Students’ Performance Prediction using Artificial Neural Network

N Ahmad1*, N Hassan1, H Jaafar1 and N I M Enzai1
1Faculty of Electrical Engineering, University Teknologi MARA Cawangan Terengganu
Kampus Dungun, 23000, Dungun, Terengganu, Malaysia

*Corresponding author: norha291@uitm.edu.my

Abstract. This paper proposed a method for predicting diploma students’ performance of Faculty of Electrical Engineering (EE), Universiti Teknologi MARA (UiTM) Terengganu. Data of 59 first semester students from Electrical Engineering (EE) were obtained to predict students’ academic performance. A predictive model based on a machine learning technique was employed. The predictive model utilizes Artificial Neural Network (ANN) technique that was developed to predict the actual performance of first semester students based on Sijil Pelajaran Malaysia (SPM) results, first semester results, and the interest level of the participants towards EE course. The findings have shown that the developed model could accurately predict the actual result of the first semester students successfully with minimal errors.

1. Introduction
Educational quality is essential for a country’s development. With the use of systems for admission, academic information, learning management system and e-learning, the data amount in education domain has been increasing rapidly. The data collected from students are usually used only for decision making. However, most of the data remain unused due to its complexity and large volume of data sets. Therefore, there is a need for analysis of educational data, especially to predict the students’ performance.

It is necessary to build a system that can continuously keep track of the students’ academic performance and accurately predict their future performance; such as when are they likely to graduate and what are their estimated final grade point averages (GPAs).

In this paper, we propose a method for predicting diploma students’ performance of Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM) Terengganu. The scope of this paper is the prediction of students who could achieve Dean’s List (GPA of 3.5 and above) but the general framework can be used for other performance prediction tasks.

The rest of this paper is organized as follows. Section 2 discusses the related work. Section 3 details the methodology. Section 4 describes results and discussion and section 5 concludes the paper.
2. Literature Review

There are various factors and parameters which could affect the students’ academic performance. The factors could be internal or could be caused by surrounding or external.

Social media such as Facebook has been investigated by [1] in term of its impact towards academic performance. As social media has become inevitable, its usage can’t be avoided even though there’s concern that it could cause distraction in learning process. A systematic review of Facebook usage impact was conducted using Mixed Method Appraisal Tool. The findings indicate both positive and negative effects of Facebook usage on the academic performance of university students. However, the work outlined mechanisms to better utilize Facebook for improving students’ performance academically.

Factors of resilience and engagement of students in relation to their academic performance are studied in [2]. 748 full-time first-year students of the Administration and Business Management degree at two public Business faculties in Northern Spain were involved in the study. The study aims to figure out whether the dimensions of resilience and engagement play important part in ensuring their ability to survive and continue their study until the end. The findings indeed verified that two dimensions of resilience and engagement predict the students’ academic performance. Interestingly, gender also affects the difference in importance of these variables.

Based on the hypothesis that Emotional Intelligence (EI) could affect academic performance, [3] conducted a meta-analysis by searching nine databases for related researches on the topic of interest: ProQuest Dissertations, ERIC, Google Scholar, ISI Web of Science, Medline, and Theses, PsycINFO, ScienceDirect, PubMed, and Scopus. Students with higher EI are expected to be more able to control the negative emotions, manage the social surroundings, and more competent in humanities subjects. The Inclusion and Exclusion Criteria for the meta-analyses are: (a) written in English; (b) published in either a test manual or journal article; (c) direct appraisal of academic performance indices or self-reported. It is found that EI has moderate impact towards producing better academic performance, in particular for skills-based EI tasks.

The knowledge, attitudes and behavioural skills (KAB) model is the basis of [4] in predicting the student academic performance. To enhance the modelling accuracy, variable selection and dimensionality reduction techniques are proposed. A real data set of ten factors that are related to both attitude and skill-related behaviours of 320 first-year students was utilized to conduct evaluation. The prediction accuracy of student achievement was improved by employing the proposed methodologies.

Physical activity, aerobic fitness, and motor skills are believed to have some effect towards executive functions and academic achievement. Cross-sectional data from 697 children from 57 schools in Norway was analysed using a linear mixed model. Even though the impact of physical activity on academic achievement was found to be weak, the results still indicate positive relationship [5].

Motivational beliefs and learning strategy factors in enhancing student learning outcomes are studied by [6]. This work also adopts Motivated Strategies for Learning Questionnaire (MSLQ) as their methodology like [7]. SPSS 17.0 was utilized to perform analyses of data obtained from 323 participants that comprised of Liberian junior and senior high school student. The results indicate that the students’ motivational beliefs and learning strategies give moderate impact on their learning outcomes. Rather, the improvement of education quality is more required by the students.

To study the effect of family background factors combined with students’ previous academic performance on their tertiary first year performance, machine learning algorithms: WEKA and multilayer perceptron have been employed by [8]. This work constructs a model of students’ academic performance to act as building block for other students which consequently forecasts the overall performance. The research sample involved students’ enrolment forms from Babcock University, Nigeria. Among the extracted information are age, gender parent occupations and education, parents’ marital status, religion, and previous educational performances of related subjects. Family background factors as well as culture and region were found to play significant role in the first-year students’ performance.
In the era of Internet and mobility, impact of mobile phone usage has been the research focus of [9]. The research motivation stems from the idea that mobile phone usage during class could distract the students even though mobile phone has its benefits in making the teaching and learning process more interesting. Structured questionnaires and additional interviews were conducted among 45 students of Federal Polytechnic, Ilaro to obtain their personal perspectives. Based on the responses and data analysis, mobile phones give significant negative impact as the students’ attention deteriorated which leads to poor academic performance. However, if the mobile phone usage is managed properly, the academic performance of students can actually be improved.

The effect of students’ exposure to green space and vegetation was also explored by [10]. The relation of tree cover, tree diversity, and tree species composition with students’ academic performance was examined in 387 schools across the Toronto District School Board (TDSB). School-level geospatial and tree inventory data and statistical analysis were employed as research techniques. Schools which showed the highest level of external challenges experienced the most profound effect of tree cover and species composition on student performance [10].

With the objective of assisting in improving students’ performance based on predicted grades, machine learning techniques were utilized by [11] to predict students’ grades in 24 different courses of different semesters. 225 Electrical Engineering undergraduate students were involved in the study. Collaborative Filtering (CF), Matrix Factorization (MF), and Restricted Boltzmann Machines (RBM) techniques were used for systematic analysis and discussed for students’ grade prediction. Out of the three techniques, RBM showed the best results in students’ performance prediction.

Another promising technique, Artificial Neural Network (ANN) is utilized by [12-14] for students’ performance prediction. ANN emulates the biological processing ability by determining the upcoming events based on past data [15],[12] forecasted the undergraduate students’ CGPA of a Chinese university based on their entrance examination results and socio-economic background using ANN. Among the findings are: female students scored better and rural or urban areas did not affect the scores. Meanwhile, [13] took into account factors such as high school score, scores of science and technology related subjects, CGPA obtained during first year, and gender of students in the Faculty of Engineering and Information Technology in Al-Azhar University of Gaza to feed the ANN model. The performance of more than 80% students was able to be predicted correctly. Blended learning course is selected by [14] for ANN implementation. The ANN model considers four learning activities, namely email interaction, wiki content collaboration, files view measurement and online quizzes. Correct classification rate (CCR) of 98.3% was obtained for students’ performance prediction.

Based on the reviewed literature, there are many works that have been done with regards to prediction of student academic performance. This indicates that the need for such works are really crucial in ensuring that the students will be able to perform from the very beginning of their study. ANN technique in particular has garnered wide attention due to its success in prediction, classification, and forecasting in various areas including education, which is the focus of this paper.

3. Methodology

To predict students’ performance, a predictive model based on a machine learning technique was employed. The predictive model utilizes Artificial Neural Network (ANN) technique that was developed to predict the actual performance of first semester students. The details of the prediction model will be discussed in the next section.

3.1 Data collection and Neural Network Model Development

The flow of the overall proposed system could be seen in figure 1. The work comprised data collection, data pre-processing, network/model development, and performance validation. Data of 59 first semester students from Electrical Engineering (EE) were obtained for this study. A survey that consists of Sijil Pelajaran Malaysia (SPM) results, first semester results, and the interest level of the participants towards EE course were collected. These collected parameters will be set as input and target for ANN topology to develop the prediction model shown in table 1. The input variable factors were transformed into a
format that is suitable for neural network analysis. The domain of the input variables used in this study is shown in table 1. The reasons for choosing those parameters as the input data/independent parameters for modeling are because all the chosen subjects are mandatory to be eligible for enrollment in EE course, and students’ interest could be considered as a potential indicator to acquire a good result.

![Figure 1. Block diagram of the students’ performance prediction](image)

Table 1. Input and Output variable

| Input variable       | Domain                          | Output variable     |
|----------------------|---------------------------------|---------------------|
| Additional Mathematics | A+ = 6, A = 5, A- = 4, B+ = 3, B = 2, C = 1 | First-semester CGPA |
| Mathematics          |                                 |                     |
| Physics              |                                 |                     |
| Interest in Electrical Engineering courses | Really interested = 5, Interested = 4, Neutral = 3, Less interested = 2, not interested = 1 |                     |

For efficient training process, it is recommended to pre-process the samples so that the input and target dataset will fall in the same range of values and become reliable. A reliable dataset is important because it can affect the accuracy and performance of the developed model. Furthermore, a dataset could also affect the learning process of the model during training. To get a clean dataset, two types of data pre-processing techniques were utilized in this study, namely data selection and data inspection. For the model development process, ANN was adopted to develop a model that can predict the potential of a student to acquire Dean List status for actual result in the first semester. The structure of the model is depicted in figure 2. ANN is a universal function approximator that generally consists of an input layer, several hidden layers, and an output layer. The networks are very sensitive to the training hyperparameters such as number of the hidden layer, number of hidden neuron in one hidden layer, and activation function. Proper hyper-parameters must be chosen to produce a high-quality network. ANN with one hidden layer trained with backpropagation technique is developed in Matlab. The initial layer where the input comes into ANN is called the input layer. It is immediately followed by one hidden layer while the last layer where the output is produced is called the output layer. To develop the prediction model, the collected data will be divided into two parts. The first part is the training and learning process which is to create a nominal model with regards to the loaded data. The second part is the efficiency evaluation of the developed model and the data for this process which is called unlearning data or unseen data.

In the training phase, data was divided into three groups: 70% of the data were used to train and construct the network, 15% to validate the network that is developed by the training data set, and the remaining 15% will be utilized to reconfirm the prediction ability of the developed network. Several values of the hidden neuron are varied by using systematic trial and error method in training state to develop a reliable network. To achieve an optimal and desirable set of weights to pre-defined network architecture, learning and training processes are needed. These processes are based on an optimization approach to adjust the network weights and bias concerning the set of input-output. In this study, the network is trained by Lavernberg-Marquardt (LM) Back Propagation algorithm, since it offers a more efficient training scheme compared to gradient descent [16]. This training method falls under the variation of Newton’s method which updates weights and biases value based on LM optimization. The optimal hidden neurons value for this work is 20.
Input variable

Mathematics
Additional Mathematics
Physics
Interest in EE course

Output variable

above 3.50
or below 3.50

Figure 2. Neural Network Structure for students’ performance prediction

4. Results and Discussion

Four components must be considered to evaluate the performance of a model in the training state or during the model development process: confusion matrix training, Receiver Operating Characteristic (ROC), cross entropy performance, and error histogram plot.

A confusion matrix is known as a contingency table or an error matrix. It is a specific table layout that allows visualization of the performance of an algorithm as supervised learning. For confusion matrix mapping shown in figure 3, the network outputs could be said as an accurate model since it shows the higher numbers of correct responses in the green regions. The number of incorrect responses is lower as shown in the red regions. The lower right blue region indicates the overall accuracy of the developed model which is 93.2%.

Figure 3. Confusion matrix

Receiver Operating Characteristic (ROC) shown in figure 4 is a plot of the true positive rate (sensitivity) versus the false-positive rates (1 - specificity) as a threshold is varied. The plot shows that the model almost had a perfect test for the training phase since the value is approaching 100% sensitivity and 100% specificity. Thus, it is coherent with the performance mapping in the confusion matrix.
Figure 4. Receiver Operating Characteristic (ROC)

Figure 5 and figure 6 show the cross-entropy result and error histogram respectively. Cross-entropy is a loss function to optimize the classification model and the best cross-entropy performance for this model is 0.10664. The lower the cross-entropy value, the better the resulting model. In the error histogram, the significant error mapping range is between −0.5752 and 0.5752, and most of the errors are tabulated at the center of the plot which is near zero value. Based on the above-mentioned model performance evaluator, it indicates that most of the training data are successfully discriminated during the training process.

Figure 5. Cross entropy network performance

Figure 6. Error histogram

To make sure the developed model is robust and reliable; an additional step or process is conducted. During this process, the network is introduced to the data that is never encountered during the training process. The comparison of the real output of unlearned (data that has never been seen by the network during the training process) data set and predicted output from the developed model is shown in table 2. The errors between the real value and predicted value of five students are calculated to see the network prediction ability when dealing with the unseen data set. All errors are shown in table 2 which depicts low values. This means that the target output is discriminated very well.
Table 2. The validation process for five students

| Student | Input [add math, math, Physic, interest] | Real output | Prediction output | error |
|---------|-----------------------------------------|-------------|-------------------|-------|
| Student 1 | 5,5,5,5 | 1 | 0.9320 | 0.068 |
| Student 2 | 3,3,3,5 | 1 | 0.6069 | 0.393 |
| Student 3 | 3,2,2,4 | 0 | 0.0066 | 0.0066 |
| Student 4 | 2,2,2,2 | 0 | 0.0066 | 0.0066 |
| Student 5 | 2,5,3,4 | 1 | 0.9071 | 0.0929 |

5. Conclusion

This research aims to predict the students’ performance using a supervised learning technique which is ANN. Based on the students’ SPM results and interest in EE course, the findings have shown that the developed model could accurately predict the actual result of the first semester students successfully with minimal errors. This framework is also expected to be adopted as a tool to make an early prediction of Dean List candidates from the first semester students. Although the developed model shows reasonably positive results, further studies and investigation are necessary to understand the correlation between cause and effect of every single parameter involved in this study. For future work, input variables like family factors and financial issue of students can be added. This research could contribute to education quality management through systematic students’ performance monitoring.

References

[1] Nadeem S, Dhool M A, Butt R M, Rasheed A and Bashir A 2020 A Systematic Review on the Impact of Facebook Usage on Academic Performance Rev. Manag. Sci. 2(1) 17-39.
[2] Ayala J C and Manzano G 2018 Academic performance of first-year university students: The influence of resilience and engagement High. Educ. Res. Dev. 37(7) 1321-1335.
[3] MacCann C, Jiang Y, Brown L E, Double K S, Bucich M and Minbhashian A 2020 Emotional intelligence predicts academic performance: A meta-analysis Psychol. Bull. 146(2) 50.
[4] Al-Sheeb B A, Hamouda A and Abdella G M 2019 Modeling of student academic achievement in engineering education using cognitive and non-cognitive factors J. Appl. Res. High. Educ. 11(2) 178-198.
[5] Aadland K N, Moe V F, Aadland E, Anderssen S A, Resaland G K and Ommundsen Y 2017 Relationships between physical activity, sedentary time, aerobic fitness, motor skills and executive function and academic performance in children Ment. Health Phys. Act. 12 10-18.
[6] Gbollie C and Keamu H P 2017 Student academic performance: The role of motivation, strategies, and perceived factors hindering Liberian junior and senior high school students learning Educ. Res. Int. 2017 1-11.
[7] Pardo A, Han F and Ellis R A 2016 Combining university student self-regulated learning indicators and engagement with online learning events to predict academic performance IEEE Trans. Learn. Technol. 10(1) 82-92.
[8] Goga M, Kuyoro S and Goga N 2015 A recommender for improving the student academic performance Procedia Soc. Behav. Sci. 180 1481-1488.
[9] Soyemi J, Oloruntoba S and Okafor B 2015 Analysis of mobile phone impact on student academic performance in tertiary institution Int. J. Emerging Technol. Adv. Eng. 5(1) 361-365.
[10] Sivarajah S, Smith S M and Thomas S C 2018 Tree cover and species composition effects on academic performance of primary school student PLoS One 13(2).
[11] Iqbal Z, Qadir J, Mian A N and Kamiran F 2017 Machine learning based student grade prediction: A case study arXiv preprint arXiv 1708.08744
[12] Lau E, Sun L and Yang Q 2019 Modelling, prediction and classification of student academic performance using artificial neural networks *SN Appl. Scien.* 1(9) 1-10.

[13] Abu-Naser S S, Zaqout I S, Abu Ghosh M, Atallah R R and Alajrami E 2015 Predicting student performance using artificial neural network: In the faculty of engineering and information technology *Int. J. Hybrid Inf. Technol.* 221-228.

[14] Zacharis N Z 2016 Predicting student academic performance in blended learning using Artificial Neural Networks *Int. J. Artif. Intell. Appl.* 7(5) 17-29.

[15] Arsad P M and Buniyamin N 2013 A neural network students' performance prediction model (NNSPPM) *IEEE Int. Conf. on Smart Instrum. Measure. and Appli. (ICSIMA)* 1-5.

[16] Munakata T 2008 *Neural Networks: Fundamentals and the Backpropagation Model:* Springer p 7-36.