TECHNOSTRESS MANAGEMENT THROUGH DATA MINING

Mohamed Salah-Eddine, Hassan 1st University, Morocco Mustapha Belaissaoui, Hassan 1st University, Morocco Alae El Alami, Hassan 1st University, Morocco Karima Salah-Eddine, Mohammed V University, Morocco

ABSTRACT

No one can deny that the arrival of the new computer era has brought with it many benefits and facilities to the human being, to such a point that if we want to justify it, we can give an endless list of examples, but like with many other things in life, there is also a dark side of computers. Employees and users who are confronted to an excessive use of Information and Communication Technologies (ICTs) of Information Systems (IS) in many aspects of their current life, are more prone to Technostress; a new type of stress that we will detect and assess through the use of Data Mining and we will classify patients by using some learning algorithms depending on to their type of stress, the data that we will use is resulting from a Technostress Survey that had already been performed, The use of artificial intelligence in the monitoring of mental health is a new approach in this field which we hope will be of help to both physicians and patients in the near future, making it easier for organizations to use the same diagnostic method but this time by making it computer-assisted, and making detection easier without the need for the physician presence at all times.

Keywords: Information and Communication Technologies, Information Systems, Computer Technology, Coping Strategies, Technostress, Data Mining, Artificial Intelligence.

INTRODUCTION

In this current work we are interested in one of the drawbacks of ICTs which is known as Technostress and which is a type of stress resulting of inappropriate use of ICTs by employees in organizations (Tarafdar et al., 2017; Ragu-Nathan et al., 2008), in our previous works (Salah-Eddine & Belaissaoui, 2016; Salah-Eddine, et al., 2018) we have discussed in detail the phenomenon of Technostress and its relation with the Coping Strategies and based on the existing literature on this subject, we have conceived a conceptual process for organizations to allow ICTs employees overcome Technostress, by reducing and/or eliminating its negative effects as a way to increase job satisfaction (Somvir, 2012) and all this by using ICTs themselves for the first time, in order to cut with the old methods used in the field of Technostress.

Before conceiving our conceptual process, we have noticed that the causality relationship between Technostress-Creators (TCs), which represents ICTs circumstances that have the potential to create job-related strain outcomes among employees in organizations (Tarafdar et al., 2017; William & Cooper, 2002) and between Technostress which is the consequence of TCs associated with the repeated use of the Coping Strategies (Lazarus, 1966; Lazarus & Folkman, 1984) that health specialists use to face Technostress, have given us a quick overview of what we can develop from the relation between these three components, as shown in Figure 1. In Salah-Eddine & Belaissaoui (2016) we have simply begun by gathering the parts of Technostress, Coping Strategies and job satisfaction in a hierarchical way to facilitate the understanding of the whole phenomenon of Technostress from the moment it is detected to the moment it is cured.



FIGURE 1 THE SUMMARIZED CONCEPTUAL MODEL

BACKGROUND

This summarized conceptual model has been, despite its simplicity, the foundation stone in the process of our research development. This thereafter helped us elaborate the Conceptual Model for overcoming Technostress and Getting Job Satisfaction (Salah-Eddine & Belaissaoui, 2016) as shown in the Figure 2. Anyone who observes this new model will notice that it was built on the backbone of the above summarized conceptual model and inspired from the relation between Technostress and Coping and elaborated in a new approach that would help overcome Technostress in a visible logical way but still making use of some ICTs resources. We have actually added these ICTs resources on purpose to remedy the negative effects caused previously by ICTs themselves as an Arabic proverb says "*Treat the disease with what caused it*".

After creating this conceptual model as shown in Figure 2. We transformed it into an activity process as shown in Figure 3. Which allowed us to emphasize the treatments and to represent the behaviors of our method, by producing it in the form of a flow chart that made it easily intelligible (Salah-Eddine et al., 2018).



FIGURE 2 THE CONCEPTUAL MODEL FOR OVERCOMING TECHNOSTRESS AND GETTING JOB SATISFACTION

This research is particularly carried on to confirm the effectiveness of our previous work concerning The Technostress Management Process as shown in Figure 3. And for that reason, we will proceed by steps; this step which concerns our current study will only focus on the Technostress Detection operation as shown in Figure 3. And which is detailed in Figure 4. Which also represents the Detection Sub-Process (Salah-Eddine et al., 2018) and as long as this operation focuses on data, then we will be interested in the classification of data.



TECHNOSTRESS MANAGEMENT PROCESS

1532-5806-22-1-127



FIGURE 4 DETECTION SUB-PROCESS

METHODOLOGY

To analyze the data to predict best results, we know that artificial intelligence provides many software solutions that have multiple classification filtering data processing algorithms, for learning and visualization.

The classification provides a prediction model from training data and test data. These data are screened by a classification algorithm, which through the combination of mathematical tools and computational methods produced a new model capable of classifying any data with the same data classes.

The analysis of data in the field of medicine is more and more frequent in order to specify diagnoses, refine research methods and provide appropriate supplies of equipment according to the importance of the pathologies that appear. The intertwining of our environment and health is becoming more and more linked, hence the birth of Technostress.

Our study is particularly interested in the classification of data. The classification provides a prediction model from training data and test data. These data are screened by a classification algorithm that through the combination of mathematical tools and computational methods produced a new model capable of classifying any data with the same data classes.

The first phase of our study is the data preprocessing phase which is often the most laborious and requires the most time, this is particularly due to the lack of structure and the large amount of noise present in the raw data of use.

Preprocessing consists of cleaning and structuring the data contained in files in order to prepare them for a future analysis. The objectives of the preprocessing stage are to transfer these data into an environment that is easier to exploit. (In our case, this is the case of a database in .arff format).

3 databases were used (Techno-stress.arff, Techno-consulting.arff and Techno-consulting-test.arff).

% 1. Title: Techno-stress Database

%

% 2. definition:

%Techno-overload

%I1_1—I am forced by this technology to work much faster.

%I1_2—I am forced by this technology to do more work than I can handle.

%I1_3—I am forced by this technology to work with very tight time schedules.

%I1_4—I am forced to change my work habits to adapt to new technologies.

%I1_5—I have a higher workload because of increased technology complexity. %Techno-invasion

%I1_8—I spend less time with my family due to this technology.

%I1_9—I have to be in touch with my work even during my vacation due to this technology.

%I1_10—I have to sacrifice my vacation and weekend time to keep current on new technologies.

%I1_11—I feel my personal life is being invaded by this technology.

%Techno-complexity

%I1_12—I do not know enough about this technology to handle my job satisfactorily.

%I1_13—I need a long time to understand and use new technologies.

%I1_14—I do not find enough time to study and upgrade my technology skills.

 $I1_15$ —I find new recruits to this organization know more about computer technology than I do. ?

%I1_16—I often find it too complex for me to understand and use new technologies.

%Techno-insecruty

%I1_17—I feel constant threat to my job security due to new technologies.

%I1_18—I have to constantly update my skills to avoid being replaced.

%I1_19—I am threatened by coworkers with newer technology skills.

%I1_20—I do not share my knowledge with my coworkers for fear of being replaced. ?

%I1_21—I feel there is less sharing of knowledge among coworkers for fear of being replaced.

%Techno-uncertainty

%I1_22—There are always new developments in the technologies we use in our organization. ?

%I1_23—There are constant changes in computer software in our organization.

%I1_24—There are constant changes in computer hardware in our organization.

%I1_25—There are frequent upgrades in computer networks in our organization.

% %

1532-5806-22-1-127

@RELATION Techno-stress
@ATTRIBUTE I1_1 numeric
@ATTRIBUTE I1_2 numeric
@ATTRIBUTE I1_3 numeric
@ATTRIBUTE I1_4 numeric
@ATTRIBUTE I1_5 numeric
@ATTRIBUTE I1_8 numeric
@ATTRIBUTE I1_9 numeric
@ATTRIBUTE I1_10 numeric
@ATTRIBUTE I1_11 numeric
@ATTRIBUTE I1_12 numeric
@ATTRIBUTE I1_13 numeric
@ATTRIBUTE I1_14 numeric
@ATTRIBUTE I1_15 numeric
@ATTRIBUTE I1_16 numeric
@ATTRIBUTE I1_17 numeric
@ATTRIBUTE I1_18 numeric
@ATTRIBUTE II_19 numeric
@ATTRIBUTE II_20 numeric
@ATTRIBUTE II_21 numeric
@ATTRIBUTE II_22 numeric
@ATTRIBUTE II_23 numeric
@ATTRIBUTE II_24 numeric
WATTRIBUTE II_23 humenc
@DATA
00100100101000010101000
0.0.0.1.1.1.0.0.1.0.0.1.1.1.0.0.0.0.0.0
% 1. Title: Techno-consulting Database
@relation Techno-consulting
@attribute Consultation {0,1,2,3,4,5,6,7}
@attribute FirstConsultation numeric
@attribute LastConsultation numeric
<pre>@attribute sick {1,0}</pre>
@data
4,200210,200601,0
5,200301,200601,1

Technostress database was based on the results we got from a survey that we carried out based on Ragu-Nathan Technostress creators questionnaire (Ragu-Nathan et al., 2008) and in which took part a number of colleagues and relatives who despite their different works they all use ICTs. The majority of the participants are civil servants in Hassan I University of Settat, including:

- 1. Faculty of Science and Technology-Settat.
- 2. National School of Commerce and Management-Settat.
- 3. Higher School of Technology-Berrechid.
- 4. Higher Institute of Health Sciences-Settat.
- 5. Faculty of Economic and Social Legal Sciences of Settat.

Techno-consulting section was divided into two distinct parts so as to realize the tests which are primordial for the validation of the matrix of confusion that will guarantee the effectiveness of the selected model and at the same time guarantee that the database can be enriched by new records without impacting the new results.

Data Mining at the Heart of the Decision-Making Process

Literally, Data Mining means "*digging data*", yet talking about extracting knowledge from the data would be more accurate. The term initially refers to a central stage of the process known as KDD (Knowledge Discovery in Databases), but in use it has acquired a broader meaning. Today Data Mining refers to all the computer techniques, tools and applications, allowing to automatically discovering new knowledge within large databases (Lajnef & Ayed, 2005).

In what follows, we will quote the methods that we will perform and understand the role of each model and stains it will accomplish. We will use our databases techno-stress.arff and techno-consulting.arff.

These methods are classification and clustering techniques, they are considered as strategies and main divisions of data mining processes. These strategies are essential in managing algorithms for data analysis. Precisely, these two processes split data into sets.

Classification

Classification is a data mining function that refers items in a collection to target categories or classes. The objective of classification is to predict precisely the target class for any case in the data. To approve the classification model we have to compare the predicted values to recognize target values in a set of test data. The historical data for a classification project is regularly divided into two data sets: one for building the model; the other for testing the model¹.

The classification is also known as the classification tree that creates a step-by-step guide that determines the output of the new data instance. The nodes of the tree represent the decisions that must be made by the function. This model can be used by any instance of data and this is supposedly its best advantage.

The model is built by known values, each time we have new unknown output values, we present it to the model to produce the expected output. This concept is called overffiting, which is used to predict future unknowns, hence the need for our Techno-consulting-test.arff database.

The data that we will use for the classification are summed up on medical consultations for users. The attributes we used in our database are:

- 1. Consultation: The number of consultations that the user has performed.
- 2. First Consultation: The date of the first consultation.
- 3. Last Consultation: The date of the last consultation.
- 4. Sick: The result of the diagnostic.

And the database used is Techno-consulting.arff.



TECHNO-CONSULTING CLASSIFICATION DATA

As shown in the Figure 5. We used a database with 3000 records that we will use to establish a classification models and then choose the appropriate model for our case.

J48 Decision Tree

Decision tree J48 is the implementation of algorithm ID3 (Iterative Dichotomiser 3) developed by the WEKA project team (Kumbhar, 2016; Vadhanam et al., 2006). It is a supervised classification algorithm based on class examples.

Dichotomisation is a separation of two completely opposite things². It means that the algorithm separate attributes iteratively into two groups, the first group is the dominant attribute, second group is the rest of attributes for cause to construct a tree. Then, it calculates the entropy and information gains of each attribute. After founding the dominant attribute, we will put it as a decision node in the tree. After that, entropy and gain scores will be calculated another time among the other attributes. So the dominant attribute is found another time. So on until the decision-making.

The Iterative Dichotomiser 3 algorithm is as follow (Wang, 2017):

73

$$Ent(D) = -\sum_{d=1}^{k} pklog2Pk$$
$$Gain(D, a) = Ent(D) - \sum_{v=1}^{V} \frac{DV}{D}, Ent(DV)$$

D: The number of training samples:

```
Algorithm: Iterative Dichotomiser 3 algorithm
Input: Attribute set A = \{a1, a2, \dots, ad\} & Training set D = \{(x1, y1), \dots, xn\}
(x\overline{2}, y\overline{2}), \cdots, (xn, yn)
Output: decision tree
1 Generate node
2 If example from D \in to the same class. Then
3 The node is labeled as class C leaf node; Return
4 End if
5 If A = \varphi or the values in A= the values in D, Then
6 The node is labeled as leaf node; Return
7 End if
8 Entropy is chosen as a heuristic strategy to select the optimal splitting
attribute ai
9 Foreach aiv in ai do
10 Generate a branch for node
11 If Dv.isEmpty() Then
12 The branch node is labeled as a leaf node
13 Else Take Tree Generate (Dv, A \setminus \{ai\}) as the node divided
14 End if
15 End for
```

The information obtained after the application of the algorithm:

```
=== Run information ===
            weka.classifiers.trees.J48-C 0.25-M 2
Scheme:
Relation:
           Techno-consulting
Instances:
           3000
Attributes: 4
Consultation
FirstConsultation
LastConsultation
sick
Test mode: evaluate on training data
=== Classifier model (full training set) ===
J48 pruned tree
 _____
FirstConsultation<=200011
  Consultation=0: 1 (271.0/114.0)
  Consultation=1
```

1532-5806-22-1-127

```
LastConsultation<=200512: 0 (69.0/21.0)
  LastConsultation>200512: 1 (69.0/27.0)
  Consultation=2: 1 (194.0/84.0)
  Consultation=3: 1 (109.0/38.0)
  Consultation=4
  LastConsultation <= 200511: 0 (54.0/22.0)
LastConsultation>200511: 1 (105.0/40.0)
  Consultation=5
 | LastConsultation<=200505
   \perp LastConsultation<=200504: 0 (8.0)
  | | LastConsultation>200504
 | | | FirstConsultation<=199712: 1 (2.0)
 | | | FirstConsultation>199712: 0 (3.0)
 LastConsultation>200505: 1 (185.0/78.0)
  Consultation=6
  LastConsultation<=200507
      FirstConsultation <= 199812: 0 (8.0)
   FirstConsultation>199812
   | | | FirstConsultation<=200001: 1 (4.0/1.0)
| | | FirstConsultation>200001: 0 (3.0)
 LastConsultation>200507: 1 (107.0/43.0)
Consultation=7: 1 (115.0/40.0)
FirstConsultation>200011
 Consultation=0
| | FirstConsultation<=200412: 1 (297.0/135.0)
| | FirstConsultation>200412: 0 (113.0/41.0)
 Consultation=1: 0 (122.0/51.0)
  Consultation=2: 0 (196.0/79.0)
  Consultation=3: 1 (139.0/69.0)
  Consultation=4: 0 (221.0/98.0)
  Consultation=5
  LastConsultation<=200512: 0 (177.0/77.0)
    LastConsultation>200512
      FirstConsultation<=200306: 0 (46.0/17.0)
  | | FirstConsultation>200306: 1 (88.0/30.0)
  Consultation=6: 0 (143.0/59.0)
  Consultation=7
   LastConsultation<=200508: 1 (34.0/11.0)
    LastConsultation>200508: 0 (118.0/51.0)
```

C	lassifier output											
Γ.												
	Number of Leaves	s: 2	8								A.	
	Size of the tree : 43											
	Time taken to build model: 0.03 seconds											
	Evaluation on training set											
	Time taken to test model on training data: 0.02 seconds											
	=== Summary ===											
	Correctly Classi	ified Inst	ances	1774		59.1333	8					
	incorrectly clas	ssiriea in	stances	1226		40.8667	5					
	Kappa statistic			0.18	07							
	Mean absolute en	ror		0.47	73							
	Root mean square	ed error		0.48	85							
	Relative absolute error 95.4768 %											
	Root relative so	quared err	or	97.71								
	Total Number of Instances 3000											
	=== Detailed Accuracy By Class ===											
		TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class		
		0,662	0,481	0,587	0,662	0,622	0,182	0,616	0,599	1		
		0,519	0,338	0,597	0,519	0,555	0,182	0,616	0,596	0		
	Weighted Avg.	0,591	0,411	0,592	0,591	0,589	0,182	0,616	0,597			
	=== Confusion Matrix ===											
	a b < classified as											
	710 765	b = 0									$ \ge $	
											v	
											_	

FIGURE 6 OUTPUT FROM THE CLASSIFICATION MODEL

In the Figure 6. The confusion matrix shows us that we have 516 as false positives and 710 as false negatives. A false negative is a test result that indicates that a person does not have an illness. As a result, a false positive test result indicates that a person has a disease. According to the accuracy rate of 59.1% the model is a pretty good model¹.

To finalize our classification, we will run the test *via* our Techno-consulting-test.arff database as shown in the Figure 7. To confirm that the accuracy of the model is almost the same for all learning.



FIGURE 7 TECHNO-CONSULTING-TEST CLASSIFICATION DATA

76

1532-5806-22-1-127

٥	Weka Explorer -	□ ×							
Preprocess Classify Cluster Associate	Select attributes Visualize								
Classifier									
Choose J48 -C 0.25 -M 2									
Test options	Classifier output								
 Use training set 	Evaluation on training set								
O Supplied test set Set	Time taken to test model on training data: 0 seconds								
Cross-validation Folds 10	Summary								
O Percentage split % 66									
More options	Correctly Classified Instances 852 56.8 %								
	Kappa statistic 0.1315								
(Nom) sick	Root mean aguared error 0.4867								
Olaut Olau	Relative absolute error 97.3854 %								
Beauti list (right elick for entione)	Total Number of Instances 1500								
Result list (fight-click for options)	Detailed Accuracy By Class								
16:27:21 - trees.J48									
10.20.33 - 0003.340	IP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class 0,437 0,306 0,578 0,437 0,498 0,135 0,581 0,538 1								
	0,694 0,563 0,562 0,694 0,621 0,135 0,581 0,563 0								
	Weighted Avg. 0,568 0,437 0,570 0,568 0,561 0,135 0,581 0,551								
	=== Confusion Matrix ===								
	a b < classified as								
	321 414 a = 1 234 531 b = 0								
	[T							
Status									
ок	Log	×0							
	, —								

FIGURE 8 CLASSIFICATION TREE TEST

Based on the comparison of the "*Correctly Classified Instances*" as shown in Figure 7. And Figure 8. Element of the 56.8% test with the 59.1% training set, we notice that the accuracy is almost identical, which guarantees that the prediction of future unknowns will not change the results. However, the accuracy of the model remains unsatisfactory ranking the data correctly between 50% and 60%, which is a minimal rate in our case study. Therefore, we will test new models until we reach satisfactory results

Random Forest

A random forest is a classifier consisting of a collection of tree-structured classifiers $\{h(x, \theta k), k=1\}$ where the $\{\theta k\}$ are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x (Breiman, 1999; Louppe, 2014).

Random forest is a supervised algorithm that creates a random forest that is a set of decision trees merged together for better accuracy and prediction, which can be used for classification and regression³.

Journal of Management Information and Decision Sciences

)			W	leka Explore	r						×
Preprocess Classify Cluster Associate	Select attributes Vis	sualize									
lassifier											
Choose RandomForest -P 100 -I 100 -nu	um-slots 1 -K 0 -M 1.0 -V	0.001 - 5 1									_
aet optione	Classifier output										
 Use training set 	Evaluation o	n traini	ng set	-							A
O Supplied test set Set	Time taken to te	st model	on traini	ing data: 0.	77 secon	is					
O Cross-validation Folds 10											
Percentage colif. % 66	Summary										
C Percentage spint 50 00	Correctly Classi	2663	88.7667	3							
More options	Incorrectly Classified Instances			337	337 11.2333 %						
	Kappa statistic	0.71	0.7752								
(Nom) sick	Mean absolute er	0.22	0.2264								
	Relative absolut	45,2992 %									
Start Stop	Root relative sq	58.8396 %									
equit list (right click for options)	Total Number of Instances 3000										
court inst (right-click for options)	(1) and	_									
16:27:21 - trees.J48	Decalled Acc	didby by	C1655	-							
16:28:33 - trees.J48		TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class	
16:30:39 - trees.RandomForest		0,900	0,125	0,882	0,900	0,891	0,775	0,966	0,968	1	
	Weighted Lug	0,875	0,100	0,894	0,875	0,885	0,775	0,966	0,967	0	
	werghoed hvy.	0,000	0,113	0,000	0,000	0,000	0,775	0,900	0,500		
	=== Confusion Matrix ===										
	a p < Classified as										
	184 1291	b = 0									
											٧
tatus											
OK.										Lo	a and x
UN .											

FIGURE 9 OUTPUT FROM THE CLASSIFICATION RF MODEL

The confusion matrix as shown in the Figure 9. Shows us that we have 153 as false positive and 184 as false negative. False negatives are people who do not suffer from any illness; false positives indicate people with an illness. According to the accuracy rate of 88.7% it is a good model.

•	Weka Explorer –										
Preprocess Classify Cluster Associate	Select attributes Visualize										
Classifier											
Choose RendemEarant R 100 1100 aur											
Test options	Classifier output										
 Use training set 	Evaluation on training set	A									
O Supplied test set Set	Time taken to test model on training data: 0.3 seconds										
Cross-validation Folds 10	Time taken to test model on training data, 0.5 seconds										
O Percentage split % 66	Summary										
Here entines	Correctly Classified Instances 1391 92.7333 %										
	Incorrectly Classified Instances 109 7.2667 % Keroe statistic 0.8546										
	Mean absolute error 0.213										
(Nom) sick	Root mean squared error 0.2667										
Start Ston	Relative absolute error 42.6222 %										
	Total Number of Instances 1500										
Result list (right-click for options)	Persiled Registers Br. Class										
16:27:21 - trees.J48	Detailed Accuracy by Class										
16:28:33 - trees.J48	TP Rate FF Rate Precision Recall F-Measure MCC ROC Area PRC Area Class										
16:30:39 - trees.RandomForest	0,924 0,069 0,928 0,924 0,926 0,855 0,984 0,985 1										
16:32:16 - trees RandomForest	Weighted Avg. 0,927 0,073 0,927 0,927 0,927 0,855 0,984 0,985										
	Confinion Matrix										
	a b < classified as										
	53 712 b = 0										
		v									
Status											
ОК	Log	x0									

FIGURE 10 CLASSIFICATION TEST FOR RF

1532-5806-22-1-127

As shown in the Figure 10. The comparison of the "*Correctly Classified Instances*" element of the 92.7% test with the 88.7% training set, shows us that the accuracies are almost identical, which guarantees that the prediction of future unknowns will not change the results.

Clustering

Clustering is a statistical analysis method used to organize raw data. Within each group, the data are grouped according to common features. Clustering is especially used to segment, classify a database, or extract knowledge to solicit subsets of data that are difficult to identify.

Formalization:

Either a database $S = \{xi, i = 1 \dots N\}$ such as xi is a tuple.

Clustering problem: determining a function $f: S \rightarrow \{1, ..., k\}$.

*A classCi, j=1k contains the tuples taking the value j on the function f.

*Requires a measure of similarity sim.

*Determine $f: X \to \{1, ..., k\}$ as for each class $Cj, \forall xu, xv \in Cj \land xw \notin Cj$:

(xu, xv) > (xu, xw) and (xu, xv) > (xv, xw)

Dataset Techno-stress.arff was used to be able to group the problems defined in the database in the form of questionnaire. We seek to extract this data by finding patterns in the data and using clusters to determine whether certain behaviors are emerging among users. There are 100 data records in this sample and each column describes user responses to the stress caused by the technology.

With this dataset, clusters were created by the use of k-Means. K-Means is an unsupervised algorithm for non-hierarchical clustering. It allows the grouping in k separate clusters the observations of the data set. Thus similar data will end up in the same cluster.

To consolidate a data set into k distinct cluster, the K-Means algorithm needs a way to compare the degree of similarity between different observations⁵. We will have the choice between two methods: The Euclidean distance (Krislock & Wolkowicz, 2012) and The Manhattan distance (Dulong, 1995). For high dimensional vectors, one might find that Manhattan works better than the Euclidean distance. That is why we are going to work with the Euclidean distance in our case.

Euclidean distance: $\sum_{i=1}^{n} |xi - yi|$

5 clusters were used to extract 5 categories to assign them to 5 types of processing.

RESULTS



FIGURE 11 CLUSTER OUTPUT WITH A VISUAL INSPECTION

The resulting output in the Figure 11. Shows us that clusters match with a "1" means that users share the same value and a "0" means that users in this cluster have a value of zero for this attribute⁴.

Each cluster shows us a type of behavior among users, from which we can draw some conclusions:

- 1. Cluster 0: This group represents only 10 people. We cannot therefore make a weak conclusion against them knowing that they are not statistically relevant.
- 2. Cluster 1: These are the patients who have trouble keeping up with new technologies (which we can classify as Belonging to the old school).
- 3. Cluster 2: This group represents only 10 people. We cannot therefore make a weak conclusion against them knowing that they are not statistically relevant.
- 4. Cluster 3: These are the patients that we will classify as part of laziness intelligent, who finds excuses for not working so much. These people are often remarkably ingenious; when they are given a task, they always find the simplest and the least tiring way to perform it. This leads to thinking more and to working one's mind to find the optimal way to perform the tasks. Their logic for success is not to work harder but simply smarter.
- 5. Cluster 4: This group represents patients who do not need treatment, but simply training sessions to adapt to new technologies and keep pace with the group.

DISCUSION

We want to indicate that this long-term solution will be almost costless since it will concerns only the maintenance and updates of the application (Somvir, 2012), using this solution will eliminate or at least reduce the need of spending money on repeated physician diagnosis on every patient, and the professional secrecy of the organization and staff is respected by keeping all collected data shared locally, Without forgetting that this data can be used by the same

organization managers, in order to acquire more knowledge about the relationship of their employees towards organizational ICTs.

In a survey made by Microsoft, results shows that interference between humans and computers generally lead to Technostress, about 20,000 workers from all around Europe were questioned, and the findings says that 20% was the average of European employees that feel highly engaged at work in our current workplaces (William & Cooper, 2002), But in a strong digital culture, this increases dramatically with four times as many employees saying they felt engaged, the same Research indicated that Technostress is a matter of serious discussion now and employees' health cannot be put at stake because of it⁶.

CONCLUSIONS

The aim behind this first phase of our research is to use data mining as a tool to confirm the accuracy of our computerized solution, from the beginning of our conceptual model until the Technostress detection step. We did an employee-based case study and a TCs classification by type. Moreover, we applied a set of processing's to the results *via* learning algorithms which we have classified in accordance with the five categories: Techno-overload, Techno-invasion, Techno-insecurity, Techno-uncertainty and Techno-complexity. This study is very important for mental health specialists because, in addition to their traditional evolutionary and diagnostic methods, this solution will be of great help to them in the Technostress detection process and in the precision of specific CTs for each patient and for any eventual comparison, but the main target for our current research is organizations and specifically the organizations whose employees are daily confronted to the use of ICTs which often results in the appearance of Technostress among them.

In our next research, we will focus on the second and last phase of our conceptual model which will analyze the coping strategies as a way to gain time, precision in prescriptions and to provide organizations with a dynamic coping referential that would fit the ICTs of each organization without the need for repetitive mental health checking's, thus helping every category of these TCs to automatically get the most appropriate treatment using artificial intelligence.

ENDNOTES

- 1. https://docs.oracle.com/cd/B28359_01/datamine.111/b28129/classify.htm#i1005746
- 2. http://data-mining.business-intelligence.uoc.edu/home/j48-decision-tree
- 3. https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm
- 4. <u>http://www.ieee.ma/uaesb/pdf/distances-in-classification.pdf</u>
- 5. <u>http://stanford.edu/~cpiech/cs221/handouts/kmeans.html</u>
- 6. <u>https://pulse.microsoft.com/uploads/prod/2018/04/Work-Productivity-Digital-Culture-your-competitive-advantage-Full-research.pdf</u>

REFERENCES

Breiman, L. (1999). Random forests. UC Berkeley TR56.

- Dulong, C. (1995). Apparatus and method for determining the Manhattan distance between two points. U.S. Patent No 5,384,722, 24.
- Krislock, N., & Wolkowicz, H. (2012). Euclidean distance matrices and applications. Handbook on semidefinite, conic and polynomial optimization. Springer, Boston.
- Kumbhar, V.S., Oza, K.S., & Kamat, R.K. (2016). Web mining: A Synergic approach resorting to classifications and clustering. River Publishers.
- Lajnef, M.A., Ayed, M.B. (2005). Maximum convergence of the process of data mining and design-evaluation HMI: adaptation of the model. *Proceedings of the 17th Conference on l'Interaction Homme-Machine. ACM*.
- Lazarus, R.S. (1966). Psychological stress and the coping process, (Second Edition). New York, Mc Graw Hills.
- Lazarus, R.S., & Folkman, S. (1984). Stress, appraisal and coping. New York, Springer Publishing Company.
- Louppe, G. (2014). Understanding random forests: From theory to practice.
- Ragu-Nathan, T.S., Tarafdar, M., Ragu-Nathan, B., & Tu, Q. (2008). The consequences of technostress for end users in organizations: Conceptual development and empirical validation. *Information Systems Research*, 19(4), 417-433.
- Salah-Eddine, M., & Belaissaoui, M. (2016). Technostress, coping and job satisfaction model of information systems. *International Conference on Computational Science and Computational Intelligence (CSCI)*.
- Salah-Eddine, M., El-Hamlaoui, M., & Belaissaoui, M. (2018). Computerised method of coping with technostress in organizational situation. *Information Management and Processing (ICIMP) International Conference on*, 130-134.
- Somvir, S.K. (2012). Job satisfaction among library professionals in Haryana state. *International Journal of Scientific and Research Publications*, 2(5), 1-5.
- Tarafdar, M., Tu, Q., Ragu-Nathan, B.S., & Ragu-Nathan, T.S. (2007). The impact of technostress on role stress and productivity. *Journal of Management Information Systems*, 24(1), 301-328.
- Vadhanam, B., Rebecca, J., Mohan, S., & Ramalingam, V.V. (2016). Performance comparison of various decision tree algorithms for classification of advertisement and non-advertisement videos. *Indian Journal of Science* and Technology, 9(1), 48-65.
- Wang, Y., Li, Y., & Song, Y. (2017). Improvement of ID3 algorithm based on simplified information entropy and coordination degree. *Algorithms*, 10(4), 100-124.
- William, S., & Cooper, L. (2002). Managing workplace stress: A best practice blueprint, (First Edition). John Wiley & Sons, West Sussex.