**MBA Admission Insights and Predictions:**

**Analyzing and Modeling Admission Outcomes**

**Project Category**: Education Analytics

**Team Information**:

* Shiling Zheng (sz298)
* Zhehan Wang (zw311)
* Xinyu Qian (xq44)
* Xiaozhu Yu (xy165)
* Zairan Xiang (zx122)

**Problem Statement**

The problem we are addressing focuses on understanding the factors that influence MBA admissions, particularly examining which features—such as GPA, GMAT scores, work experience, gender, race, and international status—most impact admission decisions. A key point of interest is whether the admissions process is fair for international students and applicants from diverse racial backgrounds. Understanding these factors is crucial for both educational institutions striving for equitable admissions policies and applicants seeking to navigate the competitive process.

**Challenges and Dataset**

However, when collecting features of MBA admission data, some important factors such as extracurricular activities and achievements couldn't be included, because they are difficult to encode into numerical values. Additionally, we couldn't exclude bias from the collected data, which would add noise to the data and lead to a less reliable prediction model. Finally, we find our MBA admission prediction dataset (Class 2025) on Kaggle, which includes 10 features and 6,194 data pieces.

**Dataset Link** https://www.kaggle.com/datasets/taweilo/mba-admission-dataset?select=MBA.csv

**Method/Algorithm**

Before applying machine learning model to our data, we will perform some data preprocessing such as encoding categorical variables using one-hot encoding, scaling numerical features to ensure they are on similar scale, handling imbalance classes and missing data. We will then use Logistic Regression as our baseline model. It is interpretable, allowing us to understand how each feature like GPA and GMAT influences the admission decision. Next, we will explore several different models from the sklearn to improve the baseline model's performance. These models will also help us determine the importance of different features in the admission process.

* Random Forest Classifier: an ensemble method that can capture complex relationships between features. It will also allow us to identify important features using feature importance scores.
* Support Vector Machine (SVM): A robust model for classification tasks, especially in high-dimensional spaces. We'll experiment with kernel functions (linear, radial) to find the best fit for our data
* K-Nearest Neighbors (KNN): A simple clustering algorithm. We will tune the number of neighbors and evaluate whether a non-parametric method like KNN performs better than others.
* Neural Network (MLPClassifier): A simple multi-layer perceptron can be used as a final model to see how well a NN captures complex patterns in the data.

**Literature Review**

Research has identified several key factors influencing MBA admissions, such as GPA, GMAT scores and working years [1]. It was found that fluency in spoken English played a crucial role in MBA admissions for non-native English-speaking students [2]. Studies have also shown that classification methods may be more suitable for the problem of admission prediction, while the results of regression and neural network models only perform at least as well as classification models [3].

**Evaluation**

We will evaluate the model's performance using both qualitative and quantitative methods. Key performance metrics such as accuracy, precision, recall, and F1-score will assess the model's effectiveness in classifying admissions. We will also use fairness metrics to assess whether the models treat international students and applicants from different racial backgrounds fairly. Statistical tests like chi-square will examine feature significance, and confusion matrices will help identify misclassifications. Visual aids, including feature importance plots and ROC curves, will provide insight into how factors like GPA, GMAT, and work experience influence predictions, uncovering biases or trends in admissions outcomes.

**Reference**

[1] Pratt, W. R. (2015) Predicting MBA Student Success and Streamlining the Admissions Process, *Journal of Education for Business*, 90(5), pp. 247–254. <https://doi.org/10.1080/08832323.2015.1027164>

[2] Wu, X., Wu, J. (2020) Criteria evaluation and selection in non-native language MBA students admission based on machine learning methods. *J Ambient Intell Human Comput* 11, 3521–3533. <https://doi.org/10.1007/s12652-019-01490-0>

[3] Wilson, R. L., & Hardgrave, B. C. (1995). Predicting Graduate Student Success in an MBA Program: Regression Versus Classification. *Educational and Psychological Measurement*, *55*(2), 186-195. <https://doi.org/10.1177/0013164495055002003>