Predicting Student Outcomes with Machine Learning

By: Elijah Lopez

November 15, 2024

Summary & Recommendations

1. Which factors impact student outcomes the most?
2. How can we flag at-risk students and offer support?

1

Encourage students to take daytime classes.

2

Research effects of displacement.

3

Build awareness of scholarship opportunities.

4

Monitor student performance across semesters.

5

Use XGBoost to flag at risk students for intervention.

Source: CleanPNG

Outline

1. Business Problem
2. Data & Methods
3. Statistical Results
4. Machine Learning
5. Conclusions

Business Problem

Students

Institutions

Donors

Metric of Success: Student Outcomes (Dropout, Enrolled, Graduate)

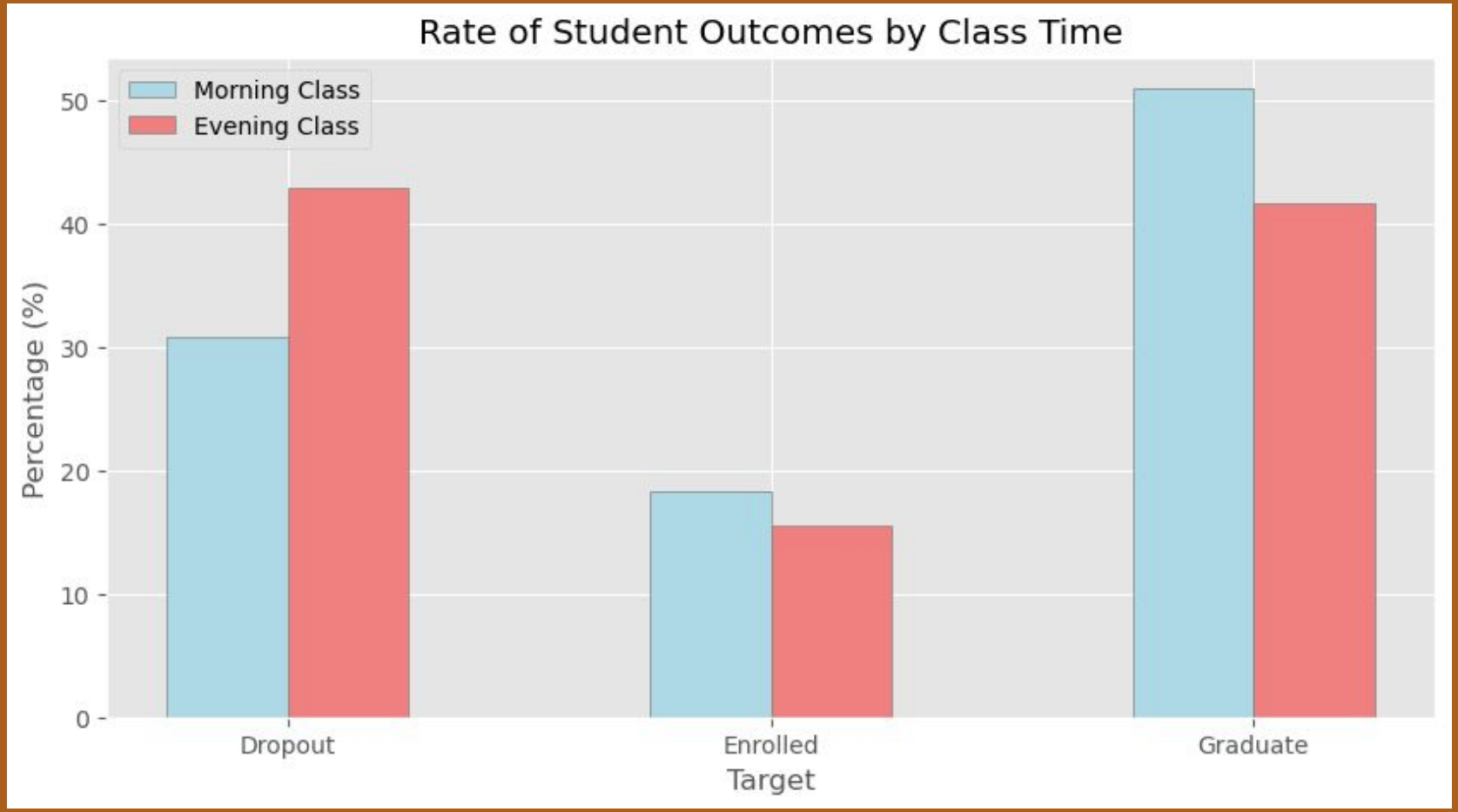
Image Source: CleanPNG

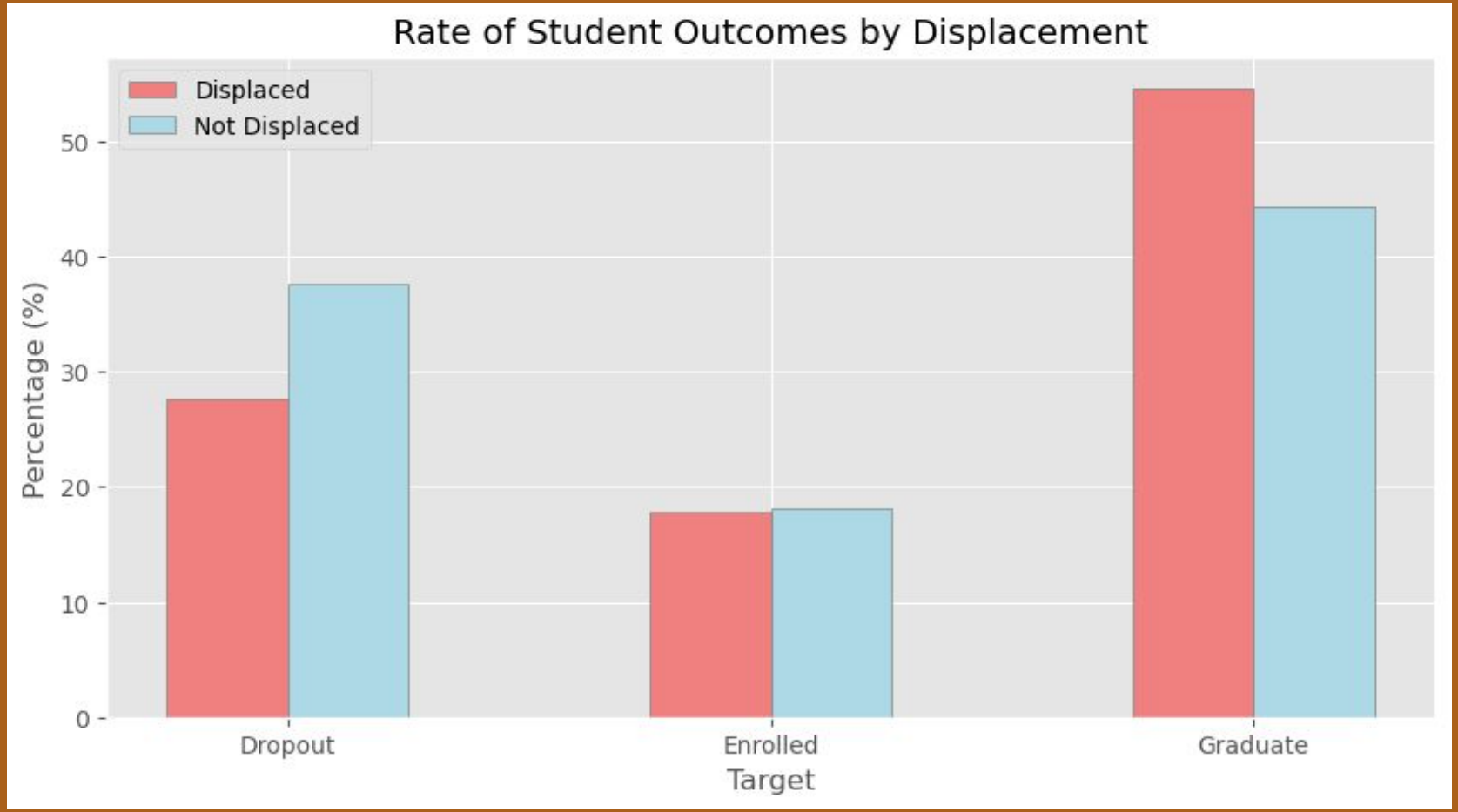
Image Source: CleanPNG

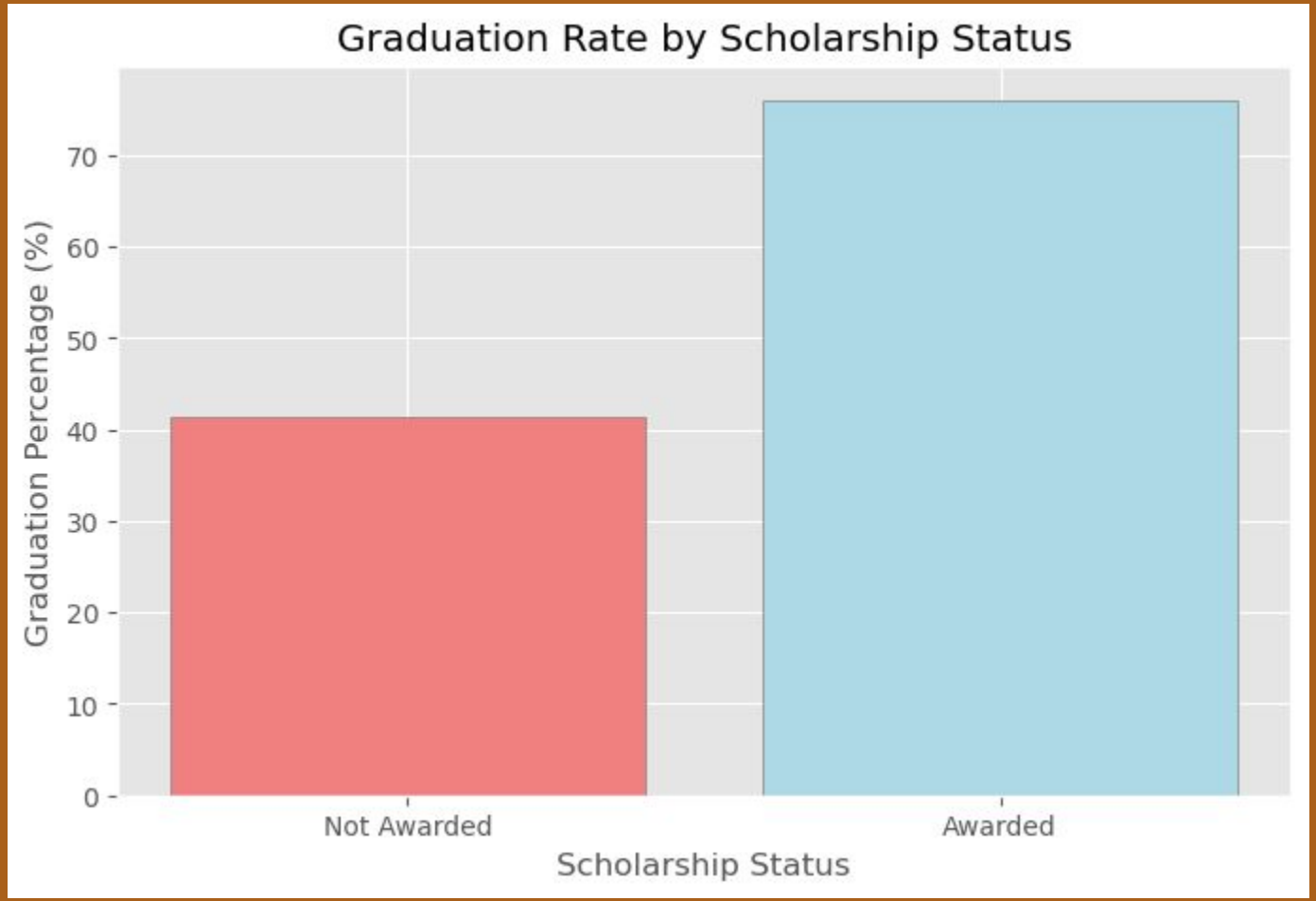
Data & Methods

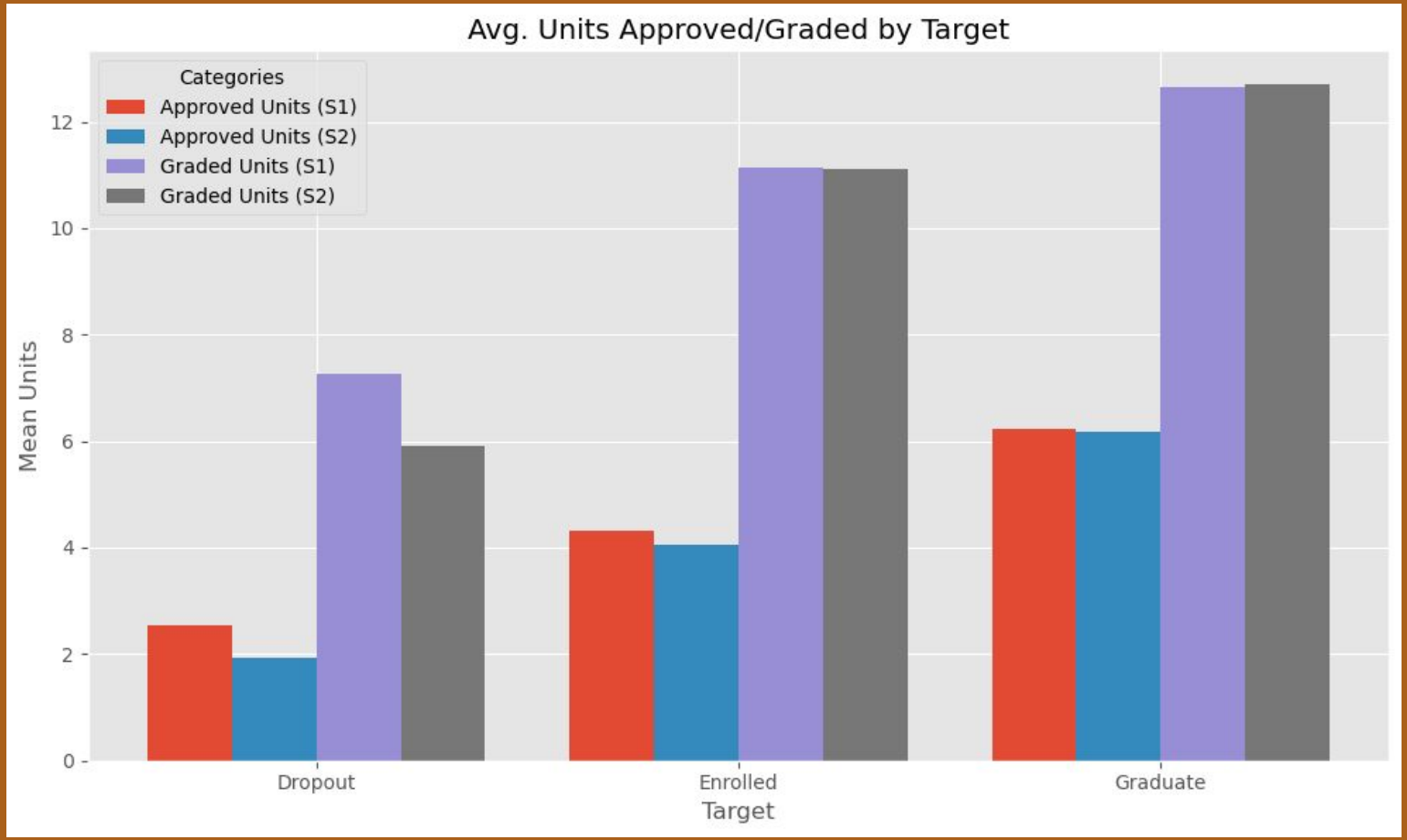
Data: UCI ML Repo

- **~4,400 students**
- **Measured on 35 features + outcome**
  - **Demographics (parents & students)**
  - **Student performance**
- **Inference & Statistical Testing**
- **Machine Learning**


Rate of Student Outcomes by Class Time
Morning Class
Evening Class
Percentage (%)
0
10
20
30
40
50
Dropout
Enrolled
Graduate
Target


Rate of Student Outcomes by Displacement
Displaced
Not Displaced
Percentage (%)
0
10
20
30
40
50
Dropout
Enrolled
Graduate
Target


Graduation Rate by Scholarship Status
Graduation Percentage (%)
70
60
50
40
30
20
10
0
Not Awarded
Awarded
Scholarship Status


Avg. Units Approved/Graded by Target
Categories
Approved Units (S1)
Approved Units (S2)
Graded Units (S1)
Graded Units (S2)
Mean Units
0
2
4
6
8
10
12
Dropout
Enrolled
Graduate
Target

Machine Learning Method

Model Selection

- Logistic Regression
- K-Nearest Neighbors
- Decision Tree
- XGBoost
- Support Vector Machine

Data Handling

- Scaling
- Target encoding
- Recoding
- One hot encoding

Other Processing

- Resampling
- Feature Selection
- Hyperparameter Tuning

Machine Learning Results

79 %

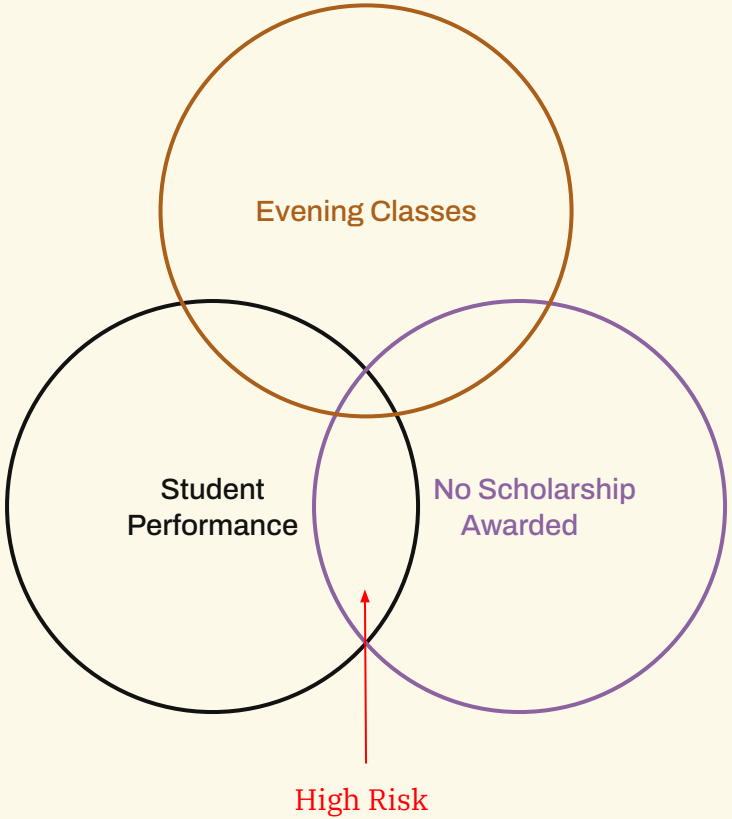
Training Accuracy

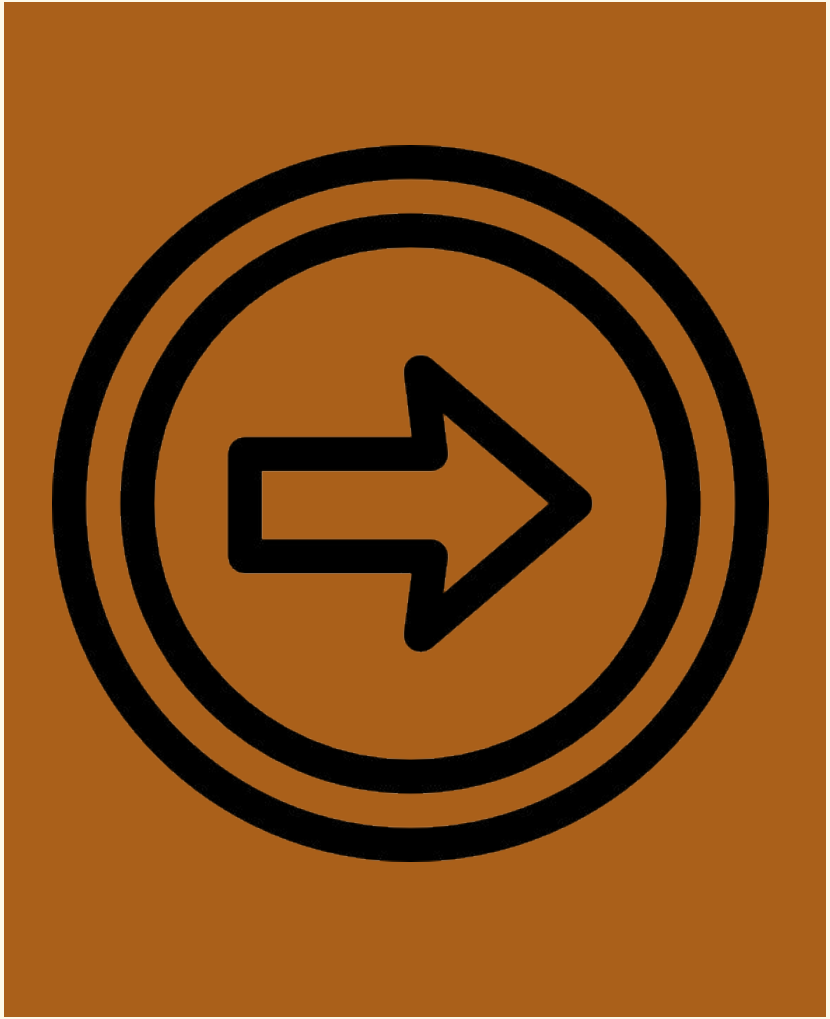
76 %

Testing Accuracy

Recommendations

1. **Encourage students to take morning classes.**
2. **Research effects of displacement.**
3. **Build awareness of scholarship opportunities.**
4. **Monitor student performance across semesters.**
5. **Use XGBoost to flag at risk students for intervention.**



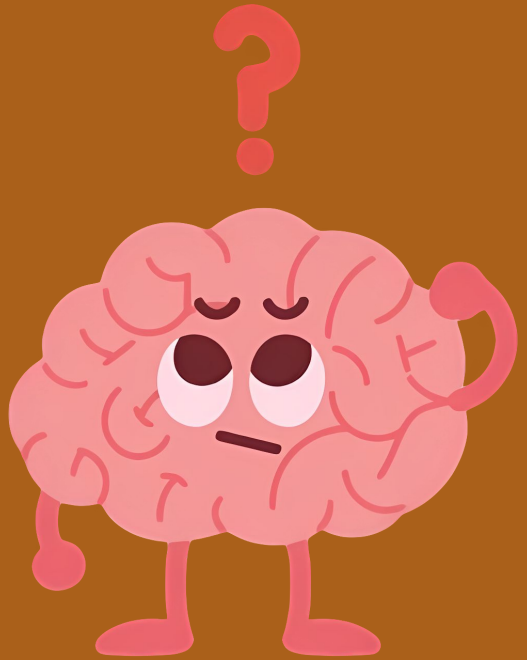


Next Steps

1. Use predictive modeling to identify at-risk students.
2. Meet with students. Offer support.
3. Gather more data to solve class imbalance and improve model.
4. Automate model for ease of use and deploy.

Image Source: CleanPNG

Thank you!



Elijah Lopez | elijahlopez94@gmail.com

Image Source: CleanPNG