

Paper Title*

1st Given Name Surname

dept. name of organization (of Aff.)

name of organization (of Aff.)

City, Country

email address

2nd Given Name Surname

dept. name of organization (of Aff.)

name of organization (of Aff.)

City, Country

email address

3rd Given Name Surname

dept. name of organization (of Aff.)

name of organization (of Aff.)

City, Country

email address

4th Given Name Surname

dept. name of organization (of Aff.)

name of organization (of Aff.)

City, Country

email address

5th Given Name Surname

dept. name of organization (of Aff.)

name of organization (of Aff.)

City, Country

email address

6th Given Name Surname

dept. name of organization (of Aff.)

name of organization (of Aff.)

City, Country

email address

Abstract—This document is a model and instructions for L^AT_EX. This and the IEEEtran.cls file define the components of your paper [title, text, heads, etc.]. *CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract.

Index Terms—component, formatting, style, styling, insert

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? 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.

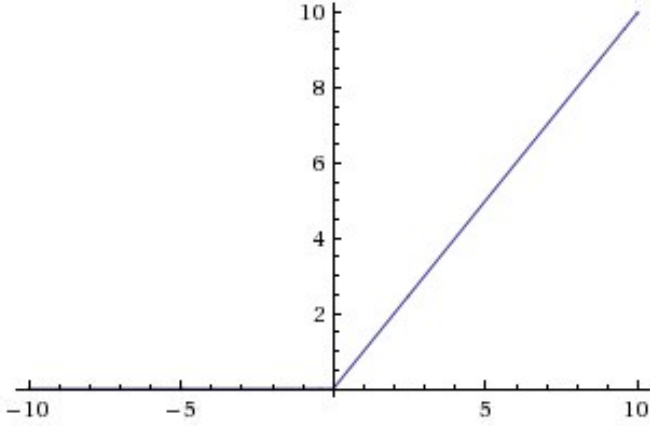


Fig. 1. A graph showing the form of ReLU [6].

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

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

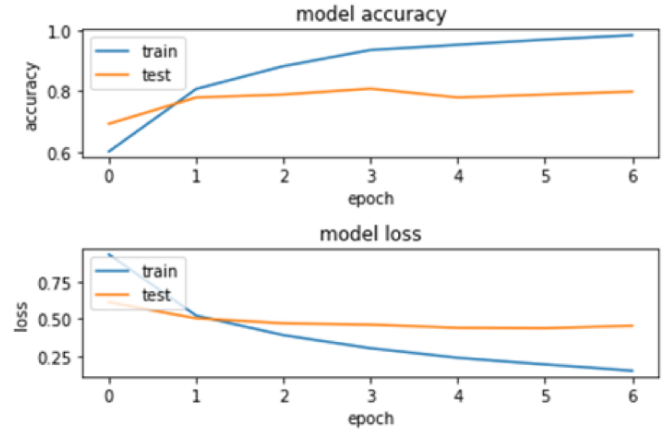


Fig. 2. Initial test results.

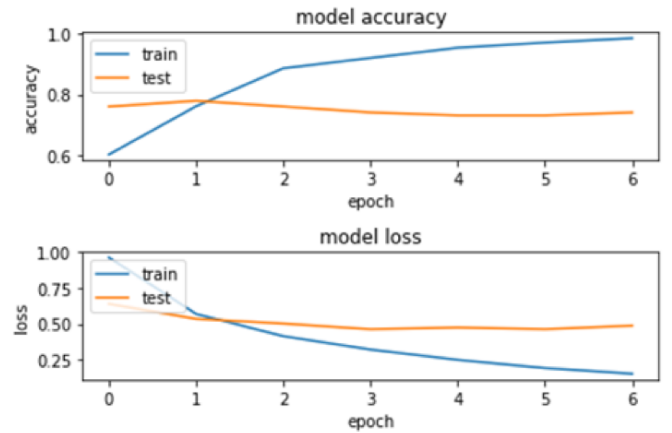


Fig. 3. Test results without grade 3.

in network accuracy. Figure 3 is from the network that trained on the data without the grade three feature.

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.

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.

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

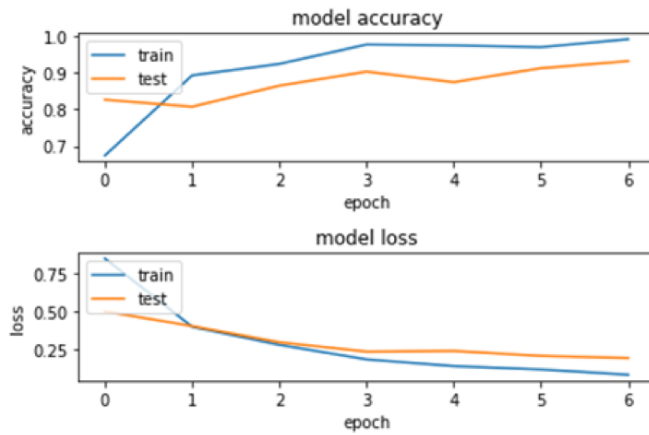


Fig. 4. Test results without absences.

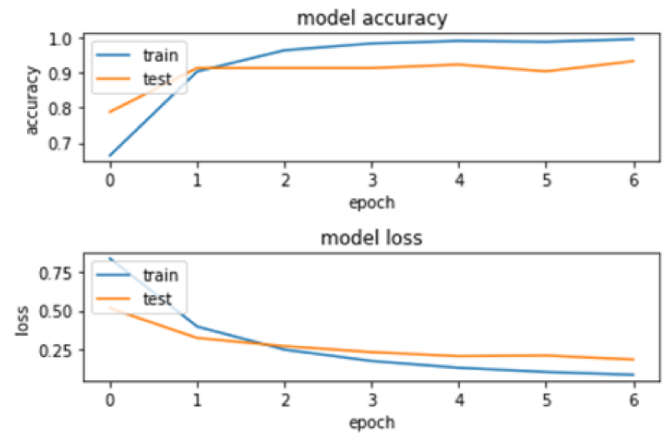


Fig. 6. Test results without travel time.

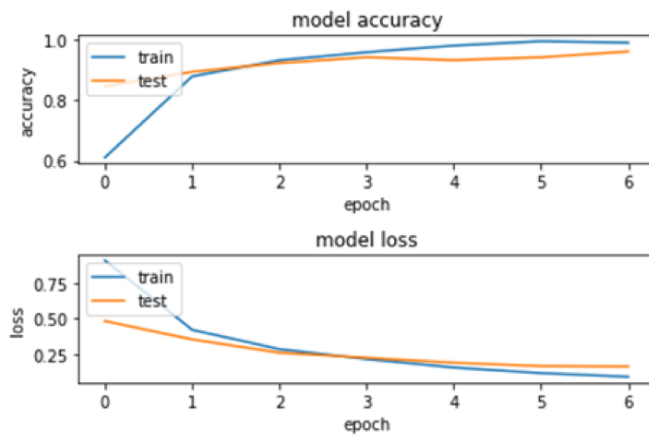


Fig. 5. Test results without age.

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

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

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

V. DISCUSSION

Start typing here [?].

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.