*SindhUniv. Res. Jour. (Sci. Ser.) Vol. 52 (03) 285-288 (2020)*
http://doi.org/10.26692/sujo/2020.09.42
SINDHUNIVERSITYRESEARCHJOURNAL(SCIENCE SERIES)

# Smart Mental Stress Predictive System for Healthcare Using Data Mining Techniques

N. A. KANASRO, F. A. SURAHIO, H. U. ABBASI, M.U.R. MAREE

Department of Computer Science, Sindh University Campus Larkana, Pakistan

Received 24th December 2019 and Revised 19th May 2020

**Abstract:** A huge amount of medical data is available in healthcare that collects from patients. Often receives two to three mental health symptoms like, heart beating and blood sugar, a patient is to be considered that suffering from mental disaster which is not confirm enough. This research developed and provides a model of Mental Stress Predictive System (MSPS) by using advance data mining technique (DMT), namely Decision tree and Naïve Bayes that would help organization to predict the patient's mental health. For this, various other factors have taken such as, chest pain, head ache, heart palpitation etc, and according to these factors, MSPS predicts the patient status. The MSPS is implemented under the .Net framework. Thus, results show that both DMT has distinct strength and works better than traditional decision system. Lift chart and classification matrix method has used for the sake of calculating correct prediction.

**Keywords:** Mental Stress Predictive System MPSP, Data Mining Technique (DMT), Web Data Mining (WDM)

## I. INTRODUCTION

The process of mining is being continued with several algorithms such as, Apriori, SVM, C4 and Decision Tree etc (Liao, *et al.*, 2012) (Yoo, *et al.*, 2012). Organizations are hiring professionals for the sake of to filter and predict passive amount of data through the applications. It has been seen that DMT helps to reveal the hidden pattern as well as relationship from the huge databases. The combination of ML, statistical analysis and database is working together for data extraction (Liu, Bing, and Lei Zhang. (2012). The worth of DMT can be seen through its rich online and offline applications that are used specific purpose.

For instance, few are used to predict data and plenty applications are just for filtration of data. It does not mean that DMT are particularly used in healthcare organization but it is also rendering services in mathematics domain, genetics, marketing and cybernetic (Ngai *et al.*, 2009).

Web Data Mining (WDM) is another type of mining that is helping consumer relationship management which integrates data accompanied by previous mining methods and techniques over the internet. In addition to this, versions of WDM are increasing for giving facility to set hands on prediction features.

Plenty of technique performs different functionalities in various domains for classifying pattern and scalable execution of DM algorithms. From the prediction perspective, it is found significant a lot unless the objective is to predict unknown variables.

It does not mean that known variable does not predict by the mining process. But, it just works for things that a user knows and need to predict for future.

Similarly, objective of this research project is to develop MSPS that helps to organization for patient's mental health prediction.

## 2. RELATED WORK

An Application is provided by Srinivas, *et al.*, 2010) using Naïve Bayes (NB), Neural Network (NN) and Decision tree (DT). It helps to organization to reveal pattern of relationship between medical reasons related to heart patient. Thus, the results were compared to each other and received 84.14% of accuracy with NB. Likewise, another mining application has given by (Aljumah, *et al.*, 2013) for analysis of diabetic treatment and used oracle miner tool. Author used regression technique that was capable to predict the patient's treatment. The dataset classified effectiveness of various treatments through chosen younger and old age people, the results described that younger has treatment need immediately than older. Slightly forward, an application of medicine data mining in healthcare presented by (Niaksu, 2015) and given various methods to develop application based on cardiology patients. The author was given a distinct method namely CRISP-MED-DV and through this method, the application was capable to predict cardio patients.

Corresponding author: Nadeem Ahmed Kanasro, Email: nadeem.kanasro@usindh.edu.pk
*IMCS, University of Sindh
Centre for Environmental Sciences, Faculty of Natural Sciences, University of Sindh, Jamshoro, Pakistan

Till now, all previous contributions shows that majority of work has been done in healthcare for the diabetic patients and connection is being continued, as, a learning algorithm was applied on set of healthcare data for predicting kidney disease based on the (DT) (Boukenze, *et al.,* 2016):. The study selected confusion matrix and this method, received 63% of accuracy into prediction account. Similarly in healthcare, heart patients were predicted via classification tree and achieved 90% results. Various scenarios were described in the study and at the end the author had reached on the best scenario along given percentage of accuracy (Peixoto, *et al.,* 2017) Later, lungs cancer patient diagnosis system (LPDS) has given by the (Omar, et al., 2018). The system was generated via the two major observations; (1) efforts are needed for improvement as well as productivity, accuracy and (2) DM tools cannot use without trainings. The author had used traditional method used MySQL accompanied oracle database.

Weka employed as tool in which the results were calculated and used J8 algorithm for classification of data. Regression was adopted for the prediction and thus the study proved the efficiency of the LPDS in healthcare. The author explained the MT from various aspects.

Again, it is observed that, plenty contributions are available regarding to healthcare and often shown DM is frequently used in medical domain and information is being extracted for identifying heart, kidney and lungs patients. During navigation of contribution, several research papers found related to mental disorders and almost they were using various devices for analysis of mental health.

In this paper, anxiety patient's data collected and predicted patients who are suffering from mental stress.

## 3. METHODOLOGY

The process starts with six steps; business, data understanding, preparation of data, modeling, evaluation and deployment. The first step focuses business understanding the goals and needs from business corner, thus it converts into DM problem definition, and designing primarily strategy to get objectives. In the second step, the raw data is used and proceeds to understand data. Also, classify its quality, increase opening insight and detect subsets for the information that is hidden.

Thus, the data preparation step creates the final dataset that would be supplied into modeling tool. This comprises tables, records and attribute, also data cleaning and transformation. In the fourth step that is modeling, it selects and applies different techniques adjust parameters to optimal values. In the fifth evaluation step, model evaluates for ensuring achieved business objectives. In the final step, the needed tasks are used by models. As, without any language it is not possible to build or access the model contents so, for this Data Mining Extension-(DMX) is used. It is an analysis service of the SQL server product. Furthermore, it gave us facility to create, train, predict and access the model. All parameters set to default setting except for parameters " Minimum Support is equal to 1" for the DT and Minimum Dependency Probability is = to 0.005 for NB".

The models (Trained) were evaluated against dataset for accuracy before deployment in the Chandka Medical College & Hospital Larkana (CMCHL). With the help of Lift chart and classification matrix, the models were validated.

Consequently, the visualizations are incorporated based on graphics that enhances analysis and explanation of results. The description of the attributes is given in **(Table.1).**The attributes are presenting key role for this research because prediction is based on these attribute.

**Table.1 shows selected 12 attributes**

| **Predictable Attributes** | **Description** | **Key Attributes** | **Input Attributes** | **Description** |
|---|---|---|---|---|
| **Diagnose Disease** | Value 0: Low for less than 50% and Value 1: high (greater than 50%) Diameter | Patient No | Sex | 0: Female , 1: Male |
| | | | Chest Pain | Value:1 for type Sharp (for 5 seconds), Value 0 for stab |
| | | | Fasting Blood Sugar | Value 1 for >120 mg, Value 0 for <120 mg |
| | | | Mouth Dry | Value 1 for abnormal and Value 0 for normal |
| | | | Hands Tingling | 0 Value for Normal and Value 1 for Frequent |
| | | | Short Breath | Value 0 for often and Value 1 for repeadly |
| | | | Cold | Value 0 for Low and Value 1 for High |
| | | | Serum | Value 0 for <138 and 1 for >138 Mg |
| | | | OldPeak | 0 for normal, 1 for low, 2 for high |
| | | | Age in Year | 18+ |

**(Table.1)** shows the attributes selected for prediction. In fact, these values are the anxiety grounded patients means those who are suffering from the general anxiety disorder (Kanwar, *et al.,* 2013) Willgoss, *et al.,* 2013).

This disorder often finds in people that are why, GAD patients are selected for this research project. Thus, two symptoms like hands tingling and chest pain to be counted as major symptoms. The dataset of patients is collected from CMCHL database.

## 4. RESULTS

The dataset contains 808 patient records and these were obtained from CMCHL database along various 12 attributes. The attributes are already given in above **(Table.1).** The records of patients were split similarly into two datasets; training (405) and testing (404 records). Thus, to avoid bias, each set was selected randomly for records. The consistency is an important perspective so for this, only categorical fields (attributes) were used for two models. Similarly, non-categorical fields were transformed to categorical information. The major predictable attribute "diagnosis" assigned with value "1" for mental stress patients and value "0" for normal patients. "Patient No" attribute was used as primary and core key and rest are input fields. Consequently, problems of missing, inconsistent and duplicate data have been resolved.

The test process was done via the models that are really effective and proved through two methods lift chart and classification metric by adding add-in in the excel 2013.

### *A. Predictable value with Lift Chart*

To govern if there was enough or sufficient information to acquire patterns in response to the predictable attribute. The columns of model were mapped in the test dataset columns. Thus, column and its state that predicts patients with mental stress predict value =1 to be considered and is given in Figure.1 that shows output.

The X (axis) shows the percentage of test data used to compare predictions however the Y shows the values of predicted to the particular state. From the chart green and blue lines display the result random guess and isolate for ideal model. The purple and red lines display results of NB and DT. The ideal model shows by the green top line and capture 100% of the population that was targeted for mental stress patients using 46% of the dataset for test.

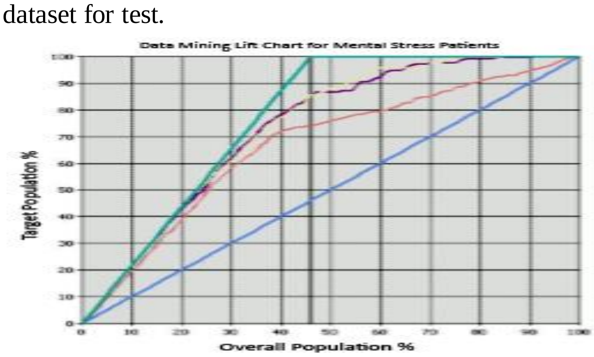


**Fig. 1. Predictable Value with Lift Chart**

The random line is shown by bottom blue line which is the always 45 degree line across the chart lift. It displays that if it is randomly guess the results for individual case 50% of the dataset in test phase then two models lines (yellow and red) fall among the ideal and guess model lines and these are showing that two have enough information to acquire patterns in reply to the predictable states.

### *B. No-Predictable value with Lift Chart*

This step is similar to previous one but the difference is the state of predictable column left blank. For a guess model it does not include a line. It notifies well from the correction of predicting attributes. Figure.2 clarifies the output. The percentage of dataset shows in X axis that was used to compare prediction however Y shows correct predictions percentage. Blue, purple and line red display model that is ideal from the NB and DT. It shows performance of all model states. The blue line (ideal model) is at 45 degree. It means, if 50% test is preceded with dataset, 50% of dataset is predicted correctly from test dataset.

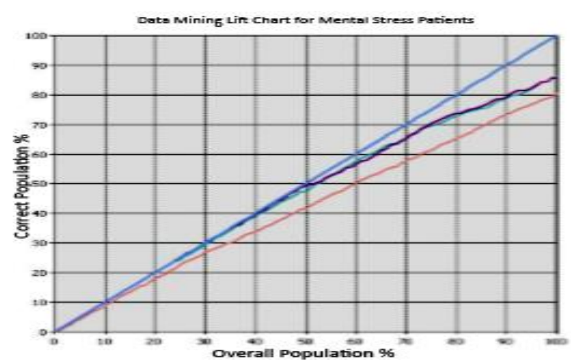


**Fig. 2. No Predictable Value with Lift Chart Result**

If the entire population is proceed with NB then the performance of the NB is noted highest than DT and found 86.12% however DT 80.4%. Also, with less than 50% of population process lift lines always to be higher than for DT. NB is making better high percentage than DT from the correct prediction corner.

### *C. Classification Matrix*

The frequency of correct and incorrect prediction is displaying in classification matrix. It compares value that is actual with predictable value in trained model from test model. In this instance, 208 patients with mental stress contained in test dataset however 246 normal patients. The results of classification matrix are given in Figure.3 with two models. The row represents predicted value and column refers actual value; "1" for mental stress and "0" for normal mean who are not suffering from mental stress or disorder. The predicted values by model are shown in the left most-column and the diagonal shows the correct prediction values.

**Counts for Decision Tree on Diagnosis Group**

| | Predicted | 0 (Actual) | 1 (Actual) |
|---|---|---|---|
| | 0 | 219 | 62 |
| | 1 | 27 | 146 |

**Counts for Naive Bayes on Diagnosis Group**

| | Predicted | 0 (Actual) | 1 (Actual) |
|---|---|---|---|
| | 0 | 211 | 28 |
| | 1 | 35 | 180 |

**Fig.3. Results of two Models in Classification Matrix**

This Table.2 summarizes the result of two models. NB proved as effective with highest correction and 86.53% found with correct prediction of patients from the mental stress grounded. Thus, DT is effective for predicting normal patients and got 89% compare to first model. The models results are given in **(Table.2).**

**Table. 2. Results of Two Models**

| Model Type | Prediction Attribute | No of Patient Cases | Prediction |
|---|---|---|---|
| Decision Tree | +WMS,+PHMS | 146 | Correct |
| | -WMS,+PHMS | 27 | Incorrect |
| | -WMS,-PHMS | 219 | Correct |
| | +WMS, -PHMS | 62 | Incorrect |
| | | | |
| Naïve Bayes | +WMS,+PHMS | 180 | Correct |
| | -WMS,+PHMS | 35 | Incorrect |
| | -WMS,-PHMS | 211 | Correct |
| | +WMS, -PHMS | 28 | Incorrect |

Myth
+ WMS ( Patients with Mental Stress)
- WMS (Patients with no Mental Stress)
+PHMS (Patients Predicted as having Mental Stress)
-PHMS (Patients Predicted as having no Mental Stress)

**(Table.2)** shows the summarize view of the two models. The association rules were applied for prediction and to valid the relationships with patient disease, the conditions were applied. For instance, chest type=1 and mouth dry=1 etc. The size of the dataset is slight small and that is the reason to achieve good results. Huge dataset can give better results. The patients have predicted via these two models accompanied with decisions and applied association rules.

## 5. CONCLUSION

MSPS has provided in this research for the sake of prediction of mental stress patients who are suffering from general disorder anxiety by using two data mining classification techniques; NB and DT.DMX is used for accessing and generating models. The provided system is using DMX in web based interface and the results are given with the help of two methods; lift chart and classification matrix. The effective predict patient model is NB according to results followed by DT. In this research, mental stress patients have predicted and got 86.53% with NB however the DT gave effective percentage with normal patients as compare to NB.

## REFERENCES:

Aljumah, A. A., M. G. Ahamad, and M. K. Siddiqui. (2013): "Application of data mining: Diabetes health care in young and old patients." Journal of King Saud University-Computer and Information Sci. 2: 127-136.

Boukenze, B., H. Mousannif, and A. Haqiq. (2016): "Predictive analytics in healthcare system using data mining techniques." Comput Sci Inf Technol 1, 1-9.

Kanwar, A., S. Malik, L. J. Prokop, L. A. Sim, D. Feldstein, Z. Wang, and M. H. Murad. (2013) "The association between anxiety disorders and suicidal behaviors: A systematic review and meta-analysis." Depression and Anxiety 30, no. 10: 917-929.

Liao, Shu-Hsien, Pei-Hui Chu, and Pei-Yuan Hsiao. (2012) "Data mining techniques and applications–A decade review from 2000 to 2011." Expert systems with applications, Vol.39, no. 12 , Pp.11303-11311.

Liu, Bing, and Lei Zhang. (2012) "A survey of opinion mining and sentiment analysis." In Mining text data, pp. 415-463. Springer, Boston, MA,

Niaksu, O. (2015) "CRISP data mining methodology extension for medical domain." Baltic Journal of Modern Computing 3, no. 2, 92Pp.

Ngai, Eric W T., Li Xiu, and C. Dorothy and K. Chau. (2009) "Application of data mining techniques in customer relationship management: A literature review and classification." Expert systems with applications 36, no. 2: 2592-2602.

Omar, Y., A. Tasleem, M. Pasquier, and A. Sagahyroon, (2018). Lung Cancer Prognosis System using Data Mining Techniques. In *HEALTHINF* 361-368.

Peixoto, R., L. Ribeiro, F. Portela, M. F. Santos, and F. Rua. (2017) "Predicting Resurgery in Intensive Care- A data Mining Approach." Procedia Computer Science113: 577-584.

Srinivas, K., B. Kavihta Rani, and A. Govrdhan. (2010)" Applications of data mining techniques in healthcare and prediction of heart attacks." International Journal on Computer Science and Engineering (IJCSE) 2, 02: 250-255.

Willgoss, T. G. and A. M. Yohannes. (2013) "Anxiety disorders in patients with COPD: a systematic review." Respiratory care 58, no. 5: 858-866.

Yoo, I., P. Alafaireet, M. Marinov, K. Pena-Hernandez, R. Gopidi, Jia-Fu Chang, and L. Hua. (2012) "Data mining in healthcare and biomedicine: a survey of the literature." Journal of medical systems 36, no. 4 2431-2448.