

The Third Information Systems International Conference

A Review on Predicting Student's Performance using Data Mining Techniques

Amirah Mohamed Shahiri^{a,*}, Wahidah Husain^a, Nur'aini Abdul Rashid^a,

^a*School of Computer Sciences Universiti Sains Malaysia 11800 USM, Penang, Malaysia*

Abstract

Predicting students performance becomes more challenging due to the large volume of data in educational databases. Currently in Malaysia, the lack of existing system to analyze and monitor the student progress and performance is not being addressed. There are two main reasons of why this is happening. First, the study on existing prediction methods is still insufficient to identify the most suitable methods for predicting the performance of students in Malaysian institutions. Second is due to the lack of investigations on the factors affecting students achievements in particular courses within Malaysian context. Therefore, a systematical literature review on predicting student performance by using data mining techniques is proposed to improve students achievements. The main objective of this paper is to provide an overview on the data mining techniques that have been used to predict students performance. This paper also focuses on how the prediction algorithm can be used to identify the most important attributes in a students data. We could actually improve students achievement and success more effectively in an efficient way using educational data mining techniques. It could bring the benefits and impacts to students, educators and academic institutions.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of organizing committee of Information Systems International Conference (ISICO2015)

Keywords: Student performance, educational data mining, performance prediction

1. Introduction

Students performance is an essential part in higher learning institutions. This is because one of the criteria for a high quality university is based on its excellent record of academic achievements [1]. There are a lot of definitions on students performance based on the previous literature. Usamah et al. (2013) stated that students performance can be obtained by measuring the learning assessment and co-curriculum [2]. However, most of the studies mentioned about graduation being the measure of students success.

Generally, most of higher learning institutions in Malaysia used the final grades to evaluate students performance. Final grades are based on course structure, assessment mark, final exam score and also extracurricular activities [2]. The evaluation is important to maintain students performances and the effectiveness of learning process. By analyzing students performance, a strategic program can be well planned during their period of studies in an institution [3].

Currently, there are many techniques being proposed to evaluate students performance. Data mining is one of the most popular techniques to analyze students performance. Data mining has been widely applied in educational

*Corresponding author

Email address: amirahshahiri@gmail.com (Amirah Mohamed Shahiri)

area recently [4]. It is called educational data mining. Educational data mining is a process used to extract useful information and patterns from a huge educational database [5]. The useful information and patterns can be used in predicting students performance. As a result, it would assist the educators in providing an effective teaching approach.

Besides, educators could also monitor their students achievements. Students could improve their learning activities, allowing the administration to improve the systems performance. Thus, the application of data mining techniques can be focused on specific needs with different entities. In order to encounter the problems, a systematically review is proposed. The proposed systematically review is to support the objectives of this study, which are:

1. To study and identify the gaps in existing prediction methods.
2. To study and identify the variables used in analyzing students performance.
3. To study the existing prediction methods for predicting students performance.

The next section will focused on the methodology of survey in predicting students performance. Then, a discussion on research questions will be described in Section 3. In Section 4, the detail results on the existing prediction methods are discussed. Lastly, the conclusion and future work are outlined in Section 5.

2. Methodology

The reasoning for performing systematic relational review is to find suitable methods for existing parameter, to fulfill the gaps in existing research and to place a new research activity in the suitable context [6]. The aim for systematic review in the current literature is to support the proposed research questions. Next, subsection will be identifying the research questions to guide the results. This is also useful to identify the scope of the study.

2.1. Research Questions

Right research questions are important to understand the existing study of predicting students performance. By following the Kitchenhams steps in structuring the research questions, which consists of Population, Intervention, Outcome and Context (PIOC) [6][7], the criteria of research question are shown in Table 1.

Table 1 Research question criteria

Criteria	Details
Population	University (student performance)
Intervention	Methods/ techniques for prediction
Outcome	Prediction accuracy, successful prediction techniques
Context	Academic institutions. All types of empirical studies such as preliminary studies, questionnaires, experiments, and case studies.

Therefore, the research questions proposed in this study are:

- Q1: What are the important attributes used in predicting students performance?
- Q2: What are the prediction methods used for students performance?

However, it is better to start with a pilot study before going into the depth of this study. The purpose of doing the pilot study is to investigate the appropriateness of the research questions with the objectives of this study. Next, the study will explain the search strategy for conducting pilot study.

2.2. Search Strategy

In a systematic review, a well-planned search strategy is very important so that every relevant piece of work can be found in the search results. Therefore, an extensive search for research papers was conducted to try answering the proposed research questions. The search terms used in this systematic review were developed using the Kitchenham et al. (2010) steps [7]. The resulting search strings are as follows: (student performance) AND (systems OR application OR method OR process OR system OR technique OR methodology OR procedure) AND (educational data mining) AND

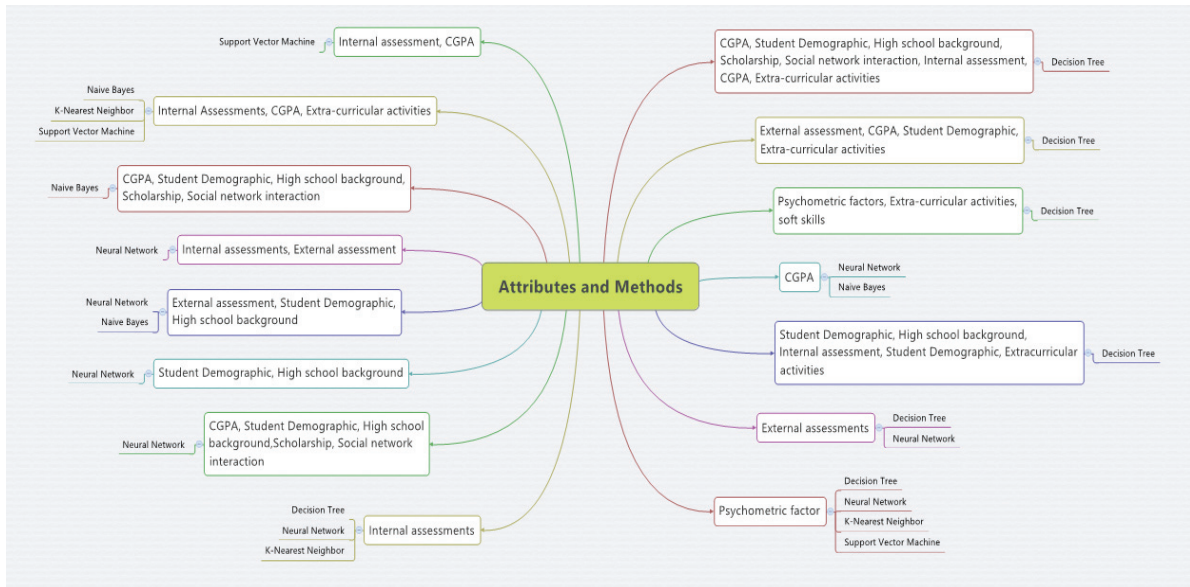


Fig. 1 List of common attributes and methods used in predicting student's performance

(prediction OR estimation OR assessment). The search strategy contains the following decisions: Searched databases: IEEE Xplore, Springer Link, Science Direct, ACM digital Library. Search items: Journal articles, workshops papers and conference papers. Search applied on: Full text - to avoid exclusion of papers that do not include the keywords in the title or abstract, but are still relevant to the review. Publication period: Since 2002. This search was limited until the early 2015 since that is the time this literature performed. Hence, any paper published after January 2015 is not included.

3. Important Factors on Predicting Student's Performance

This section will intensively discuss the important factors on predicting students performance. There are two main factors in predicting students performances, which are attributes and prediction methods. Fig 1. shows a graphical representation for list of common attributes and list of methods used in predicting student's performance. First step will be focused on the important attributes used in predicting student performance and second step will be focused on the prediction methods used in predicting students performance.

3.1. The important attributes used in predicting student's performance

The systematical literature review is used to identify the important attributes in predicting students performance. The attributes that have been frequently used is cumulative grade point average (CGPA) and internal assessment. Ten of thirty papers have used CGPA as their main attributes to predict students performance [5, 8, 9, 10, 3, 11, 12, 13, 14, 15, 16]. The main idea of why most of the researchers are using CGPA is because it has a tangible value for future educational and career mobility. It can also be considered as an indication of realized academic potential [2]. Through the coefficient correlation analysis, the result shows that CGPA is the most significant input variable by 0.87 compared to other variables [3]. Besides, in Christian and Ayub study [14], CGPA is the most influence attributes in determining the survival of students in their study, whether they can complete their study or not. In this study, internal assessment was classified as assignment mark, quizzes, lab work, class test and attendance. All attributes will be grouped in one attribute called internal assessment. The attributes are mostly used among the researchers to predict students performance [5, 17, 18, 19, 20, 21, 10, 22, 23, 12].

Next, the most often attribute being used is students demographic and external assessments. Students demographic includes gender, age, family background, and disability [2, 18, 9, 3, 24, 11, 25, 13, 14]. While external assessments

is identified as a mark obtained in final exam for a particular subject [5, 17, 19, 26, 27, 24, 28, 13, 29]. The reason of why most of the researchers used students demographic such as gender is because they have different styles of female and male students in their learning process [2]. Study done by Meit et al. (2007) found that most of female students have various positive learning styles and behaviors compared to male students [30]. Female students are more discipline and dutiful in their studies, self-directed, always preserved and focused. In other side, female students have an effective learning strategies in their study [31]. They have self-motivation, organization and rehearsal that were effectively used by them. Thus, it is proven that gender is one of important attributes influencing students performance.

The three attributes mostly used in predicting students performance are extra-curricular activities [5, 18, 12, 13, 32], high school background [9, 24, 11, 25] and social interaction network [9, 33, 26, 21, 34]. There are five out of thirty studies that used each one of these attributes.

There are also several researchers in another study who have used psychometric factor to predict students performance [35, 36, 32, 37]. A psychometric factor is identified as student interest, study behavior, engage time, and family support. They have used this attributes to make a system to look very clear, simple and user friendly. It helps the lecturer to evaluate students achievement based on their personal interest and behavior [12]. However, these attributes are rarely to apply in predicting students performance by several researchers because it focuses more on qualitative data and it is also hard to get a valid data from respondents.

3.2. The prediction methods used for student performance

In educational data mining method, predictive modeling is usually used in predicting student performance. In order to build the predictive modeling, there are several tasks used, which are classification, regression and categorization. The most popular task to predict students performance is classification. There are several algorithms under classification task that have been applied to predict students performance. Among the algorithms used are Decision tree, Artificial Neural Networks, Naive Bayes, K-Nearest Neighbor and Support Vector Machine. Next, the specific application of data mining techniques grouped by algorithms in predicting student performance will be described in the next section.

3.2.1. Decision Tree

Decision Tree is one of a popular technique for prediction. Most of researchers have used this technique because of its simplicity and comprehensibility to uncover small or large data structure and predict the value [8, 9, 13]. Romero et al. (2008) said that the decision tree models are easily understood because of their reasoning process and can be directly converted into set of IF-THEN rules [22]. As shown in Table 2, there are approximately ten (10) papers that have used Decision Tree as their method to evaluate students performance. Examples of previous studies using Decision Tree method are predicting drop out features of students data for academic performance [8], predicting third semester performance of MCA students [32] and also predicting the suitable career for a student through their behavioral patterns [18]. The students performance evaluation is based on features extracted from logged data in an education web-based system. The examples of dataset are students final grades [23], final cumulative grade point average (CGPA) [3] and marks obtained in particular courses [22]. All this datasets were studied and analyzed to find out the main attributes or factors that may affects the students performance [28, 13]. Then, the suitable data mining algorithm will be investigated to predict students performance [25]. Mayilvaganan and Kapalnadevi (2014), have compared the classification techniques for predicting students performance in their study [12]. Meanwhile, Gray et al. (2014) investigated the accuracy of classification models to predict learners progression in tertiary education [36].

3.2.2. Neural Network

Neural network is another popular technique used in educational data mining. The advantage of neural network is that it has the ability to detect all possible interactions between predictors variables [36]. Neural network could also do a complete detection without having any doubt even in complex nonlinear relationship between dependent and independent variables [29]. Therefore, neural network technique is selected as one of the best prediction method. Through the meta-analysis study, eight (8) papers have been published using Neural Network method. The papers present an Artificial Neural Network model to predict students performance [38] [29]. The attributes analyzed by Neural Network are admission data [24], students attitude towards self-regulated learning and academic performance [19]. The rest are same papers in addition with Decision Tree method where researchers have used both techniques

Table 2 Result accuracy using Decision Tree method

Methods	Attributes	Results	Authors
Decision Tree	Internal assessments	76%	Romero et al. (2008) [22]
	Psychometric factors	65%	Gray et al. (2014)[36]
	External assessment	85%	Bunkar et al. (2012)[28]
	CGPA	91%	Jishan et al. (2015)[16]
	CGPA, Student Demographic, High school background, Scholarship, Social network interaction	73%	Osmanbegovic and Suljic (2008) [9]
	Internal assessment, CGPA, Extra-curricular activities	66%	Mayilvaganan and Kapalnadevi (2014) [12]
	Student Demographic, High school background	65%	Ramesh et al. (2013) [25]
	Internal assessment, Student Demographic, Extra-curricular activities	90%	Elakia et al. (2014)[18]
	External assessment, CGPA, Student Demographic, Extra-curricular activities	90%	Natek and Zwilling (2014)[13]
	Psychometric factors, Extra-curricular activities, soft skills	88%	Mishra et al. (2014) [32]

to compare which one is the best prediction method for analyzing students performance. The results of prediction accuracy are summarized in Table 3.

Table 3 Result accuracy using Neural Network method

Methods	Attributes	Results (Accuracy)	Authors
Neural Network	Internal assessments	81%	Wang and Mirovic (2002) [38]
	Psychometric factors	69%	Gray et al. (2014)[36]
	External assessment	97%	Arsad et al. (2013)[29]
	CGPA	75%	Jishan et al. (2015)[16]
	CGPA, Student Demographic, High school background, Scholarship, Social network interaction	71%	Osmanbegovic and Suljic (2008) [9]
	Student Demographic, High school background	72%	Ramesh et al. (2013) [25]
	External assessment, Student Demographic, High school background	74%	Oladokun et al. (2008)[24]
	Internal assessments, External assessment	98%	Anupama and Vijayalakshmi (2012) [19]

3.2.3. Naive Bayes

Naive Bayes algorithm is also an option for researchers to make a prediction. Among thirty (30) papers, there are four (4) papers that have used Naive Bayes algorithms to estimate students performance. The objective of all these four (4) papers is to find the most effective prediction technique in predicting students performance by making comparisons [9, 12, 25, 16]. Their research showed that Naive Bayes has used all of attributes contained in the data. Then, it analyzed each one of them to show the importance and independency of each attributes [9]. The result is shown in Table 4.

Table 4 Result accuracy using Naive Bayes method

Methods	Attributes	Results	Authors
Naive Bayes	CGPA, Student Demographic, High school background, Scholarship, Social network interaction	76%	Osmanbegovic and Suljic (2008) [9]
	Student Demographic, High school background	50%	Ramesh et al. (2013) [25]
	CGPA	75%	Jishan et al. (2015)[16]
	Internal assessment, CGPA, Extra-curricular activities	73%	Mayilvaganan and Kapalnadevi (2014) [12]

3.2.4. K-Nearest Neighbor

As depicts in Table 5, all three papers studied in this research showed that K-Nearest Neighbor gave the best performance with the good accuracy. According to Bigdoli et al. (2003), K-Nearest Neighbor method had taken less time to identify the students performance as a slow learner, average learner, good learner and excellent learner [23, 12]. K-Nearest Neighbor gives a good accuracy in estimating the detailed pattern for learner's progression in tertiary education [36].

Table 5 Result accuracy using K-Nearest Neighbor method

Methods	Attributes	Results	Authors
K-Nearest Neighbor	Psychometric factors	69%	Gray et al. (2014) [36]
	Internal assessment, CGPA, Extra-curricular activities	83%	Mayilvaganan and Kapalnadevi (2014) [12]
	Internal assessment	82%	Bigdoli et al. (2003)[23]

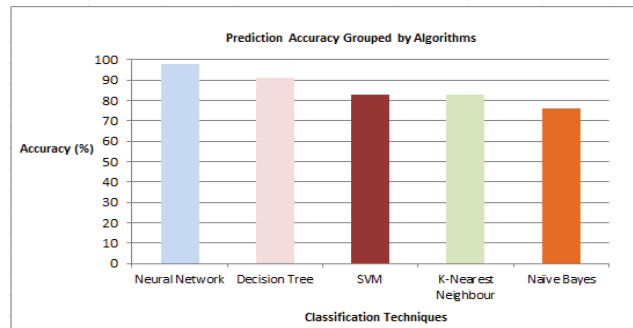


Fig. 2 Prediction accuracy grouped by algorithms since 2002-2015

3.2.5. Support Vector Machine

Support Vector Machine is a supervised learning method used for classification. There are three papers that have used Support Vector Machine as their method to predict students performance. Hamalainen et al. (2006) had chose Support Vector Machine as their prediction technique because it suited well in small datasets [10]. Sembiring et al. (2011) stated that Support Vector Machine has a good generalization ability and faster than other methods. [35]. Meanwhile, the study done by Gray et al (2014) demonstrated that Support Vector Machine method has acquired the highest prediction accuracy in identifying students at risk of failing [36]. Table 6 shows the result of prediction accuracy.

Table 6 Result accuracy using Support Vector Machine method

Methods	Attributes	Results	Authors
Support Vector Machine	Psychometric factors	83%	Sembiring et al. (2011) [35]
	Internal assessment, CGPA, Extra-curricular activities	80%	Mayilvaganan and Kapalnadevi (2014) [12]
	Internal assessment, CGPA	80%	Hamalainen et al. (2006) [10]

4. Discussion

This section will discuss the results analysis of the recent works in predicting students performance. This meta-analysis is based on the highest accuracy of prediction methods and also the main important factors that may influence the students performance. Fig. 2 shows the prediction accuracy that uses classification method grouped by algorithms for predicting students performance since 2002 to 2015.

By looking at the graph in Fig. 2, Neural Network has the highest prediction accuracy by (98%) followed by Decision Tree by (91%). Next, Support Vector Machine and K-Nearest Neighbor gave the same accuracy, which is (83%). Lastly, the method that has lower prediction accuracy is Naive Bayes by (76%).

The result on prediction accuracy is depending on the attributes or features that were used during the prediction process. Neural Network method gave the highest prediction accuracy because of the influence from main attributes. This attributes are hybridization of two features, which are internal and external assessments [5]. With the use of only one variable, which is external assessments, the accuracy is decreased by (1%) [29]. The third most used variable is internal assessments that gave the result of (81%) accuracy [38]. It shows that external assessment, which is the marks obtained in final examination, plays an important role in predicting students performance. While, significant variable

that gave the least impact on student performance is psychometric factors with the accuracy of only (69%) [36]. The psychometric factors usually use qualitative data, so it is difficult for Neural Network algorithm to make a prediction instead of using quantitative data. However, Neural Network method still has less maximum error prediction. The maximum error of prediction is less than (10%) [24]. Another advantage of Neural Network is the ability to capture nonlinear relationships easily. It is also referred as adaptive system due to its ability to readily update the historical data like a human brain. So, the model always functions beyond the knowledge base. In addition, the strength of neural network is the ability to learn from a limited set of data [11].

Second higher prediction accuracy is the Decision Tree method by (91%) of performance accuracy [16]. In Decision Tree method, the factor that leads to the highest accuracy in predicting students performance is CGPA. There are another two studies supporting this statement by which when they include CGPA as their main features, the result prediction was about (90%) of performance accuracy [18, 13]. It can be concluded that Decision Tree is able to handle both numerical and categorical data [12], perform well in large dataset [28] and easy to be understood and interpreted the relationship between variables [39, 32]. Besides that, the less important attribute in predicting student performance is psychometric factors [36] with the accuracy result of only (65%). It shows that Decision Tree is not suitable to predict students performance using psychometric variables.

Next is Support Vector Machine with the performance accuracy around (83%). Based on the analysis, psychometric factors are the most suitable attributes in predicting students performance with Support Vector Machine method [10]. However, the result decreased to (73%) performance accuracy when extra-curricular activities were included as another feature. Conversely, K-Nearest Neighbor showed high accuracy (83%) with the combination of three attributes, which are internal assessment, CGPA and extra-curricular activities in predicting students performance [12]. When comparing with the other two methods, which are Decision Tree and Nave Bayes the accuracy results are lower than K-Nearest Neighbor method [12]. Another study have also used extra-curricular activities as their attribute, however, they have combined it with another attributes, therefore the prediction accuracy is higher than just using one attribute [13][32].

Lastly, the method that has lowest prediction accuracy is Naive Bayes by (76%)[9]. The variables used are CGPA, student demographic, high school background, scholarship, social network interaction. All this attributes were also used in Neural Network method and Decision Tree method but the result showed that Naive Bayes gave highest accuracy compared to Neural Network and Decision tree. This is because the attributes used are significant with each other when using Naive Bayes as a prediction method.

5. Conclusion and Future Work

Predicting students performance is mostly useful to help the educators and learners improving their learning and teaching process. This paper has reviewed previous studies on predicting students performance with various analytical methods. Most of the researchers have used cumulative grade point average (CGPA) and internal assessment as data sets. While for prediction techniques, the classification method is frequently used in educational data mining area. Under the classification techniques, Neural Network and Decision Tree are the two methods highly used by the researchers for predicting students performance. In conclusion, the meta-analysis on predicting students performance has motivated us to carry out further research to be applied in our environment. It will help the educational system to monitor the students performance in a systematic way.

Acknowledgements

We would like to show our gratitude to Universiti Sains Malaysia (USM) for supporting this research.

References

- [1] M. of Education Malaysia, [National higher education strategic plan](http://www.moe.gov.my/v/pelan-pembangunan-pendidikan-malaysia-2013-2025) (2015). URL <http://www.moe.gov.my/v/pelan-pembangunan-pendidikan-malaysia-2013-2025>
- [2] U. bin Mat, N. Buniyamin, P. M. Arsad, R. Kassim, An overview of using academic analytics to predict and improve students' achievement: A proposed proactive intelligent intervention, in: Engineering Education (ICEED), 2013 IEEE 5th Conference on, IEEE, 2013, pp. 126–130.

- [3] Z. Ibrahim, D. Rusli, Predicting students academic performance: comparing artificial neural network, decision tree and linear regression, in: 21st Annual SAS Malaysia Forum, 5th September, 2007.
- [4] C. Romero, S. Ventura, *Educational data mining: A review of the state of the art*, Trans. Sys. Man Cyber Part C 40 (6) (2010) 601–618. doi:10.1109/TSMCC.2010.2053532. URL <http://dx.doi.org/10.1109/TSMCC.2010.2053532>
- [5] D. M. D. Angeline, Association rule generation for student performance analysis using apriori algorithm, The SIJ Transactions on Computer Science Engineering & its Applications (CSEA) 1 (1) (2013) p12–16.
- [6] M. W. Sukumar Letchuman, Mac Roper, Pragmatic cost estimation for web applications.
- [7] B. Kitchenham, R. Pretorius, D. Budgen, O. Pearl Brereton, M. Turner, M. Niazi, S. Linkman, *Systematic literature reviews in software engineering - a tertiary study*, Inf. Softw. Technol. 52 (8) (2010) 792–805. doi:10.1016/j.infsof.2010.03.006. URL <http://dx.doi.org/10.1016/j.infsof.2010.03.006>
- [8] M. M. Quadri, N. Kalyankar, Drop out feature of student data for academic performance using decision tree techniques, Global Journal of Computer Science and Technology 10 (2).
- [9] E. Osmanbegović, M. Suljić, Data mining approach for predicting student performance, Economic Review 10 (1).
- [10] W. Hämmäläinen, M. Vinni, Comparison of machine learning methods for intelligent tutoring systems, in: Intelligent Tutoring Systems, Springer, 2006, pp. 525–534.
- [11] M. M. A. Tair, A. M. El-Halees, Mining educational data to improve students performance: a case study, International Journal of Information 2 (2).
- [12] M. Mayilvaganan, D. Kalpanadevi, Comparison of classification techniques for predicting the performance of students academic environment, in: Communication and Network Technologies (ICCNT), 2014 International Conference on, IEEE, 2014, pp. 113–118.
- [13] S. Natek, M. Zwilling, Student data mining solution–knowledge management system related to higher education institutions, Expert systems with applications 41 (14) (2014) 6400–6407.
- [14] T. M. Christian, M. Ayub, Exploration of classification using nbtree for predicting students' performance, in: Data and Software Engineering (ICODSE), 2014 International Conference on, IEEE, 2014, pp. 1–6.
- [15] K. F. Li, D. Rusk, F. Song, Predicting student academic performance, in: Complex, Intelligent, and Software Intensive Systems (CISIS), 2013 Seventh International Conference on, IEEE, 2013, pp. 27–33.
- [16] S. T. Jishan, R. I. Rashu, N. Haque, R. M. Rahman, Improving accuracy of students final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique, Decision Analytics 2 (1) (2015) 1–25.
- [17] S. Parack, Z. Zahid, F. Merchant, Application of data mining in educational databases for predicting academic trends and patterns, in: Technology Enhanced Education (ICTEE), 2012 IEEE International Conference on, IEEE, 2012, pp. 1–4.
- [18] G. Elakia, N. J. Aarthi, Application of data mining in educational database for predicting behavioural patterns of the students, Elakia et al./IJCSIT International Journal of Computer Science and Information Technologies 5 (3) (2014) 4649–4652.
- [19] D. M. S. Anupama Kumar, Appraising the significance of self regulated learning in higher education using neural networks, International Journal of Engineering Research and Development Volume 1 (Issue 1) (2012) 09–15.
- [20] Z. K. Papamitsiou, V. Terzis, A. A. Economides, Temporal learning analytics for computer based testing, in: Proceedings of the Fourth International Conference on Learning Analytics And Knowledge, ACM, 2014, pp. 31–35.
- [21] B. K. P. Conrad Tucker, A. Divinsky, Mining student-generated textual data in moocs and quantifying their effects on student performance and learning outcomes, in: 2014 ASEE Annual Conference, Indianapolis, Indiana, Indianapolis, Indiana, 2014, <https://peer.asee.org/22840>.
- [22] C. Romero, S. Ventura, P. G. Espejo, C. Hervás, Data mining algorithms to classify students, in: Educational Data Mining 2008, 2008.
- [23] B. M. Bidgoli, D. Kashy, G. Kortemeyer, W. Punch, Predicting student performance: An application of data mining methods with the educational web-based system lon-capa, in: Proceedings of ASEE/IEEE frontiers in education conference, 2003.
- [24] V. Oladokun, A. Adebajo, O. Charles-Owaba, Predicting students academic performance using artificial neural network: A case study of an engineering course, The Pacific Journal of Science and Technology 9 (1) (2008) 72–79.
- [25] V. Ramesh, P. Parkavi, K. Ramar, Predicting student performance: a statistical and data mining approach, International Journal of Computer Applications 63 (8) (2013) 35–39.
- [26] A. Bogarín, C. Romero, R. Cerezo, M. Sánchez-Santillán, Clustering for improving educational process mining, in: Proceedings of the Fourth International Conference on Learning Analytics And Knowledge, ACM, 2014, pp. 11–15.
- [27] C. Coffrin, L. Corrin, P. de Barba, G. Kennedy, Visualizing patterns of student engagement and performance in moocs, in: Proceedings of the fourth international conference on learning analytics and knowledge, ACM, 2014, pp. 83–92.
- [28] K. Bunkar, U. K. Singh, B. Pandya, R. Bunkar, Data mining: Prediction for performance improvement of graduate students using classification, in: Wireless and Optical Communications Networks (WOCN), 2012 Ninth International Conference on, IEEE, 2012, pp. 1–5.
- [29] P. M. Arsal, N. Buniyamin, J.-I. A. Manan, A neural network students' performance prediction model (nnsppm), in: Smart Instrumentation, Measurement and Applications (ICSIMA), 2013 IEEE International Conference on, IEEE, 2013, pp. 1–5.
- [30] S. S. Meit, N. J. Borges, B. A. Cubic, H. R. Seibel, Personality differences in incoming male and female medical students., Online Submission.
- [31] A. Simsek, J. Balaban, Learning strategies of successful and unsuccessful university students., Online Submission 1 (1) (2010) 36–45.
- [32] T. Mishra, D. Kumar, S. Gupta, *Mining students' data for prediction performance*, in: Proceedings of the 2014 Fourth International Conference on Advanced Computing & Communication Technologies, ACCT '14, IEEE Computer Society, Washington, DC, USA, 2014, pp. 255–262. doi:10.1109/ACCT.2014.105. URL <http://dx.doi.org/10.1109/ACCT.2014.105>
- [33] C. Romero, M.-I. López, J.-M. Luna, S. Ventura, Predicting students' final performance from participation in on-line discussion forums, Computers & Education 68 (2013) 458–472.
- [34] N. Thai-Nghe, T. Horváth, L. Schmidt-Thieme, Factorization models for forecasting student performance, in: Educational Data Mining 2011, 2010.
- [35] S. Sembiring, M. Zarlis, D. Hartama, S. Ramliana, E. Wani, Prediction of student academic performance by an application of data mining techniques, in: International Conference on Management and Artificial Intelligence IPEDR, Vol. 6, 2011, pp. 110–114.

- [36] G. Gray, C. McGuinness, P. Owende, An application of classification models to predict learner progression in tertiary education, in: *Advance Computing Conference (IACC)*, 2014 IEEE International, IEEE, 2014, pp. 549–554.
- [37] I. Hidayah, A. E. Permanasari, N. Ratwastuti, Student classification for academic performance prediction using neuro fuzzy in a conventional classroom, in: *Information Technology and Electrical Engineering (ICITEE)*, 2013 International Conference on, IEEE, 2013, pp. 221–225.
- [38] T. Wang, A. Mitrovic, Using neural networks to predict student's performance, in: *Computers in Education*, 2002. *Proceedings. International Conference on*, IEEE, 2002, pp. 969–973.
- [39] M. Wook, Y. H. Yahaya, N. Wahab, M. R. M. Isa, N. F. Awang, H. Y. Seong, Predicting ndum student's academic performance using data mining techniques, in: *Computer and Electrical Engineering*, 2009. *ICCEE'09. Second International Conference on*, Vol. 2, IEEE, 2009, pp. 357–361.