Student Performance Factors

Study Habits

Consistent study habits correlate with better performance.

Collaboration

Teamwork and peer support enhance understanding.

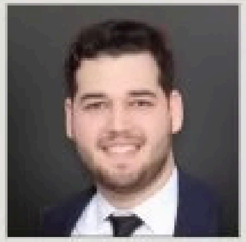
Teacher Support

Guidance and encouragement from educators matter.

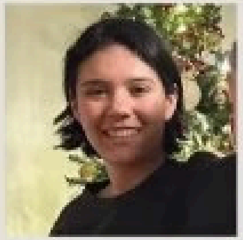
INTRODUCTION OF
OUR TEAM

https://linktr.ee/byte_by_byte_project2

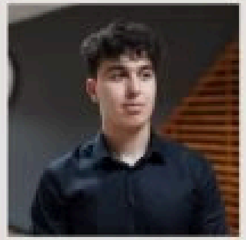

Shawn Gibson


Andres Cuervo


Zoey Huang


Juliana Uribe


Jason Nwaneti


Ahmet Sari

WE'RE STRIVING TO ANSWER:

1 What factors can be used to predict a student's academic success or dropout?

2 Answering this Question is beneficial for:

Students

Educators

Institutions

Policymakers

Society

WHAT WE ARE USING TO PREDICT STUDENT PERFORMANCE

Marital status	4424 non-null	int64
Application mode	4424 non-null	int64
Application order	4424 non-null	int64
Course	4424 non-null	int64
Daytime/evening attendance	4424 non-null	int64
Previous qualification	4424 non-null	int64
Nacionality	4424 non-null	int64
Mother's qualification	4424 non-null	int64
Father's qualification	4424 non-null	int64
Mother's occupation	4424 non-null	int64

Implementing Factors such as:

GDP

Scholarship Holder

Unemployment Rate

Mother / Father Occupation

Inflation Rate

Debt

Why Student Prediction Factors

1

Relevance

As students, we're interested in exploring factors that impact grades, making this topic directly relatable and applicable to our own academic experiences.

2

Data Balance

It has a good mix of categorical and numerical data, allowing for diverse analytical techniques.

3

Insight Potential

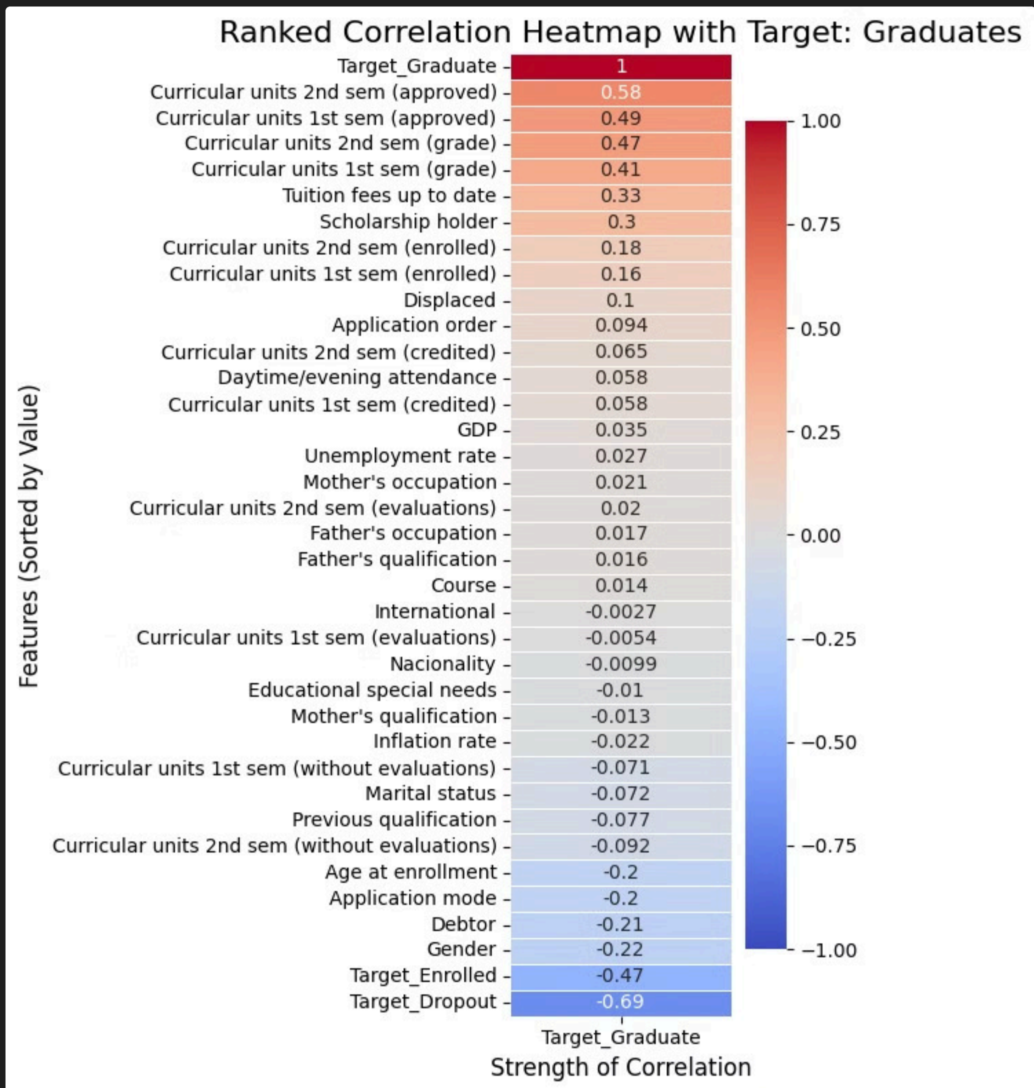
With its extensive variables, we can draw significant conclusions that may influence educational strategies.

4

Variable Richness

The dataset contains a wide range of both categorical and numerical variables, enabling compressive analyses.

CORRELATION BETWEEN FACTORS AND GRADUATING



KEY CORRELATIONS:

Financial Stability is crucial:

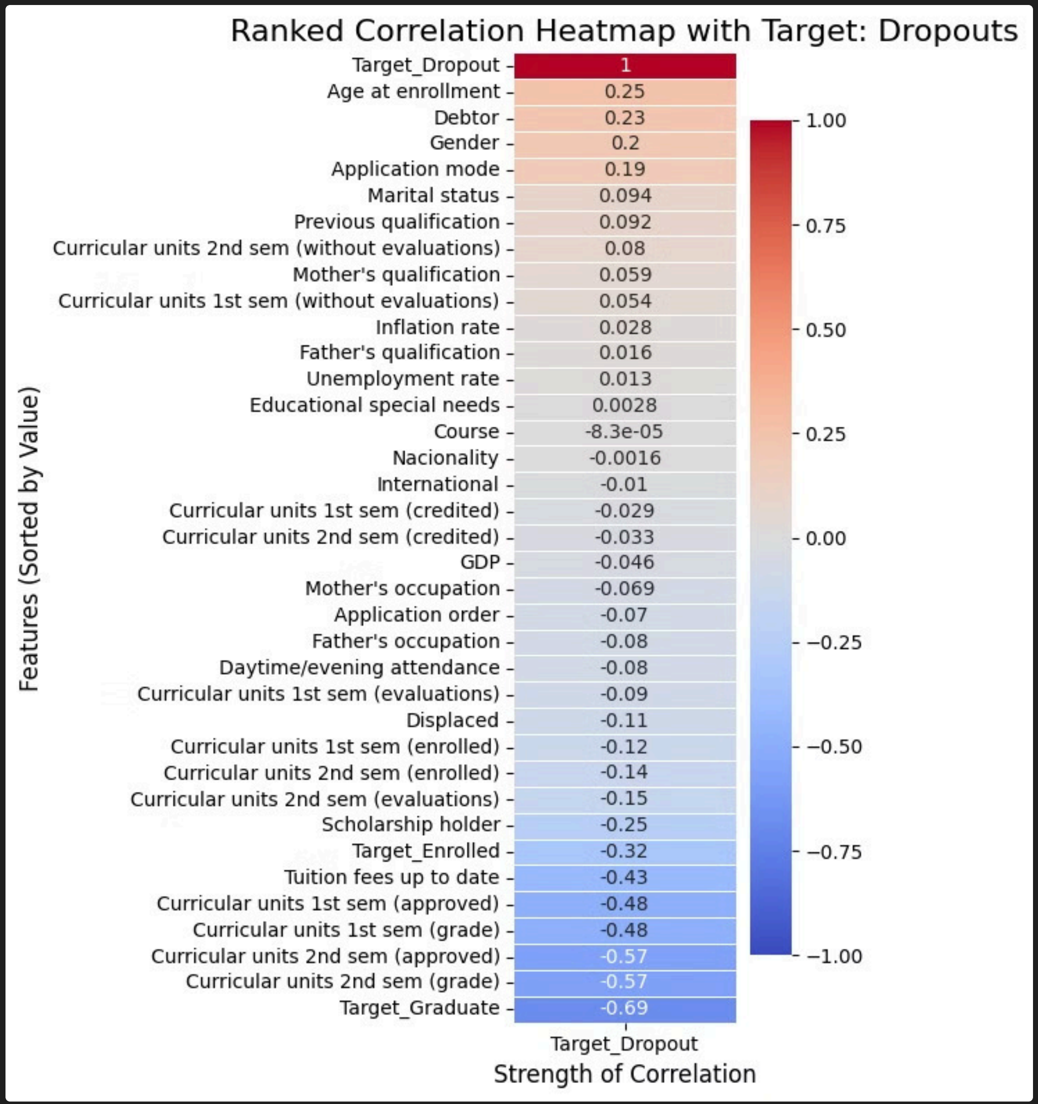
Up-to-date tuition fees (0.33) indicate a strong link between financial stability and graduation likelihood.

Scholarship holders (0.30) show a positive correlation, highlighting the impact of financial aid.

Enrollment and Persistence:

Engagement in curricular units (0.16 - 0.18) demonstrates a moderate correlation, emphasizing the importance of continuous coursework.

CORRELATION BETWEEN FACTORS AND DROPPING OUT



KEY CORRELATIONS:

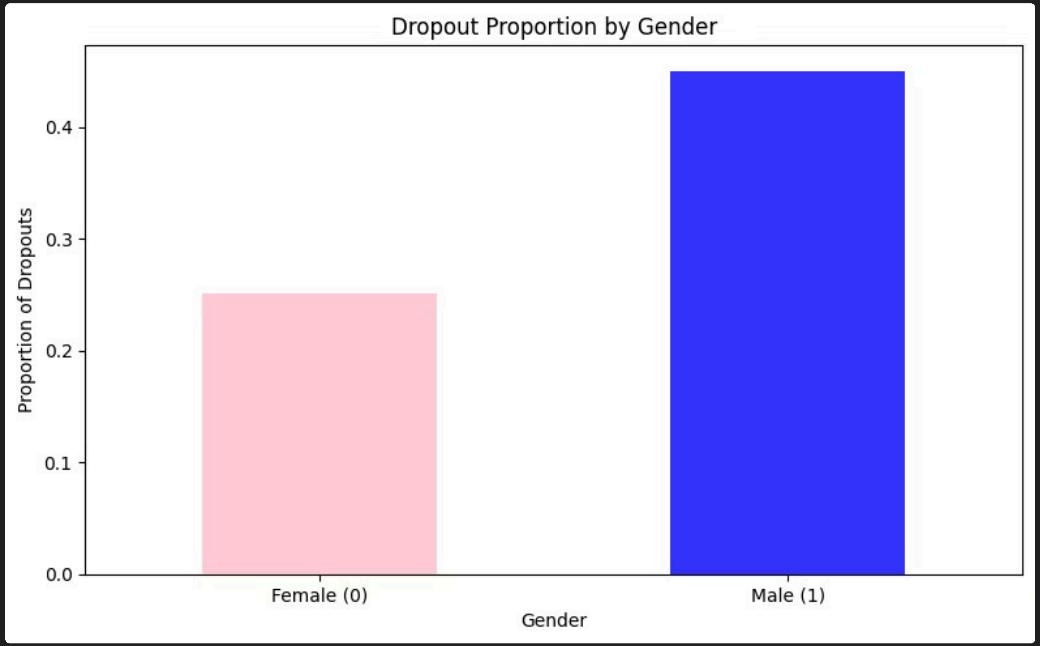
Age of Enrollment (0.25): Older students may face challenges increasing dropout risk.

Debtor Status (0.23): Financial difficulties are a major dropout risk factor.

Gender (0.20): Gender-related issues may influence dropout rates; further investigation needed.

Curricular Performance (-0.48 to -0.57): Academic success is a significant protective factor against dropping out.

DROPOUT PROBABILITY BASED ON GENDER

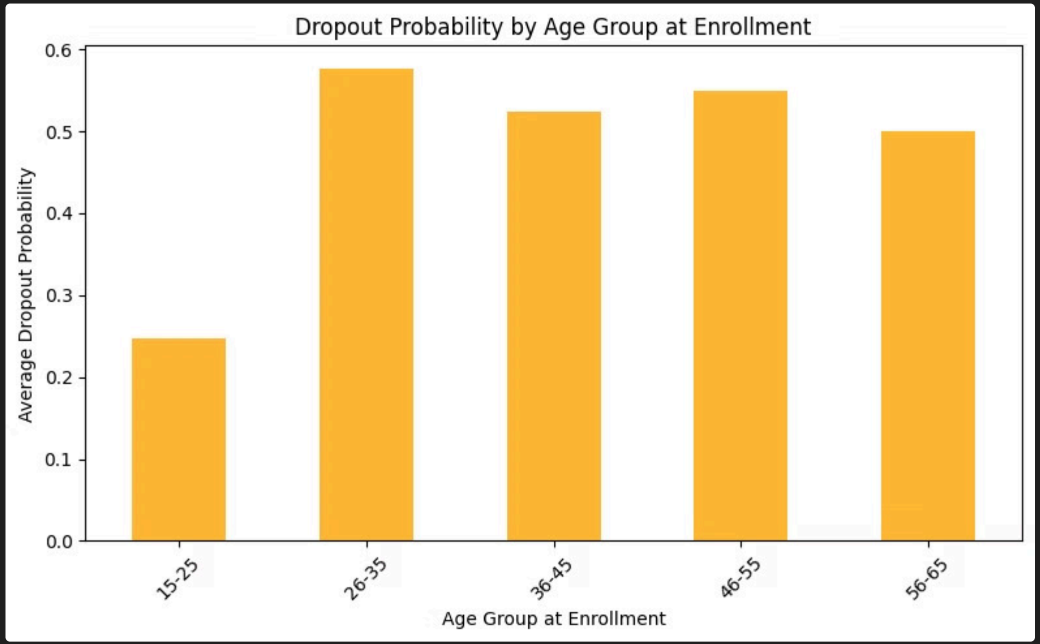


MALES EXHIBIT A HIGHER PROBABILITY OF DROPPING OUT COMPARED TO FEMALES.

THE DROPOUT RATE FOR MALES EXCEEDS 40%, WHEREAS IT IS ONLY 25% FOR FEMALES.

THESE GENDER DIFFERENCES MAY STEM FROM ACADEMIC, FINANCIAL, OR PERSONAL FACTORS.

DROPOUT PROBABILITY BASED ON AGE REGISTRATION



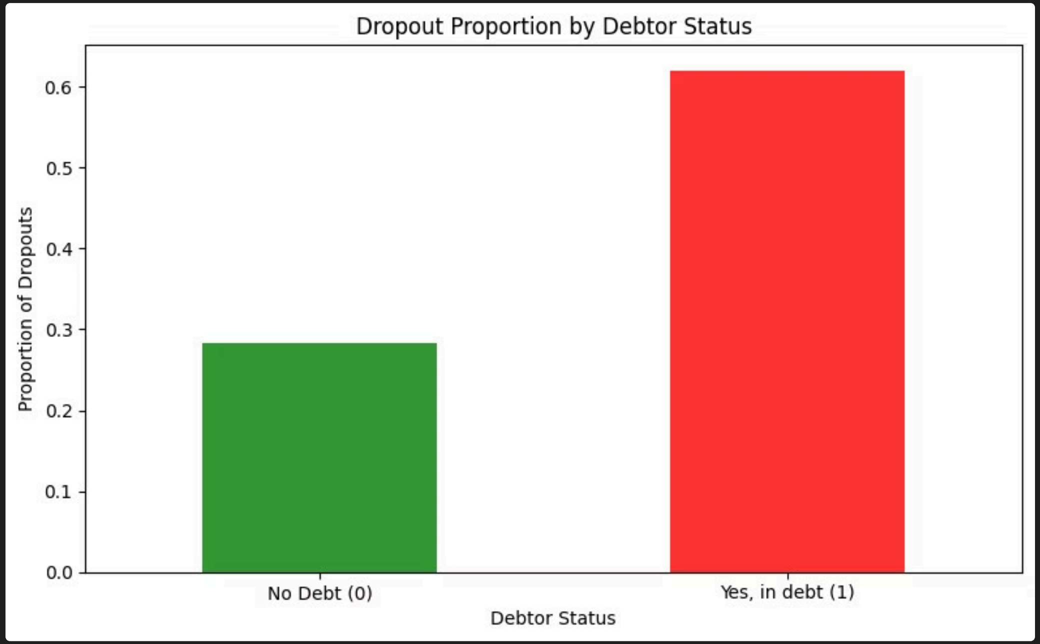
Method:

- Visualization of Dropout Proportion by Age Group

Key Insights:

- Dropout probability rises with age, peaking in the 26-35 and 46-55 age brackets.
- Younger students have the lowest dropout rates, indicating greater academic stability.
- Older students show consistently high dropout probability, potentially due to external pressures like work or family.

DROPOUT PROBABILITY BASED ON DEBT



Method:

- Constructed a visualization to observe Dropout Proportion by Debt status using Matplotlib library.

Key Insights:

- Students with debt have a significantly higher dropout rate (~0.62) compared to those without debt (~0.30).
- Financial instability is a strong predictor of dropout risk, highlighting the need for better financial aid or support systems.
- Reducing student debt through scholarships, grants, or flexible payment options could improve retention rates.

PREDICTIVE MODELING- Graduate

GOAL: EXPERIMENT WITH PREDICTION MODELS AND DECIDE WHICH IS BEST TO USE FOR PREDICTING GRADUATION/DROPOUT.

1. XGBOOST < -- WINNER!
2. RANDOM FOREST CLASSIFIER
3. SUPPORT VECTOR MACHINE
4. ANN (ARTIFICIAL NEURAL NETWORK)

Random Forest Classifier:

1. **Balanced Performance**: Delivered an optimal trade-off between precision and recall, achieving **high F1 scores** (87.1%), showcasing strong predictive power for both classes.

2. **Effective on Imbalanced Data**: Handled class imbalance effectively after resampling, maintaining high sensitivity (90.3% recall for graduates).

3. **Scalable and Efficient**: Performed well with minimal hyperparameter tuning, demonstrating its adaptability across datasets.

PREDICTIVE MODELING- Dropout

GOAL: EXPERIMENT WITH PREDICTION MODELS AND DECIDE WHICH IS BEST TO USE FOR PREDICTING GRADUATION/DROPOUT.

1. RANDOM FOREST CLASSIFIER < -- WINNER!
2. XGBOOST
3. SUPPORT VECTOR MACHINE
4. ANN (ARTIFICIAL NEURAL NETWORK)

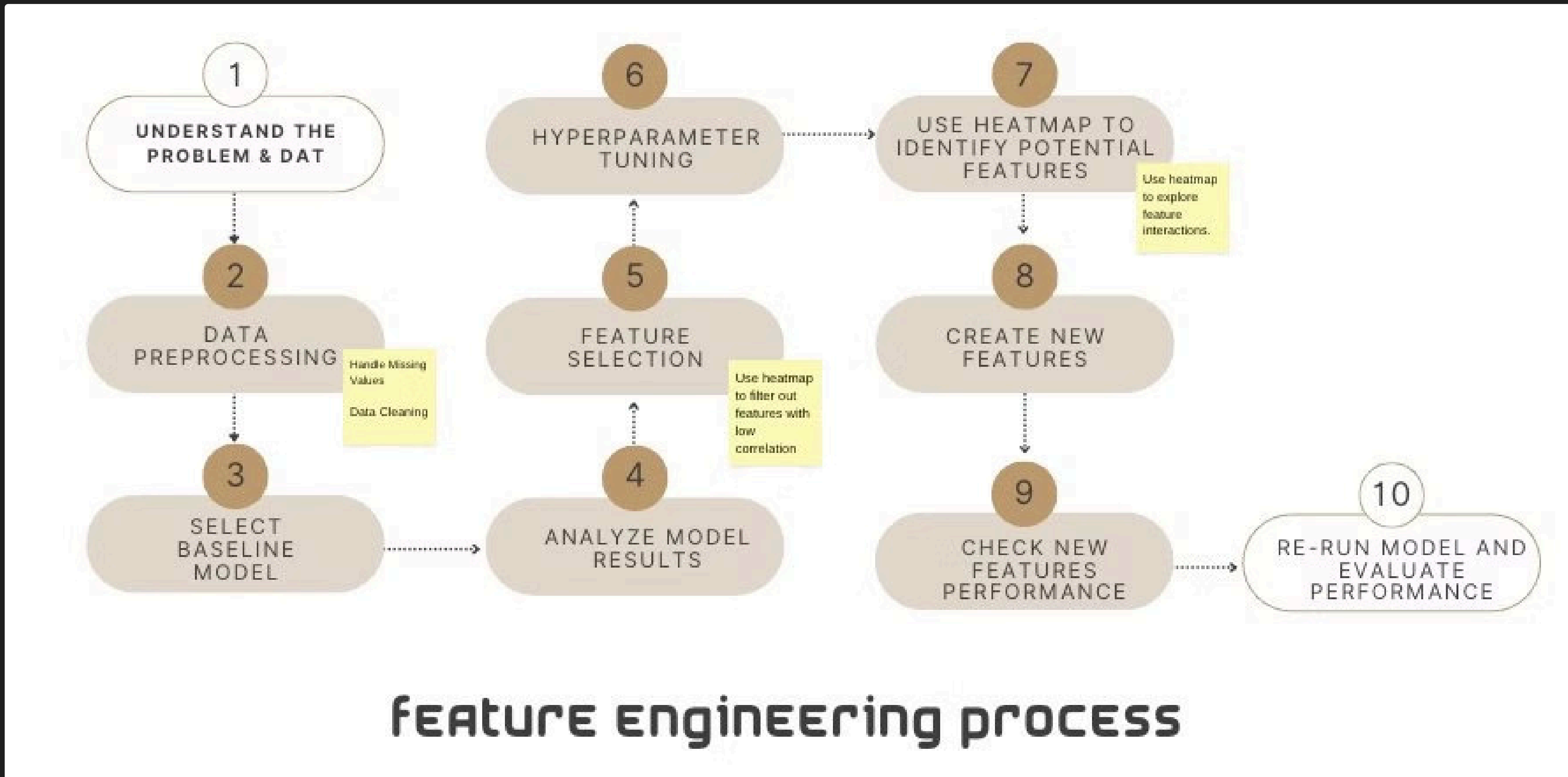
Random Forest Classifier:

1 **Balanced Performance**: Delivered an optimal trade-off between precision and recall, achieving **high F1 scores** (87.1%), showcasing strong predictive power for both classes.

2 **Effective on Imbalanced Data**: Handled class imbalance effectively after resampling, maintaining high sensitivity (90.3% recall for graduates).

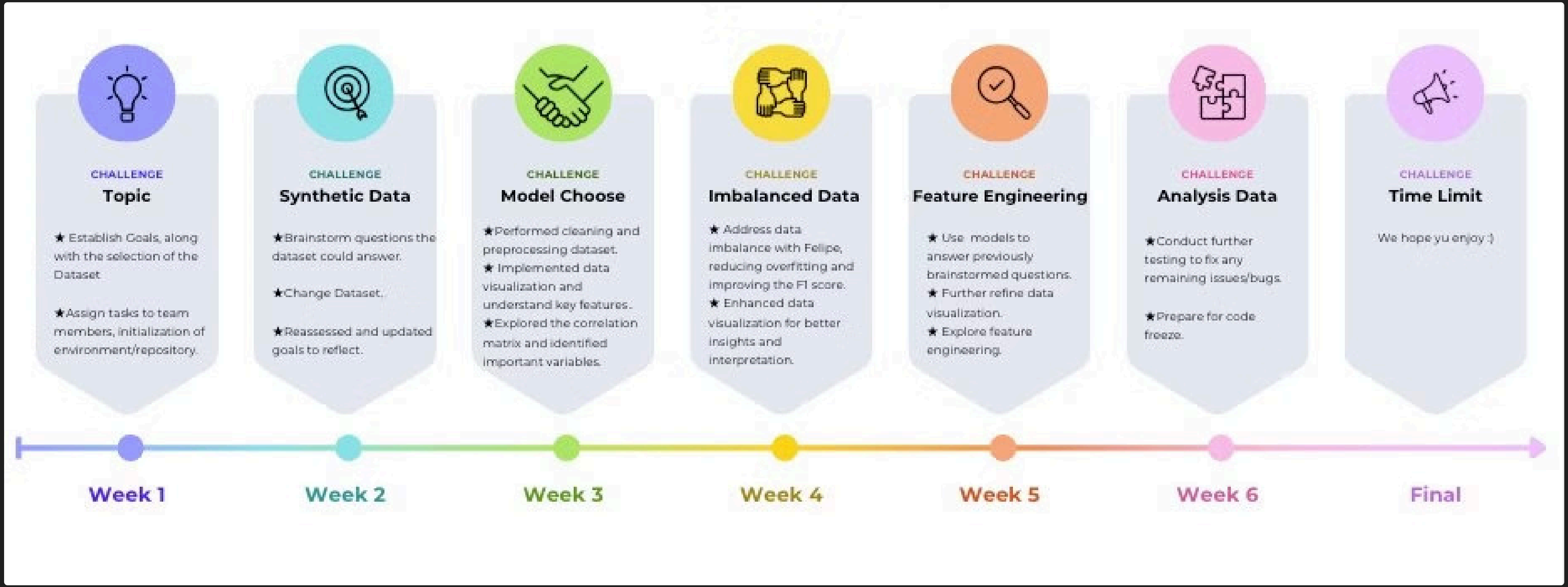
3 **Scalable and Efficient**: Performed well with minimal hyperparameter tuning, demonstrating its adaptability across datasets.

Feature Engineering



Made with Gamma

Timeline



CHALLENGE
Topic
★ Establish Goals, along with the selection of the Dataset
★Assign tasks to team members, initialization of environment/repository.
CHALLENGE
Synthetic Data
★Brainstorm questions the dataset could answer.
★Change Dataset.
★Reassessed and updated goals to reflect.
CHALLENGE
Model Choose
★Performed cleaning and preprocessing dataset.
★ Implemented data visualization and understand key features.
★Explored the correlation matrix and identified important variables.
CHALLENGE
Imbalanced Data
★ Address data imbalance with Felipe, reducing overfitting and improving the F1 score.
★ Enhanced data visualization for better insights and interpretation.
CHALLENGE
Feature Engineering
★ Use models to answer previously brainstormed questions.
★ Further refine data visualization.
★ Explore feature engineering.
CHALLENGE
Analysis Data
★Conduct further testing to fix any remaining issues/bugs.
★Prepare for code freeze.
CHALLENGE
Time Limit
We hope yu enjoy :)
Week 1
Week 2
Week 3
Week 4
Week 5
Week 6
Final

Comparison

Stage-Drop Out	Model	Accuracy	+/-	F1 Score	+/-
Original	ANN	0.86		0.77	
Drop low important	ANN	0.87	↑	0.78	↑
Fine-Tuned	ANN	0.86	↓	0.85	↑
Feature Engineering	ANN				
Original	Random Forest	0.86		0.77	
Drop low important	Random Forest	0.87	↑	0.79	↑
Fine-Tuned	Random Forest	0.87	↑	0.79	↑
Feature Engineering	Random Forest	0.87		0.83	↑
Original	SVM	0.85		0.77	
Drop low important	SVM	0.88	↑	0.80	↑
Fine-Tuned	SVM	0.88		0.80	
Feature Engineering	SVM	0.87	↓	0.82	↑
Original	XGBoost	0.85		0.77	
Drop low important 34->30	XGBoost	0.86	↑	0.79	↑
Fine-Tuned	XGBoost	0.87	↑	0.79	
Feature Engineering	XGBoost	0.87		0.82	↑

Conclusion

-
-
-
-

Any Questions?

unknown linkunknown linkunknown link

Thank You!