



## Iliou Machine Learning Data Preprocessing Method for Stress Level Prediction

Theodoros Iliou, Georgia Konstantopoulou, Ioannis Stephanakis, Konstantinos Anastasopoulos, Dimitrios Lymberopoulos, George Anastassopoulos

### ► To cite this version:

Theodoros Iliou, Georgia Konstantopoulou, Ioannis Stephanakis, Konstantinos Anastasopoulos, Dimitrios Lymberopoulos, et al.. Iliou Machine Learning Data Preprocessing Method for Stress Level Prediction. 14th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI), May 2018, Rhodes, Greece. pp.351-361, 10.1007/978-3-319-92007-8\_30 . hal-01821068

**HAL Id: hal-01821068**

**<https://inria.hal.science/hal-01821068>**

Submitted on 22 Jun 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Iliou Machine Learning Data Preprocessing Method for Stress Level Prediction

Theodoros Iliou<sup>1</sup>, Georgia Konstantopoulou<sup>2</sup>, Ioannis Stephanakis<sup>3</sup>, Konstantinos Anastasopoulos<sup>4</sup>, Dimitrios Lymberopoulos<sup>4</sup> and George Anastassopoulos<sup>1</sup>

<sup>1</sup> Medical Informatics Laboratory, Medical School, Democritus University of Thrace, Alexandroupolis, 68100, Greece, [tiliou@med.duth.gr](mailto:tiliou@med.duth.gr), [anasta@med.duth.gr](mailto:anasta@med.duth.gr)

<sup>2</sup> Special Office for Health Consulting Services, University of Patras, Greece, [gkonstantop@upatras.gr](mailto:gkonstantop@upatras.gr)

<sup>3</sup> Hellenic Telecommunication Organization S.A. (OTE), 99 Kifissias Avenue, GR-151 24, Athens, Greece, [stephan@ote.gr](mailto:stephan@ote.gr)

<sup>4</sup> Department of Electrical Engineer, University of Patras, Greece, Wire Communications Lab, [anastasofpv@gmail.com](mailto:anastasofpv@gmail.com), [dlympero@upatras.gr](mailto:dlympero@upatras.gr)

**Abstract.** Data pre-processing is an important step in the data mining process. Data preparation and filtering steps can take considerable amount of processing time. Data pre-processing includes cleaning, normalization, transformation, feature extraction and selection. In this paper, Iliou and PCA data preprocessing methods evaluated in a data set of 103 students, aged 18-25, who were experiencing anxiety problems. The performance of Iliou and PCA data preprocessing methods was evaluated using the 10-fold cross validation method assessing seven classification algorithms, IB1, J48, Random Forest, MLP, SMO, JRip and FURIA, respectively. The classification results indicate that Iliou data preprocessing algorithm consistently and substantially outperforms PCA data preprocessing method, achieving 98.6% against 92.2% classification performance, respectively.

**Keywords:** Data preprocessing, machine learning; data mining; classification algorithms; stress; anxiety disorder; panic disorder.

## 1. Introduction

Anxiety disorders are among the most prevalent mental disorders that may occur in the general population and they can lead to chronic characteristics, associated with significant percentages of mortality [3].

This paper is structured as follows: Section 2 describes the Diagnostic criteria, Section 3 presents the Cognitive models, Section 4 presents the Beck Anxiety Inventory (BAI), Section 5 describes our dataset, Section 6 presents the experimental results of this study, while Section 7 concludes this paper and describes future work.

## 2. Diagnostic Criteria

According to DSM-5, anxiety of separation, selective mutism, specific phobia, social anxiety disorder (or social phobia), panic disorder, agoraphobia, generalized anxiety disorder, substance-induced or drug-induced anxiety disorder, anxiety disorder due to another physical state, another predetermined anxiety disorder and unspecified anxiety disorder fall in the category of anxiety disorders [10].

Panic attack is a sudden onset of intense fear or intense discomfort that culminates in minutes, during which can occur at least four of the following symptoms (abrupt onset may occur from calm situation or stressful situation): palpitations, heart "pounding", or accelerated heart rate, transpiration, trembling or intense fear, breathlessness or feeling of suffocation, choking feeling, pain or discomfort in the chest, nausea or abdominal discomfort, dizziness, instability or fainting, chills or feeling of warmth, hallucinations (numbness or tingling), derealization (feeling unreal) or depersonalisation (feeling posting by itself), fear of losing control or oncoming madness, fear of death.

Generalized anxiety disorder is characterized from feelings of excessive anxiety and worry (fearful expectation), which are present quite often and during a period of at least six days for a number of events or activities (such as work and school performance). The person feels that it is difficult to control his worry, while stress and anxiety are associated with at least three of the following symptoms (with some of the symptoms to be present for more days during the last six months) [1] [2] [10]: nervousness or anxiety or stress feeling, feeling of an unusual upset, difficulty in concentrating or feeling that the mind is emptied, irritability, muscle tension.

## 3. Cognitive Models

The theory of anxiety disorders through the cognitive-behavioral model has been the subject of extensive clinical research in recent years. Moreover, the use of this model in therapy has demonstrated so far promising indicators of high efficiency [4-15].

Hoehn-Saric & McLeod in 1988 [11], and also Freeman, in 1990 described stress as a globally known experience that function like a warning safety mechanism. This mechanism emits warning signals during hazardous conditions, associated with uncertainty, while it seems to malfunction in a number of cases, like for example when: (a) stress is too intense, (b) stress lasts also after the exposure to the risk (c) occurs in situations in which risk or threat are absent, or (d) occurs without any particular reason.

#### **4. Beck Anxiety Inventory**

Beck Anxiety Inventory (BAI), created by Aaron T. Beck and his colleagues, is a 21-item, multiple-choice, self-report inventory that is used to measure severity of anxiety in children and in adults [6]. BAI questions are associated with symptoms of anxiety that the subject experienced during the past week (including the day of the BAI), such as numbness and tingling, sweating, and fear that the worst is going to happen. It can be administered to individuals over 17 years old and it takes around 5 to 10 minutes in order to be completed. Several studies have found BAI to be an accurate measurement of anxiety symptoms in children and adults [17].

Somatic subscale is more emphasized on the BAI, as there are 15 out of 21 items measuring physiological symptoms, while the rest of the items concern cognitive, affective, and behavioral components of anxiety. Therefore, BAI functions more adequately in anxiety disorders with a high somatic component, like for example panic disorder. On the other hand, BAI is not appropriate for disorders, such as social phobia or obsessive-compulsive disorder, which have a stronger cognitive or behavioral component. Many questions of the Beck Anxiety Inventory include physiological symptoms, such as palpitations, indigestion, and trouble breathing. Because of this, it has been shown to elevate anxiety measures in those with physical illnesses like postural orthostatic tachycardia syndrome, when other measures did not [18].

The Beck Anxiety Inventory (BAI) and the Anxiety Disorders Interview Schedule (ADIS-IV) were administered to 193 adults with a primary diagnosis of generalized anxiety disorder (GAD), specific or social phobia, panic disorder with or without agoraphobia, obsessive-compulsive disorder (OCD), and no psychiatric diagnosis, at a major Midwestern university recruited from an anxiety research and treatment center. The results of this study support previous findings that the strongest quality of the BAI is its ability to assess panic symptomatology and can be used as an efficient screening tool for distinguishing between individuals with and without panic disorder [16]. The study of Steer RA, Kumar G, Ranieri WF (1995), administered the Beck Anxiety Inventory to 105 outpatients between 13 and 17 years old who were diagnosed with various types of psychiatric disorders and the results are supporting the use of the inventory for evaluating self-reported anxiety in outpatient adolescents [29].

The Beck Anxiety Inventory (BAI), created by Aaron T. Beck and other colleagues, is a 21-question multiple-choice self-report inventory that is used for measuring the severity of anxiety in children and adults. It is a reliable tool. The BAI contains 21 questions, each answer being scored on a scale value of 0 (not at all) to 3 (severely).

Higher total scores indicate more severe anxiety symptoms. The standardized cutoffs are:

- 0–9: normal to minimal anxiety
- 10–18: mild to moderate anxiety
- 19–29: moderate to severe anxiety
- 30–63: severe anxiety

## 5. Data Collection

Questionnaires were completed by 103 students, aged 18-25, who were experiencing anxiety problems and had visited the Office of the Special Consulting Health Services, of Patras University, from the September 2014 until the June 2016.

A result between 0-21 indicates low levels of stress. Although this is usually positive, someone should consider whether the result is realistic or if the person is in refusal concerning its problems. A result between 22-35 indicates moderate stress levels, meaning that the body is trying to tell something. The person should look for patterns in relation to the appearance of symptoms. For example, whether symptoms appear before a specific activity (e.g., a business meeting) and must learn to manage the stress before this activity. No panic is needed, but certainly some technical manager stress will help. A result over 36 points shows high levels of stress and is certainly a reason to ask for supply management. Remember that a high level of anxiety is not a sign of weakness or personal failure. It is something that should be addressed in order to prevent the impact to the mental and physical level of someone [9].

Questionnaires were completed by 103 students, aged 18-25, who were experiencing anxiety problems and had visited the Office of the Special Consulting Health Services, of Patras University, from the September 2014 until the June 2016. It was performed by one person, the psychologist of the office. Though the BAI was developed to minimize its overlap with the depression scale as measured by the Beck Depression Inventory, a correlation of  $r=.66$  ( $p<.01$ ) between the BAI and BDI-II was seen among psychiatric outpatients, suggesting that the BAI and the BDI-II equally discriminate between anxiety and depression. Another study indicates that, in primary care patients with different anxiety disorders including social phobia, panic disorder, panic disorder with or without agoraphobia, agoraphobia, or generalized anxiety disorder, the BAI seemed to measure the severity of depression. This suggests that perhaps the BAI cannot adequately differentiate between depression and anxiety in a primary care population. We have uploaded the original data as well.

## 6. Data preprocessing

The set of techniques used prior to the application of a data mining method are named as data preprocessing for data mining [19] and it is known to be one of the most meaningful issues within the famous Knowledge Discovery from Data process [20]. Since real world data are generally imperfect, incomplete (lacking attribute values,

lacking certain attributes of interest, or containing only aggregate data), noisy (containing errors or outliers), inconsistent (containing discrepancies in codes or names) and contain redundancies, is not directly applicable for a starting a data mining process. We must also mention the fast growing of data generation rates and their size in business, industrial, academic and science applications. The bigger amounts of data collected require more sophisticated mechanisms to analyze it. Data preprocessing is able to adapt the data to the requirements posed by each data mining algorithm, enabling to process data that would be unfeasible otherwise. Data preprocessing mainly includes:



**Figure 1. Data Preprocessing techniques**

- (i) Data cleaning: fill in missing values, smooth noisy data, identify or remove outliers, and resolve inconsistencies.
- (ii) Data integration: using multiple databases or files.
- (iii) Data transformation: normalization and aggregation.
- (iv) Data reduction: reducing the features of the initial dataset but producing the same or similar analytical results.
- (v) Data discretization: part of data reduction, replacing numerical attributes with nominal ones.

Indicatively, we describe the below data preprocessing methods:

- Principal component analysis (PCA) tries to find a rotation such that the set of possibly correlated features transforms into a set of linearly uncorrelated features. The columns used in this orthogonal transformation are called principal components [22]. This method is also designed for matrices with a low number of features.
- Iliou Preprocessing method transforms the initial dataset into a completely new dataset with  $4*m$  attributes (in big data, Iliou method achieves feature reduction), where  $m$  is the number of classes of the dataset output. In this paper we modified our method in order to achieve better results and less complexity.

Iliou method uses Statistics and Linear Algebra methods like the dot (inner) product of dataset vectors in conjunction with some descriptive statistics [24]. The new dataset achieves better classification performance comparing to other preprocessing methods. We used Iliou method for classification or prediction problems [24, 25]. Below follows the detailed description of the method:

### Step 1

Let's assume that a dataset of a machine learning problem named dataset1 is chosen, with n instances (rows), k variables (columns) and m classes. Then, a new set (called from now-on Basic-Set or BS) is created randomly selecting 10% of data from dataset1, consisting of d instances and m classes. The remaining 90% of dataset1 is called Rest-Set.

### Step 2

Afterwards, inner (dot) product of every Basic Set instance (row) with the remaining rows of the Basic Set is computed (An inner product is a generalization of the dot product. In a vector space, it is a way to multiply vectors together, with the result of this multiplication being a scalar). Then, follows the calculation of mean and median values of the inner product result for every instance of each class with the rest instances of its class (Mean\_class<sub>m</sub>\_row<sub>x</sub> and Median\_class<sub>m</sub>\_row<sub>x</sub> respectively, see equation 1 and 2), producing totally m+m=2m new variables, namely Total\_Mean<sub>1</sub>, Total\_Mean<sub>2</sub>, ..., Total\_Mean<sub>m</sub> and Total\_Median<sub>1</sub>, Total\_Median<sub>2</sub>, ..., Total\_Median<sub>m</sub>. For each one of these variables, the method produces d values.

The Mean\_class<sub>m</sub>\_row<sub>x</sub> and Median\_class<sub>m</sub>\_row<sub>x</sub> values are calculated as shown in equation (1) and equation (2) respectively, where m is the name of the class, x (ranges from 1 to d) is the row of the BS and m<sub>1</sub>, m<sub>2</sub> ... m<sub>k</sub> is the first, second...and last row of m class of the BS.

$$\text{Mean\_class}_m\_row_x = \sum_{i=1}^k \text{Mean}(\text{row}_x / \text{row}_i) \quad \text{Eq. 1}$$

$$\text{Median\_class}_m\_row_x = \sum_{i=1}^k \text{Median}(\text{row}_x / \text{row}_i) \quad \text{Eq. 2}$$

Apart from the above, the Total\_Mean and Total\_Median values are calculated based on equation (3) and (4) respectively, where m is the name of the class and d is the sum of rows in the BS. Finally, m total\_Mean and m total\_MEDIAN values are generated, one for every class of the BS.

$$\text{Total\_Mean}_m = \frac{(\text{Mean\_class}_m\text{-row}_1 + \text{Mean\_class}_m\text{-row}_2 + \dots + \text{Mean\_class}_m\text{-row}_d)}{d} \quad \text{Eq. 3}$$

$$\text{Total\_Median}_m = \frac{(\text{Median\_class}_m\text{-row}_1 + \text{Median\_class}_m\text{-row}_2 + \dots + \text{Median\_class}_m\text{-row}_d)}{d} \quad \text{Eq. 4}$$

### Step 3

Assuming that Rest-Set from step 2 has  $r$  instances (rows) and  $m$  classes, a similar to step 2 approach follows. Specifically, inner (dot) product of every single Rest-Set row with every single row of the Basic Set is performed. Then, the mean and median values of the inner (dot) product result of every row for each class are calculated ( $\text{RS\_Mean\_class}_m\text{-row}_j$  and  $\text{RS\_Median\_class}_m\text{-row}_j$  respectively), producing new  $m+m=2m$  variables for every row of the Rest Set. As a result, we have  $r$  values for  $\text{RS\_Mean\_class}_m\text{-row}_j$ , and  $r$  values for  $\text{RS\_Median\_class}_m\text{-row}_j$ .

Similarly to step 2, we compute mean and medial values ( $\text{RS\_Mean\_class}_m\text{-row}_x$  and  $\text{RS\_Median\_class}_m\text{-row}_x$  respectively) for every class.

Apart from the above, the  $\text{Final\_Mean}_m\text{-row}_j$  and  $\text{Final\_Median}_m\text{-row}_j$  values are also calculated as shown in equation (5) and equation (6) respectively ( $m$  is the name of the class and  $j$  (from 1 to  $r$ ) is the row of the Rest set).

$$\text{Final\_Mean}_m\text{-row}_j = \text{total\_Mean}_m\text{ (step 3)} - \text{RS\_Mean\_class}_m\text{-row}_j \quad \text{Eq. 5}$$

$$\text{Final\_Median}_m\text{-row}_j = \text{total\_Median}_m\text{ (step 3)} - \text{RS\_Median\_class}_m\text{-row}_j \quad \text{Eq. 6}$$

Finally,  $m$   $\text{Final\_Mean}_m\text{-row}_j$  and  $\text{Final\_Median}_m\text{-row}_j$  values result, one for every class  $m$  and every row  $j$  of the Rest set.

### Step 4

The rows (variables)  $\text{RS\_Mean\_class}_m\text{-row}_j$ ,  $\text{RS\_Median\_class}_m\text{-row}_j$ ,  $\text{Final\_Mean}_m\text{-row}_j$  and  $\text{Final\_Median}_m\text{-row}_j$  for every class are selected from previous step and then are placed in a new table.

The method ends with the transposition of the Table we described in previous step and the final dataset is now ready to be forwarded in any classification schema. Concluding the description of the proposed method, it is evident that the final dataset consists of 4 variables, namely  $\text{RS\_Mean\_class}_m\text{-row}_j$ ,  $\text{RS\_Median\_class}_m\text{-row}_j$ ,  $\text{Final\_Mean}_m\text{-row}_j$  and  $\text{Final\_Median}_m\text{-row}_j$  for every class of the initial dataset. Thus, if the original dataset has  $m$  classes, the final dataset will have  $4*m$  variables.

## 7. Experimental results

The experiments conducted using seven classification schemes (Table 1): IB1 (Nearest-neighbour classifier), J48 (C4.5 algorithm implementation), Random Forest, MLP (Multilayer Perceptron), SMO (Support Vector Machines), JRip (Repeated Incremental Pruning to Produce Error Reduction) and FURIA (Fuzzy Unordered Rule Induction Algorithm), respectively. In order to estimate how accurately the above predictive models will perform in practice and to assess how the results will generalize to an independent data set, we used the repeated 10-fold cross validation technique [26]. The experiments conducted using WEKA 3.8 data mining software [27] by their default WEKA parameters. We evaluated the performance of the classification schemes using Precision, Recall, Kappa statistics, Weighted Avg ROC area, Weighted Avg Precision Recall Curve (PRC) area, Matthews correlation coefficient (MCC) and Root mean squared error metrics (Table 1) [28-31]. Hence, we evaluated the performance of Iliou preprocessing method by comparing against Principal Component Analysis (PCA) and initial data without preprocessing.

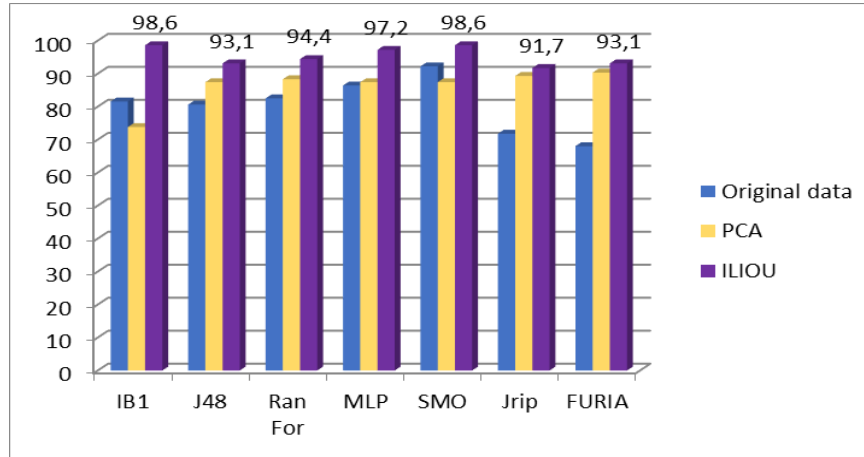
In literature, Galatzer-Levy, Isaac R. et al. (2014) they did experiments classifying trauma survivors according to their stress achieving classification performance 76-80% using SVM, AdaBoost and Random Forest classification algorithms [32]. Purnendu Shekhar Panday (2017) based on heart beat tried to predict whether a person is in stress or not achieving 68% accuracy in classification process [33]. In [34] a real-life, unconstrained study carried out with 30 employees within two organisations and the classification results using ensemble approach increased the accuracy by  $\approx 10\%$  to 71.58% compared to not using any transfer learning technique.

In equations 7 and 8 the formulas for Precision and Recall metrics are presented. The True Positive (TP) is the number of items correctly labeled as belonging to the positive class. The items that correctly labeled as not belonging to the positive class, they are called True Negative (TN). In the case that items which were not labeled as belonging to the positive class by the classifier but should have been, they are called False Negatives (FN). Finally, the items incorrectly labeled as belonging to the class, they are called False Positives (FP). Thus, the number of true positives, true negatives, false negatives and false positives add up to 100% of the set.

$$\text{Precision} = (\text{true positives}) / (\text{true positives} + \text{false positives}) \quad \text{Eq. 7}$$

$$\text{Recall} = (\text{true positives}) / (\text{true positives} + \text{false negatives}) \quad \text{Eq. 8}$$

As we can observe in Table 1, initial (raw) data have achieved the best classification results 92% with Support Vector Machines (SMO) algorithm, PCA transformed data achieved the best classification results 90% and 89% with FURIA and JRip algorithm respectively. Finally, Iliou method achieved 98% classification performance with SMO and IB1 algorithm.



**Figure 2. Experimental Results**

## 8. Conclusions

In this paper we focused on stress level prediction using initial dataset without preprocessing and datasets being preprocessed by PCA and Iliou data preprocessing methods, respectively. Our experimental evaluation has shown that the Iliou data preprocessing algorithm consistently and substantially outperforms PCA data preprocessing method in classification performance. Moreover, table 1 reveals that Iliou method achieved better classification results than initial dataset. In our point of view, Iliou method can be used for significantly improve classification algorithms performance in every stress level dataset.

In future work, it would be preferable to make the same experiments in larger datasets using more classifiers. In addition, Iliou data preprocessing method could be extended to classification algorithm.

## References

1. American Psychiatric Association (APA, 1987). Diagnostic and statistical manual of mental disorders (third edition - revised), American Psychiatric Association, Washington, DC.
2. American Psychiatric Association (APA, 1994). Diagnostic and statistical manual of mental disorders (fourth edition), American Psychiatric Association, Washington, DC.
3. Armstrong KA, Khawaja NG (2002). "Gender differences in anxiety: An investigation of the symptoms, cognitions, and sensitivity towards anxiety in a nonclinical population". Behavioural and Cognitive Psychotherapy. 30: 227–231

4. Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioural change. *Psychological review*, 84, 191-215.
5. Beck, A.T., Emery, G. (with Greenberg, R.) (1985). *Anxiety disorders and phobias*. Basic Books, New York.
6. Beck, A.T., Epstein, N., Brown, G., Steer, R.A. (1988). An inventory for measuring clinical anxiety: psychometric properties. *Journal of consulting and clinical psychology*, 56(6), 893-897
7. Blackburn, I., Davidson, K. (1990). *Cognitive therapy for depression and anxiety*. Blackwell Scientific Publications, Oxford
8. Clark, D.M., Beck, A.T. (1988). Cognitive Approaches. In C.G. Last & M. Hersen (eds), *Handbook of Anxiety Disorders*, Pergamon Press, Oxford.
9. Creamer M, Foran J, Bell R (1995). "The Beck Anxiety Inventory in a nonclinical sample". *Behav Res Ther*. 33 (4): 477–485
10. *Diagnostic and Statistical Manual of Mental Disorders (DSM -5)*, Fifth Edition. Edited by American Psychiatric Association 2013
11. Hoehn-Saric, R., McLeod, D.R. (1988). Panic and Generalized Anxiety Disorders. In C.G. Last & M. Hersen (eds), *Handbook of Anxiety Disorders*, Pergamon Press, Oxford.
12. Kendall, P.C., Hollon, S.D. (1989). Anxious self-talk: development of the anxious self-statements questionnaire. *Cognitive therapy and research*, 13(1), 81-93.
13. Kendall, P.C., Howard, B.L., Hays, R.C. (1989). Self-referent speech and psychotherapy: the balance of positive and negative thinking. *Cognitive therapy and research*, 13(6), 583-598.
14. Kendall, P.C., Ingram, R. (1987). The future for cognitive assessment of anxiety: Let's get specific.
15. L. Michelson & L.M. Ascher (eds), *Anxiety and stress disorders: cognitive-behavioral assessment and treatment*, The Guildford Press, London.
16. Leyfer, OT; Ruberg, JL; Woodruff-Borden, J (2006). "Examination of the utility of the Beck Anxiety Inventory and its factors as a screener for anxiety disorders.". *Journal of anxiety disorders*. 20 (4): 444–58
17. Raj, Satish R (2006-04-01). "The Postural Tachycardia Syndrome (POTS): Pathophysiology, Diagnosis & Management". *Indian Pacing and Electrophysiology Journal*. 6 (2): 84–99.
18. Pyle, D., 1999. *Data Preparation for Data Mining*. Morgan Kaufmann Publishers, Los Altos, California.
19. M. Balasubramanian, E. L. Schwartz, The Isomap Algorithm and Topological Stability. *Science* 4 January 2002: Vol. 295 no. 5552 p. 7.

20. Hall MA. Correlation-based feature selection for machine learning. Waikato University, Department of Computer Science. 1999.
21. Duntelman GH. Principal Components Analysis. A Sage Publications. Thousand Oaks: SAGE Publications; 1989.
22. Vafaie, H., & Imam, I. F. (1994). Feature selection methods: genetic algorithms vs. greedy-like search. In Proceedings of international conference on fuzzy and intelligent control systems.
23. T. Iliou, C. Anagnostopoulos, I. Stephanakis, G. Anastassopoulos, "A novel data preprocessing method for boosting neural network performance: A case study in osteoporosis prediction", Information Sciences, Volume 380, 2017, Pages 92-100.
24. T. Iliou, C. Anagnostopoulos, M. Nerantzaki, and G. Anastassopoulos (2015), "A Novel Machine Learning Data Preprocessing Method for Enhancing Classification Algorithms Performance." In Proc. of the 16th International Conference on EANN, ACM, Article 11, 5 pages.
25. R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection". Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, vol. 2, no. 12, 1995, pp. 1137–1143.
26. Waikato Environment for Knowledge Analysis, Data Mining Software in Java, available online: <http://www.cs.waikato.ac.nz/ml/index.html>, Accessed 11/4/2017.
27. Matthews, B. W. (1975). "Comparison of the predicted and observed secondary structure of T4 phage lysozyme". Biochimica et Biophysica Acta (BBA) - Protein Structure. 405 (2): 442–451. doi:10.1016/0005-2795(75)90109-9.
28. Lehmann, E. L.; Casella, George (1998). Theory of Point Estimation (2nd ed.). New York: Springer. ISBN 0-387-98502-6. MR 1639875.
29. Powers, David M W (2011). "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation". *Journal of Machine Learning Technologies*.
30. Steer RA, Kumar G, Ranieri WF (1995). "Use of the Beck Anxiety Inventory with adolescent psychiatric outpatients". Psychological Reports. 76: 459–465. doi:10.2466/pr0.1995.76.2.459.
31. Smeeton, N.C. (1985). "Early History of the Kappa Statistic". Biometrics. 41: 795. JSTOR 2531300.
32. Galatzer-Levy, Isaac R. et al.(2014). "Quantitative forecasting of PTSD from early trauma responses: A Machine Learning application". Journal of Psychiatric Research , Volume 59 , 68 – 76.
33. Purnendu Shekhar Panday "Machine Learning and IOT for Prediction and Detection of Stress" 2017 17th International Conference on Computational Science and Its Applications (ICCSA) DOI:10.1109/ICCSA.2017.8000018.

34. AlbanMaxhuni, PabloHernandez, Leal., EnriqueSucar, Venet Osmani, Eduardo Morales, Oscar Mayora. "Stress modelling and prediction in presence of scarce data". In: Ambient Intelligence for Health, Springer International Publishing, Puerto Varas, Chile, pp. 224-336 (2015).

**Table 1.** Classification Results

Original data									
	Pre %	Rec %	k	ROC	PRC Area	TP Rate	FP Rate	MCC	Root MSE
<b>IB1</b>	82.1	81.55	0.7	0.852	0.749	0.816	0.105	0.706	0.3455
<b>J48</b>	80.6	80.6	0.681	0.847	0.715	0.806	0.105	0.701	0.3493
<b>Ran</b>	82.3	82.5	0.709	0.954	0.905	0.825	0.114	0.716	0.277
<b>For</b>									
<b>MLP</b>	85.5	86.4	0.772	0.975	0.939	0.864	0.069	0.798	0.2388
<b>SMO</b>	92.4	<b>92.2</b>	0.872	0.955	0.888	0.922	0.034	0.887	0.3022
<b>Jrip</b>	73.5	71.8	0.54	0.796	0.705	0.718	0.165	0.555	0.4015
<b>FURIA</b>	67.1	68	0.45	0.788	0.664	0.680	0.222	0.468	0.4186
PCA									
	Pre	Rec	k	ROC	PRC Area	TP Rate	FP Rate	MCC	Root MSE
<b>IB1</b>	73.4	73.8	0.56	0.768	0.639	0.738	0.179	0.568	0.41
<b>J48</b>	87.3	87.4	0.79	0.901	0.836	0.874	0.077	0.797	0.27
<b>Ran</b>	88.4	88.3	0.80	0.969	0.935	0.883	0.079	0.818	0.26
<b>For</b>									
<b>MLP</b>	87.6	87.4	0.793	0.964	0.917	0.874	0.057	0.815	0.27
<b>SMO</b>	89.8	87.4	0.78	0.937	0.835	0.874	0.088	0.811	0.3196
<b>Jrip</b>	89.4	89.3	0.82	0.917	0.850	0.893	0.064	0.829	0.25
<b>FURIA</b>	90.5	<b>90.3</b>	0.84	0.953	0.906	0.903	0.056	0.843	0.2325
ILIOU method									
	Pre	Rec	k	ROC	PRC Area	TP Rate	FP Rate	MCC	Root MSE
<b>IB1</b>	98.7	<b>98.6</b>	0.97	0.990	0.977	0.986	0.002	0.981	0.09
<b>J48</b>	93.4	93.1	0.88	0.947	0.897	0.931	0.037	0.890	0.21
<b>Ran</b>	94.5	94.4	0.9	0.998	0.996	0.944	0.035	0.908	0.14
<b>For</b>									
<b>MLP</b>	97.4	97.2	0.95	1	1	0.972	0.014	0.957	0.09
<b>SMO</b>	98.7	<b>98.6</b>	0.97	0.993	0.979	0.986	0.002	0.981	0.27
<b>Jrip</b>	92.5	91.7	0.86	0.959	0.918	0.917	0.051	0.873	0.23
<b>FURIA</b>	93.2	93.1	0.88	0.985	0.972	0.931	0.047	0.887	0.17