Deloitte.

eMerge Education

August 2023

Roberto Aguilar, Brenda Alicia Delgado, Jorge Eduardo Cortes, Mercedez Vela, Daniel Jackson

AGENDA & TABLE OF CONTENTS

Overview

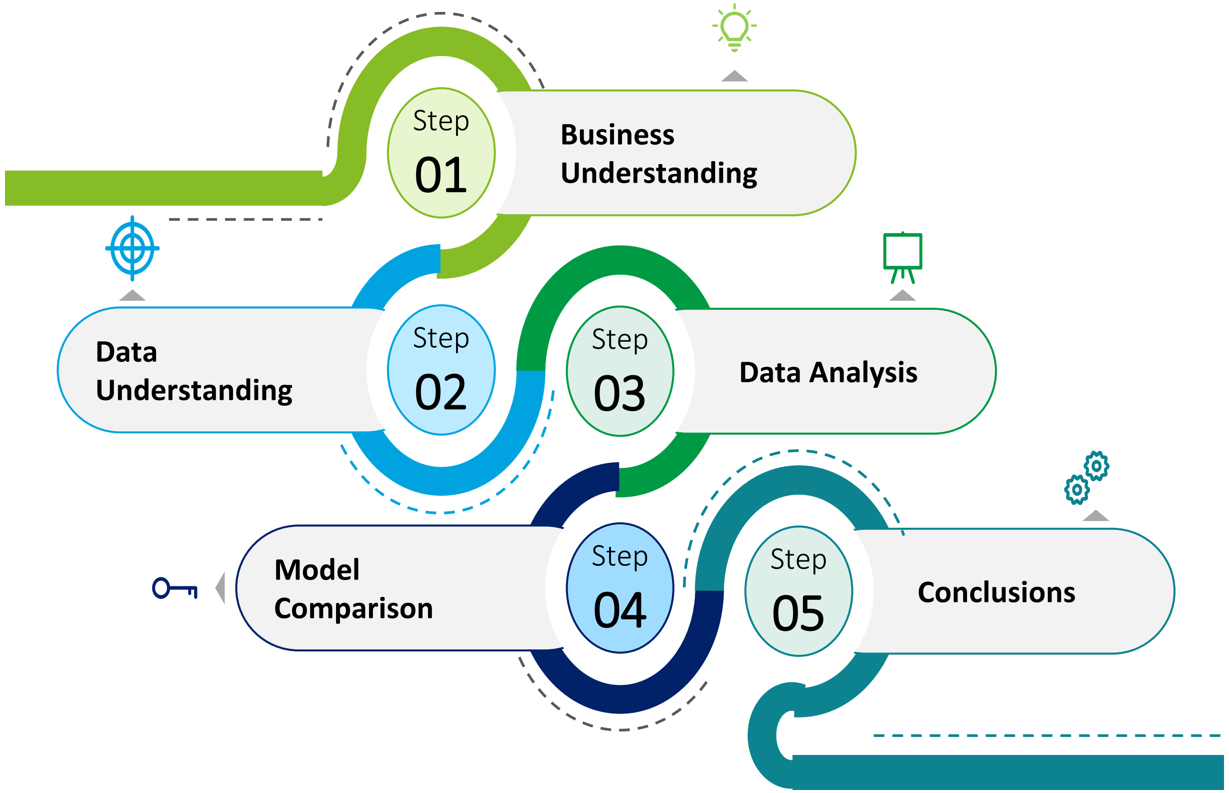
1 Business Understanding

2 Data Understanding

3 Data Analysis

4 Evaluation & Model Comparison

5 Conclusions & Areas for Improvement



Business Understanding

- The persisting dropout rates among higher education students is attributed to a lack of proper identification of root causes within student's lives from the Institution, as well as the institutions limited resources and time

- Build a predictive model to anticipate potential dropouts
- Assess the performance of different supervised machine learning algorithms for this specific task
- Compare the effectiveness of various supervised machine learning algorithms in accurately predicting student dropouts to properly allocate time and resources

- These algorithms can be used for early detection and forecasting of potential student dropouts or students at the risk of quitting schools efficiently so that corrective strategies may be applied

Data Understanding

Data

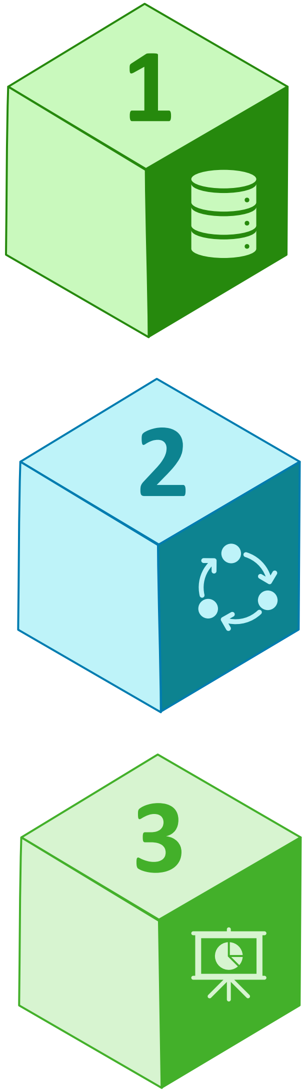
- One dataset that contains information gathered from multiple disconnected databases within a higher education institution, focusing on students enrolled in various undergraduate degree programs.
- Includes information known at the time of student enrollment – academic path, demographics, and social-economic factors.
- Three category classification task (dropout, enrolled, and graduate)
- 36 Features, 19 were Categorical Data

Method

- Handling of outliers and anomalies
- Scaling for features that had a non-linear scale

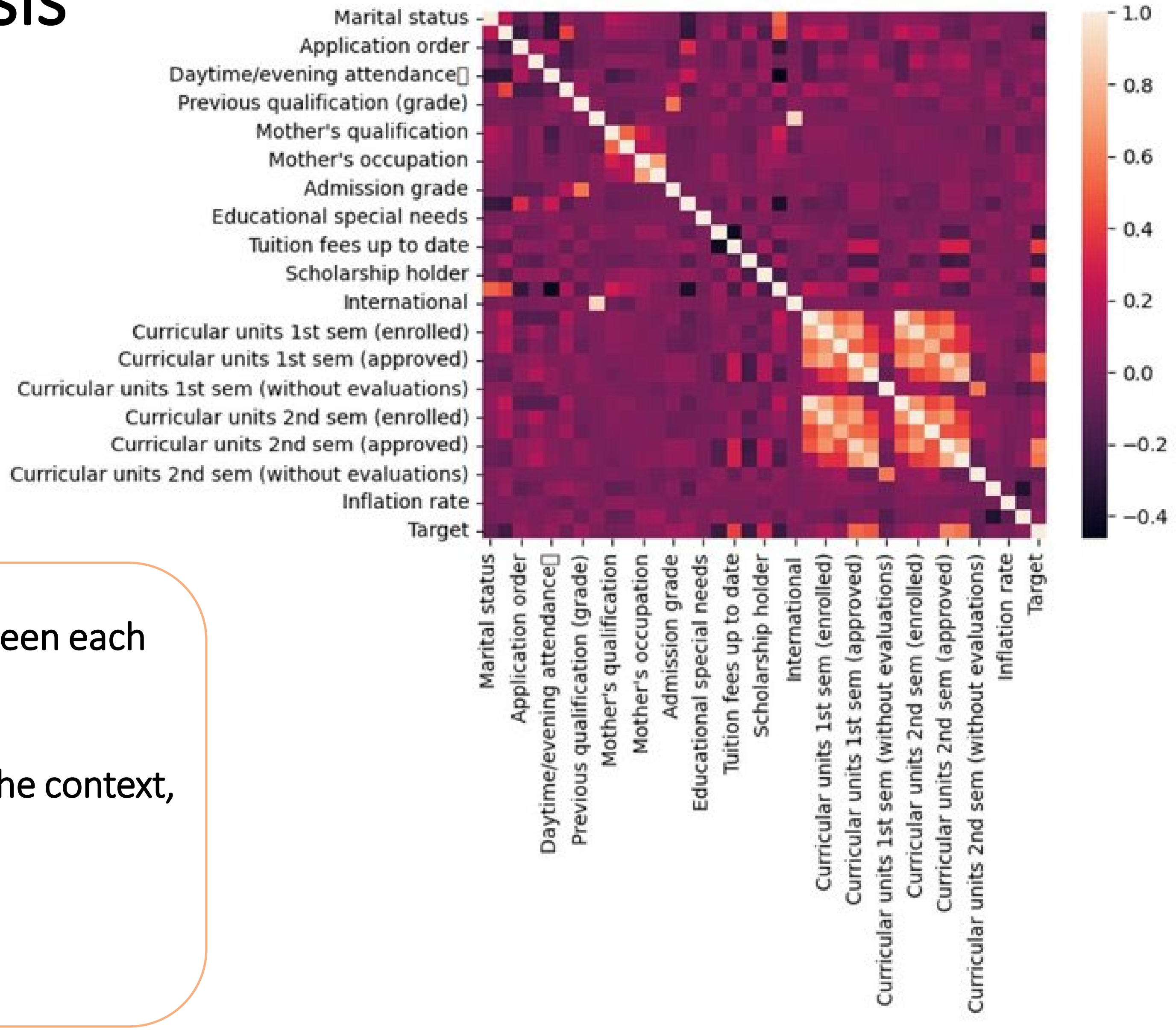
Results

- Kept the 36 Features
- Two category classification task (Dropout, and graduate)



Dataset: Predict students' dropout and academic success - UCI Machine Learning Repository

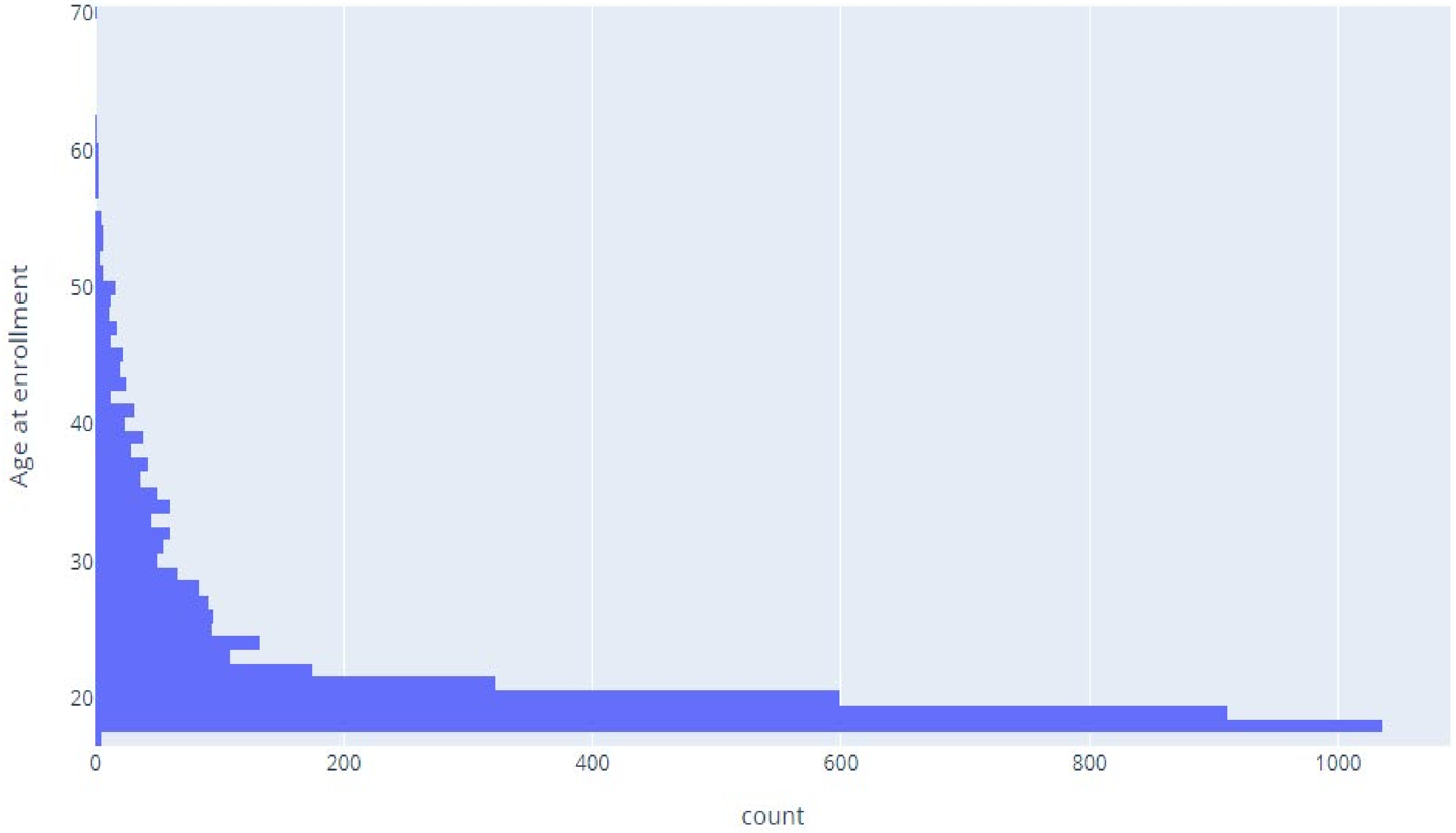
Exploratory Data Analysis



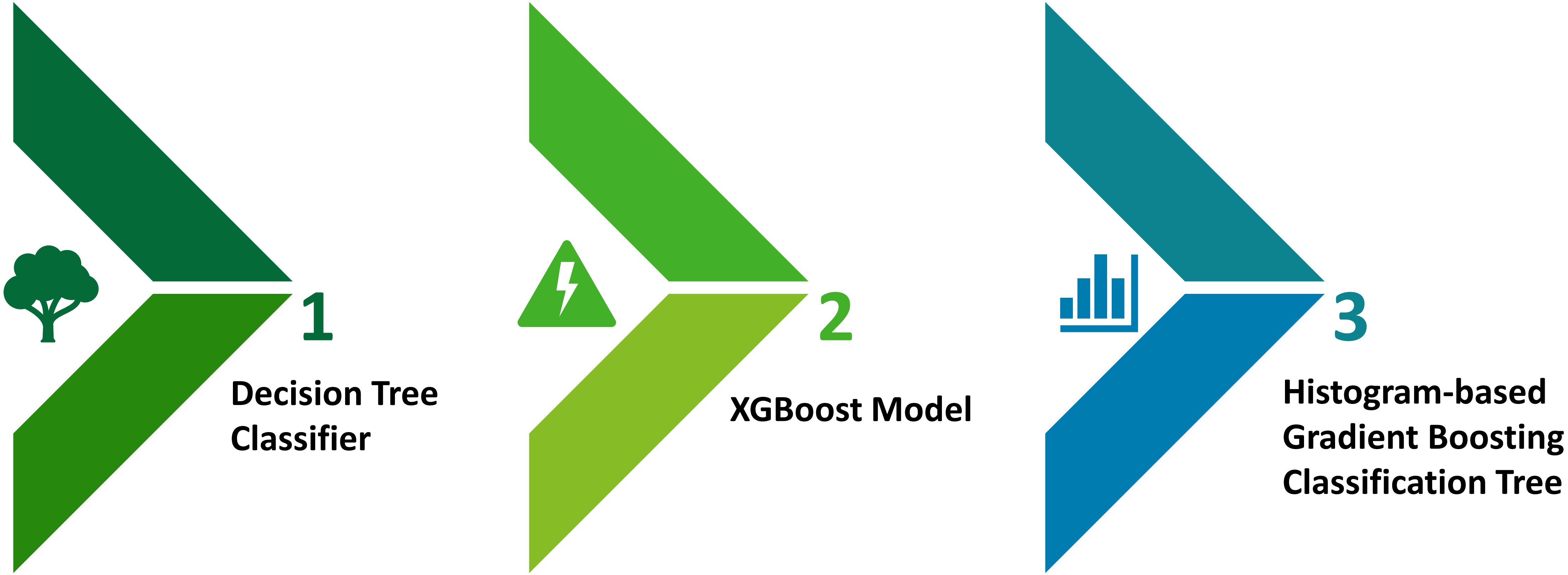
- All features do not have high correlation between each other.
- High correlated features are explained given the context, and business understanding.

Exploratory Data Analysis (Continued)

This result indicates that the majority of enrollments occur at ages 19 and 20, while enrollments at age 30 and upwards are relatively infrequent. The reasons for this distribution could be influenced by various factors, such as educational requirements, age at enrollment, special needs, etc.



Model Development



1

Decision Tree Classifier

2

XGBoost Model

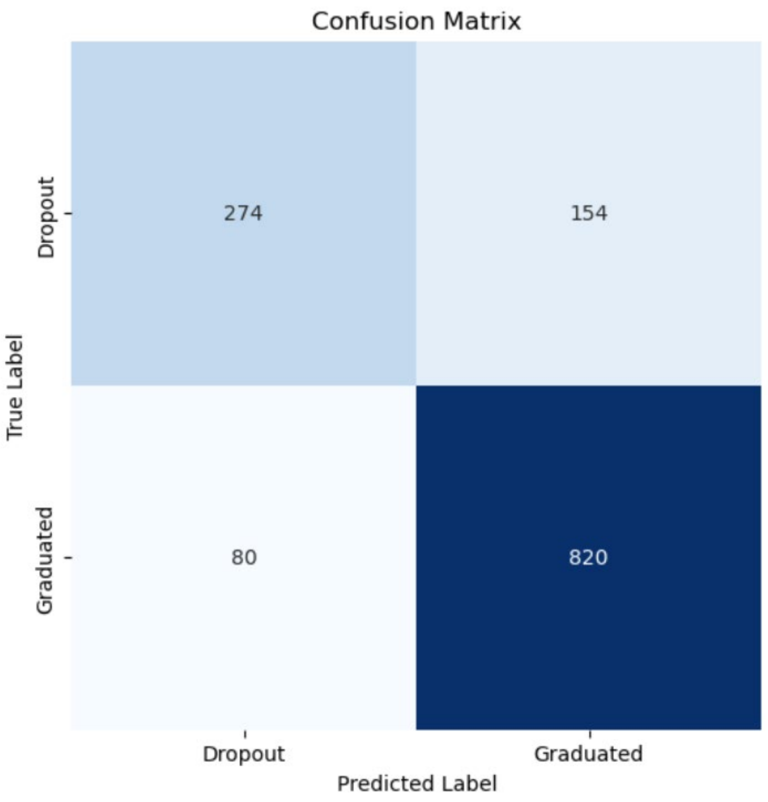
3

Histogram-based Gradient Boosting Classification Tree

Model Evaluation (Accuracy Score)

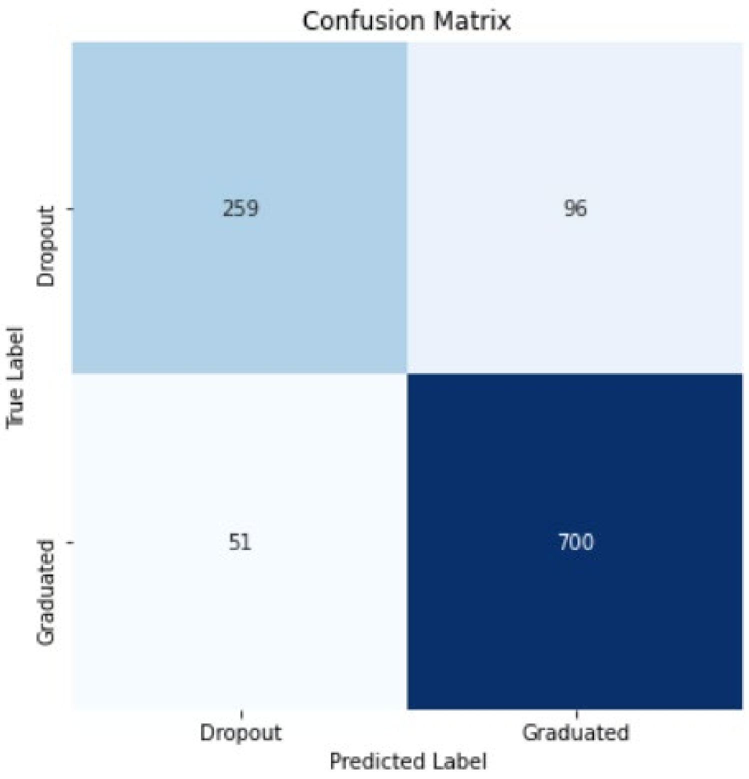
Decision Tree Classifier

82.3795%



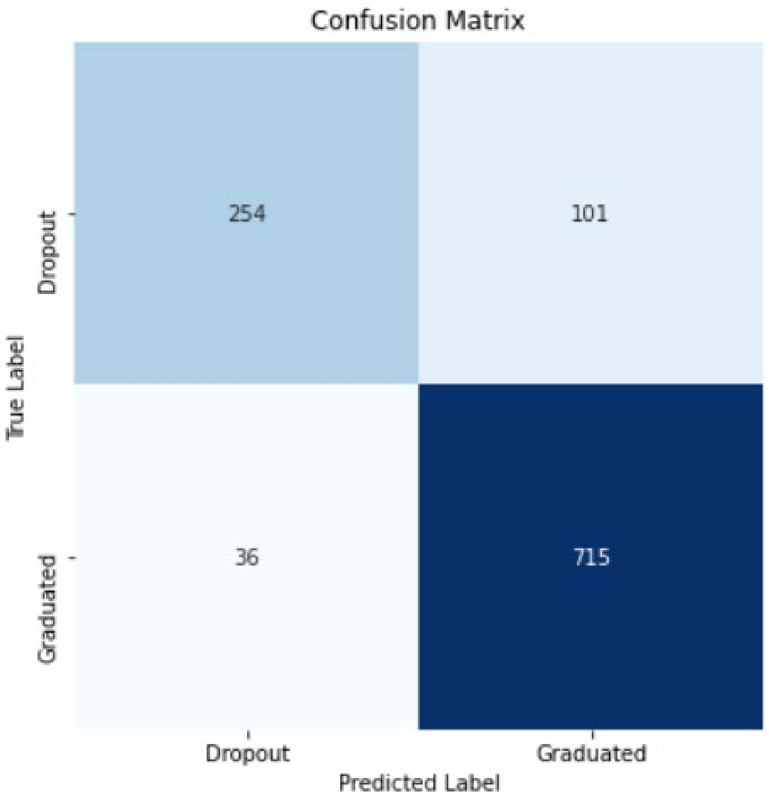
XG-Boost Model

86.7%



Histogram-based Gradient Boosting Classification Tree

87%

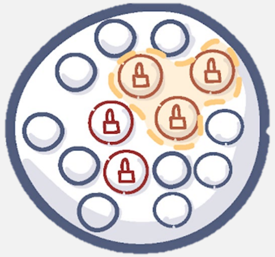


Precision vs Recall Trade-off

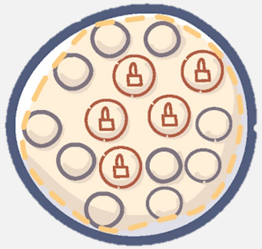
What **threshold** to use to determine if an instance is positive or negative?

Threshold

High threshold; High Precision, Low Recall



Low threshold; Low Precision, High Recall



Our Goal

High Threshold

To:

- **Ensure efficient resource allocation**
- **Minimize false positives**

Conclusions & Areas of Improvement

WHO

- Which students that were predicted to dropout, should be given resources
- This will include human intervention among the predicted top 10%

BIAS

- Variables the model learns from such as socioeconomic status, often have bias associated to them

TIMING

- Model will not be used as an acceptance metric, only as an intervention method to minimize ethical issues surrounding bias
- Human intervention will take time, thus seen as a limitation

LEGAL

- Potential legal repercussions could be students who are actually struggling, that the model did not predict, do not receive resources
- Human intervention will help mitigate this risk
- Clearly outlining the institutes limited resources helps mitigate risk

HUMAN INTERVENTION

- Human intervention is key to mitigating any bias the model has developed
- There will be a case-by-case evaluation for each student

Thank you! Any Questions?

- Roberto Aguilar
- Brenda Alicia Delgado
- Jorge Eduardo Cortes
- Mercedez Vela
- Daniel Jackson