of each independent variable, $X_1, X_2, \ldots, X_n$ are the actual values of each independent variable.

2.3. Data Sources and Variable Explanations

The data for this study is sourced from a special survey on the elderly population in a certain region, focusing on the living and elderly care needs of the silver haired population. The data covers four dimensions: economy, health, family, and society. Economic status includes indicators of economic ability such as income and savings; Health status includes physical and mental condition as well as self-care ability; Family support focuses on cohabitation and daily assistance; The social environment involves community service, transportation, and social participation. Multidimensional data analysis helps to comprehensively analyze the influencing factors of elderly care mode selection.

3. Empirical analysis

3.1. Descriptive Statistics.

Data shows that home-based elderly care is the preferred choice for the elderly due to emotional and environmental dependence. Economic strength and family support are key influencing factors. High income individuals not only pursue home comfort, but also are willing to pay for professional nursing care in institutions; When family support is sufficient, elderly people tend to prefer home-based elderly care.

3.2. Traditional statistical analysis results

Table 1. Correlation Coefficient

variable	Economic status	health status	family support	social environment score	gender	age	education level	marital status
Economic situation	1	-0.35	0.45	0.2	0.1	-0.2	0.25	0.15
health status	-0.35	1	-0.6	0.1	-0.05	0.15	-0.1	-0.1
Family support	0.45	-0.6	1	0.15	-0.02	-0.08	0.05	0.03
Social Environment Rating	0.2	0.1	0.15	1	0.08	-0.1	0.15	0.05
gender	0.1	-0.05	-0.02	0.08	1	-0.03	-0.02	0.02
age	-0.2	0.15	-0.08	-0.1	-0.03	1	-0.2	0.1
degree of education	0.25	-0.1	0.05	0.15	-0.02	-0.2	1	0.05
marital status	0.15	-0.1	0.03	0.05	0.02	0.1	0.05	1

Through the analysis of the correlation coefficient table in Table 1, it was found that there is a strong positive correlation between economic status and family support ($\mathrm{r} = 0.45$), and elderly people with better economic conditions are more likely to receive family support; There is a negative correlation between health status and family support ($r = -0.6$), and elderly people with poor health rely more on family care; Economic status is also negatively correlated with health status ($r = -0.35$), and elderly people with poor economic conditions face more health problems. In addition, the social environment score is weakly positively correlated with economic status ($\mathrm{r} = 0.2$) and educational level ($\mathrm{r} = 0.15$). These correlations reveal the factors that influence the elderly care mode choices of the silver haired population, providing a basis for formulating precise policies and service strategies.

To evaluate the impact of different factors on the choice of elderly care mode, we used a logistic regression model for analysis. Logistic regression is a statistical method widely used in classification problems, suitable for studying the relationship between binary or multi class dependent variables and one or more independent variables.

For the logistic regression model in this study, the specific formula is:

$$\mathrm{Logit}\,(\mathrm{P}\,(\text{home - based care})) = -1.20 + 0.75\,(\text{economic status}) - 0.45\,(\text{health status}) - 0.80\,(\text{family support}) + 0.20\,(\text{social environment score}).$$

Logistic regression analysis results is shown in Table 2.

Table 2. Logistic regression analysis results

independent variable	B	SE	Wald χ^2	df	p	Exp(B)
Economic situation	0.75	0.12	39.06	1	<0.001	2.12
health status	-0.45	0.10	20.25	1	<0.001	0.64
Family support	-0.80	0.15	28.44	1	<0.001	0.45
Social Environment Rating	0.20	0.08	6.25	1	0.01	1.22

These findings emphasize that economic ability and family support are key factors influencing the choice of elderly care models. Policy makers and service providers should fully consider these factors, optimize the design and resource allocation of elderly care services, in order to better meet the needs of the elderly and improve service quality.

3.3. Machine learning model results

When conducting in-depth analysis of the data, we constructed two machine learning models, decision tree and random forest, to optimize predictive performance and further understand the key factors affecting the elderly care mode selection of the silver haired population. Performance Comparison of Machine Learning Models is shown in Table 3.

Table 3. Performance Comparison of Machine Learning Models

Model name	Accuracy	Precision	Recall	F1 Score
Decision Tree Model	78%	0.76	0.79	0.77
Random Forest	85%	0.84	0.86	0.85

The random forest model performs better than the decision tree model in all indicators, especially with an accuracy of 85%, indicating that the model can more effectively predict which elderly people tend to choose based on input features.

The random forest model not only provides highly accurate classification results, but also allows us to identify which factors are most critical in determining the choice of elderly care mode through feature importance scores. Feature importance is calculated based on the contribution of each feature to the model's prediction accuracy, with higher scores indicating that the feature is more important.

Table 4. Ranking of Feature Importance in Random Forest Model

Feature Name	Feature Importance Score
Economic situation	0.35
Family support	0.25
health status	0.15
Social Environment Rating	0.10
Other factors	0.15

Ranking of Feature Importance in Random Forest Model is shown in Table 4. Economic status is the most significant factor, followed by family support, health status, and social environment score.

4. Conclusions and outlooks

This study combines traditional statistics and machine learning to explore the factors influencing the selection of elderly care models: logistic regression determines economic ability and family support as key factors, and the accuracy of the random forest model reaches 85%, with economic factors having the most significant impact. Comprehensive methods can improve prediction accuracy and provide suggestions for optimizing the allocation of elderly care resources. In the future, more factors and advanced algorithms can be explored.

While the study offers valuable insights into eldercare decision-making, its impact is constrained by regional data, technical oversights, and untapped variables. Future work should prioritize explainable AI, longitudinal designs, and culturally adaptive models to drive scalable, equitable eldercare solutions. Bridging these gaps will transform theoretical findings into actionable, real-world policies that truly resonate with the aging population's diverse needs.

References.

[1] Zhang Wenjuan, Wei Meng. Research on the willingness and influencing factors of urban elderly to receive institutional care [J]. Population and Economy, 2014 (6).

[2] Wu Haixia, Wu Fan Comparative Analysis of Elderly Care Intentions between Only Child and Multiple Child Families [J]. Population Journal, 2022, 44 (2): 85-98.

[3] Yan Shiqi, You Xin, Jin Xingyu. Research on the Capacity Enhancement of Elderly Care Workers and the Elderly Care Needs [J]. Nursing Science, 2025, 14: 439.

[4] Deng Xin, Qiao Hong. Overview of the Current Situation, Challenges, and Countermeasures of Rural Family Elderly Care Models in China [J]. Aging Research, 2025, 12:39.

[5] Yang Xinyue, Xu Xinglong. Analysis of Elderly Care Willingness and Its Influencing Factors in Zhenjiang City [J]. Journal of Nanjing Medical University (Social Sciences Edition), 2024 (1): 54-59.

[6] MODEL O N A H U. Analysis of the Willingness and Influencing Factors of Middle aged and Elderly People to Participate in Mutual Aid Pension through "Time Bank" Based on Anderson Model [J]. Social Science and Management Volume 8, Issue 2, 2024, July December, 2024, 8 (2): 49.

[7] Zheng Xiuyun, Feng Xiyun, Fan Xia. Research progress on the willingness of middle-aged and elderly people to retire and its influencing factors in China [J]. Journal of Yan'an University (Medical Science Edition), 2022, 20 (1): 81-84.

[8] Zhu Hailong, Ouyang Pan. The Transformation and Reflection on the Concept of Elderly Care [J]. Journal of Social Sciences, Hunan Normal University, 2015, 44 (1): 88-97.

[9] Haijian, Wenjie, Ran. Research on Future Elderly Care Models [J]. Management World, 2020, 36 (3): 89-103.

[10] Jia Yan, Lan Zhiyong, Liu Runze. Precision Elderly Care: A New Elderly Care Model Driven by Big Data [J]. Journal of Public Administration, 2020, 17 (2): 95-103.

[11] Chen Weitao. Harmony in Diversity ": A Comparative Study of the Concepts of Family based Elderly Care, Home based Elderly Care, Community based Elderly Care, and Institutional Elderly Care [J]. Guangxi Social Sciences, 2021, 9: 144-150.

[12] Feng Tieying, Nan Yan. Practice logic and path selection of the integrated medical and elderly care model: analysis based on typical cases of the national elderly care service industry [J]. Journal of Public Administration, 2020, 17 (3): 113-125.