

# **Predicting Student Dropout Risk Using Educational Data**

## Abstract

The increasing rates of student dropout pose significant challenges to educational institutions and hinder student success and retention. Identifying students at risk of dropping out early can enable educators to implement timely interventions and support strategies, thereby improving student retention and academic achievement. This paper presents the development of a Django-based application designed to predict student dropout risk using educational data and machine learning models. The application aims to assist educators in identifying at-risk students and implementing targeted support programs to enhance student retention and success.

The application leverages the Django web framework, known for its robustness, scalability, and security features, to create a comprehensive platform for managing and analyzing educational data. Django's capabilities in handling complex data workflows, secure data management, and dynamic user interfaces make it an ideal choice for developing a system that supports educational institutions in their efforts to address student dropout issues.

Central to the application is its data handling module, which manages and processes student educational data. This module integrates with various data sources to collect relevant information, including attendance records, academic performance, and engagement metrics. The data handling module ensures that the data is securely processed and managed, adhering to best practices for data protection and privacy. It includes features for data validation, cleaning, and transformation, preparing the data for analysis and model training.

Preprocessing of educational data is a critical step in the application, involving the extraction of relevant features that are indicative of dropout risk. This step includes tasks such as aggregating attendance records, analyzing grade trends, and evaluating engagement levels. By preprocessing the data, the application ensures that the features used for dropout prediction are accurate and meaningful, enhancing the performance of the machine learning models.

The core functionality of the application is based on machine learning models that predict dropout risk. These models are trained to analyze student data and identify patterns associated with higher dropout probabilities. By using classification models such as Decision Trees and Gradient Boosting, the application can predict the likelihood of a student dropping out based on their educational data. The trained models provide predictions that are used to assess dropout risk and guide intervention efforts.

The application features an interactive dashboard that provides educators with insights into dropout risk predictions. This dashboard displays risk levels for individual students, highlighting those who are at higher risk of dropping out. It includes visualizations such as risk scores, trend analyses, and historical comparisons, offering educators valuable information for decision-making. The dashboard also provides recommendations for targeted support programs, such as counseling, academic tutoring, or personalized learning plans.

In addition to dropout risk predictions, the application generates reports that detail the findings and insights from the analysis. These reports include information on risk factors, prediction accuracy, and recommended interventions. The reports support educators in understanding the factors contributing to dropout risk and in developing strategies to address these issues. The application's reporting capabilities help institutions evaluate the effectiveness of their intervention programs and make data-driven decisions.

The architecture of the application is designed to be modular and adaptable, allowing for future enhancements and the integration of additional features. Potential developments include incorporating new machine learning models for improved prediction accuracy, integrating with institutional databases for real-time data updates, and expanding the platform's capabilities to support additional educational metrics and indicators.

Security and privacy are paramount in the development of the application, given the sensitivity of educational data. The platform implements robust security measures to protect student information, including data encryption, secure authentication, and access controls. Compliance with educational data protection regulations and best practices ensures that the application maintains the confidentiality and integrity of student data.

In summary, this paper outlines the development of a Django-based application for predicting student dropout risk using educational data and machine learning models. By integrating advanced data handling, preprocessing, and predictive analytics, the platform aims to support educators in identifying at-risk students and implementing effective intervention strategies. The application's features contribute to improved student

retention, enhanced academic support, and better overall educational outcomes, advancing the field of educational data analysis and dropout prevention.