Bulletin of Electrical Engineering and Informatics
Vol. 8, No. 4, December 2019, pp. 1556~1562
ISSN: 2302-9285, DOI: 10.11591/eei.v8i4.1432

Adaboost-multilayer perceptron to predict the student's performance in software engineering

Ahmad Firdaus Zainal Abidin[1], Mohd Faaizie Darmawan[2], Mohd Zamri Osman[3], Shahid Anwar[4], Shahreen Kasim[5], Arda Yunianta[6], Tole Sutikno[7]
[1,2,3,4]Faculty of Computer Systems & Software Engineering, Universiti Malaysia Pahang, 26300 Gambang, Pahang, Malaysia
[5]Faculty of Computer Science & Information Technology, Universiti Tun Hussein Onn Malaysia, Parit Raja, Johor, Malaysia
[6]Faculty of Computing and Information Technology in Rabigh, King Abdulaziz University, Jeddah, Saudi Arabia
[7]Faculty of Industrial Technology, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

Article Info

Article history:

Received Jun 19, 2019
Revised Sep 2, 2019
Accepted Oct 10, 2019

Keywords:

Adaboost
Education
Machine learning
Multilayer perceptron
Software engineering

ABSTRACT

Software Engineering (SE) course is one of the backbones of today's computer technology sophistication. Effective theoretical and practical learning of this course is essential to computer students. However, there are many students fail in this course. There are many aspects that influence a student's performance. Currently, student performance analysis methods just focus on historical achievement and assessment methods given in the class. Need more research to predict student's performance to overcome the problem of student failing. The objective of this research is to perform a prediction for student's performance in the SE using enhanced Multilayer Perceptron (MLP) machine learning classification with Adaboost. This research also investigates the requirements of each student before registering in this course. This research achieved 87.76 percent accuracy in classifying the performance of SE students.

Corresponding Author:

Ahmad Firdaus Zainal Abidin,
Faculty of Computer Systems & Software Engineering,
Universiti Malaysia Pahang, 26300 Gambang, Pahang, Malaysia.
Email: firdausza@ump.edu.my

1. INTRODUCTION

Software engineering (SE) is one of the important courses to address real-life problem solving and improve the poor quality of software [1-3]. It provides great solution to decrease the complexity of the project, manage project time and budget, and to ensure that the project developed systematically, measurable, and within specification [2]. Software engineering helps to decrease the time to human daily tasks by providing the user to press single button only to finish multiple jobs at one time. Furthermore, all the IT areas including graphics, networking and computer maintenance need software to assists the user in completing jobs [4-6].

However, many students in SE still unable to graduate successfully and fail to meet the industry demands [7]. Many aspects influence this student's performance such as learning resources, learning environment, teaching and learning process in the classroom and academic background [8]. The assessment and qualification aspects of student performance in SE can be different in every country depending on the education system applied in that country. Consequently, there is a need to conduct research to predict the performance of the students (pass or fail) before they register in SE course. One of the techniques to assess student's performance through the e-learning system is using meaningful learning concept [9]. However, this research is focused on the prediction of student performance using Adaboost and Multilayer Perceptron on machine learning implementation [10-13].

The aim of this research is to propose Adaboost-MLP to predict the SE student's performance. The contributions of this study are as follows:

- In order to develop an outstanding predictive performance in machine learning, this paper utilized real and genuine case data from the SE students from Universiti Malaysia Pahang (UMP) [14]. With this real case data, this study able to train machine learning and develop a good prediction model.
- To investigates multiple features in SE students, for instance, Malaysian University English Test (MUET's) results, entry qualifications, gender and status (graduate or fail).
- This study used adaboost, a type of boost that converts the multilayer perceptron into a strong learner for efficient machine learning results.

This paper organized as follows. Section 2 is a literature review to discuss, analyze make a comparison between existing works and this proposed research. Section 3 is to illustrate the methodology in the experiment. The results derived from the experiment are given in Section 4. Finally, Section 5 delivers a conclusion from the results.

2. LITERATURE REVIEW

This section explains some aspects that influence student performance in Software Engineering (SE) course. Furthermore, it also discusses machine learning information in this paper. Last but not least, this chapter will explain about related work for this research.

2.1. Aspects that influence student performance

In the Software Engineering course, there are three aspects that influence student performance. The first aspect related to the subject course. The history of subject courses that already taken by student one of the important aspects because this will be a prerequisite skill and knowledge that need to prepare by students before they take SE course. The second aspect is about the student. This is important because every student has their own skill to understand and solve a problem on every course they take. And the third aspect is about the teacher. The teacher is one of the important objects that implement a learning process in the classroom.

2.1.1. Subject course

In order for the students to excel successfully in SE course, they need to learn the basics of programming. For instance, the basic syntax, structure, and style gradually. One of the crucial parts of programming is to interpret the algorithm into code. If the algorithm is correct, then it will produce the right program and achieve the required objectives. Therefore, many students tend to fail. The students who fail in the examination need to repeat this subject. This is because SE subject is the prerequisite for the next programming subject which enquires higher skill levels than the subject before [15, 16].

2.1.2. Student

Student's interest plays an important role in learning the programming subjects. Otherwise, it will be difficult if they uninterested in SE subjects. Furthermore, different students have different motivations. Learning style's also different for each student. Certain students are more to organize discussions in a group while some other students prefer to study alone [15, 17].

2.1.3. Teacher

Teachers play an important role in delivering the knowledge efficiently to the students (Gomes and Mendes, 2007). Teachers are responsible to make clear explained to the students and suggest to them solution for student's problems. Teachers should have expertise in controlling class situations as this can affect effective learning to students [15, 18].

2.2. Machine learning technology

Machine learning emerged from the computer that has the ability to "learn" some specific task to solve a real-life problem. The aim of machine learning solution is to create some prediction and take a decision for some future improvement [19]. Machine learning technology has been applied in many domains and areas. There are some existing researches that used machine learning implementation in the prediction of student performance [20-22]. However, this research is focused on how to predict student performance using enhanced Multilayer Perceptron (MLP) machine learning classification with Adaboost.

Boosting is a method to improve the machine learning algorithm to gain outstanding accuracy. Adaboost, which is also knowns as Adaptive boosting, is one of the boosting method that introduced in year of 1995 by Freund and Schapire. Compare to other previous boosting methods, Adaboost has resolved multiple cases in practical difficulties. In order to boost and produced precise classification result, it learn of

the weak algorithm and repeatedly multiple times in series of round until gain the good score. In each time it repeats, it generates new weak prediction rule until satisfy the possible outcome.

Meanwhile, this paper used Multilayer Perceptron (MLP) in prediction, as MLP is one of the famous algorithms in Artificial Neural Network (ANN) category. It has contributed a good result in many research areas. MLP is a neural network that inspired from a neuron of a human brain, which consists of set of outputs from a set of inputs, in multiple layers. Therefore, this research is to discover the combination of MLP and Adaboost algorithm to improve the prediction of chemical student either graduate or fail.

2.3. Analysis and comparison of existing works

This section also compares and explains the previous and our propose studies in investigating the student's performance in software engineering. Table 1 summarizes the related work done for a student's performance evaluation. It shows that the two studies use statistical methods to evaluate the student's performance while two studies including ours use machine learning (ML). Moreover, the previous study used Random Forest (RF) without Adaboost, while this study proposed Adaboost with MLP. In addition, we used new fresh dataset from Universiti Malaysia Pahang [14] which is the real case study. The dataset includes the SE students that registered in 2013 and graduated (including fail) in 2017. As this paper used the real case study from UMP, therefore this study is able to contribute to predicting the SE performance in UMP, which contributes to public institutions in Malaysia as well.

Table 1. Comparison with previous and propose studies

Related Works	Research titles	Domains	Methods
[23, 24]	Impact of Pre-University Factors on the Motivation and Performance of Undergraduate Students in Software Engineering	To evaluate the impact of factors prior to university on the performance and motivation of undergraduate freshmen students in Software Engineering	Statistical analysis.
[25]	Major problems in basic programming that influence student performance	To identify problems and causes faced by programming students	Statistical Package for the Social Science (SPSS) Software version 19.
[26]	Using the Random Forest Classifier to Assess and Predict Student Learning of Software Engineering Teamwork	Develop effective machine-learning-based methods for assessment and early prediction of student learning effectiveness in software Engineering teamwork.	Random Forest (RF) machine learning (ML)
Current work	To Predict the Student's Performance of Software Engineering by Using Machine Learning	Detect the performance of students in Software Engineering Course	Used Natural Inspired Machine Learning (Adaboost-MLP) to detect performance (fail or graduated)

3. METHOD

This section discussed the methodology of the undertaken study. Figure 1 depicts the methodology in this study. Initially, we acquired the dataset of software engineering students from the academic center of UMP to implement this study. In order to filter the dataset, we utilized R-Studio. Meanwhile, Waikato Environment for Knowledge Analysis (WEKA) is used to perform the experiments and evaluations. Then we show a timeline of activities during this study and the software and hardware used in this study.

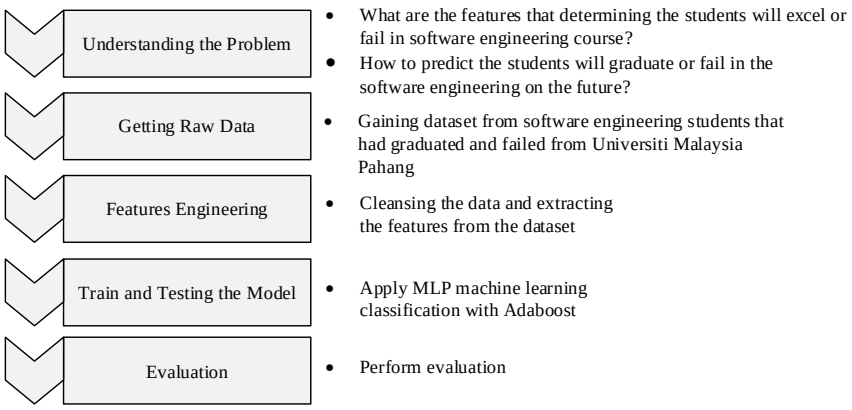

Figure 1. Methodology

3.1. Understanding the problem

In this modern era, the country's future is dependent on information technology (IT). However, the young generation in our country is an individual who will drive the country's technology in the future. Software education is a key area for technological development. Although, the performance of Software Engineering students in programming subjects can be seen decreasing with a high ratio [24]. Based on the journal from UKM [25], students that are weak in the programming subject have low-level confidence to finish individual tasks and they depend on the help of other students. These students also have very minimum initiative, especially those who are moderate and weak in the performance. They always depend on other sources such as answer scheme, the assistance from the lecturers and friends to help them solve the tasks. In another point of view, many students unable to discover that SE course is suitable for them and they can graduate successfully in SE. These problem statements derived various questions. Therefore, the research questions of this study are as follows:

- What are the features that determine the students who will graduate or fail in software engineering course? For instance: previous MUET results, entry qualification from matriculation, Malaysian Higher School Certificate (Sijil Tinggi Persekolahan Malaysia, STPM), Diploma?
- How to discover the students will graduate or fail in software engineering in the future?

3.2. Getting raw data

Aforementioned that we collected the dataset from the academic center of UMP [14], which is a real case and genuine data consists results of software engineering students. This dataset consists the information about MUET's results, entry qualifications, gender, and status. It is important to note that this study is the first research that scrutinized this dataset. As this study is to predict the performance of SE students, therefore there is a need to utilize a decent data from the SE which indicates a real situation.

3.3. Feature engineering

In features engineering, this research needs to clean the data as it comes from raw form [27]. The cleansing data involves deleting information that ambiguous, useless, incomplete and missing. For instance, state to whom the student belongs to, the age of the student, and contact number. Once this step is done, we were able to extract the features and provides many discoveries in a student's status (graduate and fail).

Table 2. MUET Bands with users level

Band	1	2	3	4	5	6
User	Very Limited	Limited	Modest	Satisfactory	Proficient	Highly Proficient

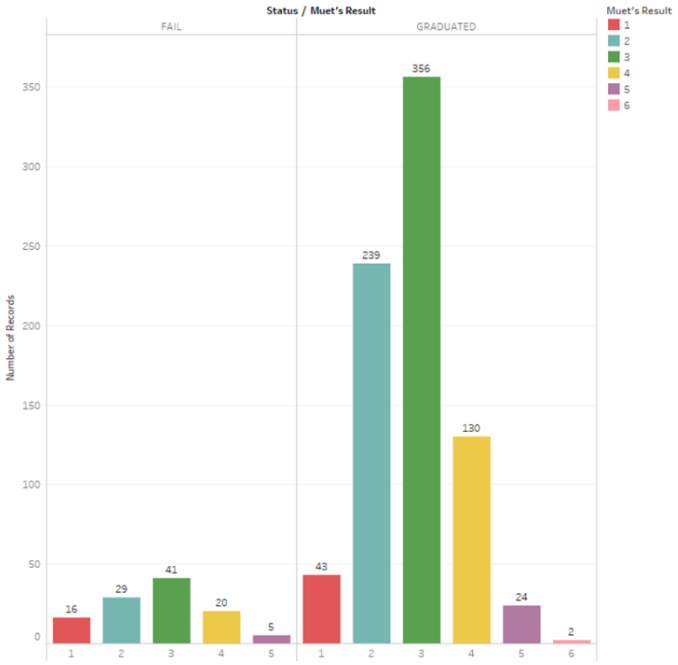

Figure 2. Feature analysis between status and MUET's results

Figure 2 shows the combination between status (graduate and fail) and MUET's result. MUET is an English skill test administered by the Malaysian Examination Council (MEC) [28, 29]. Table 2 portrays the MUET bands with their user's level. Band 1 represents those students who are very limited in English; Band 2 represents those who limited to use English language. Similarly, Band 3 represents modest students, while Band 4 shows the satisfactory one. Band 5 is the second last band in this table; it shows the English proficient students. However, Band 6 indicates that the students were very fluently in practicing English language, they are considered highly proficient. Based on the graph, it depicts that most of the students that obtained band 3 in MUET result (modest English user) were able to graduate successfully in software engineering. On the other hand, the lowest number of MUET result (band 6–highly proficient user) was the fewest student graduate in software engineering. This indicates that the students are able to graduate in software engineering even though they were modest in English.

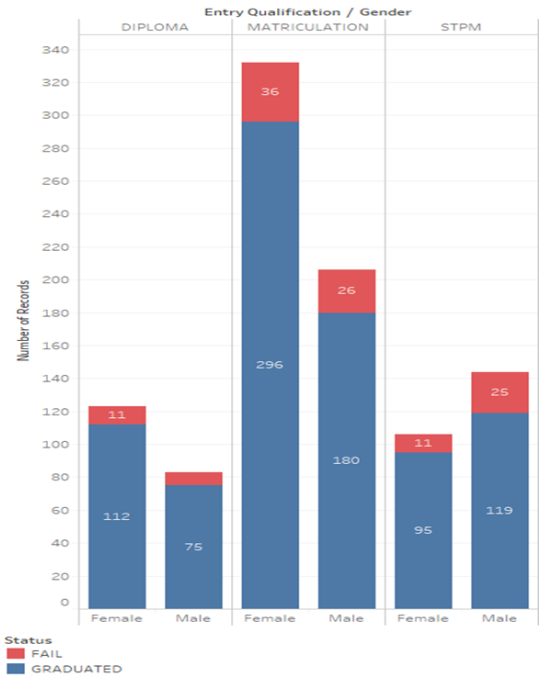

Figure 3. Feature analysis between status, entry qualification and gender

The combination features between status (graduate or fail), entry qualification (Diploma, Matriculation and STPM) and gender (male or female) shown in Figure 3. It shows that 991 students in total got enrolled in SE. Figure 3 depicts that most of the students get register in SE on the base of matriculation while STPM also known as Malaysian Higher School Certificate is on second and diploma is on third place. It shows that more than half of the students got enrolled in the base of matriculation including 330 female students while 210 male students. However, less than half in total got register on the base of remaining two entry qualification including 255 students have STPM while 203 had diploma. This derived assumption that students from matriculation may have a passion for this course or following their friends to register for this course.

On the other hand, in the gender aspect, the number of female students who got registered on the base of Diploma and Matriculation are higher than male students. However, the enrollment of male students is higher in case of STPM entry qualification. The number of failed and graduated students is shown with red and blue bar graph. In the Diploma and Matriculation, the number of female students who failed in SE is higher than the male students, which is 36 in Matriculation and 11 in Diploma rather than 26 male students failed in Matriculation and 6 in Diploma. Furthermore, these graphs show that female who has Diploma and Matriculation are more interested in SE as compare to female from STPM, whereas the number of females from STPM is only 95, while the male is 119. In conclusion, the number of female students is more interested in SE course as compared to male students. However, it also summed up with a higher number of female students failed in SE as compared to male students.

3.4. Train and testing the model

In the interest to discover the performance of the predictions, the predictive dataset needs to be trained before utilizing the algorithms. A few test options provide by WEKA such as Cross-validation, supplied training set, Use training set, Split percentage. This study used cross-validation testing as training and testing sets are various parts and data in testing part is excluded from the training set. Cross-validation is used to overcome the problem of overfitting and makes the predictions more general.

The dataset will be trained with Adaboost to boost the MLP algorithm (Adaboost-MLP) to discover how accurate the predictive model is. The predictive train data is generalized to get a more widely applicable classifier. In this study we used k-fold cross-validation. In this study, we used 10-fold cross-validation while the data is divided in 10 times. In order to get precise accuracy k-1 folds are used for training while one-fold is used for testing.

3.5. Evaluation

This study evaluates the Adaboost–MLP performance in predicting the software engineering students in accuracy benchmark, which shows the percentage of the correctness in classifying the performance either graduate or fail.

4. RESULTS

This section discussed the results from the Adaboost–MLP prediction in SE performance either graduate or fail. The results in accurateness are 87.76 percent in classifying the performance either graduate or fail. With this Adaboost–MLP model, this study able to predict the students before they register to enroll SE course in Universiti Malaysia Pahang or any institution. If the outcome of the prediction result is failed, we advised the student to concentrate and pay more attention to SE course or change any course that interests them.

5. CONCLUSION

In a nutshell, the objective of this study was achieved which is to predict the students in the Software Engineering course in the future either quit or graduate by using enhanced Multilayer Perceptron (MLP) machine learning classification with Adaboost. This study achieved 87.76 percent accuracy in classifying the performance of software engineering students. This study is to predict the student's performance in future, so they would able to reconsider their decision before registering the software engineering course. For the predicted student that will fail, we suggest advising the students to register another course that more suitable for them instead wasting their 4 years' time studying in software engineering. Lastly, we may encourage the students that would excellence in software engineering to register the course.

ACKNOWLEDGEMENTS

This work was supported by Universiti Malaysia Pahang, under the Grant Faculty of Computer Systems and Software Engineering (FSK1000), RDU180360.

REFERENCES

[1] Y.-G. Guéhéneuc and F. Khomh, "Empirical Software Engineering," in Handbook of Software Engineering, S. Cha, R. N. Taylor, and K. Kang, Eds., ed Cham: Springer International Publishing, 2019, pp. 285-320.
[2] D. P. F. Möller, "Systems and Software Engineering," in Guide to Computing Fundamentals in Cyber-Physical Systems: Concepts, Design Methods, and Applications, ed Cham: Springer International Publishing, 2016, pp. 235-305.
[3] A. M. Moreno, M.-I. Sanchez-Segura, F. Medina-Dominguez, and L. Carvajal, "Balancing software engineering education and industrial needs," *Journal of Systems and Software*, vol. 85, no. 7, pp. 1607–1620, Jul. 2012.
[4] R. Pérez-Castillo, I. G. R. De Guzmán, F. García, and M. Piattini, "A Practical Teaching Experience about Software Reengineering," *Procedia-Social and Behavioral Sciences*, vol. 83, no. iii, pp. 254–260, Jul. 2013.
[5] S. Surakka, "What Subjects and Skills are Important for Software Developers?" Magazine Communications of the ACM-The patent holder's dilemma: buy, sell, or troll?, vol. 50, no. 1, pp. 73–78, 2007.
[6] H. Saiedian, "Software engineering challenges of the 'Net' generation," *Journal of Systems and Software*, vol. 82, no. 4, pp. 551–552, Apr. 2009.
[7] D. Budny and J. Tartt, "Do Engineering Students Fail Because they Don'TTt Know How to Fail?," in *Proceedings-Frontiers in Education Conference*, 2009.

[8] S. Helal, J. Li, L. Liu, E. Ebrahimie, S. Dawson, D. J. Murray, et al., "Predicting academic performance by considering student heterogeneity," Knowledge-Based Systems, vol. 161, pp. 134-146, 2018/12/01/ 2018.
[9] A. Yunianta, N. Yusof, M. S. Othman, and D. Octaviani, "Analysis and Categorization of e-Learning Activities based on Meaningful Learning Characteristics," in World Academy of Science, Engineering and Technology, 2012, pp. 811-816.
[10] A. Firdaus and N. B. Anuar, "Root-exploit Malware Detection using Static Analysis and Machine Learning," in *Proceedings of the Fourth International Conference on Computer Science & Computational Mathematics (ICCSCM 2015)*, Langkawi, Malaysia, 2015, pp. 177–183.
[11] A. Firdaus, N. B. Anuar, A. Karim, and M. F. A. Razak, "Discovering optimal features using static analysis and genetic search based method for android malware detection," *Frontiers of Information Technology & Electronic Engineering*, vol. 19, no. 06, pp. 1–27, 2017.
[12] M. A. Ismail, V. Mezhuyev, S. Deris, M. S. Mohamad, S. Kasim, and R. R. Saedudin, "Multi-objective Optimization of Biochemical System Production Using an Improve Newton Competitive Differential Evolution Method," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 7, no. 4–2, p. 1535, 2017.
[13] M. A. Ismail, V. Mezhuyev, K. Moorthy, S. Kasim, and A. O. Ibrahim, "Optimisation of biochemical systems production using hybrid of newton method, differential evolution algorithm and cooperative coevolution algorithm," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 8, no. 1, pp. 27–35, 2017.
[14] U. M. Pahang, "Bahagian Pengurusan Akademik (BPA)," 2018. [Online]. Available: http://bpa.ump.edu.my/index.php/ms/. [Accessed: 21-Dec-2018].
[15] M. Rahmat, S. Shahrani, R. Latih, N. F. M. Yatim, N. F. A. Zainal, and R. A. Rahman, "Major Problems in Basic Programming that Influence Student Performance," *Procedia-Social and Behavioral Sciences*, vol. 59, pp. 287-296, 2012.
[16] D. L. Parnas, "Software Engineering Programs Are Not Computer Science Programs," in *IEEE Software*, vol. 16, no. 6, pp. 19-30, Nov.-Dec. 1999.
[17] T. Punter, R. L. Krikhaar, and R. J. Bril, "Software engineering technology innovation–Turning research results into industrial success," *Journal of Systems and Software*, vol. 82, no. 6, pp. 993–1003, Jun. 2009.
[18] D. Shao, B. Luo, E. Ding and Q. Liu, "An introductory software engineering course for software engineering program," *2011 24th IEEE-CS Conference on Software Engineering Education and Training (CSEE&T)*, Honolulu, HI, 2011, pp. 411-415.
[19] R. Fernandes de Mello and M. Antonelli Ponti, "A Brief Review on Machine Learning," in Machine Learning: A Practical Approach on the Statistical Learning Theory, ed Cham: Springer International Publishing, 2018, pp. 1-74.
[20] X. Xu, J. Wang, H. Peng, and R. Wu, "Prediction of academic performance associated with internet usage behaviors using machine learning algorithms," *Computers in Human Behavior*, vol. 98, pp. 166-173, 2019/09/01.
[21] S. Qu, K. Li, S. Zhang and Y. Wang, "Predicting Achievement of Students in Smart Campus," in *IEEE Access*, vol. 6, pp. 60264-60273, 2018.
[22] J. Xu, K. H. Moon, and M. v. d. Schaar, "A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, pp. 742-753, 2017.
[23] R. M. Felder, K. Forrest, L. Baker-Ward, E. J. Dietz, and P. H. Mohr, "A longitudinal study of engineering student performance and Retention. I. Success and failure in the introductory course," *Journal of Engineering Education*, vol. 82, no. 1, pp. 15–21, 1993.
[24] P. Schoeffel, R. S. Wazlawick and V. Ramos, "Impact of Pre-University Factors on the Motivation and Performance of Undergraduate Students in Software Engineering," *2017 IEEE 30th Conference on Software Engineering Education and Training (CSEE&T)*, Savannah, GA, 2017, pp. 266-275.
[25] M. Rahmat et al., "Major problems in basic programming that influence student performance," Procedia - Social and Behavioral Sciences vol. 59, pp. 287–296, 2012.
[26] D. Petkovic et al., "Using the random forest classifier to assess and predict student learning of Software Engineering Teamwork," *Proceedings-Frontiers in Education Conference, FIE*, vol. 2016–November, 2016.
[27] M. K. Khaleel, M. A. Ismail, U. Yunan, and S. Kasim, "Review on Intrusion Detection System Based on The Goal of The Detection System," *International Journal of Integrated Engineering*, vol. 10, no. 6, pp. 197–202, 2018.
[28] EduAdvisor, "Everything You Need To Know About MUET in Malaysia," 2018. [Online]. Available: https://eduadvisor.my/muet/. [Accessed: 21-Dec-2018].
[29] MPM, "Malaysians Examination Council," 2018. [Online]. Available: http://portal.mpm.edu.my/en/muet1. [Accessed: 21-Dec-2018].