Analyzing the Retirement Satisfaction Predictors among Men and Women Using a Multi-Layer Feed Forward Neural Network and Decision Trees

Ehsan Ardjmand, Gary R. Weckman, Diana Schwerha
Department of Industrial and Systems Engineering
Ohio University
Athens, Ohio USA
Email: ardjmand@ohio.edu, weckmang@ohio.edu, schwerha@ohio.edu

Andrew P. Snow
School of Information and Telecommunication Systems
Ohio University
Athens, Ohio USA
Email: snowa@ohio.edu

***Abstract*— In this article, we will analyze the effect of different retirement satisfaction predictors on each other and the retirement satisfaction level among men and women. The following factors will be used as predicators of retirement satisfaction: health; wealth; smoking and drinking habits; education; faith; income; impact of health on activities of daily living (ADL); frequency of activities; and the number of people in a household. A set of 858 retired men and 1179 retired women from a 2012 Health and Retirement Study database have been chosen and analyzed. A neural network was trained for each gender in order to predict retirement satisfaction; it also generated a decision tree that symbolizes the retirement satisfaction and its predictors. The results demonstrate that health, age, smoking habits, income, and wealth are the most significant predictors for both genders, while for men, education also plays an important role in retirement satisfaction.**

Keywords- Retirement Satisfaction; Artificial Neural Networks; Multi-Layer Perceptron; Decision Tree

I. INTRODUCTION

As the population of retired people is growing, retirement satisfaction has become a significant issue in aging and retirement research. It is predicted that around 24 percent of the United States' work force in 2018 will be at least 55 years old [1]. In addition to positive changes in lifestyle, retirement—as a major alteration in life for the elderly—can be the source of many negative experiences, such as loneliness, anxiety, and sometimes even psychological disorders [2].

There is a large body of research on factors which may have an effect on retirement satisfaction—among which health and wealth, as the two most important predictors, have been shown to have a positive correlation with this kind of satisfaction [3-8]. A positive psychological condition is also shown to have a positive correlation with retirement satisfaction [6].

Sexuality is also another analyzed factor in literature. Although there are many studies focusing only on men or women in terms of retirement satisfaction, the studies show that there is no significant difference among men and women in this category [6, 9-15].

Voluntary retirement, engagement in social activities, higher educational level, and having a spousal partner also can have a positive effect on retirement satisfaction [8, 12, 15-21].

Although the retirement satisfaction factors have been analyzed extensively in literature, the inter-relational effect of these factors remains an unchallenged problem. For example, we know that wealth and health have a positive correlation with retirement satisfaction [5], but how will a high level of wealth and a low level of health affect retirement satisfaction simultaneously? Additionally, what level of each factor is the threshold at which retirement satisfaction may be altered?

In this paper, using the data of 858 retired men and 1179 retired women from the 2012 Health and Retirement Study database, we predict the retirement satisfaction level as a dependent variable and the health, wealth, smoking and drinking habits, education, faith, income, impact of health on instrumental and regular ADLs, frequency of activities, and number of people in a household as independent variables by using a multi-layer perceptron neural network. We then try to illustrate the effect of different levels of independent variables on retirement satisfaction simultaneously by using a decision tree for both men and women.

In Section 2, we explain the method and data we use for analysis. In Section 3, the results of analyzing retirement satisfaction as an outcome of predictor variables are presented for both men and women. In Section 4, the overall conclusion is stated.

II. DATA AND METHODOLOGY

A. Health and Retirement Study

The data for this research came from the 2012 Health and Retirement Study (HRS), which was launched in 1992. The total number of randomly considered retired people chosen from HRS for this study was 2037, which consisted of 858 men and 1179 women. Notice that only the respondents with no missing values in both dependent and independent variables were considered in this study.

The dependent variable is considered to be retirement satisfaction. If a person is reported to be retired in 2012 he/she is asked the G136 question, “All in all, would you say that your retirement has turned out to be very satisfying, moderately satisfying, or not at all satisfying?" The answer to this question is supposed to capture the retirement satisfaction level for retirees.

The independent variables in this research are the age (in months); years of education; belief in a higher power; self-report of health (based on a 5-point scale in which 1 shows

Copyright (c) IARIA, 2016. ISBN: 978-1-61208-457-2

excellent health and 5 shows very poor health); a binary variable which shows if the health limits the ability to work or not; level of difficulty in pursuing the ADLs (based on a 6-point scale in which 0 shows no difficulty and 5 shows someone is unable to perform ADL); mental health (based on a 9-point scale in which 0 is excellent and 8 is very poor); a set of binary variables that show if the person has blood pressure, diabetes, cancer, lung disease, heart problem and/or arthritis; frequency of vigorous, moderate, and light activity; a binary variable that shows if the person smokes or not; the number of alcoholic drinks consumed per week; wealth; income; and the number of people living in a household.

B. Methodology

In this research for modeling retirement satisfaction and other independent variables, we use a multi-layer feed forward neural network. For illustrating this relationship in a symbolic structure, we will use a decision tree technique proposed by Craven [22] and modified by Young [23].

1) Artificial Neural Networks (ANN)

ANNs are mathematical models that mimic the human brain. Besides being considered a "black-box" model, ANNs also have the limitation of requiring a large amount of training and cross-validation data, i.e., typically three times more training samples than network weights [24]. However, a systematic way of modelling complex non-linear patterns is proposed [25]. Since their resurgence in the 1980s, ANNs have been applied to a variety of problem domains such as speech recognition [26] and generation [27], symbolic learning [28], robotic design [29], medical diagnostics [30], game playing [31], healthcare systems [32], bankruptcy [33], credit cards [34], and estimation of functions as in forecasting [35-37]. Theoretically, it is possible to prove that a three-layered NN can estimate the value of a function with desirable accuracy [38, 39]. Since the relationship of retirement satisfaction and other independent variables is not necessarily linear and can be considered highly complex, feed forward neural networks can be a useful tool for predicting the value of retirement satisfaction.

There are many types of ANN topologies that have been comprehensively documented [40], and they range in their use and complexity. One of the most widely used neural networks (NN) is the feed forward neural network (FNN). For example, Figure 1 shows the general structure of a FNN. The network shown is fully connected, since each layer is connected via previous layers. The first hidden layer's neurons are connected to the second hidden layer's, and the second hidden layer's neurons are connected with all of the output layer's neurons.

There are two main paradigms of ANN training--supervised and unsupervised learning. The primary difference between the two learning schemes is that in supervised learning, known outputs, or--"targets"--are used to adjust the network's weights. In unsupervised learning, there is not a known output, and the method functions as a clustering algorithm.

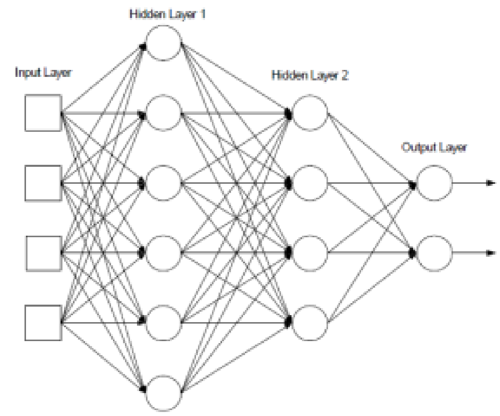

Figure 1. Feed Forward Neural Network [41].

2) Decision trees and TREPAN

One of the main drawbacks of neural networks is the lack of explanation capability [42]. In order to represent the knowledge about retirement satisfaction learned by a neural network, we use decision trees. Decision trees classify data through recursive partitioning of the dataset into mutually exclusive subsets [43], which best explain the variation in the dependent variable under observation [44]. A decision tree model consists of logical tests, which result in possible classifying consequences. Decision trees have been used to aid decision makers in many real-world problems. For example, Leech [45] applied a decision tree to a chemical nuclear power plant process involving continuous feedback systems. Another use led a manufacturing company to reduce inventory levels and improve processing efficiencies, which saved the company ten million dollars in operation expenses a year [46].

TREPAN is a novel rule-extraction algorithm [47] that utilizes the behavior of a trained ANN. Given a trained ANN, TREPAN extracts decision trees that provide a close approximation to the function represented by the network when there are issues of accurately calculating tree partitions, which are caused by limited sample sizes. TREPAN uses a concept of recursive partitioning similar to other decision tree algorithms; however, in contrast to the depth-first growth used by other decision tree algorithms, TREPAN expands using the "best first" principle. For conventional induction algorithms, the amount of training data decreases as a decision tree grows. Thus, there is less data at the bottom of the tree able to determine class labels accurately. In contrast, TREPAN uses an "oracle" to answer queries that determine decision tree splits better when sample instances are limited. One important aspect of this feature is the user-determined parameter called minimum sample. TREPAN ensures that splits are determined with a minimum number of sample instances. If the number of instances at a particular node, *m*, is less than the minimum sample allowed, TREPAN will make membership queries equal to the minimum sample from the ANN oracle in order to artificially create sample instances to meet the minimum sample requirement.

TREPAN uses an entropy-based criterion called "information gain" to determine the best position in which to

ISBN: 978-1-61208-457-2

partition the dataset. TREPAN uses M-of-N expressions as it splits upon the dataset. In this case, N rules are created. The algorithm also determines a value for M, which represents the minimum conditions that must be met, which in turn dictates the preceding node or final classification. This approach allows multiple features to be present in one node. To prevent testing of all the possible M-of-N combinations, TREPAN makes use of the heuristic "beam search" process. This process begins by selecting the best binary split at a given node based upon information gain. Additional splitting conditions are determined based on the initial rule's "complement" [48].

When sample instances are sparse, TREPAN interacts with an ANN oracle by means of membership queries. The goal of a membership query is to determine a new instance among a group of instances. To create appropriate sample instances, distributions of attribute values are created that conform to the decision tree constraints [49]. Once the ranges are determined, random pulls are made from the attributes' distribution in order for the oracle to accurately estimate the classification output label.

3) Retirement Satisfaction Model

For this study, we train a feed forward neural network with two hidden layers. There are 15 processing units in the first layer and 10 processing units in the second hidden layer, as well as tangent hyperbolic and linear transfer functions for the hidden and output layers, respectively, that use back propagation algorithms in NeuroSolutions 6.20 software. The output of the network, i.e., retirement satisfaction – is a continuous number. In order to convert the output of the network into the categorical scale of retirement satisfaction, we divide the output into three categories of $(-\infty,1.66]$, $(1.66,2.33]$, and $(2.33,+\infty]$, which are equivalent to not satisfying, moderately satisfying, and very satisfying. Notice that in the data we use the numbers 1, 2, and 3 to represent satisfying, moderately satisfying, and very satisfying, respectively.

III. RESULTS

Figure 2 and Figure 3 show the decision tree obtained for men and women regarding the relationships of the independent variables and retirement satisfaction. Notice that every rectangular shape in the decision tree shows a condition that, if met, the right branch should be followed. The left branch is for the case in which the condition is rejected. The oval shapes show the consecutive retirement satisfaction level in each branch.

As it is depicted in Figures 2 and 3, not all of the variables are involved in predicting retirement satisfaction. The reason is partially because of the low correlation of some independent variables and retirement satisfaction, as well as the overwhelming impact of these important variables on the latter that makes the other factors neutral. Another reason is the structure of the decision tree itself. By generating a decision tree, we are trying to extract the knowledge of the neural network, and the generated tree is formed in a way to represent the most possible knowledge in the form of rules according to the neural network, which can cause us to ignore some of the inputs.

A. Comparison with Literature

All of the extracted rules in decision trees are consistent with the results in literature. Age has a positive correlation with retirement satisfaction [3]. This effect can be seen by following branches that point to older ages and comparing them to the other branches in Figures 2 and 3. High levels of mental and physical health correspond to higher retirement satisfaction [3, 4, 6-8]. Higher levels of wealth and income also correspond to higher retirement satisfaction [3, 5-7]. Years of education have a positive correlation with retirement satisfaction [20].

B. New Findings

In addition to the result comparisons to previous literature, some new patterns can be deduced from the decision tree. Compared to women, the years spent in education for men is an important factor. In Figure 2, one of the parameters that affects the retirement satisfaction in men is education level. However, in Figure 3 the education level is not a condition in defining the retirement satisfaction, which shows that for women, it is not an important parameter.

Since for men the wealth appears in higher levels of the decision tree, it follows that, compared to women, wealth for men is a more important factor. Following the same logic, we can see that compared to men, mental health is a stronger predictor for women. In addition, for women with poor health, wealth is not a predictor at all. Despite this, for men with poor overall health, age cannot predict the retirement satisfaction.

Among all the health conditions analyzed, only diabetes plays a significant role in explaining retirement satisfaction. In both decision trees, i.e., men's and women's – having diabetes can cause lower retirement satisfaction, except where the income level is rather high. Although poor conditions of physical and mental health for both men and women can cause low retirement satisfaction, a high amount of wealth and income can ameliorate this situation.

IV. CONCLUSION

In this paper, using the 2012 data of the Health and Retirement Study for 858 retired men and 1179 retired women, we trained a feed forward neural network to predict the retirement satisfaction, considering health, wealth, smoking and drinking habits, education, faith, income, impact of health on ADLs, frequency of activities, and the number of people in a household as independent variables. The knowledge of neural networks was represented in the form of a decision tree.

The results show a very high consistency with previous findings in literature. Additionally, some new knowledge regarding retirement satisfaction was also revealed in the form of rules in the decision tree. It was shown that, compared to men, years of education is more important to women in regards to retirement satisfaction. Under the condition of poor health, age is an important predictor of retirement satisfaction for women. Among all the health-related diseases, diabetes plays the most important role in terms of predicting retirement satisfaction. Additionally, a poor health condition can be negated by higher income or wealth.

To the best of our knowledge, the use of decision trees in retirement satisfaction is introduced for the very first time in this article. The results show that this technique can be a very powerful method for revealing hidden relationships between the various predictors of retirement satisfaction.

REFERENCES

[1] D. Schwerha, C. Ritter, S. Robinson, R. W. Griffeth, and D. Fried, "*Integrating ergonomic factors into the decision to retire,"* Human Resource Management Review, 2011. 21(3), pp. 220-227.

[2] F. J. Floyd, et al., "*Assessing retirement satisfaction and perceptions of retirement experiences,"* Psychology and Aging, 1992. 7(4), pp. 609.

[3] K. A. Bender, "*An analysis of well-being in retirement: The role of pensions, health, and 'voluntariness' of retirement,"* The Journal of Socio-Economics, 2012. 41(4), pp. 424-433.

[4] L. T. Dorfman, "*Health conditions and perceived quality of life in retirement,"* Health & Social Work, 1995. 20(3), pp. 192-199.

[5] C. W. Panis, "*Annuities and retirement well-being,"* Pension design and structure: New lessons from behavioral finance, 2004, pp. 259-74.

[6] C. A. Price and S. Balaswamy, "*Beyond health and wealth: Predictors of women's retirement satisfaction,"* The International Journal of Aging and Human Development, 2009. 68(3), pp. 195-214.

[7] M. Reis and D. P. Gold, "*Retirement, personality, and life satisfaction: A review and two models,"* Journal of applied Gerontology, 1993. 12(2), pp. 261-282.

[8] N. Schmitt, J. K. White, B. W. Coyle, and J. Rauschenberger, "*Retirement and life satisfaction,"* Academy of Management Journal, 1979. 22(2), pp. 282-291.

[9] T. M. Calasanti, "*Gender and life satisfaction in retirement: An assessment of the male model,"* The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 1996. 51(1), pp. S18-S29.

[10] E. Fouquereau, A. Fernandez, and E. Mullet, "*Evaluation of determinants of retirement satisfaction among workers and retired people,"* Social Behavior and Personality: an international journal, 2001. 29(8), pp. 777-785.

[11] K. Hanson and S. Wapner, "*Transition to retirement: Gender differences,"* The International Journal of Aging & Human Development, 1994.

[12] C. A. Price and E. Joo, "*Exploring the relationship between marital status and women's retirement satisfaction,"* The International Journal of Aging and Human Development, 2005. 61(1), pp. 37-55.

[13] K. Seccombe and G. R. Lee, "*Gender differences in retirement satisfaction and its antecedents,"* Research on Aging, 1986. 8(3), pp. 426-440.

[14] K. F. Slevin and C. Wingrove, "*Women in retirement: A review and critique of empirical research since 1976,"* Sociological Inquiry, 1995. 65(1), pp. 1-21.

[15] M. E. Szinovacz, "*Preferred retirement timing and retirement satisfaction in women,"* The International Journal of Aging & Human Development, 1987.

[16] E. Bonsang and T. J. Klein, "*Retirement and subjective well-being,"* Journal of Economic Behavior & Organization, 2012. 83(3), pp. 311-329.

[17] B. Butrica, "*Satisfaction and engagement in retirement,"* 2006.

[18] L. T. Dorfman and M. M. Moffett, "*Retirement satisfaction in married and widowed rural women,"* The Gerontologist, 1987. 27(2), pp. 215-221.

[19] E. Fouquereau, A. Fernandez, A. M. Fonseca, M. C. Paul, and V. Uotinen, "*Perceptions of and satisfaction with retirement: A comparison of six European Union countries,"* Psychology and Aging, 2005. 20(3), pp. 524.

[20] Y. Kremer, "*Predictors of retirement satisfaction: a path model,"* International journal of aging & human development, 1983. 20(2), pp. 113-121.

[21] C. Kupperbusch, R. W. Levenson, and R. Ebling, "*Predicting husbands' and wives' retirement satisfaction from the emotional qualities of marital interaction,"* Journal of Social and Personal Relationships, 2003. 20(3), pp. 335-354.

[22] M. W. Craven, "*Extracting comprehensible models from trained neural networks*." 1996, UNIVERSITY OF WISCONSIN–MADISON.

[23] W. A. Young, et al., "*An investigation of TREPAN utilising a continuous oracle model,"* International Journal of Data Analysis Techniques and Strategies, 2011. 3(4), pp. 325-352.

[24] M. M. Nelson and W. T. Illingworth, "*A practical guide to neural nets."* Vol. 1. 1991, Addison-Wesley Reading, MA.

[25] K.-L. Du, "*Clustering: A neural network approach,"* Neural Networks, 2010. 23(1), pp. 89-107.

[26] A. Waibel, "*Modular construction of time-delay neural networks for speech recognition,"* Neural computation, 1989. 1(1), pp. 39-46.

[27] T. J. Sejnowski and C. R. Rosenberg, "*NETtalk: A parallel network that learns to read aloud."* 1988, MIT Press.

[28] J. W. Shavlik, R. J. Mooney, and G. G. Towell, "*Symbolic and neural learning algorithms: An experimental comparison,"* Machine learning, 1991. 6(2), pp. 111-143.

[29] G.-T. Hsu and R. Simmons. "*Learning footfall evaluation for a walking robot,"* in *Proceedings of the Eigth International Workshop on Machine Learning*, 2014, pp. 303-307.

[30] M. Jabri, et al. "*ANN based classification for heart defibrillators,"* in *Advances in neural information processing systems*, 1992, pp. 637-644.

[31] G. Tesauro, "*Practical issues in temporal difference learning."* 1992, Springer.

[32] G. G. Towell and J. W. Shavlik, "*Extracting refined rules from knowledge-based neural networks,"* Machine learning, 1993. 13(1), pp. 71-101.

[33] D. K. Chandra, V. Ravi, and P. Ravisankar, "*Support vector machine and wavelet neural network hybrid: application to bankruptcy prediction in banks,"* International Journal of Data Mining, Modelling and Management, 2010. 2(1), pp. 1-21.

[34] N. Naveen, V. Ravi, and D. A. Kumar, "*Application of fuzzyARTMAP for churn prediction in bank credit cards,"* International Journal of Information and Decision Sciences, 2009. 1(4), pp. 428-444.

[35] R. S. Gutierrez, A. O. Solis, and S. Mukhopadhyay, "*Lumpy demand forecasting using neural networks,"* International Journal of Production Economics, 2008. 111(2), pp. 409-420.

[36] R. Kuo, "*A sales forecasting system based on fuzzy neural network with initial weights generated by genetic algorithm,"* European Journal of Operational Research, 2001. 129(3), pp. 496-517.

 ISBN: 978-1-61208-457-2

[37] G. Zhang, B. E. Patuwo, and M. Y. Hu, "*Forecasting with artificial neural networks:: The state of the art,*" International journal of forecasting, 1998. 14(1), pp. 35-62.
[38] K.-I. Funahashi, "*On the approximate realization of continuous mappings by neural networks,*" Neural networks, 1989. 2(3), pp. 183-192.
[39] K. Hornik, M. Stinchcombe, and H. White, "*Multilayer feedforward networks are universal approximators,*" Neural networks, 1989. 2(5), pp. 359-366.
[40] M. Kubat, "*Neural networks: a comprehensive foundation by Simon Haykin, Macmillan, 1994, ISBN 0-02-352781-7.*" 1999, Cambridge Univ Press.
[41] G. R. Weckman, W. Young, S. Hernández, M. Rangwala, and V. Ghai, "*Extracting Knowledge from Carbon Dioxide Corrosion Inhibition with Artificial Neural Networks,*" International Journal of Industrial Engineering: Theory, Applications and Practice, 2010. 17(1).
[42] B. Baesens, R. Setiono, C. Mues, and J. Vanthienen, "*Using neural network rule extraction and decision tables for credit-risk evaluation,*" Management science, 2003. 49(3), pp. 312-329.
[43] G. Liepins, R. Goeltz, and R. Rush, "*Machine learning techniques for natural-resource data-analysis,*" Ai Applications in Natural Resource Management, 1990. 4(3), pp. 9-18.
[44] D. Biggs, B. De Ville, and E. Suen, "*A method of choosing multiway partitions for classification and decision trees,*" Journal of Applied Statistics, 1991. 18(1), pp. 49-62.
[45] W. Leech, "*A rule-based process control method with feedback,*" ISA transactions, 1986. 26(2), pp. 73-78.
[46] P. Langley and H. A. Simon, "*Applications of machine learning and rule induction,*" Communications of the ACM, 1995. 38(11), pp. 54-64.
[47] M. Craven and J. W. Shavlik. "*Using Sampling and Queries to Extract Rules from Trained Neural Networks,*" in *ICML*, 1994, pp. 37-45, Citeseer.
[48] D. Martens, J. Huysmans, R. Setiono, J. Vanthienen, and B. Baesens, "*Rule extraction from support vector machines: an overview of issues and application in credit scoring*," in *Rule extraction from support vector machines*. 2008, Springer, pp. 33-63.
[49] K. Krawiec, R. Słowiński, and I. Szcześniak. "*Pedagogical Method for Extraction of Symbolic Knowledge,*" in *Rough Sets and Current Trends in Computing*, 1998, pp. 436-443, Springer.

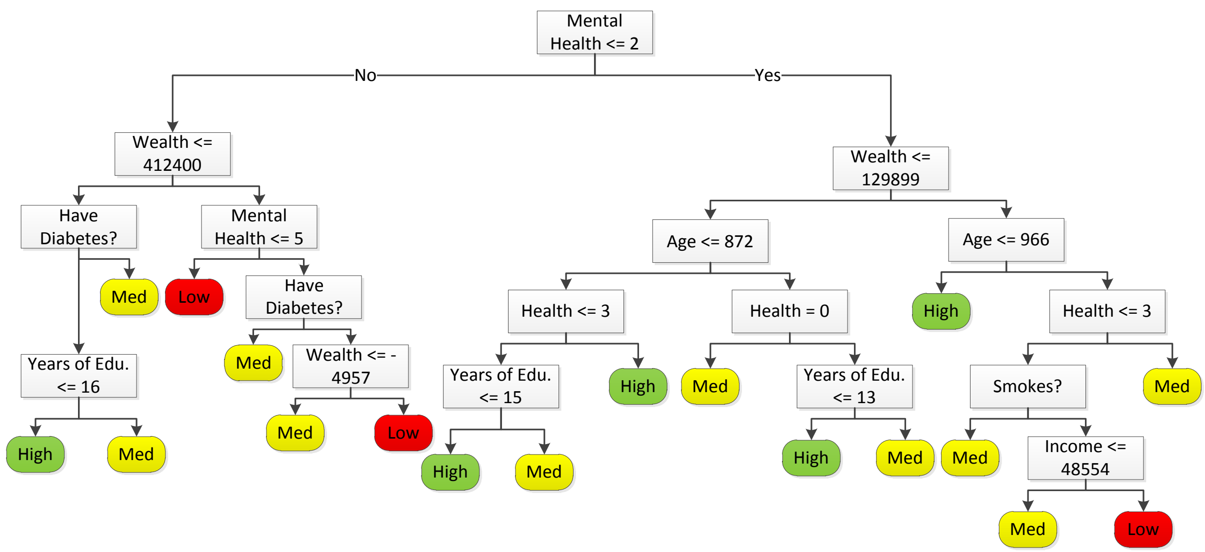

Figure 2. Decision Tree of Retirement Satisfaction for Men.

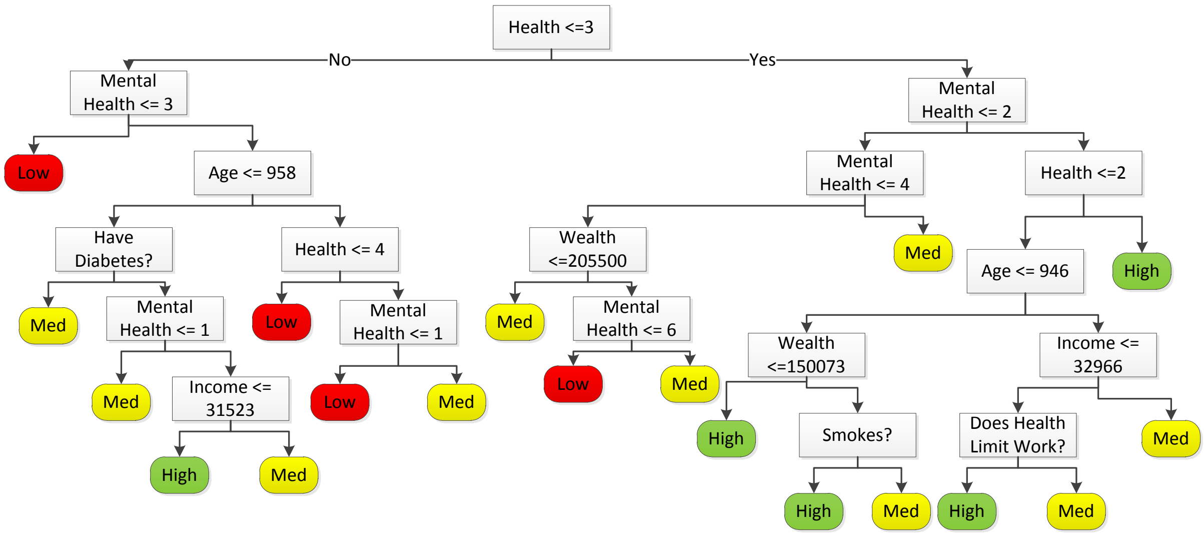

Figure 3. Decision Tree of Retirement Satisfaction for Women.