Excelerate

Student Engagement and Churn Prediction

AI-Powered

Team Members:
Hammad Mengrani, Kareem Mamdouh, Laiba Jawaid, Priyanka Kute, Reeba Joshi, Rifat Islam, Sumaya Mateen

8th September, 2025

Agenda

- Introduction & Business Problem
- Dataset Overview
- Data Preprocessing
- Exploratory Data Analysis (EDA)
- Churn Analysis
- Model Building
- Model Comparison
- Summary of Strategic Value
- Churn Factors & Recommendations
- Limitations & Future Work
- Conclusion

Introduction & problem Statement

- **Strong Initial Enthusiasm**
  - Students enroll with high motivation to gain skills and career growth.
  - Start with strong expectations, ambition, and promise of opportunities.
- **Unexpected Dropouts**
  - 31.8% of students withdraw silently without warning signals.
  - No clear indicators → difficult to predict or intervene.
  - Dropouts driven by multiple factors (personal, academic, platform-related).
  - Leads to loss of potential and wasted platform resources.

- **Core Problem**
  - High enrollments, but low course completions.
  - Urgent need for predictive solutions to detect risks early and reduce churn.

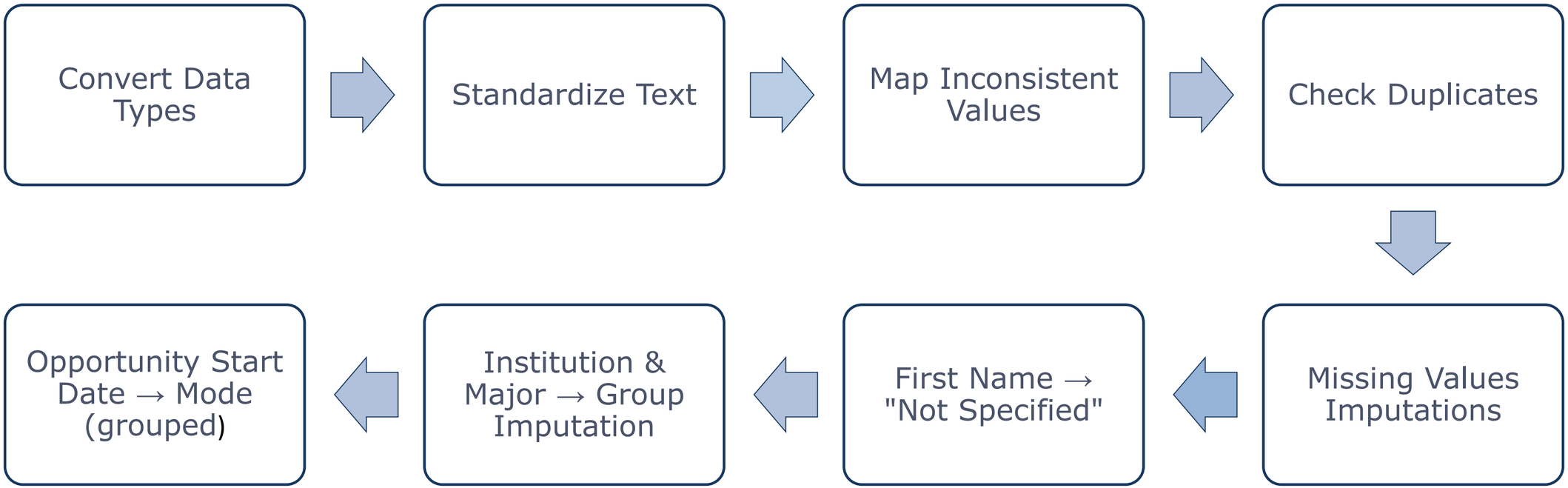
Dataset Overview

Total Records	Total Countries	Success Rate	Peak Signups
8558	70	47.6%	946

- **Scale:** 8,558 learner applications
- **Coverage:** 70 countries, grouped into 5 regions, 23 programs
- **Success Rate:** 47.6% overall with significant demographic variations
- **Key Insight:** Peak activity was in August 2023 with 946 signups

Data Preprocessing

- **Data Cleaning:**



Data Preprocessing

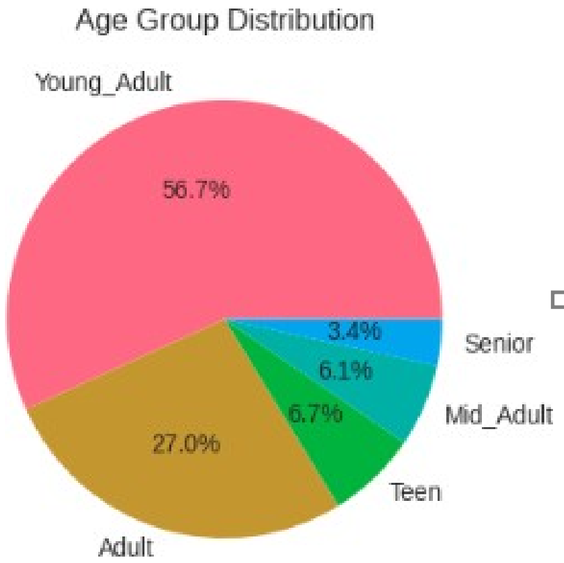
- **Feature Engineering:**

Temporal Features:	Demographic Features:	Behavioral Features:	Composite Features:
Capture time-based patterns	Classify Learners by background	Reflect learner activity & success	Combine multiple attributes
Like: Quick_Applicant, Signup_to_Apply_Days	Like: Age_Group, Region	Like: Engagement_Score, Opportunity_Popularity	Like: Age_Major_Combo, Country_Gender

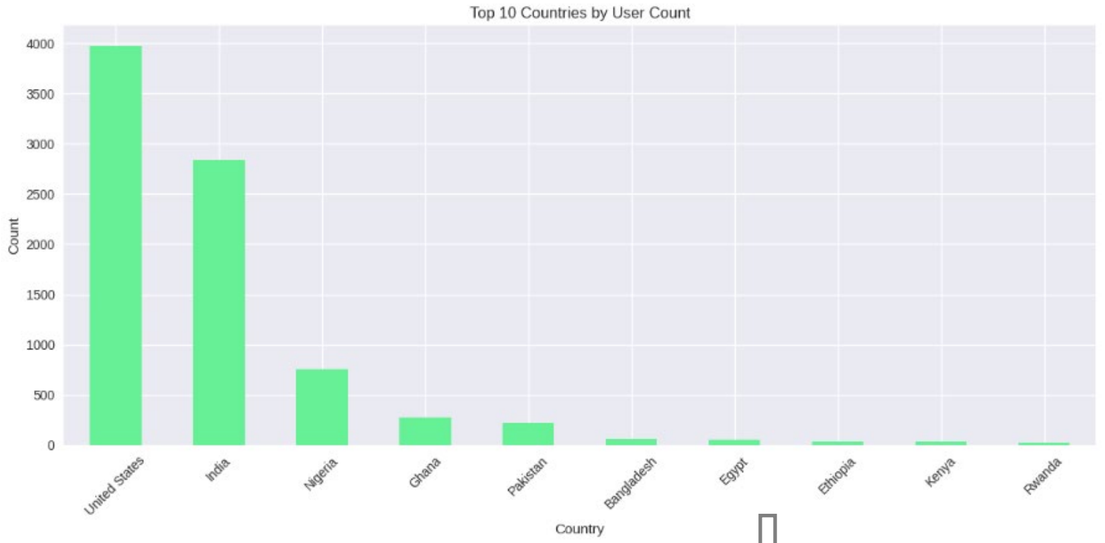
Exploratory Data Analysis

Data Quality	Descriptive Analysis	Visualization & Insights
Validate Structure	Numerical stats	Plot trends & correlation
Handle missing & duplicate	Categorical distributions	Extract key insights

Demographical Insights

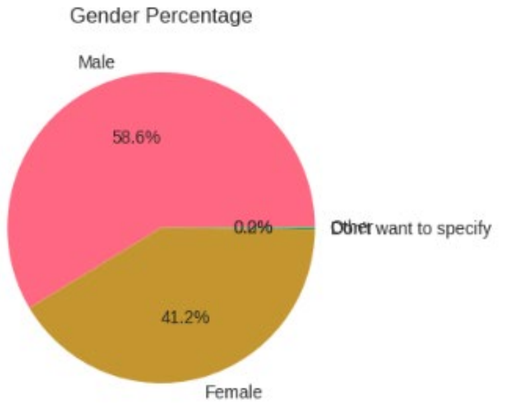


Insight:
Young Adults dominate (56.7%); Teens and Seniors are underserved (<10%), offering key growth potential.



Insight:
Enrollment is heavily concentrated in the United States and India, indicating strong market presence.

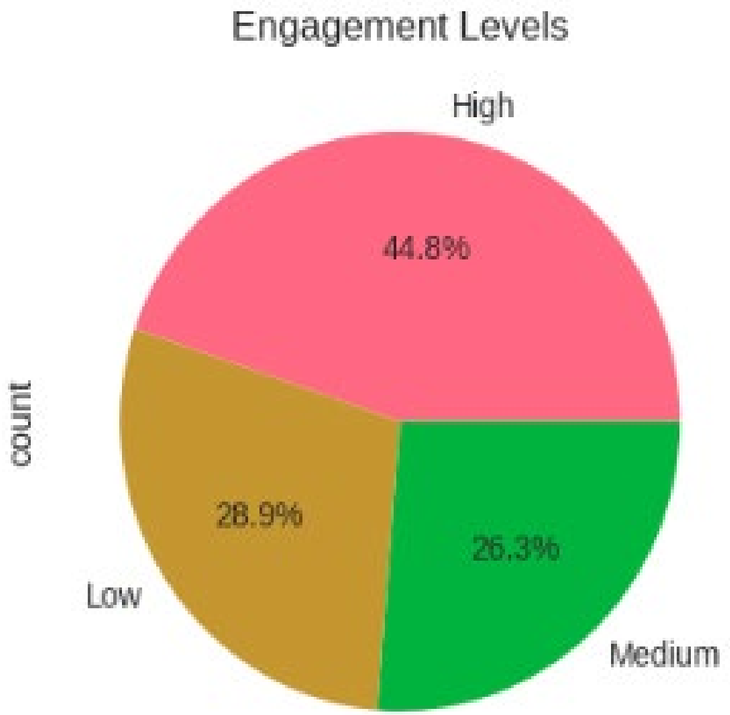
Insight:
The majority of users are male, making up over half of the total user base.



Engagement Distribution

Engagement Metrics:

- Average Engagement Score: 0.47
- Engagement Level Distribution:
  - High: 3,835 (44.8%)
  - Low: 2,472 (28.9%)
  - Medium: 2,251 (26.3%)



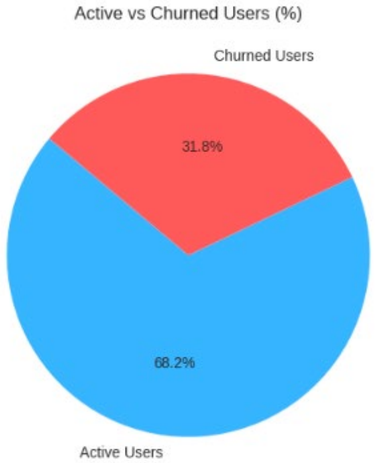
Insights:

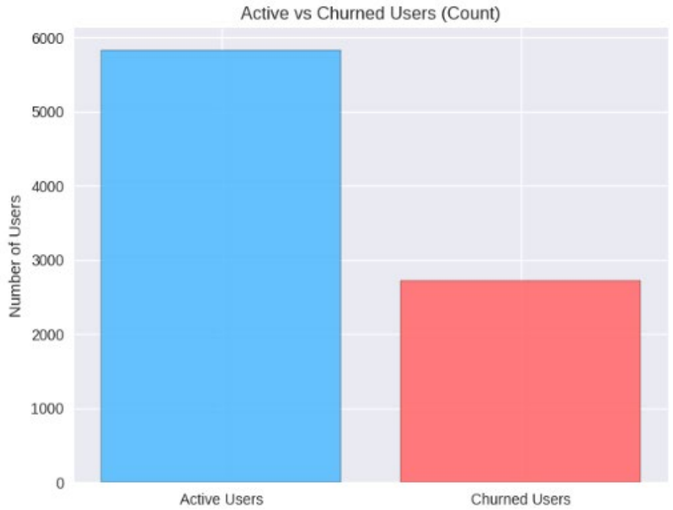
- Retained learners have significantly higher engagement scores.
- Churned users apply more, but have a lower success rate.

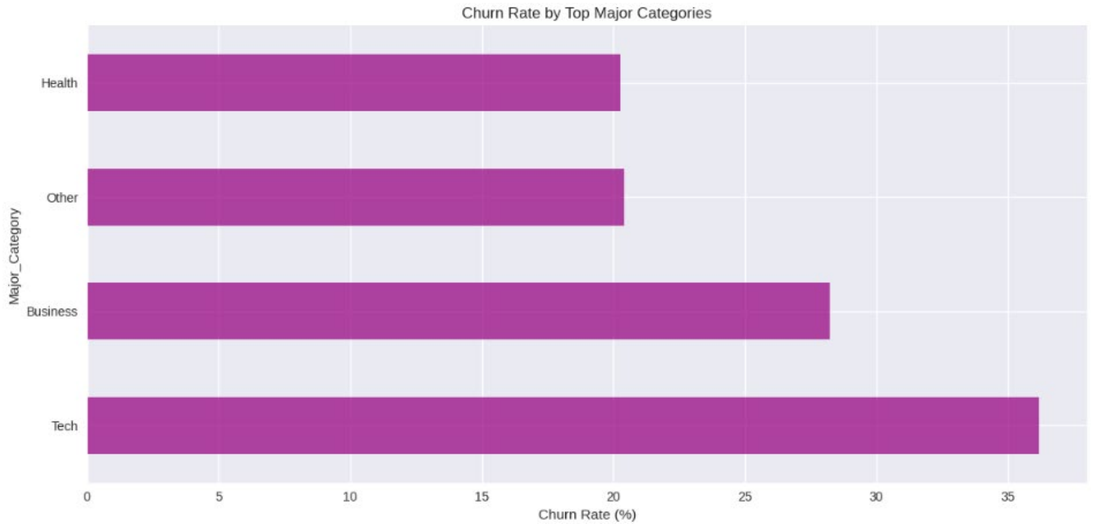
Churn Analysis

Understanding & Defining Churn:

- Challenge: Multiple churn definitions (status codes, behaviors, demographics)
- Approach: Rule-based clustering with Kmeans
- Key Features for Thresholds:
    - Engagement_Score
    - Applications_per_User
    - User_Success_Rate
    - Apply_vs_Start_Gap
- **Churn Rules:**
    - Low/no engagement
    - Many unsuccessful attempts
    - Dropped before course start
    - Very low success despite many applications



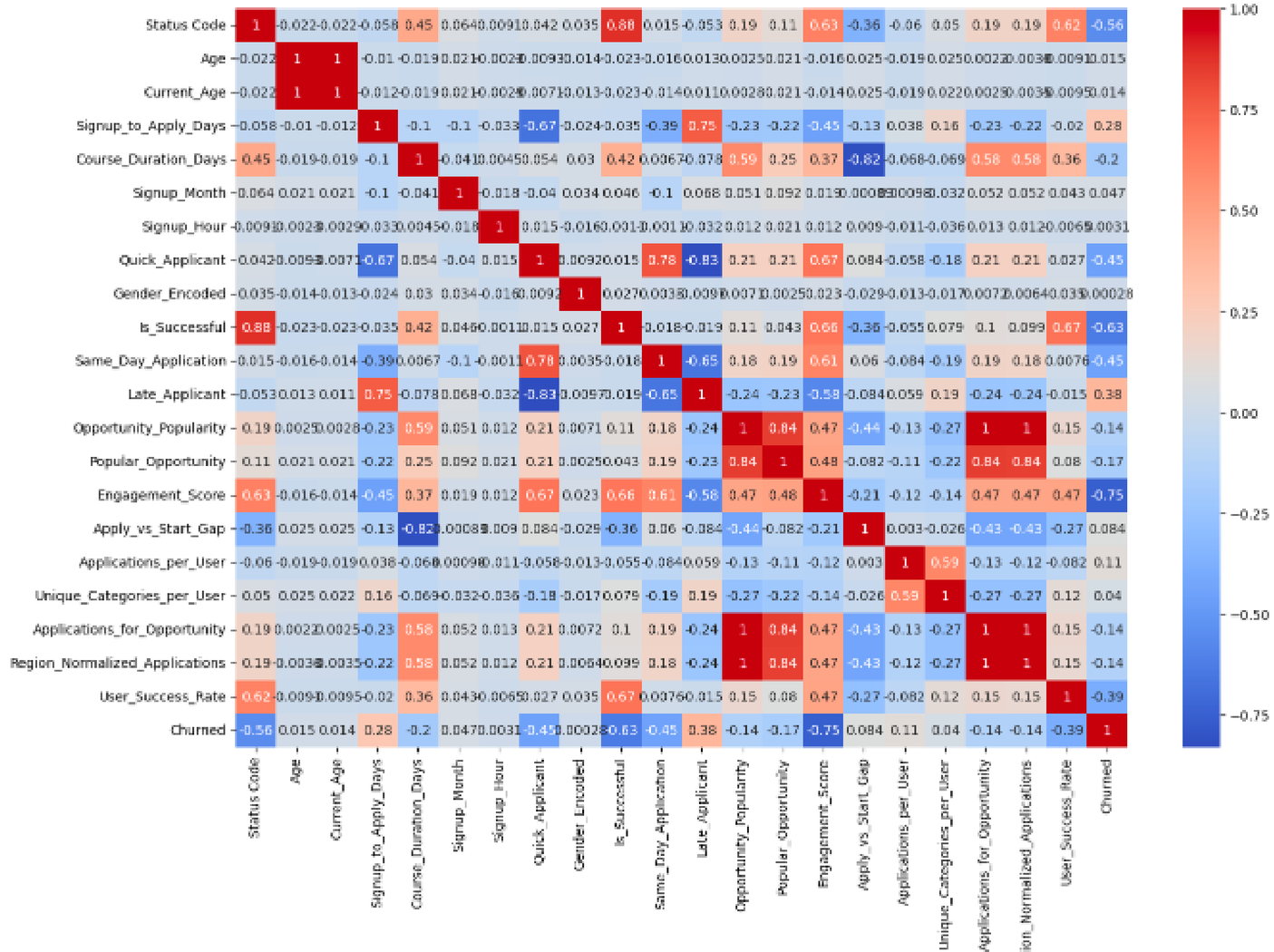




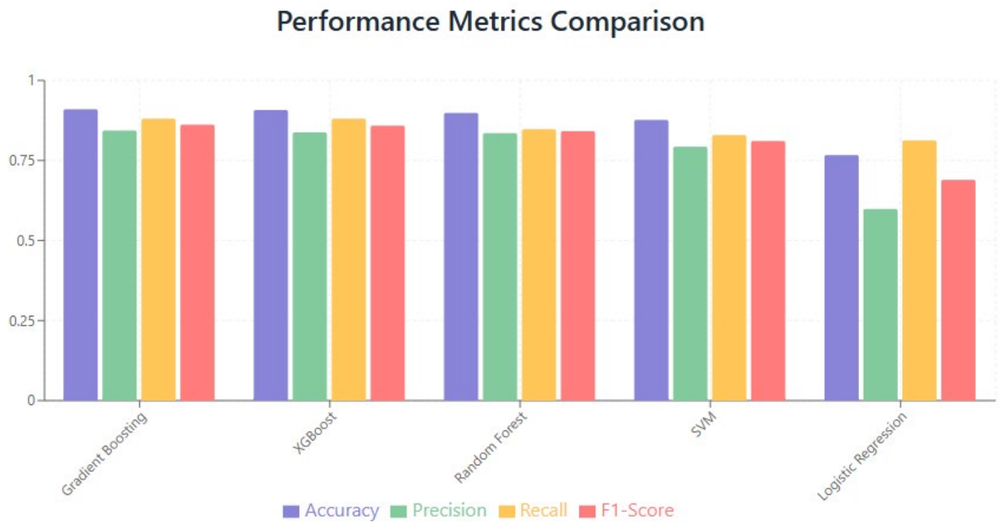
Insight:
Nearly one-third (31.8%) of users have churned, indicating a significant challenge with user retention that requires immediate attention.

Model Building for Churn Prediction

- **Feature Selection to Avoid Leakage:**
  - Correlation analysis for churn-related features
  - Removed status-based features (status_code, status_description)
  - Checking Circular dependency
  - Retained features providing predictive value without leakage
- **Models Tested**: Logistic Regression, SVM, Random Forest, Gradient Boosting, XGBoost
- **Final Model:** Gradient Boosting Classifier and gives 91% accuracy, 0.86 F1-score



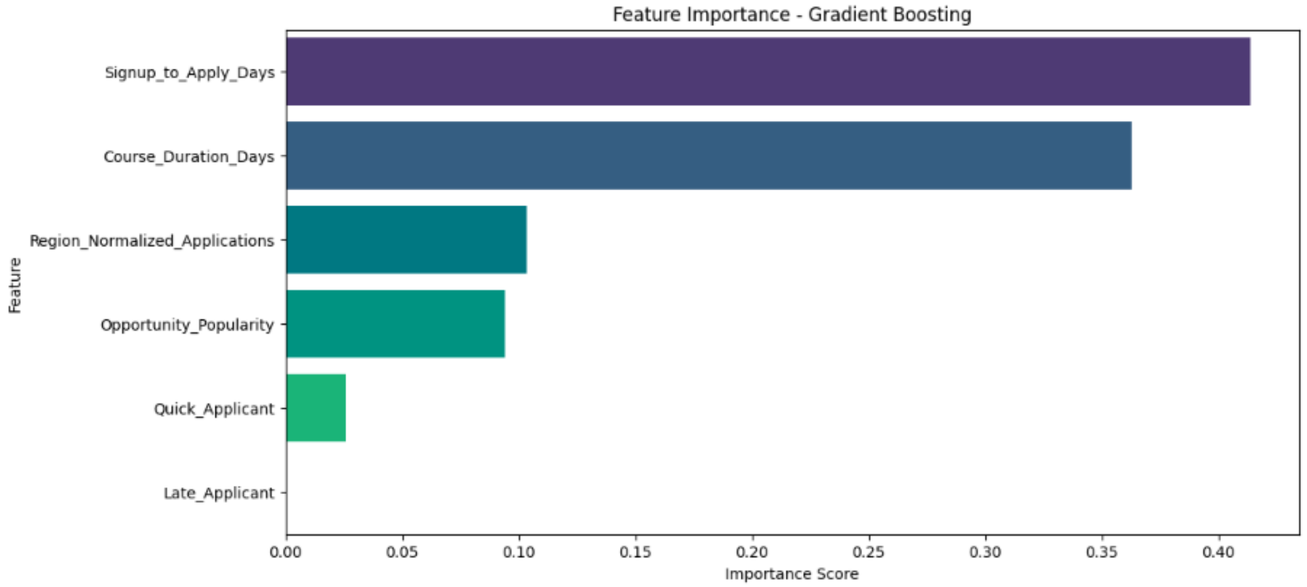
Model Comparison



Model Rankings

1	Gradient Boosting F1-Score: 0.8618	91.0% Accuracy
2	XGBoost F1-Score: 0.8587	90.8% Accuracy
3	Random Forest F1-Score: 0.8415	89.8% Accuracy
4	SVM F1-Score: 0.8108	87.7% Accuracy
5	Logistic Regression F1-Score: 0.6895	76.7% Accuracy

Churn Factors



Churn Factors & Recommendations

- **Delayed First Application = Early Dropout Risk**

Strategy: Automate a multi-email onboarding sequence with a video tour and personalized opportunity suggestions to drive immediate action.

- **Long Course Duration = Higher Overwhelm & Churn**

Strategy: Restructure long courses into weekly modules, each with a micro-certificate, to prevent learner overwhelm.

- **Low Regional Application Rate = Localized Disengagement**

Strategy: Develop region-specific content and community hubs to increase local relevance and engagement.

- **Chasing Popular Opportunities = Mismatch-Driven Churn**

Strategy: Enhance the recommendation algorithm with collaborative filtering to surface trending and peer-approved opportunities.

- **Instant Applications = Rushed Decisions, Higher Regret**

Strategy: Improve early onboarding to ensure users understand the platform before applying.

- **Late Applications = Indecision Signals, Likely Churn**

Strategy: Use AI-driven prompts to re-engage users who show signs of indecision early in their journey.

Limitations & Future Plans

- **Current Gaps:** Missing socioeconomic data, Motivation factors, European representation
- **Future Plans:**
  - Deep learning models
  - Real-time analytics
  - A/B testing
  - Social network analysis

Conclusion

- **Identified Core Problem:** Mostly churn rate driven by application delays, course length, and engagement gaps**.**
- **Built Predictive Solution:** Gradient Boosting model achieves 91% accuracy in forecasting student churn.
- **Delivered Actionable Strategies:** Personalized interventions, adaptive content, and targeted campaigns to boost retention.
- **Enabled Proactive Approach:** Shift from reactive support to data-driven, preemptive student success initiatives.

Thank You!