Prediction on Student's Dropout and Academic Success

Neural Networks and Ensemble Methods

Utsav Shakya

COMP 4980: Machine Learning

8/12/2025

Introduction

The goal of this project is to get hands deep into machine learning, exploring different application and component of machine learning to analysis data, explore and developing a model. For this project, I have learned and performed PCA for feature reduction, conducted decision tree analysis, implemented a Neural network model and use of ensemble methods (gradient boosting and adaboost). I compared the performance of my MLP, against tree based boosting approaches (Gradient boosting and ada boosting) for performing classification. This ***Link*** will take you to my code in Collab.

Data Description

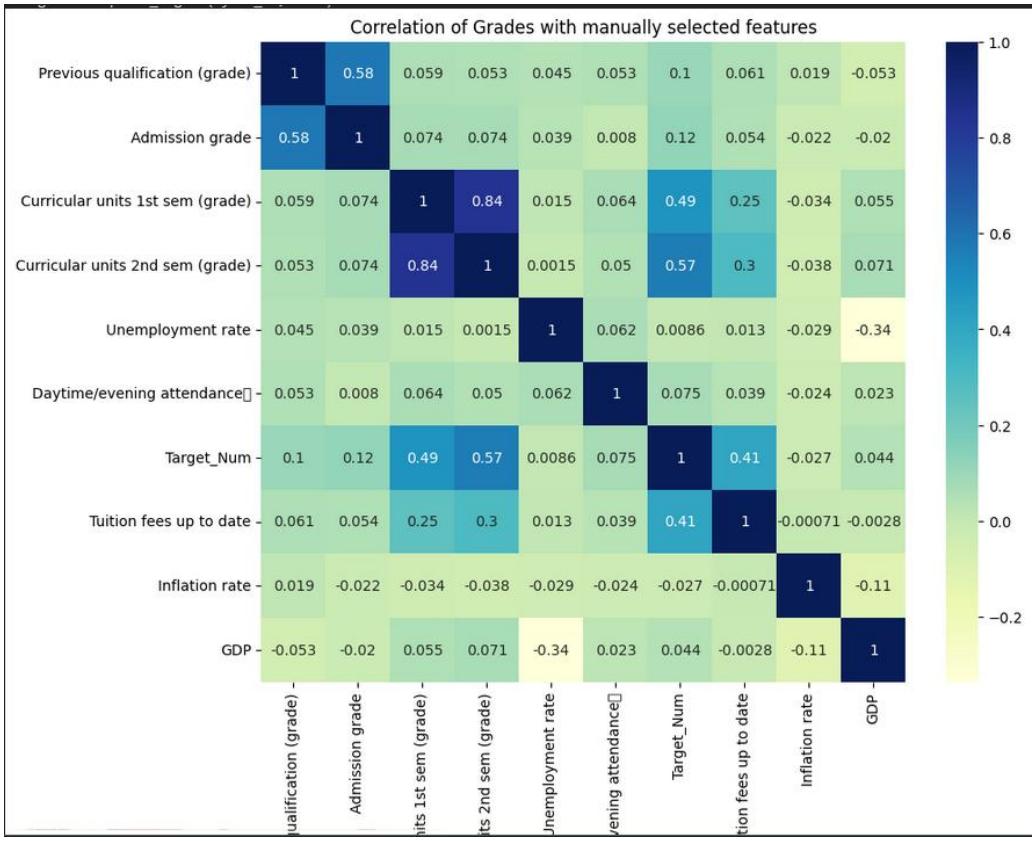
I got my dataset from UCI library called "Predict Students' Dropout and Academic Sucess". The dataset was created to help reduce academic dropout in higher education by identifying failing students early. This dataset is supported by program SATDAP - Capacitação da Administração Pública, Portugal.

This dataset is a tabular data (CSV format) and contains 4424 data (rows) related to individual student's achievement in their high school with 36 features (eg: course, grades etc). This dataset contains mix data related to academic performance and socio-economic. The target of this dataset is divided into three sections: Dropout, Enrolled and Graduate. This dataset has been thoroughly checked by the authors and contains no missing values (I also checked manually).

Data Analysis

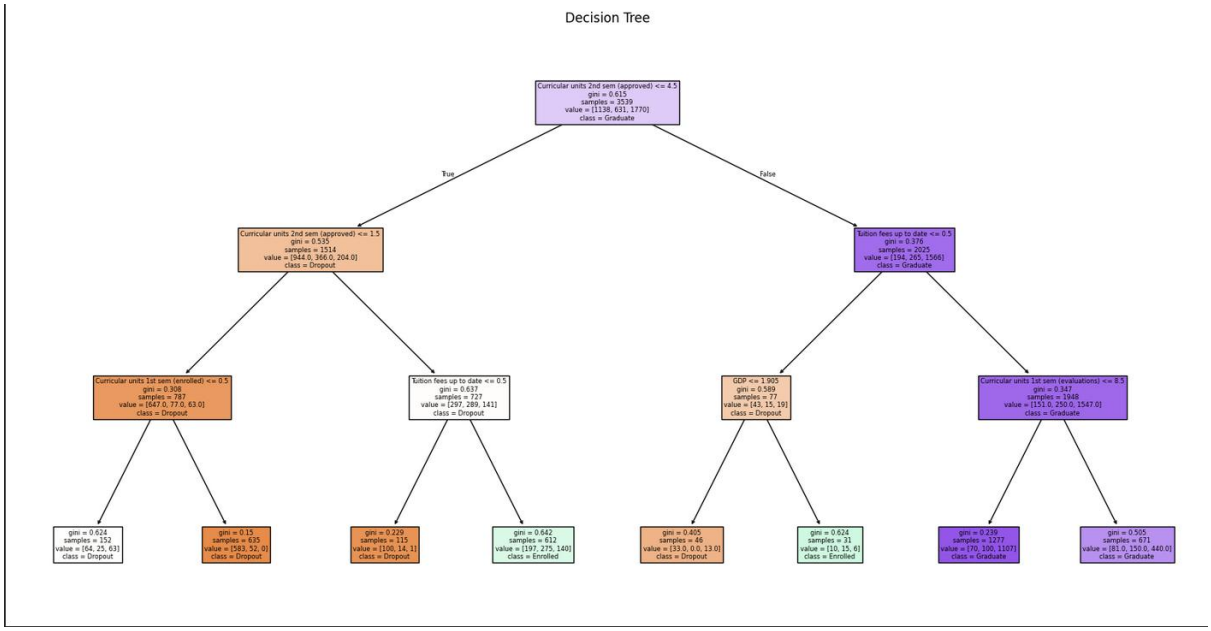
The first thing I did after performing data exploration is that I made a copy of my data and added a new column called 'Target_num', which basically maps dropout: 0, enrolled: 1 and graduate: 2, this way my target is not string but int, which would make life easier for classifying. Then after checking the description of my data, I found out that the average age of students at enrollment is around 23 years old with average previous qualification grade of 132 (current being 126). The grade range for qualification is from range 0-190 whereas the grade range for academic grades for semesters is on a scale of 0-20. This shows that the data are varied, and will need feature scaling.

I also manually selected features for heatmap that I assumed would have high correlate with features specially, with target_num. For here I found out that "Curricular units 2nd sem (grade)" and "curricular units 1st sem (grade)" had the highest correlation with our target (meaning factor related to academic correlated more with our target). Factors such as GDP and inflation rate (economic factors) were the least correlated with our target.



Data Exploration

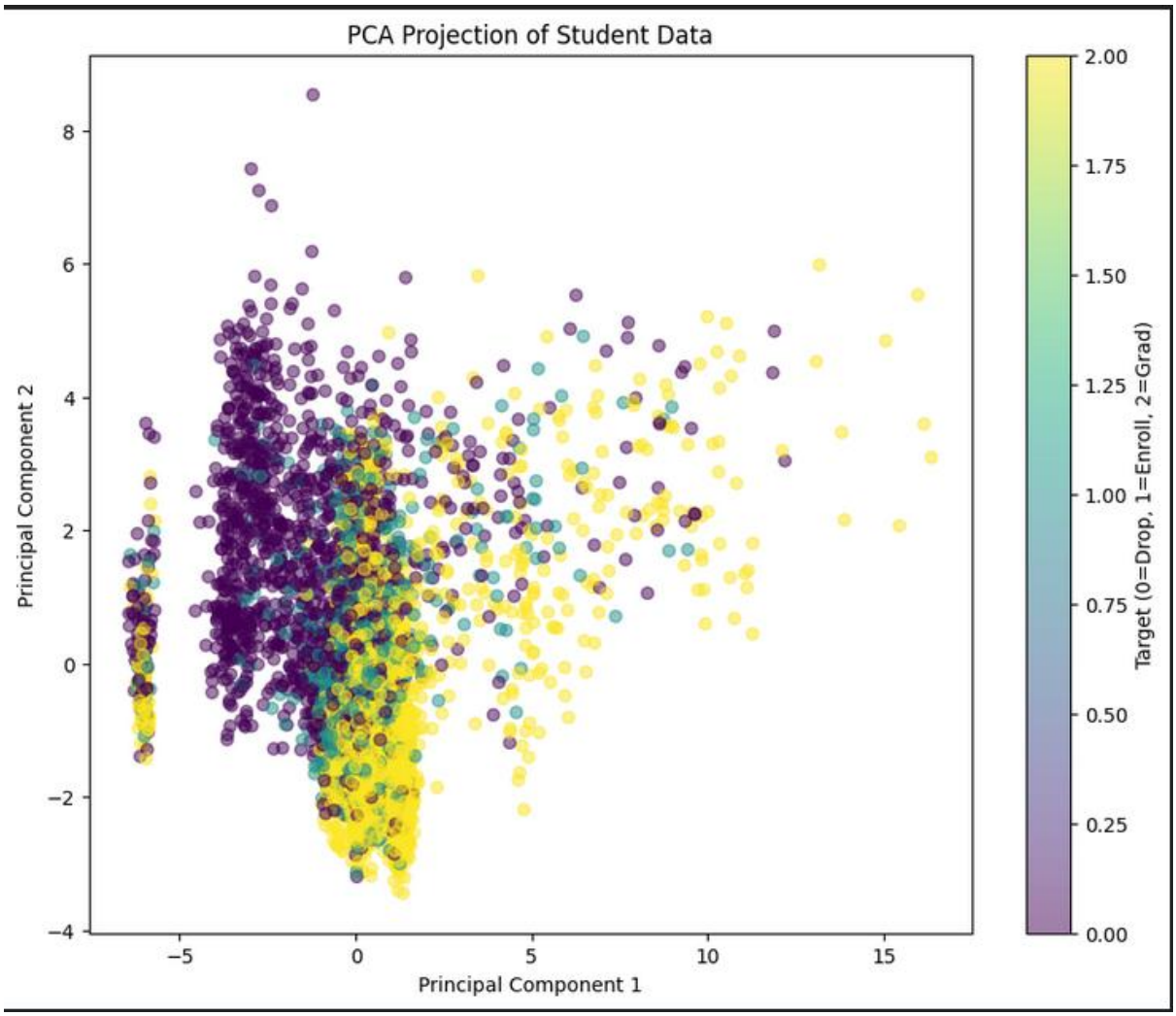
For data exploration, I first perform a simple decision tree of only 3 depths (any more than that was causing too hard to read). The decision tree revealed that curricular unit in the 2nd semester is the most influential for student's outcome. Student having kiw academic engagement and are not up to date tuition fee are likely to drop out than that with strong grade with up-to-date tuition (graduate or enrolled).



After than I performed PCA to check if I could reduce the features while having 96% variance in the data. After performing PCA, the data required 27 principal components to explain 96% of the variance in the data. This shows us that the data is spread out and not just concentrated in few components (even though grades did seems have higher contribution).

For visualization, I performed a scatter plot for the first two principal components (PC1, and PC2)

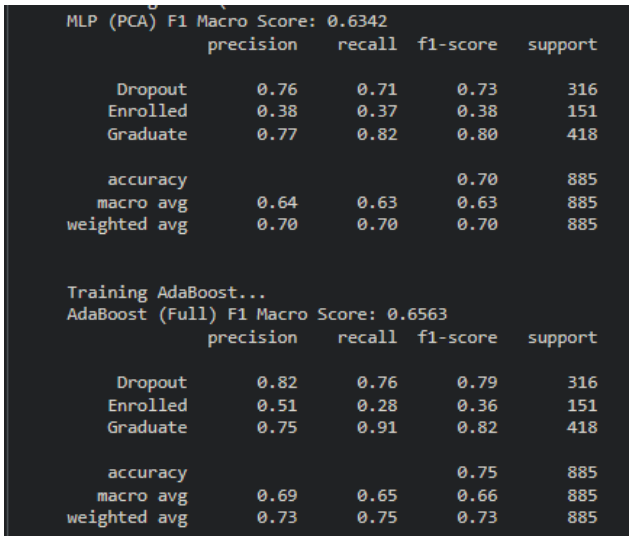
PC1 is mostly affected by the student's academic success as where as PC2 contains mostly dominated by socioeconomic. This shows us that student academic success is more influential for a student success/dropout than financial or personal status (age at enrollment, marital status). The AI did suggest me a scatterplot for PC1 vs PC2, but I did not properly understand it other than the fact that they have lots of overlap, specially for graduate (yellow), as they are highly concentrated in one small area compared to dropped (purple).



Experimental Methods, Results & Analysis

For my experiment , I initialized hypothesis that “MLP using PCA-reduced feature would outperform adaBoost for classification”, but my hypothesis was proven wrong.

Firstly, I had 3 sets of train data; PCA and fully scaled, non scaled data, and scaled data (but no PCA). Then I performed a simple MLP using PCA and adaBoost using non scaled original data’s. I got a score of 0.63 for MLP whereas adaBoost got a score of 0.65. both performed similar for dropout and graduate but MLP struggled way more with enrolled compared to that of AdaBoost as shown in diagram below.

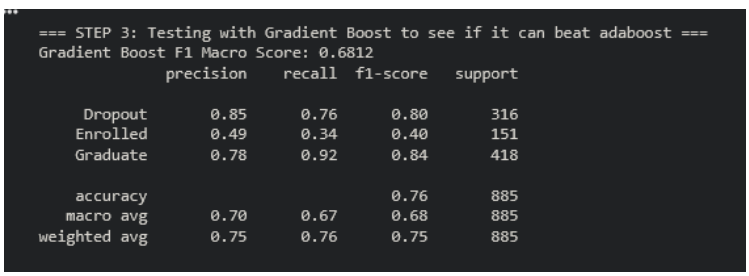


```
MLP (PCA) F1 Macro Score: 0.6342
              precision    recall  f1-score   support

     Dropout       0.76      0.71      0.73       316
    Enrolled       0.38      0.37      0.38       151
    Graduate       0.77      0.82      0.80       418

    accuracy                           0.70       885
   macro avg       0.64      0.63      0.63       885
weighted avg       0.70      0.70      0.70       885


Training AdaBoost...
AdaBoost (Full) F1 Macro Score: 0.6563
              precision    recall  f1-score   support

     Dropout       0.82      0.76      0.79       316
    Enrolled       0.51      0.28      0.36       151
    Graduate       0.75      0.91      0.82       418

    accuracy                           0.75       885
   macro avg       0.69      0.65      0.66       885
weighted avg       0.73      0.75      0.73       885
```

To improve my score I then performed a simple MLP using scaled data (no PCA), which surprising improved the score (just slightly), jumping from 0.63 to 0.64 which is a good progress but still not better than AdaBoost with 0.65.

Now I improved my MLP added more layers and other instructions, and I tested out with both PCA and non PCA scaled data. I got a score of 0.63 with PCA on updated MLP where as I got 0.64 with fully scaled all features data (got better by 0.0010) which is still less than AdaBoost. This suggests that Tree-based ensemble are better for this data than MLP.

For final, I tested out using Gradient Boost, and to check if I would get a better score than AdaBoost on full scaled dataset. I got a score 0.68, which successfully beat not only my MLP but also AdaBoost. Other than enrolled, Gradient Boost outperformed AdaBoost.



```
=== STEP 3: Testing with Gradient Boost to see if it can beat adaboost ===
Gradient Boost F1 Macro Score: 0.6812
              precision    recall  f1-score   support

     Dropout       0.85      0.76      0.80       316
    Enrolled       0.49      0.34      0.40       151
    Graduate       0.78      0.92      0.84       418

    accuracy                           0.76       885
   macro avg       0.70      0.67      0.68       885
weighted avg       0.75      0.76      0.75       885
```

After performing my experiment, my hypothesis was proven wrong, as MLP with PCA is not better than AdaBoost tree classifier (and certainly does not perform better than Gradient Boost). The final score I stuck with was 0.68 as that is the highest I could achieve. The score would have been much better if I stuck with binary classification (graduate, dropout), but my being multi classification, the highest score I could achieve was 0.68. I used macro-f1 to record the score **as** macro f1 score returns average without considering the proportion for each label in the dataset. I used Stratified Cross-validation because I am performing classification, and that my dataset is imbalanced.

This tool can be used in most schools and high schools, which can be used to a warning system that would alert the teachers (or parents), about their students if they are not able to perform well, and prevent students from failing/dropping out by having a proper counseling or other activities that could help the student salvage his later academic performance.

Reference

1. Stack Overflow. *Macro vs micro vs weighted vs samples F1 score* [Internet]. 2019 Apr 18 [cited 2025 Dec 8]. Available from: https://stackoverflow.com/questions/55740220/macro-vs-micro-vs-weighted-vs-samples-f1-sc
2. Sun P. *Stratified K-Fold Cross Validation: When Balance Matters* [Internet]. Medium. 2019 Apr 18 [cited 2025 Dec 8]. Available from: https://medium.com/@pacosun/stratified-k-fold-cross-validation-when-balance-matters-c28b9a7cb9bc