International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-1, November 2019

Pregnancy Leave Wages Compensation based on Data Mining: Clients of Iran's Social Security Organization

Amir Rajaei, Mahshid Sedighi

Abstract: ***Data mining is an interdisciplinary science which exploits different methods including statistics, pattern recognition, machine learning, and database to extract the knowledge hidden in huge datasets. In this paper, we sought to develop a model for paying pregnancy period wages compensation to the Social Security Organization (SSO) clients by using data mining techniques. The SSO is a public insurance organization, the main mission of which is to cover the stipendiary workers (mandatory) and self-employed people (optional). In order to develop the proposed model, 5931 samples were selected randomly from 11504 clients. Then the K-Means clustering algorithm was employed to divide data into cluster 1, consisting of 2732 samples, and cluster 2, consisting of 3199 samples. In each cluster, the data were divided into training and test sets with a ratio of 90 to 10. Then a multi-layer perceptron neural network was trained separately for each cluster. This paper utilized the MLP network model. The tanh transfer function was used as the activation function in the hidden activation layer. Numerous tests were conducted to develop the best neural network structure with the lowest error rate. It consisted of two hidden layers. There were 5 neurons in the first layer and 4 neurons in the second. Therefore, the neural network structure was in the 5-4-1 format. Finally, the best model was selected by using the error evaluation method. The MAPE and R2 criteria were employed to evaluate the proposed model. Regarding the test data, the result was 0.96 for cluster 1 and 0.95 for cluster 2. The proposed method produced a lower error rate than the other existing models.***

Keywords : K-Means Clustering, Particle Swarm Optimization, Neural Network, Social Security Organization, Data Mining

I. INTRODUCTION

Nowadays, the ever-increasing amounts of various data created in datasets of organizations and institutions have opened up a variety of new opportunities available for businesses in engineering sciences. In most organizations, data and information are rapidly collected and stored. However, one could claim that despite a large amount of data being available, we face a lack of knowledge most of the time to make decisions in organizations.

Revised Manuscript Received on November 30, 2019.
***** Correspondence Author
Amir Rajaei*, Department of Computer Engineering, Velayat University, Iranshahr, Iran.
Mahshid Sedighie, Department of Computer Engineering, Velayat University, Iranshahr, Iran.

In different areas such as data mining, the advent and use of new technologies have led to a revolution in big data.

Data mining is an interdisciplinary science which utilizes different methods including statistics, pattern recognition, machine learning, and databases to extract the knowledge hidden in massive datasets [1].

Mechanization of systems leads to a large archive of collected data, the amount of which is daily increasing and now creating a huge reservoir of information in the social security organization. This information mainly pertains to daily operations of the social security organization such as insurance premium, invalidity, and retirement. At higher levels, managerial reports are also prepared for decision-making. The answers to managerial questions are also provided by utilizing useful patterns found in the collected data. Multiple descriptions and prediction methods can be employed to extract appropriate patterns and rules from data records. Under the supervision of an expert, the results can greatly help decision-makers. Therefore, like any other exploration techniques, data mining must look for a treasure hidden from human sight. Using data mining as a science extraction approach, we can mine an ocean of data to obtain the valuable science pearls. This new science is among the top ten developing sciences which will make the next decade face a technological evolution. In other words, data mining techniques have been regarded as ways to address the needs of organizations in extracting knowledge from large amounts of data. An organization which deals with large amounts of data is the social security organization [2].

The *17*th century marks the historic beginning of the social security system. At the dawn of that century, the matter of poverty and lack of economic procurements for people became more prevalent. Therefore, a series of actions were taken by the governments in order to support the low-income working classes. The first instance can be found in the actions of Henry the Fourth, King of England in 1604, who ordered to take a fraction of each mine's income and use it to buy medicine and treat the injured miners. The social security phrase, which is the future of economic security and social security insurance, was first used by the *US* Federal Government in a 1935 bill. Later, in Article 22 of the Universal Declaration of Human Rights approved by United Nations General Assembly in 1948, a sentence was included to declare that everyone as a member of the society has the right to social security [3].

Moreover, the history of social security in Iran goes back as far as the approval of the first national employment law in 1922, according to which, an institution was founded for the retirement period of workers. In this law, three principles can be seen for social security: providing a pension for people who lose their ability to work after serving for a specific period of time, a special pension for people who become disabled or decrepit because of an accident, and the employers supporting the families of a servant who passes away. In fact, the social security organization is a public insurance organization and its main mission is to cover daily-paid workers (obligatory) and professionals and self-employed people (optional) [5]. Commitments of the social security organization to the insured clients whom it covers. They are divided into two categories of long-term and short-term aids. Wages compensation is money paid during periods of pregnancy, illness, and temporary inability to work when a client is not paid. According to this law, an insurance company supports a client during these periods. Pregnancy wages compensation is one of the short-term obligations of the social security organization toward its clients. In most social security systems in the world, it is given to the clients who, according to the regulations of this law, face the issues of pregnancy and giving birth in return for paying a premium settled in the law [3].

According to the law, a female client can use pregnancy aid in case she has paid the insurance premium for sixty days during the year prior to childbirth. The pregnancy aid for each day of resting is two-thirds of the average daily wages of the client in the last 90 days prior to the resting period. This aid is paid for 6 months at most, approximately a total of 26 weeks before and after childbirth, but does not include the first three days of payment [3][4].

Considering the vulnerability of the working class in the country, comprehensive studies must be conducted from different points of view in order to create (and modify if already available) insurance regulations. As discussed earlier, the wages compensation for each social class must be determined according to up-to-date economic patterns in order to provide economic security for both parties of a contract. A group of clients who are always subject to consideration and various efforts in order to be provide with economic security because of their vulnerability for cultural reasons are pregnant clients [4][5]. Therefore, this paper aims to use data mining techniques to develop a useful model for paying pregnancy period wages compensation and determining the duration of paying this compensation to social security organization clients.

Moreover, the minor objectives are:

- Identification of effective factors in the payment model of pregnancy insurance wages of social insured persons.
- Providing a solution to assist the advisor in identifying eligible people.
- Useful usage of huge volumes of data in SoS databases.

II. Literature Review

According to the related works in this area, data mining is useful in various fields, in which many papers have been published. By publishing these results, the importance and urgency of the subject are posed more than ever before. For this purpose, the following section reviews a few similar papers.

Olson *et al.*, used data mining to predict economical bankruptcy by using decision tree, logistic regression, and neural network techniques. The results indicated that the neural network model produced good outputs on the dataset, but the model was not transparent. Therefore, the decision tree model was selected to present better and more comprehensible results [6].

Chen *et al*., used data mining techniques to discover the factors affecting work injuries in the Taiwanese construction industry. For the sake of analysis, classification and regression methods were used. The results indicated that the factors related to work injuries included injury source, project eligibility, unsafe conditions, incident location, project type, and company size. These findings also showed that the factors related to project eligibility included project type, incident type, project contract value, unsafe actions, unsafe conditions, and company size. An advantage of this model is the accurate analysis of linear models, whereas its disadvantages included misbehavior toward nonlinear models [7].

Predicting spatial and temporal changes in satellite photos using a data mining approach was implemented by Boulila *et al*., to provide more accurate and reliable information about changes on the surface of the earth in satellite images. Change prediction is a challenging task of testing a society remotely. They studied the prediction of urban change through soft computing. A fuzzy decision tree was used for predicting spatial and temporal changes. As a result, it created a series of rules that were accurate and easily interpretable. Predicting the future behavior of the users is one of its advantages [8]. Dejaeger *et al.*, used data mining to develop a model for student satisfaction in educational environments in order to attract more students to that environment. In this paper, artificial neural networks were used to support the strategic decision-making process. An advantage of this model was the discovery of hidden patterns, whereas its disadvantages included its lower accuracy compared to conventional statistical methods [9].

Hijazi *et al*., proposed a method using data mining techniques to screen age-based macular degeneration, which is the main reason for blindness in adults. They employed an artificial neural network and a regression method to classify network images in order to perform *AMD* screening. The results demonstrated that methods based on neural networks were more accurate. An advantage of this method was the discovery of hidden patterns, whereas its disadvantages included the ambiguity of its calculations [10].

The factors affecting the value of intangible assets of a company were identified through data mining by Yen *et al*., This is useful for creditors or people who want to invest in a company. First, five feature selection methods were compared through principal component analysis, stepwise regression, decision tree, association rules, and the genetic algorithm methods. Then, the multi-layer perceptron neural network method was employed to evaluate the effectiveness of the features identified by these methods.

According to the results, the decision tree method had more desirable outcomes than other methods [11].

Farahi *et al*., used data mining to estimate the retirement duration of social security organization clients. For this purpose, the best prediction model for predicting retirement duration of social security organization clients was selected by using a decision-tree based on actual data. It was also determined that clients could be classified with their retirement durations qualitatively analyzed. An advantage of this model was its high accuracy, whereas its disadvantage included its lower accuracy compared to the regression method [5].

Tavakkoli *et al*., employed data mining to predict the turning away patterns of insurance clients. The researchers dug into the databases of a public stock insurance company in the fire insurance category. The results indicated that the client attraction medium was the main predicting factor for clients remaining at or turning away from the company. On the next levels, purchase history and the application of the insured locations were considered turning away prediction factors. The best advantage of this method was the discovery of hidden patterns, whereas the model ambiguity was one of its disadvantages [12].

The literature review showed that no significant research has unfortunately been conducted on using data mining science for services presentable to the clients of social welfare provider organizations and institutions. In other words, no studies have been reported on the duration and use of wages compensation neither for a pregnancy nor for an illness period. Therefore, the proposed model is introduced in the next section.

III. PROPOSED METHOD

This is a descriptive survey. Descriptive research contains the events which have already happened and might be linked with the current state. This type of research describes and interprets the state of current relations when the current condition is set as the research subject. The survey method is a descriptive technique. This method consists of gathering information directly from a group of people. A survey research studies a limited set of variables regarding a large number of people. In fact, this type of research performs a sort of scrolling on a sample or the totality of society in order to describe its attitudes, thoughts, behavior, and features. Accordingly, this study is an applied survey.

Society refers to a group of people who share common traits, distinguishing them from other groups. A sample is a subset of the society which contains some members selected from a statistical population. Therefore, in this paper, the statistical population includes all the people who receive pregnancy wages compensation from the SSO in Iran. It also includes the people receiving pregnancy compensation from the SSO in Kerman. In this paper, the random sampling method was employed to select n samples from a population of N members having an equal chance of being selected [13]. Considering the number of people receiving compensation in the statistical population from 2000 to 2015, 5931 samples were selected randomly from 11504 individuals [17].

The *K*-Means algorithm [16] was integrated with an artificial neural network [15] along with the *PSO* (*Particle Swarm Optimization*) algorithm [14] to analyze the information. More specifically, the *K*-Means algorithm was used to cluster the data. For modeling, an artificial neural network was used along with the *PSO*. By reviewing different algorithms, it was observed that each of these algorithms had its own advantages and disadvantages. Therefore, it was decided to integrate them and develop an algorithm with a better performance. Thus, they were integrated so that their advantages were used to cover the disadvantages of each other. The *PSO* was employed to train the neural network parameters. The following section addresses the steps to the implementation of the proposed method.

A. Steps to the Implementation of the Proposed Model

1. The first step was to collect the data required to develop a model for paying wages compensation and pregnancy period of the clients based on the *SSO* databases.
2. The second step was to use the *K*-Means algorithm for clustering the collected data and other related operations.
3. The third step was to train the neural network model by using the *PSO* algorithm for each cluster separately on the training data available in that cluster.

$$d(x_i, m_j) = \|x_i - m_j\| = \sqrt{\sum_{i=1}^{m} (x_{ik}, m_{jk})^2} \qquad (1)$$

4. The fourth step was to evaluate the proposed model and compare it to some other existing models.

B. The *K*-Means Clustering Algorithm

Clustering is the classification of the data into similar groups. Data clustering is based on maximizing the similarity inside the groups and minimizing the similarity between the groups [19]. This algorithm takes parameter *k* as an input and partitions *n* objects into *k* clusters so that the similarity level in clusters is high when the similarity of objects outside clusters are kept low [18]. The similarity of each object with respect to the average of the objects in its cluster (cluster center) is measured. This algorithm works as follows:

Input: *k*, number of cluster, database *x* consisting of *n* objects $x_i \in R^m$, $X=\{x_1, x_2, \ldots, x_n\}$

Output: a set of *k* clusters that minimize the squared error criterion.

Algorithm:

1. Select *k* random data points as centres of the initial clusters
2. Assign each sample to a cluster based on its distance from cluster centres. We consider the Euclidean distance as the distance criterion, obtained from the following equation.
3. Where x_i is the i^{th} input data and m_j is the center of the j^{th} cluster. The above data will be placed in the cluster with the least distance from its center.
4. Update cluster centres, *i.e*., each one becomes the average of the cluster members.
5. With the new centres, return to step 2 and repeat the above process until the clusters stop changing[16].

C. Using the *PSO* in Training the Neural Network

After the introduction of different artificial intelligence algorithms, the new approach is to use a combination of these algorithms in order to obtain a better performance [21]. Therefore, the novelty of this paper is in that we use the particle swarm optimization algorithm to train the neural network weight parameters [22] which has been rarely seen in previous papers. The overall framework of combining the neural network and particle swarm optimization algorithms is presented in Figure 1.

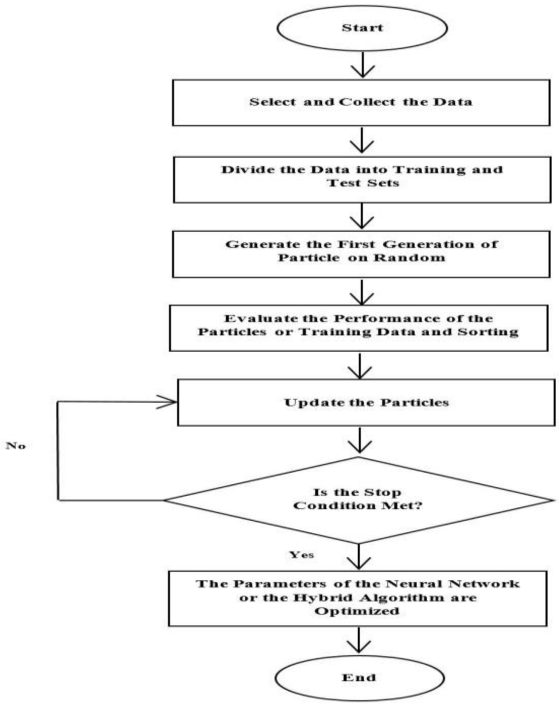

Fig. 1 The Hybrid *PSO-NN* Algorithm

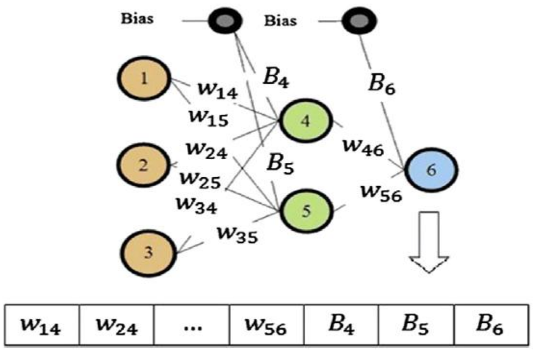

Fig. 2 Presentation of the *PSO* Algorithm Coding

1. Optimizing the Neural Network Parameters using the Proposed Method

The optimization process of the neural network weights in a few different steps and based on the overall framework are as follows:

Step 1: Introducing How the Particles Are Presented:

Considering the overall structure of the network based on all particle parameters (bias weight), *PSO* is defined as an array with the length equal to the number of weights and network biases. Figure 2 shows the *PSO* particle coding for a three-layer neural network with three input variables.

Step 2: Generating the Initial Population at Random:

In this step, N_{POP} initial solutions (the number of required particles, each of which corresponds to a neural network) are generated at random. The velocity of each particle (solution) is also generated at random. The velocity vector is as long as the solution vector (the array containing the weights and the biases which presents the complete network structure) as well.

Step 3: Evaluating the Generated Particles in the Population

The evaluation function used for this purpose is the *MSE* (*Mean Squared Error*) function which is mainly used in optimizing neural networks and is presented as follows:

$$MSE(C_j) = \frac{1}{N}\sum_{i=1}^{N}(Y_i - P_i)^2 \qquad (2)$$

Where C_j is the j^{th} particle of the population, Y_i is the output obtained from the neural network with weight C_j for the i^{th} sample of the training set,

P_i is the real output of the i^{th} sample of the training set, and N is the number of samples in the training set.

Step 4: Updating the Particles in PSO

In this step, the *PSO* particles are sorted based on the *MSE* criterion for each particle (demonstrating the quality and performance of each particle). Then, the best group and personal experiences are determined for all of the particles. The *PSO* algorithm transfers each particle to another region of the solution space at lower error expectancy based on the updated velocity and the location of each particle. The velocity updating for each particle is carried out with a retracting coefficient in the *PSO* method.

Step 5: Evaluating the Termination Condition

After re-evaluating and sorting particles based on the *MSE* criterion, if the number of iterations is equal to the predetermined value, the algorithm execution process will be terminated. Otherwise, it proceeds to the corresponding step and repeats the above process. With the termination of the algorithm execution, the best particle in the last generation is then regarded as the final solution provided by the prediction system.

Considering the advantages of neural networks in modelling complex and non-linear relationships, the development of a smart hybrid model for prediction the relationships between data was in order. Due to the complexity of the relationships between variables, using data pre-processing methods for reaching a more accurate model was taken into consideration. At the same time, clustering models were also used [20] [21].

An applications of clustering methods is to use them in developing modular models. In these models, the prediction task is performed by multiple models, each of which runs on a segment of the data space. Clustering techniques make it possible to partition the problem space into homogeneous subspaces and use the data in each subspace to train the model. Several papers have analyzed the development of smart modular models. In all of them, the models were able to improve the accuracy of non-modular models.

Therefore, in this paper considered using three data clustering technique for developing modular predicting models. For this purpose, the *K*-Means algorithm was used to partition the data into homogeneous spaces. We used the neural network model to develop the prediction model. We also trained a neural network by using the *PSO*. The hybrid optimization method was employed to obtain neural network parameters. Figure 3 shows the schematic of the hybrid smart model.

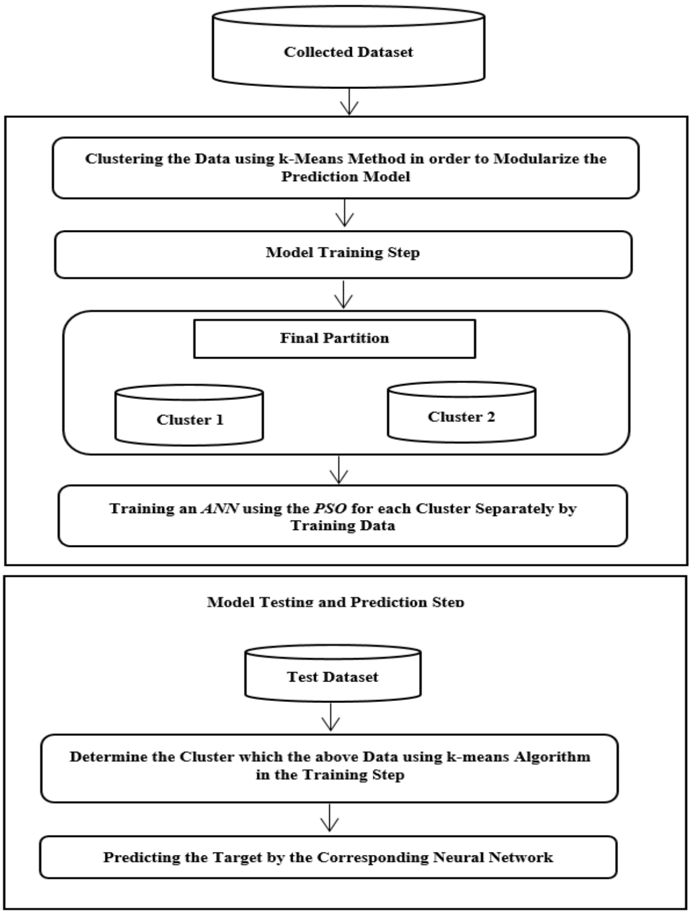

Fig. 3. Overall Schematic of the Proposed Model

IV. RESULT ANALYSIS

Once a small amount of data is gathered for research, it is necessary to first organize and summarize them in such a way that they become meaningfully comprehensible and coherent. Often, the most useful and, at the same time, the first step in organizing data is to sort them with respect to a logical criterion. In a suitable summarization of descriptive statistical methods, the features of a set of information can be accurately described. Table 1 indicates the main statistics of the variables used in this paper.

Table 1. The Main Statistics of the Variables Used

Field	Min	Max	Mean	Std.Dev
Job	0	1	-	-
Age	15	60	32.017	8.381
Leave	0	1	-	-
Family	0	4	1.026	1.220
Education	0	1	-	-
Mount Histogram	1	334	50.173	43.653
Workshop Activity	1	48	-	-
Workshop Rate	0	1	-	-
Workshop Character	0	1	-	-
Workshop Recognition	1	4	-	-
Week-Bar	1	26	22.364	12.494

The above statistics are the data corresponding to the *SSO* clients, collected from the databases of this organization. These statistics are reviewed separately.

Job: this statistic determines the nature of a client's job from a governmental or non-governmental perspective. This statistic has two values: 0 meaning governmental and 1 meaning nongovernmental.

Age: this statistic determines a client's age in the pregnancy period.

Family: this statistic refers to the number of people supported by a client during the pregnancy period.

Education: this statistic indicates either a client is literate or illiterate, taking the values of 0 and 1, respectively.

Mount_Histogram: this statistic shows the number of months a client has been supported prior to pregnancy.

Leave: this statistic determines whether a female client has any medical needs to use medical leave during the pregnancy period. Values 0 and 1 mean needing medical leave and not needing medical leave, respectively.

Workshop-Activity: this statistic refers to the type of workshop activity that a client was employed in during the pregnancy.

Workshop-Rate: this statistic shows the insurance premium rate of the workshop where a client works. It is either 0 meaning full or 1 meaning subject to governmental assistance.

Workshop-Character: this statistic describes the workshop condition which is either 0 meaning natural person or 1 meaning juridical person.

Workshop-Recognition: this statistic expresses how the workshop is insured. It takes one of the following four values:

- Meaning recognition through presenting insurance premium list
- Meaning recognition through inspection from the workshop
- Meaning recognition through the employer of the workshop
- Meaning presenting the contract corresponding to the workshop

Week_Bar: this statistic expresses how many weeks a client is entitled to receive the pregnancy insurance payment.

This paper aims to develop an effective model for predicting the value of the Week_Bar variable. In order to prepare the data, 6 reports were created separately first. The birth dates were converted to age, and prenatal records of clients were determined. Finally, by combining the created records in a single report, a number of samples with missing values in some variables were deleted. After pre-processing and deleting the data with missing values, 5931 samples remained. They were utilized to carry out the modelling process. The following tables present value scattering for each of the above variables.

Table 2. The Scattering Information Corresponding to the Job Variable

Job		
Value	Proportion %	Count
0	43.04	2562
1	56.96	3390

Table 3. The Scattering Information Corresponding to the Leave Variable

Leave		
Value	Proportion %	Count
0	29.86	1777
1	70.14	4175

Table 4. The Scattering Information Corresponding to the Workshop-rate Variable

Workshop-Rate		
Value	Proportion %	Count
0	96.30	5732
1	3.7	220

Table 5. The Scattering Information Corresponding to the Education Variable

Education		
Value	Proportion %	Count
0	1.9	113
1	98.1	5839

Table 6. The Scattering Information Corresponding to the Workshop-Character Variable

Workshop-Character		
Value	Proportion %	Count
0	49.58	2951
1	50.42	3001

Table 7. The Scattering Information Corresponding to the Workshop -Recognition Variable

Workshop -Recognition		
Value	**Proportion %**	**Count**
1	33.77	2010
2	26.46	1575
3	3.16	188
4	36.61	2179

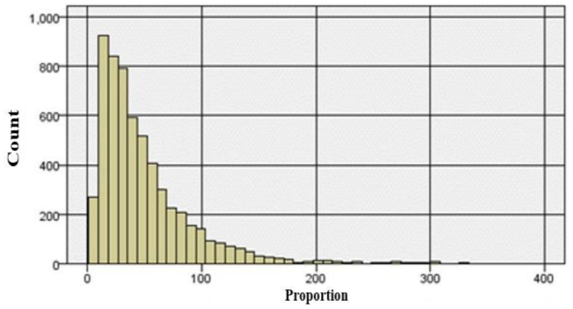

Fig. 4. Scattering Diagram of the Continuous Variable Mount Histogram

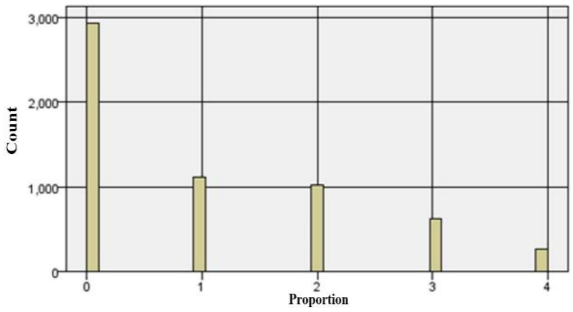

Fig. 5. Scattering Diagram of the Continuous Variable Family

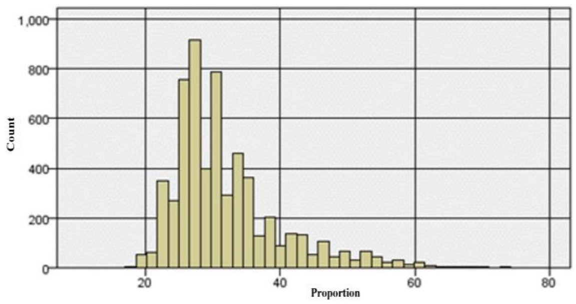

Fig. 6. Scattering Diagram of the Continuous Variable Age

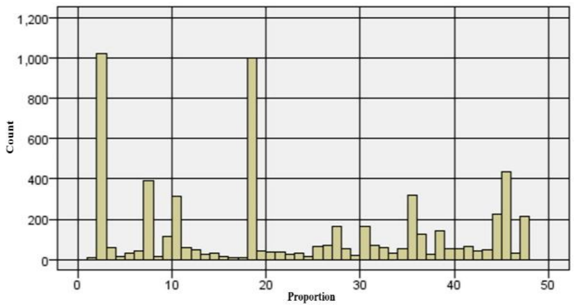

Fig. 7. Scattering Diagram of the Continuous Variable Workshop-Activity

A. Using Clustering to Partition the Data and Create a Modular Model

In order to develop a modular prediction model, it was decided to divide the entire data into homogeneous subsets by using the *K*-Means clustering algorithm. To determine each cluster, a separate prediction model can be obtained for that cluster. Considering the number of data points in our study, the clustering process was performed with 2 and 3 clusters. It was then revealed that using 2 clusters was better because it partitioned the data into clusters appropriately and that data separation rate was higher with 2 clusters.

Table 8. The Number of Samples Assigned to each Cluster

Cluster	Description	Number of Samples in Clusters
Cluster n1	C-*k*means-1	2732
Cluster n2	C-*k*means-2	3199

Table 9. Cluster Centers

Attribute	Cluster 1	Cluster 2
Job	0.943	0.253
Age	35.984	28.655
Leave	0.992	0.455
Family	2.073	0.134
Education	0.965	0.995
Mount Histogram	63.249	39.111
Workshop Activity	20.856	22.829
Workshop Rate	0.053	0.024
Workshop Character	0.554	0.464
Workshop Recognition	2.473	2.383

PCA (*Principal Component Analysis*) [23] was conducted to obtain the clustering results. This analysis is a method for pattern recognition and compression in data to highlight their similarities and differences. In this paper, the *PCA* technique helped utilize the 10 main variables and create 3 new components, which were the clusters resulting from the *K*-Means algorithm [16]. The clustering results are demonstrated using these 3 new components.

Figure 8 presents how the samples are clustered using the *K*-Means algorithm. As can be seen, the data placed in the same cluster are homogeneous and very similar to each other. Using homogeneous data for developing a prediction model for each cluster can improve the prediction accuracy of the model. Also, the figure demonstrates that the *K*-Means clustering algorithm has been able to carry out the partitioning operation of the data into homogeneous subsets very well.

One of the methods which can be used to assess the quality of a clustering algorithm is the *ANOVA* statistical test which is based on the variables used in clustering. This test evaluates the equality of sample averages in each cluster and for each variable [24]. The following tables present the results of this test.

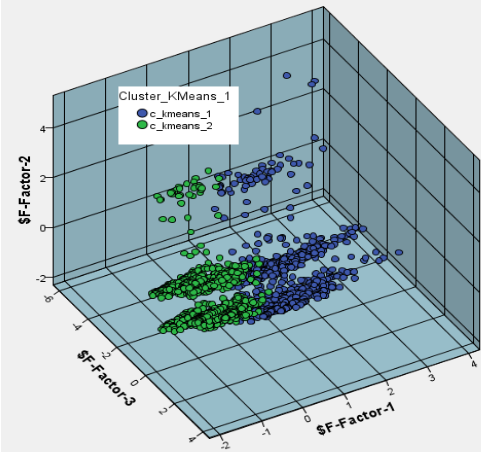

Fig.8. Representing the Clusters using the *PCA* Technique

Table 10. Results of the *ANOVA* Statistical Test for the Job Variable

Attribute-Y	Attribute-X	Description				Statistical Test		
Job	**Cluster-*Kmeans*-1**	**Value**	**Examples**	**Average**	**Std-Dev**	**Variance Decomposition**		
		***Kmeans*.1**	2732	0.94	0.23	**Source**	**SoS**	**D.F.**
		***Kmeans*.2**	3199	0.25	0.43	**BSS**	702.32	1
		All	5931	0.57	0.49	**WSS**	750.62	5929
						TSS	1452.94	5930
						Significance Level		
						Statistics	**Value**	**Prob.**
						Fisher's F	5547.5	0.00

Table 11. Results of the *ANOVA* Statistical Test for the Age Variable

Attribute-Y	Attribute-X	Value	Examples	Average	Std-Dev	Variance Decomposition		
Age	**Cluster-*Kmeans*-1**	***Kmeans*.1**	2732	35.98	9.49	**Source**	**SoS**	**D.F.**
		***Kmeans*.2**	3199	28.65	5.36	**BSS**	79152.2	1
		All	5931	32.03	8.39	**WSS**	337905.0	5929
						TSS	417057.2	5930
						Significance Level		
						Statistics	**Value**	**Prob.**
						Fisher's F	5547.5	0.00

Table 12. Results of the *ANOVA* Statistical Test for the Leave Variable

Attribute-Y	Attribute-X	Value	Examples	Average	Std-Dev	Variance Decomposition		
Leave	**Cluster-*Kmeans*-1**	***Kmeans*C1**	2732	0.99	0.09	**Source**	**SoS**	**D.F.**
		***Kmeans*C2**	3199	0.45	0.50	**BSS**	425.11	1
		All	5931	0.70	0.46	**WSS**	815.04	5929
						TSS	1240.16	5930
						Significance Level		
						Statistics	**Value**	**Prob.**
						Fisher's F	3092.47	0.00

Table 13. Results of the *ANOVA* Statistical Test for the Family Variable

Attribute-Y	Attribute-X	Value	Examples	Average	Std-Dev	Variance Decomposition		
Family	**Cluster-*Kmeans*-1**	***Kmeans*C1**	2732	2.07	1.02	**Source**	**SoS**	**D.F.**
		***Kmeans*C2**	3199	0.13	0.37	**BSS**	5542.83	1
		All	5931	1.03	1.22	**WSS**	3289.68	5929
						TSS	8832.52	5930
						Significance Level		
						Statistics	**Value**	**Prob.**
						Fisher's F	9989.87	0.00

Table 14. Results of the *ANOVA* Statistical Test for the Education Variable

Attribute-Y	Attribute-X	Value	Examples	Average	Std-Dev	Variance Decomposition		
Education	**Cluster-*Kmeans*1**	***Kmeans*C1**	2732	0.96	0.18	**Source**	**SoS**	**D.F.**
		***Kmeans*C2**	3199	0.99	0.07	**BSS**	1.31	1
		All	5931	0.98	0.14	**WSS**	109.54	5929
						TSS	110.85	5930
						Significance Level		
						Statistics	**Value**	**Prob.**
						Fisher's F	70.95	0.00

Retrieval Number: A4286119119/2019©BEIESP
DOI: 10.35940/ijitee.A4286.119119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Table 15. Results of the *ANOVA* Statistical Test for the Mount-Histogram Variable

Mount Histogram	**Cluster-*Kmeans*1**	**Value**	**Examples**	**Average**	**Std-Dev**	**Variance Decomposition**		
		KmeansC1	2732	63.25	54.22	**Source**	**SoS**	**D.F.**
		KmeansC2	3199	39.11	27.61	**BSS**	858560.48	1
		All	5931	50.23	43.70	**WSS**	1046580.9	5929
						TSS	1136141.3	5930
						Significance Level		
						Statistics	**Value**	**Prob.**
						Fisher's F	486.30	0.00

Table 16. Results of the *ANOVA* Statistical Test for the Workshop-Activity Variable

Workshop Activity	**Cluster-*Kmeans*1**	**Value**	**Examples**	**Average**	**Std-Dev**	**Variance Decomposition**		
		KmeansC1	2732	20.86	15.34	**Source**	**SoS**	**D.F.**
		KmeansC2	3199	22.83	15.38	**BSS**	5736.85	1
		All	5931	21.92	15.39	**WSS**	139867.90	5929
						TSS	140414.75	5930
						Significance Level		
						Statistics	**Value**	**Prob.**
						Fisher's F	24.32	0.00

Table 17. Results of the *ANOVA* Statistical Test for the Workshop-Rate Variable

Workshop Rate	**Cluster-*Kmeans*1**	**Value**	**Examples**	**Average**	**Std-Dev**	**Variance Decomposition**		
		KmeansC1	2732	0.05	0.22	**Source**	**SoS**	**D.F.**
		KmeansC2	3199	0.02	0.15	**BSS**	1.23	1
		All	5931	0.04	0.19	**WSS**	210.60	5929
						TSS	211.84	5930
						Significance Level		
						Statistics	**Value**	**Prob.**
						Fisher's F	34.77	0.00

Table 18. Results of the *ANOVA* Statistical Test for the Workshop-Character Variable

Work Character	**Cluster-*Kmeans*-1**	**Value**	**Examples**	**Average**	**Std-Dev**	**Variance Decomposition**		
		KmeansC1	2732	0.55	0.50	**Source**	**SoS**	**D.F.**
		KmeansC2	3199	0.46	0.49	**BSS**	11.94	1
		All	5931	0.51	0.50	**WSS**	1470.60	5929
						TSS	1482.54	5930
						Significance Level		
						Statistics	**Value**	**Prob.**
						Fisher's F	48.14	0.00

Table 19. Results of the *ANOVA* Statistical Test for the Workshop-Recognition Variable

Workshop Recognition	**Cluster-*Kmeans*-1**	**Value**	**Examples**	**Average**	**Std-Dev**	**Variance Decomposition**		
		KmeansC1	2732	2.47	1.20	**Source**	**SoS**	**D.F.**
		KmeansC2	3199	2.38	1.40	**BSS**	11.60	1
		All	5931	2.42	1.28	**WSS**	9773.48	5929
						TSS	9785.47	5930
						Significance Level		
						Statistics	**Value**	**Prob.**
						Fisher's F	9785.44	0.01

As presented in the results table of the *ANOVA* test, the *K*-Means algorithm has been able to create clusters with a good separation rate where, in all variables, the zero assumption of the *ANOVA* test gets dismissed.

B. Training the Neural Network using the Data in Each Cluster

The neural network, used for prediction, is a number of perceptron layers trained through the error backpropagation method. Data normalization is used before training the neural network. Considering the fact that the neural network in this paper used the *tanh* function (hyperbolic tangent), generating an output within [-1,1], and since the scale of input features are different from each other, we took all the input features to the [-1,1] interval and normalized the output feature to the [-0.9,0.9] interval. Normalization was carried out by using the following equation.

$$X^{*}_{new} = \left(\frac{X_{old} - X_{min}}{X_{max} - X_{min}}\right)\left(X^{*}_{max} - X^{*}_{min}\right) + X^{*}_{min} \quad (3)$$

Where X^*_{max}, X^*_{min} are the maximum and minimum value of the data on the new scale and X_{max}, X_{min} are the maximum and minimum value of the data on the original scale. X^*_{new} presents the data value in the converted scale [21]. Clearly, by reversing this equation we can obtain the data value on the old scale. Therefore, the neural network gets trained in the converted scale. After performing the prediction, the real value of the prediction is obtained by taking this value to the original scale [21].

In this paper, the datasets were partitioned into two homogeneous subsets in order to use the *K*-Means clustering algorithm. The first cluster contained 2732 samples, whereas the second one had 3199 samples. In each cluster, %90 of data was used for training, whereas the remaining %10 was used for testing the neural network. By conducting multiple tests, the best neural network structure with the least amount of error was indicated to be a neural network with two hidden layers. There were 5 neurons in its first layer, whereas the second layer contained 4 neurons. Therefore, the structure of the neural network was in the 5-4-1 format (Figure 9) [21].

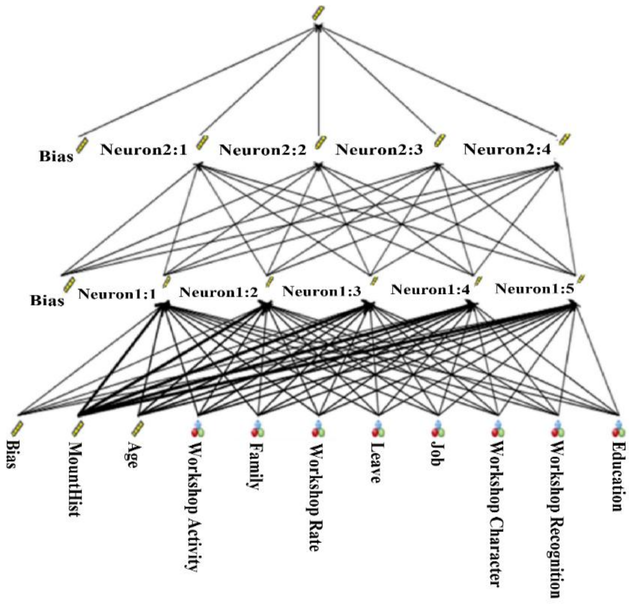

Fig.9. The Presented Neural Network

Table 20 shows the neural network parameters. Table 21 indicates the best parameters of the *PSO* algorithm used for training the neural network.

Table 20. The Best values of the *ANN* Algorithm Parameters for Implementation

	Parameter	Value
Training	Learning Rate	0.2
	Minimum Error	1e-10
	# of Epochs	1000
Hidden Layer	# of Neurons	10
	Transformation Function	Hyperbolic Tangent Sigmoid
Output Layer	Transformation Function	Hyperbolic Tangent Sigmoid

Table 21. The Best Parameter Values of the *PSO-NN* Algorithm for Implementation

	Parameter	Value
Training	C1	2.2
	C2	2.3
	#Population	120
	# of Epochs	4000
Hidden Layer	# of Neurons	The Network has Two Hidden Layers (4,5)
	Transformation Function	Hyperbolic Tangent Sigmoid
Output Layer	Transformation Function	Hyperbolic Tangent Sigmoid

Figures 10 and 11 present the convergence graphs of training and testing neural networks for predicting clusters 1 and 2. The horizontal axis represents training repetitions when the vertical axis is the *RMSE* (*Root-Mean Squared Error*) training error, obtained by using the following equation.

$$RMSE = \sqrt{\frac{1}{N} \sum_{k=1}^{N} (Y_K - P_K)^2} \qquad (4)$$

Where Y_k is the real value and P_k represents the value predicted by the proposed model for the k^{th} sample.

Training the neural network continues until we reach the intended lowest training error, whereas the error is also low for the test data. (in the following graphs, the red curve presents the test data error when the black curve shows the error for the training data in different repetitions of the neural network training process) [21].

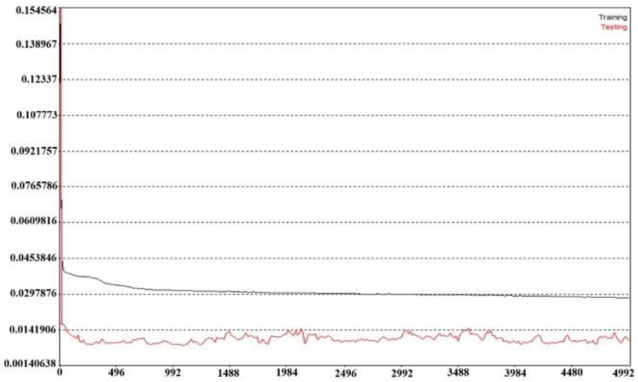

Fig. 10. The Neural Network Training Curve for Cluster 1

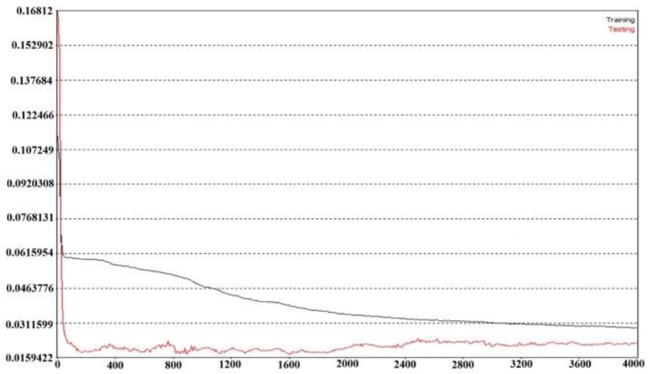

Fig.11. The Neural Network Training Curve for Cluster 2

The above figure shows that the neural network was trained well and converged by not facing the overfitting problem. In case the network faces the overfitting problem, the network error is low on the training data but remains high on the test set [21]. The following tables present the weight values of the trained neural network and the distribution of the weights.

C. Evaluation of the Proposed Method

In order to evaluate the proposed model and compare it with other models, it was decided to use the *MAPE* (*Mean Absolute Percentage Error*) statistic and the R^2 coefficient of determination index for training and test data.

$$MAPE = 100 \times \frac{1}{N} \sum_{i=1}^{N} \frac{|Y_i - P_i|}{Y_i} \qquad (5)$$

Where Y_i is the real value and P_i is the predicted value while N represents the number of data points. The following table presents the performance of the neural network for clusters 1 and 2 in the prediction of training and test data.

Table 22. The Weight Distribution of the Input Layer of the Neural Network with *PSO* for Cluster 1

To the 1st Hidden Layer	From the Input Layer										
	Bias	1st Neuron	2nd Neuron	3rd Neuron	4th Neuron	5th Neuron	6th Neuron	7th Neuron	8th Neuron	9th Neuron	10th Neuron
1st Neuron	1.4	6.3	9.4	-2.66	-4.4	2.6	0.28	0.6	5.6	1.2	-4.1
2nd Neuron	1.3	9.4	9.4	2.5	-2.0	1.8	11.4	2.4	3.2	-7.9	1.4
3rd Neuron	2.5	1.9	1.9	1.9	-2.1	2.6	-8.2	12.7	7.8	3.7	4.1
4th Neuron	-4.5	-2.7	-2.7	-2.7	-2.6	5.8	-10.5	1.11	2.8	-7.9	-1.6
5th Neuron	0.7	8.3	8.3	8.3	6.4	3.4	7.0	6.0	-4.6	3.6	0.7

Table 23. The Weight Distribution of the First Hidden Layer of the Neural Network with *PSO* for Cluster 1

To the 2nd Hidden Layer	From the 1st Hidden Layer					
	Bias	1st Neuron	2nd Neuron	3rd Neuron	4th Neuron	5th Neuron
1st Neuron	-2.4	-0.9	1.1	0.1	-1.9	-6.5
2nd Neuron	0.9	-4.1	7.0	-9.8	-8.2	-3.0
3rd Neuron	-1.1	3.0	-1.3	-8.6	-2.7	-11.5
4th Neuron	-3.3	-1.6	-2.9	0.8	2.7	-5.0

Table 24. The Weight Distribution of the Second Hidden Layer of the Neural Network with *PSO* for Cluster 1

To the 2nd Hidden Layer	From the 2nd Hidden Layer				
	Bias	1st Neuron	2nd Neuron	3rd Neuron	4th Neuron
1st Neuron	-0.9	-2.6	1.7	-4.6	-1.2

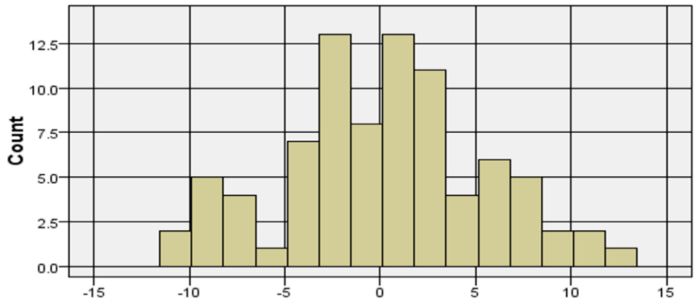

Fig. 12. The Weight Distribution of Different Layers of the Neural Network Trained with *PSO* for Cluster 1

Table 25. The Weight Distribution of the Input Layer of the Neural Network with *PSO* for Cluster 2

To the 1st Hidden Layer	From the Input Layer										
	Bias	1st Neuron	2nd Neuron	3rd Neuron	4th Neuron	5th Neuron	6th Neuron	7th Neuron	8th Neuron	9th Neuron	10th Neuron
1st Neuron	-4.6	-0.5	-1.6	-6.1	-11.9	-1.6	5.9	0.28	2.4	-0.7	-2.4
2nd Neuron	2.9	2.0	16.1	2.2	-1.3	-4.2	5.1	-2.7	-0.4	6.3	-0.4
3rd Neuron	18.7	-0.0	-0.1	-0.8	1.2	6.8	33.1	-0.1	0.3	0.0	-0.1
4th Neuron	-2.0	-2.0	-2.7	2.9	-3.5	-3.6	12.3	-2.5	5.8	1.2	-3.5
5th Neuron	-4.3	0.2	-2.1	0.1	-0.1	-3.6	10.0	-0.4	1.1	-1.5	5.3

Table 26. The Weight Distribution of the First Hidden Layer of the Neural Network with *PSO* for Cluster 2

To the 2nd Hidden Layer	From the 1st Hidden Layer					
	Bias	1st Neuron	2nd Neuron	3rd Neuron	4th Neuron	5th Neuron
1st Neuron	-2.9	-3.8	-8.8	1.1	-4.7	8.2
2nd Neuron	-6.1	4.1	3.3	5.8	-6.3	-0.4
3rd Neuron	-0.1	4.9	-0.8	-15.2	-6.4	0.6
4th Neuron	-9.9	5.5	3.5	-3.3	4.4	-1.3

Table 27. The Weight Distribution of the Second Hidden Layer of the Neural Network with *PSO* for Cluster 2

To the 2nd Hidden Layer	From the 2nd Hidden Layer				
	Bias	1st Neuron	2nd Neuron	3rd Neuron	4th Neuron
1st Neuron	-0.1	2.1	0.6	-1.3	3.6

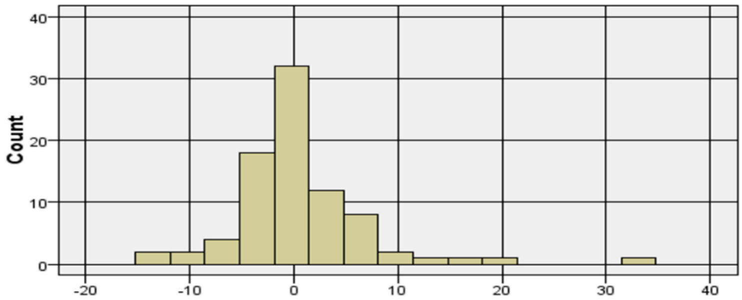

Fig.13. The Weight Distribution of Different Layers of the Neural Network Trained with *PSO* for Cluster 2

Table 28. Evaluation Statistics of the Neural Network Model Trained using *PSO* for the Training and Test Data

Cluster	Train Set		Test Set	
	MAPE	R^2	MAPE	R^2
Cluster 1	17.46	0.98	8.55	0.96
Cluster 2	24.8	0.91	11.63	0.95

To evaluate the influence of using the clustering method for developing a modular model, we implemented a non-modular model without clustering (*ANN+PS*) for all of the data. The *MLP* neural network, trained by using the error backpropagation algorithm (*MLP+BP*), and the radial basis function neural network were also used for performance comparison to the model developed in this paper. The results can be seen in the following table. (in the *RBF* neural network, the number of Gaussian neurons in the hidden layer is 10. Also, the structure of the *MLP* neural network is 5-4-1).

Table 29. Training and Test Data Error in the Three Different Methods

Model	MAPE	
	Train	Test
PSO+ANN (Without Clustering)	30.17	18.23
Modular PSO+ANN (Proposed Model)	21.41	10.21
MLP +BP	37.12	29.78
RBF	32.92	25.14

Accordingly, the developed modular model carries out the pregnancy duration wages prediction with high accuracy for training and test data and can be used as a suitable solution for the prediction of the number of months for pregnancy wages. An output of the proposed model is the sorting of input variables based on their effects on modelling the pregnancy duration wages as the output variable. The following figures show the above ranking for the models corresponding to cluster 1 and cluster 2.

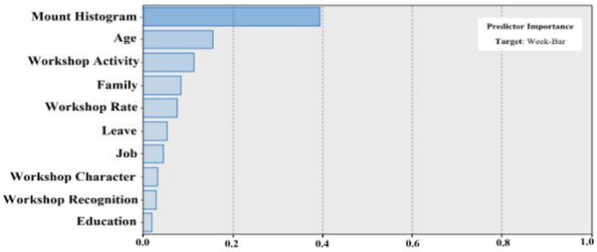

Fig. 14. Variable ranking Based on their Effect on the Pregnancy Stipend Duration for the Samples in Cluster 1

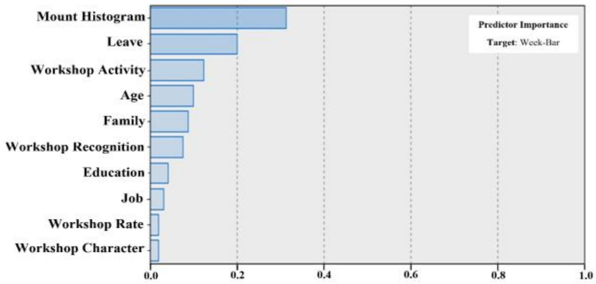

Fig. 15. Variable Ranking Based on their Effect on the Pregnancy Stipend Duration for the Samples in Cluster 2

The table also shows an example of the data used in this paper along with their real values and the values obtained from prediction. The *Week-Bar* column represents the real values, whereas the *$N-Week* column indicates the values obtained from the prediction model used in this paper. Also, column *APE* presented the error values. Error values show that there is no significant difference between the real values and the values obtained from model prediction. This shows that the model presented in this paper verifies the current method used in the social security organization for predicting the number of months for receiving pregnancy wages.

Retrieval Number: A4286119119/2019©BEIESP
DOI: 10.35940/ijitee.A4286.119119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Table 30. An Example of the Data Used in this Paper and the Results Obtained from the Prediction

Gender	Age	Marital Status	Family	Education	Mount Histogram	Workshop Activity	Workshop Rate	Workshop Character	Workshop Recognition	Mount Bik	$N-Mount	APE
1	24	1	1	1	28	2	0	0	4	13	13	0
1	24	1	2	1	10	2	0	1	2	13	13	0
1	46	1	3	1	180	18	0	1	2	23	22	0.04
1	49	1	3	1	153	18	0	0	2	20	21	0.05
0	51	1	2	1	168	3	0	1	4	20	19	0.05
0	52	1	2	1	76	38	0	1	1	19	18	0.05
1	35	1	3	1	126	18	0	1	3	19	18	0.05

V. CONCLUSION

In this paper, we sought to develop a model for paying pregnancy period wages compensation to the social security organization clients by using data mining techniques. The *SSO* commitments toward its clients are classified as one of the two long-term and short-term help categories. Pregnancy period wages compensation is one of the short-term commitments of the *SSO* toward its clients throughout the social security systems all over the world, in return for paying the insurance premium specified in the law, granted to the clients, according to the regulations face pregnancy and childbirth. In order to develop the proposed model, 5931 samples were selected from 11504 individuals at random. Then, using the *K*-Means clustering algorithm, data were divided into cluster 1, consisting of 2732 samples, and cluster 2, consisting of 3199 samples. Then, the data of each cluster were divided into training and test sets with a ratio of 90 to 10, and a multi-layer perceptron neural network was trained for each cluster separately. This paper utilized the *MLP* network model. The *tanh* transfer function was used as the activation function of the hidden activation layer. By performing several tests, the best neural network structure with the smallest amount of error was developed as a neural network with two hidden layers. There were 5 neurons in its first layer and 4 neurons in the second layer. Therefore, the neural network structure was in the 5-4-1 format. Finally, the best model was selected by using the error evaluation method. To evaluate the proposed model, *MAPE* and R^2 criteria were used. For the test data, the result was 0.96 for cluster 1 and 0.95 for cluster 2. The proposed method showed a lower error rate than the other existing models.

REFERENCES

1. H. Hassani, X. Huang, and E. Silva, "Digitalisation and Big Data Mining in Banking", Big Data and Cognitive Computing, vol.2, no.18, pp.1-13, 2018.
2. B. Minaie and F. Asghari, "Applying Data Mining to Discover the Rating Model and Behavioural Analysis Bank Customers", Proceedings of the Second Conference of Data Mining, Iran, 2008.
3. Office of Law and Regulations of the Social Security Organization Legal Aid, "Complete Set of Social Security Laws and Regulations", Jangal Publisher, 2015.
4. E. Khoshnoud, "The Effect of Main Dimension of Job on Organization Commitment of Social Security Staff", MSc Dissertation in Management, Islamic Azad University of Sanandaj, 2011(In Persian)
5. A. Farahi, A. Esaei, T. Majidi and M.R. Kangavari, "Using Data Mining to Estimate the Retirement Life of Social Security Organization Insureds by Creating a Data Warehouse", Proceedings of 5th International Conference of Information and Communication Technology Management, 2008.
6. D., Olson, D., Delen, and Y., Meng, "Comparative Analysis of Data Mining Methods for Bankruptcy Prediction", Decision Support Systems, vol. 52, pp.464-473, 2012.
7. C.W., Cheng, S.S., Leu, Y.M., Cheng and T.C., Wu, "Applying Data Mining Techniques to Explore Factors Contributing to Occupational
8. Injuries in Taiwan's Construction Industry", Accident Analysis and Prevention, vol.48, pp.214-222, 2012.
9. W., Boulila, I.R., Farah, K.S., Ettabaa, and B., Solaiman, "A Data Mining Based Approach to Predict Spatiotemporal Changes in Satellite Images", International Journal of Applied Earth Observation and Geo-information, vol. 13, pp. 83-92, 2011.
10. K., Dejaeger, F.G., Goethals, A., Giangreco, and L., Mola, "Gaining Insight into Student Satisfaction using Comprehensible Data Mining Techniques", European Journal of Operational Research, vol.218, pp. 548-562, 2012.
11. M.H.A., Hijazi, F., Coenen, and Y., Zheng, "Data Mining Techniques for the Screening of Age-Related Macular Degeneration", Knowledge-Based Systems, vol.29, pp. 83-92, 2012.
12. D.C., Yen, Y.H., Lu, and C.F., Tsai, "Determinants of Intangible Assets Value: The Data Mining Approach", Knowledge-Based Systems, vol.31, pp.67-77, 2012.
13. A., Tavakoli, S., Mortazavi, Z., Hosseini, and M., Kahani, "Applying Data Mining Process to Predict Customer Turnover Patterns in Insurance", Journal of Business Management Perspective, vol.4, no. 37, pp.41-55, 2010.
14. H. DanieFard, S.M. Alvani, and A. Azar, "Quantitative Research in Management: A Comprehensive Approach", Saffar Publisher, 2009.
15. J., Kennedy, and R.C., Eberhart, "Particle Swarms Optimization", Proceedings of the IEEE International Conference on Neural Networks, PP.1942-1948,1995.
16. C., Bishop, "Neural Networks for Pattern Recognition", Oxford University Press, 1995.
17. U.M., Fayad, G., Piatetsky-Shapiro, P., Smyth, and R., Uthurusamy, "Advances in Knowledge Discovery and Data Mining", MIT Press, 1996.
18. A., Amidi, "Sampling Methods1", 5th Edition, PayamNoor University Press, 2002.
19. M., Kargari, and M., Sepehri, "Stores Clustering using a Data Mining Approach for Distributing Automotive Spare-parts to Reduce Transportation Costs", Expert Systems with Applications, vol.39, pp.474-478, 2012.
20. [19] M.J.A., Berry, and G.S., Linoff, "Data Mining Techniques for Marketing, Sales and Customer Relationship Management", 2nd Edition, Willey, 2004.
21. E., Hadavandi, H., Shavandi, A., Ghanbari, and S., Abbasian-Naghneh, "Developing a Hybrid Artificial Intelligence Model for Outpatient Visits Forecasting in Hospitals", Applied Soft Computing, vol.12, pp. 700-711, 2012.
22. K., Kumar, and G.S.M., Thakur, "Advanced Applications of Neural Networks and Artificial Intelligence: A Review", International Journal of Information Technology and Computer Science, vol.6, pp. 57-68, 2012.
23. S., Nabavi-Kerizi, M., Abadi, and E., Kabir, "A PSO-based Weighting Method for Linear Combination of Neural Networks", Computers and Electrical Engineering, vol.36, pp. 886-894, 2010.
24. J., Shahrabi, A., ZolghadrShojaei, "Advanced Data Mining: Concepts and Algorithms", 2nd Edition, Iranian Students Booking Agency, Academic Jihad Press, 2011.
25. J., Behbodian, "Statistics and Preliminary Probability", 25th Edition, ImamReza University Press, 2005(In Persian).