IOP Conference Series: Materials Science and Engineering

PAPER • OPEN ACCESS

The prediction of undergraduate student performance in chemistry course using multilayer perceptron

To cite this article: Che Akmal Che Yahaya *et al* 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **769** 012027

The prediction of undergraduate student performance in chemistry course using multilayer perceptron

Che Akmal Che Yahaya, Che Yahaya Yaakub, Ahmad Firdaus Zainal Abidin, Mohd Faizal Ab Razak, Nuresa Fatin Hasbullah and Mohamad Fadli Zolkipli

Faculty of Computing, College of Computing and Applied Sciences, Universiti Malaysia Pahang, 26300 Kuantan Pahang Malaysia

E-mail: pcn18003@stdmail.ump.edu.my, aqemal.yahaya@gmail.com

Abstract. Chemical industry are key elements for changing crude materials to our ordinary objective merchandise. This has achieved an immense move in how things work. The disappointment pace of an understudy in a science course is additionally expanding with regards to requesting the compound specialists. Understudies who enlist the science course regularly bomb in the first or consequent semesters. Moreover, understudies are likewise unfit to comprehend in the event that they can adjust and graduate effectively in this program. The objective of this exploration is to foresee the future utilization of improved science for understudies to fizzled or graduate by upgraded Multilayer Perceptron (MLP) arrangement with Adaboost. The exactness of the outcomes is 92.23% percent**.**

Keywords: chemistry, undergraduate, education, machine learning, multilayer perceptron, neural network

1. Introduction

The chemistry course is an important education that teaches a student the interaction between matter and energy surrounding us [1]. It involves the world around us, such as medicine, cooking, cleaning, and environmental issues and also material that involves electronic devices. Many decades of studies in chemistry have resulted in the growth of several thousand molecular descriptors describing a variety of possibly any compound characteristics. Therefore, chemistry one important subject that can serve as one of the knowledge that can be implemented in routine activities. Among another subject, this research chooses chemistry subject because numerous times of science research has prompted the improvement of a few thousand atomic descriptors that depict a scope of properties of possibly any compound in our daily life [2].

However, for academic purposes these subjects tend more students to fail at their school level. Among 97,095 SPM candidates, 3.9 percent of them were failed [3]. In addition, many students unable to discover that they have the talent in chemistry courses and able to pass in flying colors. Hence, it is important to predict a student either graduates or fail before they register in this course.

By predicting the student performance (fail or graduate), it will highlight the students that have suitable in the chemistry course which will consume for 3 years. On the other hand, this prediction also will identify the students that attempt to fail and therefore the student may change other courses or to remind the student to give more concentration in a chemistry course. Accordingly, these review adopts machine learning to predict a student either suitable in the first place before entering the course. For conducted the situation, machine learning Multilayer Perceptron (MLP) has been chosen because it surpasses all the classifiers in the experiment [4]. While Adaboost for well-known boosting arrangement calculation, and it one of the learning type classifiers. It can be classified grouping calculation that improves the expectation execution during the preparation information, which blunders to improve the MLP [4]. Multilayer Perceptron is a part of artificial intelligence consciousness. It applies to a variety of methods for assessing dataset-based capacities. Such capacities, called Multilayer Perceptron (MLP), would then be able to be utilized to foresee future results. In this unique situation, a calculation is a predefined arrangement of steps that accepts a lot of information as sources of info and changes it through scientific tasks. MLP of different types uses models and calculations as structure hinders for making ecological forecasts and surmisings. The MLP algorithm will discover designs in the information, at that point anticipate the result of something that has never been considered.

This study proposes Adaboost-multilayer perceptron (MLP) to predict the performance of a student in a chemistry course. The contributions from this research as follows:

a) In order to develop outstanding performance based on machine learning predictive, implemented the real datasets from Universiti Malaysia Pahang (UMP) students in Pahang, Malaysia. With the real dataset and true situation, is able to provide great input for the machine learning training.
b) To investigate the multiple features based on student entry qualifications (matriculation, STPM, and diploma), MUET results, gender (male or female), Malaysian citizenship (Bumiputera), states in Malaysia and status (graduate or fail).
c) Implement Adaboost as an addition to MLP as a strong learner machine learning for efficient results.

The structure of the research is as follows. Section 2 reviews the related works and compare with the proposed study. Section 3 explains the methodology in the experiment. Section 4 provides results derived from experiment. Finally, section 5 delivers a conclusion from the results.

2. Literature reviews

Each paragraph discusses the factors that have an effect on the student's success in the chemical course. In addition, the machine learning methodology used will be clarified briefly and addressed in relation to the previous work for this study.

There are several factors that impact student performance in gaining the results:

a) *The lack of interest in Chemistry course*. Chemistry is a subject of great significance to science. Many students are afraid to learn chemistry because it is difficult to understand. This situation makes students uninterested in studying and completing the course. Students tend, however, to perform better if they are interested in the subject.

b) *Problems with language and communications*. Some of the words used in the science subject vary from those used in other subjects and in ordinary communication. In contrast, the subject of science is mostly in English. This makes it even more difficult for students to learn and understand that English is poor. It causes problems for students to understand the material of the chemical topic well [2]. It includes the ambiguity of the form of the sentence and the language used by lecturers and educators.

c) *Mismatch Technique by Lecturers/Teachers*. Different lecturers and instructors have their own style of teaching. Some of them favored practical instruction and some are not. They might be happy with their way, but not with the students. Teachers or lecturers must be able to approach students in a constructive and comfortable way so that they can research and learn better.

IOP Conf. Series: Materials Science and Engineering **769** (2020) 012027 doi:10.1088/1757-899X/769/1/012027

2.1. Machine learning
Machine Learning is a class of calculation that empowers applications to be progressively exact in anticipating results without being unequivocally redone. The fundamental explanation of ML is to make computations that can get input data and use quantifiable assessment to foresee execution while reviving yields as new data winds up available.

Machine learning algorithms is regularly named supervised or not supervised. The supervised algorithm requires an information researcher or information investigator with ML skill to give both information and yield, notwithstanding giving criticism on the precision of forecasts during calculation training [5]. Information researchers choose what factors or highlights of the model ought to be assessed and used to make expectations. When learning is done, the calculation must apply what has been figured out how to new information.

Unsupervised algorithms do not need to be equipped with the required result information. Alternatively, they use an iterative approach called in-depth training to analyse information and come to conclusions. Unsupervised learning algorithms, also known as neural networks, are used for more complex computing tasks than supervised learning systems, including image recognition, speech-to-text, and natural language generation. These neural networks work by millions of examples of training data and automatically finding sometimes small comparisons between many variables [6]. When equipped, the algorithm can use its association datasets to interpret new data. Such algorithms have only become possible in the age of big data because they require a massive amount of training data.

2.2. Types of machine learning
Just as there are almost infinite applications for machine learning, there are unlimited of machine learning algorithms. They range from relatively easy to highly complex. Here are some of the most widely used versions:

This class of machine learning algorithm involves identifying a correlation -- generally between two variables -- and using that correlation to make predictions about future data points.

a) *Decision trees*: these models use perceptions of specific activities and decide the best way to land at the ideal result [7].
b) *K-means clustering*: This model gatherings a given number of datasets that focus on a particular number of bunches dependent on comparative attributes.
c) *Neural networks*: Such profound learning models utilize a lot of preparing information to set up a relationship between numerous factors to figure out how to process approaching information later on.
d) *Reinforcement learning*: This territory of profound learning includes models that go through numerous endeavors to finish a system. Steps that produce good results will be remunerated and steps that produce undesired results will be punished until the calculation learns the ideal procedure.

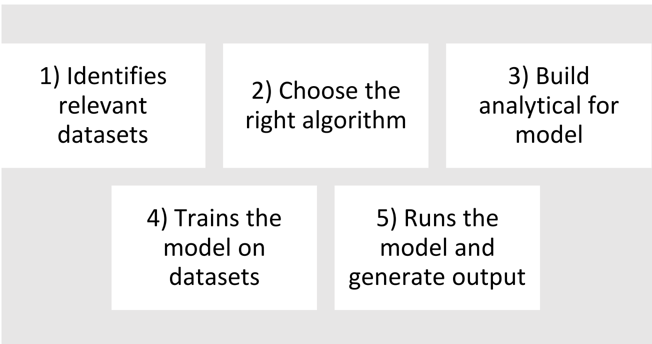

Figure 1. The machine learning methodology

2.3. Artificial neural network

A neural network system (ANN) is one of the most ordinarily utilized calculations. It is a PC model that has been utilized as a PC program. Complex connections among info and yield information groupings can be demonstrated utilizing ANNs. It was likewise used to recognize shrouded designs in informational indexes when the quantitative methodology, which is an express model dependent on the multifaceted nature of the issue, is tedious or isn't as of now conceivable. ANNs supplant these obscure practical associations with evaluated (roughly) adaptively assembled capacities. [4].

Neural systems are a lot of calculations, displayed freely after the human mind, which is intended to perceive designs. They translate tactile information through a sort of machine discernment, marking or grouping crude info. The examples they perceive are numerical, contained in vectors, into which all certifiable information, be it pictures, sound, content or time arrangement, can be deciphered

Neural networks help us cluster and classify the categories. It helps to group unlabelled data according to similarities among the example inputs, and they classify data when they have a labelled dataset to train on. Neural networks can also extract features that are fed to other algorithms for clustering and classification; thus neural networks as components of larger machine-learning applications involving algorithms for reinforcement learning, classification, and regression [5].

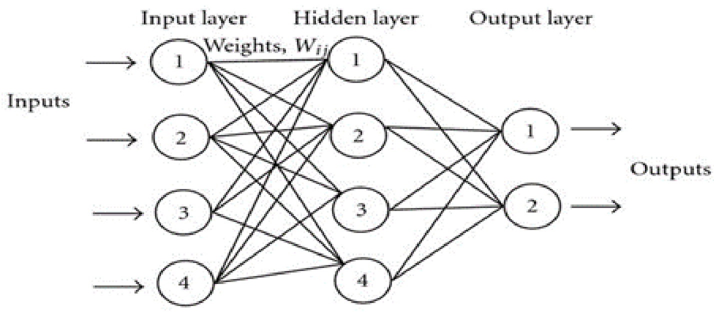

Figure 2. Schematic represent layer ANN

2.4. Boosting

Some random learning calculation precision can be improved by utilizing a general technique that is boosting. Adaboost (Adaptive Boosting) was distributed in 1995 by Freund and Schapire. A large number of the common sense troubles of the prior boosting calculations have been survived. The more solid and predictable outcome is acquired by learning the powerless calculation more than once in a progression of rounds that are given by the promoter calculation. After each time the basic learning calculation is named, it delivers another feeble forecast law. Likewise, Adaboost is utilized to test the improved classifier for the recognizable proof of understudy execution in the concoction course. There are a few focal points and detriments of boosting. Improves the consequences of the examination and upgrades the product for improved execution. The disadvantages of boosting are touchy to clamours and hard to actualize on the ongoing system.

2.5. Multilayer perceptron

The multilayer perceptron is one of the neural system framework's directed learning calculations. The multilayer perceptron comprises of a system of straightforward interconnected neurons or hubs and is completely connected to every hub associated with every hub in the following and past layers [8].

IOP Conf. Series: Materials Science and Engineering **769** (2020) 012027 doi:10.1088/1757-899X/769/1/012027

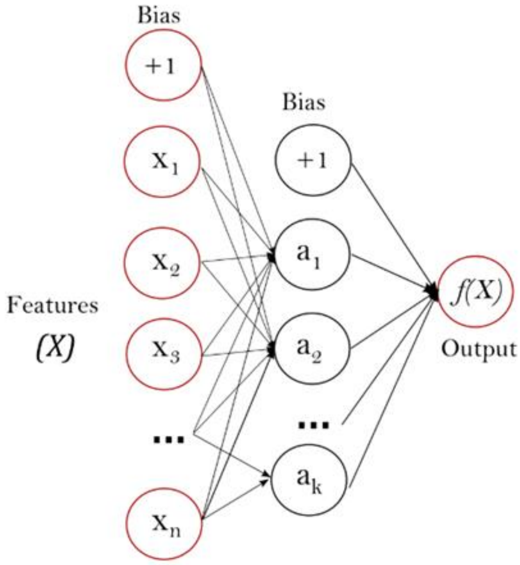

Figure 3: Multilayer perceptron

Advantages of using Multilayer Perceptron that can train non-linear models and learn models in real-time using partial fit and MLP drawbacks if hidden layers have a non-convex loss function where more than one local minimum occurs. As a consequence, different random weight initializations can lead to different accuracy of validation. Including MLP, it involves tuning a range of hyper parameters, such as the number of hidden neurons, layers, and iterations, and it is difficult to implement and interpret. The students ' information is based on their research history and the list of studies will be explained [9].

2.6. Related works

On the examination of the related work, the discoveries of past investigations, which likewise utilized explicit ML classifiers, contrasted from the test data[10]. For the examination, the Multilayer Perceptron (MLP) has a superior result, 92.5 percent, than different classifiers, including gullible Bayes and Forests, to assess the number. More investigation utilizing five separate classifiers uncovers that utilitarian trees (FT) are of the most elevated dependability (95 percent) and genuine positive worth (TRP) (96.7 percent).

In the past examination utilizing a similar Fisher test and the wellness test, the potential danger of utilizing the test is that the information utilized ought to be the recurrence or tally as opposed to the next kind. Meanwhile, the multilayer perceptron is equipped for perusing and testing information regardless of whether the data isn't in recurrence or checks. In the other case presented by the NVivo program, it is basic and proficient to utilize. It likewise improves the exactness of subjective investigations [11].

The cross-validation test is reasonable for the technique for identifying the exhibition of the understudies utilizing the Multilayer Perceptron (MLP) in the Adaboost algorithm. Cross-validation overlays the information multiple times and prepares the information to accomplish precise exactness. In light of the preparation dataset check, the exactness of the Adaboost calculation is 92.23%. It utilized to address the issue of over-fitting and to make forecasts increasingly broad [12].

Table 1. Comparison with previous paper and propose studies

Related works	Research titles	Proposed works	Methods
[1]	A longitudinal study of engineering student performance and Retention. I. Success and failure in the introductory course.	Detect the performance of students whether success or failure.	Used Fisher's exact test for independence between two categorical variables.
[2]	The inaccuracy of self-evaluation variable of student risk of failing the first year in chemistry.	Predict the risk of failing for the first-year students in a chemistry course.	Goodness-of-Fit test of Hosmer and Lemeshow (1991) using SPSS Statistics 17.0.
[3]	Root-exploit Malware Detection using Static Analysis and Machine Learning	Detect root-exploit malware in Android OS.	Identify the root-malware using three different machine learning classifiers.
[4]	Discovering optimal features using static analysis and a genetic search-based method for Android malware detection	To search for the best and smallest number of features to enhance accuracy and reduce the amount of complexity, noise, and irrelevant data in detecting malware.	Used five machine learning classifiers, Naïve Bayes (NB), functional trees (FT), J48, random forest (RF), and multilayer perceptron (MLP).
Current work	To Predict the Future of Undergraduate Student's Performance in Chemistry by Using ML	Detect the performance of students in the chemical course	Used Multilayer Perceptron in to detect performance (fail or success)

3. Methodology

The strategy tended to concern the structure of the test look into and the use of the techniques. The points of this work were to be utilized as a source of perspective for exploratory examination. It is to guarantee that the objectives of this exploration are accomplished. The machine learning programming called WEKA variant 3.8 is utilized for the investigation of the usage of characterization systems. It will likewise be talked about the stages forecast students ' success in the chemical course. Then we're showing the machine learning process and method used.

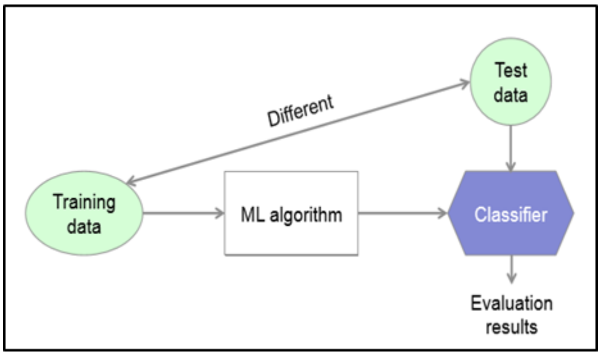

Figure 4. Machine learning process

So as to do an evaluation dependent on Weka tools as a software, it is essential to foresee the nature of understudies in the chemistry course whether to fall flat. Preparing information and test information are being set up to foresee understudy execution. Prescient preparing information will be utilized to get familiar with the ML algorithm and will at that point be founded on the preparation information of the framework. Next, the assessment result is gotten by assessing the information produced utilizing the MLP classifier strategy.

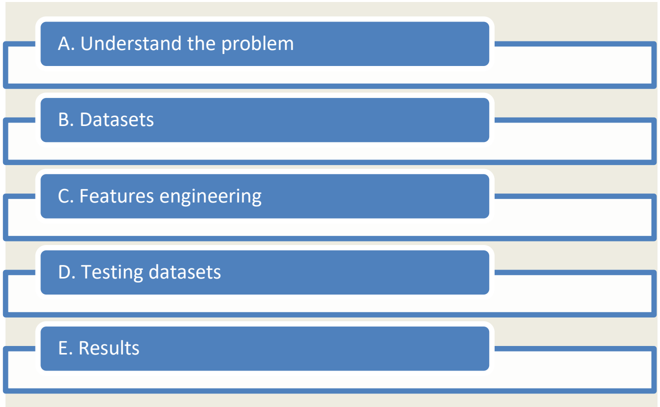

Figure 5. Methodology used

3.1. Understand the problems
The problems occur when the situation become hard to solve. For this situation, students who have initiatives can use other varieties way to solve the problems.

The problem statements from the previous section derived various questions. Hence, the research question derived as follows:
1) What are the features that determining the students will excellence or fail in a chemistry course? For instance: previous MUET results, entry qualification from matriculation, STPM, Diploma?
2) How to discover the students will fail in chemistry in the future?
3) How to discover the students will achieve excellent results in chemistry in the future?

3.2 Dataset
To predict the performance in a chemistry course, therefore, there is a need to utilize a decent dataset from students as a real situation. It collected datasets from the academic center of University Malaysia Pahang (UMP) which counted years from 2002 until 2015. It is important to note that this research is the first experimental data that scrutinizes this dataset.

3.3 Features
In features engineering, this research is in a raw form. The cleansing step data involves ambiguous, incomplete, and missing. Once it does, we were able to extract the features and provides many discoveries in student status (graduate and fail).

IOP Conf. Series: Materials Science and Engineering **769** (2020) 012027 doi:10.1088/1757-899X/769/1/012027

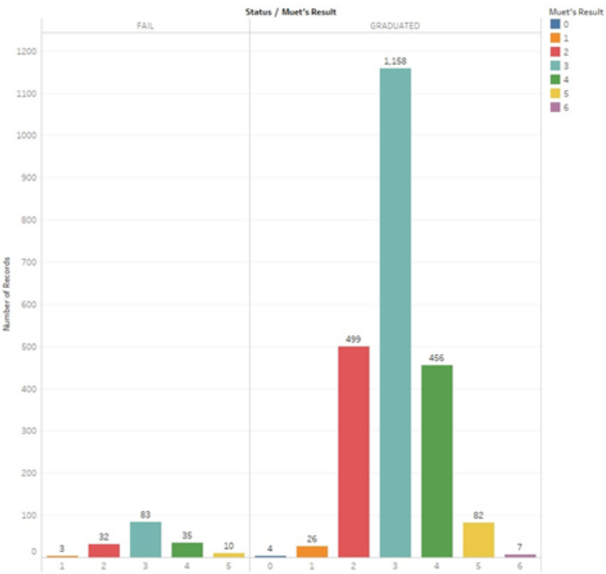

Figure 6. Features analysis by entry qualification, MUET

Figure 6 demonstrates the status of male and female students, whether to fail or succeed on the basis of their enrolment qualifications and MUET tests. Malaysia University English Test (MUET) is an English skill test administered by the Malaysian Examination Council (MEC) [15]. In this test, band 1 is the score for the student that hardly able to use the English language, however, the band 6 indicates that the students were very fluently in practicing the English language. Based on Student qualification before registering a chemistry course

In order to register the undergraduate course (bachelor degree) in Malaysia, there are three types of qualifications: (1) Diploma; (2) STPM; and (3) Matriculation.

A diploma is one of the qualification one step below an undergraduate degree, in Malaysia Diploma courses, are often taken up by students looking to enter the workforce at the earliest possible time, by equipping them with practical skills and knowledge for their specific industry. Entry into a Diploma course only requires students to have completed their high-school education, making it a very popular option amongst students in Malaysia. If students decide to change their minds at a later point, they also have the option to transfer to Year 2 of their chosen Undergraduate Degree as well, since Diploma courses typically range between 24 to 36 months [11].

Sijil Tinggi Persekolahan Malaysia (STPM, English: Malaysian Higher School Certificate) is a pre-university examination carried out by students in Malaysia. Formerly recognized as the Higher School Certificate (HSC). The HSC was a predecessor to the GCE A-Level in the United Kingdom and is still the name of the pre-university exam in some Australian states.

Since 1982, the STPM has been administered by the Malaysian Examinations Council (MEC), which has also been running the Malaysian University English Test (MUET) since 1999. However, national exams such as the Ujian Penilaian Sekolah Rendah and Sijil Pelajaran Malaysia are all set and reviewed by the Malaysian Examinations Syndicate (MES). Nevertheless, both the MEC and the MES are under the Ministry of Education [12].

STPM is one of the two main pre-university systems for admission to public universities in Malaysia. The other is a one-year enrolment program run by the Ministry of Education. STPM is not the only credential approved in addition to the Master Program and the Malaysian Higher Islamic Religious Certification (STAM). Technically, applicants can apply for admission to degree-level courses with a variety of pre-university examinations considered to be comparable to STPM, including A-level. Nevertheless, all those applying for university must have approved the MUET. STPM is internationally recognized by many universities, especially those in the Commonwealth of Nations, the United States and the Republic of Ireland. Many universities find STPM results to be equal to GCE A-Level results [13].

The Malaysian Matriculation Program (in Malay language i.e. Program Matrikulasi Malaysia or all the more regularly known as just Matrikulasi) is a 1 to the 2-year pre-college course that enables understudies to seek after a degree upon effective consummation. Matrikulasi is one of the most looked for after pre-college programs among Malaysians as it is an incredibly practical course into tertiary instruction. Understudies just need to pay a little enrollment charge, and the remainder of the expenses are borne by the Malaysian government. Registration understudies additionally get a stipend each semester for their everyday costs. Ordinarily, understudies who picked a possibility for Matriculation will keep on pursuing their degree with neighborhood open establishments. So, Malaysian Matriculation is likewise perceived by a few remote colleges, for example, those from Australia, United Kingdom, Indonesia, New Zealand, and a few other Commonwealth countries [14]. The chart demonstrates that the vast majority of the understudies who got Band 3 in MUET (unobtrusive English client) had the option to effectively graduate in science. On the other hand, the lowest number of MUET results (band 6–a highly qualified user) was the lowest student graduate in chemistry. It suggests that students are able to graduate in chemistry even though they were moderate users of English. This indicates which Band 3 enrolment students in MUET take the most of the courses. The students of Band 3 in MUET is the Standard English that the students have adept at. For most of the universities in Malaysia, the prerequisite for a student to pursue his or her studies is to get Band 3 into their MUET performance. The students who got the 5th band and 6 the course often takes place as they are knowledgeable in English, which is the main language used in universities. This is seen as their strength because it will help them learn and understand English quickly.

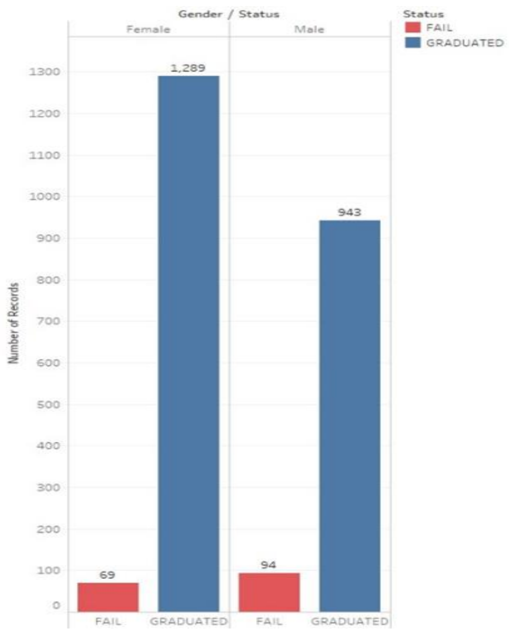

Figure 7. Entry qualification based on gender

Figure 7 shows the status of the student's status based on their gender. Female students are more advance to pass the course compare to male students.

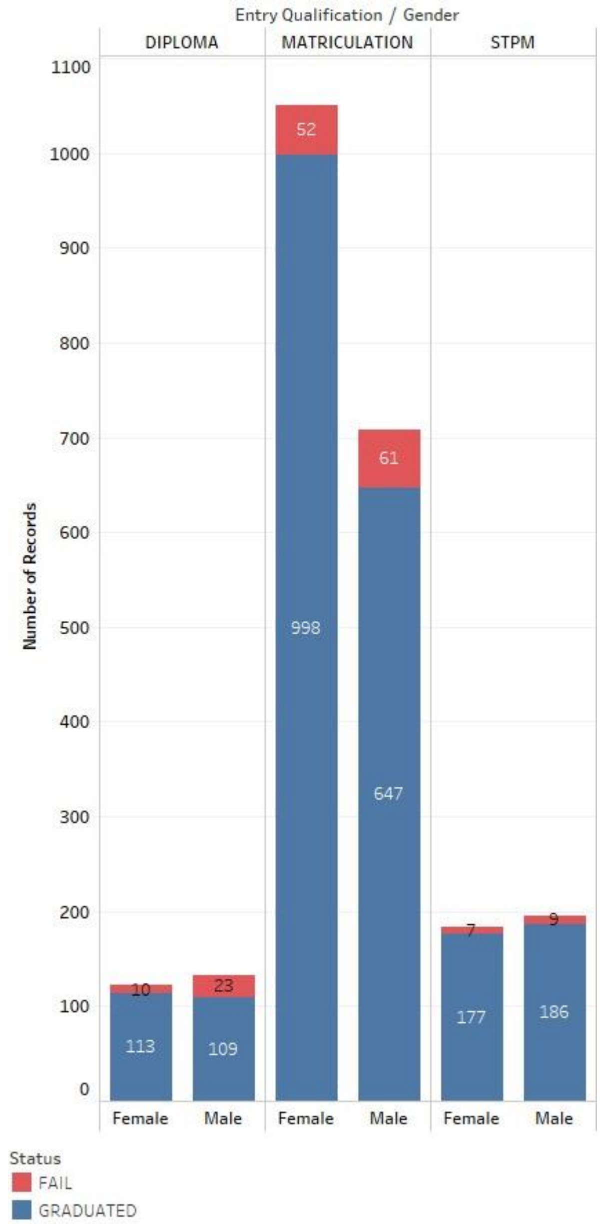

Figure 8. Result of student's performance by entry qualification and gender

Figure 8 shows the status of the student's status based on their entry qualification and gender. Female students from matriculation have the highest value in passing the course rather than students from the diploma**.**

File Edit View

Latest2.csv.arff

Relation: Latest3

No.	1: ENTRY_QUALIFICATION Nominal	2: RESULT_MUET Numeric	3: GENDER Nominal	4: STATE Nominal	5: BUMIPUTRA Nominal	6: STATUS Nominal
1	MATRICULATION	1.0	M	Teren...	B	FAIL
2	STPM	5.0	M	Wilaya...	N	GRADU...
3	MATRICULATION	3.0	F	Sabah	B	GRADU...
4	MATRICULATION	4.0	M	Saraw...	B	GRADU...
5	DIPLOMA	3.0	M	Selan...	B	GRADU...
6	MATRICULATION	2.0	M	Perak	B	GRADU...
7	MATRICULATION	3.0	M	Teren...	B	GRADU...
8	DIPLOMA	2.0	M	Selan...	B	GRADU...
9	MATRICULATION	3.0	M	Wilaya...	B	FAIL
10	MATRICULATION	4.0	M	Perak	B	GRADU...
11	MATRICULATION	3.0	M	Pahang	B	GRADU...
12	MATRICULATION	2.0	M	Teren...	B	GRADU...
13	DIPLOMA	3.0	M	Teren...	B	GRADU...
14	MATRICULATION	3.0	M	Kelant...	B	GRADU...
15	MATRICULATION	2.0	M	Teren...	B	GRADU...
16	MATRICULATION	4.0	M	Selan...	B	GRADU...
17	MATRICULATION	5.0	M	Wilaya...	B	FAIL
18	MATRICULATION	3.0	M	Selan...	B	FAIL
19	MATRICULATION	2.0	M	Perak	B	GRADU...
20	MATRICULATION	3.0	M	Perak	B	GRADU...
21	MATRICULATION	3.0	M	Kelant...	B	GRADU...
22	MATRICULATION	3.0	M	Johor	B	GRADU...
23	MATRICULATION	3.0	F	Sabah	N	GRADU...
24	MATRICULATION	2.0	M	Pahang	B	GRADU...
25	DIPLOMA	1.0	M	Pahang	B	GRADU...
26	MATRICULATION	3.0	F	Pahang	B	GRADU...
27	MATRICULATION	3.0	F	Wilaya...	B	GRADU...
28	MATRICULATION	4.0	F	Perak	B	GRADU...
29	MATRICULATION	3.0	M	Johor	B	GRADU...
30	MATRICULATION	3.0	M	Selan...	B	GRADU...
31	MATRICULATION	2.0	F	Perak	B	GRADU...
32	MATRICULATION	3.0	F	Selan...	B	GRADU...
33	MATRICULATION	4.0	M	Selan...	B	GRADU...
34	MATRICULATION	3.0	F	Perak	B	GRADU...
35	MATRICULATION	3.0	M	Perak	B	GRADU...
36	MATRICULATION	4.0	F	Neger...	B	GRADU...
37	DIPLOMA	3.0	M	Kelant...	B	GRADU...
38	MATRICULATION	3.0	M	Kelant...	B	GRADU...
39	MATRICULATION	2.0	M	Kelant...	B	GRADU...

Figure 9. Show the features of data for each student

The data that is used for the training set filter before it evaluates using Cross-Validation test. In order to evaluate the accuracy of the performance of the students, test data need to be prepared. The test data are based on student's features such as entry qualification, result MUET, gender, state, *bumiputra*, and status. The data is prepared in .arff file name and then will be evaluated using the algorithm model. Figure 9 shows an example of the data.

Classifier output

```
Time taken to build model: 11.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        2209               92.2338 %
Incorrectly Classified Instances       186                7.7662 %
Kappa statistic                          0.0199
Mean absolute error                      0.1222
Root mean squared error                  0.2676
Relative absolute error                 96.044  %
Root relative squared error            106.2587 %
Total Number of Instances             2395

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.025    0.012    0.129      0.025    0.041      0.028    0.568     0.093     FAIL
                 0.988    0.975    0.933      0.988    0.960      0.028    0.568     0.944     GRADUATED
Weighted Avg.    0.922    0.910    0.878      0.922    0.897      0.028    0.568     0.886

=== Confusion Matrix ===

    a    b   <-- classified as
    4  159 |    a = FAIL
   27 2205 |    b = GRADUATED
```

Figure 10. Result training and testing using WEKA

IOP Conf. Series: Materials Science and Engineering **769** (2020) 012027 doi:10.1088/1757-899X/769/1/012027

3.4 Testing datasets
Figure 10 indicates the likelihood calculation of the student's success or failure that depends on the characteristics of each student. It is documented after the data has been educated. The outcomes of the training data were analyzed using cross-validation, the training set given, the training set used, the percentage split. Implementation is used cross-validation testing as the training and testing set are different parts and the information in the test part is omitted from the training set. Cross-validation is utilized to beat the issue of over-fitting and to make expectations increasingly succinct.

Datasets will be outfitted with the Adaboost-MLP calculation to discover how precise the prescient model is. Adaboost classifier consolidates low classifier calculations to enable a solid classifier. By entering two classifiers, the reliability of better results will be increased [14]. Predictive train information is generalized to make the classifier more broadly available. Cross-validation is folded 10 times and train data to achieve correct accuracy.

3.5 Evaluation
To evaluate the Adaboost-MLP in predicting the student accuracy benchmark, which shows the percentage of the correctness in classifying the performance either graduate or fail.

4. Result
This section defines the results of the prediction of the performance of the students by using the data being collected from UMP. The Adaboost-MLP technique predicted 92.23 percent accuracy in classifying the performance of students before they enroll in the chemistry course. Hence, with this Adaboost – MLP model, before register chemistry course, this study able to predict either the student graduates or fail in this course in the near future.

5. Conclusion
This section offers an end on anticipating the accomplishment of understudies in a science course utilizing machine learning. Eventually, the goal is to test and apply an MLP calculation to anticipate understudy disappointment or graduate execution dependent on highlights (ENTRY QUALIFICATION, RESULT, GENDER, STATE, BUMIPUTRA STATUS). In validating the forecasting model, cross-validation method has been implemented. This work shows that Adaboost-MLP is applicable to predict student performance and the student status who will fail or graduate the course. If the prediction result is failed, the students are encouraged to focus more during the class sessions if they are still interested in chemistry course. On the other hand, the students are suggested to enroll for other courses that interest them. If the prediction result is graduated, this means that the student may have a talent in chemistry and be able to pass through in flying colors. Nevertheless, this estimation also depends on the student's ability and commitment to research the length of the chemistry course.

Acknowledgment
This work supported by Universiti Malaysia Pahang, under research grant by Faculty of Computing, Universiti Malaysia Pahang (RDU180360). The authors would like to thank the Academic Management Division (BPA), Gambang campus for providing them the dataset.

References
[1] A. M. Helmenstine, "Why study chemistry?" 2019.
[2] G. B. Goh, N. O. Hodas, and A. Vishnu, "Deep learning for computational chemistry," J. Comput. Chem., vol. 38, no. 16, pp. 1291–1307, 2017.
[3] Berita Harian, "Pendidikan Negara pada landasan betul," 16-Mar-2018.
[4] M. F. A. Razak, N. B. Anuar, F. Othman, A. Firdaus, F. Afifi, and R. Salleh, "Bio-inspired for Features Optimization and Malware Detection," Arab. J. Sci. Eng., 2017.
[5] "Mobile Malware Detection by Analyzing Official Market Metadata," Comput. Secur.
[6] A. Firdaus, N. B. Anuar, M. F. A. Razak, and A. K. Sangaiah, "Bio-inspired computational

paradigm for feature investigation and malware detection: interactive analytics," Multimed. Tools Appl., vol. 76, no. 280, pp. 1–37, 2017.
[7] B. Sanz, I. Santos, C. Laorden, X. Ugarte-Pedrero, P. G. Bringas, and G. Alvarez, "PUMA: Permission Usage to detect Malware in Android," in Advances in Intelligent Systems and Computing, 2013, pp. 289–298.
[8] P. P. K. Chan and W. K. Song, "Static detection of Android malware by using permissions and API calls," in International Conference on Machine Learning and Cybernetics, Lanzhou, China, 2015, vol. 1, pp. 82–87.
[9] S. Arshad, M. A. Shah, A. Wahid, A. Mehmood, H. Song, and H. Yu, "SAMADroid: A Novel 3-Level Hybrid Malware Detection Model for Android Operating System," IEEE Access, vol. 6, pp. 4321–4339, 2018.
[10] N. Pehlivan, U.; Baltaci, N. ; Acarturk, C. ; Baykal, "The analysis of feature selection methods and classification algorithms in permission based Android malware detection," Comput. Intell. Cyber Secur., pp. 1–8, 2014.
[11] L. Ye, C. Shuniak, R. Oueini, J. Robert, and S. Lewis, "Can they succeed? Exploring at-risk students' study habits in college general chemistry," Chem. Educ. Res. Pract., vol. 17, no. 4, pp. 878–892, 2016.
[12] H. Fröhlich, O. Chapelle, and B. Schölkopf, "Feature Selection for Support Vector Machines by Means of Genetic Algorithms," in Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'03), Sacramento, California, USA, 2003, pp. 142–148.
[13] EduAdvisor, "Everything You Need To Know About MUET in Malaysia," 2018.
[14] S. Sheen, R. Anitha, and V. Natarajan, "Android based malware detection using a multifeature collaborative decision fusion approach," Neurocomputing, vol. 151, pp. 905–912, 2015.
[15] Eduadvisor",Matriculation 2018.[Online]. Available: https://eduadvisor.my/matriculation/ [Accessed:21-Jan-2019]
[16] MPM, "Malaysians Examination Council," 2018. [Online]. Available: http://portal.mpm.edu.my/en/muet1. [Accessed: 21-Jan-2019].