A Prediction of Chinese Elderlies' Mental Health Based on CHARLS

Jingru Li; Siyuan Liu; Suyu Liu

Abstract:

As the trend of aging surge how to overcome the psychological problems faced by the elderly becomes more and more of a challenge. Using a sample of approximately 20,000 Chinese seniors, this report attempts to predict the degree of psychological health of the elderly through numerous factors. In this report, we first build a linear model and use methods such as Random Forest, Tree, and Gradient Boosting to make predictions. Finally, it reveals that age, satisfaction with life, sleep time, and education can help us to predict the mental health of the elderly. And, by further classifying the samples, we also find that the mental health of elderlies of different genders and different living backgrounds is influenced by similar factors but in the different level.

1. Introduction

The aging population is increasing rapidly. According to UN, the proportion of older adults is estimated to be 22% of the world's total population. By that time, one quarter of the population in Europe and North America will be 65 years old or older. Physical and mental challenges that they are facing should be recognized. Shocking numbers of adults aged 60 and over suffer from a mental disorder --- approximately 15%.

There is no doubt that psychological problems in older adults will affect their physical health and thus strain the clinical and social infrastructure that supports them. The health effects of psychological problems are not limited to older adults but can also threaten the health of family members and caregivers who spend significant time and energy caring for these older adults. At the same time, psychological problems in the aging population can have additional hidden costs to the economy and society. For example, caregivers are overwhelmed at both work and home, community activities lose the participation of older adults, and the government tries to control the budget for the care of the elderly. These issues create a public health care dilemma that is likely to deteriorate as the surging aging population.

It's these issues that reflect the importance of focusing on the mental health of the elderly. Understanding the factors that cause mental health problems in older adults is critical to developing effective interventions and policies. To focus on them is to focus on our own future. That's also where the motivation for this report comes from.

To better capture reality and compare samples, we focused on China, where the rate of aging has soared in recent years. The purpose of this report is to use data from the CHARLS (China Health and Retirement Longitudinal Study), a nationally representative longitudinal survey of Chinese adults aged 45 and older, to predict the mental health status of older adults in China. The survey collects data on a range of health, social, economic, and demographic factors. We will examine the effects of various factors, including age, gender, education, income, social support, and health status, on mental health outcomes among the elderly. By identifying factors that affect seniors' mental health, we can develop targeted interventions to improve their mental health and well-being and contribute to healthy aging.

2. Method

2.1 Dataset

The data used in this analysis is cross-sectional data from the China Health and Retirement Longitudinal Study (CHARLS) the national survey of wave four in 2018. CHARLS is a longitudinal survey that aims to be representative of the residents in China aged 45 and older, with no upper age limit. CHARLS is harmonized with leading international research studies in the Health and Retirement Study (HRS) model, which collects information about income, work, assets, pension plans, health insurance, etc. This contains data on 18117 elders.

For our variable to predict, in CHARLS they asked 10 questions about mental health, so we use these 10 variables to estimate mental health. The range of the answer is from "1 rarely or none", "2 some or a little", "3 occasionally or moderate amount of time", and "4 most of the time". To measure mental health, we used the function below:

$$Mental\_health = 28 - bothered\_by\_things - had\_trouble\_keeping\_mind - depressed - fearful - sleep\_restless - lonely - not\_get\_on + effort + hopeful + happy)$$

in this equation, we set 28 as the benchmark for judging, a high score in mental health means elders have better mental health.

We choose variables that may have an impact on the mental health of older people from the database by hand. Here is a list of the variables:

- Retire: if people have retired
- Life_sf: life satisfaction
- Health_sf: health satisfaction
- Marriage_sf: marriage satisfaction
- Children_sf: children's satisfaction
- Air_sf: air quality satisfaction
- Residential_address: respondents live in urban or rural
- Education: The education level of the respondents
- Marital_satus: Living together or living alone
- Self_health: self-report health
- Sleeptime: the length of sleep time each day
- Intensivesport: if respondents do intensive sport more than 10 minutes each day
- Moderatesport: if respondents do moderatesport sport more than 10 minutes each day
- Friendscom: Whether to interact with friends
- Activity: whether to attend activity monthly
- Smoke: still smoke or not
- Health_problem: whether suffering from any of the twelve chronic diseases[1]
- Disability Whether disabled or not
- Pensiontype: the type of pension
- Pension: the amount of pension
- Age: respondents' age in 2018

2.2 Method

Our goal is to build the best predictive model of older people's mental health, to verify whether each variable could have a significant effect on mental health, we first ran OLS regressions

---

[1] Hypertension, Dyslipidemia, Diabetes, Cancer, Chronic Lung Diseases, Liver Disease, Heart Attack, Stroke, Kidney disease, Stomach Disease, Emotional Problems, Memory-Related Disease, Arthritis, Asthma.

on mental health, the result will show the p-value of each variable, then we will select significant variables from which to build our predictive models.

We choose six methods of building predictive models, which are linear model, stepwise selection, single tree model, single tree pruned model, random forest, and Gradient boosting.

(1) Linear model: Linear modeling is the most widely used tool in the world for fitting a predictive model of the form:

$$Y = f(x) + e$$

They are used throughout the worlds of science and industry, and they can directly show the relationship between variables and predicted values.

(2) Stepwise selection: Stepwise selection model will do OLS regression on both candidate variables and interactions (the scope). Start with any working model containing some subset of these variables. Ideally, this should be a reasonable guess at a good model.

(3) Single tree model: Regression trees are for numerical (as opposed to categorical) outcomes. We can estimate E(y|x) by this model:

y = mental health of elder; x = significant effect variables

(4) Single tree pruned model: To improve the single tree model, single tree pruned model examines every pair of "sibling" leaf nodes and checks the increase in loss from "pruning" that split; Prune the "least useful" split, for example, the prune that yields the smallest increase in loss (decrease in fit). It might be able to provide us with better predictions than the single tree model.

(5) Random forest model: “random forest model” starts from bagging, it combines the output of multiple decision trees to reach a single result. Its ease of use and flexibility have fueled its adoption, as it handles regression problems. We are able to find out the importance of each variable and its partial effect on the outcome.

(6) Gradient boosting is a machine learning technique used in regression, among others. It gives a prediction model in the form of an ensemble of weak prediction models, which are typically decision trees.

Our evaluation criteria for these six models are Root-mean-square deviation. It measures the difference between the predicted values and the actual values in a dataset, lower RMSE indicates a better prediction model.

After choosing the best prediction model for mental health, we will do robustness tests on the model, by dividing the model into four groups with gender and region which are, urban male, urban female, rural male, and rural female. Then using these groups to explore predictive models for different types of older people's mental health by the prediction model we have chosen.

3. Result

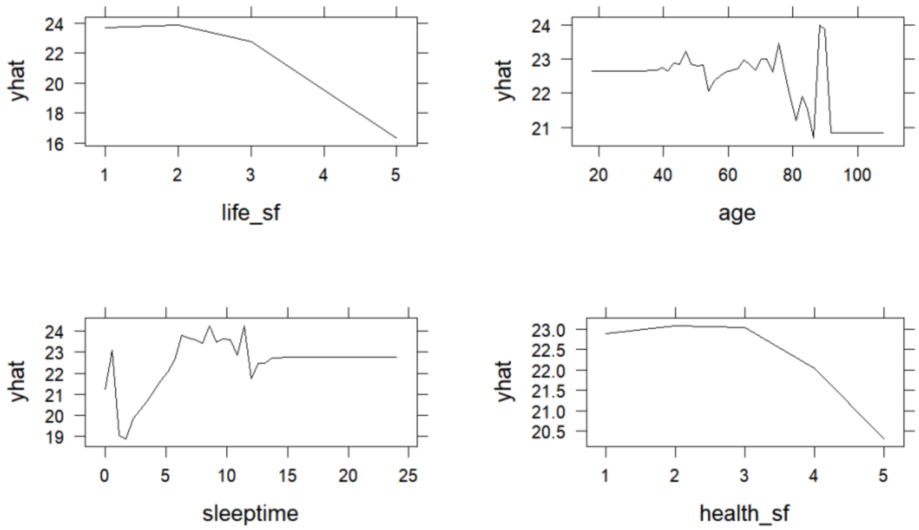
We used a variety of models to predict mental health outcomes in older adults. And here is the RMSE for these models：

	RMSE
Linear Regression	5.648879
Stepwise	5.582323
Single Tree	5.749408
Single Tree Pruned	5.844889
Random Forest	5.632340
Gradient Boosting	5.606734

These models are generally all good at predicting the mental health outcomes. And we can see that Gradient Boosting has the smallest RMSE. So, we choose Gradient Boosting model to have a further look at our case. While we cannot explain the meaning of the interaction term of stepwise, so we don't consider this model in the further study.

In Gradient Boosting models, relative influence is a measure of variable importance that reflects the contribution of each input variable to the model's overall predictive performance (Appendix. Table 1).

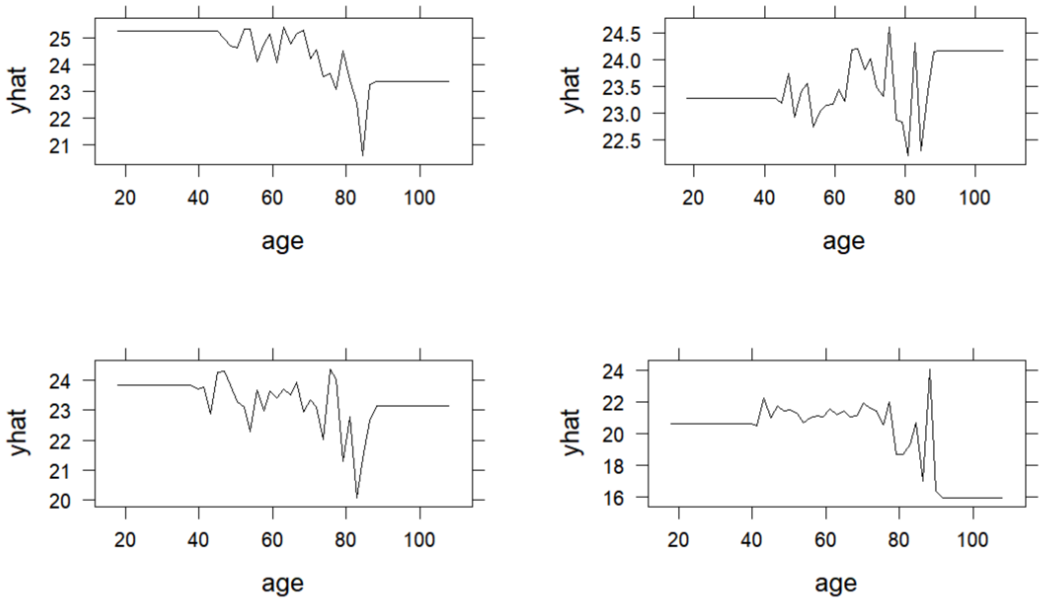
Life satisfaction, age, sleep time and self_health are the Top 4 important variables. So, we have a look at their partial dependence plot:



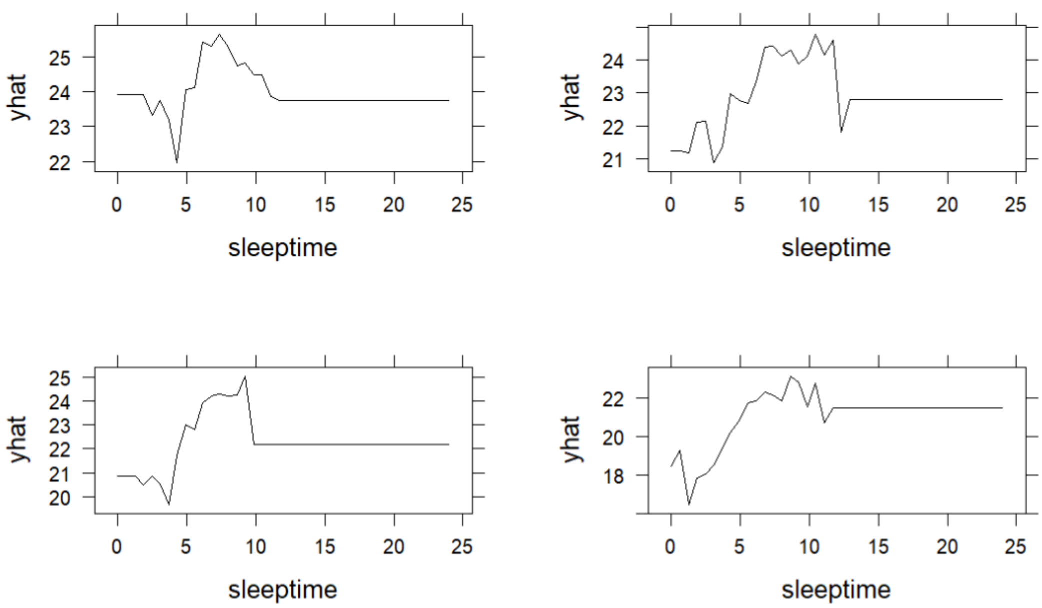
Mental health increase as life satisfaction and health satisfaction increases (5 is the worst rating and 1 is the best rating). When the sleep time is between 6-10h, their mental health is better. For age, there is no obvious and significant trend.

Next, we classified the sample by residential address (urban or rural) and gender, and divided them into four categories: urban male, urban female, rural male, rural female.

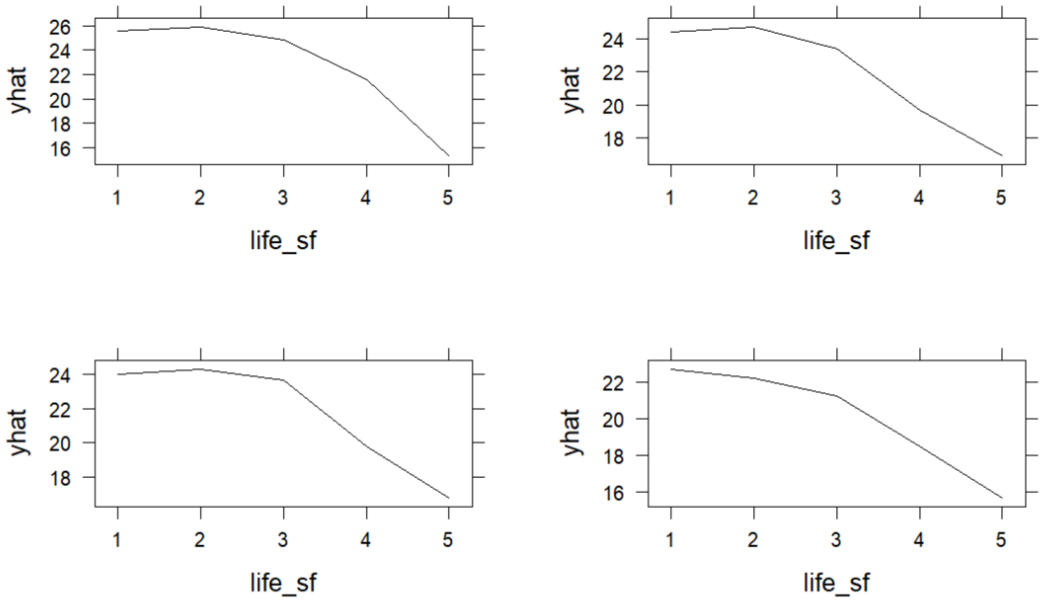
After running the gradient boosting model on each subset and get the relative influence table (Appendix. Table 2-5), age, sleeptime, life_sf is the top 3 important for these four subset, similar to the overall sample.Comparing the education for female, education is more important for urban female than rural female. Smoke is important for male while not for female. Then we have a look at the Top3 partial dependence plots:



Note. The partial dependence plots of age for the urban male, rural male, urban female, and rural female.



Note. The partial dependence plots of sleep time for the urban male, rural male, urban female, and rural female.



Note. The partial dependence plots of life satisfaction for the urban male, rural male, urban female, and rural female.

Based on the plots, there is no obvious and significant trend. Because the samples of older adults below40 and above 80 ages are too few, we can focus on the average mental health between 40-80 age. It is clear that average mental health of rural female is the worst. For the sleep time, when sleeping for 6-10 hours per day, the mental health of the elderly is better. For life satisfaction, there is no significant difference in their mental health with 1 and 2 points. As life satisfaction goes down from 3 (1 is the best and 5 is the worst), their mental health goes down.

4. Conclusion

Based on the above prediction model, we find that:

(1) Satisfaction with life, marriage, own health, and children all influence the psychological status of the elderly. When they are satisfied with their current life, they are also psychologically healthier. Although age significantly affects the prediction of mental health, we cannot directly determine the level of mental health of older adults by their age. Sleep duration within a certain time range affects mental health, but more than 6 hours of sleep is not helpful in determining psychological condition.

(2) By further classifying urban and rural, male and female, we observe that there are different performances of the influencing factors for different groups. For urban senior citizens, physical activity plays an important role in maintaining mental health, but rural aged people do not rely on sports to keep their emotions. This is also related to the fact that there is a gap between urban and rural infrastructure development in China at present. Urban communities generally build activity centers for social interaction and engagement for the elderly, but rural areas lack such infrastructure.

Education has the most significant impact on the mental health of urban women. This is also related to the limited social roles and expectations of women compared to men. In particular, older urban women experience both traditional gender roles, such as caregiving and family responsibilities, but also access to educational opportunities and skills. Thus, education has a particularly strong impact on them.

To sum up, we predict the mental health of the elderly in terms of marriage, children, education, gender, and life background, and finally select the Gradient Boosting model with the smallest RMSE. In addition to this, the samples are further classified and analyzed. The mental health of the elderly is affected by various factors. The conclusions of this report also provide a certain reference for the psychological evaluation of the psychological status of the elderly in the future.

Appendix:

Table 1

	var <chr>	rel.inf <dbl>
life_sf	life_sf	18.7905292
age	age	16.5823790
sleeptime	sleeptime	14.1501972
self_health	self_health	8.8449029
health_sf	health_sf	8.4919051
education	education	6.3282341
marrage_sf	marrage_sf	6.1548244
children_sf	children_sf	4.0463935
marital_status	marital_status	2.0877702
smoke	smoke	1.8856271

Table 2: male_urban

	var <chr>	rel.inf <dbl>
age	age	21.4871539
life_sf	life_sf	13.8857347
sleeptime	sleeptime	12.1430689
self_health	self_health	8.2094333
education	education	6.4311199
marrage_sf	marrage_sf	5.5593858
smoke	smoke	5.4356879
health_sf	health_sf	4.9359833
children_sf	children_sf	4.2796708
activity	activity	2.8272648

Table 3: male_rural

	var <chr>	rel.inf <dbl>
age	age	22.1895600
sleeptime	sleeptime	13.8114447
life_sf	life_sf	13.3437556
health_sf	health_sf	7.1654669
education	education	6.8815367
self_health	self_health	6.5613573
marrage_sf	marrage_sf	5.7914676
children_sf	children_sf	4.6352664
smoke	smoke	3.8568206
marital_status	marital_status	2.6754453

Table 4: female_urban

	var <chr>	rel.inf <dbl>
age	age	30.2379520
sleeptime	sleeptime	11.9100629
life_sf	life_sf	10.0978224
education	education	7.0426990
health_sf	health_sf	6.8975453
marrage_sf	marrage_sf	5.6709633
self_health	self_health	5.4192596
children_sf	children_sf	4.3793548
marital_status	marital_status	3.6421893
activity	activity	2.2403619

Table 5: Female_rural

	var <chr>	rel.inf <dbl>
age	age	22.2460272
sleeptime	sleeptime	15.7054076
life_sf	life_sf	14.2324345
self_health	self_health	8.7852028
health_sf	health_sf	8.6254490
marrage_sf	marrage_sf	6.1907929
children_sf	children_sf	4.6389163
education	education	3.8309682
health_problem	health_problem	2.2268463
marital_status	marital_status	2.0095184