Failure Prediction using Artificial Neural Networks

Brian Justin S. Ramos
College of Engineering
De La Salle University
2041 Taft Avenue
Malate, Manila,
Philippines
brian_justin_ramos@dlsu.edu.ph

Jules A. Danting
College of Engineering
De La Salle University
2041 Taft Avenue
Malate, Manila,
Philippines
jules_danting@dlsu.edu.ph

Luis Paolo de Vera
College of Engineering
De La Salle University
2041 Taft Avenue
Malate, Manila,
Philippines
luis_paolo_devera@dlsu.edu.ph

Ejnar Azarraga
College of Engineering
De La Salle University
2041 Taft Avenue
Malate, Manila,
Philippines
ejnar_azarraga@dlsu.edu.ph

I. INTRODUCTION

There are many problems students are facing and one of the major concern is either passing or failing a subject so what causes failures? That is where our project comes in. Most of the students nowadays try to cram the night before their exams so does that affect their scores? How about the travel time many students wake up early in the morning just to travel to their own schools and because of this plenty of students try to absorb the topics the teachers discuss but is too sleepy to absorb the said topics. Also another way for students to be affected by travelling is that when traveling is very time consuming for the student especially during rush hours which can lead to severe road congestion affecting the student's studying time. But many students as well are just lazy to study and spend most of their time watching, playing games, listening to music, and other miscellaneous tasks. So what does affect the student's scores in their quizzes and that is what our project is all about trying to understand what causes a student to either pass or fail.

II. TRAINING DATA

The data that the group will use is based on a paper on students behaviors and attitudes ranging from 0 - 5 on how they think it affects their studies in a daily basis. With 0 as negligible in terms of weight and 5 as the highest factor. A sample of 100 will be used. There will be 9 inputs for the training data. Each input are based on the common area's wherein most students might be affected such as feeling sleepy during class hours and the time allotted for studying is limited. The first input is feeling sleepy during class, this is very common and is a factor since it affects the student's ability to absorb the lessons being discussed, the second input when students only study when there are quizzes, this affects their grade depending on their ability to absorb the topic fast. The third input indicates that students prefer to do other things such as listening to music, watching television, etc. rather than studying. The fourth input indicates the location of the student in this case if they live far the travel time might affect their time to study. The fifth input shows that is the student is constantly disturbed thus affecting their concentration on studying. The sixth input show whether the study strictly follows the time schedule. The seventh input indicates the weight in which the student has time to study at home. The eighth input is feeling hungry during class shows that they have lesser concentration on the discussion due to hunger. The ninth input indicates the location or the environment of the classroom were discussions are being held.

III. NEURAL NETWORK IMPLEMENTATION

To solve the problem stated above, our group will be using Artificial Neural Networks. Artificial Neural Networks (ANN) are best used in tackling multiple inputs non-linear hypothesis problems such as the problem discussed from the previous sections. Compared to other methods available, ANN method is the best choice because this method works almost similar as how does the human brain function which in turns gives us optimal, accurate, and desirable results.

In implementing the solution, our group will be using MATLAB's nftool (Neural Network Fitting Tool). 100 samples of data each having 9 variables as well as its corresponding outputs will be used for training the neural network. A 100x9 matrix will contain the input data and a 100x1 matrix will contain the outputs for each sample and will be used as the target inputs on the nftool. The samples will be divided into three groups: 60% of the samples will be used for training, 20% of the samples will be used for Validation, and the remaining 20% will be used for Testing. According to the MATLAB's nftool the training samples are presented to the network during training, and the network is adjusted according to its error. The validation are used to measure network generalization, and to halt training when generalization stops improving, the testing samples have no effect on training and are used as independent measure of network performance during and after training. The training algorithm to be used is Levenberg–Marquardt back propagation.

To obtain better accuracy, we will use a separate dataset named testinput and use our network to predict the output. Accuracy is obtained by counting the number of correct predictions over the number of data samples. We are expecting to get at least 80% accuracy by using this method.

IV. RESULTS AND DISCUSSION

	Samples	MSE	R
Training:	60	5.65284e-2	8.28328e-1
Validation:	20	5.89810e-2	7.55165e-1
Testing:	20	5.89852e-2	7.81628e-1

Fig. 1. Training Results

Based from the figure above, 60% of the samples were used for the actual training while both validation and testing used 20% of the samples each as the group believe that this is the optimal distribution of samples. In order to arrive with this distribution of data, the group manually adjusted the distribution of samples on both the testing field and the validation field. Giving the validation field more than 20% of the samples overfits the model and giving it less than 20% would result to an underestimated output.

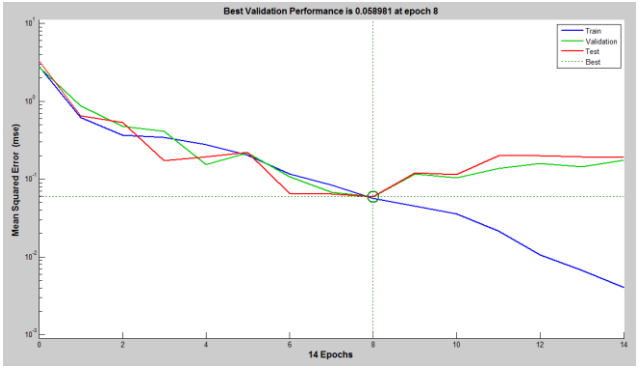

Fig. 2. MSE Plot

From the figure above, it can be seen that after 8 iterations, validation, training, and testing shows 0.058981, 0.056528, 0.058985 errors respectively which can be considered as the best performance output based from the data used. At this point, training, validation, and testing could be considered as "error free" as their mean square values are all approaching 0, further iterations will give us increasing errors for training and validation while testing will be continuously decreasing. Therefore, we can say that 8 iterations will be needed in order to get the correct output.

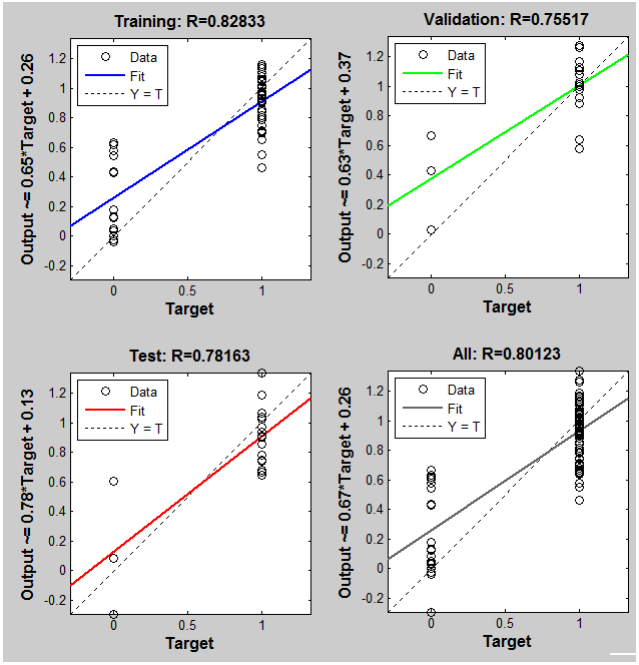

Fig. 3. Training, Testing, Validation Plots

From figure 3, we can see that the regression (R) values of training, testing, and validation are 0.82833, 0.78163, and 0.75517 respectively, these values when rounded up will be equal to 1, meaning the output and the target shows a close relationship not just a random relationship. The final plot is the resulting plot when 100% of the data were used in the regression test giving us a result of 0.80123 which when rounded up will also give us a value of 1. This means that our output and our target have almost the same result.

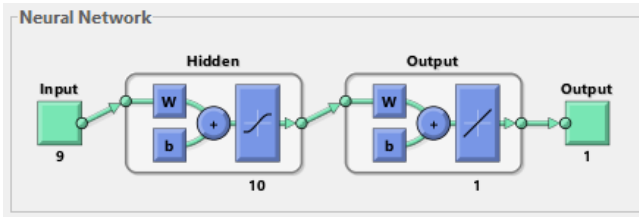

Fig. 4. Neural Network

Figure 4 shows the neural network consisting of 9 inputs, 10 hidden neurons, 1 output neuron, and 1 output that will determine whether the student will pass (1) or fail (0). 10 hidden neurons were used in testing as it generates a result that has 90% accuracy. Increasing the value of the hidden neurons (ex. 100 hidden neurons) will yield better result but the group decided to use only 10 hidden neurons because it is enough to meet and even surpass the goal of 80% accuracy, the required % of accuracy appointed by the instructor.

```
accuracy =
    90
```

Fig. 5. Accuracy

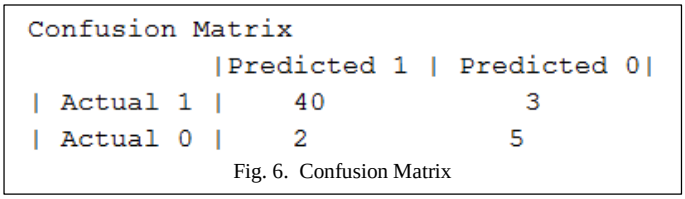

Fig. 6. Confusion Matrix

The confusion matrix that the group have attained based on the results. The group managed to get 40 samples predicted that actually passed and 5 samples predicted that actually failed. The overall accuracy of the classifier is (40 + 5) / 50 * 100 % = 90 % of the samples is correct just as shown in the output accuracy in Fig. 5.. The misclassification rate indicates a 10 % error. The precision or when it predicts that the student will fail is 5 / (3 + 5) * 100 % = 62.5 % and 40 / (40 + 2) * 100 % = 95.24 % chance of predicting that a student will pass.

V. CONCLUSION

In Conclusion we were able to make a program that Predicts if the student will fail according to the set of data asked and then implemented using Artificial Neural Network and successfully having 90% Accuracy this program can be applied to help students avoid failing their classes when they unknowingly get in one of those criteria because many students fail today not knowing that their study environment, study habits or other elements such as travel time and work load affects their studies and health. We recommend that this program be further improvised by increasing the accuracy and widening the criteria set for the student to be able to adjust the program according to their own preferences and even implement how they can improve themselves in order to pass future subjects.

ACKNOWLEDGMENT

The author acknowledges the assistance provided by his group team, professor during the development of this project.

REFERENCES

1. Why people usually fail the exams *by Prof. Randall Edna Wells Handy, Esq*. (n.d.). Retrieved from

http://academic.udayton.edu/legaled/barpass/other/resource02.htm

2. Why do many students who study hard fail the exam anyway? (n.d.). Retrieved from *https://www.quora.com/Why-do-many-students-who-study-hard-fail-the-exam-anyway*

3. Why do first year students fail to progress to their second year? An academic staff perspective (n.d.). *Véronique Johnston Department of Mathematics, Napier University* .

Retrieved from

http://www.leeds.ac.uk/educol/documents/000000453.htm

4. WHYDO STUDENTS FAIL? ACADEMIC LEADERS' PERSPECTIVES

Abour H. Cherif, Gerald E. Adams, Farahnaz Movahedzadeh, and Margaret A. Martyn

Retrieved from

http://cop.hlcommission.org/Teaching-and-Learning/cherif15.html

5. Exam pressure: 'You don't need to be scared of failure. There's more to life than grades'(n.d).

Retrieved from *http://www.telegraph.co.uk/women/womens-life/11061310/Exam-pressure-You-dont-need-to-be-scared-of-failure.-Theres-more-to-life-than-A-grades.html*

6. Artificial Neural Networks in Hardware: A Survey Janardan Misra

Retrieved from *http://www.researchgate.net/publication/223938078_Artificial_neural_networks_in_hardware_A_survey_of_two_decades_of_progress*

7. *Artificial Intelligence Topics*, [online] Available: http://www.aaai.org/AITopics/pmwiki/pmwiki.php/AITopics/HomePage

8. F. Glover "Future paths for integer programming and links to artificial intelligence", *Comput. Oper. Res.*, vol. 13, no. 5, pp.533 -549 1986

9. D. E. Goldberg *Genetic Algorithms in Search, Optimization, and Machine Learning*, 1989 :Addison-Wesley