Education and Student Outcomes
Grade Range from Student Performance/Interaction Logs

Jacob Bettinger **Tyler Wolstone** **Saitej Bommidi** **Jason Manhart** **Joseph Miller**
College of Engineering, Michigan State University
`bettin12@msu.edu, wolstone@msu.edu, bommidis@msu.edu`
`manhartj@msu.edu, mill3642@msu.edu`

Abstract

Predicting student academic performance is an important problem in education, especially since many institutions and courses tend to rely on online learning methods. A benefit of these methods is that they collect detailed data that reflect how students engage with course materials over time. In this project, our aim is to predict whether a student will pass or fail using the UCI Student Performance Dataset and Open University learning analytics data. We plan to compare several baseline learning models, including logistic regression and decision trees, with a neural network that will capture patterns within student activity. We will analyze which student behaviors are most predictive of academic success and discuss how findings could be used responsibly in educational settings.

1 Introduction and Motivation

The problem we are addressing is whether an individual is predicted to pass or fail and whether that individual can take meaningful action to change this outcome. This is interesting from a machine learning perspective because this model does not just have to be for students. This can be used in many other fields, not just a student prediction. This is also important for many reasons, one of them being the ability to be proactive to prevent students from failing. If there is a student who has the ability to pass the class, but is going out often and not studying as much as they need to be. If a student's grades start to slip and that student starts to fall behind, the intervention that they receive is reactive. This student then has to get themselves out of the hole they are in, and sometimes their effort is not enough for them to pass. This project is guided by the question of what factors contribute most to predicting student passing or failing outcomes.

2 Related Work

Prior work in educational data mining has examined student performance prediction from two main angles, structured demographic features and behavioral engagement data. Cortez and Silva (2008) demonstrated that traditional supervised learning models can achieve strong predictive accuracy using tabular features from the UCI Student Performance dataset. More recent work using datasets such as OULAD has shown that incorporating engagement patterns from virtual learning environments allows for earlier and more accurate identification of at risk students (Kuzilek et al., 2017). Deep learning approaches extend this by modeling temporal learning behavior to capture patterns that simpler models may miss. While these studies establish the value of both structured and behavioral predictors, they often analyze them in isolation. Our project builds on this foundation by systematically comparing baseline and neural models across both demographic and engagement based features, with the goal of more clearly understanding how feature representation and model complexity interact to influence predictive performance.

3 Dataset

- **Source and Size.** We will use the Student Performance dataset from the UCI Machine Learning Repository. The dataset consists of two subsets collected from Portuguese secondary schools: `student-mat.csv`, which contains records for 395 students enrolled in Mathematics courses, and `student-por.csv`, which contains records for 649 students enrolled in Portuguese language courses. Each record corresponds to an individual student and includes demographic, family, social, and school-related attributes along with academic outcomes.

- **Type of Data.** The dataset is tabular and contains a mix of categorical and numerical features. Categorical variables include attributes such as school, gender, and internet access, while numerical variables include age, weekly study time, and number of absences.

- **Labels and Prediction Target.** The prediction task is binary classification, where the goal is to predict whether a student passes or fails a course. The dataset provides a final grade variable (`G3`) ranging from 0 to 20. Following standard practice, we define a passing grade as $\texttt{G3} \geq 10$ and a failing grade as $\texttt{G3} < 10$.

- **Preprocessing.** We will convert the final grade into a binary pass/fail label. Categorical variables will be encoded using one-hot encoding, and binary yes/no features will be converted into 0/1 values. Numerical features will be scaled as needed to support model training.

- **Ethical Considerations.** This dataset is appropriate for studying student performance as it captures a range of academic and background factors related to learning outcomes. However, predictions about student success can be sensitive and may reflect underlying biases. In this project, our goal is not to make high-stakes decisions, but to explore how early indicators of academic difficulty can be identified. We will carefully consider the role of demographic features and discuss the responsible use and limitations of predictive models in educational settings.

4 Methodology

We model student performance prediction as a binary classification task, where the objective is to predict whether a student will pass or fail a course based on demographic and academic features. We first implement several baseline supervised learning models, including Logistic Regression, Support Vector Machines (SVM), and Naive Bayes. These models are well suited for tabular data and provide interpretable and reliable baselines for comparison.

In addition to the baseline approaches, we implement a neural network model using a Long Short-Term Memory (LSTM) architecture. Although the dataset is tabular, academic features can be organized to reflect progression over the course of the term, allowing the model to learn relationships between different indicators of student performance. The LSTM output is passed through a fully connected layer with a sigmoid activation to produce a pass or fail prediction.

All models use the same preprocessed input features. Categorical variables are one-hot encoded, yes/no features are converted to binary values, and numerical features are normalized. Model hyperparameters are tuned using a validation set to ensure fair and consistent comparisons.

5 Evaluation Plan

We evaluate model performance using standard binary classification metrics, including accuracy, precision, recall, and F1-score. Accuracy measures overall correctness, while precision and recall help evaluate how well the models identify students who are at risk of failing. The F1-score is used to balance precision and recall.

The dataset is split into training, validation, and test sets using an 80/10/10 split. The training set is used to fit the models, the validation set is used for hyperparameter tuning, and final results are reported on the test set. We also perform basic error analysis to examine common misclassification cases and briefly discuss potential fairness concerns related to demographic features.

6 Team Members and Responsibilities

- **Jacob Bettinger**: Model implementation, experiments
- **Tyler Wolstone**: Data preprocessing, evaluation
- **Saitej Bommidi**: Literature review, writing
- **Jason Manhart**: Analysis, visualization
- **Joseph Miller**: Presentation and report coordination

References

Paulo Cortez and Alice Silva. 2008. Using data mining to predict secondary school student performance. In *Proceedings of the 5th Annual Future Business Technology Conference*, Porto, Portugal.

Jakub Kuzilek, Martin Hlosta, and Zdenek Zdrahal. 2017. Open university learning analytics dataset. *Scientific Data*, 4:170171.

(Cortez and Silva, 2008) (Kuzilek et al., 2017)