International Journal of Engineering &Technology, 7 (4) (2018) 6873-6877

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet.v7i4.28338
Research paper

Support vector machines and twin support vector machines for classification of schizophrenia data

Zuherman Rustam [1*] , Theresia V. Rampisela [2]

[1] Department of Mathematics, University of Indonesia, Depok 164242, Indonesia
**Corresponding author E-mail: rustam@ui.ac.id*

Abstract

Schizophrenia is a disorder characterized by disturbances in thoughts, perceptions, and behaviors, making it a severe, chronic mental illness. Unfortunately, it is difficult to diagnose it due to lack of physical test, apart from its symptoms resembling that of several other mental illnesses. A former study has used the Northwestern University Schizophrenia Data to identify Schizophrenics from non-Schizophrenics using Support Vector Machines (SVM). Contrastingly, this research used a method that has never been used in Schizophrenia-related problems, the Twin SVM approach. The strategy was employed in classifying the data and comparing the results between various studies. In addition, it successfully classified Schizophrenia data in an accurate manner compared to the SVM method previously used. Precisely, the Twin SVM with Gaussian kernel produced the best final accuracy in classifying Schizophrenia data, 91.0%.

Keywords: *Classification Problem; Machine Learning; Schizophrenia; Support Vector Machines; Twin Support Vector Machines.*

1. Introduction

Schizophrenia is severe and chronic mental disorder marked by disturbance in thinking, perception, and behavior. The symptoms of this condition are grouped into three; positive, negative, and cognitive. The first one, positive symptoms, includes nonexistent brain activities such as hallucination and delusion. On the contrary, negative symptoms encompass brain acts that should exist but are lacking, for example, the absence of emotion and apathy. Lastly, cognitive symptoms relate to memory disturbances and concentration difficulties [1]. These symptoms significantly impair the quality of life and the productivity of Schizophrenics [2].
In the recent past, about ten percent of Schizophrenics have died from suicide [3], an occurrence that has reduced the life expectance of the victims up to three times lower than that of the population in its entirety [2]. Moreover, 20-40% of Schizophrenics have at least attempted suicide [3]. Even though the root cause of the disorder has not been exactly determined, it can be suffered by anyone [2]. The fact that there is no radiology, laboratory, or psychometric test to diagnose the disease, and the similarities of its symptoms with other mental disorders, for example, Major Depression and Bipolar, makes its diagnosis quite difficult [1]. To this end, it is factual to conclude that Schizophrenia needs an easy, objective, and accurate diagnostic method that will ensure an efficient treatment to the victims. One of the best diagnostic approaches is Machine learning.
Machine learning is an appropriate strategy, capable of studying the patterns of training data and use the results obtained to predict the testing variables [4]. With this approach, the use of psychological evaluations to detect Schizophrenia is a classification problem. Notably, several studies have tried to utilize machine learning in diagnosing this disorder and its related matters. Most of them have employed the following machine learning approaches: Fisher Linear Discriminant Analysis [5], Linear Discriminant Analysis and k-Nearest Neighbor [6], Support Vector Machines [7], [8], Elastic Net, and Least Absolute Shrinkage and Selection Operator [8]. Nevertheless, only Support Vector Machines and Elastic Net made use of psychological evaluation results.
The Support Vector Machine (SVM) is a binary classification method which creates a model that can generalize well [9] with optimum global solution [10]. In SVM, a hyperplane is created to separate the two target values so that the nearest distance between data and margin is maximized. In this research, SVM was used since it has a better generalization potential and accuracy [9]. Other than being applied in Schizophrenia-related studies, the method has also been used in other classification problems, for instance, in brain cancer [11], insolvency prediction in insurance companies [12], policyholders' satisfactory [13], and in intrusion detection system [14].
Besides, there is another machine learning method, Twin Support Vector Machines (Twin SVM) that has not been tried in solving Schizophrenia-related problem, though it is an improvement to the SVM method [15]. However, it has been used in various studies, including in the classification of Alzheimer's disease [16]. In this study, there was comparison in the performance of SVM, as in a former research [7], against Twin SVM in classifying Schizophrenia data by means of simulations.

2. Classification of schizophrenia data

2.1. Schizophrenia data

This study used Northwestern University Schizophrenia Data [17], specifically the clinical variables from the database. The variables included the group, demographic (gender, dominant hand, ethnic, race, and age), and questionnaires statistics from the Scale for the Assessment of Negative Symptoms (SANS) [18] and Scale for the Assessment of Positive Symptoms (SAPS) [19]. There were a total of 392 observations and 65 variables in the data, out of which 171 samples were labeled "Schizophrenics" and 221 "non-

Schizophrenics". In addition to SANS and SAPS, there were other scales used to evaluate Schizophrenia symptoms, the Brief Psychiatric Rating [20], Calgary Depression Scale for Schizophrenia [21], and Positive and Negative Syndrome [22].

The SANS is an inventory used to evaluate the negative symptoms of Schizophrenia. It is divided into five main parts with a total of 25 different symptoms, measured from 0 (no symptoms) to 5 (severe). The five parts covered the following niches: emotional reaction decline, alogia, avolition and apathy, anhedonia and asociality, and attention [19]. Besides, SANS is closely related to SAPS, a scale used to measure the positive symptoms of Schizophrenia. However, it was made up of four different parts consisting of 34 distinct symptoms measured in the same way as SANS. These parts include hallucination, delusion, bizarre behavior, and thought disorder [19].

2.2. Kernel function

Kernel function was defined by $K(\boldsymbol{x_i}, \boldsymbol{x_j}) = < \varphi(\boldsymbol{x_i}), \varphi(\boldsymbol{x_j}) >$, where $\varphi(x)$ is a function that maps $\boldsymbol{x} \in \mathbb{R}^d$ to feature space $\mathbb{F}$. Therefore, when the inner product $< \varphi(\boldsymbol{x_i}), \varphi(\boldsymbol{x_j}) >$ appears in classification algorithm, it can be replaced by $K(\boldsymbol{x_i}, \boldsymbol{x_j})$ [15]. By using kernel function, nonlinear data is expected to be separated linearly in a higher dimension. This research used the Gaussian kernel, $K(\boldsymbol{x_i}, \boldsymbol{x_j}) = exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$ [24].

2.3. Support vector machines (SVM)

Let a set of m row vectors, $\boldsymbol{A_i}, (i = 1,2, \dots, m) \in \mathbb{R}^d$ represent the patterns to be classified, and $y_i \in \{1, -1\}$ denotes the class of the i-th pattern. For the data that can be separated linearly, $\boldsymbol{w} \in \mathbb{R}^d$ and $b \in \mathbb{R}$ are determined so that: $\boldsymbol{A_i w} \geq 1 - b$ for $y_i = 1$ and $\boldsymbol{A_i w} \leq -1 - b$ for $y_i = -1$ [25]

The $\boldsymbol{w^T x} + b = 0$ is in between two other planes: $\boldsymbol{w^T x} + b = 1$ and $\boldsymbol{w^T x} + b = -1$. The plane (1) separates the two classes by a margin of $\frac{1}{\|w\|}$ on each side. By maximizing the margin, the SVM model is obtained and is equivalent to the following problem [15]:

(SVM 1) $$\min_{w,b} \frac{1}{2}\boldsymbol{w^T w}$$

Subject to:

$$\boldsymbol{A_i w} \leq -1 - b, y_i = -1 \tag{1}$$

$$\boldsymbol{A_i w} \geq 1 - b, y_i = 1 \tag{2}$$

If the two classes are not able to be linearly separated, (1) and (2) may have errors on several patterns and therefore can be modified into:

$$\boldsymbol{A_i w} + q_i \geq 1 - b \; for \; y_i = 1 \tag{3}$$

$$\boldsymbol{A_i w} - q_i \leq -1 - b \; for \; y_i = -1 \tag{4}$$

$$q_i \geq 0, i = 1,2, \dots, m \tag{5}$$

The variable q_i represents the error of the i-th sample. Since this classification method separates data with some errors, it is considered a soft margin model that dependents on the value of q_i. The class of the testing sample is obtained by determining the sign of $\boldsymbol{w^T x + b}$ [15]

The formulation of SVM with soft margin is given by:

(SVM 2) $$min \; C\boldsymbol{e^T} q + \frac{1}{2}\boldsymbol{w^T w}$$

Subject to:

$$\boldsymbol{A_i w} + q_i \geq 1 - b \; for \; y_i = 1 \tag{5}$$

$$\boldsymbol{A_i w} - q_i \leq -1 - b \; for \; y_i = -1 \tag{6}$$

$$q_i \geq 0, i = 1,2, \dots, m \tag{7}$$

The variable $\boldsymbol{e}$ denotes a vector whose entries are all 1, while $(.)^T$ represents transpose, as the scalar C depicts a trade-off parameter. A smaller value of C means that the model would prioritize the minimization of **w**, while a larger one would opt for the minimization of $\boldsymbol{q}$ [15]. By solving (SVM 2), the variable w and b can be calculated using [12]:

$$\boldsymbol{w} = \sum_{i=1}^{n} a_i^* y_i \, x_i \tag{8}$$

$$b = \frac{1}{N_s} \sum_{i \in S} (y_i - \sum_{m \in S} (a_m y_m x_m)) \tag{9}$$

2.4. Twin support vector machines (twin SVM)

Twin SVM forms two hyperplanes unparallel to each other: $\boldsymbol{x^T w^{(1)}} + b^{(1)} = 0$ dan $\boldsymbol{x^T w^{(2)}} + b^{(2)} = 0$. On each plane, data from each class lie. Let the data from class 1 and -1 be notated with matrices A and B and the number of patterns from each class be m_1 and m_2 respectively, and n the number of features. [15]

The Twin SVM model is obtained by solving these quadratic programming problems:

(Twin SVM + 1) :

$$min_{\boldsymbol{w}^{(2)}, b^{(2)}, q^{(1)}} \frac{1}{2} \left(\boldsymbol{A w^{(1)}} + \boldsymbol{e_1} b^{(1)}\right)^T \left(\boldsymbol{A w^{(1)}} + \boldsymbol{e_1} b^{(1)}\right) + C_1 \boldsymbol{e_2^T} q^{(2)}$$

Subject to:

$$-(\boldsymbol{A w^{(2)}} + \boldsymbol{e_1} b^{(2)}) + q^{(1)} \geq \boldsymbol{e_1} \tag{10}$$

$$q^{(1)} \geq 0 \tag{11}$$

(Twin SVM − 1) :

$$min_{\boldsymbol{w}^{(1)}, b^{(1)}, q^{(2)}} \frac{1}{2} \left(\boldsymbol{B w^{(2)}} + \boldsymbol{e_2} b^{(2)}\right)^T \left(\boldsymbol{B w^{(2)}} + \boldsymbol{e_2} b^{(2)}\right) + C_2 \boldsymbol{e_1^T} q^{(1)}$$

Subject to:

$$-(\boldsymbol{B}\boldsymbol{w}^{(1)} + \boldsymbol{e_2} b^{(1)}) + q^{(2)} \geq \boldsymbol{e_2} \quad (12)$$

$$q^{(2)} \geq 0 \quad (13)$$

Where $C_1, C_2 > 0$ are trade-off parameters, and $\boldsymbol{e_1}, \boldsymbol{e_2}$ are vectors with only the number 1 as their entries [15]

The Twin SVM algorithm generated two hyperplanes, each for one class. It classified variables according to the hyperplane nearer to the data. The first term in the objective function is the quadratic sum of distance between a hyperplane and data of one class and thus, minimizing such term would result in a hyperplane being nearer to the data from one class. In addition, the second term minimizes the classification error caused by data from the other class. The constraints given also require each hyperplane to be at least one unit away from the data belonging to the other class [15].

A new data $\mathbf{x} \in \mathrm{R}^n$ will be classified as a class r (r $= 1{,}2$) based on which hyperplane is closest, or in mathematical equation:

$$min_{r=1,2} \boldsymbol{x}^{\boldsymbol{T}} \boldsymbol{w}^{(r)} + b^{(r)} = \left| \boldsymbol{x}^{\boldsymbol{T}} \boldsymbol{w}^{(l)} + b^{(l)} \right| \quad (14)$$

Where $|\cdot|$ is the perpendicular distance between $\boldsymbol{x}$ and the plane [15].

2.5. Optimization of parameters

In this study, using the grid search method, several parameters were optimized. The grid search method works by finding procedurally the combination of parameters that generates the optimum model [25]. The optimized parameters in the research were: C on the interval $[10^{-3}, 10^{3}]$ by logarithmic scale on the SVM method, and C_1, C_2 on $10^j, j = -5, -4, \dots, 1$ for the Twin SVM method. Unlike the previous works, here the parameter σ^2 was also optimized: $\sigma^2 = 10^k, k = -3, -2, \dots, 3$ for the Gaussian kernel in SVM and $\sigma^2 = 10^k, k = -3, 0, 1, 3$ in Twin SVM.

2.6. Validation of model performance

Hold-Out Validation was conducted to validate the model performance. In this method, the data is randomly separated into two parts: training and testing data. The testing data is be used to validate the model. Hold-Out Validation is easy and fast [26]. In this study, Hold-Out Validation was done on different data percentages. To overcome the shortcomings of this method, that is the overdependence on the data used for training and testing [26], simulations were performed for up to 10 times with a different set of data for each percentage of the variable used.

2.7. Evaluation of model performance

Model performance evaluation was measured by accuracy, a measure of how often the classification model makes a correct prediction. Simply put, accuracy is the ratio between the number of correct predictions and their total number [26]. A classification model performance is considered good if it has a high accuracy [27].

3. Results and analysis of simulation

The results and analysis of classification of Schizophrenia data with SVM and Twin SVM is presented in this section. The software MATLAB R2017a was used to create the SVM and Twin SVM models. The later used the program found on http://jayadeva.net/wp-content/uploads/2015/03/Twin-SVM-Code.zip with some modifications.

A study [7] used the same methodology on the same Schizophrenia data. The following were the results of using SVM with linear kernel juxtaposed with the simulation results from the newly-done research using SVM with Gaussian kernel with optimized parameter and Twin SVM: with reference to Figure 1, it can be seen that an optimized value of σ^2 may improve the classification result significantly when using a low percentage of training data (10-30%) in SVM. Interestingly, different values of Gaussian kernel parameter produced significantly dissimilar results when using a high percentage of training data (80-90%) in Twin SVM as seen in Figure 2.

From Table 1, more training data used in building the model resulted in a higher final accuracy. The SVM model with linear kernel that produced the best final accuracy for the classification of Schizophrenia data was the one that used 70% training data with a mean of 90.1%. In contrast, the worst model was the one that used 10% training data with a mean of 88.1% [7]. Moreover, the SVM model with optimized Gaussian kernel parameter that performed best in the classification of Schizophrenia variables was the one with 70% or 80% training data, producing a mean of 90.1% accuracy. On the other hand, the worst model was the one using 10% training data with a mean of only 88.3%. The optimized Gaussian SVM model also managed to perform better than in a past study on the models that used 20-40% training data.

The results of Schizophrenia data classification using Twin SVM had a similar trend compared to the one using SVM. From Table I, it can also be seen that the Twin SVM method in general achieved a higher accuracy in classifying Schizophrenia data when using a higher number of training data, though there were few exceptions. Its model with linear kernel giving the best final accuracy for the classification of Schizophrenia data utilized 80% training data with a mean of 90.4%, while the worst model was the one with either 10% or 30% training data with a mean of 89.0%. In a similar way, the Twin SVM model with Gaussian kernel with the classification of Schizophrenia data was the one with 80% training data, producing a mean accuracy of 90.3%, whereas the worst model used either 10, 30, or 40% training data with a mean of only 88.9%.

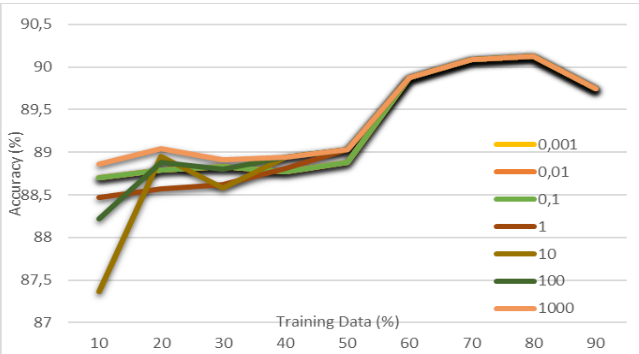

Fig. 1: Accuracy of Schizophrenia Data Classification Using SVM with Gaussian Kernel Using Different Values of Kernel Parameter.

Fig. 2: Accuracy of Schizophrenia Data Classification Using Twin SVM with Gaussian Kernel Using Different Values of Kernel Parameter.

Table 1: Accuracy of Schizophrenia Data Classification Using SVM and Twin SVM with Linear and Gaussian Kernel

Training Data (%)	Linear SVM (%)	Gaussian SVM (%)	Linear Twin SVM (%)	Gaussian Twin SVM (%)
10	88.1 ± 1.47	88.3 ± 0.55	89.0 ± 0.59	88.9 ± 0.00
20	89.0 ± 1.22	88.9 ± 0.10	89.1 ± 1.32	89.0 ± 0.00
30	88.8 ± 1.09	88.8 ± 0.11	89.0 ± 1.35	88.9 ± 0.00
40	88.8 ± 1.27	88.9 ± 0.09	89.1 ± 1.24	88.9 ± 0.00
50	89.0 ± 1.39	89.0 ± 0.08	89.1 ± 1.40	89.1 ± 0.05
60	89.9 ± 1.68	89.9 ± 0.00	89.9 ± 1.72	89.9 ± 0.07
70	90.1 ± 2.30	90.1 ± 0.00	90.3 ± 2.12	90.1 ± 0.00
80	90.1 ± 2.79	90.1 ± 0.00	90.4 ± 2.61	90.3 ± 0.25
90	89.8 ± 4.48	89.8 ± 0.00	90.3 ± 4.16	90.1 ± 0.63

4. Discussion

On the other hand, the result of classifying Schizophrenia data with Twin SVM has a similar trend compared to the classification results using SVM; from Figure 3 it can also be seen that the Twin SVM method in general also achieves a higher accuracy in classifying Schizophrenia data when it uses a higher number of training data, with the exception of some cases. The linear Twin SVM model that uses 80% training data gives the best accuracy with a mean accuracy of 90.4%, while the worst model uses 10% or 30% training data with a mean of 89.0%. Likewise, the Twin SVM model with Gaussian kernel that performs the best for the classification of Schizophrenia data is the one that uses 80% training data, producing a mean accuracy of 90.3%, while the worst ones use 10, 30, or 40% training data with a mean of only 88.9%.

Fig. 3: Comparison of Schizophrenia Data Classification Accuracy on Different Percentages of Training Data.

Comparing the results from the two methods with two different kernels and parameters, it is clear that the Twin SVM models outperformed the SVM method since it achieved a better accuracy in classifying Schizophrenia data using the linear and Gaussian kernel. Even the worst-performing Twin SVM models still performed better than the worst-performing SVM methods.

5. Conclusion and recommendations

The classification problem of Schizophrenia data using the Northwestern University Schizophrenia Data can be solved by means of machine learning. There were three types of simulations done in this study: SVM with Gaussian kernel, Twin SVM with linear and Gaussian kernel. In each of them, grid search method was used to optimize the parameters of the model as well as the kernel parameter, while Hold-Out Validation was used to verify the model.

This study established that the Twin SVM method with linear and Gaussian kernel has a similar result, even though the Twin SVM approach with linear kernel performs slightly better the one with Gaussian kernel on average. When compared to a past study that used the same data and methodology but with the SVM method, the Twin SVM method used here posted a better accuracy.

For future studies, it will be prudent to have more simulations using diverse dataset, kernels, and evaluation as well as validation methods. With this research, it is our hope that the results may be helpful, especially in the medical field, in diagnosis of Schizophrenia. This will facilitate the provision of the appropriate treatment and enhance the quality of the lives of the victims.

Acknowledgement

The authors would like to express gratitude to the Indonesian Ministry of Research and Higher Education for financial support, with PDUPT 2018 research grant scheme ID number 389/UN2.R3.1/HKP05.00/2018. Also, sincere appreciations for the support in data collection and sharing goes to NIMH grant 1R01 MH084803.

References

[1] American Psychiatric Association 2013 *Diagnostic and statistical manual of mental disorders, fifth edition* (Arlington, VA: American Psychiatric Association) https://doi.org/10.1176/appi.books.9780890425596.

[2] World Health Organization 2018 *Schizophrenia Key Facts* retrieved from http://www.who.int/en/news-room/fact-sheets/detail/schizophrenia on 17 May 2018.

[3] Schizophrenia.com 2010 *Schizophrenia symptoms and diagnosis of Schizophrenia* retrieved from http://schizophrenia.com/szfacts.htm on 1 March 2018.

[4] Witten I H & Frank E 2005 *Data mining: practical machine learning tools and techniques, second edition* (Burlington: Elsevier Inc).

[5] Neuhaus A H, Popescu F C, Bates J A, Goldberg, T E, & Malhotra, A K 2013 Single-subject classification of schizophrenia using event-related potentials obtained during auditory and visual oddball paradigms *European Archives of Psychiatry and Clinical Neuroscience* 263(241). https://doi.org/10.1007/s00406-012-0326-7.

[6] Ahn M, Hong J H, & Jun S C 2012 Feasibility of approaches combining sensor and source features in brain-computer interface *Journal of neuroscience methods* 204(1), pp 168–78. https://doi.org/10.1016/j.jneumeth.2011.11.002.

[7] Rampisela T V & Rustam Z 2018 Classification of Schizophrenia data using Support Vector Machine (SVM), *Journal of Physics: Conference Series 1108 012044* https://doi.org/10.1088/1742-6596/1108/1/012044.

[8] Hettige N C, Nguyen T B, Yuan C, Rajakulendran T, Baddour J, Bhagwat N, Bani-Fatemi A, Voineskos A N, Mallar Chakravarty M, De Luca V 2017 Classification of suicide attempters in schizophrenia using sociocultural and clinical features: A machine learning approach *Gen Hosp Psychiatry* 47, pp 20-8 https://doi.org/10.1016/j.genhosppsych.2017.03.001.

[9] Cortes C & Vapnik V N 1995 Support-vector networks *Machine Learning* 20(3), pp 273–97 https://doi.org/10.1007/BF00994018.

[10] Burges C J C & Crisp D J 1999 Uniqueness of the SVM solution *Neural Information Processing Systems* 12, pp 223-9.

[11] Panca V & Rustam Z 2017 Application of machine learning on brain cancer multiclass classification *AIP Conference Proceedings* 1862 (1), 030133 https://doi.org/10.1063/1.4991237.

[12] Rustam Z & Yaurita F 2018 Insolvency prediction in insurance companies using Support Vector Machines and Fuzzy Kernel C-Means *Journal of Physics: Conference Series 1028(1),* 012118 https://doi.org/10.1088/1742-6596/1028/1/012118.

[13] Rustam Z & Ariantari N P A A 2018 Support Vector Machines for classifying policyholders satistfactorily in automobile insurance *Journal of Physics: Conference Series 1028(1), 012005* https://doi.org/10.1088/1742-6596/1028/1/012005.

[14] Rustam Z & Zahras D 2018 Comparison between Support Vector Machine and Fuzzy C-Means as classifier for intrusion detection system *Journal of Physics: Conference Series 1028(1), 012227* https://doi.org/10.1088/1742-6596/1028/1/012227.

[15] Jayadeva, Khemchandani R, & Chandra S 2007 Twin Support Vector Machines for pattern classification *IEEE Transactions on Pattern*

Analysis and Machine Intelligence 29(5), pp 905–10 https://doi.org/10.1109/TPAMI.2007.1068.
[16] Alam S, Kwon G-R, Kim, J-I, & Park C S, 2017 Twin SVM-Based Classification of Alzheimer's Disease Using Complex Dual-Tree Wavelet Principal Coefficients and LDA *Journal of Healthcare Engineering,* vol. 2017, https://doi.org/10.1155/2017/8750506.
[17] Wang L, Kogan A, Cobia D, Alpert K, Kolasny A, Miller M I, & Marcus, D 2013 Northwestern University Schizophrenia Data and Software Tool (NUSDAST) *Frontiers in Neuroinformatics* 7, p 25 https://doi.org/10.3389/fninf.2013.00025.
[18] Andreasen N C 1984 *Scale for the Assessment of Negative Symptoms* (SANS) (Iowa City: University of Iowa).
[19] Andreasen N C 1984 *Scale for the Assessment of Positive Symptoms* (SApS) (Iowa City: University of Iowa).
[20] Overall J E & Gorham D R 1962 The Brief Psychiatric Rating Scal *Psychological Reports* 10(3), 799-812. https://doi.org/10.2466/pr0.1962.10.3.799.
[21] Addington D, Addington J & Maticka-Tyndale E 1993 Assessing depression in schizophrenia: the Calgary Depression Scale *British Journal of Psychiatry Supplementary* (22) 39-44. https://doi.org/10.1192/S0007125000292581.
[22] Kay S R, Fizbein A, Opler L 1987 The positive and negative syndrome scale (PANSS) for schizophrenia *Schizophrenia Bulletin* 13(2):261-76 https://doi.org/10.1093/schbul/13.2.261.
[23] Witten I H, Frank E, & Hall M A 2011 *Data mining: practical machine learning tools and techniques, third edition* (Burlington: Elsevier Inc).
[24] Xue H, Yang Q, Chen S 2009 *The top ten algorithms in data mining* (New York: Chapman and Hall/CRC).
[25] Lin C-J, Hsu C-W, & Chang C-C 2003 *A Practical Guide to Support Vector Classification.*
[26] Zheng A 2015 Evaluating machine learning models: a beginner's guide to key concepts and pitfalls (Sebastopol, CA: O'Reilly Media, Inc).
[27] Tan P N, Steinbach M, & Kumar V 2006 Introduction to Data Mining (Boston: Pearson).