Proceedings of the 5th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/99/20251730

Monitoring the Mental State of the Elderly in the Community Based on Machine Learning

Tongjia Hu[1,3,*], Chi-Han Lee[2,4]

[1]Shanghai University Of Engineering Science, ShangHai, China
[2]School of professional Study Appied Analytics, Columbia University In the City of New York, NYC, USA.

[3]jxhutongjia@163.com
[4]cl4069@columbia.edu
*corresponding author

Abstract. The rapid aging of the population has brought increasing attention to the mental health of elderly individuals. This study focuses on the mental well-being of elderly residents in several communities in Beijing. Through comprehensive surveys conducted in key communities with significant elderly populations, this paper gathered data to assess their mental health status. Subsequently, this paper employed various machine learning models, including Random Forest, Support Vector Machine, and XGBoost, to classify and predict mental health outcomes. The findings of this paper indicate that the Random Forest model outperformed the other models in accurately identifying mental health issues. Key predictors included factors such as whether the individual had received psychological counseling, age, social support, physical health status, and frequency of social interactions. This study provides empirical evidence that can guide healthcare providers and policymakers in developing more effective mental health management strategies for the elderly.

Keywords: mental health issues, machine learning, elderly individuals, community.

1. Introduction

In recent years, the rapid aging of the population has brought the mental health of elderly individuals into sharp focus. Mental health not only directly impacts the quality of life of the elderly but also indirectly influences the well-being of their families and communities. Although previous research has explored the relationship between aging, stress, and mental health, studies specifically addressing the unique challenges faced by the elderly population remain relatively scarce. Therefore, it is both theoretically significant and practically valuable to systematically investigate the mental health of elderly individuals and the factors that influence it.

This study aims to uncover the underlying factors that affect the mental health of elderly residents in Beijing communities by assessing their mental health status. The goal is to provide a scientific foundation for healthcare providers and policymakers to better monitor and enhance the mental well-being of elderly citizens. We conducted extensive data collection and analysis through questionnaires distributed in several major communities in Beijing. By utilizing community information networks and engaging directly with elderly residents, we applied multiple machine learning models to classify and

DOI: 10.54254/2755-2721/99/20251730

predict mental health issues. This approach enabled us to identify key characteristics influencing the mental health of the elderly and to propose targeted intervention strategies.

2. Literature review

The mental health of elderly individuals has garnered increasing attention alongside the rapid aging of the population. Numerous studies have demonstrated that aging-related stress and mental health challenges significantly affect the quality of life and overall well-being of the elderly. Additionally, imbalances between social engagement, health, and personal life have been identified as critical factors contributing to mental health problems among the elderly [1]. The unique challenges of aging, such as declining physical health and increased isolation, make this population a focal point for research into mental health issues. Previous studies suggest that the physical and emotional strains inherent in aging may exacerbate psychological stress. Besides health-related stress, other factors such as social support, financial security, and community engagement also play a role in influencing the mental health of elderly individuals [2]. These factors can vary significantly across different living environments and thus require context-specific analysis.

A considerable body of research has employed questionnaires and data analysis to assess the mental health of elderly individuals. For instance, the Social Support Scale developed by Xiao Shuiyuan is widely recognized for its reliability in evaluating stress and mental health in the elderly population [3]. In recent years, there has been a growing trend towards using machine learning models—such as Random Forest, Support Vector Machine, and others—to enhance the accuracy and effectiveness of predictions [4-6].

Regarding mental health interventions, studies have recommended improving community environments, offering psychological support, and enhancing social engagement as effective strategies to improve the mental well-being of the elderly. For example, establishing robust community support systems and reducing social isolation have been shown to significantly improve the mental health outcomes of elderly individuals [7].

3. Data and Models

3.1. Data Introduction

The dataset utilized in this study was derived from a 2023 survey on mental health among elderly residents in Beijing communities, conducted by the authors. The survey aimed to assess the impact of community environments on mental health perceptions and the prevalence of mental health disorders, while also identifying various factors that influence the mental health of elderly individuals. Data collection focused primarily on several major communities in Beijing with significant elderly populations. These communities were specifically targeted for distributing the questionnaire.

The data were gathered through questionnaires and encompass 23 features, including variables such as age, gender, living arrangements, health status, social support, financial security, and level of community engagement. The dependent variable was binary, indicating whether the respondent was experiencing mental health issues. Figure 1 illustrates the correlation coefficients between the feature vectors.

DOI: 10.54254/2755-2721/99/20251730

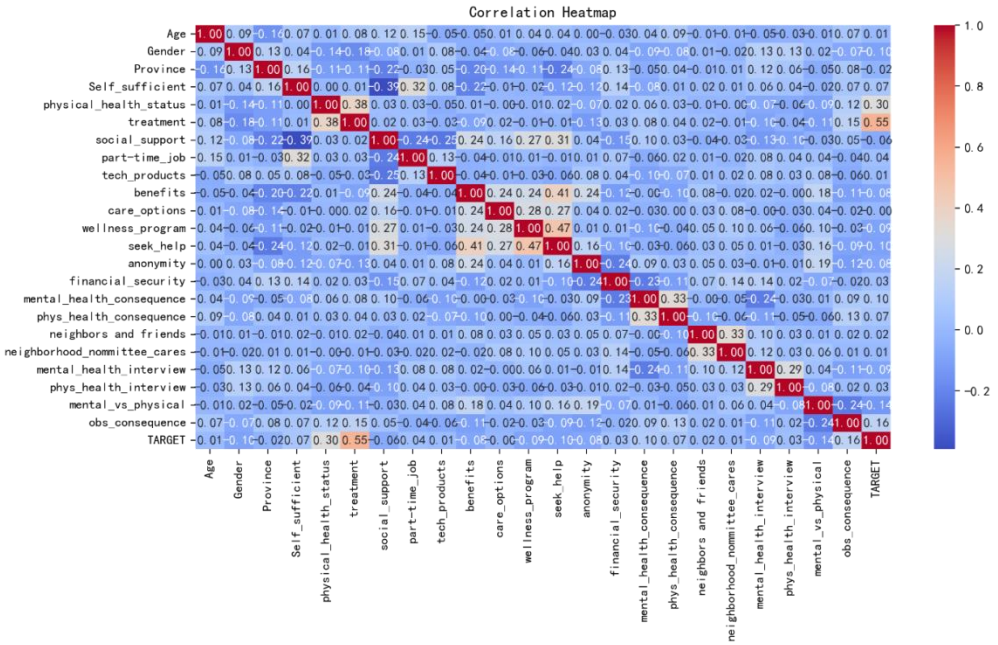

Figure 1. Correlation Heatmap

3.2. Model Introduction

A decision tree is a commonly used classification and regression model that partitions a dataset into subsets through a series of decision rules to achieve prediction of a target variable. Decision tree models are represented in a tree structure, in which each internal node represents a judgment condition for a feature, each branch represents a judgment result, and each leaf node represents the final prediction result. Common decision tree algorithms include ID3, C4.5 and CART (Classification and Regression Trees). The advantage of decision trees is that their model structure is simple and intuitive, easy to understand and interpret. However, decision trees are prone to overfitting the data, especially when the dataset is small or has a large number of features.

Support vector machine is a supervised learning model for classification and regression. The core idea is to separate different classes of samples by finding an optimal hyperplane in the feature space. SVM achieves optimal classification performance by maximizing the classification interval. When dealing with nonlinear problems, SVM uses kernel functions to map the data into a high-dimensional space, which allows the data to be linearly partitioned in that space. SVM has strong generalization ability, but the training time is longer on large-scale datasets.

The K-nearest neighbor algorithm is an instance-based learning method for classification and regression. The basic idea of the algorithm is that for a sample to be classified, the K neighboring samples with the smallest distance are selected by calculating their distances from all the samples in the training set, and the prediction is made based on the categories of these neighboring samples. In classification tasks, KNN usually uses the majority voting method to determine the prediction category; in regression tasks, KNN takes the average of the neighboring samples as the prediction result. The KNN model is simple and easy to implement, but the computational complexity is high, especially when the amount of data is large, and it is sensitive to noise and outliers.

Random Forest is an integrated learning method that performs classification or regression by constructing multiple decision trees and voting or averaging their predictions. Each decision tree is trained using Bootstrap sampling (self-sampling) and feature random selection, which makes each tree have a certain degree of variability, thus improving the overall performance and generalization ability of the model. Random Forest can effectively reduce the overfitting phenomenon and has better

DOI: 10.54254/2755-2721/99/20251730

robustness and higher prediction accuracy. It also provides valuable information for the evaluation of feature importance.

XGBoost is an efficient gradient boosting algorithm with strong classification and regression capabilities. The core idea is to train the model by incrementally building a series of weak learners (usually decision trees), each trained against the residuals of the previous round of the model. XGBoost introduces a regularization term on top of traditional gradient boosting in order to control the complexity of the model and avoid overfitting. It also leverages feature splitting best practices to significantly improve model training speed and performance. XGBoost has performed well in many machine learning competitions and is widely used in real-world problems.

GBDT is a decision tree integration model based on gradient boosting algorithm. The basic idea is to gradually improve the prediction accuracy of the model by constructing multiple decision trees, each of which learns and corrects the prediction error of the previous tree. GBDT updates the model parameters by minimizing the gradient of the loss function, so that each tree is trained on the residuals of the previous model. GBDT's strength lies in its flexibility and high efficiency, which enables it to handle a variety of data types and provide high prediction performance. Similar to XGBoost, GBDT is also robust but may take longer to compute when dealing with large-scale datasets.

4. Modeling and Analysis

This study employs a range of machine learning models, including Decision Trees, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Random Forests, XGBoost, and Gradient Boosted Decision Trees (GBDT), to classify and predict mental health outcomes among elderly individuals in Beijing communities. By comparing the performance metrics of these models, this paper identified the most effective approach for this dataset and determined the key features that significantly impact the mental health of elderly residents.

Table 1. Classifiers Accuracy Score

Model	precision	accuracy	recall	f1-score
RandomForest	0.77	0.76	0.76	0.76
DecisionTree	0.69	0.69	0.69	0.69
SVM	0.78	0.78	0.78	0.78
KNN	0.68	0.68	0.68	0.68
XGBoost	0.73	0.73	0.73	0.73
GBDT	0.76	0.76	0.76	0.76

The performance evaluation using the ROC and PR curves reveals variations in the models' effectiveness for classification tasks. The Area Under the ROC Curve (AUC) reflects the model's overall classification capability, while the PR curve focuses more on the accuracy of predicting the positive class, specifically mental health problems.

According to the ROC AUC values, the Random Forest model demonstrates the highest performance with an AUC of 0.81, indicating superior overall classification capability in detecting mental health issues among S&T workers. XGBoost and GBDT closely follow, with AUC values of 0.80 and 0.78, respectively, highlighting their effectiveness in handling complex data. Conversely, Decision Tree and KNN models perform relatively poorly, with AUC values of 0.69 and 0.72, suggesting these models struggle to capture intricate patterns within the data.

DOI: 10.54254/2755-2721/99/20251730

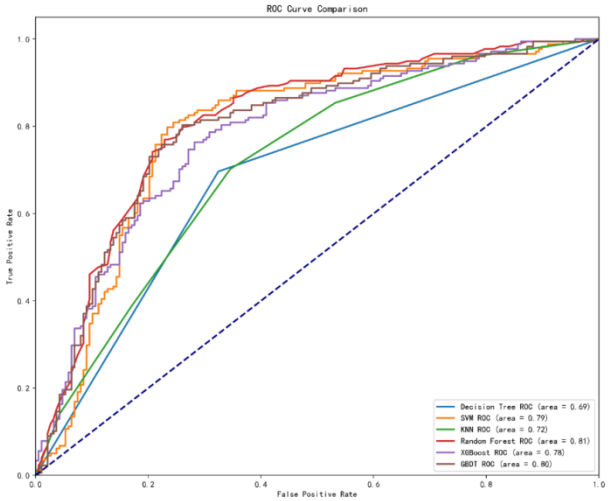

Figure 2. ROC Curve

In the PR curve analysis, despite the Decision Tree model's weaker performance in ROC curves, it achieves an AUC of 0.76 in PR curves, indicating its robustness in handling positive class data. This may stem from Decision Trees' advantages in dealing with imbalanced datasets. Random Forest and XGBoost also exhibit strong classification abilities, with PR AUC values of 0.75 and 0.74, respectively. In contrast, the SVM model has a PR AUC of 0.71, reflecting slightly lower performance, potentially due to challenges in managing the data's nonlinear relationships.

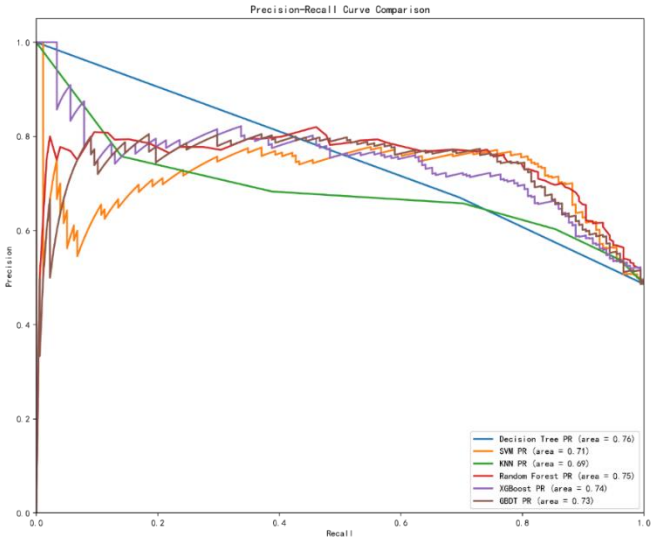

Figure 3. PR Curve

Further model analysis identifies key features influencing the mental health of the elderly. The Random Forest model, with its strong feature selection capabilities, reveals that whether an individual has received psychological counseling or treatment is the most influential factor. Other significant predictors include age, social support, financial security, physical health status, and the frequency of social interactions with family and friends.

DOI: 10.54254/2755-2721/99/20251730

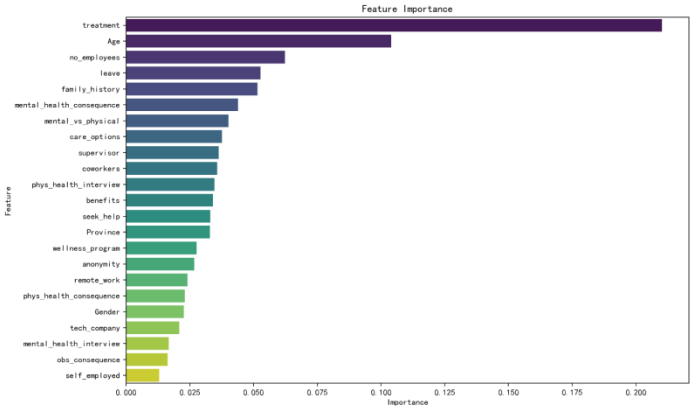

Figure 4. Feature Importance

5. Conclusion and Future Directions

The Random Forest and XGBoost models demonstrated superior performance in predicting mental health issues among elderly individuals, particularly in identifying those at risk. This suggests that more complex machine learning models are better equipped to capture the underlying patterns and relationships in high-dimensional data. Key features influencing mental health include psychological counseling and treatment history, age, social support, physical health status, financial security, and frequency of social interactions with family and friends. However, while some models excel in specific areas—such as Decision Trees on PR curves—their overall performance is still limited by the complexity of the data. Therefore, practical applications should consider integrating multiple models to achieve more accurate and robust predictions.

In addition to employing complex machine learning models like Random Forest and XGBoost to predict mental health issues, future efforts should focus on creating accessible and proactive mental health detection and management systems for elderly individuals. These systems could integrate wearable health monitoring devices, community-based mental health screenings, and online platforms that allow for continuous psychological evaluation. By leveraging real-time data from multiple sources, healthcare providers can more effectively identify early signs of mental health decline and intervene promptly.

References

[1] Maslach, C., Schaufeli, W. B., & Leiter, M. P. (2001). "Job burnout." Annual Review of Psychology, 52, 397-422.

[2] Halbesleben, J. R., & Buckley, M. R. (2014). "Burnout in organizational life." Journal of Management, 35(3), 837-865.

[3] Xiao Shuiyuan. Theoretical basis and research application of the Social Support Rating Scale[J]. Journal of Clinical Psychiatry,1994,(02):98-100.

[4] Breiman, L. (2001). "Random forests." Machine Learning, 45(1), 5-32.

[5] Cortes, C., & Vapnik, V. (1995). "Support-vector networks." Machine Learning, 20(3), 273-297.

[6] Chen, T., & Guestrin, C. (2016). "XGBoost: A scalable tree boosting system." Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785- 794.

[7] Bakker, A. B., & Demerouti, E. (2007). "The job demands-resources model: state of the art." Journal of Managerial Psychology, 22(3), 309-328.