A Model for the Mental Health Evaluation of Skilled Workers in Technology-Based Companies

Moein Yousefi (mo.yousefi@atu.ac.ir)
Allameh Tabatabai University https://orcid.org/0000-0002-8362-7620

Iman Raeesi Vanani
Allameh Tabatabai University

Research

Keywords: Mental Health, Mental Disorder, Classification, Clustering

Posted Date: May 28th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-30514/v1

Title:

A Model for the Mental Health Evaluation of Skilled Workers in Technology-Based Companies

Authors:

Moein Yousefi

MSc at Allameh Tabataba'i University

Iman Raeesi Vanani

Assistant Professor at Allameh Tabataba'i University

Abstract

Human resources are considered as one of the most valuable assets of an organization. Therefore, the top priorities of organizational managers include paying attention to manpower and providing necessary facilities to bring about satisfaction with workplace and working conditions in the light of mental and physical health. Accordingly, mental health has partly been neglected by managers because it lacks concrete and tangible aspects on the contrary to other organizational dimensions such as buildings and machinery. This study was conducted to achieve the fundamental goal of identifying the aspects and characteristics affecting the mental health of employees at technology-based companies as well as the factors resulting in mental problems and disorders. It was also decided to develop a model to predict affliction or non-affliction with mental problems among employees in the long term. For this purpose, this study focused on data mining techniques in order to identify latent relationships and introduce a prediction model. Therefore, the important and major factors were identified as individual characteristics, occupational characteristics, organizational performance features, and organizational governing culture features in mental health. The analysis and evaluation of results indicated that Classification algorithms managed to predict the mental health of employees with nearly %85 of accuracy. Besides, Clustering algorithms succeeded in dividing the samples into high-risk, medium-risk, and low-risk classes. The designed models can help organizational managers identify the factors affecting the mental health of employees and predict the chances of affliction with mental health disorders to prevent destructive harm to employees and organizations.

Keywords: Mental Health, Mental Disorder, Classification, Clustering

Introduction

One of the most important issues that have attracted human attention for long is the phenomenon of health. It is divided into two main categories of physical health and mental health. Studies have shown that mental health is considered to mean complete physical, mental and social well-being, and it is a concept beyond the lack of mental disorders (Khaghani Zadeh et al.,2016), in the sense that one cannot consider someone to have mental health by simply relying on his or her lack of mental illness; therefore, it is not possible to measure individuals' level of mental health merely through psychological variables and factors (Moosavi, 2014).

Since today's human life is tied to economic activities, the vast majority of people in the community have devoted a part of their lives to working and spending a great part of their time in workplaces around the clock (Keshavarz Afshar et al., 2016). Therefore, it is important to study the mental health of employees not only from the individual aspect but also from organizational and social aspects. Obviously, various individual, social, cultural and economic variables and factors make it difficult to analyze and predict the process of employees' mental health not only in Iran but also in many other countries in the world (Esmaeeli et al., 2015).

Lack of adequate knowledge of these factors and their impact on mental health makes it difficult for the organization management to provide the mental health of its employees properly and disregarding this issue can have consequences including destructive effects on the life of each employee and irreparable effects on the organization (Abraham et al., 2017; Alonso et al., 2018). Considering this issue, the current study tries to identify the most important factors affecting the mental health of employees by studying the previous studies and analyzing the available data through the data mining process.

To this end, data Classification and Clustering techniques were used and the most important variables affecting the target variable, i.e. mental health of each employee, were identified. Then, the clusters and categories generated were used to predict the likelihood of getting infected with mental health problems. Actually, the main questions of this study are: what aspects affect the mental health of expert staff and the emergence of psychological problems in technology-based companies? and is it possible to predict the need to treat psychological problems in expert staff at technology-based companies based on the individual aspects, occupational aspects, organization performance and the organizational governing culture on employees' health?

Attempts were also made to identify the variables, patterns, and latent relations among the data for diagnosis and prediction of employees' mental problems and diseases through data mining methods and tools.

Research Background

Mental Health

There are various definitions of mental health in different societies and cultures. This concept includes positive inner feelings, self-confidence, self-efficacy, and self-reliance as well as the discovery of competitive capacities, potential intellectual abilities, and self-actualization (Moosavi, 2014). Although it seems impossible to give a comprehensive definition of mental health because of intercultural

differences, there is a consensus that mental health is a concept beyond the lack of mental disorders. In fact, having no mental disorders cannot merely guarantee that an individual is mentally healthy (Jorm, 2000; Kessler and Üstün, 2004). Despite the fact that many individuals have no specific mental disorders, it is evident that some of them are mentally healthier than others. However, mental healthcare is obviously as important as physical healthcare (Hunta and Eisenberg, 2010).

Factors Affecting Mental Health

Different factors affect mental health. Although individual parameters are not separate from environmental parameters, there are comprehensive Classifications in which individual factors (e.g. genetic factors, gender, physiology, individual beliefs, and personal skills) and environmental factors (e.g. cultural conditions, social relationships, living environment, and workplace) are regarded as decisive factors (Milner et al., 2018). Moreover, conditions such as anxiety and depression, heart diseases, low income or job dissatisfaction, lack of education, occupational stress, unhealthy lifestyle, and sexism can intensify mental disorders. This shows that psyche and society have mutual and influential effects on human behavior and health (Kuroda and Yamamoto, 2018; Bubonya et al., 2017).

In the analysis of physical and mental characteristics of individuals, heredity and environment are the two important and decisive factors affecting an individual (Sivris and Leka, 2015). Factors such as gender, age, or a history of familial diseases are classified as individual characteristics pertaining to heredity (Ervasti et al., 2017). Other characteristics such as behavioral and personality dimensions of employees and overseers in a workplace, occupational and professional conditions, occupational expectations, organizational executive plans, and sociocultural factors in organizations are classified as environmental factors. In major sections, they are known as characteristics of a job and organization, in which an individual is employed, as well as the governing culture of that organization (Karimi, 2017; Ghaderi Charmahini, 2017; Cheshmegan Zadeh et al., 2015).

Mental Health in Workplace

It is very important to pay attention to mental health in most aspects of life such as individual, occupational, and social aspects, among which people's occupations or careers require special attention (Ali Yas and Abedini, 2016). In fact, mental health in a workplace means resistance to the emergence of psychological distress and behavioral disorders among employees and purification of the psychological workspace to meet the goals of organizational efficiency and effectiveness (Milligan-Saville et al., 2017; Slade, 2010). Therefore, managers should try to create a healthy and stress-free workspace for employees, who can feel job security and perform their tasks happily, energetically, and eagerly without encountering any unnecessary fits of anger or confrontations (Harrison and Dawson, 2016).

Hence, the concept of organizational health is introduced. According to previous studies, the durability, survival, and adaptability of an organization depend on the presence of employees who are mentally healthy in the workplace where the organization can enhance and expand its abilities to adapt better to organizational characteristics (Keshavarz Afshar et al., 2016). However, what matters is the identification and recognition of the factors affecting employee mental health to prioritize the important factors and

remind organizational managers of which factors they can control and manage to bring better results to the organization (Reavley et al., 2014; Arakida, 2016).

Data Mining

Like other information technology concepts, data mining is interpreted differently. However, if it is employed accurately, it points to sophisticated and advanced analytical tools that can automatically discover useful patterns among the data of a data warehouse (Han et al., 2011). In general, it is fair to say that data mining is an advanced form of decision support, which generates procedures, programmed rules, and patterns without needing to ask users. In other words, data mining discloses the patterns which have been ignored by users and generates answers to questions which have never been raised (Markov and Rusell., 2009; Ali Abadi, 2018)

Hence, the goal of data mining is to extract valuable information from data with the purpose of discovering knowledge. Knowledge discovery from data is sometimes used as data mining. On the contrary, data mining is often known as the core of the process of discovering knowledge from data. In this view, data mining is regarded as a basic step of knowledge discovery and is introduced as one of the most important branches of knowledge management (Chen et al., 2017).

Data mining is performed in three ways: Supervised, Unsupervised, and Reinforced (Shahrabi and Shojaei, 2011). Supervised data mining has a predetermined goal of seeking a specific pattern. However, unsupervised data mining looks for patterns or similarities between groups of information without having a specific goal or a series of predetermined classes and patterns. There is no specific information available in reinforced data mining. In supervised data mining, there is a target variable that must be classified, estimated, and predicted. Having no target variables, unsupervised data mining is responsible for finding key patterns that do not belong to a specific variable (Kunwar et al., 2016).

The data mining process is mainly related to model development. A model refers generally to an algorithm or a series of rules which relate a series of inputs to a specific target or destination. Under the right circumstances, a model can lead to the right insight. However, many of the current problems can be classified as one of the following six categories. In other words, to turn a business problem into a data mining problem, it should be turned into one of the following data mining activities (Larose, 2014):

1. Classification
2. Estimation
3. Prediction
4. Association Rules
5. Clustering
6. Description and indexing

Empirical Background

A few studies have been conducted on the use of data mining in employee health management; however, there are very few studies regarding the mental health of employees in a particular area

pertaining to model development with the purpose of identifying and predicting the chance of being afflicted with mental disorders in the workplace. This section reviews some of these studies.

In a study entitled “Mental Health and Its Relationship with Occupational Burnout and Life Satisfaction among the Staff Employed at Military Universities in Iran”, Salimi et al. analyzed a sample of 250 employees of military universities in Iran and concluded that there was a significant relationship between job satisfaction and mental health. They also found out that there was a significant relationship between occupational burnout and mental health, reasoning that it was necessary to consider these variables both in the workplace and life (Salimi et al., 2015).

Hosseini et al. analyzed the interplay of mental health, job stress, and occupational burnout among the employees in the steel industries. According to the results of their statistical tests, mental stress was directly correlated with job stress and occupational burnout (Hosseini et al., 2016). In a paper entitled “Analysis of Mental Health Status of Employees at Health and Treatment Centers of Mazandaran Province”, Salehi concluded that disregarding mental health was an important factor reducing efficiency, wasting human resources, and causing both physical and mental complications, especially in professional services (Salehi, 2015).

According to Kuroda and Yamamoto, it is widely believed that the supervisor-employee relationship can significantly affect employee health and productivity. Nevertheless, very few studies analyzed how supervisors would affect employee health and increase productivity (Kuroda and Yamamoto, 2018).

Bubonya et al. pointed out that a major part of economic expenses would result from mental problems caused by the declined productivity of workers. They benefited from national data to analyze the relationships between mental health and efficiency at the workplace. They concluded that the absence rate of employees was nearly %5 higher in people with poor mental health than in others. Moreover, occupational conditions were considered to play a relatively important role in reducing productivity. At the same time, the effects of occupational complication and stress depended on the mental health of employees, although job security and control had positive effects on their mental health (Bubonya et al., 2017).

Conceptual Model

Considering all of this, we examined the variables introduced in previous studies as well as organizational data in the area of mental health of employees in technology-based companies, and presented the research model based on the effective aspects and characteristics in the following way and then discussed them.

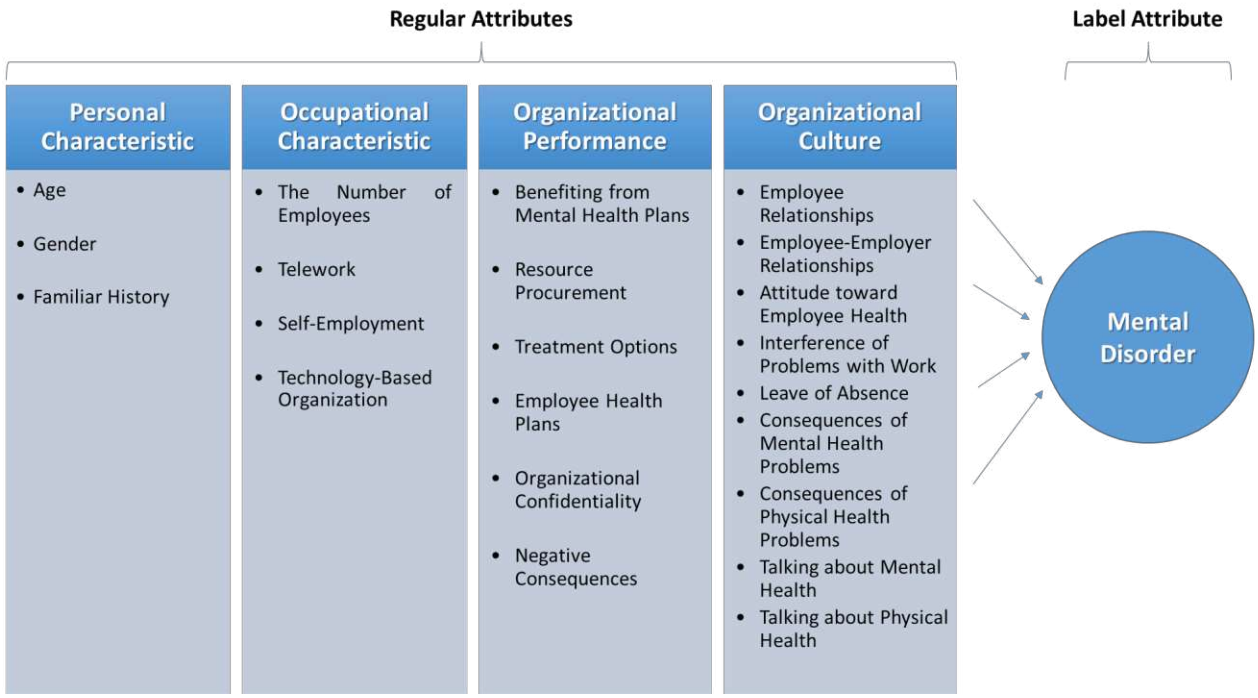

Figure 1. The proposed model

Methodology

There are several methods in the data mining process that need to be employed by the researcher to take the executive steps of the study. The most famous of these methodologies are CRISP-DM[1] and SEMMA[2] (developed by SAS Institute). IBM took a new step in this area by designing and developing its own specific research methodology. This method, designed cyclically at 10 steps, allows the researcher to move with a certain pattern at each part in order to prevent possible errors on the path (Rollins, 2015).

[1] CRoss Industry Standard Process for Data Mining
[2] Sample, Explore, Modify, Model, and Assess

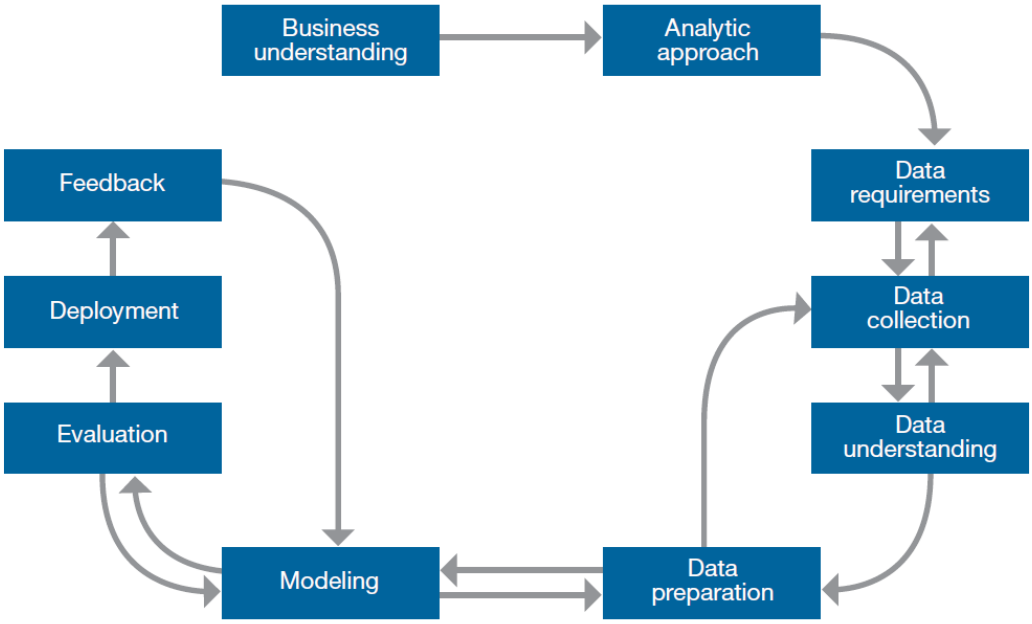

Figure 2. IBM Data Science Methodology

The IBM Data Science research method, which is a modern and practical research method, was used in this study. It is necessary to note that any organization may focus on several phases depending on its infrastructure.

A review of the above process showed that the data analysis is not just a set of techniques and methods of computational and mathematical methods, but a coherent and integrated process that begins with identifying the goals and requirements of the business environment (Phase I). Next, the problem analysis approach is adopted following the identification of the problem and possible solutions (Phase II). Then, the data are identified, collected and understood based on the requirements (Phases III, IV, and V). Next, the data are pre-processed and cleaned (Phase VI), which is one of the most important steps of the research method.

The modeling process, which has the central role of the data mining process and data analysis (Phase VII), is implemented and a model that can describe the data set is achieved. Next, the analyzed data are evaluated and validated (Phase VIII) and then the model is implemented, analyzed and developed to determine its efficiency (Phase IX). Finally, feedback is obtained from the model's performance and its potential impact, which will be the starting point of correcting and improving the model and identifying the other needs of the organization (Phase X).

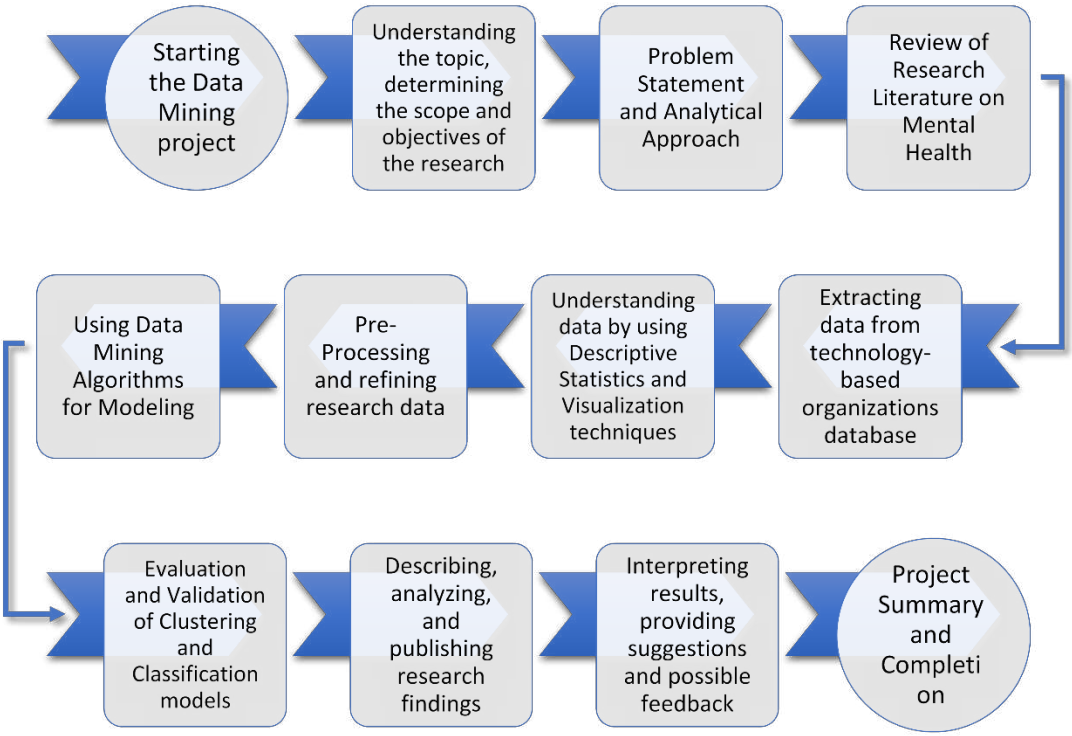

Figure 3. Executive Steps and Research Process

Data Gathering and Analysis

Data mining techniques were used in this study to identify and predict the likelihood of mental diseases among employees of technology-based organizations. The theoretical foundations of this study are formulated through the desk-based method, i.e. the study of books and articles, and the use of opinions of data mining and human resource experts. This is an applied study as it is aimed at developing practical knowledge in a particular field and moves towards scientific application (Bazargan et al., 2007).

The research scope consists of all specialized employees in technology-based companies and the data of several selected organizations in two years (between 2014 and 2016) were used for this purpose. The data of the research community were obtained through the salary and wage system, the CRM system as well as organizational questionnaires through automation. The data used to analyze the questions were extracted from the organization's database through certain algorithms and processed in the form of a data set. Desk-based and documentary research methods were also used to answer some of the questions and assumptions of this study.

There are many softwares available about data mining, including Weka, Orange, R, Python, and Rapid Miner. Rapid Miner was used in this study. It is one of the oldest and best-known tools in this respect with many advantages including high speed, simplicity of the user environment in modeling, different setting for algorithms and the aggregation of all the tools required in a package. The 3.9 version was used in this study, which is currently the latest version published by this company.

Clustering algorithms such as K-means and K-medoids, Classification algorithms such as Decision Tree, Naïve Bayes, K-Nearest Neighbors, Neural Network and other algorithms were used in this study to answer the research questions.

Research Findings

In this study, data mining was employed to extract the rules for the evaluation of employees' mental health and identify effective variables. The research data were collected from global datasets based on internal information and forms of organizations and questionnaires filled out by employees. The data were then organized and managed through several executive steps for use in model development and purgation.

In the first step, it was essential to analyze all of the existing data to extract the necessary variables. Then tables and data were merged and unified. After that, the dataset was analyzed in the number of variables and the number of elements in each variable for validation based on expert opinions. In the next steps, it was necessary to manage outliers, valueless data, repetitive data, and all of the incorrect data in general. Then a data comparison was required to compare values.

Table 1. The variables used in this study

Individual Characteristics	
Age	How old were the employees when the research data were being analyzed?
Gender	What gender were the employees?
Familial History	Was there a history of a mental disorder among their first-degree relatives?
Occupational Characteristics	
The Number of Employees	How many employees was there in the organization where participants were employed?
Telework	Were the participants employed in a telework position? Or were they present at the organization?
Self-Employment	Were the participants self-employed? Or were they employed by an organization?
Technology-Based Organization	Were organizational activities based on or related to technology?
Organizational Performance	
Benefiting from Mental Health Plans	Does the organization provide the conditions for benefiting from mental health plans?
Resource Procurement	Does the organization provide the resources required to train and identify employees in relation to mental health problems?
Treatment Options	Does the organization provide employees with treatment

	options for mental health problems?
Employee Health Plans	Does the organization regard mental health as a part of employee health plans?
Organizational Confidentiality	Should people's identities be kept confidential if they benefit from mental health plans?
Negative Consequences	Have there ever been any obvious negative consequences of mental health problems of employees at the workplace?
Organizational Governing Culture	
Employee Relationships	Are employees willing to share their mental health problems with each other?
Employee-Employer Relationships	Are employees willing to share their mental health problems with their employers?
Attitude toward Employee Health	Is the employee's physical-mental health equality assumption complied with at the organization?
Interference of Problems with Work	Do employees feel that their mental health problems interfere with their work?
Leave of Absence	Is it easy for employees to take a leave of absence to deal with their mental health problems?
Consequences of Mental Health Problems	Do employees think that talking with employers about mental health problems will have negative consequences?
Consequences of Physical Health Problems	Do employees think that talking with employers about physical health problems will have negative consequences?
Talking about Mental Health	Is it possible in the organization to talk with employees about mental health and relevant problems?
Talking about Physical Health	Is it possible in the organization to talk with employees about physical health and relevant problems?

Data Preparation

Step 1) Merging Dataset and Selecting Data:

In the data selection step, it is necessary to select the dataset which is directly related to employee mental health from a large amount of data and give it to the pre-processing step. Then, data pre-processing includes different consecutive operations, which were performed in this study for the research problem and data quality.

Step 2) Resolving the Problem of Missing Values:

There are four methods for managing missing values: deleting, estimating, ignoring, and replacing.

In this dataset, the "Replace Missing Values" operator was employed in the missing values section to purge data. After executing the operator and determining the software outputs, it was evident that there were no missing data in any of the variables in outputs. Therefore, data validity was guaranteed in this section.

Step 3) Resolving the Problem of Repetitive Data:

The repeated data are those records of the dataset which have no new information value. In fact, they include excessively repetitive or totally similar information. After data were recalled, the "Remove Duplicates" operator was selected. After execution, it was evident that some existing data were deleted because they were repetitive.

Step 4) Resolving the Problem of Outliers:

The "Detect Outlier" operator is responsible for detecting and identifying outliers in the software. Outliers have negative effects on the model performance. The operator detects outliers in a dataset concerning the number of records and their distances from K of their closest neighbors.

Modeling

This step includes a series of data mining techniques and a modeling approach implemented on the dataset pre-processed and prepared in the previous step. The goal of this step is to finally develop a model that can be employed to predict what individuals are likely to have or lack mental disorders.

Classification and Prediction Models

These methods can be employed to determine the positive or negative probability of being afflicted with mental disorders based on certain factors such as employee background and performance and their conditions in different periods. This study benefited from conventional methods such as Decision Tree, ID3 Tree, CHAID Tree, Random Forest, K-Nearest Neighbor, naïve Bayes, and Neural Networks for modeling and prediction.

According to the table of Classification algorithms comparison, the performance of each algorithm can be predicted in identifying the rules hidden in data and predicting the behavior of new data. In fact, the output of Classification algorithms is a model, new data can be entered in the model with no target variables to predict data behavior.

Table 2. Comparison of Classification Algorithms Used in this study

	Accuracy (%)	**Error** (%)	**Recall** (%)	**Precision** (%)
Decision Tree	84.99	15.01	84.49	85.70
ID3	83.91	16.09	83.37	84.07
CHAID	80.97	19.03	80.40	80.87
Random Forest	82.57	17.43	82.33	82.62
K-NN	71.58	28.42	71.48	71.46
Naïve Bayes	78.02	21.98	78.03	77.94
Neural Network	77.21	22.79	77.24	77.14

Comparing the accuracy of algorithms can lead to the point that Tree inferences have generally outperformed other methods since the prediction accuracy of Tree algorithms was over %80 when their

error rates were below %20. Among the research algorithms, the Decision Tree algorithm (%84.99) was ranked the top algorithm which could be employed to predict future data. The ID3 algorithm was ranked second (%83.91), and the Random Forest algorithm was ranked third (%82.57). Finally, the CHAID algorithm was ranked fourth (%80.97).

Then the Naïve Bayes, Neural Network, and K-Nearest Neighbor algorithms were ranked in the next positions with accuracy rates of %78.02, %77.21, and %71.58, respectively. The error rate of each of these algorithms ranged from %20 to %30. Therefore, they were weaker than the Tree algorithms. As a result, it is fair to conclude that Tree inference models outperformed other models among Classification models and that Memory-Based reasoning performed weakly. Accordingly, the Decision Tree and K-NN algorithms were ranked first and last, respectively. Thus, the Decision Tree algorithm can be employed to answer the research questions on the identification of effective factors and their effects on the target variable.

Decision Tree Algorithm

The Decision Tree is a model that can be implemented by RapidMiner. The most important feature of this model is that it develops a binary tree. In other words, records are divided into two branches in each node. Data should be put into the relevant operator to use this algorithm. After the output is obtained, the model is tested. Figure 4 shows the model implementation procedure. The model parameters are as follows:

Maximal Depth (This parameter is equal to the maximum level that the decision tree can have and depends on specific factors such as data size and type in the dataset) = 20

Confidence (This parameter determines the confidence interval used for calculating tree-pruning errors) = 0.5

Minimal Gain (This parameter determines what minimal value a node should benefit from before division) = 0.1

Minimal Leaf Size (This parameter indicates the minimal size of a leaf based on the data put in its subset) = 2

Minimal Size for Split (This parameter shows the minimal size for a split based on the data put in its subset) = 2

Number of Prepruning Alternatives (This parameter determines the number of alternative nodes for a split when it is blocked in a node due to the prepruning process) = 14

It should be mentioned the optimal values of these parameters and those of other algorithms were obtained by the “Optimize Parameters” operator through evolutionary algorithms.

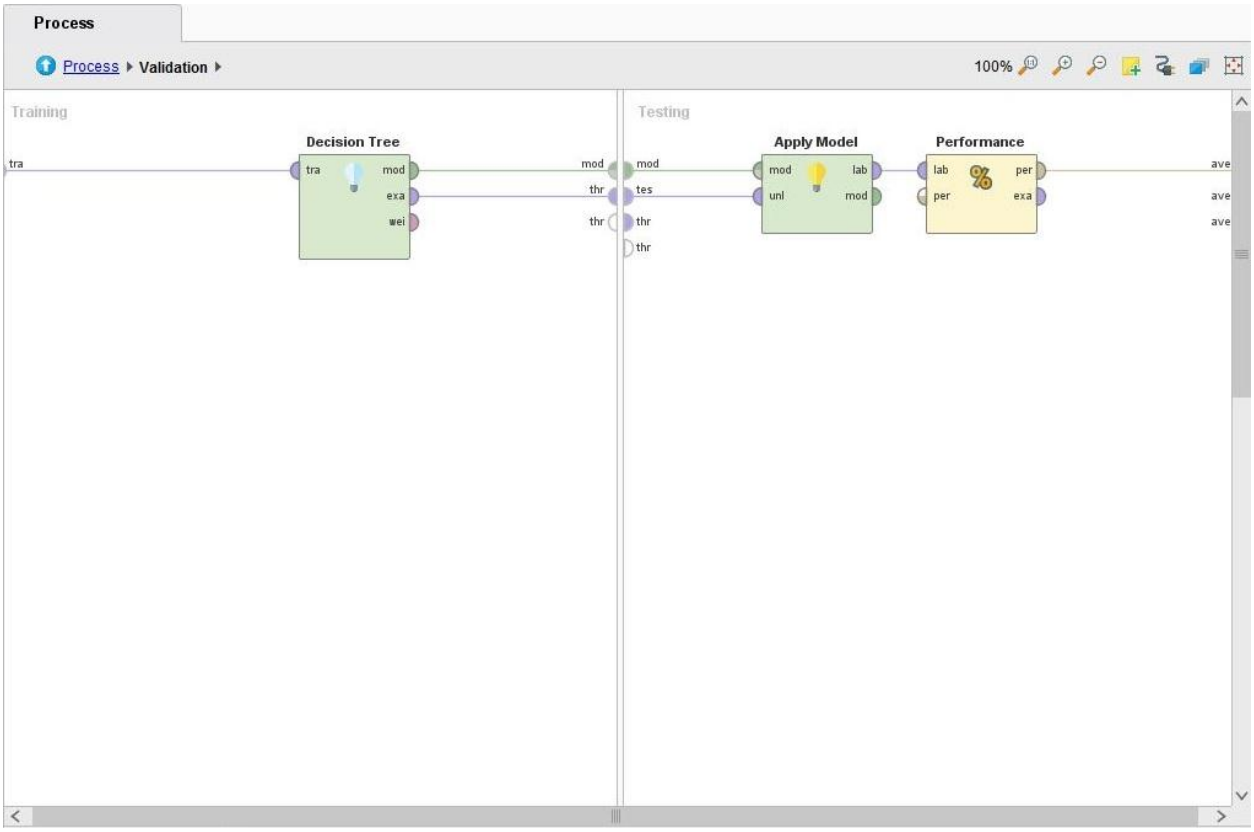

Figure 4. Decision Tree Model Implementation

After the Decision Tree algorithm was implemented, the following outputs were obtained:

accuracy: 84.99%

	true Yes	true No	class precision
pred. Yes	183	40	82.06%
pred. No	16	134	89.33%
class recall	91.96%	77.01%	

Figure 5. Decision Tree Algorithm Evaluation

According to Figure 5, this model managed to identify some of the individual characteristics, occupational characteristics, organizational performance, and organizational governing culture as the most important factors in mental health. The Classification Accuracy rate (ratio of total predictions labeled correctly by the classifier) was %84.99, and the Classification Error rate was %15.01. Moreover, the Recall rate (ratio of positive cases identified correctly) was %84.49, and the Precision rate (ratio of records which are labeled positively and whose class is really positive) was %85.70. Accordingly, it is possible to determine how to separate data by considering the target variable.

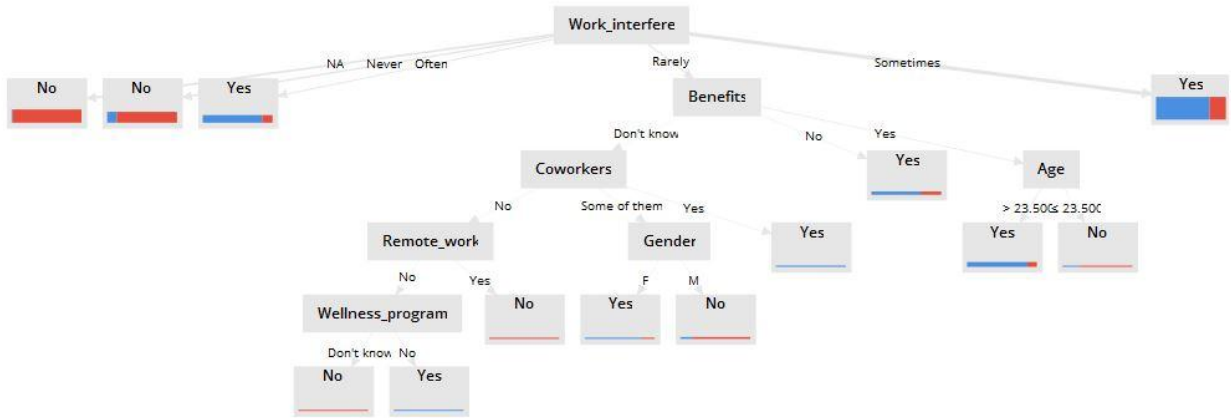

Figure 6. Decision Tree Algorithm Graph

Tree

```
Work_interfere = NA: No {Yes=4, No=257}
Work_interfere = Never: No {Yes=29, No=181}
Work_interfere = Often: Yes {Yes=120, No=21}
Work_interfere = Rarely
|   Benefits = Don't know
|   |   Coworkers = No
|   |   |   Remote_work = No
|   |   |   |   Wellness_program = Don't know: No {Yes=0, No=2}
|   |   |   |   Wellness_program = No: Yes {Yes=3, No=0}
|   |   |   Remote_work = Yes: No {Yes=0, No=4}
|   |   Coworkers = Some of them
|   |   |   Gender = F: Yes {Yes=4, No=1}
|   |   |   Gender = M: No {Yes=3, No=15}
|   |   Coworkers = Yes: Yes {Yes=6, No=0}
|   Benefits = No: Yes {Yes=33, No=14}
|   Benefits = Yes
|   |   Age > 23.500: Yes {Yes=71, No=11}
|   |   Age ≤ 23.500: No {Yes=1, No=3}
Work_interfere = Sometimes: Yes {Yes=355, No=107}
```

Figure 7. Rules Generated by Decision Tree (Effective Factors)

Clustering Models

Clustering methods are employed to obtain information on data distribution and processing. In this study, K-Means and K-Medoids algorithms were utilized for Clustering and data analysis.

Although the K-Means algorithm has a relatively appropriate accuracy in separating clusters after implementation, data averaging is not significant because the data means of each cluster are regarded as the cluster center in this algorithm. The available data of this study are mainly of two or multiple types (Binomial or Polynomial); Therefore, cluster centers could not be found among real data, something which makes the algorithm vulnerable to outliers. The K-Medoids algorithm has been proposed to cope with this problem. It determines cluster centers based on the existing main data.

K-Medoids Algorithm

This Clustering method is one of the simplest but most useful and well-known algorithms for unsupervised learning. In the K-Medoids algorithm, K members (K shows the number of clusters) are first selected as cluster centers from N members at random. Then N-K remaining members are allocated to the nearest cluster. After all of the members are allocated, cluster centers are recalculated and allocated to clusters with respect to new centers. This process is kept on until cluster centers remain fixed.

This method has regulatory parameters and switches which can be changed to optimize the model. These parameters are as follows:

Number of Branches (K)

Max Run (This parameter shows the maximum number of times the K-Medoids algorithm is run with random initialization from the starting point)

Measure Types (This parameter is employed to select measurement types to find the nearest neighbors)

Max Optimization Steps (This parameter determines the maximum number of iterations performed for one execution of the K-Medoids algorithm)

The optimal values of these parameters were determined with respect to expert opinions and algorithm conditions. In this regard, selecting the number of branches is the most important Clustering factor. If the Clustering algorithm has the longest Avg. within Centroid Distance and the lowest Avg. The similarity of each cluster with other clusters, known as the Davies Bouldin index, then it can be considered as the best Clustering outcome. Therefore, the first step was to select the best Clustering process for the available research data.

According to Table 3 and the Scree Plot diagram (Figure 8), the most important values of the abovementioned indices were obtained in the K=3 Clustering process. As a result, Clustering with three branches can be the best choice for the research data. Accordingly, the third cluster (1.725) is more homogenous than the first cluster (2.041) and the second cluster (2.524).

Table 3. K-Medoids Clustering with 2 up to 20 Branches

Number of Clusters	Davies Bouldin	Avg. Within Centroid Distance
2	0.034	2.983

3	0.053	1.965
4	0.073	1.745
5	0.076	1.591
6	0.080	1.448
7	0.082	1.402
8	0.083	1.396
9	0.076	1.303
10	0.074	1.259
11	0.078	1.222
12	0.075	1.210
13	0.080	1.182
14	0.076	1.169
15	0.078	1.128
16	0.078	1.121
17	0.081	1.130
18	0.076	1.093
19	0.076	1.067
20	0.077	1.058

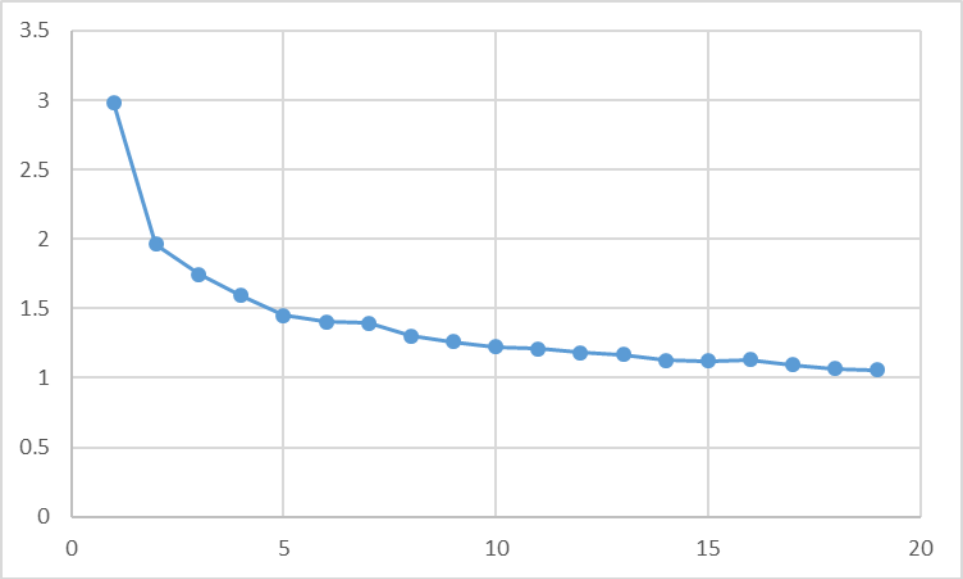

Figure 8. Scree Plot Diagram for K-Medoids

Table 4. Euclidean Distance in K-Medoids and the Number of Records Existing in each Cluster

The risk-exposed group	2.041	552 items

The potentially afflicted group	2.524	156 items
The immune group	1.725	537 items
	Avg. within centroid distance 1.965	**Total number of items 1245**

According to the outputs, the following results can be obtained in each cluster:

The risk-exposed group (Cluster 1): The first cluster, known as "The risk-exposed group" or "The high-risk group" hereinafter, includes a higher percentage of employees with mental disorders during the employment period than the other two clusters.

The potentially afflicted group (Cluster 2): Obtained from the data mining process, the second cluster is identified as "The potentially afflicted group" or "The medium-risk group," in which it is impossible to determine the chance of affliction or non-affliction with mental disorders certainly on the contrary to the first and third clusters having specific features for differentiation of members.

The immune group (Cluster 3): The final cluster, known as "The immune group" or "The low-risk group" hereinafter, includes members whose chances of being afflicted with mental disorders are very remote. In fact, their individual, occupational, and organizational conditions and workplace culture help them face a low risk of mental health problems.

Discussion

The Decision Tree, ID3, CHAID, Random Forest, K-NN, Naïve Bayes, and Neural Network algorithms were employed in two groups, i.e. "having mental health problems" and "lacking mental health problems," to determine the factors affecting the dataset of employees with mental health problems. These algorithms managed to predict the chances of affliction with mental disorders with an average accuracy of %79.89 and an average error of %20.11.

The Decision Tree algorithm outperformed the other algorithms with an accuracy of %84.99 in identifying the factors affecting the dataset; Therefore, it was selected as the winner. This algorithm succeeded in predicting and identifying seven factors as the most important factors affecting the chances of affliction with mental health disorders among employees working at technology-based organizations. These factors were "Work Interference", "Benefit from Mental Health Plans", "Age", "Coworkers", "Gender", "Telework", and "Employee Health Plans".

The ID3, CHAID, and Random Forest algorithms operated nearly similarly as the decision tree algorithm, although they were ranked in the next priorities due to the lower rates of accuracy and higher complexities. Other Classification algorithms such as Naïve Bayes, Neural Networks, and K-NN were ranked in the next positions. Considering the high accuracy and proper output of the Decision Tree method in classifying employees with mental health problems, it will be employed to analyze effective factors and predict new data. This algorithm is now discussed in detail.

The K-Means and K-Medoids algorithms were employed to identify clusters and obtain information on data distribution and processing. Considering the nature and types of research data, the K-Medoids algorithm produced more accurate and realistic results. According to the algorithm evaluation indices, it was decided that the most optimal value of the average mean distance of each cluster and their average similarity criteria were obtained in three clusters. The Clustering model was obtained concerning the fact that three clusters were selected. Furthermore, the third cluster was the most homogenous of all. After that, the first and second clusters were ranked in the next positions.

These three clusters were separated and identified as “The high-risk group”, “The medium-risk group”, and “The low-risk group”, the characteristics of which were discussed in Research Results. Identifying and analyzing six factors i.e. “Gender”, “Familial History”, “The Number of Employees”, “Work Interference”, “Telework”, “Coworkers”, it is possible to determine which cluster a member should be put into.

Conclusion and Future Research Directions

Identifying and analyzing the mental health of every member of society requires research into their personal, occupation, and social lives in which many individual and environmental factors play important roles. Since the modern human’s life is entangled with economic activities, the majority of people devote a part of their lives to work and spent many hours in the workplace every day. Therefore, not only is it important to analyze the mental health of individuals in a personal aspect, but it is also necessary from organizational and social perspectives.

Evidently, the presence of various personal, social, cultural, and economic factors and variables can make the process of analyzing and predicting the mental health of employees difficult in many countries now. The insufficient knowledge of these factors and their effects on people’s mental health have made organizational managers unable to provide their employees with mental health properly. Disregarding this problem can have dire consequences leaving irreversible impacts on organizations and destructive effects on the lives of employees, especially in technology-based organizations where expert manpower plays a precious and vital role.

Therefore, this study collected effective factors by reviewing the research literature and analyzing available data through data mining techniques to identify the most important factors affecting the mental health of employees. For this purpose, data Classification and Clustering techniques were employed to identify the most important factors affecting the target variable, i.e. mental health or mental disorders of employees. Then the resultant clusters and classes can be utilized to predicting the chances of affliction with mental health problems.

The following variables were identified by the top Classification and Clustering algorithms used in this research:

- Gender
- Age
- Familial history of mental disorders
- Interference of problems with work activities in organizations
- Telework or physical presence at organizations

- Coworkers and working groups stationed in organizations
- The number of employees working in organizations
- Organizing plans for maintaining the mental health of employees
- Benefit from mental health plans

Hence, it is fair to state that the above variables are considered as the most important factors pertaining to individual, occupational, performance and cultural characteristics of the workplace. They must be taken into account by senior managers and human resources deputies of organizations. Accordingly, the mental health of employees is significantly related to the above variables. Considering this correlation and using different mathematical and statistical techniques like data mining methods, it is possible to identify the latent relationships of data and determine to what extent each variable predicts the target variable, i.e. mental health.

Since prediction models are obtained through Classification techniques in data mining, this study benefited from different Tree inference methods, Memory-Based reasoning techniques, Bayesian theory, and Neural Networks in the form of different Classification algorithms. The Decision Tree technique outperformed all of the other methods with a relatively appropriate accuracy (nearly %85) and managed to predict the chances of affliction or non-affliction with mental disorders through the input data.

Therefore, it is possible to predict the need for mental treatment for expert employees at technology-based companies by using a dataset of individual aspects pertaining to an employee's physiological conditions, occupational aspects pertaining to the general conditions of a specific job, performance aspects pertaining to organizational decisions and executive operations, and finally cultural aspects pertaining to the governing atmosphere of an organization in relation to the mental health of employees.

Since it is difficult to analyze subjective and abstract variables (which are not mainly stored and documented in numbers and figures on the contrary to most of the organizational data and information) through conventional calculation methods, data science and data mining techniques can be employed to bring about unique and unprecedented results. Thus, it is recommended to benefit from data mining approaches in similar cases such as the physical health of employees, talent management, performance evaluation, working life quality, personality identification at the workplace, and organizational behavior management.

At the same time, it is advisable that organizations operating in mental health and treatment such as the Psychology and Counseling Organization of Iran, Iran Psychology Association, and Iran's Ministry of Health, Treatment, and Medical Education, and other affiliated organizations collect, integrate, and register mental health data in databases. Technology-based organizations are advised to record and store personal characteristics and organizational workplace information from the time when new employees enter organizations in order to seriously and continuously monitor employee health plans through data-based methods and organizational human resource manager to the best of their abilities.

It is generally difficult and highly complicated to identify and improve the aspects of mental health; therefore, it is necessary to implement extensive and comprehensive research projects in this matter. Since the factors affecting the mental health of employees are not limited to organizational and occupational conditions, organizations should put other aspects such as social environment, familial

conditions, economic status, political factors, and cultural characteristics of society into their plans because these factors have significant effects on the mental health hygiene of people.

Therefore, paying attention to the mental health and hygiene of employees from the personal and environmental aspects will be among the top priorities and competitive advantages of top organizations in the business. This requires special attention paid by the senior managers of organizations. Finally, organizational managers are advised to take account of the importance of mental health as well as the physical health of their employees and execute operational and research plans to improve them.

References

Abraham, T. H., Koenig, C. J., Zamora, K., Hill, C., Uddo, M., Kelly, A. P., ... & Seal, K. H. (2017). Situating mental health work in place: Qualitative findings from interviews with Veterans in Southeastern Louisiana and Northern California. Health & place, 47, 63-70.

Ali Abadi, M. (2018). A Review of Recent Applications of Data Mining, *6th National Conference on Innovations in Industrial Management and Engineering with Emphasis on Entrepreneurship in Industry*, Payame Noor University, Tehran.

Ali Yas, S. & Abedini, M. (2016). The Relationship between Quality of Work Life and Mental Health among Hormozgan University of Jihad. *National Conference on Development Management (Challenges and Solutions in Public and Private Organizations)*, Bandar Abbas.

Alonso, S. G., de La Torre-Díez, I., Hamrioui, S., López-Coronado, M., Barreno, D. C., Nozaleda, L. M., & Franco, M. (2018). Data mining algorithms and techniques in mental health: a systematic review. *Journal of medical systems*, *42*(9), 161.

Arakida, M. (2006). Examination of employees' mental health status and related factors in a manufacturing company. In International Congress Series (Vol. 1294, pp. 153-156). Elsevier.

Bazargan, A., Sarmad, Z. & Hejazi, E. (2007). Research Methods in Behavioral Sciences. 13th Edition, Agah Publications, Tehran.

Bubonya, M., Cobb-Clark, D. A., & Wooden, M. (2017). Mental health and productivity at work: Does what you do matter? *Labour Economics*, *46*, 150-165.

Chen, M., Hao, Y., Hwang, K., Wang, L., & Wang, L. (2017). Disease prediction by machine learning over big data from healthcare communities. *Ieee Access*, *5*, 8869-8879.

Cheshmegan Zadeh, S., Zarei, M. & Elmian, G. (2015). Investigating the Importance of Employees Mental Health at Work. *First Scientific Research Conference on Strategies for Development and Promotion of Science Education in Iran*.

Ervasti, J., Joensuu, M., Pentti, J., Oksanen, T., Ahola, K., Vahtera, J., & Virtanen, M. (2017). Prognostic factors for return to work after depression-related work disability: a systematic review and meta-analysis. *Journal of psychiatric research*, *95*, 28-36.

Esmaeeli Nasr Abadi, K., Zabihi, M. & Shekari, G. (2015). The Relationship between Mental Health with Salary Satisfaction and Workplace Satisfaction with Regarding the Moderating Role of Organizational Commitment: Case Study National Bank of Iran. *International Conference on Management Economics and Social Sciences*, Spain.

Ghaderi Charmahini, Z. (2017). Evaluation of Mental Health and its Affecting Factors. *The First Regional Conference on the Development of Community-Based Nursing and Midwifery Sciences*, Islamic Azad University, Najaf Abad.

Han, J., Kamber, M. & Pei, J. (2011). Data Mining: Concepts and Techniques (Third Edition). SanFrancisco. Morgan Kaufmann Publishers, CA, USA.

Harrison, J., & Dawson, L. (2016). Occupational health: Meeting the challenges of the next 20 years. Safety and health at work, 7(2), 143-149.

Hosseini, S., habibi, E. Barakat, S. and et al. (2016), investigating the relationship of mental health with job stress and burnout in workers of metal industries, international journal of educational and psychological researchs, july 05.

Hunt, J., & Eisenberg, D. (2010). Mental health problems and help-seeking behavior among college students. Journal of adolescent health, 46(1), 3-10.

Jorm, A. F. (2000). Mental health literacy: Public knowledge and beliefs about mental disorders. The British Journal of Psychiatry, 177(5), 396-401.

Karimi, M. (2017). A Quantitative Model of the Relationship between Mental Health and Job Burnout Considering the Moderating Role of Spiritual Intelligence. *3rd International Conference on Management, Accounting and Knowledge-Based Economics with Emphasis on Resistance Economics*, Allameh Majlesi University, Tehran.

Keshavarz Afshar, M., Amini, K. & Imani, A. (2016). A Survey of Mental Health Level and its Relationship with Some Background Variables of Nurses Working in Zanjan Teaching Hospitals, *Zanko Journal of Medical Sciences*, Kurdistan University of Medical Sciences.

Kessler, R. C., & Üstün, T. B. (2004). The world mental health (WMH) survey initiative version of the world health organization (WHO) composite international diagnostic interview (CIDI). International journal of methods in psychiatric research, 13(2), 93-121.

Khaghani Zadeh, M., Sirati Neer, M., Abdi, F. & Kaviani, H. (2016). Evaluation of Mental Health in Nurses Working in Teaching Hospitals Affiliated to Tehran University of Medical Sciences. *Journal of Mental Health Principles*, 8(32), 141-148.

Kunwar, V., Chandel, K., Sabitha, A. S., & Bansal, A. (2016). Chronic Kidney Disease analysis using data mining classification techniques. In *2016 6th International Conference-Cloud System and Big Data Engineering (Confluence)* (pp. 300-305). IEEE.

Kuroda, S., & Yamamoto, I. (2018). Good boss, bad boss, workers' mental health and productivity: Evidence from Japan. *Japan and the World Economy*.

Larose, D. (2014). Discovering Knowledge in Data: An Introduction to Data Mining. 2nd Edition. John Wiley-Interscience.

Markov, Z. & Rusell, I. (2009). An Introduction to the WEKA data mining system. Proceedings of the 11th annual SIGCSE conference on Innovation and technology in computer science education, 367-368.

Milligan-Saville, J. S., Tan, L., Gayed, A., Barnes, C., Madan, I., Dobson, M., & Harvey, S. B. (2017). Workplace mental health training for managers and its effect on sick leave in employees: a cluster randomised controlled trial. *The Lancet Psychiatry*, *4*(11), 850-858.

Milner, A., King, T., LaMontagne, A. D., Bentley, R., & Kavanagh, A. (2018). Men's work, Women's work, and mental health: A longitudinal investigation of the relationship between the gender composition of occupations and mental health. *Social Science & Medicine*, *204*, 16-22.

Moosavi, F. (2014). Mental Health of Employees in Workplace. *2nd International Conference on Management of Challenges and Solutions*, Shiraz.

Reavley, N. J., Ross, A., Martin, A., LaMontagne, A. D., & Jorm, A. F. (2014). Development of guidelines for workplace prevention of mental health problems: A Delphi consensus study with Australian professionals and employees. *Mental Health & Prevention*, *2*(1-2), 26-34.

Rollins, J.B. (2015). Foundational Methodology for Data Science. *Big Data University*, http://bigdatauniversity.com/bdu-wp/bdu-course/data-science-methodology.

Salehi, M. (2015). Determining the Mental Health Status of Health Care Workers in Mazandaran Province, Istanbul.

Salimi SH. Azad M. Abedi M. (2015), Mental health and its relationship with job burnout and life satisfaction in staff at a military university, Iranian journal of military medicine, V. 14, No. 4, Winter, pp. 289- 29.

Shahrabi, J. & Shojaei, A. (2011). Advanced Data Mining: Concepts and Algorithms. Industrial Amir Kabir University, Tehran.

Sivris, K. C., & Leka, S. (2015). Examples of holistic good practices in promoting and protecting mental health in the workplace: current and future challenges. *Safety and health at work*, *6*(4), 295-304.

Slade, M. (2010). Mental illness and well-being: the central importance of positive psychology and recovery approaches. BMC health services research, 10(1), 26.

Figures

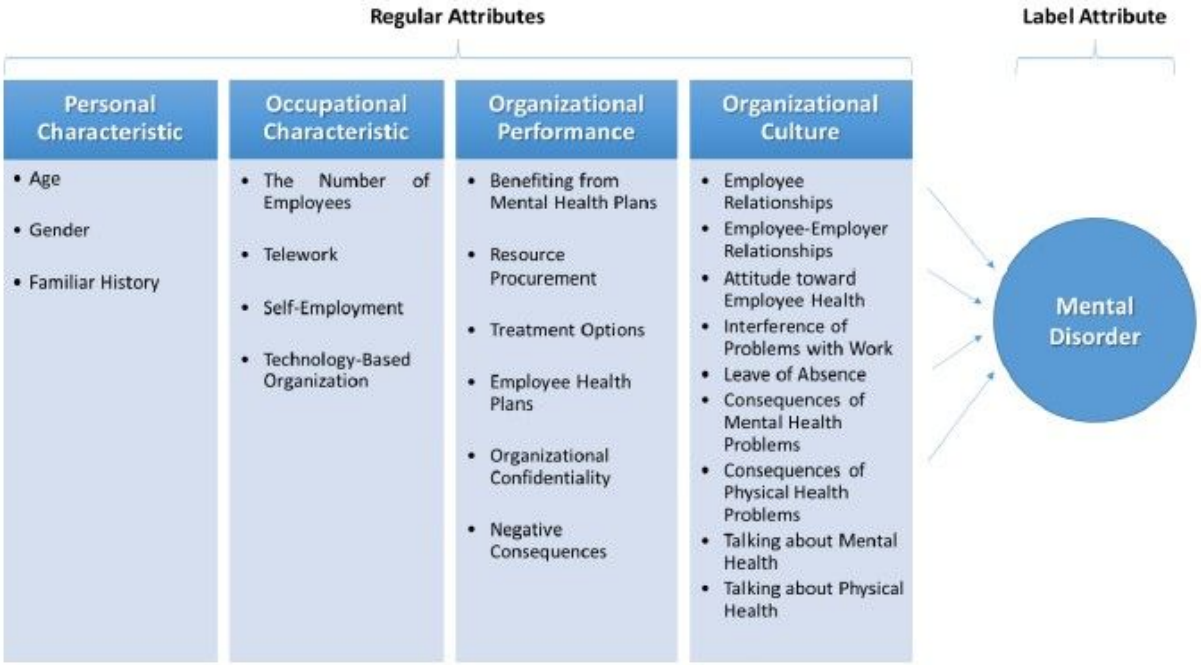

Figure 1. The proposed model

Figure 1

The proposed model

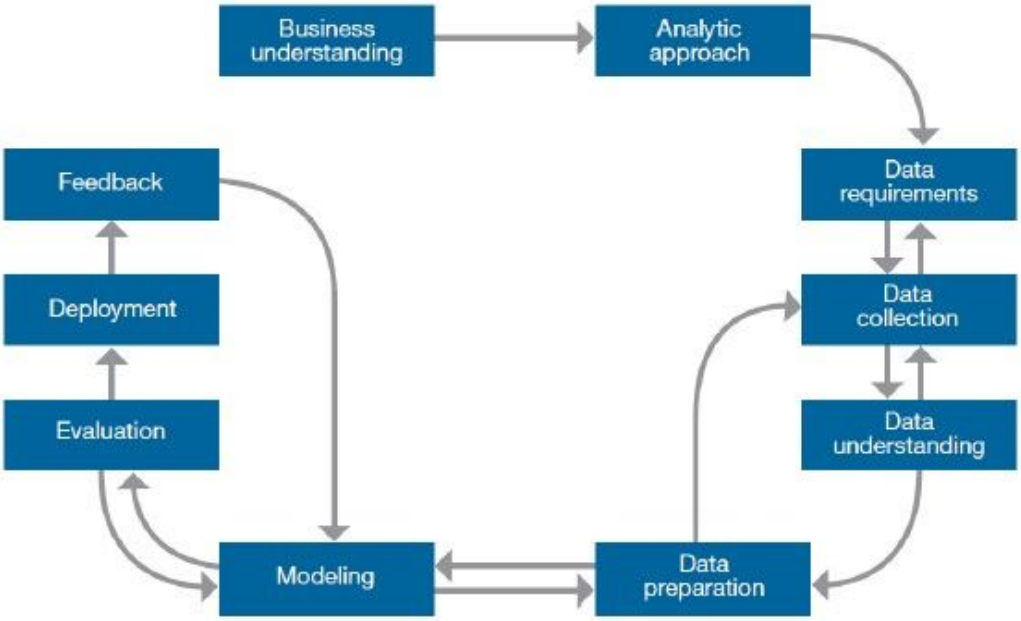

Figure 2. IBM Data Science Methodology

Figure 2

IBM Data Science Methodology

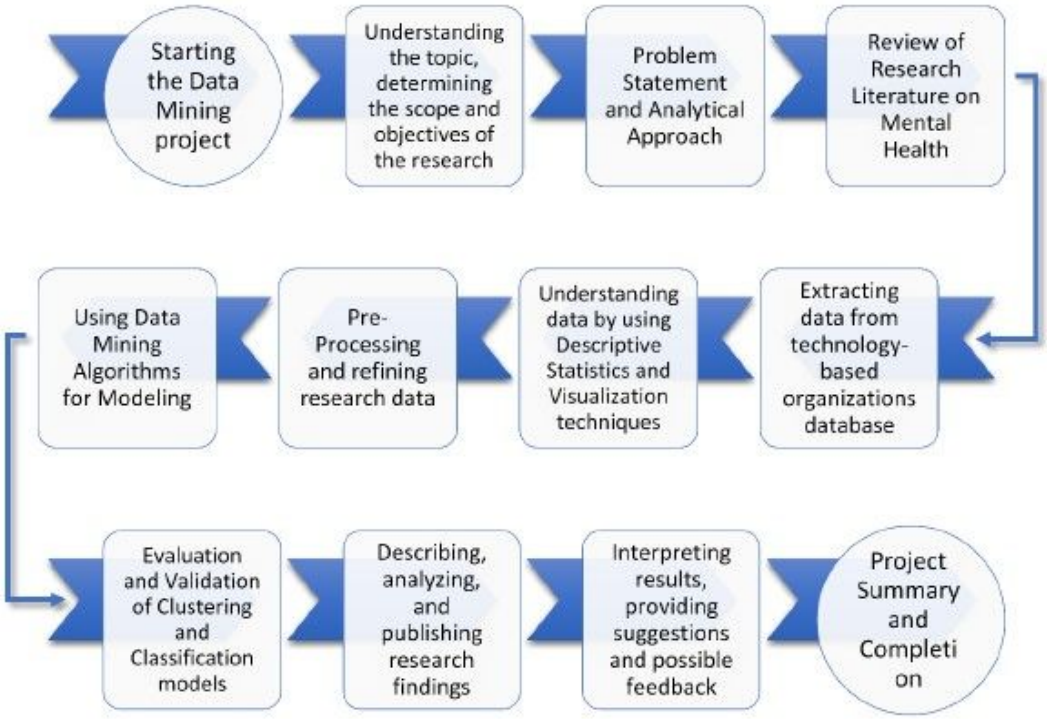

Figure 3. Executive Steps and Research Process

Figure 3

Executive Steps and Research Process

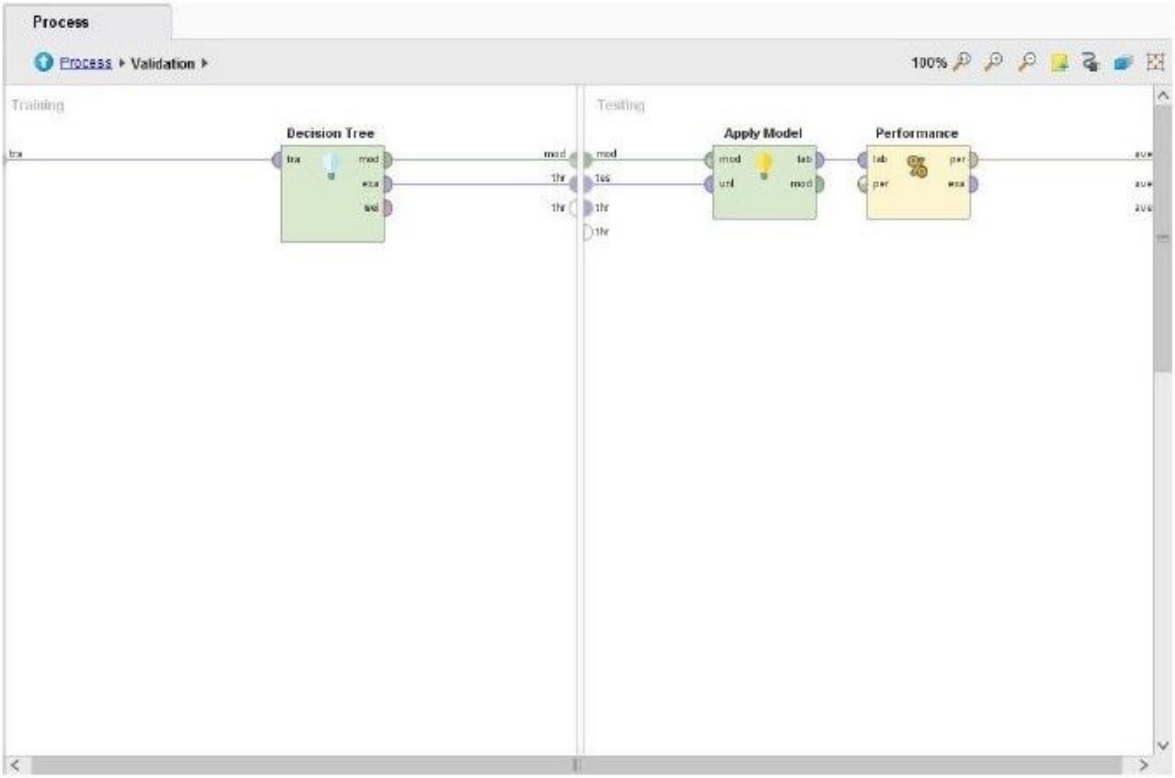

Figure 4. Decision Tree Model Implementation

Figure 4

Decision Tree Model Implementation

	true Yes	true No	class precision
pred. Yes	183	40	82.05%
pred. No	16	134	89.33%
class recall	91.98%	77.01%	

Figure 5. Decision Tree Algorithm Evaluation

Figure 5

Decision Tree Algorithm Evaluation

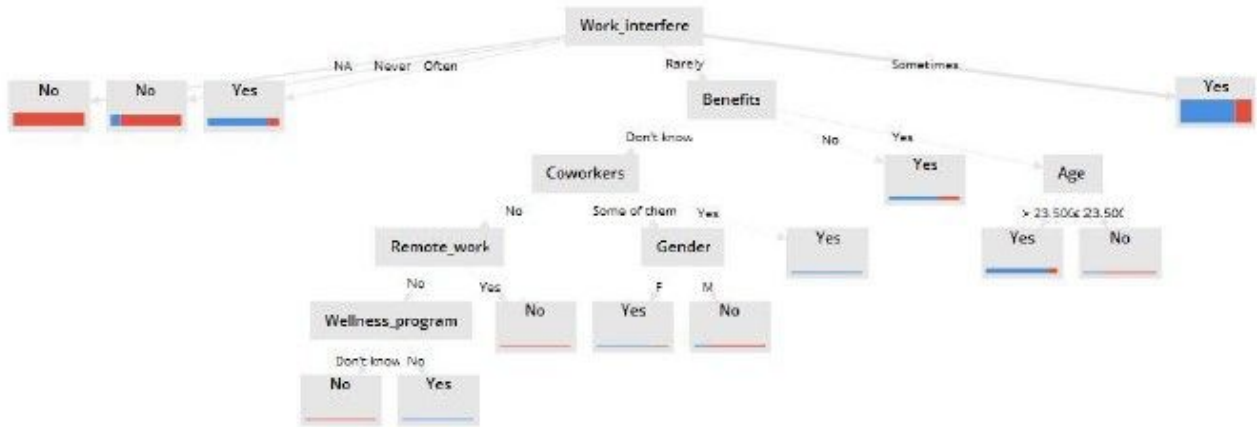

Figure 6

Decision Tree Algorithm Graph

Tree

```
Work_interfere = NA: No {Yes=4, No=257}
Work_interfere = Never: No {Yes=29, No=181}
Work_interfere = Often: Yes {Yes=120, No=21}
Work_interfere = Rarely
|   Benefits = Don't know
|   |   Coworkers = No
|   |   |   Remote_work = No
|   |   |   |   Wellness_program = Don't know: No {Yes=0, No=2}
|   |   |   |   Wellness_program = No: Yes {Yes=3, No=0}
|   |   |   Remote_work = Yes: No {Yes=0, No=4}
|   |   Coworkers = Some of them
|   |   |   Gender = F: Yes {Yes=4, No=1}
|   |   |   Gender = M: No {Yes=3, No=15}
|   |   Coworkers = Yes: Yes {Yes=6, No=0}
|   Benefits = No: Yes {Yes=33, No=14}
|   Benefits = Yes
|   |   Age > 23.500: Yes {Yes=71, No=11}
|   |   Age ≤ 23.500: No {Yes=1, No=3}
Work_interfere = Sometimes: Yes {Yes=355, No=107}
```

Figure 7. Rules Generated by Decision Tree (Effective Factors)

Figure 7

Rules Generated by Decision Tree (Effective Factors)

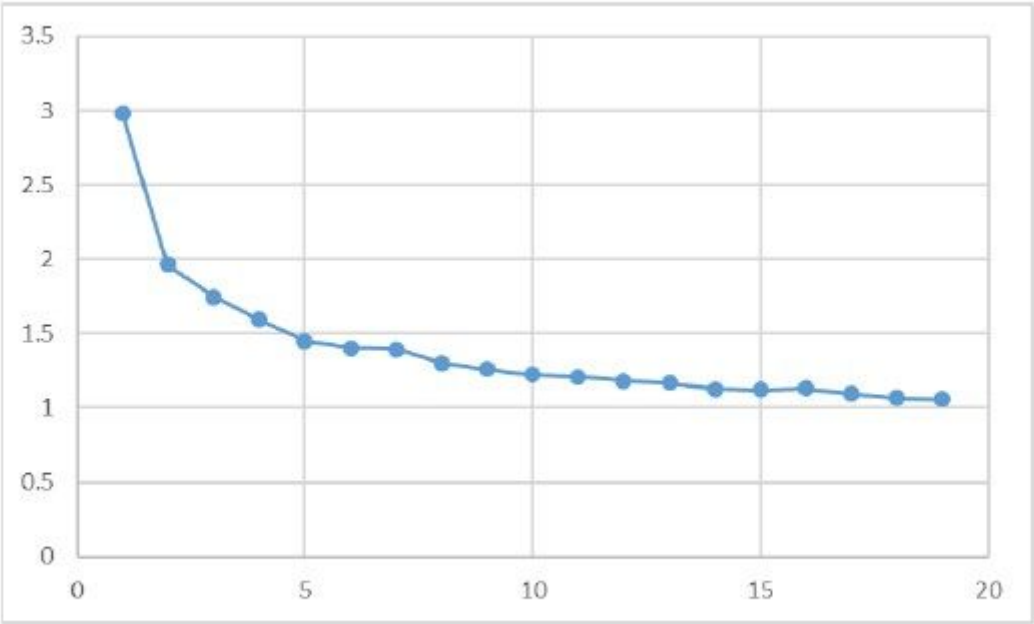

Figure 8. Scree Plot Diagram for K-Medoids

Figure 8

Scree Plot Diagram for K-Medoids