Report

### **Deep Learning Project Report**

**Team Overview**This project was undertaken by Group 28, consisting of members Wanting Mao, Tianyun Yuan, Max Xu, and Wenqi Jia. The team collaborated to apply deep learning techniques to tackle a critical problem in academic course selection, leveraging historical data to improve the decision-making process for students.

### **Problem Motivation**

Selecting appropriate courses is a crucial decision for students, yet it is fraught with challenges. Students often struggle to align course objectives with their goals and workloads, and harsh grading curves can significantly impact their GPAs. Existing course review systems offer historical data but lack reliable predictive capabilities that could help students anticipate their performance. To address this, our project aimed to develop a predictive model using deep learning techniques. By doing so, we hope to empower students to make informed course choices, reduce stress, and foster a more fulfilling educational experience.

### **Data and Features**

The dataset for this project presented several challenges, including missing, imbalanced, and sparse data, as well as the complexities associated with categorical data. Missing data was addressed by introducing a new category, "NA," for cases where the primary instructor information was unavailable. To manage the long-tailed distribution of grades—where the majority fell between A and B—we employed sparse matrix structures to optimize computational efficiency. Categorical features, such as course subjects and schedule types, were transformed using one-hot encoding to ensure they were machine-readable. The features included subject, course level, schedule type, primary instructor, and class size. Our target metric was the average GPA, calculated as a weighted average based on the grade distribution within each class.

### **Modeling Approach**

We implemented two models to predict average GPAs: a traditional linear regression model and a deep learning model. The linear regression model was developed using Scikit-Learn and optimized to minimize the residual sum of squares. Our deep learning model, built with PyTorch, utilized an architecture consisting of linear layers, ReLU activations, layer normalization, and dropout layers. Hyperparameter tuning was conducted to optimize the performance of the deep learning model, with experiments testing various combinations of optimizers (Adam, SGD, RMS Prop), batch sizes (8-128), and learning rates (0.1-0.001).

### **Results and Analysis**

Our results showed that the deep learning model significantly outperformed the linear regression model, achieving an R2R^2R2 score of 0.52. This indicated that the deep learning model was better at capturing the complex relationships and interactions in the data. To enhance the interpretability of our results, we conducted feature importance analyses using SHAP values and feature permutation techniques. SHAP analysis revealed that instructors had the most significant impact on predictions, followed by subject type and course level. Feature permutation further validated this finding, as shuffling the instructor feature resulted in the largest drop in model performance. These analyses underscored the importance of instructors in predicting average GPAs, highlighting their influence on academic outcomes.

### **Key Insights**

Through this project, we identified several critical insights. First, instructors play a pivotal role in shaping GPA outcomes, as demonstrated by their dominant impact in both SHAP and permutation analyses. Additionally, subject type and course level were also influential, albeit to a lesser extent. Our deep learning model successfully captured intricate patterns and relationships that traditional models could not, making it a robust tool for predicting student performance. This predictive capability holds significant practical value, enabling students to make data-driven course selections and mitigate the stress associated with academic uncertainty.

### **Conclusion**

In conclusion, our deep learning model provides an effective solution for predicting average GPAs, helping students make more informed course decisions. By addressing the challenges of missing and categorical data, optimizing hyperparameters, and conducting rigorous feature importance analyses, we developed a model that is both accurate and interpretable. This project not only demonstrates the power of deep learning in tackling real-world problems but also underscores its potential to enhance the academic experience for students. We hope that our work will pave the way for more personalized and data-driven approaches in education.

Script

### **Transcript for Deep Learning Project Presentation (7-Minute Read)**

**[Slide 1: Title Slide] Wenqi**Welcome to our presentation on leveraging deep learning to predict course outcomes. Today, we’ll explore how we addressed the challenges students face when choosing courses and how our model can make their decisions more informed.

TABLE OF CONTENTS

THIS IS WHO WE ARE

We are Group 28: Wanting Mao, Tianyun Yuan, Max Xu, and Wenqi Jia.

**[Slide 2: Problem Motivation] Wenqi**Choosing the right course is essential but often difficult for students. Misaligned goals, harsh grading curves, and a lack of predictive tools in existing course reviews create stress and uncertainty. We aim to solve this by predicting average GPAs using historical data, empowering students to select courses that best fit their needs while minimizing academic stress.

**[Slide 3: Objective and Impact] Wenqi**Our plan is to leverage deep learning and historical academic data to provide reliable GPA predictions. This will help students make better decisions, reduce stress, and create a more personalized academic experience. We want to foster an environment where students feel confident about their choices, leading to improved academic outcomes.

**[Slide 4: Data Challenges]**Our dataset posed several challenges:

1. Missing primary instructor data was resolved by introducing a new "NA" category.
2. Imbalanced data, where most grades fall between A and B, required sparse matrix structures for computational efficiency.
3. Categorical data, such as course subject and schedule type, needed one-hot encoding due to the lack of inherent order.

**[Slide 5: Features and Target Metric]**Kes include course subject, level, schedule type, primary instructor, and class size. Our target metric, the average GPA, is computed as a weighted average based on the number of students in each grade category.ey featur

**[Slide 6: Modeling Approach]**We implemented two models:

1. A linear regression model built using Scikit-Learn, designed to minimize residual errors.
2. A deep learning model developed in PyTorch, featuring linear layers, ReLU activations, layer normalization, and dropout layers. Hyperparameter tuning included variations in optimizers, batch sizes, and learning rates.

**[Slide 7: Results and Performance]**Our deep learning model significantly outperformed the linear regression model, achieving an R2R^2R2 score of 0.52. This shows that our model effectively captures complex patterns in the data, offering reliable predictions.

**[Slide 8: Feature Importance with SHAP Values] Max**To enhance interpretability, we analyzed feature importance using SHAP values. This approach quantifies the contribution of each feature to model predictions. The results highlighted that instructors have the largest impact, followed by subject type and course level.

**[Slide 9: Feature Permutation Analysis] Max**We validated these findings using feature permutation, which involves shuffling feature values to measure the impact on performance. Again, the instructor feature caused the largest drop in model accuracy, reinforcing its importance in GPA predictions.

**[Slide 10: Key Insights] Max**Our analysis revealed that:

1. Instructors are the most critical feature in predicting average GPAs.
2. Deep learning effectively captures relationships that traditional models cannot.
3. This tool provides practical value for students, enabling data-driven course selection.

**[Slide 11: Conclusion] Max**In summary, our deep learning model offers a robust and interpretable solution for predicting course outcomes. By addressing data challenges and conducting rigorous analyses, we developed a tool that empowers students to make informed decisions, reducing stress and improving their academic experience. We believe our approach sets a strong foundation for personalized education in the future.

**[Slide 12: Thank You]**Thank you for your time. We’re happy to answer any questions about our project.