ANALYSIS OF BEHAVIORAL DATA IN BUSINESS BURNOUT DURING ECONOMIC UPHEAVAL IN GREECE

Constantinos Halkiopoulos, Hera Antonopoulou and Evgenia Gkintoni

Entrepreneurship & Digital Innovation Laboratory, Department of Management Science and Technology, University of Patras, Greece

DOI: http://dx.doi.org/10.24327/ijrsr.2020.1101.5076

ARTICLE INFO

Article History:
Received 6th November, 2019
Received in revised form 15th November, 2019
Accepted 12th January, 2019
Published online 28th February, 2020

Key Words:
Behavioral Data, Burnout, Business, MBI, Economic Upheaval, Data Mining, R

ABSTRACT

In this research paper presents the Evaluation of burnout Index professionals working in both the private and public sector, amid the economic upheaval in Greece. Data were collected by filling working adults aged 18 years and above through a questionnaire which is designed to measure burnout at work (Maslach Burnout Inventory - MBI). This questionnaire measures three dimensions of burnout: (1) Emotional Exhaustion, (2) Depersonalization, (3) Personal Accomplishment. Subsequently, selected data for analysis by adding additional parameters such as gender, age, education, occupation and birth / residence. Then an appropriate pretreatment provided as input to a special software and Machine Learning Data Mining (R) to generate patterns and inference rules. In conclusion, the results obtained by weighting and criteria basis, evaluate and present adult rates, which in turn analyze the degree of burnout.

INTRODUCTION

The Greek economy situation today is characterized particularly unfavorable. One of the areas most affected significantly by the current reality is the workplace, with many people losing their jobs and are unable to meet basic survival needs. Thus, it is understood that the economic downturn and its impact affecting both physical and mental health. The economic upheaval seems to exert a direct influence on people's behavior, in the event of stress, insecurity and uncertainty about the future, and anger expression for the perpetrators of the situation in which the country found.

A major issue goes with an economic upheaval and burnout. This phenomenon has been addressed in recent decades many disciplines, such as psychology, sociology, economics as well as the impact on the worker is very important, both for himself and for the business in which they work. The study of the factors affecting it now becomes imperative, given that stress at work plays a very important role in the onset, progression and deterioration of various conditions related to mental health, this phenomenon deserves special attention.

*Corresponding author: Constantinos Halkiopoulos
Entrepreneurship & Digital Innovation Laboratory, Department of Management Science and Technology, University of Patras, Greece
abilities (employee offering) and also has some personal needs, requirements and expectations of the working space such as recognition, prestige, financial requirements, professional development opportunities, etc.

On the other hand, the working environment, has certain requirements and expectations of the employee (demand) because the reward for promotion of (labor supply). But when these calling and gives the practitioner does not agree to those demands and requests the working environment is created gap, which involves an imbalance or discrepancy. This has the effect of increasing chronic stressor likely that progressively leads to the appearance of burnout syndrome (Anagnostopoulos and Papadatou, 1999). Figure 1 is a diagrammatic representation of the dynamics between the various actors’ relationships that lead to the appearance of the syndrome (Schaufeli W. And Enzmann D., 1998).

**Description of Maslach Burnout Inventory (MBI)**

In this research in order to measure the phenomenon of burnout was used the Maslach Burnout Inventory (MBI). The Maslach Burnout Inventory (MBI) is the most commonly used tool to self-assess whether you might be at risk of burnout. To determine the risk of burnout, the MBI explores three components: exhaustion, depersonalization and personal achievement. While this tool may be useful, it must not be used as a scientific diagnostic technique, regardless of the results. The objective is simply to make you aware that anyone may be at risk of burnout.

According to Maslach, a questionnaire composed of 22 self-declarations, 9 for emotional exhaustion, depersonalization 5 to 8 for personal accomplishment. Each statement evaluated in 7-Likert scale, with responses from 0 = never, 1 = few times a year, 2 = once a month, 3 = few times a month, 4 = once a week, 5 = few times a week, up to 6 = every day, the above two dimensions of burnout. The higher score indicated the subscale of emotional exhaustion and depersonalization, and the higher levels of burnout (Maslach et al., 1996).

The internal consistency reliability of the three subscales are relatively high. Specifically, the index Cronbach's a for emotional exhaustion are 0.83, 0.75 for depersonalization and personal achievement 0.76.

**Description of Machine Learning & Data Mining methods**

Data Mining is an emerging knowledge discovery process of extracting previously unknown, actionable information from very large scientific and commercial databases. It is imposed by the explosive growth of such databases. Usually, a data mining process extracts rules by processing high dimensional categorical and/or numerical data. Classification, clustering and association are the most well known data mining tasks.

Classification is one of the most popular data mining tasks. Classification aims at extracting knowledge which can be used to classify data into predefined classes, described by a set of attributes.

Association rules can be used to represent frequent patterns in data, in the form of dependencies among concepts attributes.

In this paper, we consider the special case, that is known as the market basket problem, where concepts-attributes represent products and the initial database is a set of customer purchases (transactions).

Clustering involves finding a specific number of subgroups (k) within a set of n observations (data points/objects); each described by d attributes. A clustering algorithm generates cluster descriptions and assigns each observation to one cluster (exclusive assignment) or in part to many clusters (partial assignment).

**METHODOLOGY**

In this paper were applied Machine Learning and Data Mining methods in order to evaluate the burnout (MBI) of employees in Business. The methodology, that was adopted, consists of three concrete phases. During the first phase electronic questionnaires were created and posted through the website http://www.cicos.gr. Subsequently, data were collected and preprocessed from the questionnaires. The data set for analysis was consisted of demographics elements of responders, such as the gender, the birth-place, the place of present residence, educational background of both the respondents and their parents, professional occupation of parents and also of subscales of the MBI Inventory. During the third phase, the data set was analyzed based on Data Mining techniques and evaluate the results. More specifically, we utilized classification algorithms so as to manage to describe the hidden patterns underlying in the data. Decision trees are a powerful way in order to represent and facilitate statements analysis (psychological) principally, comprising successive decisions and variable results in a designated period.

**Sample**

A total of 151 employees, 72 females and 79 males, (48% female, 52% male) were recruited from several job occupations from public and private work sectors.

**Total per category**

- [18,21]: Women 24, Men 21, Total 45
- [22,24]: Women 20, Men 25, Total 45
- [>=25]: Women 28, Men 33, Total 61

**RESULTS**

Classification methods aim to identify the classes from some descriptive traits. They find utility in a wide range of human activities and particularly in automated decision making. Decision trees are a very effective method of supervised learning. It aims is the partition of a dataset into groups as homogeneous as possible in terms of the variable to be predicted. It takes as input a set of classified data, and outputs a tree that resembles to an orientation diagram where each end node (leaf) is a decision (a class) and each non-final node (internal) represents a test. Each leaf represents the decision of belonging to a class of data verifying all tests path from the root to the leaf. The tree is simpler, and technically it seems easy to use. In fact, it is more interesting to get a tree that is adapted to the probabilities of variables to be tested. Mostly balanced tree will be a good result. If a sub-tree can only lead to a unique solution, then all sub-tree can be reduced to the simple conclusion, this simplifies the process and does not change the final result. Ross Quinlan worked on this kind of decision trees.
Decision trees are built in "ctree (Conditional Inference Trees)" by using a set of training data or data sets. At each node of the tree, "ctree" chooses one attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. Its criterion is the normalized information gain (difference in entropy) that results from choosing an attribute for splitting the data. The attribute with the highest normalized information gain is chosen to make the decision. During the construction of the decision tree, it is possible to manage data for which some attributes have an unknown value by evaluating the gain or the gain ratio for such an attribute considering only the records for which this attribute is defined. Using a decision tree, it is possible to classify the records that have unknown values by estimating the probabilities of different outcomes. Ctree builds decision trees from a set of training data in the same way as ID3 or C4.5, using the concept of information entropy.

The training data is a set \( S = s_1, s_2, \ldots \) of already classified samples. Each sample \( s_i \) consists of a p-dimensional vector \((x_{1i}, x_{2i}, \ldots, x_{pi})\), where the \( x_{ji} \) represent attribute values or features of the sample, as well as the class in which \( s_i \) falls. At each node of the tree, "ctree" chooses the attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. The splitting criterion is the normalized information gain (difference in entropy). The attribute with the highest normalized information gain is chosen to make the decision. The "ctree" algorithm then recurs on the smaller sublists. In order to specify the best result, it was necessary to fit the data to the model in a proper way. This task was carried away by changing and testing the controls of "ctree".

The parameters in the control function that were altered are:

- mincriterion: The value of the test statistic (for testtype \( \equiv \) "Teststatistic"), or 1 - p-value (for other values of testtype) that must be exceeded in order to implement a split.
- mnsplit: The minimum sum of weights in a node in order to be considered for splitting.
- mtry: The number of input variables randomly sampled as candidates at each node for random forest like algorithms.
- maxdepth: The maximum depth of the tree.

Tree 1

- Depended variable: age
- Independent variables: sex +3 indexes
- ‘mincriterion’ value: 0.01
- ‘mnsplit’ value: 20L
- ‘mtry’ value: Inf (Infinite)
- ‘maxdepth’ value: Inf (Infinite)

Clustering Results

Clustering involves finding a specific number of subgroups (k) within a set of n observations (data points/objects); each described by d attributes. A clustering algorithm generates cluster descriptions and assigns each observation to one cluster (exclusive assignment) or in part to many clusters (partial assignment).

Throughout this paper, we shall refer to the output of a clustering algorithm (e.g, the medoids of the clusters) as clustering rules. To estimate the number of clusters, we used the Partitioning Around Medoids method with the help of 'pamk' algorithm in R. It prints the suggested number of clusters based on optimum average silhouette width. The optimal number of clusters that came up is 2.

Pamk performs a partitioning around medoids clustering with the number of clusters estimated by optimum average silhouette width or Calinski-Harabasz index. The Duda-Hart test is applied to decide whether there should be more than one cluster (unless 1 is excluded as number of clusters or data are dissimilarities).
Cluster means

Group

| sinesthimatiki_eksantlisisapotropoihsieillipsi_prosopikon_eptueugmaton |
|--------------------------|----------------|----------------|
| 1 | -0.6385339  | -0.6417615  | 0.2633158  |
| 2 | 0.8680070   | 0.8723945   | -0.3579449 |

Plot interpretation

Clusplot: A bivariate plot visualizing a partition (clustering) of the data. All observations are represented by points in the plot, using principal components or multidimensional scaling. Around each cluster an ellipse is drawn.

Silhouette plot: Gives the silhouette width information for each data point, average silhouette width for each cluster and for the whole data. Silhouette Width is a measure to estimate the dissimilarity between clusters. A higher silhouette width is preferred to determine the optimal number of clusters.

Mining Association Rules

Association Rule Mining is a common technique used to find associations between many variables. In Data Mining, Apriori is a classic algorithm for learning association rules. Apriori is designed to operate on databases containing transactions (for example data collected from surveys in this case). As is common in association rule mining, given a set of item sets, the algorithm attempts to find subsets which are common to at least a minimum number C of the itemsets.

Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time, and groups of candidates are tested against the data. The algorithm terminates when no further successful extensions are found. Apriori uses breadth-first search and a tree structure to count candidate item sets efficiently. It generates candidate item sets of length k from item sets of length k – 1. Then it prunes the candidates which have an infrequent sub pattern. According to the downward closure lemma, the candidate set contains all frequent k-length item sets. After that, it scans the transaction database to determine frequent item sets among the candidates.

Association rules present association or correlation between item sets. An association rule has the form of A ⊃ B, where A and B are two disjoint item sets.

The Goal: Studies whether the occurrence of one feature is related to the occurrence of others.

Three most widely used measures for selecting interesting rules are:

- **Support** is the percentage of cases in the data that contains both A and B,
- **Confidence** is the percentage of cases containing A that also contain B, and
- **Lift** is the ratio of confidence to the percentage of cases containing B.

Apriori rules visualization

Scatterplot

This visualization method draws a two dimensional scatterplot with different measures of interestingness (parameter "measure") on the axes and a third measure (parameter "shading") is represented by the color of the points. There is a special value for shading called "order" which produces a two-key plot where the color of the points represents the length (order) of the rule.

Matrix3D

Arranges the association rules as a matrix with the item sets in the antecedents on one axis, and the item sets in the consequent on the other. The interest measure is either visualized by a color (darker means a higher value for the measure) or as the height of a bar (method "matrix3D"). Specifically of our use, the parameters that were altered are:

- measure = "lift"
- control = list(reorder = TRUE)

Grouped Matrix plot
Antecedents (columns) in the matrix are grouped using clustering. Groups are represented as balloons in the matrix.

CONCLUSIONS

The results of this study showed differences both in terms of three (3) indicators of burnout (burnout) and in terms of demographic factors such as gender. Specifically, in the total sample, the professionals involved, reported high levels of emotional exhaustion and lack of personal achievements, but the depersonalization factor, not noted as significant in this sample. Regarding gender, male employees than women, appeared to occur more influenced by the phenomenon of burnout on indicators of emotional exhaustion and personal accomplishment. Following the above, the economic upheaval from which in recent years Greece is affected significantly, the phenomenon of burnout seems to influence a proportionate manner, several of the professional categories of employees, in both the private and public sectors.

References