IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-8, Aug 2018 E-ISSN: 2347-2693

Exploring The High Potential Factors That Affects Students' Academic Performance

R. Kaviyarasi [1]*, T. Balasubramanian[2]

[1,2]Department of Computer Science, Periyar University, Salem-636011, Tamilnadu, India

Corresponding author. Tel: +91 9445829228. E-mail address:arasikavi@gmail.com

Available online at: www.ijcseonline.org

Accepted: 13/Aug/2018, Published: 31/Aug/2018

Abstract— The rapid increase in student population has resulted in expansion of educational facilities at all level. Nowadays, the responsibilities of teachers are many. It is the duty of teachers to guide the students to choose their carrier field according to their abilities and aptitudes. The Data Mining field mines the educational data from large volumes of data to improve the quality of educational processes. Today's need of educational system is to develop the individual to enhance problem solving and decision making skills in addition to build their social skills. Educational Data Mining is one of the applications of Data Mining to find out the hidden patterns and knowledge in Educational Institutions. Generally, the three important groups of students have been identified: Fast Learners, Average Learners, and Slow Learners. In fact, students are probably struggles in many factors. This work focuses on finding the high potential factors that affects the performance of college students. This finding will improve the students' academic performance positively.

Keywords— Educational Data Mining; Feature Selection; Ensemble methods; ExtraTree Classifier

I. INTRODUCTION

In earlier education systems, the responsibilities of educators were limited only for teaching the lessons in classrooms to expand the knowledge of students. But today, the teachers' contribution should be improved in all over manners such as to achieve optimum development of their abilities and harmonious personality development. Hence it is the responsibilities of the academic institutions to provide proper guidance to the students' for choosing the right carrier according to their abilities and aptitudes, so that they can achieve success and obtain personal satisfaction in their life[15]. Many factors determine the level of academic performance of the students. Few are given below:-

[1]. Student abilities and their personal characteristics
[2]. Faculties abilities and their personal characteristics
[3]. Level of interaction between students and faculties
[4]. Infrastructural facilities available in the college
[5]. External environmental influences on the students'

Learners, Learning processes and Learning situations are three focal areas of education. The academic class is generally not homogeneous but heterogeneous. [12] There are Fast, Average and Slow Learners in the class. Students pass through various stages of physical development such as infancy, childhood and adolescence. These development stages have their own characteristics. If the prospective teacher knows these characteristics, then he/she can utilize the students in imparting instruction and moulding their behaviours according to the specified goal of education.

The learning ability is not the same in all. As learning depends on one's level of intelligence, interest and motivation, significantly the rate of learning differs from one to one. One may be fast learner and other takes more time to learn the same thing. The Rate of learning can be measured by the following formula:

$$\text{Rate of learning} = \frac{\text{Amount of learning proficiency achieved}}{\text{Time taken to achieve the amount of learning}} \quad (1)$$

Related studies have been carried out in this area. It identifies the poor performers and analyses the factors that affects the students' academic performance at schools, colleges and even at universities [16]. This proposed research aims to analyses what could be the reason behind the less academic performance of the students'.

The rest of the paper is organized as follows: Section II presents an Objective of the work. Section III highlights the Significance. Section IV explains the Related Work about the existing research work. Section V gives details about the Factors Affecting the Students' Performance. Section VI, VII and VIII explains the Feature Selection, Ensemble methods and Extra Tree Classifier in brief. Finally concluded the work under Conclusion Part.

II. OBJECTIVE

The main objective of this work is to explore the various factors affecting the academic performance of college students with a view to increase the individual performances and improvements in their academic level.

III. SIGNIFICANCE

Other than personal characteristics, many factor such as previous academic background, study habits, family background, self motivation, etc., affect the students' academic performance. Identifying the high potential factors can help the teachers and parents to make the students to increase their academic performance [1]. It can also create awareness to students about their responsibilities to achieve the higher studies and importance about education [13].

IV. RELATED WORK

Raified: Asif, Agathe Merceron, Syed Abbas Ali and Najmi Ghani Haider [8] used data mining methods to analyze the undergraduate students' performance. In their study, two important groups of students such as the low and high achieving students have been identified. Also their study has investigated three research questions with the aim of providing information to teachers and study programme directors that might help them to improve the educational programmes at their institutions.

Cristobal Romero, Manuel-Ignacio Lopez, Jose-Maria Luna and Sebastian Ventura [9] used several data mining approaches to improve prediction of students' final performance starting from student participation in an on-line discussion forum. With the proper format data, classification and classification via clustering techniques are applied and compared. Finally, the obtained classification models are described and compared to clustering models and additional mining association rules for each other.

Anne-Sophie, Hoffait, Michael Schyns [10] used data mining methods to present a new means of identifying freshmen's profiles likely to face major difficulties to complete their first academic year. Their study also designed algorithm to increase the accuracy of the prediction..

Ashwin Satyanarayana, Marinsz Nuckowski [11] used multiple classifiers (Decision Trees- J48, Naïve Bayes and Random Forest) to improve the quality of students' data by eliminating noisy instances and hence improving predictive accuracy. Also their paper identified association rules that influence students' outcome using a combination of rule based techniques (Apriori, Filtered Associator and Terius).

Pandey and Taruna [17] proposed the integrated multiple classifiers for the predictions of students' academic performance. A product of probability combining rule is employed to integrate the multiple classifiers that consists of three complementary algorithms, namely Decision Tree, K-Nearest Neighbour, and Aggregating One-Dependence Estimators (AODE). Their method has been applied and compared on three student performance datasets using t-test. Also this method is compared with KSTAR, OneR, ZeroR, Naive Bayes, and NB tree classifiers as well as with the individual classifiers.

R. Asif, A. Merceron and M. K. Pathan [19] used Data Mining Techniques to explore the possibility of students' performance prediction based on their academic data at an early stage of their degree program. In their study, two datasets were fed to the MLP network and other mining techniques such as Decision Tree, Rule Induction, K-Nearest Neighbour, and Naive Bayes. From this work, the result shows Naive Bayes performed best than other techniques. It also stated that students' degree performance prediction is possible without any socio-economic or demographic feature but with just their academic data (Pre-university marks and marks obtained in year 1 and year 2).

Abimbola R. Iyanda, Olufemi D. Ninan, Anuoluwapo O. Ajayi and Ogochukwu G. Anyabolu [18] compared two neural network models (Multilayer Perceptron and Generalized Regression Neural Network) with a view to identifying the best model for predicting students' academic performance based on single performance factor. In this study, only the academic factor (students' results) was considered as the single performance factor. The result obtained in this study shows that Generalized Regression Neural Network had a better accuracy although Multilayer Perceptron had prediction accuracy of 75%.

V. FACTORS AFFECTING THE STUDENTS' PERFORMANCE

The level of academic achievement of students in the classroom is determined by many factors. This is shown in Fig.1.

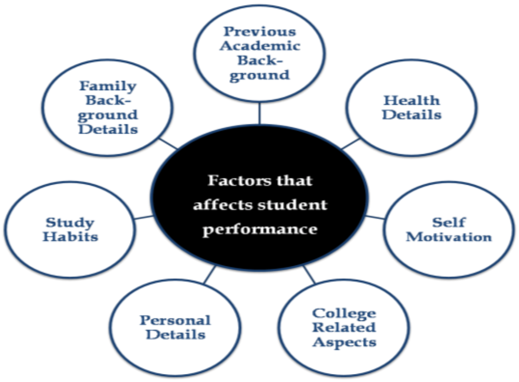
Previous Academic Back-ground
Family Back-ground Details
Health Details
Factors that affects student performance
Study Habits
Self Motivation
Personal Details
College Related Aspects

Fig.1. Factors affecting Students Performance

A total of forty five attributes are identified and listed in the dataset. These 45 attributes are specified in Table 1.

Table 1. Full list of Features in the Dataset

S. No.	**Attributes**
1.	Accommodation
2.	Taken Care By
3.	Living Location
4.	Parental Status
5.	Cohabitation Status
6.	Fathers Education
7.	Fathers Job
8.	Mothers Education
9.	Mothers Job
10.	Family size
11.	10th grade
12.	12th grade
13.	Medium
14.	School
15.	Secondary syllabus
16.	Group at Secondary
17.	Any Part Time
18.	Study Interest
19.	Reason to choose this college
20.	Travelling way
21.	Travel time
22.	Have mobile
23.	Student Using Mobile
24.	Computer/laptop at home
25.	Net access
26.	Social network id
27.	Study hours
28.	Past arrears
29.	Extra college support
30.	Extracurricular activities
31.	Extra paid classes
32.	Going outings
33.	Alcohol consumption
34.	Health status
35.	Any learning disabilities
36.	Place to study
37.	Guidance
38.	Care at home
39.	Interest in course
40.	Attention in class
41.	Quality of study materials
42.	Attendance percentage

43.	Semester percentage now
44.	Internal test 1
45.	Internal test 2

VI. FEATURE SELECTION

Feature selection is the process of selecting a particular feature from a massive collection of features. It plays an important role in machine learning and data mining. The features that contribute high for predicting variable or output can be selected automatically through feature selection methods. Feature selection is also termed as variable selection or attributes selection or variable subset selection [20].

Traditional feature selection process consists of four basic steps namely, subset generation, subset evaluation, stopping criterion and validation. Subset generation is a search process that produces candidate feature subsets for evaluation based on certain search strategy. Each candidate feature subset is evaluated and compared with previous best one based on certain evaluation. If the new subset turns to be better, it replaces the best one and this process is repeated until a given stopping condition is satisfied [21].

Feature selection is important because it reduces the dimensionality of feature space, removes redundant or irrelevant, or noisy data to increase the prediction accuracy. Filter, Wrapper and Embedded are three methods of feature selection algorithms. Speeding up a data mining algorithm, improving the data quality, improving the performance of data mining and increasing the clarity of the mining results are the significance of feature selection methods. The key benefits of feature selections are [2, 3, 6]:-

[1]. Reduce Overfitting
[2]. Improves Accuracy
[3]. Reduce Training Time

VII. ENSEMBLE METHOD

Ensemble methods are used to create stronger (i.e., more accurate) classification tree models and this can be done by combining weak classification tree models to create stronger versions. Ensemble method is a learning method that combines multiple models into one and it performs better than the standard methods. Ensembles are useful with all modeling algorithms. Ensemble Data Mining Methods is also termed as Committee Methods or Model Combiners [4]. An ensemble classifier detects noisy instances by constructing set of classifiers [11]. It increases the accuracy and reduces the variability of classification. Generally, ensemble methods improve the generalization performance of a set of classifiers in a domain.

The benefits of using an ensemble classification models are [14]:

a) the ensemble classifier is likely to have a lower error rate

b) the variance of the ensemble classifier will be lower than had we used certain unstable classification models, such as decision tress and neural networks, that have high variability

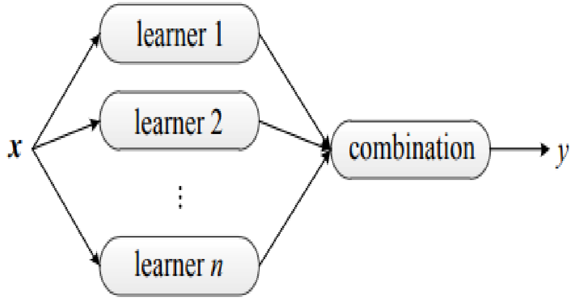

Fig.2. Ensemble Architecture

The three most popular methods for combining the predictions from different models are [5]:

[1]. *BAGGing, or Bootstrap AGGregating*. Building multiple models from different subsamples of the training dataset and uses each of them to generate a classifier for inclusion in the ensemble.

[2]. *Boosting*. Building multiple models each of which learns to fix the prediction errors of a prior model in the chain. The term boosting refers to a family of algorithms that are able to convert weak learners to strong learners [22].

[3]. *Voting*. Building multiple models and simple statistics are used to combine predictions.

The methods for constructing ensembles are

- By manipulating the training set
- By manipulating the input features
- By manipulating the class labels
- By manipulating the learning algorithm

The general procedure for ensemble method is given here [23].

Let D denote the original training data, k denote the number of base classifiers, and T be the test data.

```
for $i$ = 1 to $k$ do
Create training set $Di$ from $D$. Build a base classifier $Ci$ from D.
end for
for each test record $x \in T$ do
$C * x = Vote\ C1$ , $C2$ x , … , $Ck$ x
end for
```

VIII. EXTRA TREE CLASSIFIER

Extra-Tree method stands for **ext**remely **ra**ndomized **trees**. Extra Tree is a form of Bagging where Random trees are constructed from subsamples of the training dataset. Through ExtraTree Classifier, extra tree model is constructed [5]. Instead of computing the locally optimal feature or split combination based on information gain or the Gini impurity, for each feature under consideration, a random value is selected for the split. This value is selected from the feature's empirical range. Features that produce large values are ranked as more important than features which produce small values [23]. In our work, among forty five attributes, the high potential factors are identified and listed in the Table 2.

Table 2. Top 12 Features based on the Importance Values

Attributes	**Importance Value**
Internal test 2	0.2492
Internal test 1	0.1842
Guidance	0.0446
Have mobile	0.0400
Family size	0.0293
Extracurricular activities	0.0273
12th grade	0.0233
Alcohol consumption	0.0224
Attention in class	0.0209
Any LD	0.0189
Place to study	0.0189
Travel time	0.0180

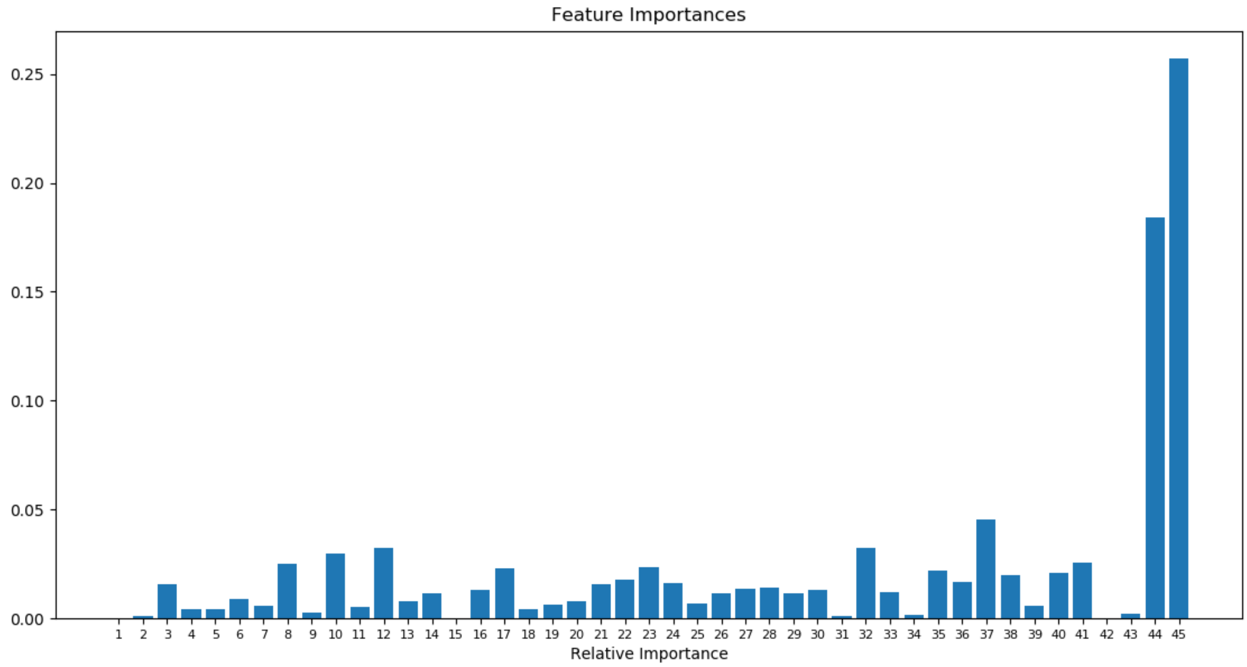

Fig.3. Feature Importance

Table 3. Main attributes with expected relation

Attributes	Expected Relation	Description
Guidance	Positive (Yes)	Guidance results in good performance
Mobile	Negative(Yes)	Usage of mobile reduces student involvement in studies
Family Size	Positive(<=5)	Family with limited members can take care the children
Extracurricular activities	Positive (Yes)	Increases study interest
12th grade	Positive(I group)	I group Students' performance will be good
Alcohol consumption	Negative(Yes)	Consuming Alcohol reduces study interest
Attention in class	Positive (Yes)	Attention in Class increases study interest
Any ld	Negative(Yes)	Learning Disability reduces study performance
Place to study	Positive (Yes)	Place to study increases study interest
Travel time	Negative(>1 Hr)	Makes tired and reduces study interest

IX. CONCLUSION

In this research work, the high potential factors that affect students' academic performance are identified. It focused on building the Extra Tree classifier that determines the feature importance. The future work of this research is to predict the performance of college Students with high accuracy using a new proposed model.

ACKNOWLEDGEMENT

The authors would like to acknowledge the support provided by the affiliated colleges under Periyar University- Salem for providing a valuable dataset of the institutions to carry out the research study.

REFERENCES

[1] Jayashree M Kudari. 2016. "Survey on the Factors Influences the Students' Academic Performance". International Journal of Emerging Research in Management &Technology. ISSN: 2278-9359 (Volume-5, Issue-6)

[2] Jason Brownlee. 2014. "Feature Selection in Python with Scikit-Learn". (July 2014). Retrieved March 21, 2018 from https://machinelearningmastery.com/feature-selection-in-python-with-scikit-learn/

[3] Jasmina NOVAKOVIĆ, Perica STRBAC, Dusan BULATOVIĆ . 2011. "Toward Optimal Feature Selection Using Ranking

[4] Methods And Classification Algorithms". Yugoslav Journal of Operations Research.21 (2011), Number 1, 119-135.DOI: 0.2298/YJOR1101119N

[5] Giovanni Seni, John F. Elder. 2010. "Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions". Morgan & Claypool publishers

[6] Jason Brownlee. 2016. "Ensemble Machine Learning Algorithms in Python with scikit-learn". (June 2016). Retrieved March 27. 2018 from https:// machinelearningmastery.com/ensemble-machine-learning-algorithms-python-scikit-learn/

[7] Sergio Ledesma, Gustavo Cerda, Gabriel Avina, Donato Hernandez, and Miguel Torre. "Feature Selection Using Artificial Neural Networks". Mexican International Conference on Artificial Intelligence. MICAI 2008: Advances in Artificial Intelligence pp 351-359 Retrieved from https://link.springer.com/chapter/10.1007/978-3-540-88636-5_34

[8] Zhi-Hua Zhou. 2012. CRC Press Taylor & Francis Group. "Ensemble Methods Foundations and Algorithms".

[9] Raheela Asif, Agathe Merceron, Syed Abbas Ali and Najmi Ghani Haider. "Analyzing undergraduate students' performance using educational data mining". Computers & Education. 113 (2017) 177-194.

[10] Cristobal Romero, Manuel – Ignacio Lopez, Jose- Maria Luna and Sebastian Ventura. "Predicting students' final Perfromance from participation in on-line discussion forums". Computers & Education 68(2013) 458-472

[11] Anne – Sophie Hoffait, Michael Schyns. "Early detection of university students with potential difficulties". Decision Support Systems 101(2017) 1-11.

[12] Ashwin Satyanarayana, Mariusz Nuckowski. "Data Mining using Ensemble Classifiers for Improved Prediction of student Academic Performance". Spring 2016 Mid- Atlantic ASEE Conference, April 8-9, 2016 GWU.

[13] Mona Zamani. "Cooperative learning: Homogeneous and heterogeneous grouping of Iranian EFL learners in a writing context". Cognet Education. Volume 3, 2016- Issue 1.Retrived from https://doi.org/10.1080/2331186X.2016.1149959

[14] Irfan Mushtaq, Shabana Nawaz Khan. "Factors Affecting Students' Academic Performance". Global Journal of Management and Business Research. Volume 12 Issue 9 Version 1.0 June 2012. Online ISSN: 2249-4588 & Print ISSN: 0975-5853

[15] Daniel T. Larose and Chantal D. Larose. 2016. "Data Mining and Predictive Analytics". Wiley India Pvt. Ltd. New Delhi.

[16] K.Nagarajan, and S. Natarajan. 2012. "Guidance and Counseling". Chennai. Ram publishers.

[17] Shoukat Ali, Zubair Haider, Fahad munir, Hamid Khan, Awais Ahmed. "Factors Contributing to the Students' Academic Performance: A case study of Islamia University Sub- Campus". American Journal of Education Research. 2013; 1(8): 283-289.

[18] Mrinal Pandey, S. Taruna. “Towards the integration of multiple classifier pertaining to the Student’s performance prediction”. Perspectives in Science (2016) 8, 364—366.

[19] Abimbola R. Iyanda, Olufemi D. Ninan, Anuoluwapo O. Ajayi, Ogochukwu G. Anyabolu, “Predicting Student Academic Performance in Computer Science Courses: A Comparison of Neural Network Models", International Journal of Modern Education and Computer Science(IJMECS), Vol.10, No.6, pp. 1-9, 2018.DOI: 10.5815/ijmecs.2018.06.01

[20] R. Asif, A. Merceron and M. K. Pathan, “Predicting Student Academic Performance at Degree Level: A Case Study”. International Journal of Intelligent Systems and Applications, 7(1):49. 2014.

[21] Aparna U.R. and Shaiju Paul, “Feature selection and extraction in data mining”. 2016 Online International Conference on Green Engineering and Technologies (IC-GET). DOI: 10.1109/GET.2016.7916845

[22] M. Dash and H.liu, “Feature Selection for Classification”. An International Journal of Intelligent Data Analysis, vol. 1, no. 3, pp. 131-156, 1997

[23] Zhou Zhi-Hua (2012). “Ensemble Methods: Foundations and Algorithms”. Chapman and Hall/CRC. p. 23. ISBN 978-1439830031.

[24] Geurts P, Ernst D, Wehenkel L (2006). "Extremely randomized trees" (PDF). Machine Learning. 63: 3–42. doi:10.1007/s10994-006-6226-1

Authors’ Profiles

R. Kaviyarasi (Corresponding author) received her MCA Degree from Karpagam College of Engineering, Coimbatore affiliated with Anna University, Chennai and M. Phil Degree with university rank from Periyar University, Salem. She has qualified in NET & SET eligibility examinations. Now, she is pursuing her Ph. D (Part time) in Periyar University, Salem. Currently she is working as a Assistant Professor in Department of Computer Science & Applications at Sri Vidya Mandir Arts and Science College, Uthangarai, Krishnagiri(Dt), Tamilnadu. Her research area is Machine Learning Techniques.

T. Balasubramanian received his Master Degree in Computer Science from Bharathidasan University, Trichy in 1996. He has received his Master of Philosophy degree from Periyar University Salem in year 2007. In the specialization of Data Mining, he received his Doctorate from Bharathiar University in 2015. He published more than 32 research papers in various International and National Journals. He also presented 8 research articles in various International Conferences and 12 papers in National and State level conferences & seminars. He has written 5 books under various domains of computer science. Currently he is working as a Associate Professor in the PG and Research Department of of Computer Science in Sri Vidya Mandir Arts and Science College, Uthangarai, Krishnagiri (Dt). He has more than 17 Years of teaching experience. His domain interest is Data mining, Big data, Network Security.