International Journal on Future Revolution in Computer Science & Communication Engineering
Volume: 3 Issue: 9
ISSN: 2454-4248
115 – 120

Self Organizing Map (SOM) based Modelling Technique for Student Academic Performance Prediction

Kapil Saxena
Research Scholar, Department of Appl. Math's and C.S.
S.A.T.I.
Vidisha, India
e-mail: special.kapil@gmail.com

Shailesh Jaloree
Department of Appl. Math's and C.S.
S.A.T.I.
Vidisha, India
e-mail: shailesh_jaloree@rediffmail.com

R.S. Thakur
Department of Computer Applications.
M.A.N.I.T.
Bhopal, India
e-mail: ramthakur2000@yahoo.com

Sachin Kamley
Department of Computer Applications.
S.A.T.I.
Vidisha, India
e-mail: skamley@gmail.com

***Abstract*—** Over the years, student academic performance mapping is considered an important issue for academic institutions and designing such system is very complicated. However, the student performances rely on various factors such as attendance, marks, family background, curriculum activities, social behavior etc. and mapping of all these attributes is very complicated. In the past, various data mining software and techniques have been proposed to classify student data set. These software's and techniques have been failed to classify student dataset correctly. Now advances of Artificial Intelligence (AI) and data mining techniques made it possible to classify student data set and draw useful patterns efficiently. In this study, real data set of Government Girls College (GGC) vidisha of 250 students is considered. The main concern of this study is to apply SOM clustering approach to classify student dataset. Finally, experimental results demonstrated that 4 clusters have been formed based on category like very good, good, average, and poor.

Keywords*-SOM;AI; DATA MINING; PREDICTION; CLASSIFICATION; GGC, CLUSTERING, MATLAB R2011A;*

*****

I. INTRODUCTION

For better quality education, student academic performance monitoring is an important task for each and every institute. However, the educational institutes are incorporating performance monitoring to their educational process in order to achieve high quality standards in the education as well as classifying good and poor students based on performance [1]. Thus, this classification will be helpful during admission time in order to get suitable candidates for academic programs. But, selection of wrong candidates or failing to make accurate predictions means unsuitable candidates being entered in the institutes. Overall, this might be degrading the education quality of the institute [2] [3].

In this direction, various soft computing techniques like Genetic Algorithm (GA), Artificial Neuro Fuzzy Inference System (ANFIS), Artificial Neural Network (ANN) and Self Organizing Map (SOM) etc. have been used in the previous study in order to make accurate predictions as well as maintaining high quality standards of education.

SOM is an unsupervised neural network clustering algorithm which is introduced by Kohonen in 1982 [4]. However, it named as SOM because there is no supervision is required. Apart from it, network has competitive learning feature i.e. it can learn by own through and mapping their weights to conform input data [5] [6]. Fig. 1 shows basic structure of SOM network.

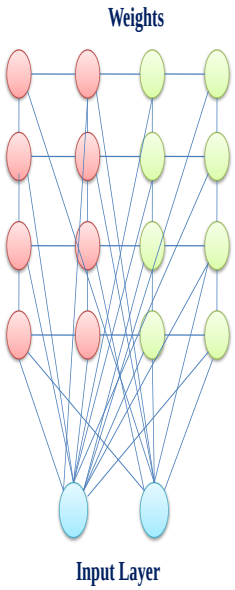

Figure 1. Basic Structure of SOM Network [6].

 ISSN: 2454-4248
115 – 120

II. LITERATURE REVIEW

In this section, we are presenting a brief literature review of some significant researchers.

Alias et al. (2015) [7] have applied SOM clustering algorithm for identifying significant patterns of student dataset using e-learning system in order to improve teaching methodology. Their, experimental results stated that proposed algorithm identify several clusters or extracted more effective results than previous methods.

Khadir et al. (2015) [8] have designed student performance monitoring system using SOM based clustering approach. However, their system could have been predicted the semester wise performance of students.

Hijazi and Naqvi (2006) [9] have monitored the performance of students by selecting data set of approximately 300 students. They have identified some attributes such as attendance, mother education, weekly study hours, family income which has significant impact on student performance.

Halees (2009) [10] has used EM-clustering learning algorithm to identify student behavior or characteristics. In their study, they have considered personal records and academic records of students. For performance monitoring they obtained the grades of students. Finally, experimented results stated that students groups are formed based on performance i.e. excellent, very good, good, poor and fail.

Yusob et al. (2004) [11] have used SOM clustering algorithm to identify learner's status i.e. beginning, intermediate and advanced. They have considered various parameters for supervision of network i.e. learning time, no. of backtracking steps etc.

Olama et al. (2014) [12] have used Multi Layer Perceptron (MLP) model to classify student performance (success or failure rate). However, the performance classification they have considered attributes such as quizzes, discussions and forum and homework.

Delgado et al. (2006) [13] have used neural network model to predict the student final grades. They have used Radial Basis Function (RBF) for feed forward neural network to predict student pass or fail from moodle logs. However, the experimental results have demonstrated the outstanding prediction accuracy.

Olokar and Deshmukh (2016) [14] have applied SOM clustering technique to predict student performance in the context of lower education to higher education in the university. Finally, proposed method could have been succeeded to extract knowledge from student dataset and overall improves the performance of students.

Teir and Halees (2012) [15] have applied data mining techniques for extraction of knowledge from educational domain to improve the performance of graduate students i.e. low grades. In their study, they have considered the data of college of science and technology. Finally, experimental results stated that they could have been succeeded to overcome problem of low grades of students.

Sathya and Abraham (2013) [16] have presented a comparative study between supervised and unsupervised algorithms in the higher education scenario. Finally, their experimental results had stated that supervised algorithm i.e. error back propagation algorithm is very efficient for non-linear real problems while unsupervised algorithm KSOM provides more accurate results than other algorithms.

Saxena et al. (2017) [17] have applied NN based modeling approach to predict student behavior. They have considered the dataset of GGC Vidisha of approximately 250 undergraduate students. Finally they have classified the student performance in terms of very good, good, average, poor.

This study enhanced by SOM based modeling approach to classify student performance.

III. DATA PREPROCESSING

This study employs Government Girls College (GGC), Vidisha (M.P.) data for study purposes. The data set consists of 12 essential attributes and 250 tuples [18]. Before data given to network, some data preprocessing steps like data cleaning, filling missing values are applied on the data set. Finally data set are prepared on excel work book format. Table 1 describes student data set in brief.

TABLE I. DESCRIPTION OF STUDENT DATA SET [18]

S.No.	Attributes	Description
1	SSC	Matriculation Marks
2	HSC	Higher Secondary Marks
3	S1-S4TH	Sem1 to Sem 4 Theory Marks
4	S1-S4PR	Sem1 to Sem 4 Practical Marks
5	S1-S4CCE	(Sem 1 to Sem 4 Continuous and Comprehensive Evaluation Marks
6	SGPA	Semester Grade Point Average Marks
7	CGPA	Cumulative Grade Point Average Marks
8	Attendance	Class Attendance Marks
9	Income	Family Income
10	WST	Weekly Studying Time
11	IAH	Internet Access at Home
12	STAS	Study Time After School

IV. PROPOSED METHODOLOGY

A. Self Organizing Map (SOM)

SOM is one of the well known popular clustering techniques that have ability to monitor the student performance by forming cluster groups. SOM has capability to map high dimensional data input into one or two dimensional data output [19]. The SOM network has feed-forward layer structure i.e. single computational layer arranged in rows and columns format where each and every neuron is fully connected to all other input neurons [20]. Fig. 2 shows flowchart of SOM network process.

 ISSN: 2454-4248
115 – 120

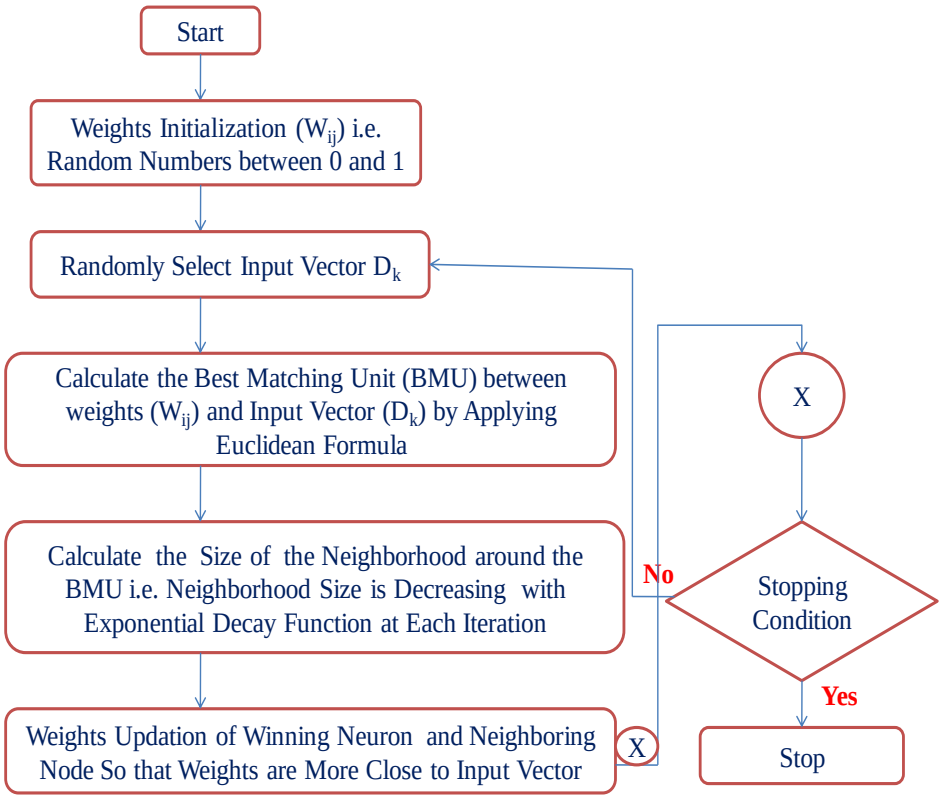

Figure 2 Flowchart of SOM Network Process [20].

V. EXPERIMENTAL RESULTS

In order to conduct experimental results, SOM back propagation learning algorithm (batch weight/bias rules) is used. However, the fully connected $12 \times 10 \times 10$ (2D 100 fully connected) architectures are used where 12 shows input neurons. Now the data set is given as input to train the network model and results of the network are always based on the input vector so therefore, small amount of training data given sequentially to train the network. Fig. 3 training state of SOM network.

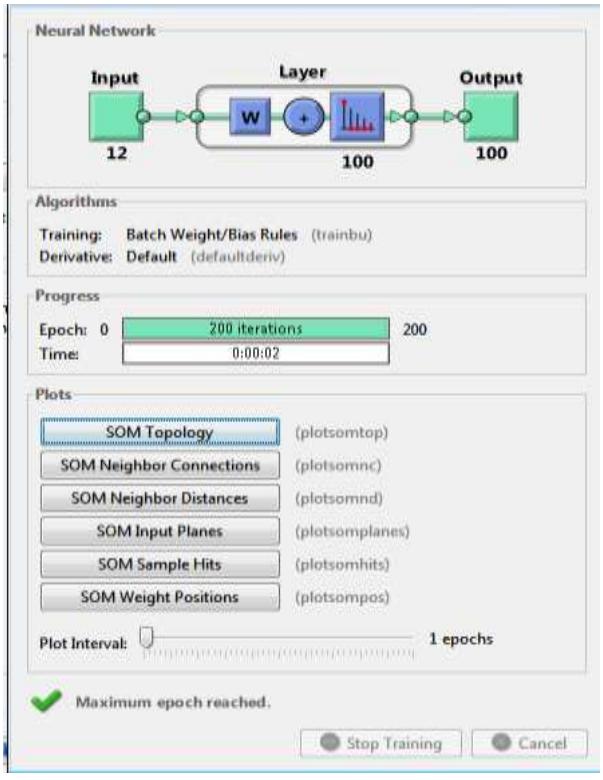

Figure 3 Training State of SOM Network.

Fig. 4 shows SOM topological mapping of dataset.

 ISSN: 2454-4248
115 – 120

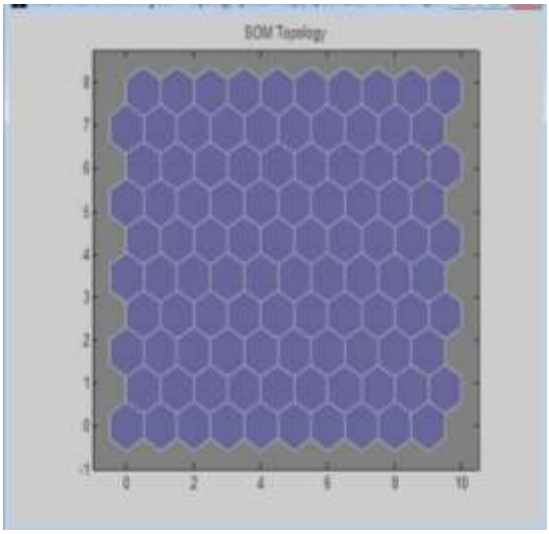

Figure 4 SOM Topological Mapping of Data Set.

Fig. 4 states that neurons are arranged in 2D (10×10) topological format. Fig. 5 shows SOM neighbor connections.

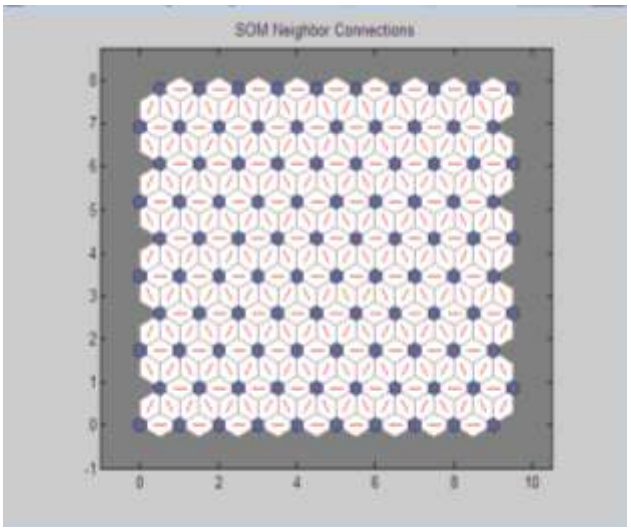

Figure 5 SOM Neighbour Connections.

Fig. 5 clearly states that SOM neighbor connections where neurons denoted as blue dark patches and their connections with their direct neighbor denoted with red line segments. Fig. 6 shows SOM neighbor weight distances.

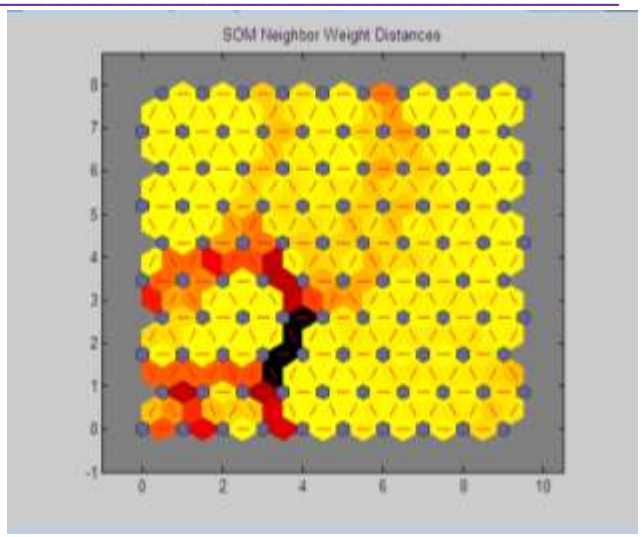

Figure 6 SOM Neighbour Weight Distances.

In Fig. 6 SOM layer neurons are depicted as standard dark center patches and their direct relations with their neighboring neurons which are shown by line segments. Therefore, the neighboring neurons are shown by with different shades of color like red and black which indicates that how close each neurons weight vector with their neighboring neurons. Moreover, dark red color represents nodes are very close with each other whereas dark black color represents nodes are further apart. Fig. 7 shows SOM input weight planes.

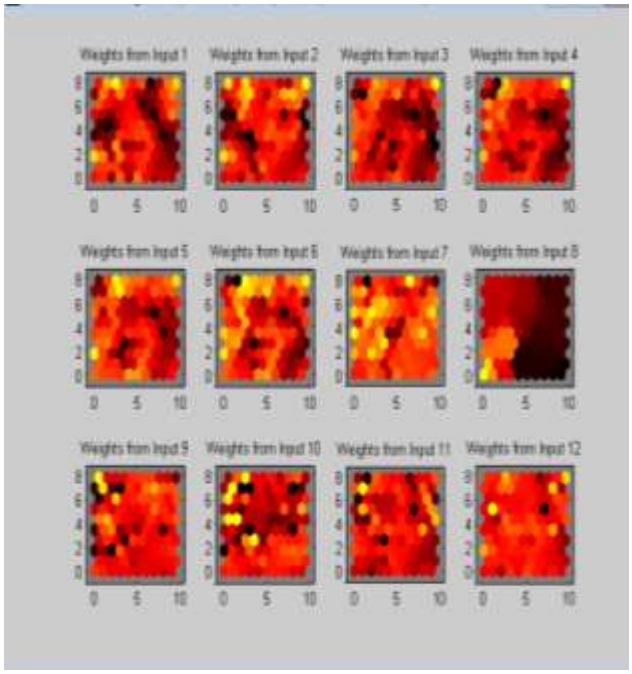

Figure 7 SOM Input Weight Planes.

Fig. 7 generates a various subplots where each subplot shows weights from the i^{th} input to the layer neurons. However, the various connections are shown with different shades of colors like red, yellow and black etc. where black color shows the no connection between neurons. Fig. 8 shows SOM sample hits of input vector.

 ISSN: 2454-4248
115 – 120

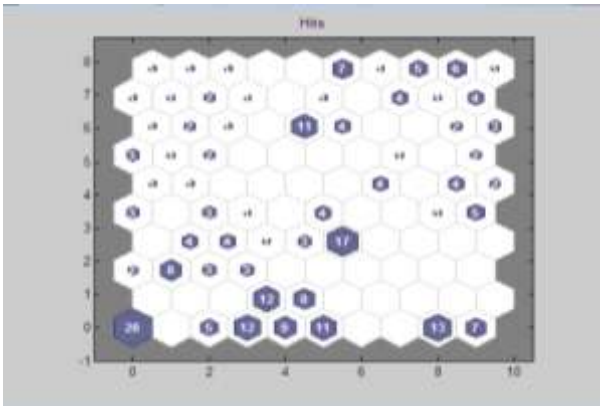

Figure 8 SOM Sample Hits of Input Vector.

Fig. 8 shows a classification of SOM layer and each neuron which is clearly indicated by cell numbers. In the next section, sum of hits for each cluster is observed based on the data set.

VI. ANALYSIS OF RESULTS

In this study, Matlab R2011a software is used to analyze the results. There are 250 tuples which contains 12 attributes. Based on the attributes, SOM clustering is used to monitor the student performance. Table 2 shows sample of student performance by forming four cluster groups i.e. very good, good, average and poor. Table 2 shows mean of each cluster for each attribute.

TABLE II. ATTRIBUTE WISE CLUSTER MEAN

Attribute	Cluster 1	Cluster 2	Cluster3	Cluster 4
SSC	390	423.4	425.5	328.1
HSC	324.2	313.4	329.4	341
S1-S4TH	268.5	256.3	344.5	357.7
S1-S4PR	281	333.8	369.3	225.64
S1-S4CCE	275.3	331.4	376.2	302.2
SGPA	61.18	60.03	63.09	61.35
CGPA	60.17	60.06	62.50	59.52
Attendance	61.8	60.5	63.24	56.6
Income	67500	80833.4	113666.6	94117.7
WST	26.83	24.72	23.93	22.54
IAH	3.1	2.58	2.76	2.45
STAS	2.78	2.91	1.84	2.5

Based on the above results, the student performance is divided among four groups for each attributes. Fig. 9 shows performance graph for student data set.

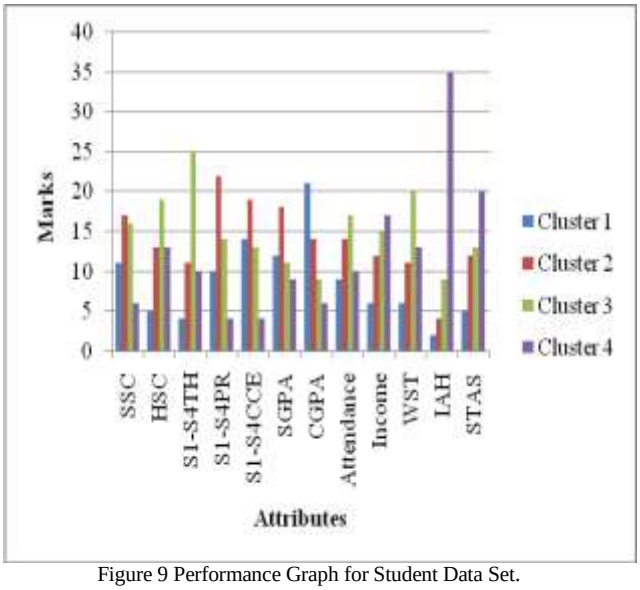

Figure 9 Performance Graph for Student Data Set.

Fig. 9 represents all attributes but describing all attributes is too much complicated. Here, we describing some attribute. Out of 50 samples, for SSC attribute, cluster 1 consists 11 students, cluster 2 consists 17 students, cluster 3 consists 16 students and cluster 4 consists 6 students which show very good, good, average and poor performance respectively. Similarly, for HSC attribute, cluster 1 consists 5 students, cluster 2 consists 13 students, cluster 3 consists 19 students and cluster 4 consists 13 students respectively. For S1-S4 TH attribute, cluster 1 consists 4 students, cluster 2 consists 11 students, cluster 3 consists 25 students and cluster 4 consists 10 students respectively. Similarly, for S1-S4CCE attribute, cluster 1 consists 14 students, cluster 2 consists 19 students, cluster 3 consists 13 students and cluster 4 consists 4 students respectively. For CGPA attribute, cluster 1 consists 21 students, cluster 2 consists 14 students, cluster 3 consists 9 students and cluster 4 consists 6 students respectively. Meanwhile, Table 3 shows percentage of hits for all clusters respectively.

TABLE III. PERCENTAGE OF HITS FOR ALL CLUSTERS

Attributes	Very Good (Cluster 1)	Good (Cluster 2)	Average (Cluster 3)	Poor (Cluster 4)
SSC	22%	34%	32%	12%
HSC	10%	26%	38%	26%
S1-S4TH	8%	22%	50%	20%
S1-S4PR	20%	44%	28%	8%
S1-S4CCE	28%	38%	26%	8%
SGPA	24%	36%	22%	18%
CGPA	42%	28%	18%	12%
Attendance	18%	28%	34%	20%
Income	12%	24%	30%	34%
WST	12%	22%	40%	26%
IAH	4%	8%	18%	70%
STAS	10%	24%	26%	40%

Table 3 clearly states the statistics for all clusters. Student hit rate for SSC attribute is 34% and 32% for good and average category respectively but for HSC attribute hit rate for average and poor category are large respectively that means in higher secondary exams students performance or success

rate is average or poor as compared to metric exams. Similarly, students have average hit rate for theory marks but in practical and CCE they have good hit percentage. Most of the time students attend college regularly which is shown by attendance hit rate which is the 34% highest for average category and 20% for poor category. Similarly, income has also impact on student performance which is the highest hit rate for poor category. Other attributes such as WST and IAH have hit rate 40% and 70 % respectively which shows students focus on weekly study is average i.e. less students focusing on weekly study. Most of the students prefer to study at the examination time. Similarly, hit rate for Internet Access at Home (IAH) is poor which shows most of the students belong to village background and their family income is also poor so they don't have internet connectivity at home. For SGPA attribute hit rate is 36% for good category but for CGPA attribute which is 42% for very good category that means in current semester students have good performance but finally students have very good performance in final exams or throughout the graduation. At last, we have also identified those students who like to study after school time i.e. The highest hit rate for poor category is 40%. The figure shows that most of the students like to play after school time.

Finally, based on the results we can say that attributes which affect student performance throughout the carrier are SSC/HSC marks, attendance, practical marks, family income, daily habit for studying at home, studying from internet. Therefore, we also observed that some students have poor performance in their school exams but they have scored high marks in graduation exams despite of poor family background, internet connectivity at home.

VII. CONCLUSION AND FUTURE SCOPES

This study presents a SOM clustering approach which is used to monitor students' performance and also will be helpful for enhancing the decision making process in academic organizations or stack holders to predict the successively students' performance semester by semester by incorporating the future academic results in the subsequence academic session. on the other side, we have identified some attributes which plays measure role in student academic career or have direct impact on performance. In future, more data samples will be considered for analytical study and some intelligent techniques or hybrid model will be considered to classify the student academic performance.

REFERENCES

[1] K.N. Shah, S. Kothuru, and S. Vairamuthu, "Clustering students' based on previous academic performance" International Journal of Engineering Research and Applications, vol. 3, no. 3, pp. 935-939, 2013.

[2] S. Borkar, and K. Rajeswari, "Predicting students' academic performance using Education Data mining", International Journal of Computer Science and Mobile Computing, vol. 2, no. 7, pp. 273-279, 2013.

[3] C. Romero, S. Ventura, C. Hervás, and P. Gonzales, "Data mining algorithms to classify students", In proceeding of International Conference on Educational Data Mining, Montreal. Canada, pp. 8-17, 2008.

[4] T. Kohonen, and O. Simula, "Engineering applications of the Self-Organizing Map", Proceeding of the IEEE, Vol. 84, No. 10, pp. 1354-1384, 1996.

[5] J. Han, and M. Kamber, "Data Mining: Concepts and Techniques", 2nd edition, 2006.

[6] S.N. Sivanandam, S. Sumathi, S.N. Deepa, "Introduction to neural networks using Matlab 6.0", Tata McGraw Hill Publishing Company Limited 7th Edition, New Delhi (India), 2006.

[7] U.F. Alias, N.B. Ahmad, and S. Hasan, "Student behavior analysis using Self Organizing Map clustering technique", ARPN Journal of Engineering and Applied Sciences, vol. 10, no. 23, pp. 17987-17995, 2015.

[8] A.S.A. Khadir, K.M. Amanullah, and P.G. Shankar, "Student's academic performance analysis using SOM", International Journal for Scientific Research and Development, vol. 3, issue (2), pp. 1037-1039, 2015.

[9] S. T. Hijazi, and R. S. M. M. Naqvi, "Factors affecting student's performance: A Case of Private Colleges", Bangladesh e-Journal of Sociology, Vol. 3, No. 1, **2006**

[10] A.E. Halees, "Mining student data to analyze learning behavior: A case study", 2009.

[11] B. Yusob, N. Haron, Ahmad, N. B. H., and S. A. Halim, "Individualizing the learning material and navigation Path in an Adaptive Hypermedia Learning System", In Proceeding on 2nd International Conference on Computer Graphics and Multimedia, Esset, Selangor, 2004.

[12] M. M. Olama, G. Thakur, A. W. McNair, and S. R. Sukumar, "Predicting student success using analytics in course learning management systems", In SPIE Sensing Technology + Applications, pp. 91220M-91220M, 2014.

[13] M. Delgado, E. Gibaja, M.C. Pegalajar, and Q. Pérez, "Predicting students' marks from Moodle Logs using neural Network Models", In Proceeding of International Conference on Current Developments InTechnology-Assisted Education. Sevilla, Spain, pp. 586-590, 2006.

[14] B.S. Olokar, and V.M. Deshmukh, "Data mining technique for prediction of academic performance of student using SOM", International Journal of Innovative Research in Computer and Communication Engineering, vol. 4, issue (11), pp. 19286-19291, 2016.

[15] M.M. Abuteir, and A.M. EI- Halees, "Mining educational data to improve students' performance: A case study", 2012.

[16] R.Sathya, and A. Abraham, "Comparison of Supervised and Unsupervised learning algorithms for pattern classification", International Journal of Advanced Research in Artificial Intelligence, vol. 2, no. 2, pp. 34-38, 2013.

[17] K.Saxena, S. Jaloree, R.S. Thakur, and S. Kamley, "Student performance monitoring system using NN based modeling", International Journal of Computer Science and Information Technology Research Excellence, vol. 7, issue (3), pp. 41-48, 2017.

[18] Student data set collected from "Government Girls College (GGC)", Vidisha (M.P.) site, 2017.

[19] Awodele, and O. Jegede, "Neural Networks and its application in Engineering", proceedings of Informing Science & IT Education Conference (InSITE) pp. 83-95, 2009.

[20] S. Ali, and K. A. Smith, "On learning algorithm selection for classification", Applied Soft Computing, Vol. 6, pp. 119–138, 2006.