JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Implement Software Solutions to Reduce Student Dropout Rates at Various Education Stages

[1]Mr.R.S.Sathyaraj,
[2]Ms.M.Muthulakshmi, [3]Ms.K.Pennita Jasmine, [4]Ms.V.Periya Sathiya, [5]Ms.K.Santhiya

[1]Professor- Computer Science And Engineering, Jayaraj Annapackiam CSI College of Engineering, Nazareth, Tamil Nadu
[2]UG- Computer Science And Engineering, Final Year, Jayaraj Annapackiam CSI College of Engineering, Nazareth, Tamil Nadu
[3]UG- Computer Science And Engineering, Final Year, Jayaraj Annapackiam CSI College of Engineering, Nazareth, Tamil Nadu
[4]UG- Computer Science And Engineering, Final Year, Jayaraj Annapackiam CSI College of Engineering, Nazareth, Tamil Nadu
[5]UG- Computer Science And Engineering, Final Year, Jayaraj Annapackiam CSI College of Engineering, Nazareth, Tamil Nadu

ABSTRACT

This project introduces a comprehensive predictive analytics solution aimed at identifying students at risk of dropping out of school education. Leveraging the power of machine learning, the solution integrates a Random Forest classifier with a Feature Tokenizer Transformer (FTT) to effectively analyze and interpret both structured and unstructured data. The primary objective is to accurately predict the likelihood of student dropouts based on a wide range of influential factors, including academic performance, attendance records, behavioral patterns, socioeconomic background, and qualitative teacher feedback. The Random Forest algorithm is utilized for its robustness and ability to handle high-dimensional data, building a reliable predictive model that can classify students into risk categories. To enhance the model's capability, the Feature Tokenizer Transformer is employed to convert complex textual data—particularly teacher comments and feedback—into quantitative features that can be incorporated into the prediction pipeline. This enables the system to capture subtle behavioral cues and emotional indicators that may not be reflected in numerical data alone. By combining traditional data analytics with advanced natural language processing (NLP) techniques, this project aims to offer an early-warning system for educators and administrators. The end goal is to support timely interventions, reduce dropout rates, and ultimately contribute to better educational outcomes and equity in student support systems.

1. INTRODUCTION

Student dropout is a persistent and critical issue in the field of education, with far-reaching consequences for both the individual and society. When students leave school prematurely, it not only disrupts their personal and professional development but also affects the overall performance and reputation of educational institutions. Addressing this challenge requires proactive strategies that go beyond traditional methods and leverage the power of data-driven insights. This project presents an innovative approach .to dropout prediction by integrating Random Forest, a robust machine learning algorithm, with a Feature Tokenizer Transformer (FTT) for natural language processing. The aim is to develop a predictive model that can accurately identify students who are at high risk of dropping out. By examining a combination of quantitative factors—such as academic performance, attendance history, disciplinary records, and socioeconomic background—alongside qualitative inputs like teacher feedback and behavioral observations, the system creates a holistic profile for each student. The Random Forest algorithm is chosen for its ability to handle complex, multi-dimensional data and provide high predictive accuracy. To enhance the model's ability to interpret unstructured text data, the FTT component is used to transform teacher comments and feedback into meaningful numerical representations. This enables the model to capture nuanced behavioral signals and emotional indicators that might otherwise go unnoticed. Through this comprehensive analysis, the project aims to equip educators and school administrators with early-warning capabilities. The ultimate goal is to enable timely interventions, foster a more supportive learning environment,

and significantly reduce student dropout rates, thereby improving student retention and educational outcomes.

2. PROBLEM IDENTIFIED

Student dropout rates are a significant challenge in educational institutions, affecting both individual academic progress and overall educational outcomes. Early identification of students at risk of dropping out is crucial for implementing timely interventions. However, many schools struggle to predict which students are most likely to drop out due to a lack of comprehensive analysis. Traditional methods of identifying at-risk students often rely on manual observation and subjective assessments, which may overlook key indicators.

3. AIM AND OBJECTIVES:

The aim of this project is to develop a predictive analytics model to analyze and predict student dropouts in school education using Random Forest and Feature Tokenizer Transformer (FTT). The model leverages various student-related factors to identify students at risk of dropping out, enabling early intervention to
improve student retention.

- To build a Random Forest model for dropout prediction.
- To evaluate model performance using key metrics.
- To identify key factors influencing student dropout.
- To provide insights for early intervention strategies.

4. LITERATURE REVIEW

1. Multilayer Fuzzy Inference System for Predicting the Risk of Dropping Out of School at the High School Level (2024)
Objective: Predict high school dropout risk using a neurofuzzy system.
Methodology: Utilizes fuzzy inference to analyze economic and social factors influencing dropout risk.
Algorithm: Neuro-Fuzzy System.

2. Deep FM-Based Predictive Model for Student Dropout in Online Classes (2023)
Objective: Predict dropout in online courses using a DeepFM model.
Methodology: Combines factorization machine and deep neural networks (DNN) for prediction. Algorithm: DeepFM (Factorization Machine + DNN).

3. Predicting Student-Teachers Dropout Risk and Early Identification: A Four-Step Logistic Regression Approach (2022)
Objective: Predict student teacher dropout risk. Methodology: Applies logistic regression on personal, academic, and socioeconomic data. Algorithm: Logistic Regression.

4. Determination of Dropout Student Profile Based on Correspondence Analysis Technique (2019)
Objective: Analyze dropout student profiles. Methodology: Uses correspondence analysis and chi-square tests to identify dropout patterns. Algorithm: Correspondence Analysis.

5. EXISTING SYSTEM

The current systems used by educational institutions to predict student dropout primarily rely on traditional methods and manual monitoring. These methods typically involve tracking student attendance, grades, and behavior but often lack advanced predictive capabilities. Most systems analyze basic statistics to identify trends, but they do not provide a comprehensive, data-driven prediction model for student dropout.

a. Manual Monitoring:

Teachers and administrators rely on their observations to identify at-risk students.

b. Basic Statistical Analysis:

The use of basic metrics like attendance rates, exam

6. DISADVANTAGES

- Limited predictive capabilities.
- Reliance on manual, subjective monitoring.
- Inability to process unstructured data (e.g., teacher feedback).
- Delayed intervention due to lack of real-time analysis.
- Limited data integration across various sources.

7. PROPOSED SYSTEM

The proposed system aims to predict student dropouts using advanced predictive analytics. It leverages Random Forest for classification and Feature Tokenizer Transformer (FTT) to process and transform unstructured data, such as teacher feedback, into useful features.

7.1 Data Integration:

Combines structured data (academic, attendance) and unstructured data (teacher feedback, behavioral comments).

7.2 Feature Tokenizer Transformer (FTT):

Processes and converts textual data into meaningful features for analysis.

7.3 Real-Time Predictions:

Provides early insights and actionable recommendations for timely intervention.

7.4 Predictive Analytics:

Uses Random Forest to accurately predict dropout risks based on multiple student factors.

8. ADVANTAGES

• Identifies at-risk students early for timely intervention.
• Combines structured and unstructured data for better predictions.
• Provides actionable insights for immediate action.
• Helps educators make informed decisions based on data analysis.
• Reduces dropout rates by proactively addressing student challenges.
• Allows for targeted interventions based on individual student needs.

9. MODULE LIST

9.1 Data Collection and Preprocessing

This module focuses on gathering both structured and unstructured data relevant to student performance and engagement. Structured data may include grades, attendance records, and socioeconomic details, while unstructured data involves qualitative inputs like teacher comments. Preprocessing involves cleaning the data, handling missing values, encoding categorical variables, and preparing textual feedback for analysis using the Feature Tokenizer Transformer (FTT).

9.2 Prediction and Classification

In this core module, a machine learning model—specifically a Random Forest classifier—is trained on the processed dataset to predict the likelihood of student dropout. The model classifies students into predefined risk categories (e.g., low, medium, high risk) based on various features. This classification enables targeted monitoring of at-risk students.

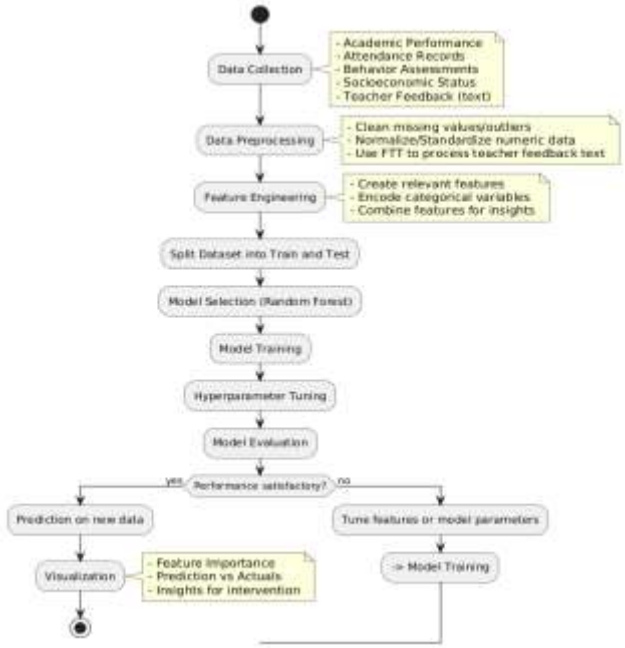
9.3 Hyperparameter tuning

To improve the model's accuracy and efficiency, this module involves optimizing the hyperparameters of the Random Forest algorithm. Techniques like grid search or random search are applied to find the best combination of parameters (e.g., number of trees, max depth, split criteria) that enhance model performance while avoiding overfitting.

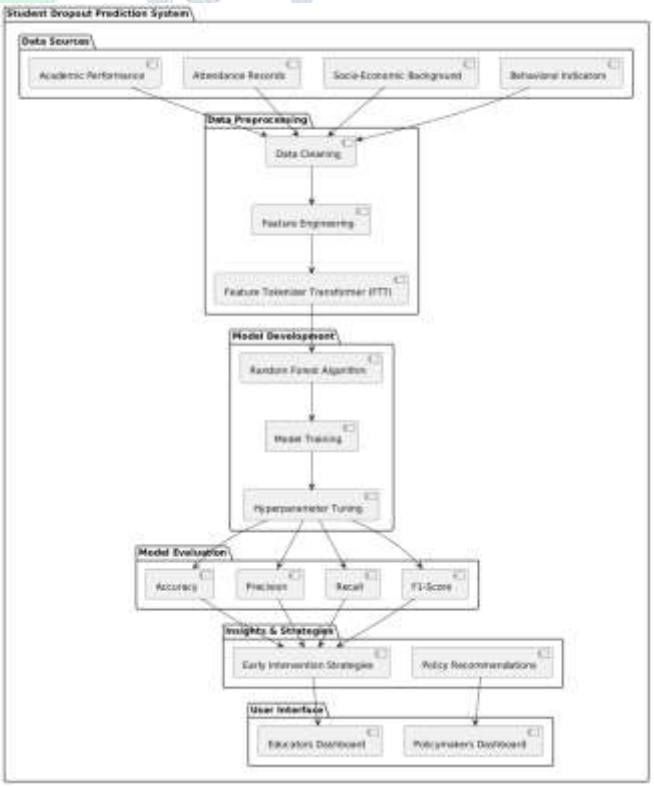
9.4 Dropout mitigation

This final module translates predictions into action. By identifying at-risk students early, educators and administrators can implement targeted interventions, such as counseling, academic support, or engagement programs. The module also includes monitoring and feedback mechanisms to assess the effectiveness of interventions and continuously refine the support strategy.

10. FLOWCHART



11.ARCHITECTURE



11. SYSTEM REQUIREMENTS

11.1 HARDWARE REQUIREMENTS

Processors	: Intel Core
Ram	: 8 GB of Ram
DRAMDisk space	: 320 GB
Operating systems	: Windows® 10

11.2 SOFTWARE REQUIREMENTS

Language	: Python 3.7.4(64-bit) or (32-bit)
UI	: HTML, CSS, Bootstrap
Web Framework	: Flask 1.1.1
Back End	: MySQL 5.
Server	: Wampserver 2i
Packages	: Pandas, Matplotlib, Sickit Learn

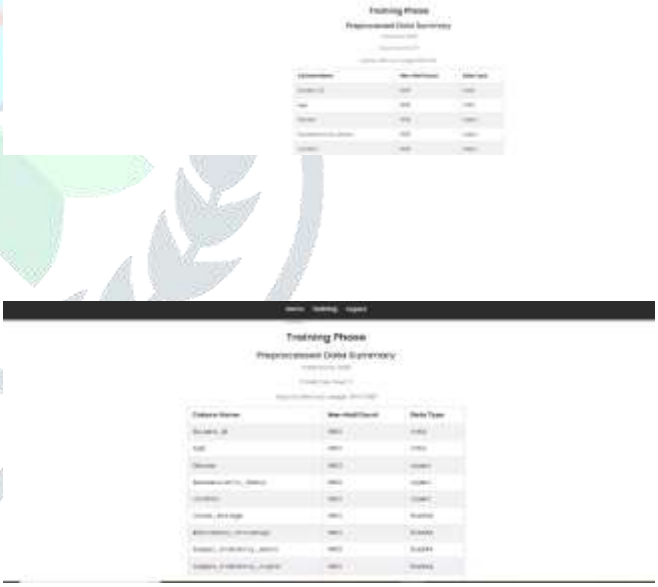
11. SCREENSHOTS

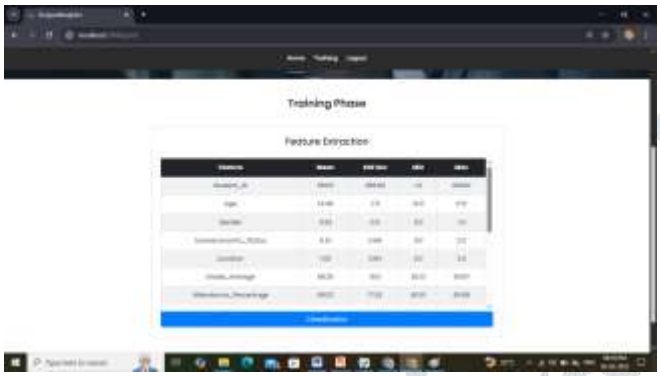
HOME PAGE :

SIGN IN PAGE:

TRAINING PHASE PAGE:

ADMIN PAGE:





USER PAGE:

SIGN IN PAGE:

USER PROFILE PAGE:

RESULT PAGE:

12 . REFERENCE

A. N. de Vasconcelos, L. A. Freires, G. D. L. Loureto, G. Fortes, J. C. A. da Costa, L. F. F. Torres, et al., "Advancing school dropout early warning systems: The IAFREE relational model for identifying at-risk students", Frontiers Psychol., vol. 14, Jul. 2023.

W. M. Ei Leen, N. A. Jalil, N. M. Salleh and I. Idris, "Dropout early warning system (DEWS) in Malaysia's primary and secondary education: A conceptual paper", Proc. Int. Conf. Inf. Syst. Intell. Appl., vol. 550, pp. 427-434, 2023.

M. Yagci, "Educational data mining: Prediction of students' academic performance using machine learning algorithms", Smart Learn. Environments, vol. 9, no. 1, Dec. 2022.

Y. Chen, Q. Zheng, S. Ji, F. Tian, H. Zhu and M. Liu, "Identifying at-risk students based on the phased prediction model", Knowl. Inf. Syst., vol. 62, no. 3, pp. 987-1003, Mar. 2020.

Z. Alharbi, J. Cornford, L. Dolder and B. De La Iglesia, "Using data mining techniques to predict students at risk of poor performance", Proc. SAI Comput. Conf. (SAI), pp. 523-531, Jul. 2016.

13. CONCLUSION

This project presents a comprehensive and data-driven solution to the critical issue of student dropouts in school education. By integrating the predictive power of the Random Forest algorithm with the natural language processing capabilities of the Feature Tokenizer Transformer (FTT), the system effectively analyzes both structured and unstructured data to identify students at risk of dropping out. The model not only achieves high accuracy in prediction but also offers valuable insights into the key factors influencing student retention, such as academic performance, attendance, behavioral patterns, socioeconomic conditions, and qualitative teacher feedback.