Prediction of Revisit of Repeated Attempted-Suicide Patients

Vuttichai Vichianchai*, Sumonta Kasemvilas**,
Sakarin Kaewhao***, and Orapin Youdkang***

Abstract

This works proposes a technique to predict the revisit of repeated attempted-suicide patients. The technique applies the factors relating to suicide and attempted suicide, which are collected from medical treatment information. The proposed technique considers the probability distribution of attempted-suicide dates of the patients in order to examine a pre-determined threshold for classifying the patients into three categories of revisit duration, i.e. (i) low, (ii) medium, and (iii) high. In addition, this work proposes a feature filtering method that can select a set of significance factors from the suicide and self-harm surveillance report (RP. 506S) of Khon Kaen Rajanagarindra Psychiatric hospital to perform the classification. There are 10,112 patients who had been in the services more than once. The filtering is performed before the threshold is determined using a Gaussian function. The experiment results show that the proposed technique is superior to the baseline for every learning algorithms, i.e. (i) k-NN, (ii) SVM, (iii) random forest and (iv) neural networks. In addition, the results obtained from random forest provide promising outcomes. The best performance (in terms of F-measure) is 91.10%, obtained from random forest.

Keyword: Data Classification, Revisit Attempted-suicide, Feature Filtering; Support Vector Machine, k-Nearest Neighbors.

1. Introduction

Suicide is the act of intentionally causing its own death [1]. The causes of suicide come from different factors and conditions such as depression and stress etc. A number of deaths and suicides have increased every year [2]. In Thailand, 16 deaths have been committed in every 100,000 population [3]. This phenomenon is one of the major problems in Thailand that must be solved and prevented. In recent years, a number of researches have been conducted to study and identify suicidal factors [4], [5]; however, these factors cannot be used directly for preventing suicide among the surveillance. In medical treatments, medical records of patients are important information to undergo treatments for mental illness patients. This information is not only used for the treatment purposes, but also can be aided for preventing further possible harms (such as hurting themselves or even suicide) by assessing and evaluating the behaviors of patients, including predicting the possibility of attempted or repeated suicides. However, with the large amount of non-qualitative data, human cannot perform an analysis efficiently, which is prone to inter and intra-reproducibility. For this reason, it is necessary to develop an algorithm that assists the analysis by deploying the data and transform them into qualitative information.

Therefore, this paper is aimed at designing and developing a technique that assess the possibility of the patients (who have repeated attempted-suicide) to revisit and receive treatment services. The revisit can be used to examine patient conditions and for treatment plans. Therefore, this work proposes a technique that applies learning algorithms to classify patients into different groups of revisit duration i.e. (i) low, (ii) medium, and (iii) high. The features rating to the revisit of patients are examined. These features are filtered to produce a final set of discriminative features before the classification is performed.

* *Mahasarakham University.*
** *Khon Kaen University.*
*** *Khon Kaen Rajanagarindra Psychiatric Hospital.*

This paper is organized as follows: Section 2 provides the proposition of related. Section 3 explains the overall data used in this study. In Section 4, the proposed technique is delineated before the experiments are demonstrated in Section 5. The conclusion of the work is given in Section 6.

2. Related work

In general, the factors of suicide can be studied by observing hidden information from selected samples of the people who have attempted suicide. A. Khamma proposed a technique that studied suicide related factors by selecting samples from the suicide population in 2006-2011 [4]. The data was collected in Sukhothai, Muang District, Si Samrong District, Si Satchanalai District and Sawankhalok District by interviewing the patients. There were 330 patients. The collected data was analyzed by using SPSS to analyze quantitative data and risk factors. The work divided the sample (patients) into two groups, i.e. suicide and suicide attempts. In this work, the interview form was modified from the standard interview form for searching death information by Khon Kaen Rajanagarindra Psychiatric hospital, i.e., 1) Suicidal groups were collected by interviewing relatives or deceased close relatives of those who died within 3 months, psychiatric nurses, District Health hospital and Public Health Officer. 2) Suicidal groups were collected by individual interviews during psychiatric treatment sessions with emphasis on ethical principles in human research. The factors and additional information were collected, which are collect age, gender, occupational status, medical history, both physical and psychiatric diseases, problem alcohol use events or event trigger of suicide, including suicide attempts. The results of the research illustrated that males between the ages 31 to 50 years, hiring or labor or farming and quarrels presented a high degree of risk for suicide.

To study the suicide rate and factors related to suicidal behaviors of the Chaophrayayommarat hospital in Suphanburi Province. A research was conducted by collecting the information of the treatment records of the patients [5]. The main objective of the research is to understand the suicidal factors to prevent possible harms. The data collected are the data of who attempt to suicides. There were two ways of collecting data, i.e., 1) Data collection from Emergency department between October 1, 2010 and September 30, 2012. There were 374 patients, including gender, age, occupation, marital status, attempted suicide, suicide, history of physical and mental health problems and information on depression. 2) In-depth interviews of suicide attempters at the medical ward between May 1 and June 30, 2013.

This study investigated five patients with suicidal behaviors. The questionnaire consists of general information, suicide attempt information, family relationship information, illness or substance use, and media mimicking. The results of the research found that 1) the factor of suicide, i.e., male, employed, in the middle age or working age, has a history of self-harm and lives in the area of drug problems, especially cigarettes and alcohol. 2) The factor of suicide attempt, i.e., jobs or a workers, teenagers and early working age, living in rural areas, levels of depression and personal problems with friends.

Risk of suicide classification using machine learning is one of the techniques that have been reported in literature [6]. There are two techniques, i.e., the k-Nearest Neighbors technique [7] and the Linear Classifiers of Gaussian [8]. The conceptual model of suicide attempts [9]. However, it cannot be analyzed because the data used in the analysis is based on interview data and cannot be tested. The problem was solved by using data sets from the Barwon hospital, Australia. The emergency and inpatient departments included 42,000 records and the psychiatric patients included 8,000 records from 2005 to 2012 [10]. This research divides the risk of suicide into three groups: low risk is thought to be suicide, moderate risk and high risk. The data set used for the test included 17,781 records of 7,746 patients, including age, gender, language, religion, occupation, marital status Indigenous status and postal code of origin. The results of the research showed that the proposed framework outperformed the risk assessment tool by medical professionals.

In addition, decision trees [11], regression log analysis [12], random forest techniques [13], gradient boosting machines [14], and deep neural networks [15] area among of the technique that are used to predict the risk of suicide in different period i.e., 15, 30, 60, 180, and 360 days. These techniques have used data from Barwon hospital in Australia [10]. There are 10,000 records for importing information. Measurements have used 25,000 records of mental illnesses resulting from drug addiction and alcohol abuse. The results of the research found that deep neural networks technique was the most effective ways of predicting risk of suicide.

3. Data description

This section provides the details of data used in this study. To evaluate the proposed technique the data is collected and prepared for subsequence processes. The detail is given as follows:

3.1 Data set

The data used in this paper is collected from the suicide and self-harm surveillance report (RP. 506S) from Khon Kaen Rajanagarindra Psychiatric hospital. The data was collected from 2008 to 2016. There are 187,253 records and 164 features (attributes) that are divided into seven groups.

1) Treatments: composing of hospital number (HN), national identification number, first name, last name, gender, age, postal code, marital status, occupation, etc.

2) Services: date of service, the first service, following up visits, home visits, depression, admission, suicidal thoughts, etc.

3) Suicide methods: drugs (overdose, insecticide, herbicides), firearms, hanging, drowning, etc.

4) Trigger events: disappointment, love, studying, poverty, etc.

5) Health problems: diabetes, hypertension, heart disease, AIDS, etc.

6) Social behavior: smoking, alcohol, addiction, gambling, games, etc.

7) Getting help: counseling, guideline documents, etc.

3.2 Errors and missing data

This data set contains 164 attributes, and has a variety of data types, such as numeric data types, numeric data types represent descriptive and textual data types. The RP. 506S has been developed and updated several versions from 2006 to 2017, which alters and changes some of the recorded data resulting inaccurate information (Such as null values, empty values). In addition data missing and non-validated data are still remained in the database. These errors are caused by the following:

1) Allow to record without complete inspection,

2) Programmer errors: invalid data validation, and

3) Inconsistency of some attributes depending on collection periods and versions.

4. Methodology

This section explains the methodology proposed in this research. A generative model will be conducted and used for assessing the prediction of repeated attempted-suicide patients. The overall process is illustrated in Figure 1.

4.1 Research methodology

This step is one of the significant steps in this research. Patient information is confidential and cannot be disclosed. Therefore, patients and hospitals must be approved in accordance with the terms and conditions before we can be used in the research. Identify the headings.

4.2 Data preparation and cleaning

There are 187,253 records and 164 attributes of data set from the suicide and self-harm surveillance report (RP. 506S). Therefore, before data is used into the experiments, the recorded data must be validated and corrected, including null values, empty values and invalid values. There are six methods of verification as follows:

1) Consulting an expert,

2) Verifying the data: for example, prefix name must be matched to gender data. In addition, zip codes can be validated based on the area of the patients,

3) Missing data can be solved by analyzing some similar records in order to rejecting the missing data,

4) Records are removed if they are not in 1-3,

5) Remove records that indicate that the patient was follow up or home visits, and

6) Removed records indicated that the patient died.

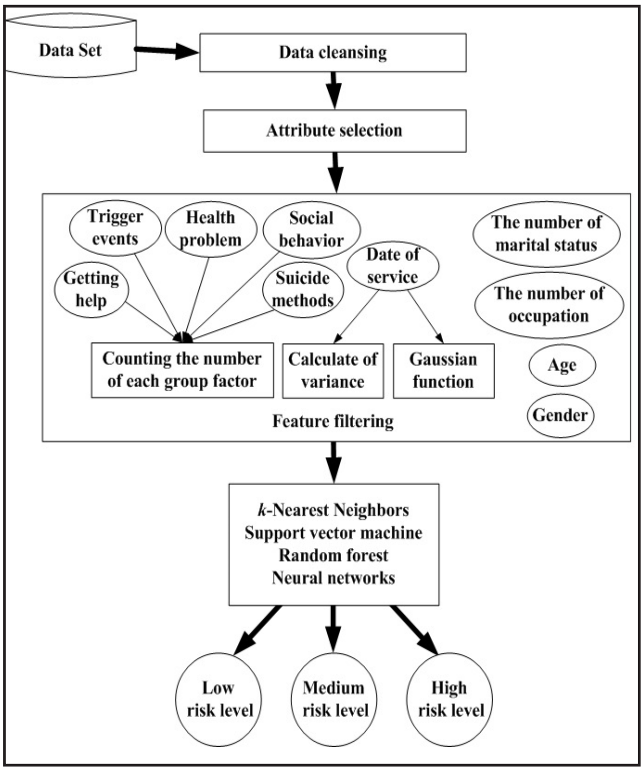

Figure 1. *Shows the conceptual framework of the research.*

4.3 Attribute selection

After completing the cleaning step, this step is recommended by four experts. There are 12 groups of factors that affect suicide (attributes), i.e., province code, gender, age, marital status, date of the suicide, depression, the suicide method, the trigger factor, the health problems, the social behavior of patient, the getting help, admit, and suicidal thinking. If the information states that a patient has died or up visit or home visit, then the information will be deleted.

4.4 Feature filtering

This step will filter the data of each patient into a single record, consisting of the province code, gender, age, the number of marital status, the number of occupations, the number of depression, the number of suicide methods, the number of trigger events to suicide, the number of health problems, the number of social behavior, the number of getting help, the average distance of date suicide, the variance of the distance of date service and the group of risk level, as follows:

1) Choose the maximum age of each patient.

Table 1. *Shown example of the patient information.*

Patient name	Age	Marriage status	Occupation	Depression
Mr. A	45	Single	Government	Yes
Mr. A	46	Married	Government	Yes
Mr. A	46	Married	Government	Yes
Miss R	41	Married	Unemployed	No
Miss R	41	Married	Unemployed	No
Miss R	42	Married	Unemployed	Yes
Miss R	42	Married	Employee	No

Table 1 demonstrates an example the data of patients who revisit the treatment. A patient can revisit and receives the treatment. With multiple visits, therefore, the current age is obtained to represent the age of the patients. For example Mr. A received the treatment with 3 different periods of time. Therefore, the most recent age will be selected to represent the age of Mr. A, which 46. Moreover, 42 will be used to represent the age of Mr. R, as shown in Table 2.

Table 2. *Shows the maximum age of the patients.*

Patient name	The Maximum age
Mr. A	46
Miss R	42

2) Counting the number of marital status of the patients. From Table 1, Mr. A has two marriage statuses, i.e. single and married. On the other hand, Miss. R has only one marriage status, which is married. Therefore, the number marital status of the patients is used in this study. This is to consider often a patient changes his/her marital status overtime. The number of marital status of patients in Table 1 can be summarized in Table 3.

Table 3. *Shows the number of marital status of the patients.*

Patient name	The number of marital status
Mr. A	2
Miss R	1

3) Counting the number of occupations of the patients. Considering Table 1, for example, Mr. A has only one occupation, which is government. On the other hand, Miss. R has two occupations, i.e. unemployed and employee. The number of occupations of patients in Table 1 can be summarized in Table 4.

Table 4. *Shows the number of occupations of the patient.*

Patient name	Number of occupations that patients have done
Mr. A	1
Miss R	2

4) Counting the number of depression value. From Table 1, the depression is assigned to either "yes" or "no". To incorporate this information to the study, the number of depression of patients who have received the treatment is taking into account by counting the number of depression overtime. The summary is in Table 5.

Table 5. *Shows the number of depression.*

Patient name	Mean of depression
Mr. A	3
Miss R	2

5) Counting the number of suicide methods

Table 6. *Shown example of the suicide methods of patient.*

Patient name	Suicide methods			
	Overdose	Firearms	Hanging	Drowning
Mr. A	Yes	No	No	No
Mr. A	Yes	No	No	No
Mr. A	No	Yes	No	No
Miss R	Yes	No	No	No
Miss R	No	No	No	No
Miss R	Yes	No	No	Yes
Miss R	No	Yes	No	No

Table 6 illustrates the method of suicide, for example, Mr. A visits and receives treatment for 3 times. Each time is reported different method of suicide. Mr. A is considered to use 2 different methods of suicide, which are overdose and firearms. On the other hand, Miss. R uses 3 different suicide methods overtime. Table 7 summaries the number of suicide methods of patients in Table 6.

Table 7. *Shows the number of suicide methods of the patients.*

Patient name	The number of suicide methods
Mr. A	2
Miss R	3

6) Counting the number trigger events

Table 8. *Shown example of the trigger events of patient.*

Patient name	Trigger events			
	Disappointment	Love	Studying	Poverty
Mr. A	Yes	No	No	No
Mr. A	Yes	No	No	No
Mr. A	No	Yes	No	No
Miss R	No	No	No	Yes
Miss R	No	No	No	Yes
Miss R	No	No	No	Yes
Miss R	No	No	No	Yes

This is to count the number of trigger events to perform suicide of patients. For example, Mr. A, has 2 main trigger events factors, i.e. dissepiment and love over the period of receiving the treatment. Miss R. has only 1 of trigger events factor. Table 9 shows the summary of the number of trigger events to perform suicide of patients.

Table 9. *Shows the number of trigger events of the patients.*

Patient name	The number of trigger events
Mr. A	**2**
Miss R	**1**

7) Counting the number of health problems relating to suicide

The number of health problem relating to suicide. Table 10 demonstrates the health problems of patients diagnosed during the period of revisiting and receiving the treatment. The number health problems can be an important piece of information that can be used to the classification. Table 11

Table 10. *Shown example of the health problems of patient.*

Patient name	Suicide methods			
	Diabetes	Hypertension	Heart disease	AIDS
Mr. A	Yes	No	No	No
Mr. A	Yes	No	No	No
Mr. A	Yes	No	No	No
Miss R	No	No	No	Yes
Miss R	No	No	No	Yes
Miss R	No	No	No	Yes
Miss R	No	No	No	Yes

summaries the number of health problems (the number of diseases of the patients) of patents from Table 10.

Table 11. *Shows the number of health problems of the patients.*

Patient name	The number of health problems
Mr. A	1
Miss R	1

8) Counting the number of getting help of patients

Table 12. *Shown example of the getting help of patient.*

Patient name	Getting help		
	Counselling	Drug psychosis	Amitriptyline
Mr. A	Yes	No	No
Mr. A	Yes	No	No
Mr. A	Yes	No	No
Miss R	No	Yes	No
Miss R	No	Yes	Yes
Miss R	No	Yes	Yes
Miss R	No	No	Yes

The number of helps gives to the patients. For example, Mr. A receives the treatment 3 times. Every time he has

Table 13. *Shows the number of getting help of the patients.*

Patient name	The number of health problems
Mr. A	1
Miss R	2

9) Calculating the distance of the service dates, then calculate the average distance of service dates. If this is the first visit, set it is 0.

Table 14. *Shown the date of service.*

Patient name	Date of the suicide	Distance of date
Mr. A	5/1/2555	0
Mr. A	2/3/2555	56
Mr. A	1/6/2555	91
Miss R	28/1/2552	0
Miss R	25/2/2552	28
Miss R	25/3/2552	29
Miss R	28/5/2552	64

From the data in Table 14, the average of distance from the date of service is shown in Table 15.

10) Calculating the mean and variance of the duration of service of each patient and used as an additional factors,

Table 15. *Shown the average and variance of the distance between dates of the suicides.*

Patient name	The mean of the distance	The variance of the distance
Mr. A	49.00	1,404.6670
Miss R	30.25	515.1875

4.5 Thresholding

This step will determine the probability of the distribution of revisit of patients, which perform as follows:

1) Randomly divide the data into two sections from data of the revisit of repeated attempted-suicide patients, i.e. the first section, there is 30% from all data to find threshold of the risk classification in suicide, and the remaining 70% is used to test the threshold of revisit.

2) Consider the date of service using the first part of the data to determine the probability and set the risk classification criteria for revisit. The expert counseling on suicide risk level classification by consideration from the date of the suicide of each patient can be divided into 3 levels, i.e. (i) Low-risk patients are those who attempt suicide with a range of very distant days, (ii) Moderate-risk patients are those who attempted-suicide with a range of days that are not very close to each other and (iii) High-risk patients are those who attempted-suicide with a very close range of dates.

Since the day of attempted-suicide, the patient does not have the same distance every time. The distribution of attempted-suicide in each patient was measured using the Gaussian function from the following Equation 1 [16].

$$f(x)=\frac{1}{\sqrt{\pi 2\sigma^2}}e^{-\left[\frac{(d-\mu)^2}{20^2}\right]} \qquad (1)$$

Where μ is the mean value of service duration of each patient and σ is the variance of the duration of dates for all date of the suicide, *f(x)* is the probability of distribution of dates of the suicide, x the current record and d is the duration of the current from the last date of the suicide.

3) Determine the probability of the second group of the distance range of attempted-suicide using the Gaussian function.

4) The probability of the second group calculated is compared with the threshold from the data in the first group to determine the risk group for suicide.

5) Data of the second group has been defined the risk of the suicide and a 10-fold cross validation using k-NN technique is performed.

6) Consider the value of precision and recall to adjust the probability range of the value of the threshold.

7) Repeat steps 1, 3 to 7 until the threshold will get good the value of precision and recall.

In this research, we have adjusted the values of 20 times, so that the appropriate threshold values, which shown in Equation 2.

$$(x)=\begin{cases} L, & f(x)\leq 0.000010 \\ M, & 0.00001 < f(x) < 0.002925 \\ H, & f(x)\geq 0.002925 \end{cases} \quad (2)$$

Where *p(x)* is the average of the probability of distribution of the current date of the suicide of the patient, *L* is a low level of risk to return to suicide, *M* is a moderate level of risk to return to suicide and *H* is a high level of risk to return to suicide [6]. The group of patients is depicted in Table 16.

Table 16. *Shown the probability of the current record and groups for revisit of the patient.*

Patient name	*p(x)*	Group
Mr. A	0.007115583	M
Miss R	0.010130435	H

Then, feature vectors are generated for each patient. There are 16 features in total (13 features are explained in Section 4.3 and the rest are: mean and variance of service duration). Finally, there are 10,112 patients in the dataset.

5. Experimental results and discussions

In order to compare the effectiveness of the proposed technique, all the factors are feed to the learning algorithms without filtering the features (so called baseline). Experiments are carried out using a cross-validation technique. A 10-fold cross validation is deployed and applies SVMs and k-Nearest Neighbors technique to perform a classification. In k-NN, the experiments vary k values (i.e. 1, 3, 5, 7, 9, 11, 13, 15, 17 and 19) in order to determine the efficiency of the technique, obtaining the best performance by determining F-measure. The results of the experiments are shown in Table 17 and 18.

Table 17. *The results of the experiments using k-NN (baseline)*

k	Precision (%)	Recall (%)	F-measure (%)
1	71.50	71.50	71.50
3	72.20	78.20	74.30
5	73.00	80.00	**74.60**
7	74.00	80.70	74.50
9	74.50	80.90	74.30
11	74.50	81.10	74.00
13	75.50	81.20	73.90
15	**76.00**	**81.30**	73.90
17	76.90	81.30	73.80
19	75.90	81.30	73.70

Table 18. *The results of the experiments using k-NN (baseline)*

k	Precision (%)	Recall (%)	F-measure (%)
1	76.10	77.40	76.70
3	77.30	81.00	**78.00**
5	77.80	81.90	77.80
7	78.10	82.30	77.50
9	78.50	82.50	77.30
11	78.40	82.40	76.80
13	**78.80**	**82.50**	76.50
15	78.70	82.40	76.30
17	78.00	82.30	75.90
19	78.30	82.30	75.70

Table 17 and Table 18 demonstrate the performance of the classification using k-NN technique. Feature filtering (the proposed technique) is superior to the baseline for all varied k. The best performance is obtained by the proposed feature filtering technique with k = 3, achieving 78.00% of F-measure.

In addition to k-NN, a SVM is deployed to learn a model and make a prediction [17]. The experiment is carried out and the results are demonstrated in Table 19.

The experimental results shown in Table 19 indicate that the feature filtering method provide better performance than the baseline, in terms of F-measure. In addition, neural networks provide the best results in terms of F-measure. The precision obtained from k-NN, SVM, random forest and

Table 19. *Shows the results of the experiments.*

Technique	Precision	Recall	F-measure
k-NN (Baseline) (k=5)	73.00	80.00	74.60
k-NN (Baseline) (k=15)	76.00	81.30	73.90
k-NN (Feature filtering) (k=3)	77.30	81.00	78.00
k-NN (Feature filtering) (k=13)	78.80	82.50	76.50
SVM (Baseline)	70.80	81.30	73.10
SVM (Feature filtering)	75.40	84.70	79.00
Random forest (Baseline)	66.80	78.60	70.60
Random forest (Feature filtering)	**93.6**	**93.10**	**91.10**
Neural networks (Baseline)	71.80	76.10	73.60
Neural networks (Feature filtering)	90.90	91.00	90.00

neural networks is marginally different. However, the recall provided by random forest is better than all techniques.

The feature filtering method reduces the number of feature in the dataset. The reduction is performed by grouping features or factor that is related. For example, there are 12 suicide methods or procedures that patients used to attempted-suicide. These factors are sparse. Therefore, isolating these features as a single factor may not discriminative enough to separate the class of the data. The filtering method ensembles the related features and thus the results obtained from the method are promising.

6. Conclusions

The data used in this paper is collect from the suicide and self-harm surveillance report (RP. 506S) from Khon Kaen Rajanagarindra Psychiatric hospital. The data was collected from 2008 to 2016. There are 187,253 records and 164 features (attributes) that are divided into seven groups. Then data preparation and cleaning and attribute selection is carried out. There are 10,112 patients who had been in the services more than once, attribute selection, feature selection, Gaussian function applying for feature filtering to determine criteria of the group.

The data is divided into 2 groups. The first group is used to determine the threshold. Then, k-NN is performed using a 10-fold cross validation. Then, a grid search is carried out to obtain the best threshold.

The experiment results show that the proposed technique is superior to the baseline for every learning algorithms, i.e. (i) k-NN, (ii) SVM, (iii) random forest and (iv) neural networks. In addition, the results obtained from random forest provide promising outcomes. The best performance (in terms of F-measure) is 91.10%, obtained from random forest.

Reducing the number of feature from 164 to 13 features provides promising result, especially when using random forest to classify. By determining the significant of the factors using information gain, the 3 top most significant features relating to the type of revisit group are occupation, trigger event and health problem respectively.

The future work will be focusing on the reducing the groups of risk of revisit by eradicating the duration of visits or receive the treatments of patients.

7. References

[1] T. L. Stedman and J. H. Dirckx. "*Stedman's concise medical dictionary for the health professions*." *Lippincott Williams & Wilkins*, 2001.

[2] W. H. Organization. "World Health Statistics 2016: Monitoring Health for the SDGs Sustainable Development Goals." *World Health Organization*, 2016.

[3] A. Mongkol. *Thailand suicide rate to one hundred thousand people in 2016*. Available Online at https://www.suicidethai.com/report/download/dl.asp?id=21, accessed on 2017

[4] A. Khamma, "Risk factors associated with suicide: a case-control study in Sukhothai Province." *J Psychiatr Assoc Thai*, Vol. 58, pp. 3-16, 2013.

[5] W. Prateeptheeranun. "Rate of Suicide and Factors Related to Suicidal Behavior at Chaophayayommarat Hospital Suphanburi Province." *The Journal of Psychiatric Nursing and MentalHealth*, Vol. 28, pp. 90-103, 2014.

[6] J. Han and M. Kamber. "Mining Frequent Patterns, Associations, and Correlations." *Data Mining: Concepts and Techniques*. San Francisco, USA: Morgan Kaufmann Publishers, 2nd ed., pp. 227-283, 2006.

[7] B. Yegnanarayana. "Artificial neural networks." *PHI Learning Pvt. Ltd.*, 2009.

[6] T. Tran, D. Phung, W. Luo, R. Harvey, M. Berk, and S. Venkatesh. "An integrated framework for suicide risk prediction." *In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge discovery and data mining*, pp. 1410-1418, 2013.

[7] Y. Wu, K. Ianakiev, and V. Govindaraju. "Improved k-nearest neighbor classification." *Pattern recognition*, Vol. 35, pp. 2311-2318, 2002.

[8] D. M. Tax, M. Van Breukelen, R. P. Duin, and J. Kittler. "Combining multiple classifiers by averaging or by multiplying?." *Pattern recognition*, Vol. 33, pp. 1475-1485, 2000.

[9] F. Ruiz, I. Valera, C. Blanco, and F. Perez-Cruz. "Bayesian nonparametric modeling of suicide attempts." *In Advances in neural information processing systems*, pp. 1853-1861, 2012.

[10] T. Nguyen, T. Tran, S. Gopakumar, D. Phung, and S. Venkatesh. "An evaluation of randomized machine learning methods for redundant data: Predicting short and medium-term suicide risk from administrative records and risk assessments." *arXiv preprint arXiv:1605.01116*, 2016.

[11] R. Lior. "Data mining with decision trees: theory and applications." *World scientific*, Vol. 81, 2014.

[12] D. Zelterman. "Bayesian artificial intelligence." *ed: Taylor & Francis*, 2005.

[13] C. Zhang and Y. Ma. *Ensemble machine learning: methods and applications*, Springer, 2012.

[14] J. Gao, Y. Su, Q. Wu, C. J. Burges, K. Svore, and E. R. T. Abib. "Boosting algorithm for ranking model adaptation." ed: Google Patents, 2012.

[15] N. Kriegeskorte. "Deep neural networks: a new framework for modeling biological vision and brain information processing." *Annual review of vision science*, Vol. 1, pp. 417-446, 2015.

[16] C. Bishop. "Pattern Recognition and Machine Learning." *Springer-Verlag New York*, 1 ed. Vol. 1 2006.

[17] Y. Ma and G. Guo. "Support vector machines applications." *Springer*, 2014.