# Machine Learning Preprocessing Method for Suicide Prediction

Theodoros Iliou, Georgia Konstantopoulou, Mandani Ntekouli, Dimitrios Lymberopoulos, Konstantinos Assimakopoulos, Dimitrios Galiatsatos, George Anastassopoulos

# Machine Learning Preprocessing Method for Suicide Prediction

Theodoros Iliou[1], Georgia Konstantopoulou[2], Mandani Ntekouli[3], Dimitrios Lymberopoulos[3], Konstantinos Assimakopoulos[4], Dimitrios Galiatsatos[1] and George Anastassopoulos[1]

[1] Medical Informatics Lab, Medical School, Democritus University of Thrace, Greece
[2] Special Office for Health Consulting Services, University of Patras, Greece
[3] Wire Communications Lab, Department of Electrical Engineer, University of Patras, Greece
[4] Department of Psychiatry, University of Patras, Greece

**Abstract.** The main objective of this study was to find a preprocessing method to enhance the effectiveness of the machine learning methods in datasets of mental patients. Specifically, the machine learning methods must have almost excellent classification results in patients with depression who have thoughts of suicide, in order to achieve the sooner the possible the appropriate treatment. In this paper, we establish a novel data preprocessing method for improving the prognosis' possibilities of a patient suffering from depression to be leaded to the suicide. For this reason, the effectiveness of many machine learning classification algorithms is measured, with and without the use of our suggested preprocessing method. The experimental results reveal that our novel proposed data preprocessing method markedly improved the overall performance on initial dataset comparing with PCA and Evolutionary search feature selection methods. So this preprocessing method can be used for significantly boost classification algorithms performance in similar datasets and can be used for suicide tendency prediction.

**Keywords:** data preprocessing, Principal Component Analysis, classification, feature selection, suicidal ideation, depression, mental illness

## 1 Introduction

Suicidal ideation is generally associated with depression and other mood disorders. However, it seems to have associations with many other psychiatric disorders, life events, and family events, all of which may increase the risk of suicidal ideation. For example, many people with borderline personality disorder exhibit recurrent suicidal behavior and suicidal ideation. One study found that 73% of patients with borderline personality disorder have attempted suicide, with the average patient having 3 or 4 attempts.

Early detection and treatment are the best ways to prevent suicidal ideation and suicide attempts. If signs, symptoms, or risk factors are detected early then the person will hopefully seek for treatment and help before attempting to take his/her own life. In a study of people who did commit suicide, 91% of them likely suffered from one or more mental illnesses. Nevertheless, only 35% of those people were treated or being treated for a mental illness. This emphasizes the importance of early detection; if a mental illness is detected, it can be treated and controlled to help prevent suicide attempts. Another study investigated strictly suicidal ideation in adolescents. This study found that depression symptoms in adolescents as early as of ninth (9) grade (14–15 years old) is a predictor of suicidal ideation.

## 2 Suicide - Suicidal Ideation

Suicide is a prevalent problem that concerns all countries in the world. However, it is rarely discussed both in the media and in everyday conversations. Many times when people make thoughts regarding one self-destructive behavior, they are attributed to the term "suicidal ideation". The suicidal ideation in some people may persist for years, and in others it may be occasional and caused by difficult events happened in their life. Suicidal thoughts that a person makes may neither be clear nor defined nor involve a very well organized suicide plan. The more persistent and intense these thoughts are, the more serious is the suicidal ideation. People who have attempted suicide even once in their life, are much more likely to try again, especially within the first year of the attempt.

The majority of people who attempt suicide show some samples of their purposes before proceeding to act. Symptoms of suicidal ideation are immediately visible, especially from those of their close environment. The sense of despair - which can be expressed through phrases like "nothing is going to change and get better" - , the feeling of helplessness, the belief that their suicide constitutes no obstacle to family life and friends, alcohol and substance abuse, the preparation of a note for their imminent suicide and tendency to accidents, such as the intentional carelessness in dangerous situations are evidence that if is perceived early can prevent these people from a possible suicide attempt.

One in two people who commit suicide had a history of depression. The rates of suicide in depressed patients are higher than in patients with other diagnosed disorders [1-3] and even higher in patients with severe depression. Undeniable is the fact that people with the disorder of depression consider themselves, the world and the future in a negative way. This indicates the relationship between suicide and the feeling of despair that they have. They believe that there are few or no alternatives for that in his life. Thus, an evaluation of depressed people should include control of suicidal behavior. The purpose of the clinical therapists is to estimate the possibility of suicidal episode, so that can be properly avoided.

## 3 What is depression?

Depression is a disorder that affects mood, thoughts and is usually accompanied by physical discomfort. It affects the eating habits of the patient, his/hers sleep, the way he/she sees himself/herself and how he/she thinks and perceives the world. When someone is diagnosed with depression, he/she often describes himself/herself as a sad, desperate, discouraged and disappointed person.

However, every day we use the term depression meaning a state of unhappiness and misery, that is most of the time transient, has less intensity and probably is caused by something relatively insignificant. This "everyday" meaning of depression differs from the depression as a disorder which is characterized by symptoms that last more than two weeks and are severe enough to interfere with daily life of a person and leads him/her to functional impairment in many aspects of it. In psychiatry, the term depression can also be referred as a mental illness, even when their symptoms do not have reached a high level of severity to obtain such a diagnosis.

For example, people that experience this kind of pessimistic, and intensively sad feelings, do not even have the strength to get up in the morning and do the basic things for surviving, like eating or sleeping. So some people with depression sleep too many hours some cannot sleep at all, while others do very irregular sleep, or they wake frequently during the night or difficulty falling asleep. The most common sleep disorder is the morning awakening, in which the person wakes up very early in the morning and cannot go back to sleep. To be more specific, someone might experience when he/she has depression has depressed mood lasting most of the day and nearly every day, for a period of two weeks, loss of pleasure and reduced interest in activities that were previously the person wanted, and he liked to do.

Helplessness, pessimism, lack of hope and concern about the future are symptoms to be depressed. The person sees everything black and believes that this will remain. Difficulty in concentrating, thinking, memory and making decisions. To have feelings and thoughts of guilt, worthlessness and low self-esteem.

Sometimes the person with depression feels so desperate that commits suicide. The suicide attempt is the most serious and dangerous complication of depression. In people with severe depression, suicide risk is particularly high.

## 4 Data collection

In this paper we establish a mechanism for detecting the possibilities of a patient suffering from depression to be leaded to the suicide. For this reason we measure, using real world statistical data, the effectiveness of all the above symptoms in each case. This cohort is the same one used in previous study [4] and concerns 91 patients who had come to the Special Office for Health Consulting Services University of Patras were diagnosed with different types of depression [5]. Patients were falling in one of the below categories: Major Depressive Disorder, Persistent depressive disorder (Dysthymia), Bipolar Disorders (I & II), Cyclothymic disorder, and Depressive disorder

not otherwise specified (DD-NOS). Our study group included both sexes, age 18-30 and their files contained history of the last 5 years.

A key element for the validity of the disorder decision method is the confirmation of the existence of each symptom, based on interviews that were done. We examined the "symptoms" and not the "points" that had the patient. Symptoms are determined by himself/herself, while the points are independent observations people make the environment and the specialist. For example, the crying may be a point and insomnia a symptom.

With the method of interviews, we recorded the symptoms and the time period that these symptoms occur (e.g., depressed mood over two weeks, or sleep disturbances over two years) in ninety-one (91) patients who were diagnosed with a mood disorder. Then, depending on the symptoms and the time they were repeated we characterized the type of disorder (e.g. Persistent Depressive Disorder – Dysthymia).

In order to achieve our goal, we analyzed all incidents concerning emotional symptomatology and more specifically, concerning about the symptoms that are associated with mood changes.

# 5 Description of machine learning methods

## 5.1 Data Pre-processing Methods

Data pre-processing is an important step in the data mining process since analysis of data that has not been carefully examined can produce misleading results. To this end, the representation and quality of data should first be ensured prior the execution of the experiments. Preprocessing tasks include data cleaning (e.g. identification or outliers' removal), data integration, data transformation (i.e. new feature generation) and data reduction. The product of a data pre-processing task is a new training set that would eventually improve the classification performance and reduce the classification time. This is due to the fact that the dimensionality of the data is reduced, which allows learning algorithms to operate faster and more effectively. In some cases, accuracy on future classification can be improved; in others, the result is a more compact, easily interpreted representation of the target concept [6].

In this paper, we used Principal Component Analysis (PCA) and a novel machine learning data preprocessing method that we have proposed in [7] in order to compare our suggested method performance with PCA.

### 5.1.1 Feature selection

In order to identify if feature (attribute) selection provides better results in our problem and optimize the classification time and performance, a feature selection evaluator, the CfsSubsetEval attribute evaluator was used. Feature selection is the process of selecting a subset of relevant features for use in model construction. The central assumption when using a feature selection technique is that the data contains many redundant or irrelevant features. For the selection of the method, the WEKA 3.8 data mining software was used [8]. WEKA offers many feature selection and feature rank-

ing methods, where each method is a combination of feature search and evaluator of currently selected features. Several combinations have been tested in order to assess the feature selection combination that gives the optimum performance for our problem. The feature evaluator and search method (offered in WEKA) that presented the best performance in the data set were (i) Correlation-based Feature Selection Sub Set Evaluator and (ii) Evolutionary Search method.

The Correlation-based Feature Selection Sub Set Evaluator (CfsSUbsetEval) evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. Subsets of features that are highly correlated with the class while having low intercorrelation are preferred. On the other hand, Evolutionary Search explores the attribute space using an Evolutionary Algorithm (EA). The EA is a (mu, lambda) one with the following operators: uniform random initialization, binary tournament selection, single point crossover, bit flip mutation and generational replacement with elitism (i.e., the best individual is always kept). The combination of the above mentioned methods proposed four from all of the features that formed originally the feature set. These features are: (i) difficulties in functioning, (ii) unworthiness/guilt, (iii) Major Depressive Disorder and (iv) Depressive disorder not otherwise specified (DD-NOS).

### 5.2 Short Description of suggested Data Pre-processing Method

The proposed method can substantially improve successful classification when applying machine learning techniques to data mining problems. It transforms the input data into a new form of data, which is more suitable and effective for the learning scheme chosen. 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.

The differences between adjacent elements of every instance of dataset1 are calculated, and the new k-1 variables are added in dataset1, creating a new dataset named dataset2 with k+(k-1) variables.

**Step 2**

Assuming that the set of attributes for every instance is a vector whose elements are the coefficients of a polynomial in descending power, step 2 estimates the derivative of the vector. The result is a new vector (one element shorter than initial one), with the coefficients of the derivative in descending power. Then, this new vector is added in dataset2 forming a new dataset named dataset3.

**Step 3**

In the third step of the proposed method, a new set (called from now-on Basic-Set) is created randomly selecting 10% of data from dataset3, consisting of d instances and m classes. The remaining 90% of dataset3 is called Rest-Set. Then, matrix right division (or slash division) of every Basic Set instance (row) with the remaining rows of

the Basic Set is computed (Slash or matrix right division B/A is roughly the same as B*inv(A), more precisely, B/A = (A'\B')' ). Then, follows the calculation of mean and median values of the division result for every instance of each class with the rest instances of its class (variables $Mean\_class_m\_row_x$ and $Median\_class_m\_row_x$ respectively), producing totally m+m=2m new variables ($Total\_Mean_1$, $Total\_Mean_2$,…, $Total\_Mean_m$ and $Total\_Median_1$, $Total\_Median_2$, …, $Total\_Median_m$ for every row of the Basic Set. Hence, we have d values for $Total\_Mean_1$, d values for $Total\_Mean_2$, …, d values for $Total\_Mean_m$ and d values for $Total\_Median_1$, d values for $Total\_Median_2$, …, d values for $Total\_Median_m$ .

Apart from the above, the Total_Mean and Total_Median values are calculated as shown in equation (1) and equation (2) respectively (m is the name of the class and d is the sum of Basic Set rows). Finally, m total_Mean and m total_MEDIAN values result, one for every class of the Basic set.

$$Total\_Mean_m = \frac{(Mean\_class_m\_row_1 + Mean\_class_m\_row_2 + \ldots + Mean\_class_m\_row_d)}{d} \quad (1)$$

$$Total\_Median_m = \frac{(Median\_class_m\_row_1 + Median\_class_m\_row_2 + \ldots + Median\_class_m\_row_d)}{d} \quad (2)$$

**Step 4**

Assuming that Rest-Set from step 3 has r instances (rows) and m classes, a similar to step 3 approach follows. Specifically, matrix right division of every single Rest-Set row with every single row of the Basic Set is performed. Then, the mean and median values of the division result of every row for each class are calculated ($RS\_Mean\_class_m\_row_j$ and $RS\_Median\_class_m\_row_j$ respectively), producing new m+m=2m variables for every row of the Rest Set. As a result, we have r values for $RS\_Mean\_class_m\_row_j$, and r values for $RS\_Median\_class_m\_row_j$.

Similarly to step 3, we compute mean and medial values ($RS\_Mean\_class_m\_row_x$ and $RS\_Median\_class_m$ respectively ) for every class.

Apart from the above, the $Final\_Mean_m\_row_j$ and $Final\_Median_m\_row_j$ values are also calculated as shown in equation (3) and equation (4) respectively (m is the name of the class and j (from 1 to r) is the row of the Rest set.

$$Final\_Mean_m\_row_j = total\_Mean_{m\ (step3)} - RS\_Mean\_class_m\_row_j \quad (3)$$

$$Final\_Median_m\_row_j = total\_Median_{m\ (step3)} - RS\_Median\_class_m\_row_j \quad (4)$$

Finally, m $Final\_Mean_m\_row_j$ and $Final\_Median_m\_row_j$ values result, one for every class m and every row j of the Rest set.

**Step 5**

The rows (variables) RS_Mean_classm_rowj, RS_Median_classm_rowj, Final_Meanm_rowj and Final_Medianm_rowj for every class are selected from previous step and then are placed in a new table [7].

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 consist of 4 variables, namely RS_Mean_class$_m$ _row$_j$, RS_Median_class$_m$ _row$_j$, Final_Mean$_m$_row$_j$ and Final_Median$_m$_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.

## 6 Experimental results

For our experiments we used the dataset described in section 4. In order to categorize subjects into two classes (suicide tendency, no-suicide tendency), several machine learning classification algorithms were tested in this paper, selected based on their popularity and frequency in biomedical engineering problems. Each classifier was tested with initial dataset, with dataset after feature selection with Evolutionary search, with transformed dataset using PCA method and finally with the transformed dataset using our suggested data preprocessing method (Table 2). In order to better investigate the generalization of the prediction models produced by the machine learning algorithms, the repeated 10-fold cross validation method was used. We used the WEKA default parameters for the classifiers that we have used. For the classifiers with the best performance (MLP) the parameters are: Hidden layers=8, learning rate=0.3, momentum=0.2, training time =500 epochs.

**Table 1.** Classification results

| **Classifiers** | **Initial Data Set (%)** | **Evolutionary Search (%)** | **PCA (%)** | **Suggested Method (%)** |
|---|---|---|---|---|
| MLP (Multilayer Perceptron) | 64.83 | 75.82 | 64.83 | 92.18 |
| MultilayerPerceptronCS | 64.83 | 75.92 | 64.83 | 92.18 |
| Radial Basis Function Classifier | 75.82 | 74.72 | 78.02 | 85.93 |
| RBFNetwork | 70.32 | 75.82 | 76.92 | 84.37 |
| FURIA (Fuzzy Logic algorithm) | 71.42 | 75.82 | 76.92 | 82.81 |
| SMO (Support Vector Machines) | 64.83 | 76.92 | 75.62 | 76.56 |
| HMM (Hidden Markov Models) | 76.92 | 76.92 | 76.53 | 76.56 |
| J48-Graft | 64.83 | 73.62 | 64.83 | 82.81 |
| Random Forest | 73.62 | 74.72 | 76.84 | 85.93 |
| IB1 | 54.94 | 76.92 | 67.03 | 93.75 |

In Table 2, we can observe that HMM classification results have not increased with any of the preprocessing or attribute selection methods we have used. Using Evolutionary search, classification results was increased almost in all classification algorithms except RBF classifier and HMM. Using PCA method, classification results

were increased in most of the classification algorithms as well. Our suggested data preprocessing method significantly increased the classification performance (93.75% with IB1 algorithm and 92.18% with MLP) and achieved the best classification results comparing with all the other methods we have used.

## 7 Conclusions

Data pre-processing is an important step in the data mining process. If there is much irrelevant and redundant information present or noisy and unreliable data, then knowledge discovery during the training phase is more difficult. Data preparation and filtering steps can take considerable amount of processing time. The experimental results reveal that our novel proposed data preprocessing method markedly improved the overall performance in initial dataset comparing with PCA and Evolutionary search feature selection method. In our point of view, our suggested method can be used to significantly boost classification algorithms performance in similar datasets and can be used for suicide tendency prediction.

In future work, it would be preferable to make the same experiments in similar datasets consisting of more records, using different classifiers and different feature selection and data preprocessing methods. In addition, our proposed data preprocessing method could be modified so as to achieve better classification performance.

## 8 References


1. Miles CP. (1977) Conditions predisposing to suicide: A review. J Nerv Ment Dis 164: 231–246
2. Angst, J., Angst, F., Stassen, H.H. (1999). Suicide Risk in Patients with Major Depressive Disorder. Journal of Clinical Psychiatry, 60:57-116.
3. American Psychiatric Association (APA) (2010). Practice Guidelines for the treatment of Patients with Major Depressive Disorder (3rd ed), Copyright 2010.
4. Dimitrios Galiatsatos, Georgia Konstantopoulou, George Anastassopoulos, Marina Nerantzaki, Konstantinos Assimakopoulos and Dimitrios Lymberopoulos, " Classification of the most Significant Psychological Symptoms in Mental Patients with Depression using Bayesian Network", Proc. of the 16th International Conference on Engineering Applications of Neural Networks (EANN 2015), 25 - 28 September 2015.
5. Diagnostic and Statistical Manual of Mental Disorders DSM-V, 5th ed., American Psychiatric Publishing, Washington DC, USA, London England, p.123-154, 2013.
6. S. B. Kotsiantis, D. Kanellopoulos, and P. E. Pintelas, “Data Preprocessing for Supervised Leaning”, World Academy of Science, Engineering and Technology, vol. 1, 2007, pp. 856-861.
7. Theodoros Iliou, Christos-Nikolaos Anagnostopoulos, Marina Nerantzaki and George Anastassopoulos, "A Novel Machine Learning Data Preprocessing Method for Enhancing Classification Algorithms Performance", Proc. of the 16th International Conference on Engineering Applications of Neural Networks (EANN 2015), 25 - 28 September 2015.
8. Waikato Environment for Knowledge Analysis, Data Mining Software in Java, available online: http://www.cs.waikato.ac.nz/ml/index.html, [Accessed 1 May 2016].