

## Project Proposal – Predicting Student Admission

### **Introduction and Background**

Each year, almost 3.1 million undergraduate students apply to graduate school (Master or Ph.D. programs) yet only 25% of those students are accepted to graduate programs. Many students expend so much time, money, and effort applying to schools only for many to receive a rejection after 6-9 months of waiting. The machine learning model this project aims to construct has the potential to change that. The model will predict the probability that an international student will gain admission to a graduate school. College boards can use this model to predict which students will get into their college. Furthermore, undergraduate students can use this model to apply only to colleges where they have a high chance of acceptance (high chance as defined by the student). This will save countless students time, money, and effort for both parties. If this model is successful at predicting international student admission, it can easily be adapted to domestic students, as the only difference in criteria between the two is the TOEFL requirement for international students.

### **Dataset**

The dataset is a CSV file of international students who applied to the UCLA graduate school. Each row represents one student, and each column represents their admission criteria (GPA, GRE score, their university's ranking, etc). International students only have the TOEFL exam as an extra criterion. The target variable is a floating-point value for the probability of admission, but a neural network will be used on the dataset instead of logistic regression. This is because admission data almost always has nonlinear correlations, and neural networks are better

at handling nonlinearities. Students' admission criteria aren't always linearly correlated to admission; many factors play into admission all at once.

## **Procedure**

Exploratory data analysis (EDA) will first be conducted on the dataset to gain better insight into the dataset and determine how well the model might perform. This will comprise histograms, bar charts, a correlation matrix, and boxplots. Once insight is gained from the EDA step, an artificial neural network model dataset can be trained on the data. The dataset must first be split into training and testing sets, then the model can be trained on the data. Since this is a neural network with a numerical target variable, its performance will be tested by its chosen loss function, the mean squared error. The lower the error, the better the model's predictions for a student's probability of admission.

## **Issues**

A few challenges must be overcome after building the model. First, because the current dataset is limited to 400 rows, more rows will be needed to confirm that the model will have a low error. Second, the current dataset only applies to international students, so a few extra features will be needed in future iterations of the model to apply it to domestic students. Finally, each college has different standards for admission criteria (i.e., an admitted student at one college might not get into another), so this must be taken into account in the model as well.

## **Ethical Concerns**

The project's goal is to save time, money, and effort, but there is one ethical concern. With an accurate prediction of who will gain admission, college boards can pick and choose who will get into their college. This can cause cases where certain demographics are accidentally left out of the admission process, as there is a strong correlation between test scores and student ethnicity. Great care must be taken to ensure a diverse set of students are admitted and no ethnicity is excluded.

## References

Mengash, H. (2020). *Using Data Mining Techniques to Predict Student Performance to Support Decision Making in University Admission Systems*. IEEE. Retrieved on May 14, 2022 from <https://ieeexplore.ieee.org/document/9042216>.

National Center for Education Statistics. (2022). Postbaccalaureate Enrollment. *Condition of Education*. U.S. Department of Education, Institute of Education Sciences. Retrieved May 14, 2022 from <https://nces.ed.gov/programs/coe/indicator/chb>.

Pangburn, D.J. (2019). *Schools are using software to help pick who gets in. What could go wrong?* Fast Company. Retrieved on May 14, 2022 from <https://www.fastcompany.com/90342596/schools-are-quietly-turning-to-ai-to-help-pick-who-gets-in-what-could-go-wrong>.