

# Student Dropout System for School Education

Avinash Patil<sup>1</sup>, Akshay Chavan<sup>2</sup>, Ajay Chaogule<sup>3</sup>, Pratik Khade<sup>4</sup>, Prof. S.J.Chaogule<sup>5</sup>

<sup>1,2,3,4</sup> Dr. J J Magdum College of Engineering, Jaysingpur

<sup>5</sup> Guide: Dr. J J Magdum College of Engineering, Jaysingpur

*Abstract- The "Student Dropout System for School Education" project aims to develop a predictive model that identifies and addresses the factors leading to student dropouts in educational institutions. By leveraging data from student demographics, academic performance, attendance, and socio-economic background, the system uses machine learning algorithms to predict the likelihood of a student dropping out. The goal is to provide early interventions by alerting teachers, counselors, and administrators about at-risk students, thereby enabling targeted support measures to improve retention rates. This system enhances the ability to provide personalized educational experiences, helping to minimize dropouts and improve overall student success.*

## I. INTRODUCTION

School dropout is a significant issue faced by educational institutions worldwide. It refers to students leaving school before completing their education, which has a negative impact on both the individual and society. High dropout rates can lead to reduced economic opportunities, higher unemployment rates, and increased social challenges.

This project aims to develop a School Dropout Analysis System to predict and analyze the factors contributing to student dropouts. By utilizing data analytics and machine learning, the system will identify students at risk of dropping out and provide early intervention measures. This proactive approach will help schools take timely actions to retain students, improve educational outcomes, and reduce dropout rates.

The system will collect data such as student attendance, academic performance, socio-economic background, and behavioral patterns. Using this data, it will generate insights into the reasons behind student dropouts and enable educators to target specific areas for improvement. By analyzing patterns and trends, the system aims to provide schools with the tools needed to prevent dropouts and ensure every student has the opportunity to succeed.

## II. UNITS

The "Student Dropout System for School Education" project aims to develop a machine learning-based model to predict and prevent student dropouts. It begins by collecting data on student demographics, academic performance, attendance, and other relevant factors. After preprocessing and cleaning the data, key features influencing dropout risk are selected. A suitable machine learning algorithm is then trained using this data, and the model's performance is evaluated and refined. Once trained, the system can predict students at risk of dropping out and alert educators to take early intervention measures. The system is deployed in a school environment, continuously updated with new data, and monitored for its effectiveness in improving student retention.

## III. HELPFUL HINTS

### Litreture Review

The issue of school dropouts is a significant challenge that has been extensively studied. Various factors contributing to dropouts have been identified, and data-driven approaches such as machine learning and predictive analytics have been proposed to tackle the problem.

1. Causes of School Dropout: According to studies, the causes of school dropouts are multifaceted. These include socioeconomic factors (poverty, family income), academic struggles (low grades, lack of motivation), personal factors (health, pregnancy, mental health issues), and external factors (peer pressure, family responsibilities). Research suggests that dropout prevention requires addressing these root causes via support systems (Rumberger, 2011).
2. Use of Data Analytics and Machine Learning: Recent literature highlights the growing use of machine learning models in predicting dropouts. Data from student attendance, academic records, socio-economic background, and engagement levels are being used to train

models that predict students at risk of dropping out. Machine learning techniques like decision trees, logistic regression, random forests, and neural networks have been successful in improving accuracy (Al-Hassan et al., 2016). Predictive models help school administrators intervene early and target the students needing support most effectively.

3. **Interactive Dashboards for Educational Data:** Dashboards are widely used in educational institutions to visualize large datasets and provide actionable insights. Research suggests that interactive visualizations help in quickly identifying patterns and trends in educational performance, attendance, and dropout rates. React.js, coupled with data visualization libraries (e.g., D3.js), has become a popular choice for creating dynamic dashboards (Yuan et al., 2019).
4. **Educational Data Mining for Dropout Prediction:** Educational data mining (EDM) techniques have been used to analyze student data and extract patterns that indicate students at risk of dropping out. Research conducted by Dekker et al. (2009) utilized data from student information systems to build a predictive model for early dropout detection. They employed decision trees and logistic regression to classify students into risk categories, proving that machine learning models can be highly accurate when predicting dropout in educational settings.
5. **Social and Academic Factors in Dropout:** Xiong et al. (2020) conducted a study to determine the interplay between social and academic factors in student dropout. Their research shows that combining social data (e.g., peer interactions, parental involvement) with academic performance data significantly improves the predictive accuracy of dropout models. They also suggest that AI can help automate the monitoring of students' behavior to identify early warning signs of dropping out.

### 3.1. Problem Definition:

The primary issue addressed in this project is the high rate of school dropouts and the challenge schools face in identifying at-risk students early enough to provide necessary interventions. Traditional methods of monitoring students' academic performance and social engagement are reactive and result in interventions only after a student has dropped out. This project aims to create a data-driven, predictive system that can help

schools proactively identify students at risk of dropping out based on factors such as attendance, academic records, socio-economic background, and behavioral patterns.

The proposed School Dropout Dashboard system will enable schools to:

- Analyze dropout reasons across different demographics.
- Predict students at risk of dropping out using machine learning models.
- Provide a user-friendly dashboard for administrators and policymakers to take proactive measures.

### 3.1 Purposed System Work

The project will be divided into the following phases:

1. **Data Collection:**
  - Collect data on students' demographics, academic performance, attendance, and dropout history.
  - Gather potential dropout reasons from school records and surveys (e.g., financial issues, academic struggles, personal problems).
2. **Data Preprocessing:**
  - Clean the collected data to handle missing or inconsistent information.
  - Feature engineering: Transform raw data into meaningful features (e.g., attendance rate, GPA).
  - Normalize and encode categorical data (e.g., gender, location).
3. **Model Training (Backend):**
  - Use machine learning models (Random Forest, Logistic Regression, Decision Trees) to train a predictive model that identifies students likely to drop out based on historical data.
  - Split the dataset into training and testing sets to evaluate model accuracy and performance.
4. **Frontend Development:**
  - Create the interactive dashboard using React.js, allowing users to visualize dropout trends, drill down into specific demographics, and view predictions for at-risk students.

5. Integration of Backend and Frontend:

- Use Python-based web frameworks (Flask/Django) to serve the machine learning model as an API.
- Integrate the backend API with the React.js frontend to display real-time predictions and visualizations.

6. Testing and Validation:

- Test the system with real or simulated school data.
- Measure the accuracy of the predictive model using metrics like precision, recall, and F1 score.

#### IV. SYSTEM ARCHITECTURE

The system architecture for the "Student Dropout Prediction" project consists of several key components:

1. **Data Collection:** Data is gathered from student records, including demographics, academic performance, attendance, and socio-economic factors.
2. **Data Preprocessing:** The collected data is cleaned, transformed, and features are selected for use in prediction models.
3. **Machine Learning Model:** Various machine learning algorithms (e.g., Decision Trees, Random Forest) are used to train the model on the preprocessed data to predict dropout risk.
4. **Prediction & Decision Layer:** The trained model predicts the likelihood of dropout for each student and classifies them into risk categories.
5. **User Interface:** An admin dashboard displays predictions and alerts for at-risk students, helping educators and counselors plan interventions.
6. **Alert & Notification System:** The system sends real-time notifications to administrators or counselors when a student is at high risk of dropping out.
7. **Feedback Loop:** The system continuously collects new data, retrains the model, and refines the predictions and interventions based on user feedback.

This architecture enables early identification of at-risk students and supports timely interventions to reduce dropout rates.

#### Techniques to be Used

For the Student Dropout Prediction System, the key techniques used are:

1. **Data Collection:** Gathering student data from management systems, surveys, and external sources.
2. **Data Preprocessing:** Cleaning data (handling missing values, encoding categorical variables), feature engineering, and normalization.
3. **Machine Learning:** Applying supervised learning algorithms like Logistic Regression, Decision Trees, Random Forest, and Gradient Boosting to predict dropout risk.
4. **Model Evaluation:** Using metrics like accuracy, precision, recall, F1-score, and ROC-AUC to assess performance.
5. **Prediction and Classification:** Classifying students into dropout risk categories and recommending interventions.
6. **Alerts and Notifications:** Sending automated email/SMS alerts to educators when a student is at high risk.
7. **User Interface:** Creating dashboards for administrators to view predictions and track interventions.
8. **Continuous Improvement:** Retraining models with new data and integrating user feedback for better accuracy.

These techniques combine to help predict, manage, and prevent student dropouts effectively.

#### V. CONCLUSION

The School Dropout Dashboard System provides an innovative solution to one of the most pressing issues in education—student dropouts. By leveraging machine learning models and data visualization techniques, the system empowers school administrators with real-time insights into dropout trends and predictions of at-risk students. This proactive approach enables early interventions that can significantly reduce dropout rates and improve overall student retention.

The project demonstrates the value of combining data science, machine learning, and intuitive visual interfaces to tackle educational challenges. As the system evolves, it has the potential to become a comprehensive tool for both educational institutions and policymakers, helping them make informed decisions that contribute to better educational outcomes.

#### APPENDIX

1. Dataset: Includes student demographics, grades, attendance, and behavior. Preprocessed with normalization and encoding.
2. Algorithms:
  - Logistic Regression, Decision Trees, Random Forest, XGBoost, SVM for dropout prediction.
3. Evaluation Metrics: Accuracy, Precision, Recall, F1-score, ROC-AUC.
4. Tools:
  - Python, Scikit-learn, XGBoost, Flask/Django, Twilio, MySQL.
5. Architecture: Web-based UI (React), backend (Flask/Django), integrated with machine learning model for predictions and alerts.
6. Challenges: Data imbalance, missing values, and model generalization.
7. Future Work: More data sources, real-time feedback, and exploring deep learning.

#### ACKNOWLEDGMENT

First of all I would like to thank Prof. *S.J. Chougule* who is presently working as an Assistant Professor of Computer Science & Engineering Department for guiding me through this project work. I am extremely grateful to him for all his invaluable guidance and kind suggestions during the Phase-I of my project work. His ever encouraging attitude, guidance and whole hearted help were biggest motivation for me in completing this project work.

I am thankful to the founder Chairman Late Dr. J. J. Magdum and the Chairman Mr. Vijayraj J. Magdum of Dr. J. J. Magdum Trust, Jaysingpur, for their encouragement. I am very grateful to Dr. S. S. Admuthe, Campus Director, Dr.G.V.Mulgund, Principal of Dr. J. J. Magdum College of Engineering, Jaysingpur for motivating me for this project work. Also I am thankful to Mr.

Prof.R.A.Bharatiya, Head of Department, Information Technology for providing necessary facilities for completion of this project work.

Lastly I thank all the people who have guided and helped me directly or indirectly.

#### REFERENCES

- [1] Rumberger, R.W. (2011). *Dropping Out: Why Students Drop Out of High School and What Can Be Done About It*. Harvard University Press.
- [2] Al-Hassan, W., Al-Shargabi, B., & Salem, A. (2016). Prediction of Students' Dropout in Higher Education Using Machine Learning Techniques. *International Journal of Advanced Computer Science and Applications (IJACSA)*.
- [3] Yuan, T., & Chen, J. (2019). Development of Education Data Dashboard Using React.js and D3.js. *Proceedings of the International Conference on Computing, Networking and Communications (ICNC), IEEE*.
- [4] Dekker, G. W., Pechenizkiy, M., & Vleeshouwers, J. M. (2009). Predicting Students Drop Out: A Case Study. *Proceedings of the 2nd International Conference on Educational Data Mining*.
- [5] Xiong, Y., Li, Y., & Ma, J. (2020). The Role of Social and Academic Factors in Predicting Student Dropout: A Machine Learning Perspective. *Journal of Educational Computing Research*.