**Week 3: Churn Analysis Report**

# Introduction

## **Objective**

## The objective of this churn analysis is to identify key factors influencing student drop-offs and provide actionable insights to improve student retention. By analyzing historical enrollment data, we aim to uncover patterns that can help institutions implement targeted interventions, ensuring better student engagement and success.

# Data Preparation

## **Data Cleaning**

## To ensure high-quality data, the following steps were taken:

## **Handling Missing Values:** Missing values were identified and either imputed using statistical methods (mean, median) or removed if necessary.

## **Removing Duplicates:** Any duplicate records were dropped to avoid data redundancy.

## **Handling Outliers:** Outliers in numerical data were detected using box plots and handled through transformation or removal if necessary.

### **Feature Selection**

### The following features were selected based on their relevance to student churn:

### **Status**

### **Gender**

### **Country**

### **Institution Name**

### **Major**

### **Opportunity Type**

### No new features were created for this analysis.

# Exploratory Data Analysis (EDA)

## **Descriptive Statistics**

## The dataset was analyzed to obtain key statistics such as mean, median, and standard deviation for numerical variables.

## Categorical variables were explored to understand the distribution of students across institutions, majors, and countries.

### **Visualization**

### To identify patterns and relationships, the following visualizations were used:

### **Bar Charts:** To show the distribution of students across different institutions, majors, and opportunity types.

### **Histograms:** To analyze student status trends.

### **Box Plots:** To detect outliers in numerical features.

### **Patterns and Trends**

Key findings from EDA:

* Certain institutions and majors had higher dropout rates.
* Gender and country also played a role in student retention, but with lower significance.

# Predictive Modeling

## **Model Selection**

#### For churn prediction, we used the Random Forest Classifier, a widely used ensemble learning algorithm. It is preferred for its robustness, ability to handle missing data, and feature importance analysis.

### **Model Training:**

### **Algorithm:** RandomForestClassifier from sklearn.ensemble

### **Parameters:** n\_estimators=100, random\_state=42

### **Training Process:** The dataset was split into training and testing sets to prevent overfitting. The model was trained using X\_train (features) and y\_train (target labels).

#### **Performance Metrics:**

#### **Accuracy:** Evaluated how well the model classified students as retained or dropped.

#### **Precision & Recall:** Assessed how many predicted drop-offs were correct.

#### **F1-score:** Balanced metric for model effectiveness.

##### **Feature Importance:**

##### The model's feature importance scores were extracted and visualized in a bar chart.

##### Institution Name and Major were the most significant predictors of churn.

##### Gender had the least impact on churn prediction.

# Churn Analysis

## **Key Factors Influencing Drop-Offs:**

## Students from certain institutions were more likely to drop out.

## The chosen major significantly influenced retention rates.

## Opportunity Type (e.g., scholarships, grants) played a moderate role.

## Gender and country had a lesser but noticeable effect.

### **Impact Analysis:**

### Students in high-dropout institutions might lack interest in extra courses.

### Specific majors may have higher difficulty levels, leading to increased dropout rates.

### Financial aid and scholarships influence retention positively.

# Recommendations

## **Strategies for Improving Retention:**

## **Targeted Support for High-Risk Institutions** – Provide additional academic and financial support.

## **Enhancing Major-Specific Guidance** – Improve advising services to help students succeed in challenging majors.

## **Gender & Country-Specific Programs** – Tailor student engagement strategies based on demographic trends.

## **Expanding Opportunities** – Increase scholarship and grant availability for students at risk of dropping out.

### **Interventions for At-Risk Students:**

### Early warning systems for students showing signs of disengagement.

### Personalized counseling sessions for students in high-dropout majors.

### Outreach programs to keep students motivated and engaged.

# Conclusion

## **Summary of Key Findings**

## The analysis identified Institution Name and Major as the most influential factors in student drop-offs, followed by Opportunity Type, Country, and Gender. Using the Random Forest Classifier, we detected key patterns that can help improve student retention through targeted interventions.

### **Future Work**

### Future efforts should explore advanced models like XGBoost and incorporate behavioral data for better accuracy. Implementing real-time predictive systems can enable timely support for at-risk students, enhancing overall retention strategies.

# Evaluation Criteria

# 

## **Data Preparation and EDA:** Thorough cleaning and exploration. **Predictive Modeling:** Used an effective model with clear evaluation metrics. **Churn Analysis:** Provided deep insights into the most influential factors. **Report Quality:** Structured, professional, and includes visuals.