

Simulated annealing for SVM parameters optimization in student's performance prediction

Esraa A. Mahareek, Abeer S. Desuky, Habiba Abdullah El-Zhni
Faculty of Science Department of Mathematics, Al-Azhar University, Egypt

Article Info

Article history:

Received Sep 29, 2020
Revised Feb 17, 2021
Accepted Apr 8, 2021

Keywords:

Bio-inspired optimization
Educational data
Educational decision making
Prediction
Simulated annealing
Students' performance
SVM parameter optimization
University education

ABSTRACT

High education is an important and critical part of education all over the world. In last year, the world has been turned increasingly to online education due to the outbreak of the Covid-19 pandemic; therefore, improving this education system became an urgent matter. Online learning systems are a primal environment for acquiring educational data which can be from different sources, especially academic institutions. These data can be mainly used to analyze and extract utilizable information to help in understanding university students' performance and identifying factors that affect it. To extract some meaningful information from these large volumes of data, academic organizations must mine the data with high accuracy. In this work, three different real datasets were selected, pre-processed, cleaned, and filtered for applying support vector machine (SVM) with multilayer perceptron kernel (MLP kernel) and optimize its parameters using simulated annealing (SA) algorithm to improve the objective function value. While examining the search space, SA has the advantage of escaping from local minima since it offers the chance for accepting the worse neighbor as a solution in a controlled manner. The results show that the designed system can determine the best SVM parameters using SA and therefore presents better model evaluation.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Esraa Alaa Mahareek
Faculty of Science Department of Mathematics
Al-Azhar University, Egypt
Email: esraa.mahareek@azhar.edu.eg

1. INTRODUCTION

COVID-19 known as a pandemic according to was declared by the World Health Organization (WHO) in the past year. This pandemic disrupted the education across the globe, as nationwide closures forced institutions to temporarily shut down. It is estimated that the closures affected about 70% of the total student population worldwide. Data mining algorithms are widely used for discovering the hidden patterns of data to help the decision-makers, it became an efficient tool to find the uncovered information from the big data. Like business organizations [1], universities are operating today in a highly dynamic and strongly competitive environment [2] and the education nowadays is not limited to classroom teaching only but it goes away to other forms such systems of online education, web-based education, seminar, project-based learning, workshops, etc. Data mining is very important in educational systems as shown in Figure 1, but all these systems can not success without accurate evaluation so, for having a successful education system a well-defined and accurate evaluation system must be maintained, the prediction of the students' performance with high accuracy is too helpful for selecting the students with low-performance levels from the beginning of

learning process. Large volumes of students' data are collected by modern universities, which are used for improving the educational process.

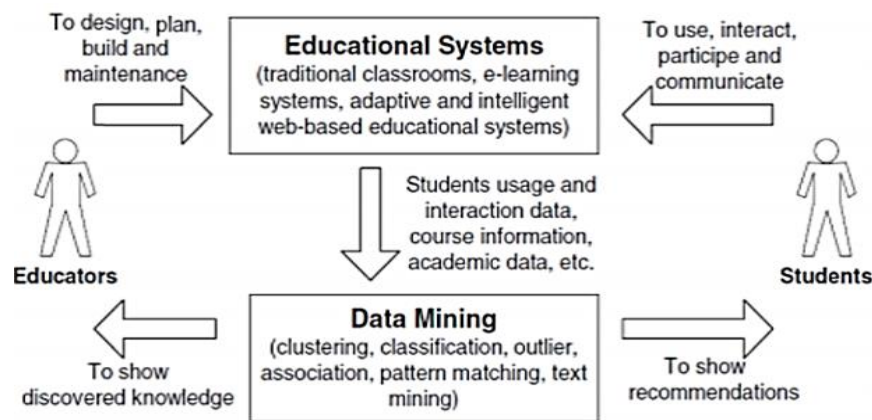


Figure 1. Data mining cycle for educational systems [1]

Both data mining and machine learning used the same methods. But there is a difference, machine learning focuses on prediction, based on known properties, whereas data mining focuses on the identification of unknown properties. Support vector machine (SVM) is a machine learning technique that builds a linear binary classifier. It defines the decision boundary between two classes [2]. Optimization is the process of achieving the best solution for a problem, there are many optimization algorithms like the standard SPSA algorithm [3], [4] which used for optimizing systems with multiple unknown parameters, Gradient descent which used for finding a local minimum of a differentiable function [5], and simulated annealing (SA) which used for approximating the global optimum of a given function in a large search space [6]. Therefore, it was chosen for this work.

SA is a popular optimization algorithm inspired from the basis of melted metals' annealing (slow cooling after heating) to crystallize their structures [7], it was invented in 1983 by Kirkpatrick *et al.* and they- and also some other researchers- analytically proved that SA can escape from the local optima and converges to the global optimum. A group of researchers studied data from students over the past decade to predict student performance. Data mining approaches and correlation analysis each of these approaches generate different levels of success. V. K. Pal and Vimal Kamlesh Kumar Bhatt [8] proposed research on the first dataset by applying the artificial neural network (deep learning) after splitting the data into two subsets, training set containing 70% of original data and test set containing the remaining 30%. The resulted accuracy for test set is 97.749% and the corresponding error rate is 2.251%. Y. K. Salal, S. M. Abdullaev and Mukesh Kumar [9] also proposed research for building classification models for the same dataset and implement algorithms like NaiveBayes with accuracy 73.1895%, decision tree (J48) with accuracy 76.57%, Randomtree with accuracy 67.95%, REPTree with accuracy 76.73%, JRp with accuracy 74.11%, OneR with accuracy 76.73%, simplelogistic with accuracy 73.65% and ZeroR with accuracy 30.97%. After implementing these algorithms on the student performance dataset, He compares the implementation result for the best model in the prediction process.

Another research proposed a classification method based on a meta-heuristic PSO algorithm to predict the students' final outcome according to their activities and the results improved by 89% [10]. D. Kabakchieva [11] also proposed an algorithm for classification by applying four different classifiers: OneR Rule Learner, Neural Network, Decision Tree, and K-Nearest Neighbour, and neural network achieved the highest classification accuracy 73.9%, followed by 72.74% for the Decision Tree and 70.49% for the k-NN model. S. Hussain, Neama Abdulaziz Dahan, Fadl Mutaheer Ba-Alwi and Najoua Ribata [12] used classification algorithms in WEKA and apply feature selection to select 12 of 33 attributes to predict the student performance.

Optimization is the process of achieving the best solution for a problem (SVM parameters) in this article using SA optimization technique help in improving the objective function (classification accuracy) value by avoiding the local minima and present comparative study for academic students' performance, each of this algorithms is compared based on its accuracy to identify the most appropriate model for this job. Comparing our results in section 7 with previous published works clearly show that our proposal SVM-SA

gives better results for accuracy, precision, sensitivity, and f-measures which improve student academic performance prediction for the decision-makers. This paper is organized as follows: In section 2 and 3 both SVM classifier and SA algorithm are explained, In section 4 the proposed SVM-SA model is described, section 5 describes the used data, section 6 evaluation measures, section 7 and 8 the results and conclusions.

2. SUPPORT VECTOR MACHINE

In 1995, SVM was originally developed using the structural risk minimization principle and Vapnik-Chervonenkis theory, it is a supervised machine learning technique that used for both classification and regression problems. SVMs are more commonly used in classification problems because it has high performance and generalization capability.

SVMs are based mainly on the idea of finding the best hyperplane that maximizes the margin (distance to nearest points) between the nearest +ve and -ve data points [13], the class boundary for linearly separable data, giving a greater chance of new data being classified correctly [14], assume the training data has the dataset $\text{data} = \{y_i, x_i\}; i=1, 2, \dots, n$, where $x_i \in \mathbb{R}^n$ represents the i -th candidate vector and target labels $y_i \in \{-1, +1\}$, represents the output label corresponding to the class of item x_i , the original formulation of the SVM algorithm seeks a linear decision surface using the formula $f(x) = w^T x + b$, where w is a dimensional coefficient vector and b is the offset [15]. The linear SVM achieves an optimal hyperplane by solving the following optimization problem:

$$\min_{w,b} \frac{1}{2} \|w\|^2 \quad s.t. \dots y_i(w^T x_i + b) \geq 1, \quad \forall i \in \{1, \dots, m\} \quad (1)$$

This quadratic optimization problem can be solved by finding the saddle point of the Lagrangian function:

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^m \alpha_i [y_i(w \cdot x_i + b) - 1] \quad s.t. \dots \alpha_i \geq 0, i = 1, \dots, m \quad (2)$$

Where α_i is Lagrange variables, after applying KKT conditions for a maximum of (2) are obtained by setting the gradient of Lagrangian with respect to the primal variables w and b to zero and by writing the complementary conditions [16]:

$$\nabla_w L = w - \sum_{i=1}^m \alpha_i y_i x_i = 0 \Rightarrow w = \sum_{i=1}^m \alpha_i y_i x_i \quad (3)$$

$$\nabla_b L = -\sum_{i=1}^m \alpha_i y_i = 0 \Rightarrow \sum_{i=1}^m \alpha_i y_i = 0 \quad (4)$$

$$\forall i, \alpha_i [y_i(w \cdot x_i + b) - 1] = 0 \Rightarrow \alpha_i = 0 \vee y_i(w \cdot x_i + b) - 1 = 0 \quad (5)$$

By (3), the weight vector w solution of the SVM problem is a linear combination of the training set vectors x_1, \dots, x_m . According to complementary conditions (6) w depends on vector x_i that corresponds the $\alpha_i \neq 0$. Which called support vectors that fully define the maximum margin hyperplane, after substitute (3) and (4) into (2) the dual form Lagrangian $L_D(\alpha)$ of (2) is derived as follows:

$$L_D(\alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad s.t. \dots \alpha_i \geq 0, i = 1, \dots, m \text{ and } \sum_{i=1}^m \alpha_i y_i = 0 \quad (6)$$

In (7), (8) and (9) presents the polynomial kernel, sigmoid kernel, and radial basis functions, respectively. These functions are used to find the optimal hyperplane, in this proposal we used sigmoid kernel (MLP) which also called feedforward ANN with three layers of neurons each neuron uses a nonlinear function for activation except the input one and also applies the concept of backpropagation during the network training [17]. The weight, bias and δ are the setting parameters of multilayer perceptron (MLP) [18].

$$\text{Polynomial kernel: } k(x_i, x_j) = (1 + x_i \cdot x_j)^d \quad (7)$$

$$\text{sigmoid kernel: } K(x_i, x_j) = \tanh(k x_i \cdot x_j - \delta), \quad (8)$$

where δ the intercept constant

$$\text{Radial basis function kernel (RBF): } k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (9)$$

where γ is the kernel parameter

There are two problems in the SVM classifier's optimization procedure [13] : 1) How to select relevant features and filter out irrelevant features to construct the SVM classifier; 2) How to properly adjust the penalty parameter C and the hyperplane parameters [19]. SVM parameters such as kernel parameters and the penalty parameter have a great influence on the accuracy and complexity of the classification models. Numerous evolutionary optimization algorithms were proposed for optimizing SVMs; in this paper, SA is proposed as an optimization algorithm which follows search strategy that improves the value of the objective function to find the best parameter settings that can highly enhance the performance of SVM classifier.

3. SIMULATED ANNEALING ALGORITHM

SA algorithm is a local search method invented to avoid local minima [7], [20]. SA's major advantage in comparison with older optimization methods is its ability to escape the local minima. This method based mainly on electing a move randomly in each stage instead of the best move (best neighbor) selection among the available moves, if the new state enhanced (reduced) the cost, it is accepted as the next state while if it caused the cost increment, it is accepted just with a P probability. P named Metropolis probability and is defined as:

$$P(\Delta E) = e^{-\frac{\Delta E}{T}} \quad (10)$$

Where ΔE represents the change in energy (value of the cost function) caused by the change in state T is the temperature or temperature-like variable that controls this probability. A "generative function" exists that denotes the way of updating variables in each attempt and indeed it is the function that specifies the speed of convergence. In typical SA, the generative function is a Gaussian or Boltzmann function:

$$g(X) = (2\pi T)^{-\frac{D}{2}} e^{-\frac{\Delta X^2}{2T}} \quad (11)$$

where D is a dimension of the search space (number of variables in the cost function). ΔX shows the rate of change of X (variables' vector). So, $X = X_0 + \Delta X$ where X_0 the current state and X the next state of variables. The temperature in the kth stage of the algorithm can be found using (12). The steps of SA algorithm [21] showed in Figure 2.

$$T_K = \frac{T_0}{\ln K} \quad (12)$$

1. Choosing an initial solution i from the group of feasible solution S.
2. Choosing the initial temperature $T_0 > 0$
3. Selecting the number of iterations $N(t)$ at each temperature
4. Selecting the final temperature T_F
5. Determining the process of the temperature reduction until it reaches T_F .
6. Setting the temperature exchange counter n to zero for each temperature
7. Creating the j solution at the neighbourhood of the i solution
8. Evaluation of the objective function at any temperature and calculate $\Delta = z(j) - z(i)$
9. Accepting the solution j , if $\Delta < 0$. Else, generation a random number then select solution j
10. Setting $n = n + 1$. If n is equal to $N(t)$ then go to 12. Otherwise, go to 7.
11. Reducing the temperature. If it reaches T_F then stop. Otherwise, go to 6.

Figure 2. SA algorithm

4. RESEARCH METHOD

Penalty and kernel are SVM parameters with a great impact on the accuracy and complexity of the model of classification. This paper proposes a novel evolutionary for the SVM model by using SVM with MLP kernel and employ SA to optimize its parameters which are expressed as $P1$ for the slope or weight and $P2$ for the intercept constant or bias where $P1 > 0$ and $P2 < 0$. By delinquency, the value is set at 1 and -1; thus, the classification error can be decreased. In this section, we describe the proposed SA-SVM model as shown in Figure 3 to find the optimal values of SVM parameters.

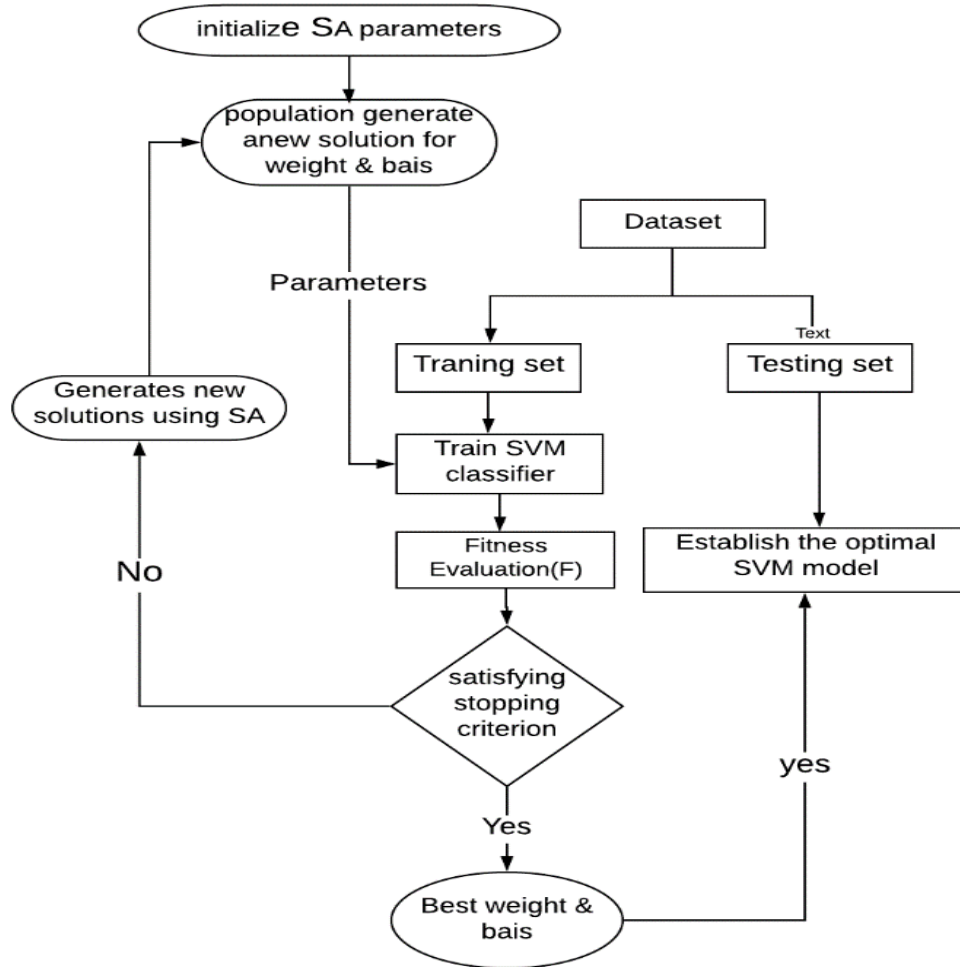


Figure 3. Flowchart of the proposed model (SA-SVM)

The main steps in the SA algorithm are: 1) generating neighbor; 2) evaluating the objective function (classification accuracy); 3) assigning an initial temperature; 4) changing the temperature; 5) cooling schemes, and 6) stopping [21]. The initial solution is one of the important components of SA which generated randomly selected among a feasible solution space in this paper. The initial solution in our algorithm is represented by a two-element vector P as (13). While $P1$ is assigned to the weight, $P2$ is assigned to the bias.

$$P = \langle P1, P2 \rangle, \text{ where } P1 \in \{0,100\}, P2 \in \{-100,0\} \quad (13)$$

$[P1 \ P2]$ is a vector specifies the MLP kernel' parameters of. The MLP kernel takes the form:

$$k = \tanh(P1 * U * V + P2) \quad \text{where } P1 > 0 \text{ and } P2 < 0. \text{ Default is } [1,-1] \quad (14)$$

A feasible solution is randomly selected to be an initial solution, The objective function is an important factor on which SA depends during its performance for evaluating the individual solutions. We formulated the objective function to depend mainly on the classification accuracy of SVM represented by the given solution. How accurately the training data is classified when the classification is conducted using the parameters presented by a solution serves as the cost for a given solution. The cost $Z(P)$ for a solution P is calculated over the training dataset (with a size of N) using (15).

$$Z(P) = \#Truly \text{ classified instances} / N \quad (15)$$

The initial temperature has also high importance as a parameter has a huge effect on the chance of selecting a bad solution. So, if the initial temperature has a high value, a solution with a bad objective function value may has a high chance of being accepted. While considering a low value for the initial

temperature increases the probability of the solution to be a local optimum, in this work the initial temperatures is chosen to be in a range from 0:500.

5. DATA PREPROCESSING

The input and output data were pre-processed by cleaning the missing values, convert the nominal data to numerical data, convert data to binary class (0 means fail, 1 means success), and splitting the dataset into two parts: training and testing datasets (ratio of 70%: 30%) without any feature selection for any dataset, we did not use cross-validation to make the comparison fair because the papers used in the comparison used the same ratio of training and testing data.

6. DATA DESCRIPTION

The first dataset consists of 649 different instances with 33 different attributes, this student's performance dataset is collected from two secondary schools of Portuguese (Gabriel Pereira (GP) and Mousinho da Silveira (MS)). The dataset contains attributes for students like academic grades, social attributes, demographic attributes, and school-related attributes. Data was collected from the students using the school reports and questionnaires. Dataset' details are shown in Table 1 [9] [22]. The second dataset is from three different colleges, Duliajan, Doomdooma, and Digboi College of Assam, India. Initially, data of twenty-two attributes were collected [12]. The third dataset is from the Common Entrance conducted by Dibrugarh University, The collected data with 12 attributes were of students who came for counseling cum admission into medical colleges of Assam in the year 2013 [23]. The three datasets are imbalanced this is due to the low repetition rate among students in the database compilation places according to Table 2.

Table 1. Information of datasets

Dataset	Instances	No. of attributes
Portuguese course	649	33
Sapfile	300	22
CEE-data	666	12

Table 2. Information about no of classes

Dataset	pass	Not pass
Portuguese course	452	197
Sapfile	224	76
CEE-data	509	157

7. MODEL EVALUATION

The classification accuracy always seizes the first look when a model is built for a classification problem as the number of instances predicted correctly from all predictions made, but the classification accuracy is not sufficient alone to evaluate a model, especially in case of imbalanced data classification. Therefore, we considered some other measurements such as sensitivity, precision, and F-measures [24]. The measurement equations used in model evaluation listed in Table 3 were: TP for True positives, TN for True negatives, FP for false positives, and FN for False negatives [25].

Table 3. Measurement equations

Measure	Formula
Accuracy, recognition rate	$\frac{TP + TN}{P + N}$
sensitivity, true positive rate, recall	$\frac{TP}{P}$
Precision	$\frac{TP}{TP + FP}$
F-measures	$2 \times \frac{Precision \times recall}{Precision + recall}$

8. RESULTS

The platform adopted to develop the SA-SVM algorithm is a laptop with the following features: Intel(R) Core (TM) i7-4600 CPU@2.10GHz, 8G RAM, a Windows 10 pro as operating system using

MATLAB software version R2015a. To evaluate the proposed algorithm, three standard classification datasets were used. The datasets were obtained from the University of California at Irvin (UCI) Machine Learning Repository, the properties of all datasets are listed in Table 1, each dataset split into two parts in which training and testing datasets with ratio of 70% and 30% respectively and the results are shown in Table 4, Table 5 and Table 6.

Table 4. CEE dataset results

CEE dataset	Accuracy	Sensitivity	Precision	F-measures
SVM (MLP kernel)	61.8%	52.8%	58%	55.29%
SVM (MLP kernel) &SA	69.34%	75.28%	63.2%	68.71%

Table 5. Portuguese course dataset results

Portuguese course dataset	Accuracy	Sensitivity	Precision	F-measures
SVM (MLP kernel)	78.35%	76.21%	97.65%	85.6%
SVM (MLP kernel) &SA	90.72%	97.56%	91.95%	94.67%

Table 6. Sapfile dataset results

Sapfile dataset	Accuracy	Sensitivity	Precision	F-measures
SVM (MLP kernel)	61.1%	55.22%	88%	67.88%
SVM (MLP kernel) &SA	67.77%	88%	73.75%	80.27%

From the Tables 4 to 8, it is clear that our proposed method shows comparative performance without feature selection to all the other classification algorithms in term of prediction accuracy. Portuguese course dataset showed 78.35% accuracy for SVM with MLP kernel, this accuracy increased to 90.72% after applying the proposed method as well as sensitivity and f-measures as shown in Table 5 and it is better than other presented classifiers' accuracy as shown in Table 7. CEE dataset showed 61.8% accuracy for SVM with MLP kernel, this accuracy increased to 69.34% after applying the proposed method as well as sensitivity and f-measures as shown in Table 4 and it is better than NaiveBayes, and decision tree (J48), ZeroR, REPTree, OneR, RandomTree, JRip, and SimpleLogistic accuracy as shown in Table 8.

Sapfile dataset showed 61.1% accuracy for SVM with MLP kernel, this accuracy increased to 67.77% after applying the proposed method as well as sensitivity and f-measures as shown in Table 6 and it is better than BayesNet accuracy as shown in Table 9. Also, the highly improvement in the f-measures for the three used datasets that reached 13% strongly proves the efficiency of our proposed SVM-SA model in dealing with the problem imbalance in data. All the forgoing confirms the effectiveness of the proposed method.

Table 7. Portuguese course dataset comparison

Method	Accuracy	Ref.	Wilcoxon rank
SVM (MLP kernel) &SA	90.72 %	Proposed	1
NaiveBayes	68.25 %	[9]	6
Decision Tree	67.79 %	[9]	7
RandomTree	53.46 %	[9]	8
REPTree	75.19 %	[9]	3
JRip	70.72 %	[9]	5
OneR	76.73 %	[9]	2
SimpleLogistic	71.34 %	[9]	4
ZeroR	30.97 %	[9]	9

Table 8. CEE dataset comparison

Method	Accuracy	Ref.	Wilcoxon rank
SVM (MLP kernel) &SA	69.34%	Proposed	1
Decision Tree (J48)	64.71%	[23]	2
Naive Bayes	57.81%	[23]	4
ZeroR	31.53%	Weka	9
REPTree	55.55%	Weka	5
SimpleLogistic	60.36%	Weka	3
OneR	51.05%	Weka	7
JRip	54.80%	Weka	6
RandomTree	33.5%	Weka	8

Table 9. Sapfile dataset comparison

Method	Accuracy	Ref.	Wilcoxon rank
SVM (MLP kernel) &SA	67.77%	proposed	1
BayesNet	65.33%	[12]	2
Naïve Bayes	51.11%	Weka	5
ZeroR	41.66%	Weka	8
REPTree	55%	Weka	4
SimpleLogistic	63.33%	Weka	3
OneR	50.66%	Weka	6
JRip	47.77%	Weka	7
RandomTree	37.61%	Weka	9

9. CONCLUSION

Machine learning techniques with educational data can be used to improve the learning process of students in higher education institutes. Different methods were developed by researchers to predict students' performance in the enrolled courses, to provide valuable information that helps in facilitating the students' retention in those courses. This information can be used by instructors to early identify students who might need assistance in their study. In our work SVM applied on three different real datasets then, a hibernation between SVM with MLP kernel and SA was used to enhance the results and finally, compared with the results of other algorithms. The results showed that the proposed method became better after applying the SA optimization technique and presents higher performance than other methods.

REFERENCES

- [1] A. Soni, Vivek Kuma, R. K. and D. Hemavathi, "Predicting student performance using data mining techniques," *International Journal of Pure and Applied Mathematics*, vol. 119, no. 12, pp. 221-227, 2018.
- [2] A. Kaur, Nitin Umesh and Barjinder Singh, "Machine Learning Approach to Predict Student Academic Performance," *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, vol. 6, no. 4, pp. 734-742, April 2018.
- [3] M. A. Ahmad, Shun-ichi Azuma and Toshiharu Sugie, "Identification of continuous-time Hammerstein systems by simultaneous perturbation stochastic approximation," *Expert Systems With Applications*, vol. 43, pp. 51-58, January 2016, doi: 10.1016/j.eswa.2015.08.041.
- [4] N. S. A. Shukor, M. A. Ahmad and M. Z. M. Tumari, "Data-driven PID tuning based on safe experimentation dynamics for control of liquid slosh," *2017 IEEE 8th Control and System Graduate Research Colloquium (ICSGRC)*, Shah Alam, Malaysia, 2017, pp. 62-66, doi: 10.1109/ICSGRC.2017.8070569.
- [5] S. S. Du, Jason D. Lee, Haochuan L, Liwei Wang and Xiyu Zhai, "Gradient Descent Finds Global Minima of Deep Neural Networks," in *International Conference on Machine learning*, Long Beach, California, vol. 97, pp. 1675-1685, 2019.
- [6] A. S. Ramli, Muhammad Ikram Mohd Rashid and Mohd Ashraf Ahmad, "Energy Management Strategy of HEV based on Simulated Annealing," *International Journal of Integrated Engineering*, vol. 12, no. 2, pp. 30-37, February 2020.
- [7] M. Vakil-Baghmisheh and A. Navarabaf, "A modified very fast Simulated Annealing algorithm," *2008 International Symposium on Telecommunications*, Tehran, Iran, 2008, pp. 61-66, doi: 10.1109/ISTEL.2008.4651272.
- [8] V. K. Pal and Vimal Kamlesh Kumar Bhatt, "Performance Prediction for Post Graduate Students using Artificial Neural Network," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 8, no. 7S2, May 2019.
- [9] Y. K. Salal, S. M. Abdullaev and Mukesh Kumar, "Educational Data Mining: Student Performance Prediction in Academic," *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 8, no. 4C, pp. 54-59, April 2019.
- [10] S. M.-. H. Hasheminejad and M. Sarvmili, "S3PSO: Students' Performance Prediction Based on Particle Swarm Optimization," *Journal of AI and Data Mining*, vol. 7, no. 1, pp. 77-96, 2019, doi: 10.22044/JADM.2018.5506.1662.
- [11] D. Kabakchieva, " Student Performance Prediction by Using Data Mining Classification Algorithms," *International Journal of Computer Science and Management Research*, vol. 1, no. 4, 2012.
- [12] S. Hussain, Neama Abdulaziz Dahan, Fadl Mutaheer Ba-Alwi and Najoua Ribata, "Educational Data Mining and Analysis of Students' Academic Performance Using WEKA," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 9, no. 2, pp. 447-459, February 2018.
- [13] F. Ye, "Evolving the SVM model based on a hybrid method using swarm optimization techniques in combination with a genetic algorithm for medical diagnosis," *Springer Science+Business Media New York*, vol. 77, no. 3, pp. 3889-3928, 2016.
- [14] F. Albu, Adelaida MATEESCU and Neculai DUMITRIU, "Architecture Selection for a Multilayer Feedforward Network," in *International Conference on Microelectronics and Computer Science*, pp. 131-134, October 1997.
- [15] F. Albu and D. Martinez, "The application of support vector machines with Gaussian kernels for overcoming co-

- channel interference," *Neural Networks for Signal Processing IX: Proceedings of the 1999 IEEE Signal Processing Society Workshop (Cat. No.98TH8468)*, Madison, WI, USA, 1999, pp. 49-57, doi: 10.1109/NNSP.1999.788122.
- [16] F. Albu, A. Mateescu, J. C. M. Mota and B. Dorizzi, "Adaptive channel equalization using neural network," *ITS'98 Proceedings. SBT/IEEE International Telecommunications Symposium (Cat. No.98EX202)*, Sao Paulo, Brazil, 1998, pp. 438-441 vol.2, doi: 10.1109/ITS.1998.718433.
- [17] K. A. A. Kamarulzaini, N. Ismail, M. H. F. Rahiman, M. N. Taib, N. A. M. Ali and S. N. Tajuddin, "Evaluation of RBF and MLP in SVM kernel tuned parameters for agarwood oil quality classification," *2018 IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA)*, Penang, Malaysia, 2018, pp. 250-254, doi: 10.1109/CSPA.2018.8368721.
- [18] D. Kaya, "Optimization of SVM Parameters with Hybrid CS-PSO Algorithms for Parkinson's Disease in LabVIEW Environment," *Hindawi*, 2019, doi: <https://doi.org/10.1155/2019/2513053>
- [19] M. N. F. b. Ariffin, Nurlaila Ismail, Nor Azah Mohd Ali, Mohd Hezri Fazalul Rahiman, Saiful Nizam Tajuddin and Mohd Nasir Taib, "Multilayer Perceptron (MLP) tuned Kernel Parameter in Support Vector Machine (SVM) for Agarwood Oil Compounds Quality Classification," *Journal of Tomography System & Sensors Application*, vol. 1, no. 1, June 2018.
- [20] . Kirkpatrick, C. D. Gelatt and M. P. Vecchi, "Optimization by simulated annealing," *science*, vol. 220, no. 4598, pp. 671-680, 1983.
- [21] A. A. Taleizadeh, Seyed Taghi Akhavan Niaki and Gholamreza Jalali Naini, "Optimizing multiproduct multiconstraint inventory control systems with stochastic period length and emergency order," *Journal of Uncertain Systems*, vol. 7, no. 1, pp. 58-71, 2013.
- [22] Mahsa Afsharizadeh, Hossein Ebrahimpour-Komleh, "Using Data Mining Techniques for Predicting Alcohol Consumption in Portuguese secondary schools," *1 st International Conference on New Research Achievements in Electrical and Computer Engineering*, 2016.
- [23] S. Hussain, Rasha Atallah, Amirrudin Kamsin and Jiten Hazarika, "Classification, Clustering and Association Rule Mining in Educational Datasets Using Data Mining Tools: A Case Study," *Computer Science On-line Conference*. Springer, pp. 196-211, 2019.
- [24] . K. Karthikeyan and P. Kavipriya, "On Improving Student Performance Prediction in Education Systems using Enhanced Data Mining Techniques," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 7, no. 5, pp. 935-941, May 2017, doi: 10.23956/ijarcsse/SV7I5/0348.
- [25] B. Sekeroglu, Kamil Dimililer and Kubra Tuncal, "Student Performance Prediction and Classification Using Machine Learning Algorithms," *8th International Conference on Educational and Information Technology*, pp. 7-11, March 2019, doi: <https://doi.org/10.1145/3318396.3318419>.