Predicting the Individuals' job satisfaction and determining the factors affecting it using the CHAID Decision Tree Data Mining Algorithm

Case Study: the National Opinion Research Center of the United States

Farhad Sheybani

Abstract - **As the general attitude of the individual about what he does, job satisfaction is the result of individual perceptions from the workplace and the factors and conditions in it; it is also influenced by his personality traits. Meanwhile, investigating the job satisfaction is of great importance in advanced societies. The present study aimed to assess the job satisfaction in the United States and evaluate the hypothesis of existence of job dissatisfaction and the factors affecting it in the studied sample. The various social data, related to job satisfaction and collected by the National Opinion Research Center of the United States, are used in this study. The sample is consisted of different people including male and female samples from nine different states in the United States. For the purpose of this study, the patterns of data were discovered, and factors affecting job satisfaction were identified using the CHAID decision tree data mining method. Finally, it was found that a small percentage of people are dissatisfied with their job.**

Keywords— **Job Satisfaction, factors affecting job satisfaction, National Opinion Research Center, data mining, CHAID decision tree.**

I. Introduction

Job satisfaction is a type of attitude according to experts, and they are defined it as the individual's attitude toward the job and simply put, how someone feels about his/her job and its various aspects.

As one of the major issues in organizational literature and one of the most important and most common research subjects in the field of organizational behavior studies [1]. more than 5,000 papers and theses have been prepared on job satisfaction [2]. There are more than 12,400 studies until 1991 only on the subject of job satisfaction [3]. Also, more than 6300 doctoral dissertations have been available in the International PhD dissertation abstract, and more than 3350 research papers have been published in this field [4].

According to the recent studies, personality is one of the components that affects job satisfaction; however, there are inadequate results about the nature, characteristics and traits leading to job satisfaction [5].

Three general categories can be speculated for the factors relating job satisfaction, each of which includes sub-factors [6].

A) Prerequisite factors that are essential for the formation of job satisfaction. Four categories of factors contribute to the formation of job satisfaction. In this case, job satisfaction is considered as a dependent variable that influences 4 factors.

Job characteristics include diversity, identity, tasks importance, autonomy, feedback, and job enhancement, role-related characteristics (role conflict and role ambiguity), group and organizational characteristics including group cohesion, community quality, commitment to participation, work pressures, Inequality in workplace, organizational structure, organizational justice, organizational climate and organizational support, the relations with the leader includes the structure of leadership intimacy, leadership considerations, leadership productivity, leader' punishment and encouragement behaviors, and exchange relations between members and leader.

B) Correlational factors of job satisfaction

Some organizational concepts, correlated with job satisfaction include organizational commitment, life satisfaction, job pressure and stress, job engagement and and job attitudes.

C) the effects and consequences of job satisfaction

These factors include three categories of motivation, civic behavior and latency behaviors such as absenteeism, turnover, and turnover intention; and performance is placed in the last category.

A combination of different categories of these factors, each of which constitute a fields in the database under study, has been used for the purpose of this study. At first, the data mining tool was used to prepare this data and then, using the CHAID decision tree in Clementine 12.0 software, factors affecting job satisfaction were investigated in the sample individuals.

The remainder of this research is structures as follows: a brief discussion of the data mining and decision tree is presented at first. The research field have been examined in the second part. The third part is assigned to the research method and tools. Finally, the extracted results are expressed and suggestions will be given in the fourth section.

II. Problem Statement

A. Data-mining

The data mining process analyzes databases and macro data for the purpose of discovery and dissemination of knowledge using mechanized and semi-mechanized methods. Such studies and explorations can be considered as the same extension and continuation of ancient and comprehensive

knowledge of statistics. The distinction is in scale, the breadth and variety of fields and applications, as well as the dimensions and size of today's data that requires machine learning, modeling, and training methods [7].

The term data mining refers to extract hidden information, or patterns and relationships in a large volume of data in one or more large databases.

Data mining also refers to the use of data analysis tools to discover patterns and valid relationships which have not been known until now. These tools may include statistical models, mathematical algorithms, and machine learning methods. Data mining is not exclusive to data collection and management and also includes information analysis and forecasting. Data-mining methods have been widely used in previous studies. In [8], researchers used a data-mining based method for protection of power systems. Many simulation results were conducted and the capability of the data-mining method was verified. In [9, 10, 11] data mining and operation research method such as Data Envelopment Analysis (DEA) can be addressed through diverse computational, and combinatorial models.

The exploratory applications that examine text or multi-media files to extract data consider various parameters including:

Association rules: Patterns upon which an event is connected to another, such as connecting a pen purchase event to the paper purchase event

Order: the pattern for analyzing the sequence of events and determining the event that leads to other events such as the birth of a baby and the purchase of milk powder.

Classification: Identifying new patterns such as the concurrency of glue and folder purchases.

Clustering: Discovering and documenting a set of unknown facts such as a geographic location of buying a branded product.

Forecasting: Discovering the patterns by which an acceptable prediction is presented of future events such as the membership in a sports club by attending sports classes.

Data-mining methods use a given dataset and try to find out the relationships between data to perform an accurate classification. Then, experts in the field can interpret the output. As an example, in [12] an intelligent method using the specialists' knowledge is used for generating electricity from waste energy in industry. This method not only shows a great example of intelligent techniques for engineering applications, but also provides the industry with a cost-effective approach.

B. Decision Tree

A decision tree is used as a tool for illustrating and analyzing decisions in decision analysis where the expected values of the competition are alternately calculated. A decision tree has three types of nodes:

1) Decision node: that typically represented by a square
2) Random node: that specifies with the circle
3) End node: Determined by a triangle.

The Fig. 1. shows the schematic view of a decision tree.

Very compact and in the form of a diagram, a decision tree can draw attention to the problem and the relationship between events. The square represents the decision, the ellipse represents the activity and the diamond represents the result. Decision trees and decision diagrams have advantages over decision support tools:

1) Simple understanding: Every person can learn the way of working with the decision tree, with little study and training.
2) Working with macro and complex data: The decision tree can easily work with complex data and decide on them.
3) Easy reuse: If the decision tree is made for an issue, different instances of that problem can be calculated with that decision tree.
4) Ability to combine with other methods: The decision tree result can be combined with other decision-making techniques and obtain better results.
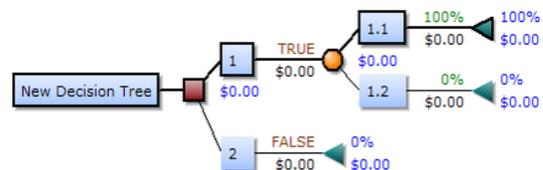5) Fig. 1. A schematic view of decision tree

Fig. 1. A schematic view of decision tree

C. The Research field

The National Opinion Research Center (NORC), founded in 1941, has headquarter located at the University of Chicago. It also has offices in Chicago, Washington, Bethesda, Maryland and Berkeley, California. Furthermore, the staff of this organization are located throughout the United States. The NORC clients include government agencies, educational institutions, a variety of charities and nonprofits, and private companies (firms). Most of the studies in the center are in national level, although some local-to-international projects are also conducted in this center. NORC creates a unique value for its customers by developing innovative and innovative solutions. In fact, NORC combines advanced technology with quality social science research with a generic benefit. The GSS project is one of the characterized research conducted by NORC since 1972, which the twenty-eighth time is conducted by 2010 in the United States. In the last third of the century, the GSS has had a profound impact on social change and the complex growth of the American community [13].

GSS is the largest project funded by the National Science Foundation's Sociology Program. In addition, the project is a source for the US Census Center, which is often the source of information analysis in the social sciences.

As said, to insert images in Word, position the cursor at the insertion point and either use Insert | Picture | From File or copy the image to the Windows clipboard and then Edit | Paste Special | Picture (with "Float over text" unchecked).

The authors of the accepted manuscripts will be given a copyright form and the form should accompany your final

submission.

III. RESEARCH METHOD AND FINDINGS

The data used in this study are prepared from the GSS research database in 2010. The database includes data on 891 men and 1153 women in 9 different states in the United States, within a SPSS file with 2044 records. This database contains 790 attributes (columns), which used for 40 attributes defined hypothesis. As stated above, the research hypothesis was to find out the amount and factors affecting the individuals' dissatisfaction in the sample and a decision tree algorithm was used for this purpose. To find patterns in the job satisfaction of individuals in the sample, the Clementine 12.0 data mining software was used and the CHAID decision tree model was used to implement this technique.

The research aimed to investigate the effective patterns on the prediction of job individuals' dissatisfaction among in this database. Therefore, the target field in this database is "job satisfaction in general". Apart from the missing values described below, this field is categorized into four categories and coded with four numbers (considering that the English equivalent of each of the fields and their respective categories, are put in the text next to their Persian translation because they are exactly in the software and the database, so that their nature is different from the English equivalent of the keywords mentioned in the footnote).

First class: Complete satisfaction with "all satisfied", code 1.

Second Class: Somewhat satisfied with "relative satisfaction", code 2.

Third floor: Relative dissatisfaction with "not too satisfied", code 3.

Fourth Floor: Dissatisfaction with "Not at all satisfied", code 4.

At first, the values of the "No Answer", "Do not Know" and "Unacceptable IAP" responses from the target field (job satisfaction) that are not effective in predicting the final goal based on the research hypothesis, were considered as "missing values" with data preparation methods. Then, the records required to explore the job satisfaction model were converted to 1165 records. After making the flow in the Clementine 12.0 software and implementing the CHAID decision tree model, the obtained tree was analyzed. A view of the tree can be seen in Figure 2, in which, there are nodes covering the highest percentage of people with relative job dissatisfaction or complete dissatisfaction (job dissatisfaction).

As shown in Fig. 2, the first node of the tree contains the information for the target field, labeled as Node 1. The information below the "Category" term are the categories related to the values of that specific field (as previously stated, the categories selected on the basis of the hypothesis are considered, and the rest of the categories are considered as missing. the percentages of each responses to each category is shown in general. As we see, the highest percentage of response was to the category of "complete satisfaction", in general. The column 'n' indicates the number

of records (individuals) who responded to the corresponding classification. Obviously, this value has a direct relationship with the percentage amount. Finally, the aggregate values are specified in the last row. The cumulative amount of records is 1161. The reason of the reduction of this number is the deletion of records responded to the missing values. The first branch of the parent Node (Target Node) in this tree is related to the field of "Respondent proud to work for employer" which indicates that this field has a greater impact in predicting the target field in this model. As we see in Figure 2, the following fields affect job dissatisfaction:

"How likely respondent make effort for new job next year",
"How often does respondent find work stressful?"
"The relation between management and employees"
And "respondent proud to work for employer".

It is concluded, from the general schema of the tree, that a small percentage of people are dissatisfied with their jobs in general and in comparison with those who are satisfied with their job and this percentage is even much lower for a state of complete dissatisfaction. So that only one node shows the frequency of this issue. However, the same number has emerged due to the existence of various factors that the decision tree has discovered for its patterns.
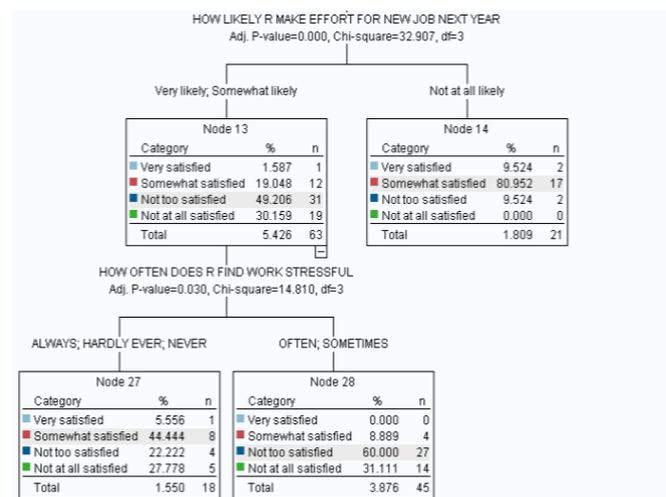

Fig. 2. The view of the decision tree created by the CHAID algorithm

It can be seen, according to the decision model made by the CHAID decision tree, that among all the records (individuals) that are located in the Node, 25 people are strongly disagree with the question of proud to work for employer. Although very few (less than 5%), however 14 out of 25 people, namely 56% are completely dissatisfied with their jobs (lack of satisfaction).

84 respondents were disagree in response to the question of proud to work for employer, of which 33 people (40 percent) have "relative dissatisfaction" with their job (Not at all satisfied); Node 3. Among the 84 people in Node 3, who are disagree with the sense of proud to work for employer, 63 people, namely 75% are looking for a new job next year; Node 13, and among them, 63 people, 31, namely about 50%

of them had "relative dissatisfaction" with their job and 30% are "dissatisfied"

Among the 63 people looking for a new job, 45 people see their jobs as stressful in most cases. 45 people, see their job stressful in most of the times. 27 out of 45 people, namely 60% have 'relative dissatisfaction' and 31% are "dissatisfied"; Node 28.

On the other side of this tree, there are people who "agreed" in response to a sense of proud to work for employer; Node 2. These people make up about 53% of the total (613 records out of 1161 records), among which, 34 people considers the relationship between the management of the organization and the employees as "Quite Bad" and "Very Bad"; Node 12 and among these 34 people, 21 people knows their job "always" and" often" stressful. Out of these 21 people, about 50% (10 people) face a "relative dissatisfaction" with the job.

Compared to other tree fields and branches, the percentage of "relative dissatisfaction" or "dissatisfaction" in the nodes affected by other factors is almost zero on average (although in most of the nodes it is more than zero) and other factors have not had much effect on these dissatisfactions, and most people have enjoyed "full satisfaction" and "relative satisfaction" in these nodes. Therefore, these nodes are neglected.

The Gain table in Clementine software shows the various values in tree nodes with different titles based on the selected target category. This table is shown in Figure 3 for reviewing the target category of "dissatisfaction."

Fig. 3. Gain table for the target category of Dissatisfaction" in the target field of job satisfaction

IV. CONCLUSIONS AND SUGGESTIONS

According to the results, it can be concluded that the lack of individuals' pride to work for their employer can have a great impact on the full job dissatisfaction. Similarly, searching for a new job can have a direct relationship with job dissatisfaction and on the other hand, the stressful work environment has had a huge impact on job dissatisfaction classifications. Also, it can be concluded from the last observations at the previous stage that the lack of a good relationship between organization management and employees may lead to job dissatisfaction.

In general, according to the research results, a small percent of people are dissatisfied with their job and this percentage is even much lower in case of people with complete dissatisfaction. However, this number can be reduced to zero in advanced societies, such as the United States by reducing the problems caused by the results obtained in the previous section.

REFERENCES

[1] P.E. Spector (2006), Industrial and Organizational Psychology: research and Practice, 4th edition, Wiley, New York.
[2] C. J. Cranny, P.C Smith, E.F. Stone, (1992), Job Satisfaction, Lexington Books, New York.
[3] P.E Spector, (1997), Job Satisfaction Thous and Oaks, CA: Sage publications Inc.
[4] A.F Kinicki, R. McKee, C, Schriesheim, K.P Carson, (2002), Assessing the Construct Validity of the Job Descriptive Index:A Review and Meta-Analysis, Journal of Applied Psychology, Vol. 87, 1, PP. 14-32.
[5] W Jouge, (2000), Persona; Chareacteristics & Job Satisfaction of Greek Teachers, Journal of Organizational Behavior.
[6] P.S, Topolosky, (2000), Linking Employee satisfaction To Business Results, Garland Publishing Inc, New York.
[7] Adriaans P., (1999), Introduction to Data Mining and Knowledge Discovery, Two Crowns Corporation.
[8] O. A. Gashteroodkhani, M. Majidi, M. Etezadi-Amoli, A. F. Nematollahi, B. Vahidi "A hybrid SVM-TT transform-based method for fault location in hybrid transmission lines with underground cables" Electric Power Systems Research, vol. 170, pp. 205-214, 2019.
[9] M. Mirmozaffari, A. Alinezhad. "Ranking of Heart Hospitals Using Cross-Efficiency and Two-stage DEA," 7th International Conference on Computer and Knowledge Engineering (ICCKE 2017), Ferdowsi University of Mashhad, 978-1-5386-0804-3/17/$31.00 ©2017 IEEE. October 26-27, 2017.
[10] M. Mirmozaffari, & A Alinezhad, (2017), Window Analysis Using Two-stage DEA in Heart Hospitals. International Conference on Innovation in Science, Engineering Computer and Technology (ISECT), Dubai (UAE) Oct.
[11] M. Mirmozaffari, (2018), Eco-Efficiency Evaluation in Two-Stage Network Structure: Case Study: Cement Companies, Iranian Journal of Optimization. Volume 11, Issue 2.
[12] O. A. Gashteroodkhani, M. Majidi, M. Etezadi-Amoli, "A Fuzzy-based Control Scheme for Recapturing Waste Energy in Water Pressure Reducing Valves" IEEE Power and Energy Society General Meeting (PESGM), pp. 1-5, Portland, OR, Aug 2018.
[13] Http://www3.norc.org/gss+website/

Farhad Sheybani was born in Astara, Guilan, Iran. He is with the department of Electrical Engineering, University of Guilan, Rasht, Iran. His research interests include data-mining, artificial intelligence and fuzzy rule based applications. (email: farhad9090@yahoo.com)