International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-2, July 2019

To Identify the Improvement Pattern of Self-financing Arts and Science College Student's Academic Performance using Classification Algorithms

R. Senthil Kumar, K. Arulanandam

Abstract: ***The aim of this research work is to identify the improvement pattern of academic performance of final year students of self-financing arts and science colleges. The data was collected from the students of nine Arts and Science Colleges. The data contains demographic, socio-economic, residence and college location, subjects, infrastructural facilities, faculty concern and self-motivation attributes. The classification algorithms like Naïve Bayes, Decision tree and CBPANN are applied on the student's data. The outcome of the research can be used to improve the academic performance students studying in self-financing arts and science colleges located in educationally backward areas. The experiment results shows that the accuracy value for Naïve Bayes algorithm is 92.63%, accuracy value for Decision Tree algorithm is 96.41% and accuracy value for CBPANN algorithm is 99.49%***

Keywords: Naïve Bayes, Decision Tree, CBPANN Algorithms and Educational Data Mining

I. INTRODUCTION

This research work does a comparative analysis among Naïve Bayes (NB), Decision Tree (DT) and Callback Priority Artificial Neural Network (CBPANN) Algorithms. This analysis is done to select the best algorithm among three algorithms. The patterns obtained are used to forecast the improvement status of student's belonging to arts and science colleges. The parameters are Grade Point Average (GPA), College Location, Father Education, Mother Education, Classification, Subject, Facilities, Faculty Concern, Skills and Attendance. The percentage of marks value of final year arts and science college students are categorized into three different 41% to 60%, 61% to 80% and 81% to 100%. Matlab software is used for experimental process. Data mining algorithms are successful in many areas like education, technology, business, medicine, economic, games, communication and agriculture. The main challenge for arts and science colleges is to analyze their student's performance.

Revised Manuscript Received on 30 July 2019.
* Correspondence Author
R. Senthil Kumar*, Research Scholar, Department of Computer Science, Periyar University, Salem, Tamilnadu, India
K. Arulanandam, Assistant Professor and Head, Department of Computer Science, GTM College, Gudiyattam, Tamilnadu, India.

We used Naïve Bayes, Decision Tree and CBPANN algorithm to analyze the data. Applications of data mining using machine learning and statistics in educational field describes educational data mining. A quick growth in educational data is an important factor in the institutions.
Quality of education is the main motto for every educational institution. Various attributes are used to forecast the student's performance. This study will construct the most confidential model for arts and science college student's performance prediction in the future. The problems in pattern recognition, classification and prediction are solved by neural network algorithm. Students, educational researchers and developers, teachers and parents are getting benefitted from educational data mining. Data mining algorithm, machine learning algorithm and statistical methods are used to analyze the data.

II. REVIEW OF LITERATURE

A scheme was presented for predicting student's academic performance of first year undergraduate computer science students. In order to select the best prediction model, the algorithms like Rule Based, Decision Tree and Naïve Bayes are applied on student data. When compare to other algorithms rule based is the best algorithm with 71.3% accuracy value. Poor and average students are given special care by the lecturers [1]. A research project was conducted at Bulgarian University. The influencing factors are number of non-success students in the first year university exams and enrollment score [2]. The study is based on prediction process from various research papers. For performance analysis there are many data mining methods are available. For performance analysis there are many efficient data mining algorithms are available. Among all Educational data mining is the best technique to handle student's prediction [3]. The present student's performance has been predicted using data mining technique called as classification rule. Compare to other classification algorithms, Naïve Bayes algorithm has highest accuracy. To predict the performance of students at end semester the information's like Assignment marks, Seminar and Attendance were collected from the students [4]. Decision tree algorithm is used to predict student's performance. To analyze the skill classification technique is used. Course evaluation questionnaire can be analyzed using data mining technique.

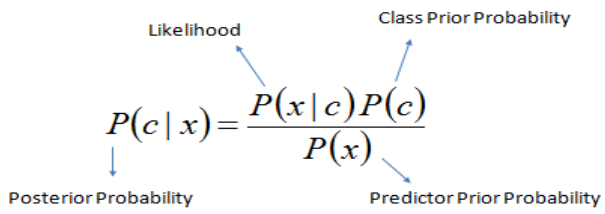
Good and average student performances and their weaknesses in particular subjects can be separated using most important variables. These variables can help instructors to improve their performance [5]. Dataset of Turkey student records can be evaluated using J48 decision tree and Naïve Bayes classifier. Most of researchers used CGPA and internal mark attribute to predict student performance which play a vital role in improving educational quality. Instructors' performance can be predicted using students achievement [6]. University of Tuzla conducted a survey among first year students during summer semester for the academic year 2010-2011. When compare to decision tree and neural network algorithms, Naïve Bayes classifier is the best method for predicting student performance [7]. The study presents result analysis of every semester. It compares classification algorithms like Naïve Bayesian algorithm, rule based algorithm and decision tree. Result shows that J48 algorithm gives lower error rate and higher accuracy [8]. To analyze student information SVM, associative classification, naïve bayes and decision tree data mining techniques are used. R studio programming tool used to convert data into decision tree model [9]. Neural network algorithm and Naïve bayes algorithm have the highest accuracy of 75%. Oversampling was imminent because the number of instances in the data set was small. To determine the model which gives highest accuracy and to improve the models through preprocessing is the primary objective [10]. For analysis and prediction the powerful tool is Data mining. The applications of data mining in various fields like prediction, loan assessment, marketing, advertising, fraud detection and education [11]. Data mining tasks on student's data gives interesting patterns and results which help both students as well as colleges. The result analysis and performance graph of student for further prediction can be given using FP-Growth algorithm. Attendance and mid-term result are used to conclude the result of students. The correlation between professor performance and student performance were found [12]. Classification technique and feature reduction are used to reduce error rate. The most affecting factors in student and academic performance are exam grades and attendance. Statistical package for social sciences (SPSS) was used to arrange and capture data. Both clustering and classification techniques are used [13].

III. MATERIALS AND METHODS

Data collection was done from the students of nine Arts and Science colleges, described by 21 parameters including residence and college location, father and mother education, liking college, faculty concern, knowledge and skill, attendance, communication skill, time spend for reading, percentage of marks, etc.

Naive Bayes Algorithm:

The popular classification algorithm is Bayesian Classifier because of its good performance of real world problems, computational efficiency and simplicity.



$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

Step 1: Read 21 input attributes from dataset in excel and load it in Matlab software
Step 2: These 21 input attributes are training data. So in training data we have 21 features or 21 Categories
Step 3: We have 2 output attributes (1) Student performance may improve and (2) Student Performance May not improve.
Step 4: The data is read column by column
Step 5: Declare the unknown in this step
Step 6: Calculate means (Add all the values together and divide by number of values in the Data set)
Step 7: Construct covariance matrix of size 21x21 for both student's performance may improve and student's performance may not improve using the formula:

$$\text{cov}(x)=1/n-1\sum(xi-\bar{x})(yi-\bar{y})$$

Step 8: Calculate determinants
Step 9: Construct inverse covariance matrix for both students' performance may improve and Student's performance may not improve data's
Step 10: Give input attribute values
Step 11: Calculate accuracy for both students' performance may improve and students' Performance may not improve data's
Step 12: Display the decision results

Machine learning programs use training data. It is also used in neural network technology. Testing data includes only input data, not the corresponding expected output. The testing data is used to assess how well our algorithm is trained. The training data includes both input data and the corresponding expected output. Sum divided by count is called mean. We can calculate mean using the formula:
Variable Name = Sum (Attribute Names)/ Total Number of data sets
Covariance Matrix can be calculated using the formula:
Variable Name = Attribute Name – Means (Attribute Name)
Determinants can be calculated using the formula: det (Matrix Name)

Decision Tree Algorithm:

The most popular technique in Educational Data Mining is Decision Tree algorithm. It gives explanations in human friendly form for decision makers to make further action.

$$Gain(S, A) = Entropy(S) - \sum_{v \epsilon Values(A)} \frac{|S_v|}{|S|}.Entropy(S_v)$$

Step 1: Read 21 input attributes from dataset in excel and load it in Matlab software.
Step 2: These 21 input attributes are training data. So in training data we have 21 features or 21 Categories
Step 3: We have 2 output attributes that are students' performance may improve and students' Performance may not improve

Step 4: Load 21 input attributes names in variable input names in Matlab
Step 5: Load 2 output attributes names in variable output names in Matlab
Step 6: Load input data's in variable X and load output data's in variable Y
Step 7: Train a regression tree using dataset
Step 8: Load input data's for training
Step 9: Plot fit against training data
Step 10: Plot Error

Plot means drawing figure window in Matlab. We are plotting figures for training data using plot command in Matlab. We are calculating data set losses for all data's in Matlab command, this is called plot error.

CBPANN (Call Back Priority Artificial Neural Network) Algorithm:

Call Back Priority Artificial Neural Network algorithm are used to predict the performance of students based on the given dataset.
Step 1: In CBPANN algorithm, we have input data's and the required output for that input Data
Step 2: The network is composed of an input layer of one neuron presenting input value, an Output layer with one neuron presenting output value and a hidden layer with 24 neurons
Step 3: The first step is to set the input and target datasets as follow

- data=xlsread('Dataset.xlsx','Sheet1');
- input=data(:,1:21)';
- target=data(:,22)';

Step 4: Then we use the neural network function to design the network

- net = nn(minmax(input))

Step 5: Set number of hidden neurons for output

- net = nn(minmax(input),[24])
- Here 24 is hidden neuron

Step 6: Train the neural network
Step 7: Training process begins and prediction is done

Quick convergence and Time efficiency are the advantages of our proposed CBPANN Algorithm. Normalization is one of the main parts of ANN learning process. If you do not normalize your inputs between you could not equally distribute importance of each input, thus naturally large values become dominant according to less value during ANN training. To get more accurate results normalization should be done.

IV RESULTS AND DISCUSSION

Based on the test results, Naïve Bayes algorithm has 92.63% accuracy value, Decision Tree algorithm shows the accuracy value of 96.41% and CBPANN algorithm shows the accuracy value of 99.49%. Among the three algorithms CBPANN algorithm shows the highest accuracy value.

Table 1: Accuracy for Naive Bayes Algorithm

S.No	Accuracy Calculation	Performance
1	True Positive(TP)	46.28
2	False Positive(FP)	3.71
3	True Negative(TN)	46.35
4	False Negative(FN)	3.64
	Accuracy	92.63

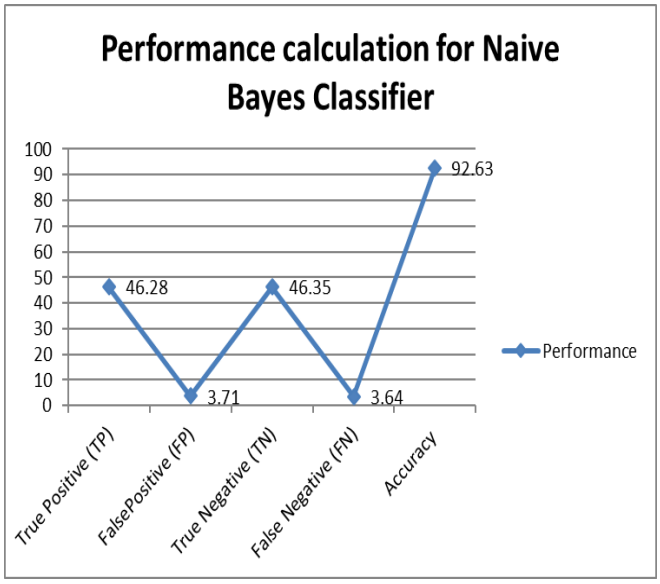


Fig. 1: Performance Calculation for Naïve Bayes Classifier

Table 2: Accuracy for Decision Tree Algorithm

S.No	Accuracy Calculation	Performance
1	True Positive(TP)	48.24
2	False Positive(FP)	1.79
3	True Negative(TN)	48.17
4	False Negative(FN)	1.79
	Accuracy	96.41

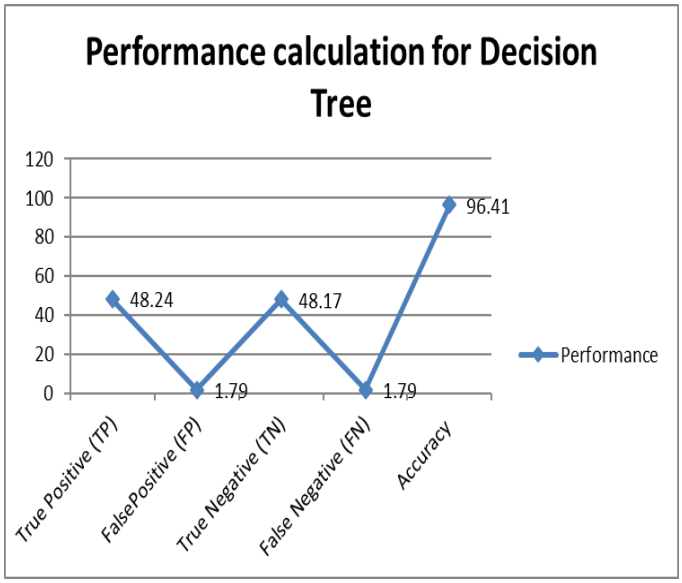


Fig. 2: Performance Calculation for Decision Tree Classifier

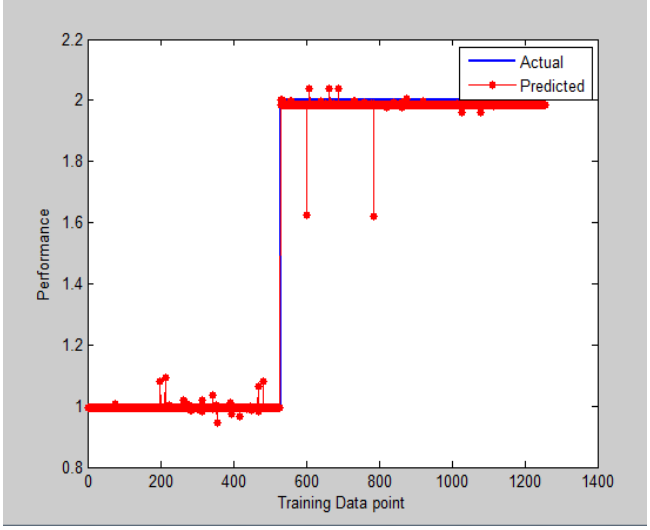


Fig. 3: Matlab result for decision tree algorithm

In Fig. 3, we have drawn Matlab figures for actual data set and predicted data set. Actual dataset are input data sets we have in Excel. Predicted dataset are the data's we got after processing input dataset through decision tree algorithm. Figure number 3 shows the deviation or difference between the data's in excel and data's after processing through decision tree algorithm. Actual data's are drawn in blue color graph in Matlab. Predicted data's drawn in red color graph in Matlab.

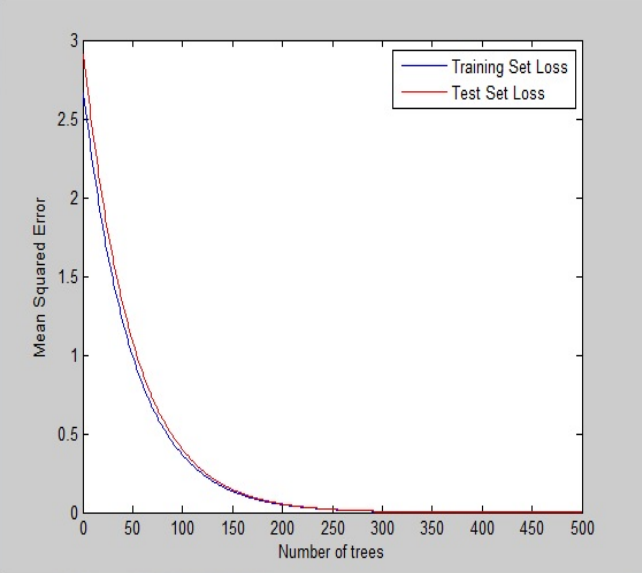


Fig 4: Matlab result for Test data and Train data

In Fig. 4, we have drawn Matlab figures for Test data (Actual data set in Excel) and Train data (training input data's in excel using algorithm).Figure shows the deviation or difference between Test data and Train data. Training data's are drawn in blue colour graph in Matlab. Test data's drawn in red colour graph in Matlab

Table 3: Accuracy for CBPANN Algorithm

S.No	Accuracy Calculation	Performance
1	True Positive(TP)	47.52
2	False Positive(FP)	0.28
3	True Negative(TN)	51.97
4	False Negative(FN)	0.21
	Accuracy	99.49

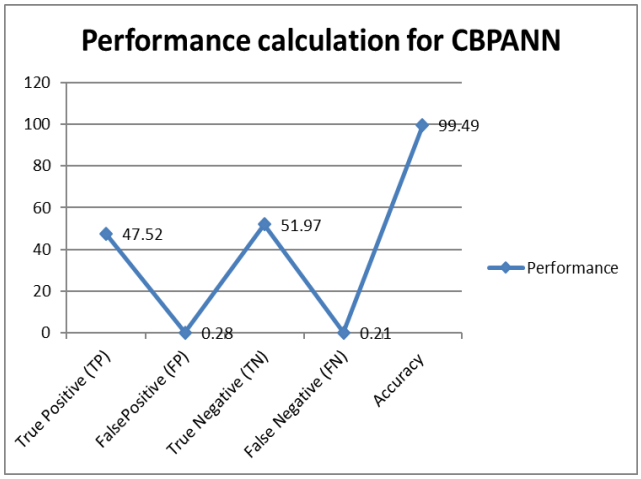


Fig. 5: Performance Calculation for CBPANN Algorithm

Table 4: Comparison Result for Naïve Bayes, Decision Tree and CBPANN Algorithms

S.No	Algorithm	Accuracy in percentage
1	Naïve Bayes (NB)	92.63
2	Decision Tree(DT)	96.41
3	CBPANN	99.49

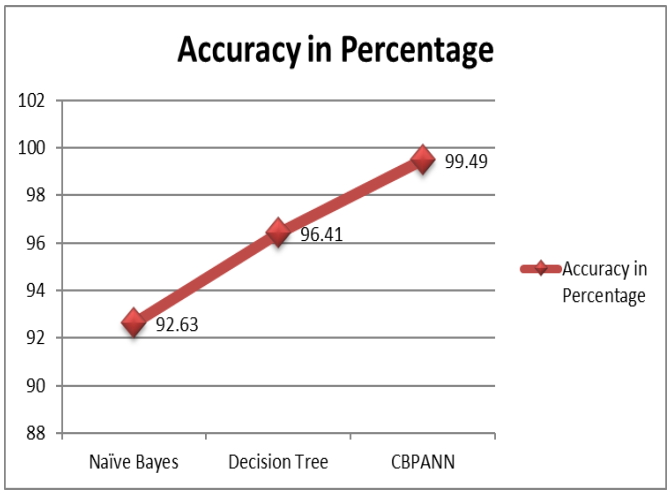


Fig. 6: Comparison Result for Naïve Bayes, Decision Tree and CBPANN Algorithms

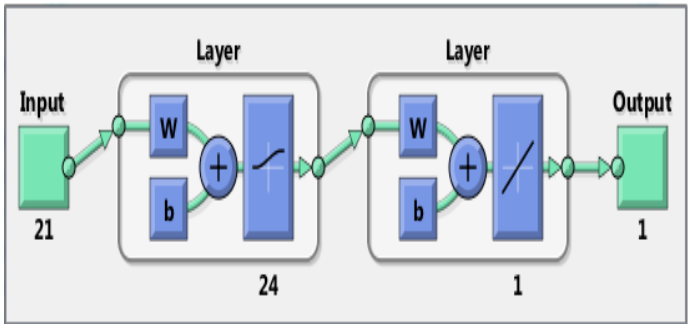


Figure 7: CBPANN Algorithm result

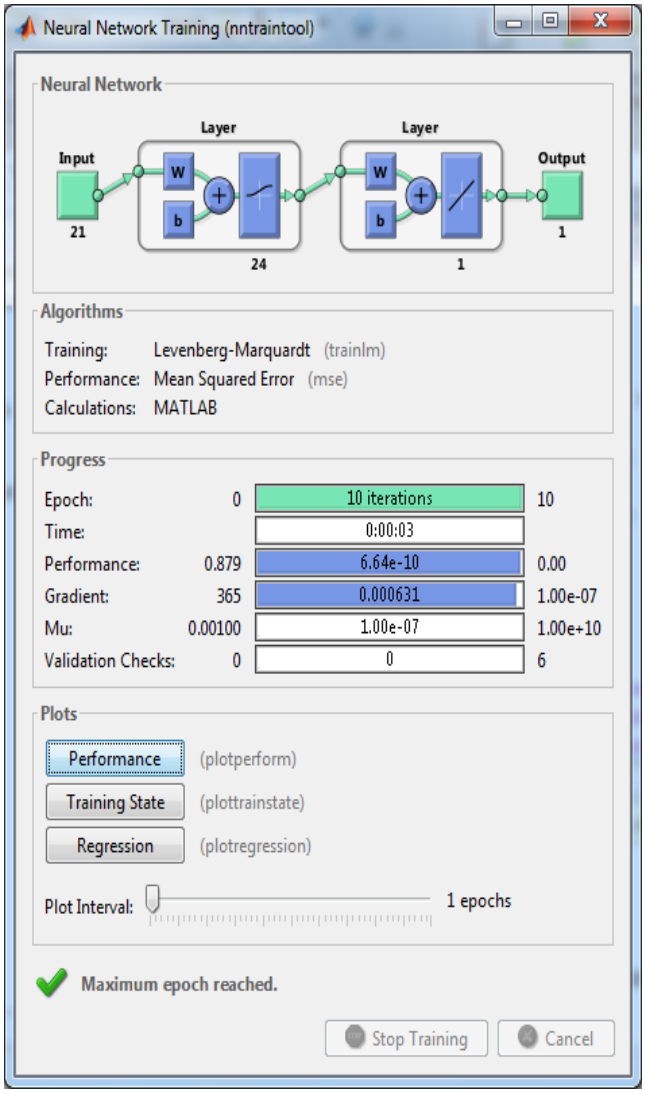


Fig. 8: CBPANN Algorithm output

In Fig. 8 CBPANN Algorithm creates its own bias and weight for input dataset. Number of neurons will be given by us in Matlab program to achieve output.

V CONCLUSION

This research study is used to identify the improvement pattern of self-financing arts and science college students. This research conducts a comparative analysis of Naïve Bayes, Decision Tree and CBPANN algorithms using Matlab software. This research work concludes that CBPANN algorithm is the best method for identifying the improvement pattern of student's performance. The test result shows that Naïve Bayes algorithm has 92.63% accuracy value, decision tree algorithm has the accuracy value of 96.41% and the CBPANN algorithm has the highest accuracy value of 99.49%. The outcome of this research can be applied in the self-financing arts and science colleges located in educationally backward areas.

REFERENCES

1. Fadhilah Ahmad et.al, The Prediction of Students' Academic Performance Using Classification Data Mining Techniques, Applied Mathematical Sciences, Vol.9, No. 129, 2015, 6415-6426
2. Dorina Kabakchieva, Predicting Student Performance by Using Data Mining Methods for Classification, Cybernetics and Information Technologies, Vol. 13, No.1, 2013, ISSN: 1311-9702
3. A. S. Arunachalam and K. Rajeswari, An inclusive survey of students performance with various data mining methods, International Journal of Engineering & Technology, No. 7, 2018, 522-525
4. Shruthi P and Chaitra B P, Student Performance Prediction in Education Sector Using Data Mining, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 3, 2016, ISSN: 2277 128X
5. Ankita A Nichat and Anjali B Raut, Predicting and Analysis of Student Performance Using Decision Tree Technique, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 5, Issue 4, 2017, ISSN: 2320-9798
6. Ahmed Mohamed Ahmed et. al, Using data mining to predict instructor performance, Procedia Computer Science, 12th International Conference o Application of Fuzzy Systems and Soft Computing 2016, 29-30, Elsevier
7. EdinOsmanbegovic and MirzaSuljic, Data Mining Approach for Predicting Student Performance, Economic Review – Journal of Economics and Business, Vol. X, Issue 1, 2012
8. Bhavesh Patel and ChetanGondaliya, Student Performance Analysis Using Data Mining Technique, International Journal of Computer Science and Mobile Computing, Vol 6, Issue 5, 2017, Pg 64-71, ISSN 2320-088X
9. Lakshmipriya K and Arunesh P.K, Predicting Student Performance Using Data Mining Classification Techniques, International Journal of Innovative Research in Science and Engineering, Vol 3, Issue 2, 2017
10. Syed TanveerJishanet. al, Improving accuracy of students' final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique, Decision Analytics, Springer, 2015
11. PoojaThakar et al., Performance Analysis and Prediction in Educational Data Mining: A Research Travelogue, International Journal of Computer Applications, Volume 110 – No. 15, 2015, 0975 – 8887
12. Ms. JuhiArora, Comprehensive Analysis of Academic Performance using Data Mining Techniques-with Special Reference to UG Class of Engineering, International Journal for Research in Applied Science & Engineering Technology, Volume 6 Issue III, 2018, ISSN: 2321-9653
13. Nawal Ali Yassein et al. Predicting Student Academic Performance in KSA using Data Mining Techniques, Journal of Information Technology & Software Engineering, Volume 7, Issue 5, 2017, ISSN: 2165-7866

AUTHORS PROFILE

Mr. R. Senthil Kumar pursed Bachelor of Science from NGM College, Bharathiar University, Coimbatore in 1991 and Master of Computer Applications from Madurai Kamaraj University, Madurai in 2002. He pursed Master of Philosophy in Computer Science from Periyar University, Salem in year 2007. He is currently pursuing Ph.D from Periyar University, Salem and currently working as Assistant Professor and Head in Department of Computer Science and BCA, RTG Arts and Science College, Polur, Tamilnadu, India since 2010. He is a member of Computer Society of India. He has published two papers in reputed international journal. His main research work focuses on Educational Data Mining. He has 10 years of teaching experience and 5 years of research experience.

Dr. K. Arulanandam pursed Bachelor of Science from University of Madras in 1997 and Master of Computer Applications from University of Madras, Chennai in 2001. He pursed Master of Philosophy in Computer Science from Manonmaniam Sundaranar University in 2003. He pursed Doctor of Philosophy in Computer Science from Vinayaka Missions University, Salem and currently working as Assistant Professor and Head in Department of Computer Science, GTM College, Gudiyattam, Tamilnadu, India since 2011. He is a life member of CSI, CSTA, IAENG, ISTE, IACSIT, IJST, ACEEE and ISCA. He has published 36 research papers in reputed national and international journals. His main research work focuses on Computer Networks. He has 17 years of teaching experience and 14 years of Research Experience.