ISSN: 0976-3104
SPECIAL ISSUE (SCMDSA)
Ramanathan et al.

THE IIOAB JOURNAL

ARTICLE OPEN ACCESS

A NOVEL GENETIC NAND PAFT MODEL FOR ENHANCING THE STUDENT GRADE PERFORMANCE SYSTEM IN HIGHER EDUCATIONAL INSTITUTIONS

Ramanathan L.[1], Angelina Geetha[2], Khalid M.[3], Swarnalatha P.[4]
[1,4] *SCOPE, VIT University, Vellore, INDIA*
[2] *Department of Computer Science & Engineering, B.S. Abdur Rahman University, Chennai, INDIA*
[3] *NITMAS & TNU, Kolkata, INDIA*

ABSTRACT

The higher education system in India is curious about the success of students in education during their study. These educational institutions are adopting several methodologies to improve the quality of education and to improve the success rate of students each year. This is used as a primary objective in improving the academic excellence of a student in his/her higher educational level. The main aim of the study is to create a model that classifies the instances correctly to predict the performance of students using PAFT methodology. The PAFT methodology consists of several attribute of a student modeled into a three-tier model that is collected based on several level testing done on a particular student. The three tier model involves his complete academic details, his/her creative and other interpersonal skills and finally the level of interest towards the present educational approach. To classify the instances correctly, Genetic algorithm is modified in its mutation level with NAND gate. The proposed classifier eliminates the regression fit problem during the selection stage with the help of Tobit regression evaluation of each individual. The classifier is also compared with other techniques like genetic-OR, genetic-AND, Multi-Layer Preceptor and artificial neural network. This classifier optimizes correctly the attributes given as an input for its processing and better learning. The PAFT methodology combined with a genetic - NAND algorithm proves successful in terms of its classification rate. This could be inferred finally that this could be utilized in institutions in determining the performance and success ratio of students as a part of their knowledge management system. Also, this model could also be used for predicting the improvement level of students for the fore coming students based on the collected data set.

Received on: 30th-Nov-2015
Revised on: 22nd-Feb-2016
Accepted on: 26– March-2016
Published on: 10th–May-2016

KEY WORDS

Data mining; Genetic NAND; Knowledge Management; Students Perception; Academic Performance

***Corresponding author: Email:** lramanathan@vit.ac.in, **Tel:** +91 9894411988

INTRODUCTION

In Knowledge management, data mining system, various techniques are employed to extract and determine the meaningful value from a large dataset. In present scenario, data mining models provide a great concern and consideration in the society and information sectors. These data mining techniques receive a greater attention in data analysis and considered as a newly emerging tool for analysis [1]. The main aim of our study is modeling the perception, attention, behaviour and forms of thoughts (PAFT) of the students in the educational sector. This is being considered as a crucial and critical factor, since educational data mining has been coming up with various innovative and improved models [2]. These models are most helpful in modeling the PAFT of students to improve their academic performance in higher educational institutions in India.

Superby, et al., (2006) considered personal details of the student, Implication and perception of student behaviour to classify them into three groups, namely low-risk, probability of success, medium-risk and high-risk with high rate of dropping out. Also, significant, influential factor is identified in these three groups that impart major part in decision making for university entry and academic performance [3]. BK Baradwaj and S Pal, (2011) performed classification tasks for evaluating the division of students with a decision tree. The factors considered are test grade and marks, laboratory marks, attendance, marks [4, 5], seminar and assignment performance and general proficiency [4]. Ramesh, et al., (2013) considered grades obtained at elementary, secondary and higher secondary level, a community of students, group study and education of parents to evaluate their academic performance [6]. Other factors considered for educational data mining to increase the students' performance include: education of mother [7], family income of students [7], dietary habits [7] and education department [8]. Academic factors include CGPA [9, 5], maximum and minimum credits in current semester [10]. These factors are taken along with additional factors more related to increasing the academic performance. Santhi Thilagam and Ananthanarayana,

THE IIOAB JOURNAL
ISSN: 0976-3104

2008 [11] proposed a multi objective function based on fuzzy implication and the dataset is collected through the computer activity done by the students. The algorithm failed to prove its effectiveness in comparing it with other algorithms. Amelia Zafra and Sebastián Ventura, 2012 [12] proposed multi-instance genetic algorithm that is used to predict the performance of student related to academics based on web based education. This research helps in identifying the pass rate and failure rate of each individual based on web based tests. The attribute considered for the student includes: total assignment number, total assignment time, total number of posts, total number of messages read, total time taken to complete the test, total questions, answered, seen and passed by the student. Ahmad Slim et al. (2014) [13] employs a Markov model in determining the performance rate of the student and applies linear regression [14] to avoid the regression fit problems. With the features used, the GPA is predicted from the intermediate test results and this helps in achieving the academic score effectively. Parneet Kaur et al. 2015 [15] compared various techniques like Multi-layer perception, SMO, J48, REP Tree, Naïve Bayes with the collected dataset. This dataset includes real world data from the high school environment that includes the personal portfolio of a student related to his name, attendance, internal grade, medium of instruction, school type, sex, private tuition, computer available at home, whether he/she is qualified or not. This is done to analyse the performance of student related to academics. The main drawback of this system is that the accuracy level of the proposed with various methods did not go more than 75%. The classical techniques and the evolutionary data mining technique called grammar based genetic algorithm (GGA) is used to build a model for predicting the performance of undergraduate students by using their past education and general record [16]. Firstly, a comparison of the accuracy of different classification algorithms used in data mining on multiple datasets and selection of an algorithm which has high accuracy for all classes to ensemble them with the help of a method by name, weak classifier to get a combination of multiple classifier is carried on[17]

Various authors suggested multiple steps [18] for building a data mining model, which involves a transformation process, data preparation with consolidation followed by data cleaning, transformation and reduction phase [19]. Berry and Linoff, (2000) defined an iterative data mining process with multiple steps that could well establish the success rate of students [20]. To analyse the data, data mining algorithm is used to understand the hidden patterns in the node and prediction of data. To do this Weka tool provides a better legitimacy in analysing the collected data [21] and support .java files is used in writing new algorithm. This tool is used as a standalone approach to read the .java files and implementing the algorithm for predicting the desired results. There are several supervised and unsupervised approaches to classify perfectly the instances; this research concentrates on improving the simplicity of an algorithm by analysing its internal parameters. This analysis is done in terms of errors occurring at each stage and rectifying it with suitable efficient technique. This helps in reducing the complexity of certain algorithm that could classify the instances correctly but with longer computation.

The research goal is to increase the prediction ability of the proposed model to improve the students in their academics. This paper also proposes a modified algorithm called Genetic NAND combined with PAFT model to form an efficient novel approach to improve the perception of learning ability in Higher Educational Institutions. The preliminary aim of this study is to collect the relevant data set related to the PAFT methodology by making a student to undergo several levels of testing. Thus, collecting suitably the relevant dataset related to study and this helps in collecting the independent variables with performance as a dependent variable. This possesses certain regression fitting problem in its selection of chromosome and that could be avoided with Tobit regression model. Also the use of Mahalanobis distance classifiers helps in finding the bit strings accurately based on the weighted distance between the selected chromosomes. Soon after this the functionality of NAND gate improves the mutation stage and avoids much complexity due to other logical gates. Here, the main outcome is the reliable success rate of the student with efficient prediction using the Genetic NAND PAFT model. The dataset is being handled with 30 different attributes that is collected from previous literatures and segregated based on PAFT model. This model is designed to work with large datasets, such that the prediction accuracy should no longer be less than previous exiting techniques. Also the simplicity of this approach helps in avoiding the hybridization of the genetic algorithm, thereby reducing the complexity of the genetic approach. This helps in avoiding the use of multi-instance learning, Neuro genetic algorithm, ensemble learning techniques etc.

The main contribution of the paper includes proposed novel genetic NAND algorithm for educational data mining, successive section discusses the results of the proposed method. Finally, it concludes that the proposed technique is efficient in improving the student's academic performance.

THE IIOAB JOURNAL
ISSN: 0976-3104

MATERIALS AND METHODS

Initially, the students are evaluated using questionnaires based on 5-point Likert scale with secondary variables collected from various literatures. The questions are designed using PAFT model with three sections with multiple questions in each section. The students are allowed to self-evaluate themselves based on the questionnaires given before the final exam. Certain sets of questions are asked to instructors regarding their behavior and other factors specified in the next section. Soon after the completion of exams, the instructors are allowed to evaluate their students based on two criteria: whether they have improved or not. Finally, the questionnaires are collected, pre-processed for cleaning the dataset and the proposed novel classification algorithm is applied to it. This new data mining approach helps in increasing the prediction rate of the student's performance using PAFT methodology. This is beneficiary in the higher Educational Institution to predict the performance of the students in future years. To achieve this, the variables collected are grouped into three sets of factors and the proposed genetic NAND algorithm accurately classifies the selected data set to produce the qualified results. Of collected 1000 samples 200 are used to train the artificial neural network. The population focused here are the students from engineering college.

Proposed PAFT system

Student's psychology is one of the major factors being considered for PAFT system. PAFT system is modeled based on the three sets of factors that involves their Perception towards learning, Attention and behavior within the institution and finally their Form of Thoughts. The ultimate aim is to increase the prediction of student's behavior towards education and thereby increasing the perception of learning using PAFT workflow.

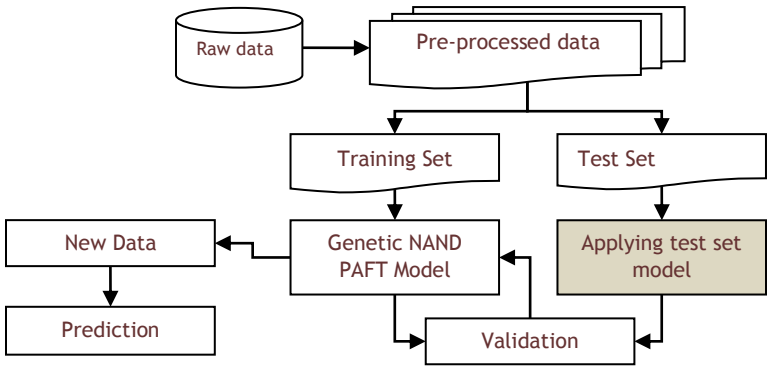

Fig: 1. Work flow of proposed PAFT model

This model has been divided based on final academic performance into three sets of factors that is being closely related to our study:

1. First factor involves stable group structural factors that relate to the personal history of the student's. This factor includes
    a. Identity of the student, academic details,
    b. Previous academic scores in primary, middle, higher secondary level, college level and
    c. Past history details of the student's family.
2. A second factor involves changing or process factors, which includes the involvement of students in relation with his studies or his behavior in studies. These factors are considered participation in other academic activities, technical relationship with his professors, marks or grade performance in periodic examination and in the final examination.
    a. Critical thinking [2]
    b. Communication and teamwork skills [2]
    c. Self-regulated learning [2]
    d. Innovative thinking [6]
    e. Systems-thinking [15]
    f. Non-routine problem-solving [15]
    g. Inquiry skills [15]
    h. Interpersonal skills [16]
    i. Skill demonstration [24]
    j. Technical capabilities include creative thinking [12],
    k. Being social [12],
    l. Patient and determined attitude [12],
    m. Leadership and problem management, crisis [12].
3. A third set of factors involves the perception of students: this set of factors considers
    a. Perception of students related to academic context
    b. His/her perception towards the course and professors

With PAFT methodology, a classification algorithm proposed is applied over these factors to find the rate of improvement in the student's performance in academics. The real-time data are collected from bachelor of computer science students from a self-

THE IIOAB JOURNAL
ISSN: 0976-3104

financing institution in Tamil Nadu. Out of the collected data, 43% were female and the remaining are male. Each student is allowed to undergo certain training based on which the data are collected, the factor one is collected from the department staffs, factor two is collected by conducting certain interpersonal skill trainings based on which they are ranked and finally the factor three is collected through the questionnaire. The average age of participants, who undergone the training was nearly 21.

Novel genetic NAND PAFT model

To overcome the problems in genetic approach like over-fitting, high optimization time, etc. A logical approach has been designed to overcome the above constraints and to increase the accuracy rate and sensitivity level of the prediction model. To achieve this, the data has to go through various processing stages that involve data cleaning, data transformation and data classification. The complete approach of data classification using Genetic NAND Algorithm is mentioned in **Figure– 2**.

Initialization

Genetic NAND algorithm [**Figure– 3**] Involves three stages that include: Fitness stage, NAND stage, Shift stage. Initially, the parameters are initialized and the random population is generated for classifying the associated dataset in the search space. To solve this problem the relationship between the dependent variable and the independent variable needs to be improved. This is done using the following relation:

$$Y = \beta_0 + \beta_0 X_2 + \beta_3 X_3 + \ldots + \beta_N X_N + \zeta$$

Where, this explains the relationship between the independent variable and the dependent variable. X belongs to the set of independent variables and β is the parameter associated with Linear Square Method. The Tobit regression is based on the transformation between the independent variable with the relation X = X* when X*< C and X = X* when X* > C. From the precise value obtained from the independent variable all the non-linear points could be removed. This could be related as

$$Y^* = \beta' X + \zeta$$

Where W is the binary variable, X is the independent variable and ζ is the error factor.

Selection of chromosome

During the selection of chromosome, the independent variable makes a challenging task in selecting the chromosome [5n]. To avoid such regression fit problem in selecting proper chromosome, it is better to select a proper regression fit method.

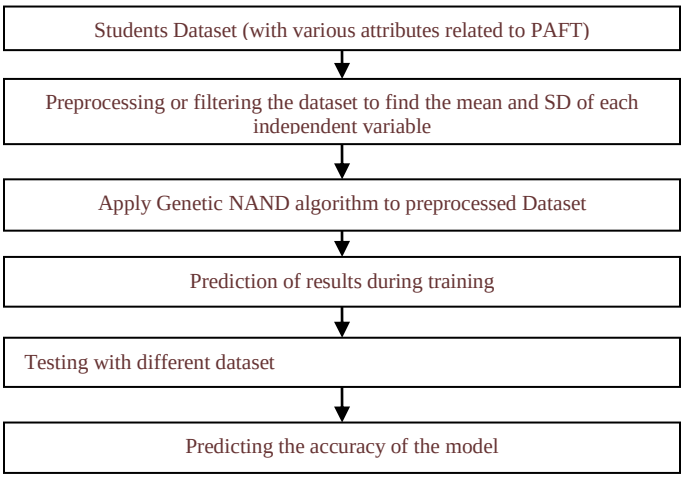

Fig: 2. Genetic NAND Approach

Fitness stage

During fitness stage, the objective function is selected or defined to calculate the best two individuals in the search space. This fitness function for a supervised approach is defined using Mahalanobis distance. In order to normalize the attributes in our system, Mahalanobis distance [17] function is taken into account which is defined as:

$$dij = [(x_i - x_j)^T M^T M (x_i - x_j)]^{1/2}$$

THE IIOAB JOURNAL
ISSN: 0976-3104

Matrix ***M*** can be evolved by using the Mahalanobis distance function between the vectors x_i and x_j. The main aim of reducing the irrelevant dataset and for better functionality of GS NAND, matrix ***M*** is chosen to be diagonal or symmetrical. Thus the chromosome is represented as a bit strings formed using the binary representation of each matrix element that belongs to diagonal or upper triangular matrix.

$$M = \begin{pmatrix} m_{11} & m_{11} & \dots & m_{1d} \\ m_{21} & m_{21} & \dots & m_{1d} \\ \dots & \dots & \dots & \dots \\ m_{d1} & m_{d1} & \dots & m_{dd} \end{pmatrix}$$

The bit string chromosomes, $B(m_{ij})$ are represented using binary representation of m_{ij}.

This matrix ***M*** represents the distance function used for activation of neurons. This increases the accuracy and reduces the network error, thus making the distance function to be good. The fitness function of this matrix ***M*** is presented using mean squared error ***E***:

$$fitness = -\log_2 E$$

This function is chosen specifically for training the datasets, such that fitness increases when the rate of error value reduces. The main work of the fitness function is to increase the distance between the chromosomes when they are nearer to zero. This increases the evolutionary pressure at the latter stages of GSOR evolution.

The attributes used here are Age, Gender, Type of study (Full time/Part time), Field, Department, Year of study, Marks scored in Cycle Test , Marks scored in Final Exam, Interest in coming to college/school, A Good Knowledge of Subjects, Ability to clear exams, Interest in Homework, Interest in listening, Ability to learn, read and Write, Quality of work Problem, Solving Team work, Dependability on his friends, Ability to learn and adapt from mistakes, Student shows understanding and sensitivity to needs and differences of others (i.e. ethnicity, religion, language, etc.), demonstrates effective written communication, demonstrates effective Oral communication, The extent to which the student effectively listens, conveys, and receives ideas, information, and direction. Proficiency in English, Ability to Draw (Perspective, Freehand Pencil Drawing), Dimensional Imagination –Skill to Draw Dimensional, Technological Skills, Extracurricular activities.

NAND stage

This stage is used for generating new chromosome individuals, which are encoded and named as phenotype. The mating or crossover is carried out when the two best chromosome.

An individual from the fitness function is made NAND with each other to form an offspring represented as:

$$fx = \overline{fx_i . fx_j}$$

Where, fx_i and fx_j are the best chromosome individuals acquired using fitness function. This creates a two new offspring from the parent chromosomes after crossover using NAND is done. These two new offspring's are put into a new generation of the population and using this logical recombining, the process is likely to attain good individuals.

Shift Stage

The newly produced offspring's are subjected to bit flipping or shift to attain a better diversity within the population and to inhibit convergence within the search space. To achieve this linear NAND register is used to perform the mutation process, where the present state output depends on its previous stage. Thus, the generated mutation individual is added to the empty population P_E. Finally, if the P_E is equal to ***P***, process is completed and finally the iteration ends with a better classification due to good chromosome individuals.

THE IIOAB JOURNAL
ISSN: 0976-3104

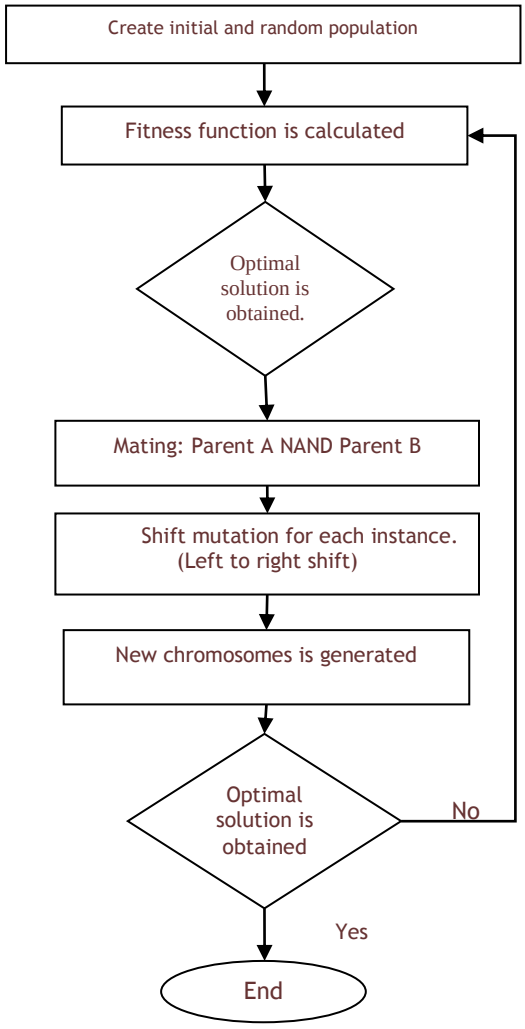

Fig:4. Genetic NAND Algorithm

RESULTS AND DISCUSSIONS

In this proposed approach, the students are allowed to undergo training based on their creative level and tested over several rounds. The students are also evaluated in terms of multiple choice questions that is being considered as an effective assessment technique for the third factor. This was conducted before the final exam date, where final year students were provided with this self-evaluation assessment. Certain questionnaires are asked to teachers regarding their improvement based on the defined set of factors. Finally, after the academic final exams the instructor gave the choice of improved or not improved decision to evaluate the student performance. Depending upon the decision of the instructor, out of collecting 1000 data samples 740 students were in improved state and rest in not improved state. The real time dataset is used as an input for the proposed technique to evaluate the performance of the students and their improvement towards academics. The dataset is filtered using an unsupervised discrete filter to remove all the duplicate sets, irrelevant and redundant dataset. All the 30 attributes were selected and given as an input to the model, since in dynamic environments like institutions the data is limited. This three set factor variable creates a full range view of substantial information regarding the prediction of student's performance in academics.

Initially, net beans are used for coding the GS-NAND algorithm and this file are converted to a weka classifier format (.arff). Then the dataset in .csv file is given as an input to weka classifier and redundant and duplicate data are removed using unsupervised filter. The output of which is given an input to testing GS- NAND classifier,

THE IIOAB JOURNAL
ISSN: 0976-3104

which is followed by a set of patterns with predicted results of the given dataset **[Figure– 1]**. Then this model is applied to validate for finding out its accuracy for various other datasets and the results were presented accordingly. Finally the results are classified into two instances class 1 and class 0: class 1 refers to qualified level and class 0 refers to not qualified level.

PAFT variables are used efficiently during training the GS-NAND algorithm, since the pre-processing removes all the irrelevant individuals from the dataset. This increases the efficiency of the training, since the accuracy increased when compared with normal genetic and other RDFF algorithms. To attain this, the following parameters of GS-NAND were taken into consideration: Population Size: 100, target fitness = 0.9, maximum number of generations = 20. This GS-NAND classifier provides the following results for the specified Elite count shown in **[Table–1, Figure–1]**.

Table: 1. Training and testing results of GSNAND

Class	Elite program	Size	Training Fitness	Validation Fitness	Error	Program weights
No	0	7	0.906	0.934	0.08	1.516
No	1	9	0.854	0.828	0.21	0.363
No	2	5	0.800	0.817	0.14	1.152
No	3	9	0.860	0.850	0.28	0.080
No	4	18	0.916	0.906	0.25	0.619
Yes	0	4	0.921	0.922	0.07	1.585
Yes	1	10	0.869	0.806	0.15	0.619
Yes	2	15	0.800	0.815	0.21	0.363
Yes	3	14	0.885	0.882	0.13	0.708
Yes	4	16	0.859	0.877	0.19	0.663

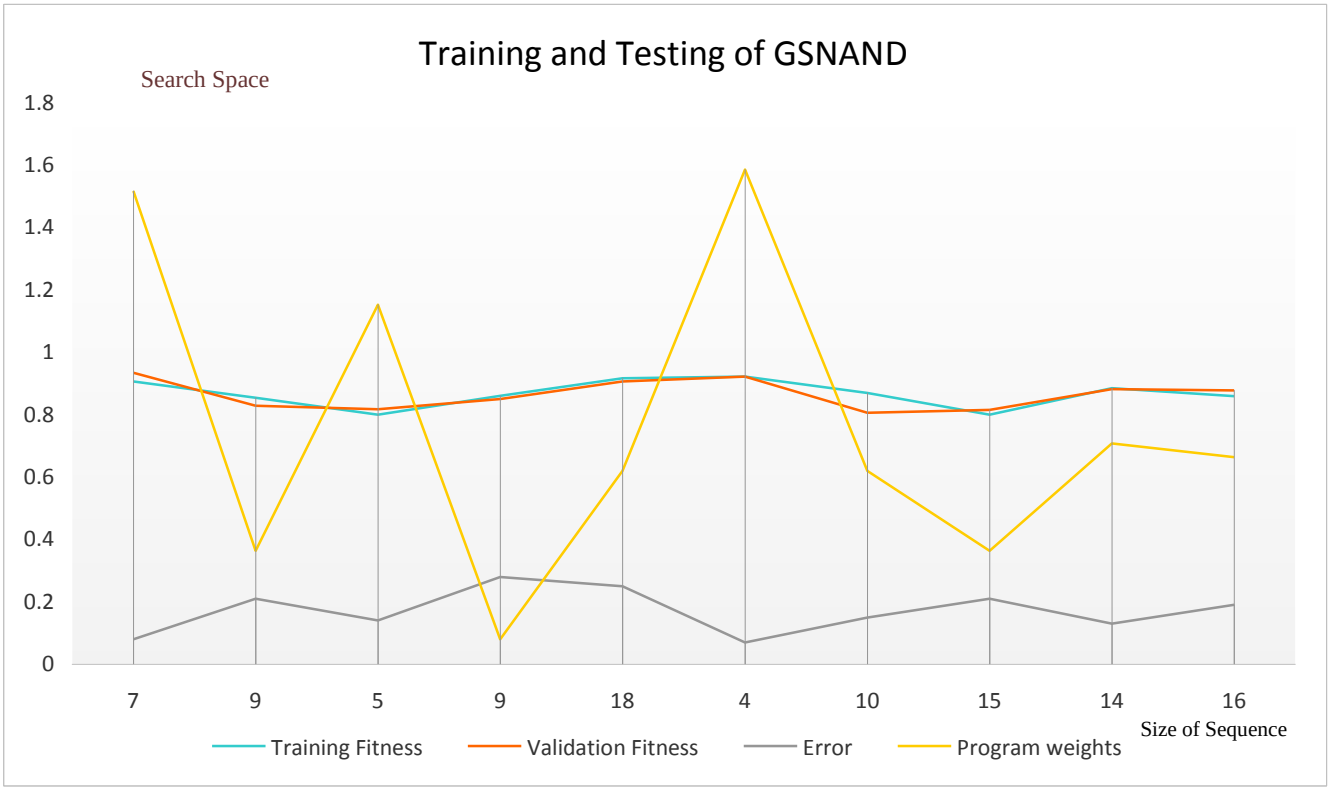

Fig: 1. Training and Testing Results of GSNAND

THE IIOAB JOURNAL
ISSN: 0976-3104

From the above **Table**, it is found that the training and testing fitness were found to be good i.e. it is greater than 9.0 for all the five elite programs of two different classes. Also, the error is reduced significantly using this proposed algorithm in the search space. This shows that the algorithm is performing well, since the rate of fitness in both testing and training phases seems promising. Thus the efficiency of the proposed model is defined in terms of comparing it with existing techniques shown in **[Table–2, Figure–2]**.

Table: 2. Evaluation of Test Performance of Classifiers for 1000 samples

Evaluation	Class	GS - OR	GS-AND	GA	MLP	Neural Network	Proposed GS-NAND
Correctly Classified Instances	-	832	844	820	790	890	980
Incorrectly Classified Instances	-	168	156	180	210	110	20
Mean absolute error	-	0.0324	0.0216	0.18	0.0434	0.1214	0.0216
Root mean squared error	-	0.2341	0.1006	0.4243	0.1152	0.1201	0.1006
True Positive Rate	1	0.854	0.859	0.759	0.042	0.94	1
	0	0.847	0.842	0.957	0.902	0.921	0.957
Precision	1	0.894	0.899	0.923	0.5	0.935	0.964
	0	0.892	0.857	0.772	0.521	0.945	1
Recall	1	0.654	0.662	0.759	0.042	0.784	1
	0	0.	0.957	0.931	0.962	0.942	0.957
F-measure	1	0.887	0.898	0.845	0.077	0.901	0.982
	0	0.892	0.899	0.845	0.676	0.903	0.978

From the above table, the ratio of the correctly classified instance to incorrectly classified instance result seems to be higher than previous techniques like Genetic Algorithm and Multilayer Perception. For the verification of proposed GAXOR approach, the classified instance is examined using primary data approach by asking an opinion about a particular student to his/her teacher. Through the collected primary data it is observed that about 840 samples correctly correlate with the classified value, whereas information regarding remaining students is not known to their teachers. From that, it is concluded that proposed GAXOR approach effectively classifies students' attitude effectively rather that existing approaches. Also, the system is compared with genetic OR and AND gates, it is found that due to the complexity of the gates the system performs inefficiently depending on the results given above. From the above instances, we can tell that the proposed GS-NAND algorithm outperforms better than l network and other tested algorithms. This is due to regression problems occurring in the selection of chromosomes or the individuals in the selection process and this is avoided with the proposed algorithm. The error rate has been reduced absolutely, thereby the accuracy has increased significantly to the level of 1 for improved (1) cases and 0.957 for not improved (0) cases. Likewise the False Positive rate is lesser than previous techniques which is 0.043 for improved (1) cases and 0 for not improved (0) cases. Similarly, the precision, recall and F-measure values are promising towards the proposed GSNAND PAFT model.

CONCLUSION

Thus, from the above experiments using Weka tool, it is found that the proposed GS-NAND PAFT model results were high and accurate when compared with previous classification techniques. Depending on the predicted result, it is concluded that the prediction rate obtained in validation or testing are highly significant in finding the academic improvement of students. Also, there is no existence of disparities since the value of false positive values is less. Thus, this model proved efficient in terms of its accuracy, precision and F-measure values. Also, it could be concluded that the PAFT factors will successfully influence the success rate of students in their academics in the fore coming years. In future, this model could be enhanced using hybridization with other machine learning algorithms. Three factors set could be increase to multiple factor set for acquiring more efficient model to predict efficiently the academic performance of the students.

THE IIOAB JOURNAL
ISSN: 0976-3104

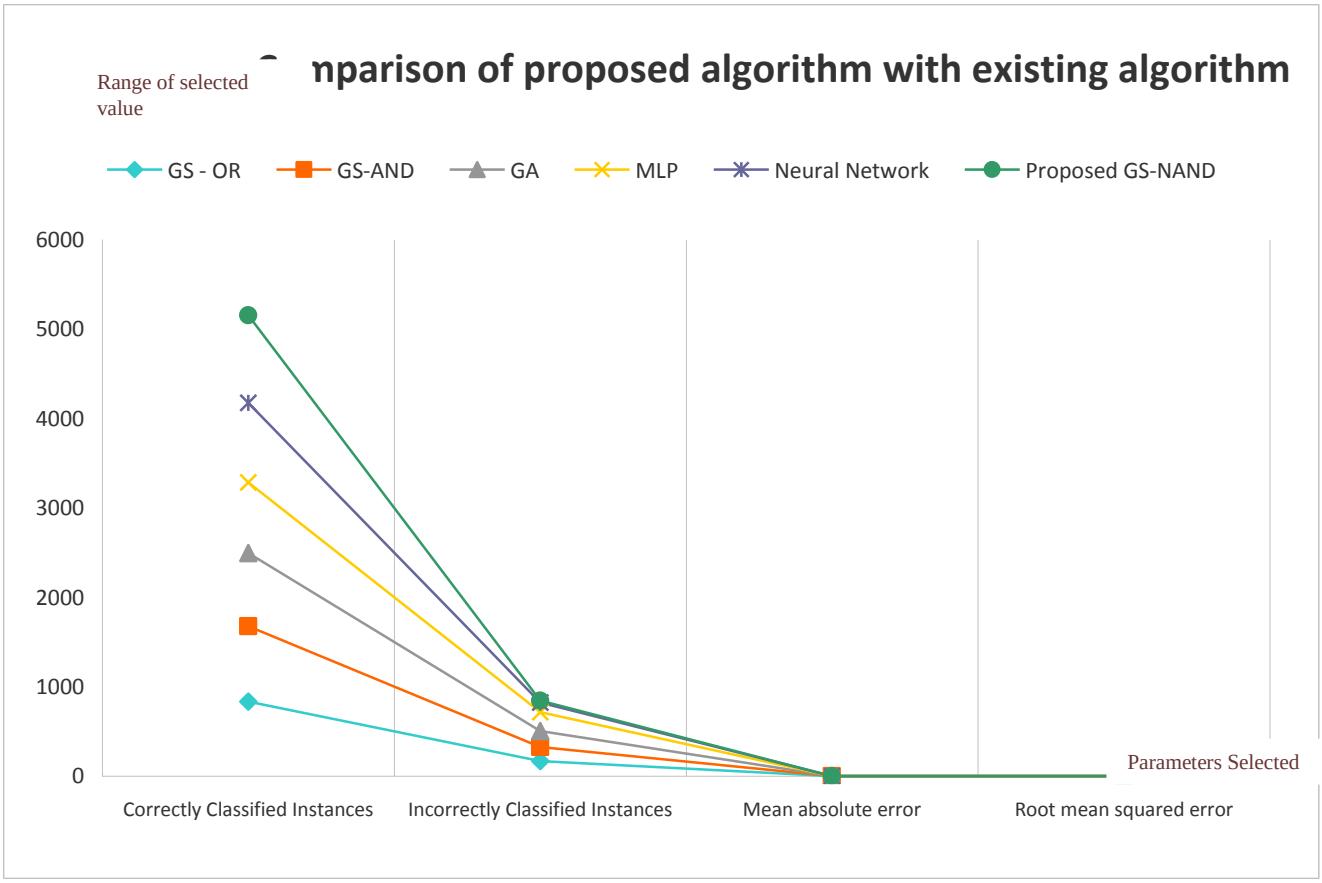

Fig: 2. Comparison of Proposed Algorithm with Existing Algorithm

CONFLICT OF INTEREST

Authors declare no conflict of interest

ACKNOWLEDGEMENT

None

FINANCIAL DISCLOSURE

No financial support was received to carry out this project

REFERENCES

[1] Srecko Natek, Moti Zwilling. [2014] Student data mining solution: knowledge management system related to higher education institutions *Expert Systems with Applications* 41: 6400–6407.

[2] Corbett A, Anderson J. [1995] Knowledge Tracing: Modeling the Acquisition of Procedural Knowledge. *User Modeling and User-Adapted Interaction* 4:253–278.

[3] Superby JF, Vandamme JP, Meskens N. [2006] Determination of factors influencing the achievement of the first-year university students using data mining methods, *Workshop on Educational Data Mining*.

[4] Bharadwaj BK, Pal S. [2011] Data Mining: A prediction for performance improvement using classification, *International Journal of Computer Science and Information Security*, 9(4): 136–140.

[5] Thilagam PS, Ananthanarayana VS. [2008] Extraction and optimization of fuzzy association rules using multi-objective genetic algorithm *Pattern Analysis and Applications*, 11(2):159–168.

[6] Ramesh V, Parkavi P, and Ramar K. [2013] Predicting Student Performance: A Statistical and Data Mining Approach. International *Journal of Computer Applications* 63(8): 35–39.

[7] Ayinde AQ, Adetunji AB Bello, M, Odeniyi OA. [2013] Performance Evaluation of Naive Bayes and Decision Stump Algorithms in Mining Students' Educational Data. *IJCSI*

THE IIOAB JOURNAL
ISSN: 0976-3104

International Journal of Computer Science Issues. 10(4): 147–151.

[8] Srimani PK, Kamath AS. [2012] Data Mining Techniques for the Performance Analysis of a Learning Model – A Case Study. *International Journal of Computer Applications* 53(5).

[9] Irfan Ajmal Khan, Jin Tak Choi. [2014] An Application of Educational Data Mining (EDM) Technique for Scholarship Prediction. *International Journal of Software Engineering and Its Applications* 8(12):31–42.

[10] Mohammed M, Abu Tair, Alaa M, El-Halees. [2012] Mining Educational Data to Improve Students' Performance: A Case Study. *International Journal of Information and Communication Technology Research* 2(2).

[11] Mohammed M, Abu Tair, Alaa M, El-Halees. [2012] Mining Educational Data to Improve Students' Performance: A Case Study. *International Journal of Information and Communication Technology Research* 2(2).

[12] Zafra A, Ventura S. [2012] Multi-instance genetic programming for predicting student performance in web based educational environments *Applied Soft Computing*, 12(8):2693–2706.

[13] Slim A, Heileman G L, Kozlick J, & Abdallah CT. [2014] Employing Markov Networks on Curriculum Graphs to Predict Student Performance. In Machine Learning and Applications (ICMLA), 2014 13th International Conference on (pp. 415–418). IEEE.

[14] Kau P, Singh M, Josan, GS. [2015] Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector, *Procedia Computer Science*, 57:500–508.

[15] Kongsakun K. [2013] An improved recommendation model using linear regression and clustering for a private university in Thailand. In Machine Learning and Cybernetics (ICMLC), 2013 International Conference on 4: 1625–1630). *IEEE*.

[16] L Ramanathan, A Geetha, and M Khalid.[2015] Mining Students' Record to Predict Their Performance in Undergraduate Degree, *Int J Appl Eng Res* ISSN, 10(1): 973–4562, L Ramanathan, S Dhanda, and D Suresh Kumar.[2013] Predicting students' performance using modified ID3 algorithm, *Int. J Eng Technol*.

[17] L Ramanathan, P Swarnalatha, K Vijayakumar, S. Kaushik, Prabu. [2014] Students Performance Prediction based on Multiple Classifiers, *Int. J Appl Eng Res* ISSN, 9(21): 973–4562.

[18] Partnership for 21st Century Skills. [2007] P21 Framework definitions. Retrieved from<http://www.p21org/storage/documents/P21 Framework Definitions.pdf/>.

[19] Turban E, Sharda R, Delen D. [2011] Decision support and business intelligent systems (9th ed.). Pearson.

[20] Berry M, Linoff G. [2000] Mastering data mining: The art and science of customer relationship management. Wiley.

[21] Sternberg, Robert J. [2006] The nature of creativity. *Creativity Research Journal* 18(1): 87–98.

APPENDIX

The samples of collecting 1000 dataset used above are shown here.

	Gender	Type of study (Full time/Part time)	Field	Department	Year of study	Marks scored in Cycle Test	Marks scored in Final Exam	Interest in coming to college/school
21	MALE	FULL TIME	ENGINEERING	CSE	4	9.1	8.9	highly interested
21	FEMALE	FULL TIME	ENGINEERING	CSE	4	8.2	7.6	interested
21	MALE	FULL TIME	ENGINEERING	CSE	4	6.3	6.3	not interested
20	MALE	FULL TIME	ENGINEERING	CSE	3	5.4	5.8	highly uninterested
20	FEMALE	FULL TIME	ENGINEERING	CSE	3	7.4	7.5	neutral
20	MALE	FULL TIME	ENGINEERING	CSE	3	6.4	6.3	neutral
20	FEMALE	FULL TIME	ENGINEERING	CSE	3	5.6	5.2	highly uninterested
20	MALE	FULL TIME	ENGINEERING	CSE	3	8.6	8.4	interested
20	FEMALE	FULL TIME	ENGINEERING	CSE	3	9.5	9.2	highly interested

A Good Knowledge of Subjects	Ability to clear exams	Interest in Homework	Interest in listening	Ability to learn, read and Write	Quality of work	Problem Solving	Team work
extreme knowledge	very high	highly interested	highly interested	very high	very high	very high	very high
knowledgeable	high	interested	interested	high	high	high	high
less knowledgeable	low	not interested	not interested	low	low	low	low
poor knowledge	very low	highly uninterested	highly uninterested	very low	very low	very low	very low

THE IIOAB JOURNAL
ISSN: 0976-3104

neutral	neutral	neutral	neutral	neutral	neutral	neutral	neutral
neutral	neutral	neutral	neutral	neutral	neutral	neutral	neutral
less knowledgeable	low	not interested	not interested	low	low	low	low
knowledgeable	high	interested	interested	high	high	high	high
extreme knowledge	very high	highly interested	highly interested	very high	very high	very high	very high

Dependability on his friends	Ability to learn and adapt from mistakes	Student shows understanding and sensitivity to needs and differences of others (i.e. ethnicity, religion, language, etc.)	demonstrates effective written communication	demonstrates effective Oral communication
highly dependable	very high	highly sensitive	highly demonstratable	demonstratable
dependable	high	sensitive	demonstratable	less demonstratable
independable	low	less sensitive	less demonstratable	poor demonstration
highly independable	very low	insensitive	poor demonstration	poor demonstration
neutral	neutral	neutral	neutral	poor demonstration
dependable	neutral	sensitive	less demonstratable	poor demonstration
neutral	low	insensitive	less demonstratable	poor demonstration
highly independable	high	sensitive	demonstratable	demonstratable
neutral	very high	neutral	highly demonstratable	highly demonstratable

ABOUT AUTHORS

Ramanathan L *has received his B.E. in Computer Science & Engineering from Bharathidasan University, Trichirappali, India and M.E in Computer Science from Sathyabama University, Chennai, India. He is currently an Assistant Professor (Selection Grade) in VIT University, Vellore, India. His area of interest is in Data Mining and Database systems, Software Egg, Cloud Computing and Virtualization. He is currently pursuing his PhD in the field of Educational Data mining. His main focus is on Prediction, Classification and Clustering for Educational systems.*

Prof. Angelina Geetha *is currently Professor at B.S.Abdur Rahman University, Chennai, India since 1995. She is Ph. D. from ANNA University Chennai, India. She is having 21 years of teaching and R&D experience. She received his Bachelor from Mother Teresa Womens University, Kodaikanal,in 1994, Master and PhD in Computer Science Engineering from Anna University, Chennai, in 2001, 2008 respectively. He published more than 35 papers in different journals, conferences, patents etc. She is member of various professional bodies like: Life Member System Society of India, Indian Society for Technical Education, senior member of IEEEetc. Her area of interest is Data Mining and Software Engineering and communication Systems.*

Prof M Khalid *has received his PhD degree from IIT Bombay, India. He is currently the Director at Neotia Institute of Technology Management & Science (NITMAS) & TNU, Kolkata, India. He is an academician having more than 35 years' experience including Professor at IIT Bombay, India, Pro-Vice Chancellor and Director at Galgotia University, Greater Noida, India, and VIT University, Vellore, India. He is author of more than 70 research papers published in International journals and conferences. He has received many more awards and memberships to his credit.*

Prof. Swarnalatha Purushotham *is an Associate Professor, in the School of Computer Science and Engineering, VIT University, at Vellore, India. She pursued her Ph.D. degree in Image Processing and Intelligent Systems. She has published more than 50 papers in International Journals/International Conference Proceedings/National Conferences. She is having 14+ years of teaching experiences. She is a member of IACSIT, CSI, ACM, IACSIT, IEEE (WIE), ACEEE. She is an Editorial board member/reviewer of International/ National Journals and Conferences. Her current research interest includes Image Processing, Remote Sensing, Artificial Intelligence and Software Engineering.*