

A novel academic performance estimation model using two stage feature selection

Pamela Chaudhury¹, Hrudaya Kumar Tripathy²

¹Department of Computer Science & Engineering, Silicon Institute of Technology, India

²KIIT, India

Article Info

Article history:

Received Nov 2, 2019

Revised Feb 15, 2020

Accepted Mar 5, 2020

Keywords:

Academic result estimation

Classification

Dummy variables

Feature selection

RBF

ABSTRACT

Educational data mining has gained tremendous interest from researchers across the globe. Using data mining techniques in the field of education several significant findings have been made. Accurate academic performance estimation has been a challenging task due to the variety of students' attributes involved. In this study we have developed a novel framework to estimate the academic performance of students. Our proposed model outperformed existing models of students' academic performance determination and gives a new direction to academic performance estimation. The proposed model can help not only to reduce the number of academic failures but also help to comprehend the factors contributing to a students' academic performance (poor, average or outstanding). Some of the techniques used were conversion of categorical attributes into dummy variables, instance segregation, classification using an optimised and improved differential evolutionary algorithm.

Copyright © 2020 Institute of Advanced Engineering and Science.

All rights reserved.

Corresponding Author:

Pamela Chaudhury,

Department of Computer Science & Engineering,

Silicon Institute of Technology,

Bhubaneswar, Odisha, India.

Email: pamela.chaudhury@gmail.com

1. INTRODUCTION

Educational Data Mining is an emergent field of research that uses data mining in the field of education. Determination of academic performance has gained the attention of several researchers in the recent times using machine learning techniques [1]. Machine learning algorithms enable educationist to predict the students' results before the examination was conducted. These results could be used to make pedagogical interventions to prevent failures of the academically underperforming students. Many researchers have used data mining techniques in the domain of education [2-9]. Most researchers have used classification algorithms to train the classification /prediction model using a dataset consisting of students' information. Some of the classification algorithms used was neural network, decision trees, support vector machines and naive bayes classifier [10, 11]. Experiments initially were conducted using these popular classification techniques. The accuracy of the classification model highly depends upon the quality of the data used as input. Data pre-processing techniques like handling of missing values, noisy data and balancing the minority class instances were used before the implementation of classification technique to improve the accuracy of the classification models [12, 13]. Feature selection has been used to identify the relevant attributes in student datasets and others [14-16].

The data collected from the students through questionnaire is often incomplete and requires pre-processing techniques to be implemented. The student dataset is generally unbalanced. A dataset is unbalanced if the class labels of data instances are not uniformly distributed. In such datasets it is found that most instances belong to one class often called the majority class and a few instances of the dataset belong to the minority

class. Several data balancing techniques are used to handle this problem. The student dataset attributes mostly consist of categorical attributes. The dataset used for such classification models consisted of academic, demographic information and psychometric from students or student database. Information considered for classification was factors like gender, age, social background, attendance percentage, previous academic record etc. A detailed literature survey was done to find out the attributes that have been used in previous research. Some researchers used attributes like age, gender, civil status, city, income, assignments completed etc. Others used attributes like internal grades and Cumulative grade point Average (CGPA)). Demographic features like parental background, health condition, were found to have an impact on academic performance [17]. Psychometric tests have been conducted by researchers to determine academic performance.

Amongst several classification algorithm, Radial Bias Function (RBF) network, Decision tree, Naive Bayes classifier and Support Vector machines (SVM) have gained significant popularity. RBF has been used for classification in several domains [18]. Support Vector machines (SVM) has been used by researchers in classification problems and has been found to outperform other classifiers [19]. Naive bayes classifier has gained its popularity due to its high accuracy despite its simplicity [20]. The attributes of the dataset have are often tested using feature selection algorithms. Feature selection is the process of selecting a subset of attributes/features from a set of features [21]. The outcome of feature selection algorithm is a sub-set of features using which the classification model can accomplish the task of generating highly accurate results.

The findings from literature survey were as following:

- a) The student dataset mostly consisted of several categorical attributes. Categorical data attributes need to be handled properly before it can be used as input for classification. The attributes like gender, mother's qualification, internet browsing pattern are categorical data. The categorical data of any dataset holds valuable information. The categorical data cannot be treated as numeric data for classification algorithms and need to be handled.
- b) All models for academic performance determination are based on classification. Some models of academic performance determination use feature selection techniques to find the significant features from the student dataset.
- c) No research has been conducted before to identify attributes specific of each class of students (based on their academic performance: poor, average and outstanding). For example, identification of specific attributes/ characteristics of the academically underperforming students has not been tried by researchers.

Hence our problem statement was to identify the characteristics of academically weaker students, the characteristics of academically average students and the characteristics of academically outstanding students using an optimised classification algorithm. Further, the intention was to compare these characteristics and address the issues attributing poor and average academic performances. The proposed solution consisted of comparison and implementation of existing EDM models. Further a novel framework for academic performance estimation is proposed. Our proposed novel model not only outperforms existing EDM models but also helps to identify specific attributes of students who are academically weak, average or outstanding. The paper is organised as follows: Section 2 describes the experimental setup. Section 3 discusses the result obtained and its implication. Section 4 concludes the paper.

2. RESEARCH METHOD

2.1. Data collection

The primary dataset consisted of students' data from KIIT University and Biju Patnaik University. Before considering data collection from students, a systematic analysis of the features that should be used in the research was done. For this purpose, we have taken the opinion of researchers from literature survey, faculties, parents and placement cell. The following attributes were considered as the relevant and necessary features required for the estimation of academic performance.

- Gender
- Background (urban, rural)
- Mother's qualification
- Mother's occupation
- Father's qualification
- Father's occupation
- Parents staying together or apart
- Joined engineering willingly or under parental pressure
- Most engineering subjects are interesting
- Most engineering subjects are complex
- Participation in extracurricular activities

- Time spent in sports or extra-curricular activities
- Study duration weekly
- Time spent with friends weekly
- sleeping time duration daily
- Quality of family relations
- Financial distress
- Health impairment
- Type of internet connectivity on smartphone
- Frequency of internet usage daily
- Frequency of internet access during class
- Does internet feel addictive
- Absent percentage
- Previous failures
- 10 class division
- 12 class division
- Internal grade1
- Internal grade 2
- Previous CGPA1
- Previous CGPA 2
- What could help to improve grades
- Any other issue affecting education

These attributes were collected from students pursuing engineering from two educational intuitions: School of computer engineering, KIIT University and Silicon Institute of Technology in Bhubaneswar, India. Data was collected in form of questionnaire. Prior permission was taken from the concerned authorities.

2.2. Data pre-processing

The data collected from students could not be directly used as input for the classification. Pre-processing the data improves the effectiveness of the classification algorithms. There were missing and incorrect values in some of the data instances. Missing and invalid data entries were replaced by the mode of the attribute values. Since most attributes in the data set are categorical data hence special techniques were used to handle the categorical data. For example, the attribute ‘father’s occupation’, could have the values: service, business and others. In order to use these values as a part of classification model, these values should be mapped into numbers. One of the frequently used methods for this would be mapping each of the unique value of the attributes to numbers say 1, 2, 3. The number of values required will depend upon the number of levels of an attribute for example number of levels is three for attribute occupation. However, converting values into numbers may not be correct procedure. For example, if attribute value service is marked as one, business as two and others as three then it implies that business is less than service and service is less than business. Working with these assumptions would give inaccurate results. Hence a search for other approaches with more accurate results was done. Dummy variable is one of the techniques. Dummy variable represents one level of categorical variable. Presence of a level is represented by 1 and absence is represented by 0. For example, instances 1, 2, 3 and 4 are shown in the table with the attribute gender and their respective class are shown in Table 1.

Table 1. Variable gender

Sl.	Gender	Class
1	Male	Fail
2	Female	Pass
3	Female	Pass
4	Male	Pass

For every level present, one dummy variable will be created. Table 2 shows dummy variables male and female created. The same principle was adopted for all categorical attribute levels were converted into dummy variables. This further can be used for classification algorithms. The dataset was unbalanced and 90% of the instances belonged to the majority class. To restore balance synthetic samples were generated using SMOTE algorithm [22].

Table 2. Dummy variable male and female created

Sl.	Male	Female	Class
1	1	0	Fail
2	0	1	Pass
3	0	1	Pass
4	1	0	Pass

2.3. RBF as a classifier

Though several classifiers were used, RBF was mostly used in our proposed work as it outperformed other classification models. RBF network is a supervised learning algorithm used to build classification models. It has been increasingly used by researchers as a popular artificial neural network in the fields of machine learning. RBF is a three-layered feed forward network as shown in Figure 1. It is a three-layered feed forward network consisting of input layer, a hidden layer and output layer. The hidden layer implements the activation function. The training process of RBF network is faster than MLP. It allows selecting suitable parameters for centre and spread of hidden units and can provide high accuracy rate comparable with classification methods like SVM.

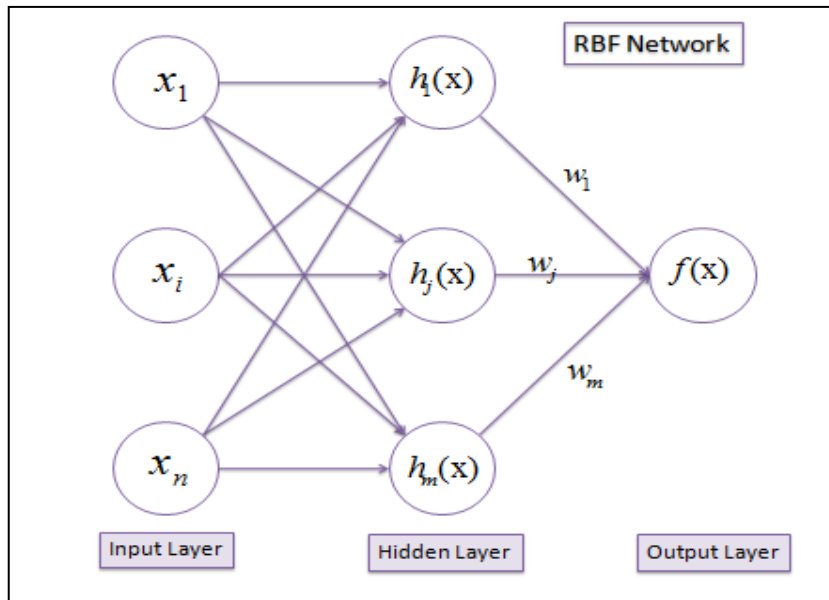


Figure 1. RBF network

The hidden layer implements the radial activation function. The output is derived from the weighted sum of hidden units where μ and σ

$$\phi_i(x) = \exp\left(\frac{\|x - \mu_i\|^2}{2\sigma^2}\right) \tag{1}$$

$$y_{net} = w_0 + \sum_{i=1}^N w_i e^{\left(\frac{-\|x - \mu_i\|^2}{2\sigma^2}\right)} \tag{2}$$

2.4. Optimised RBF with novel and improved DE

Differential Evolution is heuristic search algorithm [23] similar to the genetic algorithm which maintains a population of possible solution. There are different variants of the differential evolution. However, the authors have used a scheme based on the DE/ rand/1/bin scheme [24]. It also maintains a pool of chromosomes each encoding a possible solution. An iterative process of mutation, recombination (crossover) and selection of chromosomes continues until the desired fitness of the chromosomes is achieved. Displaying the iterative process of mutation, recombination and selection as shown in Figure 2.

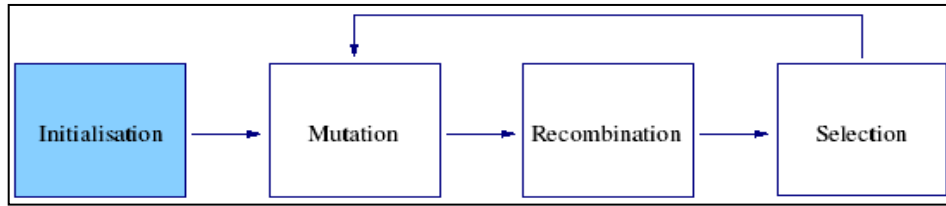


Figure 2. Displaying the iterative process of mutation, recombination and selection

A randomly generated population of size n is initialized and evaluated using a fitness function f iteratively. The mutation process creates a donor vector which is capable of encoding the solution using randomly selected members for the population. Recombination then creates a trial vector by combining two donor vectors. Selection is the process of determining whether the newly generated trial vector can form a part of the population using fitness function. The donor vector is obtained using the following equation:

$$v_v(t+1) = x_k(t) + f_m * (x_l - x_m(t)) \quad (3)$$

Where k , l and m are mutually distinct and randomly selected indices. The trial vector is created by the following process of recombination as explained in (4):

$$u_i(t+1) = \begin{cases} v_i(t+1) & \text{if } (rand \leq c_r) \text{ or } (i = rand(ind)) \\ x_i(t+1) & \text{if } (rand > c_r) \text{ and } (i \neq rand(ind)) \end{cases} \quad (4)$$

Where i is a random number generated, c_r is the crossover constant. If the child instance created or the trial vector has higher fitness function than the parent will be replaced. Then selection process helps to select the best individuals as per (5).

$$x_i(t+1) = \begin{cases} u_i(t+1) & \text{if } f(u) < f(x) \\ x_i(t+1) & \text{otherwise} \end{cases} \quad (5)$$

In the novel and improved Differential Evolution algorithm, we have modified the mutation process and altered the process of creating donor vector as in (6).

$$v_v(t+1) = x_{r_{best}} + f_m * (x_{r_1} - x_{r_3}(t)) + f_m * (x_{r_2} - x_{r_4}(t)) \quad (6)$$

where, r_1 , r_2 , r_3 and r_4 are randomly selected four indices. $x_{r_{best}}$ is the best valued individual among x_{r_1} , x_{r_2} , x_{r_3} and x_{r_4} . Also $x_{r_1} > x_{r_2}$

$$\begin{aligned} & x_{r_2} > x_{r_3} \\ & \text{and } x_{r_3} > x_{r_4} \end{aligned}$$

The reason behind modifying the donor vector is that an improved donor vector would contribute to better individuals in the population. It will further positively impact the recombination process and selection process.

2.5. Feature selection

Feature selection is the process of identification and removal of redundant attributes. These attributes do not contribute to the classification process [25-27]. The reason behind feature selection is not only to reduce computational complexity but also makes the classification model simpler and accurate. There are two types of feature selection method: filter and wrapper-based method. Filter based classification model is a process assigns a weight to each attribute [28]. The weight depends upon the statistical characteristics of the dataset. Feature selection also helps to analyse and interpret results. Feature selection has been done using filter-based feature selection method. There are several filter-based feature selection algorithms [29, 30]. The Relief Based Feature Selection, Entropy Based Feature Selection and Correlation based feature selection was used. All these algorithms have been used by several researchers for their accuracy and simplicity. Using these algorithms each of the attributes of the student dataset was investigated. Since relief, correlation and entropy are filter based feature selection methods hence they return a ranking of the features. These rankings helped to identify the prominent attributes responsible for determination of academic results. Figure 3 demonstrates our process model.

After pre-processing the data and implementing feature selection, our proposed model uses a novel differential evolution algorithm to optimize the RBFN as the classification algorithm. Further instances of each class category is segregated and second stage feature selection is implemented to understand and analyse the patterns in students' academic performance.

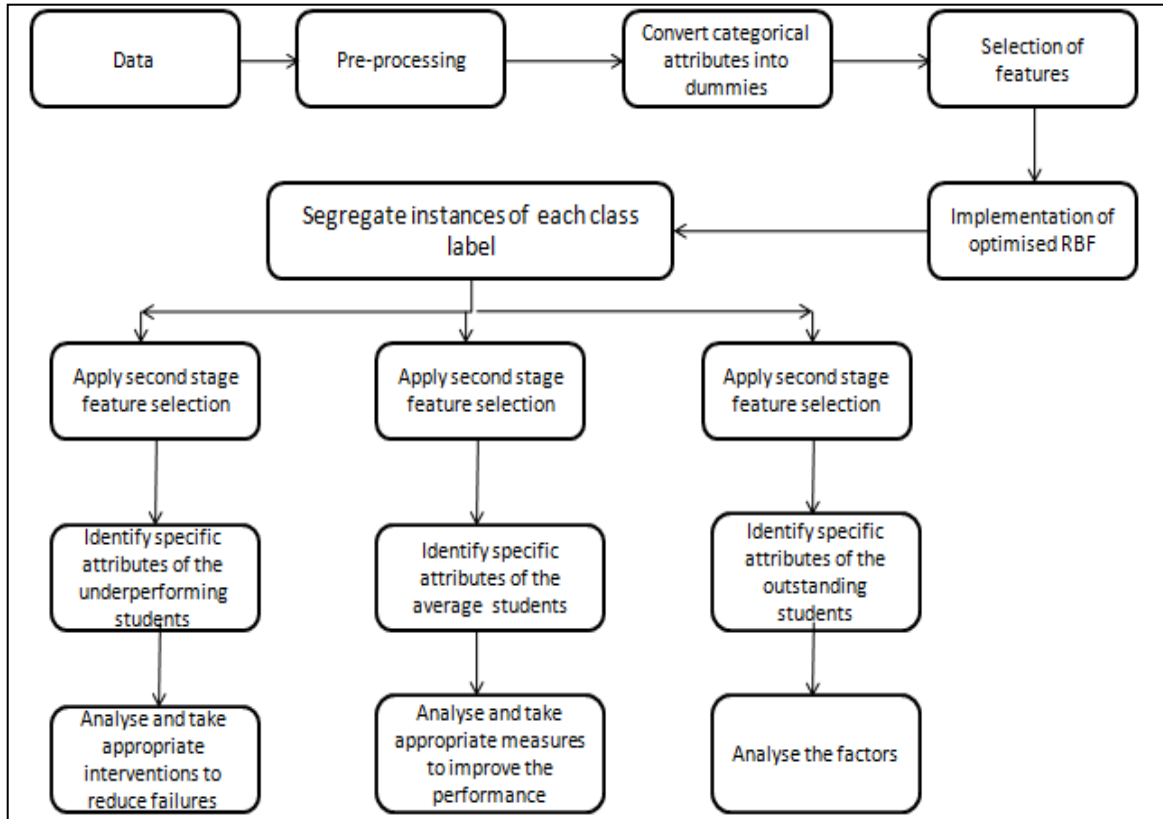


Figure 3. Our proposed model using EDM technique

3. RESULT ANALYSIS AND DISCUSSION

This section presents the results of the experiments of the proposed work as well as compares the experimental techniques used by other researchers with our model. Many works have been done to determine academic performance. Experiments conducted to determine academic performance using data mining can be outlined into the following three types:

- a. Classification based
- b. Classification with data pre-processing
- c. Classification on pre-processed data with selected features

The authors have used these methods on their data and then have compared these approaches with their proposed method. Accuracy and f-measure have been used to evaluate the effectiveness of the experiments.

$$Fmeasure = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{7}$$

Similarly Figure 4, 5, 6 displays the f-measure of classification model of type I, II and III experiments respectively. It was observed that RBF Neural Network, optimised with differential evolution, resulted in better accuracy and f measure value. They also used a novel and modified differential evolution. Therefore, the authors implemented their proposed work with this model. In their approach as discussed in Section 2, all categorical attributes were converted into dummy variables. The RBF model using dummy variables gave an accuracy of 84% which is a significant improvement over type II experiments. Our model consisted of incorporating dummy variables, usage of feature selection algorithms to remove redundant attributes. Further the RBFN classification

model was optimized using differential evolutionary algorithm. We also modified the differential evolutionary algorithm as discussed in the previous section. It has been used to find the optimum centre and spread for the RBFN network. Figure 7 shows the accuracy obtained from each model.

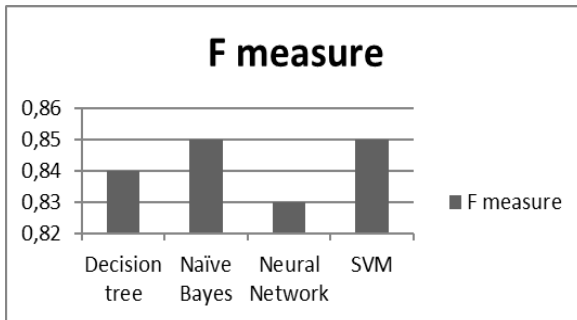


Figure 4. Depicts the effectiveness of type I experiments using the F measure

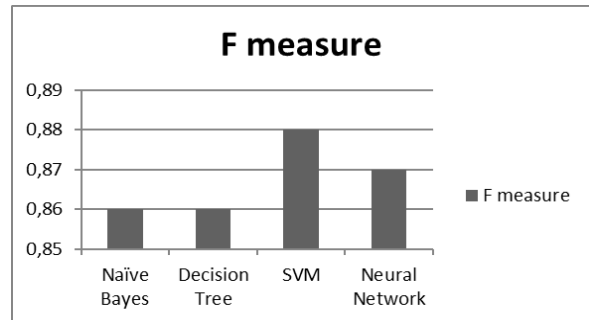


Figure 5. Depicts the effectiveness of type II experiments using the F measure

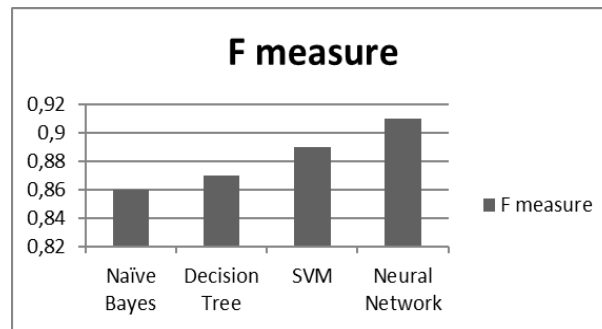


Figure 6. Depicts the effectiveness of type III experiments using the F measure

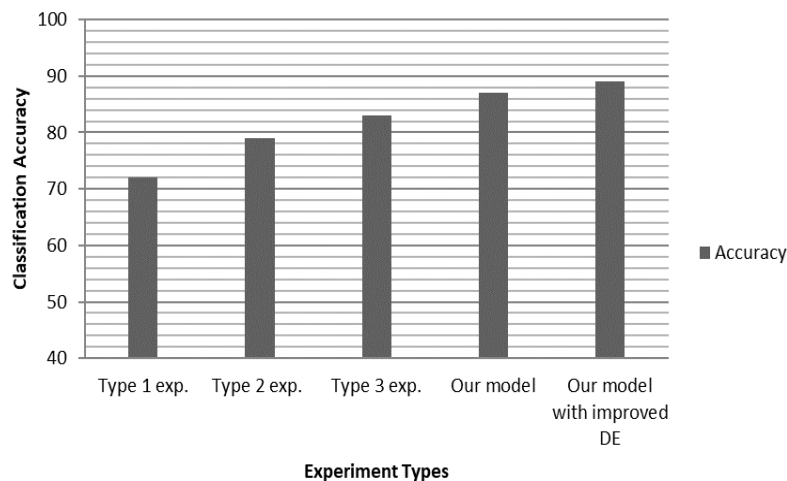


Figure 7. Accuracy of the EDM models

These results encouraged us to use and investigate feature selection algorithms. Several experiments were conducted with a subset of attributes returned from the feature selection algorithm. The subset of attributes was used as input to our model with modified DE. The accuracy of the classification model was measured and plotted with various subsets of attributes as input as depicted in Figure 8.

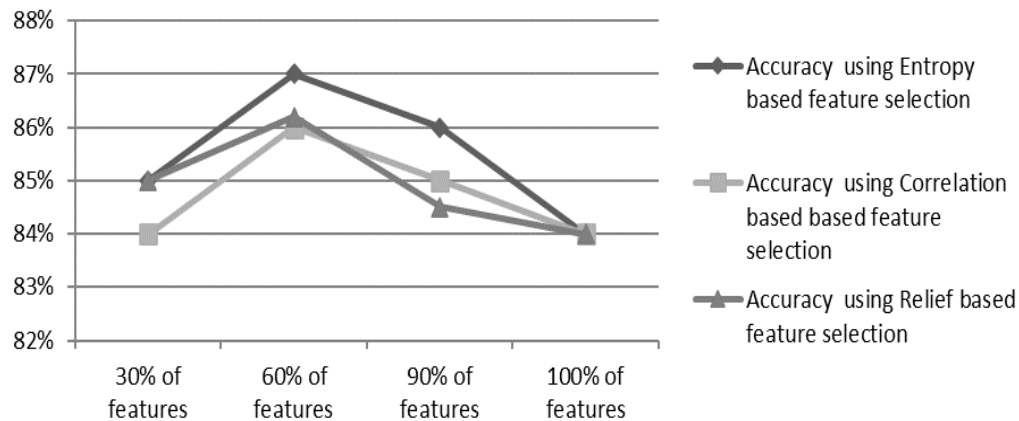


Figure 8. Depicting accuracy of the system vs percentage of features selected

The experimental results show that a subset features consisting of 60% of original features provide the best results for the classification model. An investigation on the ranking of the attributes returned by the feature selection algorithms was done. This helped to identify the attributes which highly influence academic performance. The ranking of the attributes by the feature selection algorithms are shown in Table 3. Two stage feature selection has been used in our study. In the first stage experiments were conducted with 30%, 60%, 90% and 100 % of features.

The accuracy of the resultant classification model was measured and plotted as shown in figure. It was found that the classification accuracy was highest with 60 % of the input attributes. Hence 60% of the features were used for the final classification model. The second stage of feature selection was used on individual categories/ classes. The experimental data consisted of three classes of students' academic performance: the outstanding students, the average performing students and the underperforming students. Applying filter-based feature selection in the second stage on individual classes helped us to identify the attributes of individual categories of students as shown in table and take appropriate measure. This was achieved by segregating the data instances of three categories of data. Further a ranking of attributes was done using the filter-based feature selection algorithm. The results obtained are shown in Table 4. These significant results obtained helped us to devise different approaches for students based on their academic performance. It was found that the outstanding students were self motivated, had high attendance percentage and were engaged in regular study. Most of the average students were found to have some internet addiction. They spent a lot of time with friends and mostly studied only before exams. These students accepted that a more sincerity on their part could certainly improve their grades. Once these students are identified using the classification model, they can be motivated to study regularly and to reduce the time spent on internet. The underperforming students showed a variety of attributes as a key factor to their academic performance. Lack of interest in academics, internet addiction, financial distress and family problems and very low attendance were some of the issues. Students with academic negligence were recommended remedial classes. Students with family problems and indifference were suggested counselling sessions. It was observed that internet usage has significantly increased among the student community resulting in deterioration of academic performance. Instances of internet usage were accepted by students within classrooms. This needs to be taken care by educational institutes and parents before internet addiction turns into an epidemic. Low cost of data and smartphone availability has even fuelled this addiction. Internet usage should be reduced by motivation and promoting participation in extracurricular activities

Previously all research work done so far for academic performance determination have followed a unified strategy of implementing classification and feature selection techniques on the entire dataset. The features selected in the process were used to predict results of different categories of students. This method cannot work efficiently as it treats all students equally. However, our novel framework suggests first segregation of instances of the academically weaker students, average students, the outstanding students and then application of feature selection on these segregated groups. This segregation of attributes helps to take different pedagogical interventions for different students.

Table 3. Ranking of the features by filter algorithms

Sl. No.(Rank)	Attribute	Information gain	Correlation	Relief
1	Daily Internet usage	4	6	8
2	Financial distress	5	3	2
3	Previous backlog	2	4	3
4	Participation in Extra-curricular activities	7	9	7
5	Health condition	6	5	6
6	Type of internet connectivity	10	11	9
7	10 th grade	8	7	5
8	Attendance	3	2	4
9	Study daily	12	8	10
10	Internal grade	1	1	1

Table 4. Top ten significant features of each class category

Sl. no. (Rank)	Outstanding students	Average students	Underperforming students
1	CGPA	CGPA	CGPA
2	Internal Grades	Internal Grades	Internal grades
3	Study daily	Attendance medium	Internet access time high
4	Attendance high	12 th grade medium	Attendance less
5	12 th grade high	Accept that sincerity could increase their grades	Finds most subject as complex
6	Finds most subject interesting	Internet usage high	Study only before exam
7	10 th grade high	4G connectivity	Internet addiction
8	Weekly participation in extra-curricular activities	Internet addiction sometimes	Health affecting education
9	Joined engineering by self choice	Time spent in hanging out with friends > 5hrs	12 th grade medium
10	Finds none of subject as complex	Study mostly before exam	Family relations not harmonious

4. CONCLUSION

In this paper the authors have developed a novel model to identify the academically weaker students, the average students and the outstanding students. A comparative study and analysis of the existing EDM models was done on our data set. In the proposed approach the authors have converted categorical attributes into dummy variables. Apart from use of classification model, our model proposed a two-staged feature selection technique. In the first stage of feature selection, filter-based feature selection methods were used to analyze the student attributes that affect their academic performance. In the second stage of feature selection, segregated instances of each class were used separately to identify the dominant attributes specific to that class. The findings were important and helped to comprehend academic performance pattern of each class category of students. The RBFN network model used for classification was optimized by a novel and improved differential evolution technique. Our proposed framework of student academic performance estimation could not only help to reduce the number of failures in university examination but also improve the academic performance of the average students.

REFERENCES

- [1] C. Romero and S. Ventura, "Educational Data Mining: A Review of the State of the Art," in *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 6, pp. 601-618, 2010.
- [2] Al-Otaibi & L. Houghton, "Examining the Theoretical Factors that Influence University Students to Adopt Web 2.0 Technologies: The Australian Perspective," *International Journal of Information and Communication Technology Education (IJICTE)*, vol. 11, no. 1, pp. 1-26, 2015.
- [3] R. Asif, A. Merceron, S. A. Ali, N.G. Haider, "Analyzing undergraduate students' performance using educational data mining," *Computers & Education*, vol. 113, pp. 177-194, 2017.
- [4] A.P. Ayala, "Educational data Mining: a survey and a data mining-based analysis of recent works," *Expert Systems with Applications*, vol. 41, no. 4, pp. 1432-1446, 2014.
- [5] D. D. Essel, & O.A. Wilson, "Factors Affecting University Students' Use of Moodle: An Empirical Study Based on TAM," *International Journal of Information and Communication Technology Education (IJICTE)*, vol. 13, no. 1, pp. 14-26, 2017.
- [6] G. Gray, et. al., " An Application of Classification models to predict learner progression in tertiary education," *Advance Computing Conference (IACC), IEEE International*, pp. 549-554, 2014.

- [7] M. Khanbabaei, F.M. Sobhani, M. Alborzi, R. Radfar, "Developing an integrated framework for using data mining techniques and ontology concepts for process improvement," *Journal of Systems and Software*, vol. 137, pp. 78-95, 2018.
- [8] J.P. Meriac, "Work ethic and academic performance: Predicting citizenship and counterproductive behaviour," *Learning and Individual Differences*, vol. 22, no. 4, pp. 549-553, 2012.
- [9] W. Xing, R. Guo, E. Petakovic, & S. Goggins, "Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory," *Computers in Human Behavior*, vol. 47, pp. 168-181, 2015.
- [10] B.E. Costa, B. Fonseca, M.A. Santana, F. Araújo, J. Rego, "Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses," *Computers in Human Behaviour*, vol. 73, pp. 247-256, 2017.
- [11] X. Wu, V. Kumar, J. R. Quinlan, J., "Top 10 algorithms in data mining," *Knowledge and information systems*, vol. 14, no. 1, pp. 1-37, 2008.
- [12] S. García, J. Luengo, F. Herrera, "Tutorial on practical tips of the most influential data preprocessing algorithms in data mining," *Knowledge-Based Systems*, vol. 98, pp.1-29, 2016.
- [13] J. Han, M. Kamber, & J. Pei, "Data Preprocessing Data Mining," (Third Edition) *The Morgan Kaufmann Series in Data Management Systems*, pp. 83-124.
- [14] P. Chaudhury & H.K. Tripathy, "An empirical study on attribute selection of student performance prediction model," *Int. J. Learning Technology*, vol. 12, no. 3, pp. 241- 252, 2017.
- [15] M. Marques, M. C. Nicoletti & M.P. Guimarães, "Using feature selection algorithms in educational data: Three case studies," *12th Iberian Conference on Information Systems and Technologies (CISTI)*, pp. 1-6, 2017.
- [16] M. Ramaswami & R. Bhaskaran, "A Study on Feature Selection Techniques in Educational Data Mining," *Journal of Computing*, vol. 1, no. 1, pp. 7-11, 2009.
- [17] P. Cortez, and A. Silva, "Using Data Mining to Predict Secondary School Student Performance," in *Proceedings of 5th Future Business Technology Conference (FUBUTECH 2008)*, pp. 5-12, 2008
- [18] N. Naveen, V. Ravi, C.R. Rao & N. Chauhan, "Differential evolution trained radial basis function network: application to bankruptcy prediction in banks," *International Journal of Bio-Inspired Computation*, vol. 2, no. (3-4), pp. 222-232, 2010.
- [19] C. Cortes & V. Vapnik. "Support-vector networks," *Machine Learning*, vol. 20, pp. 273-297, 1995.
- [20] P. Domingos & M. Pazzani, "On the optimality of the simple bayesian classifier under zero-one loss," *Machine Learning*, vol. 29, no (2-3), pp. 103-130, 1997.
- [21] M. Dash & H. Liu, "Feature Selection for Classification," *Intelligent Data Analysis*, vol.1, no.1, pp. 131-156, 1997.
- [22] N.V. Chawla, K.W. Bowyer, L.O. Hall, and W.P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321-357, 2002.
- [23] X. Y. Yang, "Differential Evolution, In Nature-Inspired Optimization Algorithms," *Elsevier*, pp. 89-97, 2014.
- [24] R. Storn and K. Price, "Differential Evolution-A Simple and Efficient Adaptive Scheme for Global Optimization over Continuous Spaces: Technical Report TR-05-012," *International Computer Science Institute*, Berkely, 1995.
- [25] V. Kotu & B. Deshpande, "Predictive analytics and data mining: concepts and practice with rapidminer," *Morgan Kaufmann*, Boston, pp.347-370, 2015.
- [26] M. Zaffar, et.al., "Performance analysis of feature selection algorithm for educational data mining," *IEEE Conference on Big Data and Analytics (ICBDA)*, Kuching, Malaysia pp. 7-12, 2017.
- [27] P. Chaudhury & H. K. Tripathy, "Optimising the parameters of a RBFN network for a teaching learning paradigm," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 15, no. 1, pp. 435-442, 2019.
- [28] P. Somol, B. Baesens, P. Pudil, J. Vanthienen, "Filter-versus Wrapper-based Feature Selection For Credit Scoring," *International Journal of Intelligent Systems*, vol. 20, no. 10, pp. 985-999, 2005.
- [29] A. Acharya, D. Sinha, "Application of Feature Selection Methods in Educational Data Mining," *International Journal of Computer Applications*, vol. 103, no. 2, pp. 34-38, 2014.
- [30] M. Ramaswami and R. Bhaskaran, "A Study of feature selection techniques in Educational Data Mining," *Journal of computing*, vol. 1, no. 1, pp.7-11, 2009.