Analyzing the Effects of Non-Academic Features on Student Performance

1st Gabriela Castillo-Rivera
Department of Computer Science
Middle Tennessee State University
Murfreesboro, United States
gmc3e@mtmail.mtsu.edu

2nd Jeffrey Hooper
Department of Computer Science
Middle Tennessee State University
Murfreesboro, United States
jch5u@mtmail.mtsu.edu

3rd Austin Leverette
Department of Computer Science
Middle Tennessee State University
Murfreesboro, United States
abl4c@mtmail.mtsu.edu

4th James Scruggs
Department of Computer Science
Middle Tennessee State University
Murfreesboro, United States
jms2dv@mtmail.mtsu.edu

Abstract—**Many factors contribute to the academic success of an individual. By determining what these factors are, we can improve student performance and thus the prosperity of a nation. By analyzing quantitative and qualitative data of students in secondary education, we determined what factors have the most impact on student success and how they do so. The data used was taken from a data set by Paulo Cortez comprised of questionnaire answers from students in two Portuguese schools. After cleaning the data, we used a simple feed-forward network with two layers to attempt to predict student success based on many factors. Afterwards, we removed features one at a time to attempt to find which features have the greatest impact on success. While the initial network had an accuracy of 77%, this was improved by removing data that had a negative impact on accuracy. After removing this data, the final network was able to achieve an accuracy of 92%. The most influential feature was the presence of family support. Its removal resulted in a decrease in the accuracy of 3.0769%. There was minimal to no effect on student grades. The second most influential feature is shared by four: reason for choosing school, time spent going out, free time after school, and extra paid classes. All decreased the accuracy by 2.3077% upon their individual removal. There seems to be an optimal amount of time with friends and free time for best student performance. Attending extra paid classes had a detrimental effect on grades.**

Index Terms—**neural networks, education, data analysis, secondary education, predictive technology, feed-forward, student performance, final grades**

I. INTRODUCTION

It is said that our children are our future, and for that reason we, as a society, try our best to prepare them for the world they live. Secondary education helps accomplish this by bridging the gap from primary education to either post-secondary, vocational training, or the work force. Whichever path an individual decides to take, secondary education is crucial for the prosperity of a nation. With secondary education related to unemployment and incarceration rates, the importance of education becomes apparent to a society [1]. The United States addresses this by spending 5.62% of its GDP on education – this equals more than a trillion dollars [2]. However, many students still fall short of finishing a post-secondary education. Yearly, over 1.2 million individuals in high school dropout in the United States [3]. Can we assume that there are certain contributing factors to a student's success in a secondary institution? And if we can, how do we identify what they are? What are the affects on student performance? Using a neural network, this is what we intend to find out.

II. BACKGROUND

Research in student performance has been aided many times using neural networks. Researchers at the University of Technology Mara Malaysia used a neural network to determine which early subjects in electrical engineering contributed the most towards the students success in the major. Data on 391 students was collected from the university. The input data consisted of student grades for individual subjects (six subjects per student) in an early semester while the output is predicting the cumulative grade point average. Using sigmoid for the hidden layer, purelin for the outer layer, they were able to achieve a relatively low Mean Squared Error (0.05544) [4]. However, this neural network is limited in that it only considers quantitative data.

In the European Union, Portugal ranks the lowest with student success rates. Paulo Cortez and Alice Silva, Information Systems Scientists at the University of Minho in Portugal, collected data on the secondary public-school system to see if a neural network could determine factors contributing to the high failure rates of their youth. In their study, 788 students were given questionnaires with predefined options. Along with the questionnaires, past exam grades and attendance were collected; 139 samples were discarded due to faulty sampling leaving 649 confident samples. Using a feed-forward classification neural network, they were able to achieve 72% accuracy with three classes and 62% accuracy with nine classes [5]. This study considered both qualitative and quantitative data which is why we use the same sample collected for our neural network.

III. METHODS

Our data consists of information concerning students from two schools in Portugal during the 2005-2006 school year. Information was collected both from school results along with questionnaires, which are used to determine additional details about the students environments. We are interested in grade information, which is scored on a scale of 0-20, along with other information about the students environment. Some of the information includes parental education, alcohol consumption, health status, and past class failures. We looked at student performance in Portuguese of which there are 649 records.

With some of our input data being qualitative, we had to encode it for the network. This was done by using a label encoder tool provided by Scikit-learn. This was done for all binomial and nominal data. Zeros (0) were encoded to negative ones (-1) for the binomial data to ensure proper weight updates. In addition, we needed to shuffle our dataset due to its ordered nature. Without this step, we would be training on data exclusively from one school, then testing on data from a different school.

For our networks output shape, we decided to attempt classification of students scores into 5 categories. The categories range from 0, the worst score, to 4. We use a keras sequential model with two hidden layers for our simple feed-forward network. Both of these layers have the rectified linear unit as its activation function. ReLU is used because it avoids activations of zero, which can be an issue with softmax, while still allowing for nonlinearity. The function is defined as $f(x) = max(0, x)$ [6]. An image of the function is shown in Figure 1.
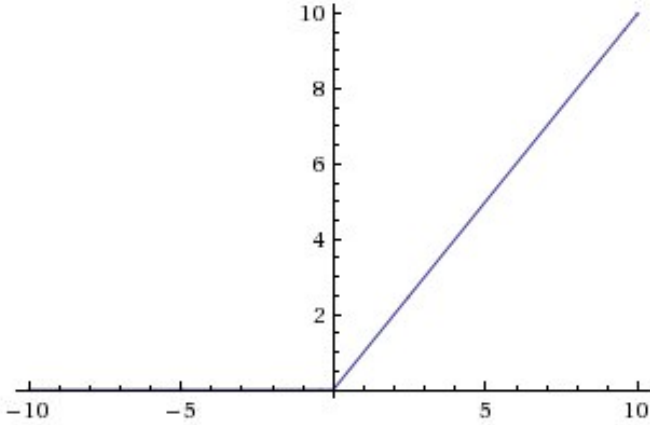

Fig. 1. A graph showing the form of ReLU [6].

The first hidden layer has 800 units while the second has 400. The final, output layer uses the softmax activation function, which allows us to predict one of several categories. The error function we selected is categorical cross entropy due to its ability to help us classify students into one of multiple different categories.

Now that we have a network which we believe can predict student scores, we split our data into two sets, one for training and one for testing of the trained network. After several trials, we decided that beyond 7 epochs, our model began overfitting, so we stopped at that point. As our main objective is determining what factors contribute the most to student success, after an initial test, we began testing the network while removing a single feature every trial. Afterwards, we attempted to determine which features had the greatest impact on our models performance. We then selected the most detrimental of those features and removed them all at the same time, hoping to improve our results.

IV. RESULTS

During the initial test of the network, a test accuracy of 77% was achieved while using all features in the data set. It was thought that this accuracy could be increased by removing redundant features. This test is depicted in figure 2
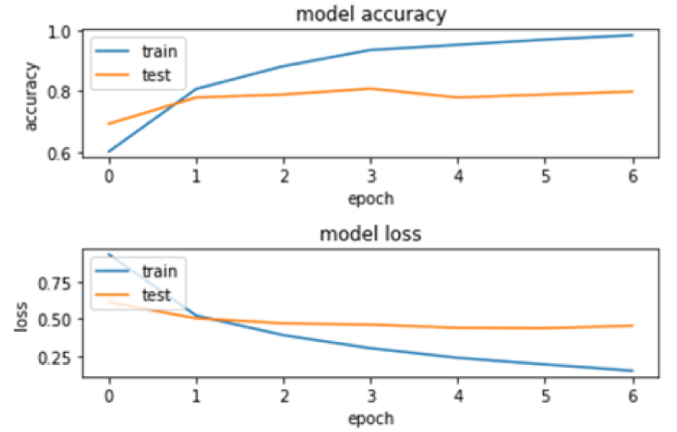

Fig. 2. Initial test results.

There were two features whose removal from the training data had little effect on the networks accuracy. The removal of the school and grade three features saw little to no increase in network accuracy. Figure 3 is from the network that trained on the data without the grade three feature.
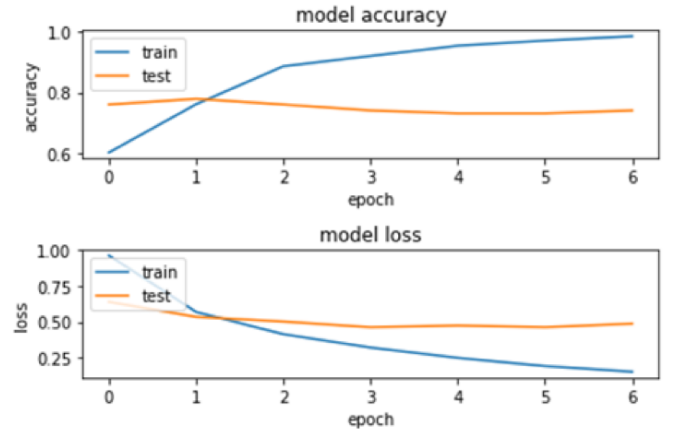

Fig. 3. Test results without grade 3.

Aside from the two features listed above, the removal of any other feature from the data set caused the networks test accuracy to increase significantly. The removal of the feature that describe the mothers occupation caused the test accuracy to increase to 88%. The same is true for the feature that describes the number of absences the student has, and the higher education feature. Figure 4 depicts network accuracy and loss with the absence feature removed.
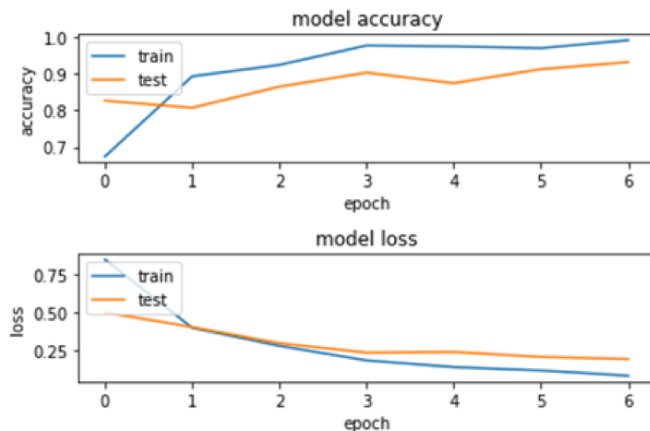

Fig. 4. Test results without absences.

The highest test accuracy achieved was 97% accuracy. This accuracy rating was reached on three different occasions of feature removal: the removal of the age feature, the removal of the address feature, and the removal of the free time feature. Figure 5 depicts the overall accuracy and loss of the network after the removal of the age feature.
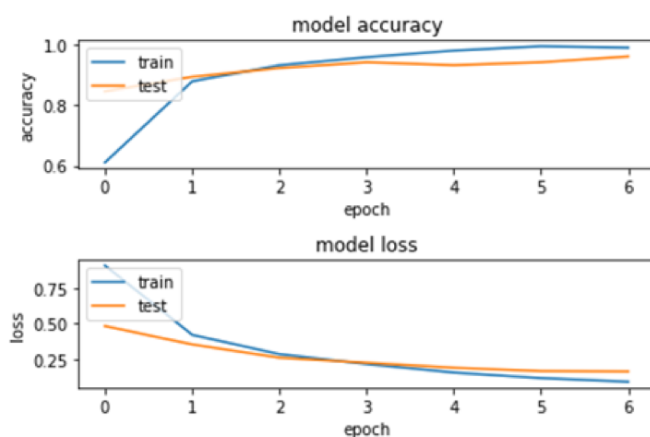

Fig. 5. Test results without age.

The removal of other features from the training set generally resulted in a test accuracy between 90% and 97%. The highest accuracy among this group is 96%, after the removal of the feature that described how much alcohol a student drank on the weekend. A test accuracy of 90% was the lowest accuracy score of this group, after the removal of the travel time feature

from the training set. Figure 6 shows the accuracy and loss of the network that was trained without the travel time feature.
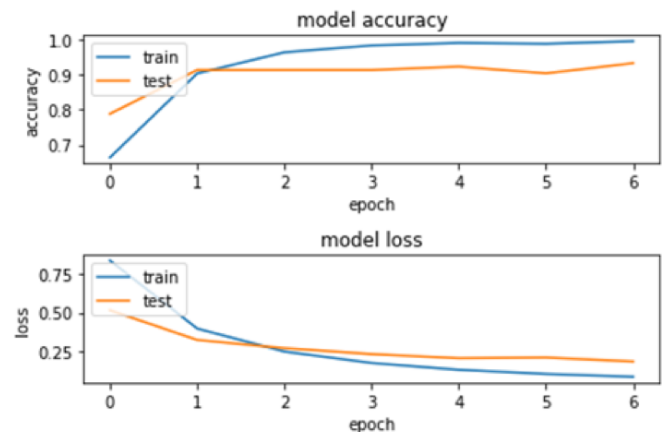

Fig. 6. Test results without travel time.

Other notable features are the grade one and grade two feature, whose removal saw an increase in test accuracy to 90% and 93% respectively. Features like internet access and romantic involvement, after removal, saw a test accuracy of 95% and 96% respectively.

Lastly, we removed several features that seemed to have the most negative effects on accuracy and trained the network on the new data set. The features removed were student internet accessibility, fathers job, mothers job, and fathers education. This test yielded an accuracy of 92% and a loss of .172.
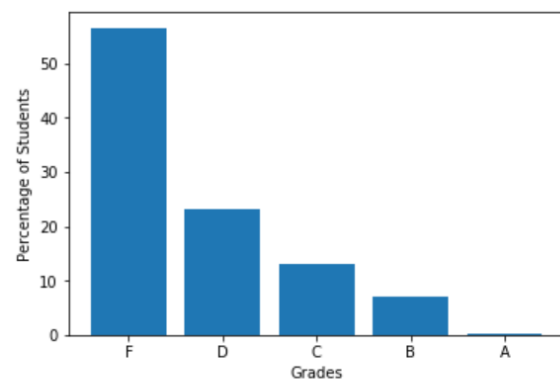
[56.53266332 23.11557789 13.06532663 7.03517588 0.25125628]

Fig. 7. Distribution of grades among students with family educational support.

Upon reaching an acceptable percentage of accuracy, we found the non-academic features that negatively impacted this accuracy the most with their removal. The most influential of these was the presence of family support which decreased the accuracy by 3.0769%. The next four most influential features - reason for choosing school, the frequency which the student goes out, the amount of free time the student has after school, and the usage of extra paid classes - all decreased the accuracy by 2.3077%.

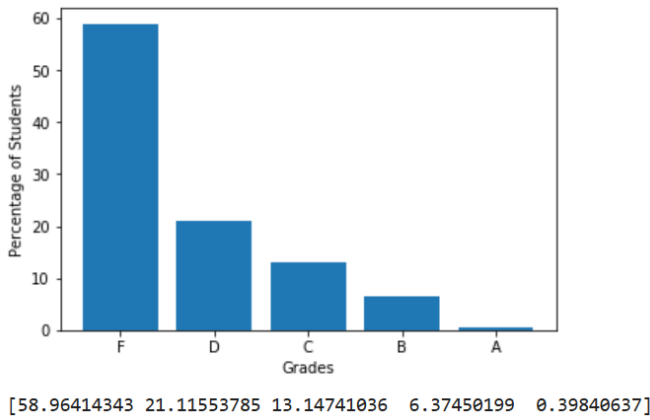[58.96414343 21.11553785 13.14741036 6.37450199 0.39840637]

Fig. 8. Distribution of grades among students without family educational support.

Let us first take a look at the distribution of grades amongst students with and without family educational supprt, Figure 7 and Figure 8. Surprisingly, there is not much difference in the distribution other than the fact that of the students wih family educational support have 2% less F's and 2% more D's. With only two students in the entire data set with a final grade of A, there is one in both groups. The same approach to visualizing these results was used on the following features.

There is more variation when looking at the distribution of grades amongst students with different reasons of choosing their current school. Students who chose their school based on course prefrence had the highest percentage of Fs at 72.222% and lowest percentage of all other grades. There were no A-students in this group nor in the group of students who chose their school because it was close to home. This group had the second highest percentage of Fs at 63.509% and second lowest percentage of all other grades. The students who chose their school based on reputation fared better: 51.007% got Fs, 29.530% got Ds, 14.755% got Cs, 4.027% for Bs, and one student got an A. Those who chose another reason fared even better: 44.755% got Fs, 24.476% got Ds, 16.783% got Bs, and one student got an A.

There is an interesting pattern when looking at the distribution of grades amongst students who go out with their friends at different frequencies. The best grades are found at the second lowest frequency: 48.966% Fs, 21.379% Ds, 15.862% Cs, 13.103% Bs, and 0.690% As (one student). Below this, there is a higher percentage of Fs and Ds and a much lower percentage of Cs (4.167%). There were no As or Bs. Among students who rate the times spent with friends as a 3, 4, or 5, the percentage of students with Fs increased as the amount of time spent with friends increased.

A similar pattern can be seen when looking at the distribution of grades amongst students with different levels of free time. Those with the second lowest amount of free time had the best grades: 42.925% Fs, 26.168% Ds, 18.892% Cs, 10.280% Bs, and one student with an A. At the lowest frequency, the percentage of students with an F increased to 60.000

Finally, lets look at the last influential feature: access to extra paid classes. Those who attended these classes seemed to do worse than their non-attending counterparts. With extra paid classes, there were 69.231% Fs, 20.513% Ds, and 10.256% C's. Without them, there were 56.721% Fs, 22.459% Ds, 13.279% Cs, 7.213% Bs, and 0.328% As.

V. Discussion

As seen in our results, removing unimportant features improved our accuracy. This highlights the importance of asking the right questions when it comes to our childrens education. Some interesting information can be derived from the evaluation of the most influential factors. It came as a surprise that the most influential non-academic feature (family educational support) did not show much of a difference in the distribution of grades. So perhaps this is not a question we should be focusing on? Furthermore, it is interesting to see that those who chose their school based on course preference performed, in general, worse than other students. Initially, one may have thought that those that participate in classes of their choosing would do better. In the future, it might be beneficial to take into consideration their reasoning for picking these courses. If they picked them because they are easier, it may show a lack of motivation. Of all the reasons close to home, school reputation, course preference, and other it is important to notice that unnamed reason produced the best results. What are these other reasons? Initially, we thought basing choice school reputation would indicate a more motivated student and thus produce higher grades. Another finding that defied expectations was the fact that those with access to extra paid class did worse than those without. What does this mean? Perhaps those with access to these classes feel as though they dont need to put in the extra work that their non-attending counterparts do. Also, if parents are paying for these classes, do they feel the need to offer other forms of support? Do this differ from parents who are unable to pay for these classes?

Other intriguing findings can be found when looking at the distribution of grades with the features that describe how often students go out with their friends and how much free time they have after school. When looking at their distributions, there seems to be an ideal frequency for both which is the second lowest. This provides evidence that students do need some free time and social interaction to perform better in school. Going out with friends too much as well as having too much free time seems to be detrimental to student performance. This begs the question, how much is too much? Also, what qualifies as free time after school? Does this account for time spent doing homework?

This bring up a possible change that could be made in the future: the ambiguity of feature descriptions. If these descriptions were more specific, our study would be more reliable and could provide us with more information. Other possibilities include collecting data from more than just two schools, choosing schools with better methods of bookkeeping, and including more grade variation (this study included and abundance of failing students and only 2 A-students). Studies

can be performed looking solely at one influential feature at a time. Influential features do not have to be the only ones analyzed. Our most influential feature did not have a large effect on final grades but alcoholic consumption, while not one of the most influential features, might prove to be more affective.

This study brought to light how some aspects affect student grades, sometimes in ways we would least expect. We hope this study would spark the conversation on how we can better education by looking at the features that affect student performance.

REFERENCES

[1] D. Mitra, "Pennsylvania's best investment: The social and economic benefits of public education." [Online]. Available: https://www.elc-pa.org/resource/pennsylvanias-best-investment-the-social-and-economic-benefits-of-public-education/

[2] "Portugal vs united states education stats compared." [Online]. Available: https://www.nationmaster.com/country-info/compare/Portugal/United-States/Education

[3] T. Miller, "Partnering for education reform," Jul 2011. [Online]. Available: https://www.ed.gov/news/speeches/partnering-education-reform

[4] P. Mohd Arsad, N. Buniyamin, and J.-L. Ab Manan, "A neural network students' performance prediction model (nnsppm)," 11 2013, pp. 1–5.

[5] P. Cortez and A. Silva, "Using data mining to predict secondary school student performance," *EUROSIS*, 01 2008.

[6] "Rectified linear units (relu) in deep learning." [Online]. Available: https://www.kaggle.com/dansbecker/rectified-linear-units-relu-in-deep-learning