
DATA DRIVEN ANALYSIS AND STUDENT PERFORMANCE PREDICTION

Open University learning Analytics Dataset



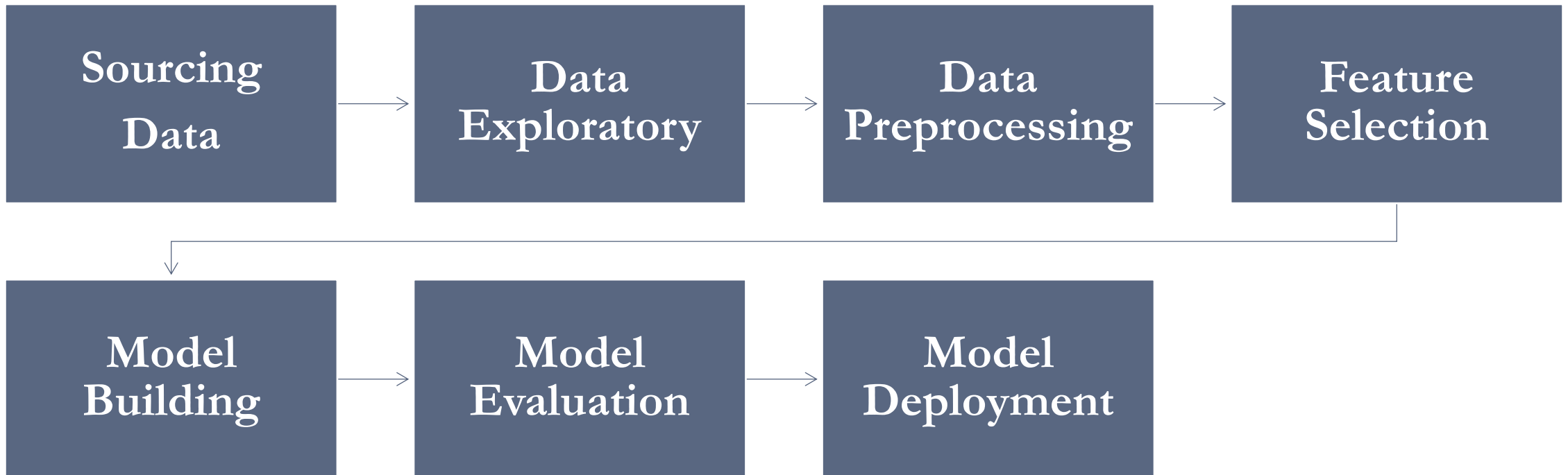
By: Lara Onipede

ANALYTICAL OBJECTIVES



- Analyze Student Information and performance factors:
 - Data Exploration
 - Data Integration
 - Visual EDA
 - Data Preprocessing and Feature Engineering
 - Build Predictive Models to predict student performance in the final exams
 - Build Predictive Models: Random Forest, Gradient boosting, Decision Tree and SVM
 - Evaluate Models
 - Create a Flask API to deploy Model
 - Test Model for prediction
-

PROJECT FLOW



DATASET

Open University Learning Analytics Dataset

Courses

Assessments

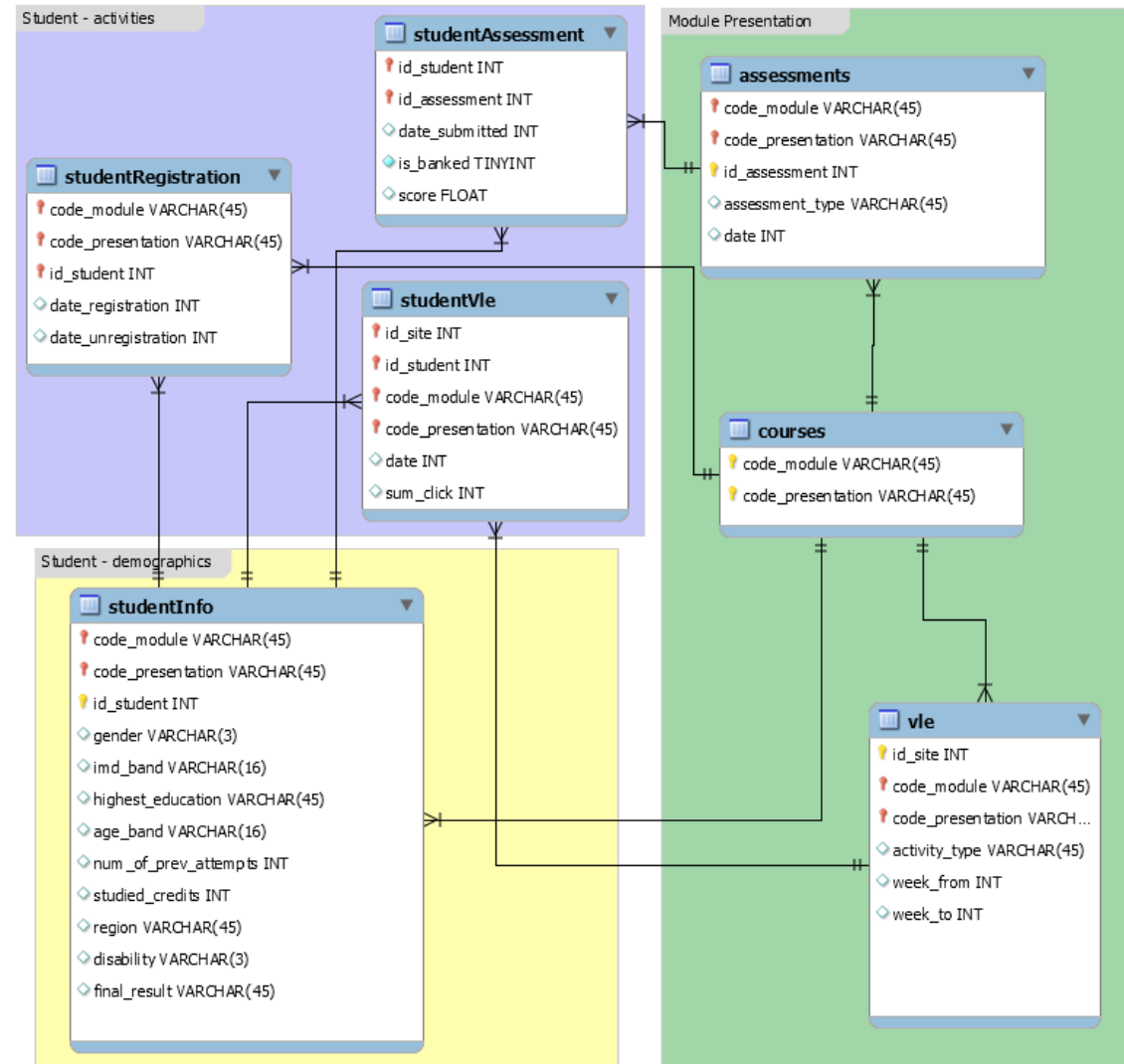
Student_Infor
mation

Student_Regis
tration

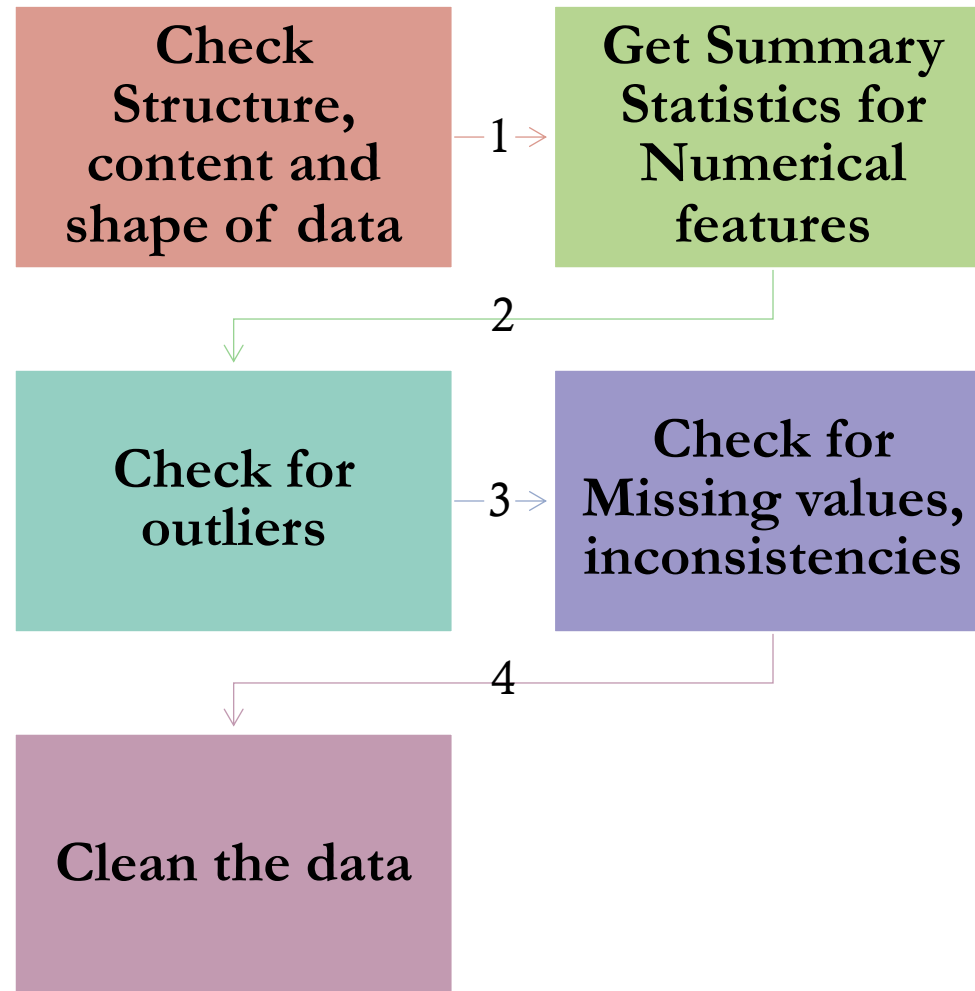
Student_Asses
sments

Vle

Student_Vle

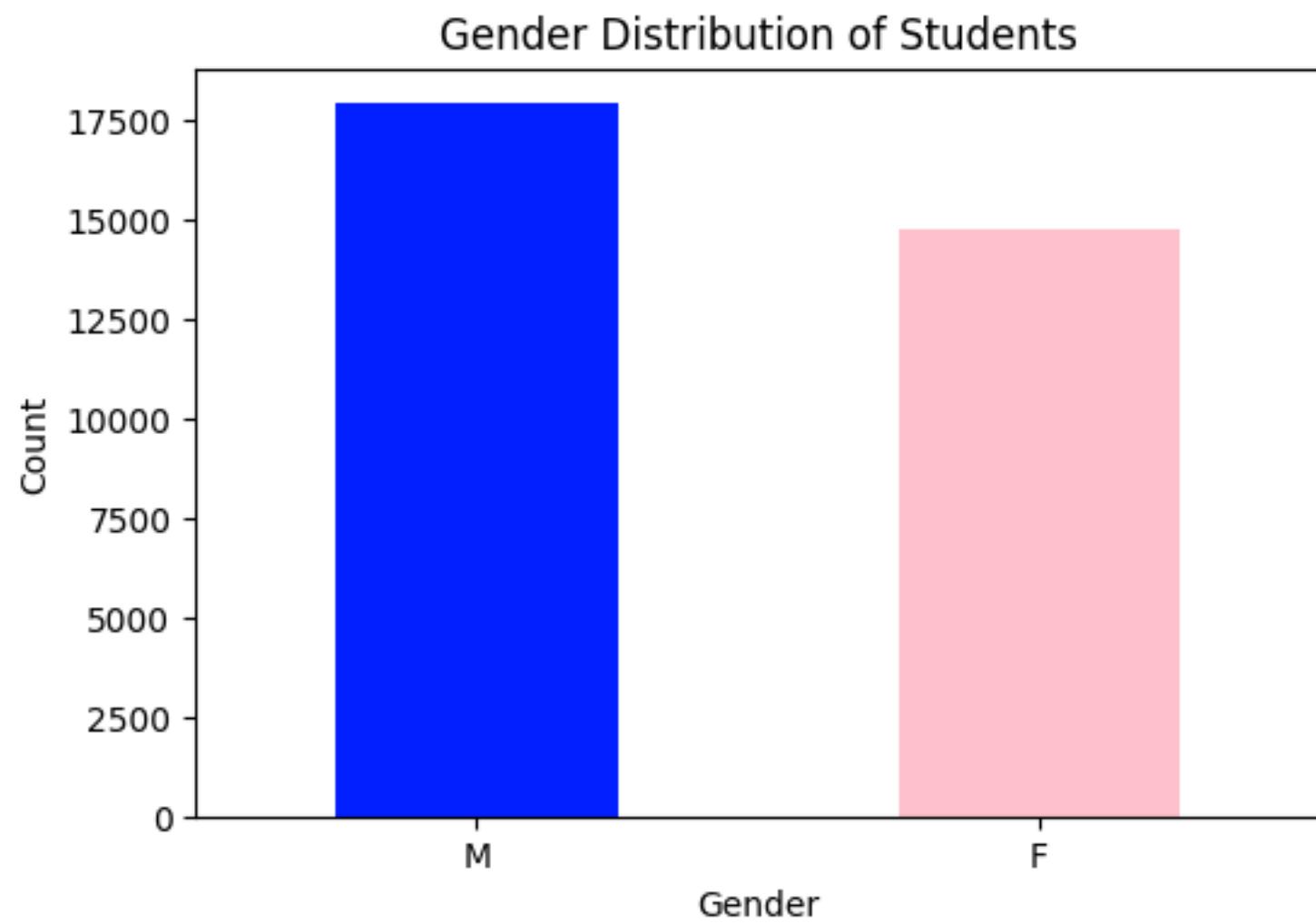


DATA EXPLORATORY AND CLEANING

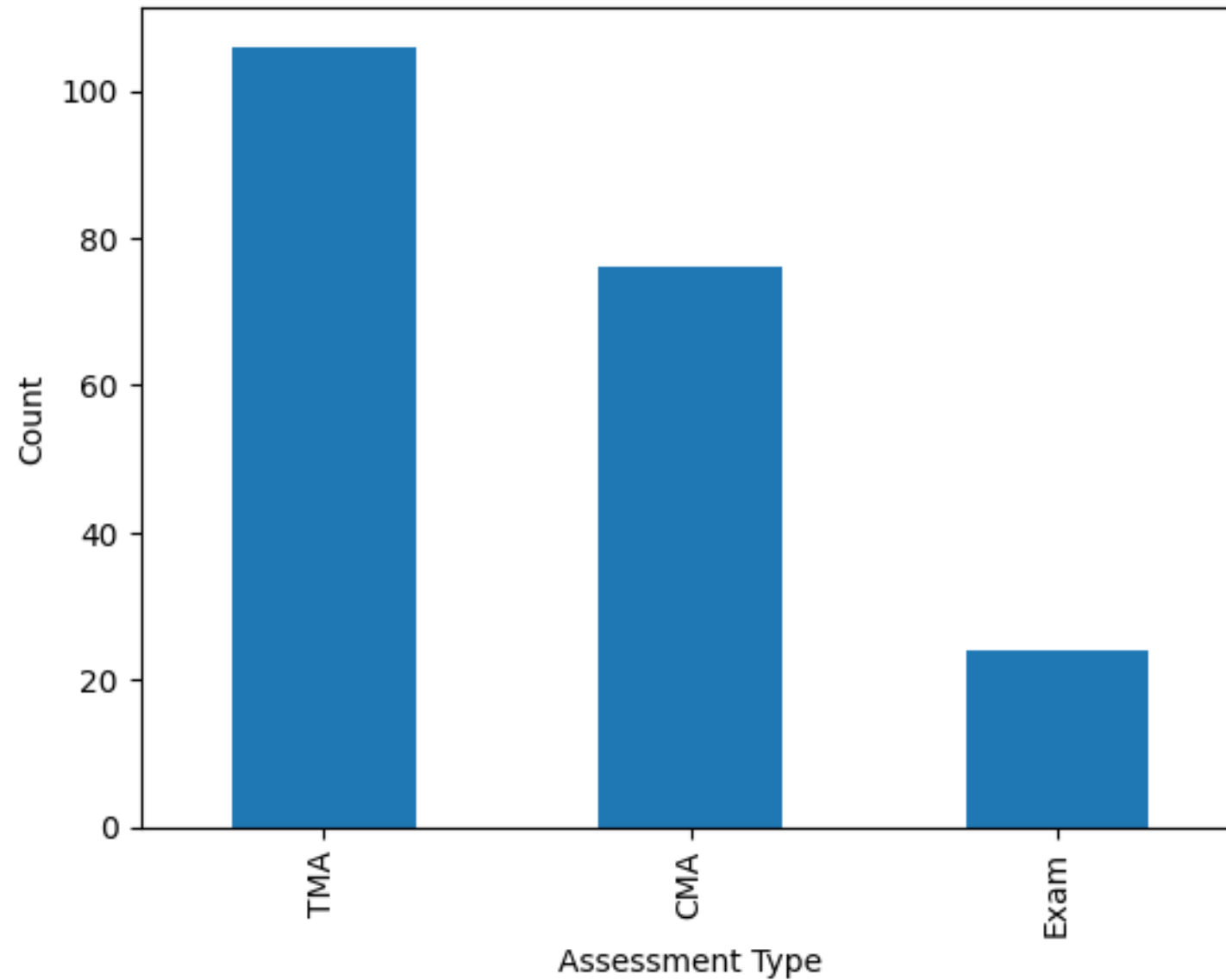


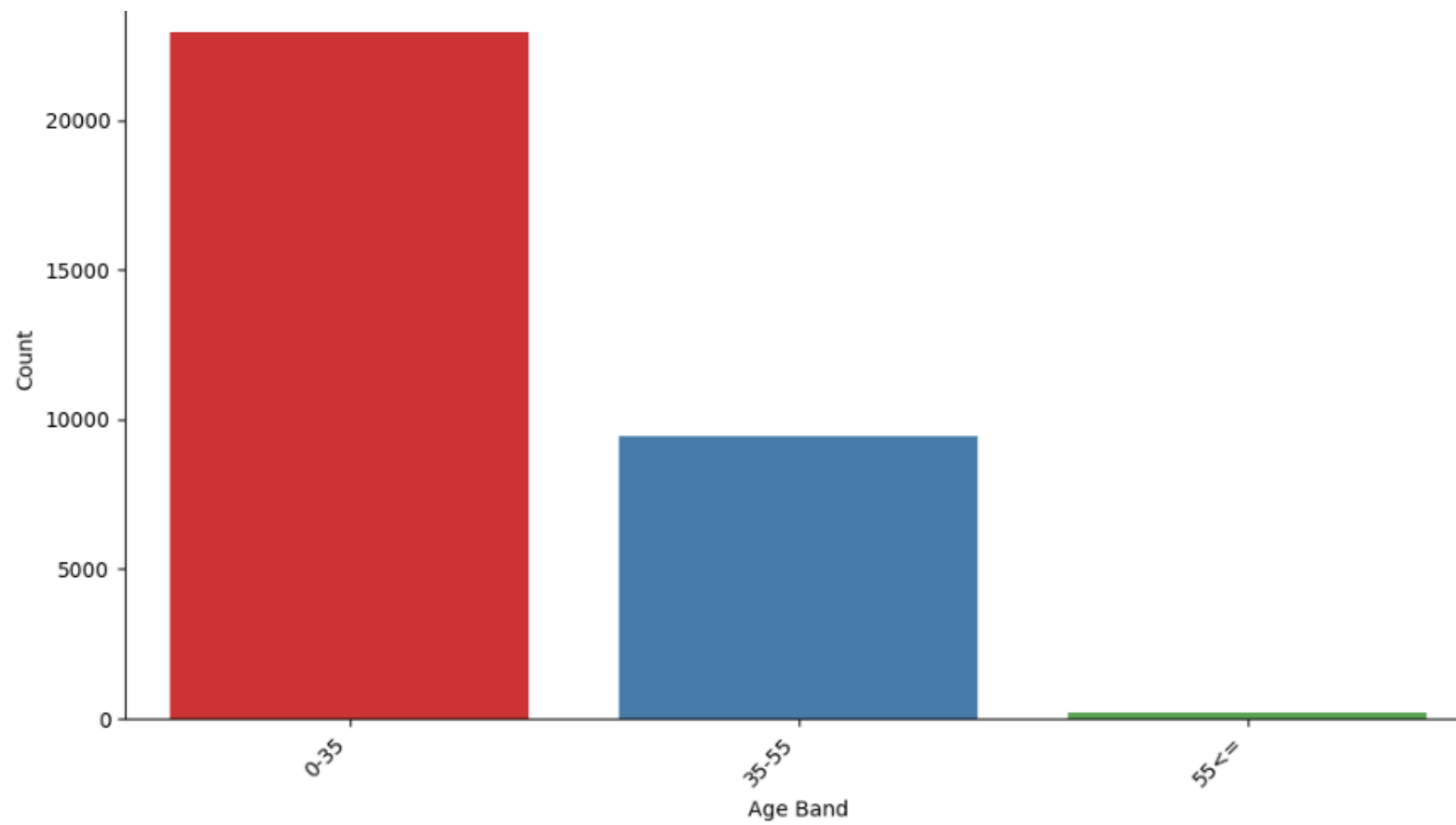


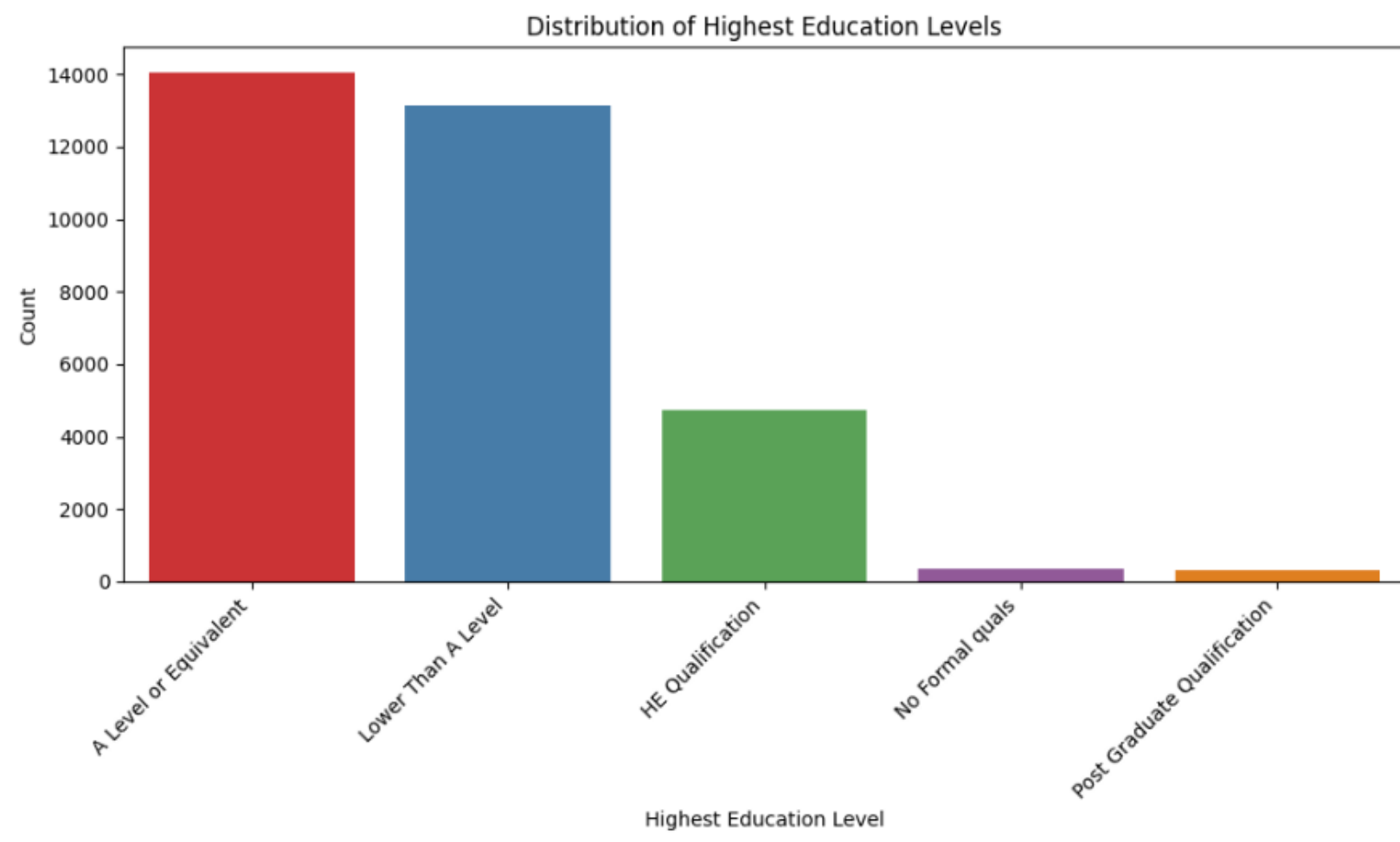
VISUAL EXPLORATORY ANALYSIS

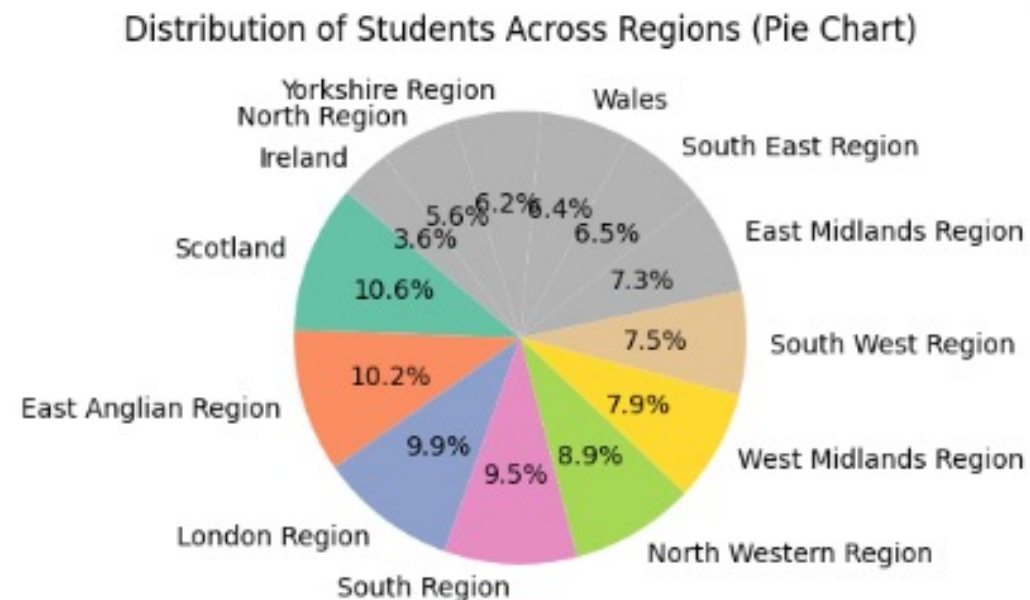
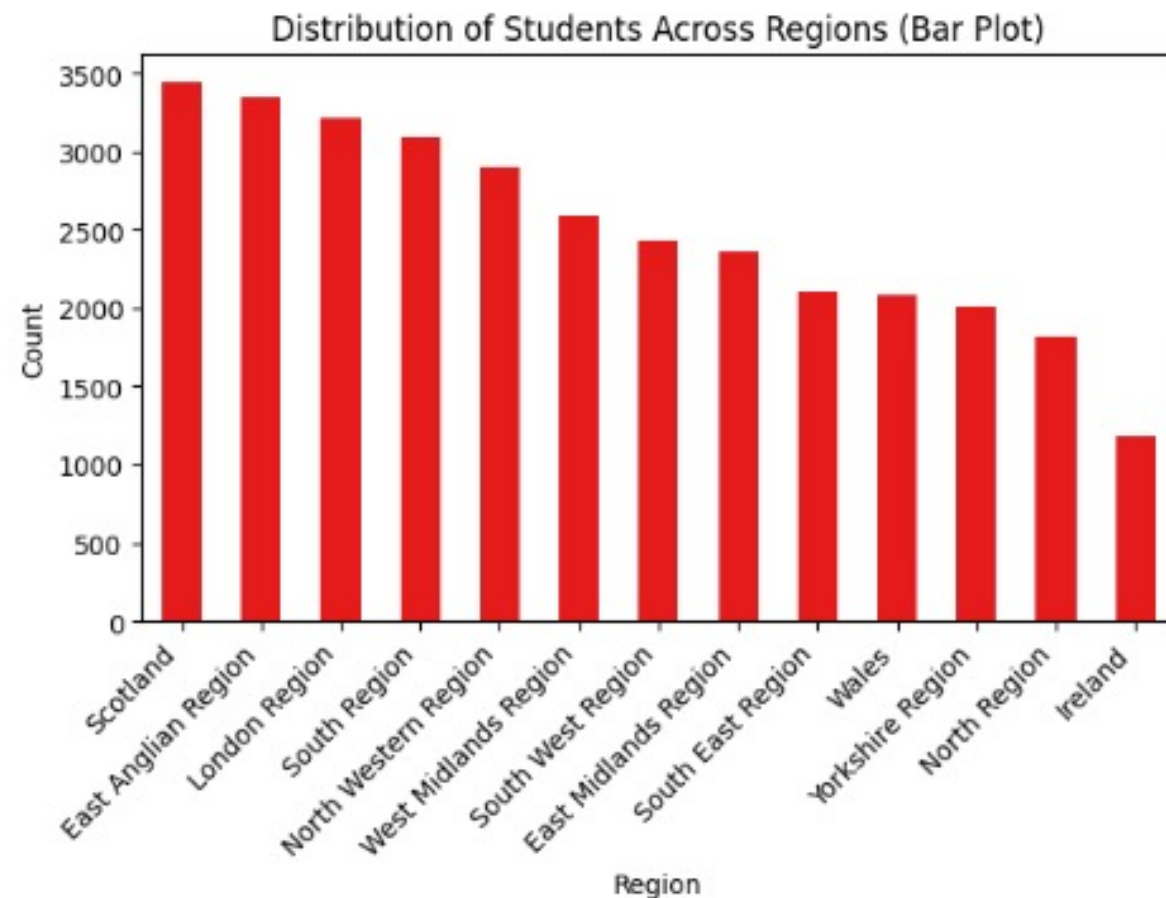


Distribution of Assessment Types









DATA PREPROCESSING

Merged Dataframes

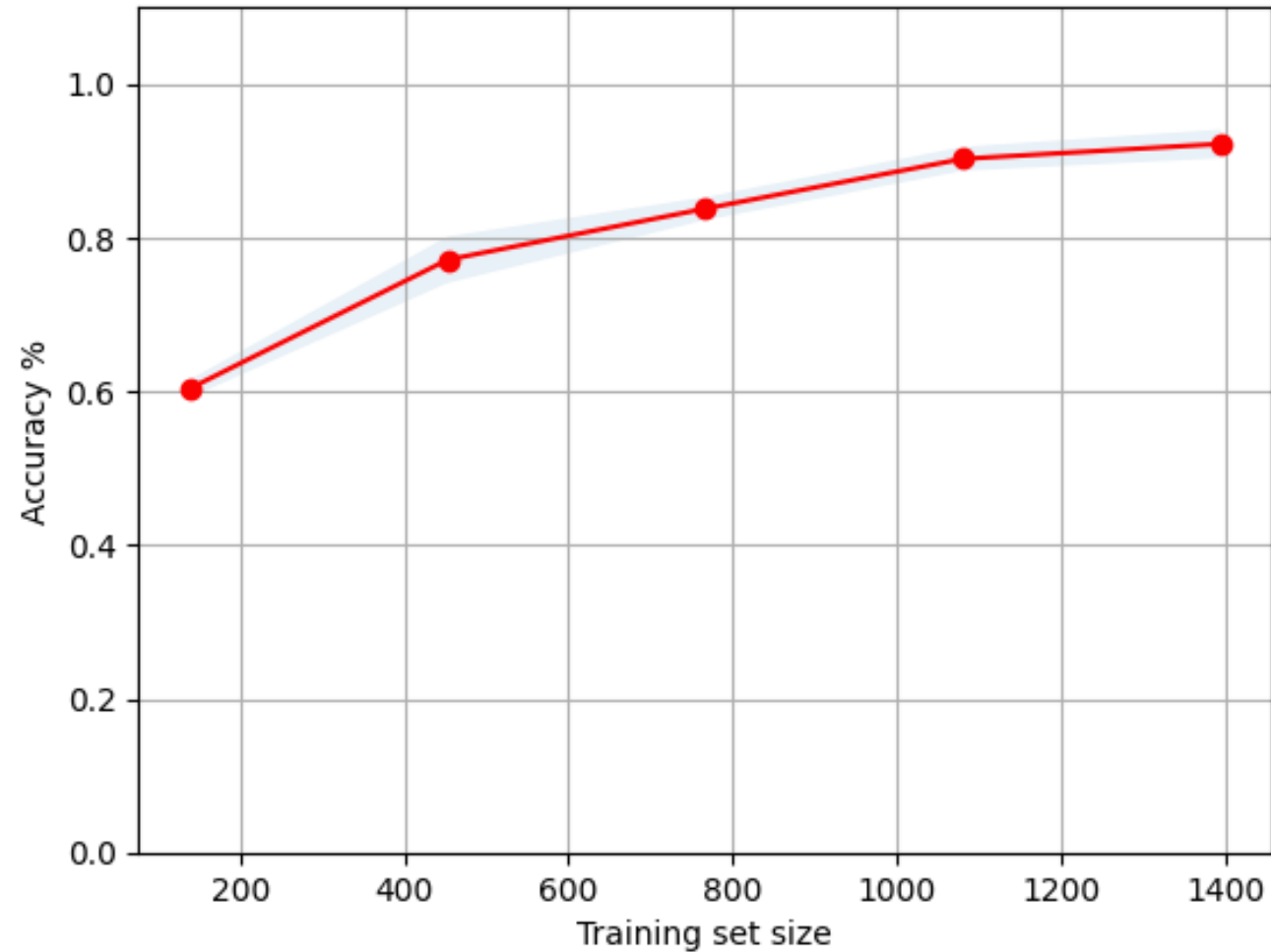
Converted categorical features to
numerically encoded integers

PREDICTIVE MODELS

ML Model/Accuracy	Decision Tree	Random Forest	Gradient Boosting	SVM
Train Accuracy	99%	99%	99.9%	65%
Test Accuracy	93%	95%	94.5%	62%

EVALUATING PREDICTIVE MODELS

Accuracy Plot Model's Performance

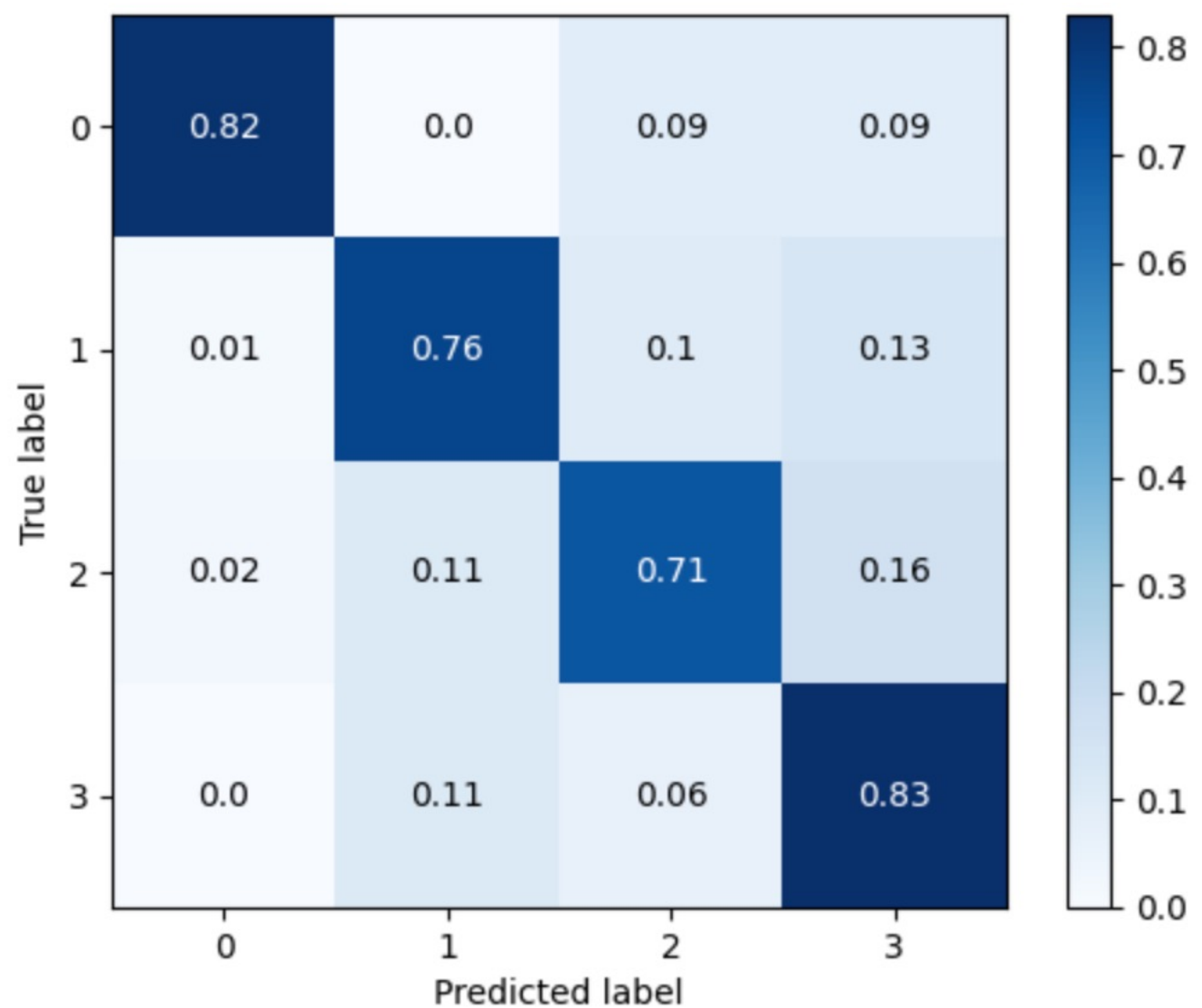


PREDICTIVE MODELS

ML Model/Accuracy	Decision Tree	Random Forest	Gradient Boosting	SVM
Train Accuracy	99%	99%	99.9%	65%
Test Accuracy	96%	94%	94.5%	60%

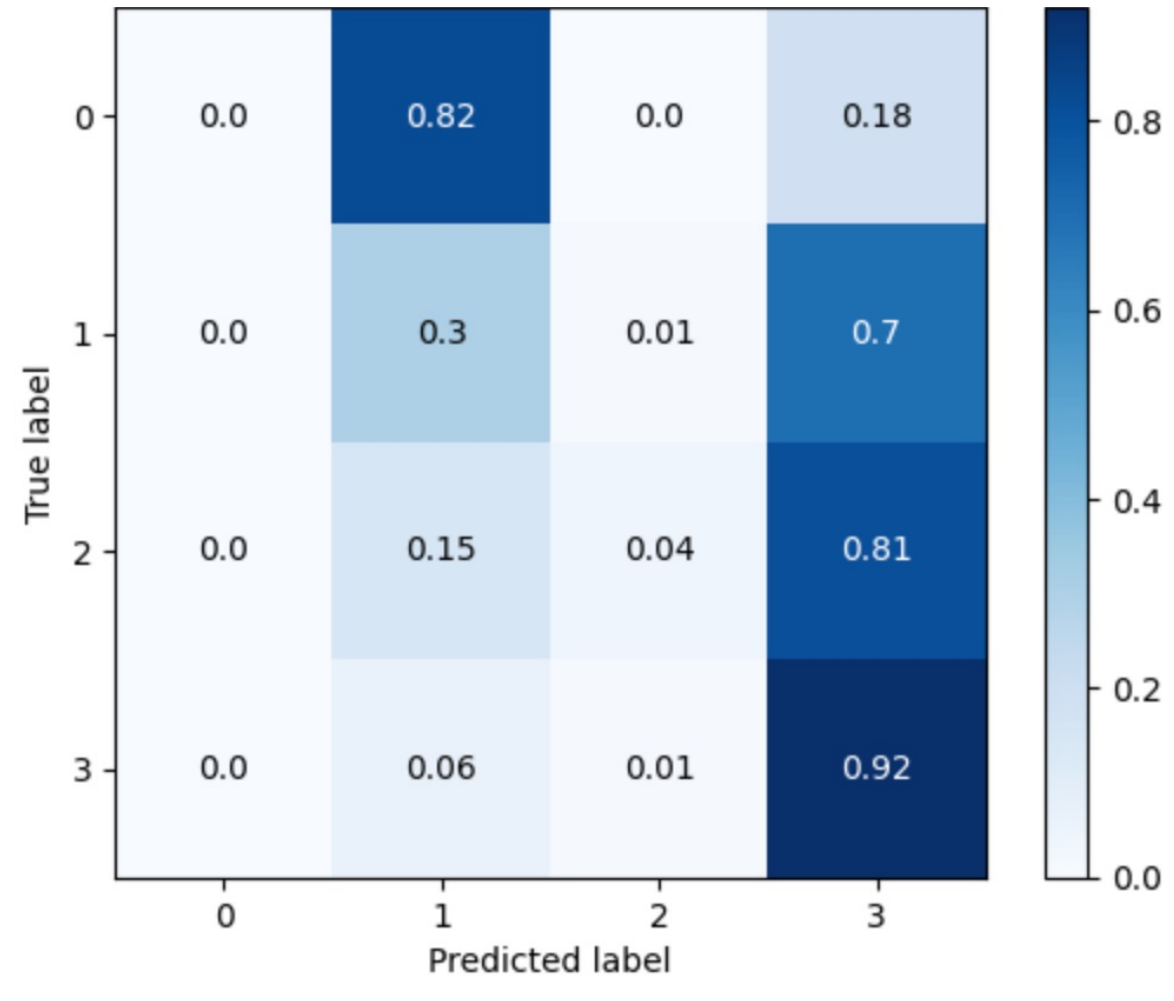
EVALUATING PREDICTIVE MODELS

Confusion Matrix for Decision Tree



EVALUATING PREDICTIVE MODELS

Confusion Matrix for SVM



RESULTS



Success factors Identification

Age-band
Highest education level
Number of previous attempts



Time Management Insights

Optimal time for best student performance is 1 hour 45 Min



Assessment Effectiveness

The Tutor Marked Assessments have the strongest Correlation with student Performance



Model Performance

Decision Tree, Random forest and Gradient Boosting all Performed well with the data while SVM isn't a suitable one

DATA DRIVEN EDUCATION

Educational Institutions

- Early Intervention
- Tailored Learning
- Resource Allocation
- Institutional Improvement

Students

- Motivation and Goal Setting
 - Enhanced Student Experience
 - Time Management
 - Career Readiness
-



FUTURE WORK

Incorporating the model
directly into an application on
the most suitable platform.





THANK YOU