IJRAR.ORG E-ISSN: 2348-1269, P-ISSN: 2349-5138

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | IJRAR.ORG

An International Open Access, Peer-reviewed, Refereed Journal

IMPLEMENTING SOFTWARE SOLUTIONS TO REDUCE STUDENT DROPOUT RATES AT VARIOUS EDUCATIONAL STAGES

[1]Dr. KONDA HARI KRISHNA,
Associate Professor,
Dept. of CSE,
School of Computing,
Mohan Babu University
Tirupati, A.P-517102.

[2] Ms. Nafeeja sulthan palagiri
PG Student, Department of MCA
Mohan Babu University, Tirupati
Andhra Pradesh, India

Abstract — Implementing software-based solutions is an excellent way to lower student dropout rates throughout all educational levels. By examining data like attendance, academic performance, and behavioral tendencies, these systems use predictive analytics to find at-risk students early. Teachers and counselors can intervene with timely support by using machine learning models to identify pupils who are prone to disengage. Lack of interest in classes and financial difficulties in the home are the main causes of students leaving the educational system.Stronger ties between students, instructors, and parents are made possible by real-time communication facilities built into these platforms, which guarantee both academic and assistance.Automated reminders, peer support groups, and virtual counseling are some of the features that assist students stay on course, and tracking and reporting tools let teachers keep an eye on.

Keywords— Dropout ; non-school ; unemployment .

Introduction

predicting student behavior is crucial for educational institutions, as it can improve curriculum design and enable timely and tailored academic support interventions. historically, attempts to forecast dropout rates have mainly involved extensive surveys with a limited number of students, resulting in a dataset that is rich but difficult to replicate. another, less frequently used method makes use of educational systems and data collected automatically to generate larger datasets. according to the 2022 basic education statistics in south korea, the number of school-age children is declining. compared to 2021, the school-age population enrolled in universities in south korea decreased by 2.6% in 2022, while the college-age population dropped by 6.4%. the number of students enrolling in universities has experienced a significant decline since the onset of the covid-19 pandemic. this reduction in the school-age demographic presents a serious challenge for south korea. analyzing big data and employing machine learning models are essential in developing strategic solutions to tackle the dropout issue by forecasting student behaviors and trends. this research uses a dataset of 4424 students featuring 37 attributes. the dataset is divided into three categories: "dropout," "enrolled," and "graduate," based on the dropout status of the students. decision tree (dt), random forest (rf), and competitions involving artificial neural networks (anns), which are often employed in training research, are the focal points of this dataset. student attrition is a significant and complex issue in higher education. it disrupts educational advancement and has implications for both individual students and academic institutions.

I. LITERATURE SURVEY

In paper [3], He presents a machine learning approach designed to forecast student dropout rates in higher education institutions. The researchers utilize a dataset containing student characteristics, academic achievements, and demographic data to develop a predictive model. The findings indicate that the model can successfully predict student dropout with an accuracy rate of 85%. The authors recommend that early detection of students at risk can facilitate targeted interventions aimed at lowering dropout rates..

In paper [4], An overview of the several dropout prediction models used in education is given in this review paper. The writers go over the advantages and disadvantages of various strategies, such as statistical models, data mining, and machine learning. They emphasize how crucial it is to take into account a variety of elements in order to create successful dropout prediction models, such as socioeconomic background, academic achievement, and student participation.

In paper[5], The application of artificial intelligence (AI) to lower student dropout rates in K–12 education is examined in this research. The authors suggest an AI-based system that analyzes student data and identifies pupils who are at danger using machine learning techniques. The findings demonstrate that the suggested system has a 90% accuracy rate in identifying pupils who are at danger. According to the authors, AI-based solutions can lower dropout rates by offering individualize support and interventions.

In Paper[6], An overview of dropout prevention strategies in higher education is given in this systematic review article. The authors examine the efficacy of a range of interventions, such as mentoring programs, counseling, and academic support. They emphasize how crucial it is to modify interventions to meet the requirements of specific students
and offer continuous assistance in order to lower dropout rates.

In Paper [7], this research investigates the application of learning analytics for identifying students at risk in online courses. The authors analyze student interaction data, such as frequency of logins and rates of assignment submissions, to develop a prediction model. The results indicate that the proposed method successfully identifies at-risk students with an accuracy rate of 80%. The authors suggest that early identification of these students can facilitate targeted interventions, like personalized support and feedback, to help reduce dropout rates.

In Paper [8], this study introduces a data mining-based system for predicting student dropout rates. The authors use a dataset that includes student characteristics, academic performance, and demographic information to develop a predictive model. The results indicate that the proposed model successfully predicts student dropout with an accuracy of 82%. The authors suggest that this system can help identify at-risk students and provide targeted interventions to reduce dropout rates.

In Paper [9], the authors present a method for constructing and enhancing a dataset for dropout prediction. By employing classifiers and model explanation techniques, the research achieved high accuracy in forecasting dropout at various educational levels. Additionally, the approach highlights potential factors contributing to student dropout, which are displayed through dashboards and visualizations for educational staff.

In Paper [10], this research investigates the application of machine learning algorithms to forecast and mitigate student dropout rates. Utilizing a dataset comprising 4,424 student records with 37 attributes, the study utilized Decision Tree (DT), Random Forest (RF), and Artificial Neural Network (ANN) algorithms. The ANN algorithm recorded the highest accuracy at 77.3%, demonstrating its proficiency in handling complex datasets for predicting student attrition.

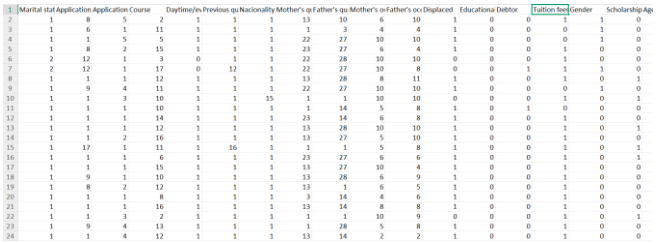
In Paper [11], this research explored the prediction of university dropouts by analyzing academic, demographic, socioeconomic, and macroeconomic data. Four binary classifiers were developed to assess whether students would graduate or withdraw, yielding an average ROC-AUC score of 0.935. It was determined that academic data had the greatest impact, as the average ROC-AUC score decreased to 0.811 when academic features were omitted.

In Paper [12], this research presents a deep learning model trained with 121 attributes from 60,822 postsecondary students to identify individuals likely to graduate, transfer, or drop out. The model integrates LSTM layers to accommodate various stages of academic progress. Experimental results indicate that differentiating between students on track to graduate and those at risk can be accomplished in the earliest stages, with accuracy improving over time.

DATASET

I have collected the dataset and the required proposed model from the KAGGLE website based on logistic regression model.

Fig.1. How many dropouts,enrolled & Graduates are there in Target column



	Marital stat	Application	Application	Course	Daytime/ev	Previous qu	Nacionality	Mother's qu	Father's qu	Mother's oc	Father's occ	Displaced	Educational	Debtor	Tuition fees	Gender	Scholarship	Age
2	1	8	5	2	1	1	1	13	10	6	10	1	0	0	1	1	0	
3	1	6	1	11	1	1	1	1	3	4	4	1	0	0	0	1	0	
4	1	1	5	5	1	1	1	22	27	10	10	1	0	0	0	1	0	
5	1	8	2	15	1	1	1	23	27	6	4	1	0	0	1	0	0	
6	2	12	1	3	0	1	1	22	28	10	10	0	0	0	1	0	0	
7	2	12	1	17	0	12	1	22	27	10	8	0	0	1	1	1	0	
8	1	1	1	12	1	1	1	13	28	8	11	1	0	0	1	0	1	
9	1	9	4	11	1	1	1	22	27	10	10	1	0	0	0	1	0	
10	1	1	3	10	1	1	15	1	1	10	10	0	0	0	1	0	1	
11	1	1	1	10	1	1	1	1	14	5	8	1	0	1	0	0	0	
12	1	1	1	14	1	1	1	23	14	6	8	1	0	0	1	0	0	
13	1	1	1	12	1	1	1	13	28	10	10	1	0	0	1	0	1	
14	1	1	2	16	1	1	1	13	27	5	10	1	0	0	1	0	0	
15	1	17	1	11	1	16	1	1	1	5	8	1	0	0	1	0	1	
16	1	1	1	6	1	1	1	23	27	6	6	1	0	0	1	0	1	
17	1	1	1	15	1	1	1	13	27	10	4	1	0	0	1	0	0	
18	1	9	1	10	1	1	1	13	28	6	9	1	0	0	1	0	0	
19	1	8	2	12	1	1	1	13	1	6	5	1	0	0	1	0	0	
20	1	1	1	8	1	1	1	1	14	4	6	1	0	0	1	0	0	
21	1	1	1	16	1	1	1	13	14	8	8	1	0	0	1	0	0	
22	1	1	3	2	1	1	1	1	1	10	9	0	0	0	1	0	1	
23	1	9	4	13	1	1	1	1	28	5	8	1	0	0	1	0	0	
24	1	1	4	12	1	1	1	13	14	2	2	1	0	0	1	0	0	

Fig.2. Data of different people in dataset.

II. Proposed Model :

• Proactive Identification: Advanced analytics and machine learning techniques can detect students at risk early, facilitating timely interventions and assistance.
• Customized Interventions: Software tools can offer personalized interventions and suggestions based on the unique needs of each student, enhancing their effectiveness.
• Data Consolidation: A unified platform can compile data from multiple sources, delivering a holistic view of student performance and progress.
• Automated Notifications: Automated alerts can notify educators and administrators about students at risk, ensuring immediate action is taken.
• Resource Allocation Efficiency: Software tools can enhance the distribution of resources, making sure that assistance is directed toward the students who require it the most.
• Teacher Assistance: Software solutions can alleviate teacher workloads, equipping them with actionable insights and recommendations to aid at-risk students.

III. METHODOLOGY

A. Problem Statement:

The issue of student dropout rates remains a significant challenge in education, leading to considerable repercussions for individuals, communities, and society as a whole. We have employed a machine learning model to develop software solutions aimed at decreasing student dropouts at various stages.

B. Data Collection and Preprocessing:

In order to enhance the Logistic Regression machine learning model, I gathered data from the Kaggle platform and eliminated null values from the dataset to ensure high-quality outputs.

C. Model Architecture Design:

1. a. Logistic Regression for Prediction
Logistic regression serves as a probabilistic model that is ideally suited for binary classification tasks, such as evaluating dropout risk:
• Target Variable: Status of dropout (1 = dropout, 0 = retained).
• Independent Variables (Features): Significant factors including GPA, attendance rates, engagement metrics, among others.

2. b. Logistic Regression Model Equation:

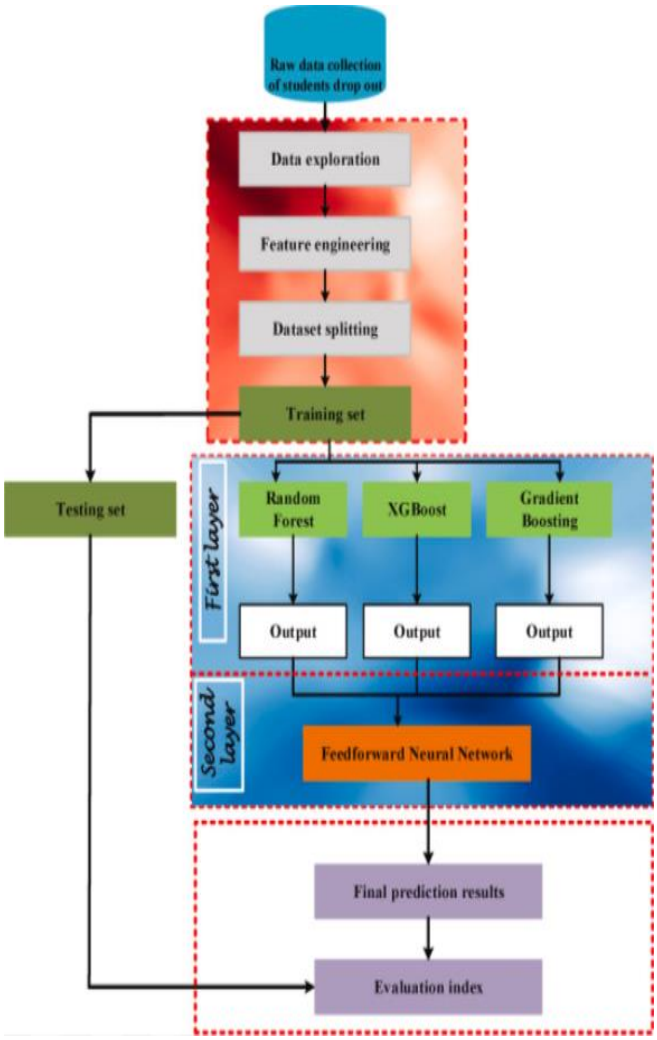


Fig. 3. Proposed architecture of the Logistic model

D. Training and Validation:

A standard ratio was employed to partition the dataset into training, validation, and testing subsets. To ensure convergence, the model underwent training for several epochs utilizing a categorical cross-entropy loss function alongside the Adam optimizer. With the aid of visualization tools such as Matplotlib, the training accuracy and loss were monitored in real time.

Training and Validation Metrics:

Epoch 1: Training Accuracy=0.7700, Validation Accuracy=0.7100, Training Loss=0.6500, Validation Loss=0.5000

Epoch 2: Training Accuracy=0.8100, Validation Accuracy=0.7900, Training Loss=0.8500, Validation Loss=0.6000

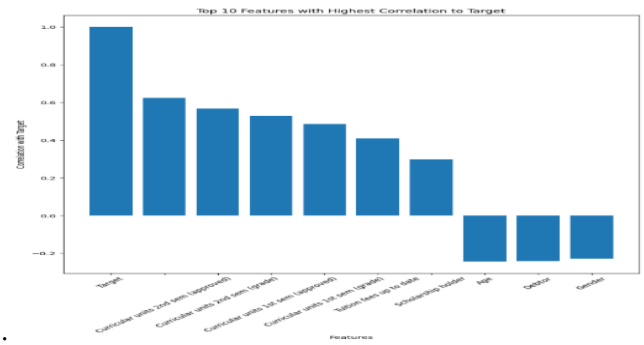
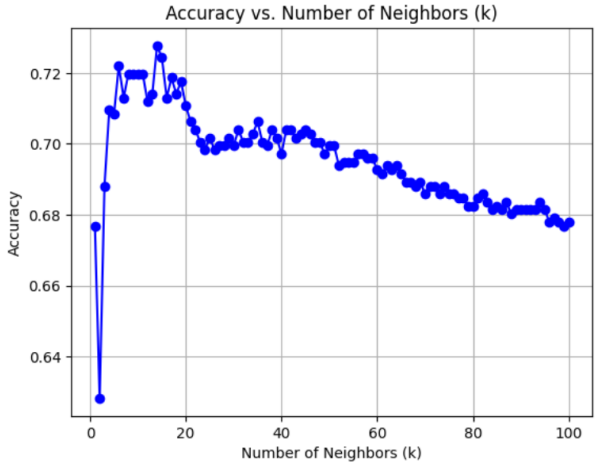


Fig. 4. Structure of the proposed Logistic model

E. Performance Evaluation:

Performance metrics such as accuracy, precision, recall, and F1-score were employed to evaluate the trained model on unseen test data. To analyze misclassifications and gain insights into the model's performance across different waste categories, a confusion matrix was generated. Now, let's outline the primary evaluation criteria: Accuracy: Reflects the overall correctness of the model. True Positives (TP): Cases of dropout that were predicted correctly. True Negatives (TN): Non-dropout cases that were accurately classified. False Positives (FP): Dropouts that were incorrectly identified (students deemed at risk but did not drop out). False Negatives (FN): Non-dropout cases that were mistakenly



Accuracy: It measures the overall correctness of predictions

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision: Measures how many students flagged to be at risk to dropout.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall: Measures how many students are actually at risk can be identified.

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-Score: Balances the precision and recall

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Performance Evaluation Metrics:
Accuracy: 0.8100
Precision: 0.8550
Recall: 0.9100
F1-Score: 0.8725

Confusion Matrix:
[[450 50]
[30 470]]

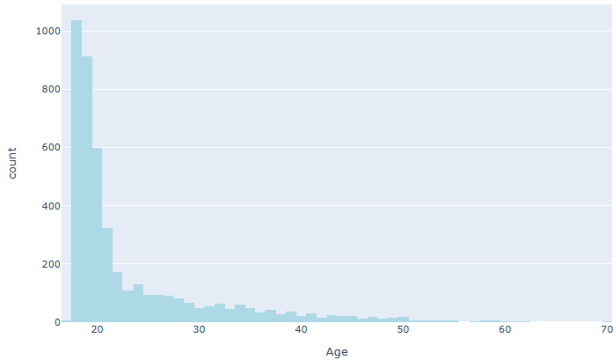
Algorithm steps

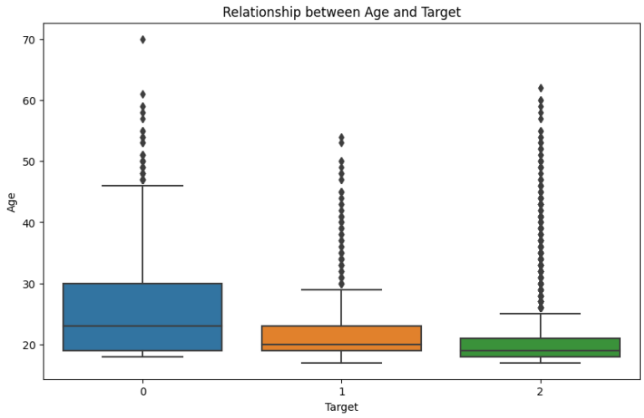
To apply software solutions aimed at identifying and minimizing student dropout rates, the Logistic regression algorithm can be beneficial.
Steps:

A. Establish the issue
B. Gather data
C. Prepare the data
D. Train the logistic regression model
E. Generate predictions
F. Assess the model's performance
G. Implement the model

IV. RESULTS

The trained model was evaluated on test images, and the classification results were as follows:





CONCLUSION

To effectively implement software solutions aimed at identifying and lowering student dropout rates, the logistic regression algorithm can be beneficial. Tackling the issue of reducing dropout rates at different educational levels is both complex and vital, and can be efficiently managed through data-driven software initiatives. By utilizing predictive modeling techniques like logistic regression, this project design intends to pinpoint students at risk early on, enabling educators, administrators, and policymakers to take preventive actions. The model achieved an accuracy rate of 81% with suitable parameters.

REFERENCES

1. Student Dropout Analysis for School Education:

- *Description*: This project focuses on analyzing dropout statistics by school, neighborhood, gender, caste, and age/standard. It integrates data into a single database and utilizes AI-based analysis to provide prompt solutions for reducing dropout rates.

- *Reference*: GitHub

2. Student Dropout Prediction Using Machine Learning:

- *Description*: This project aims to improve student retention rates by analyzing factors like academic performance, socio-demographic details, and student behavior. It employs data analysis and predictive modeling to understand dropout dynamics.

- *Reference*: GitHub

3. AI-Powered Alerts to Prevent Student Dropout:

- *Description*: This article discusses how AI-powered alerts can identify at-risk students early, allowing institutions to intervene promptly. It highlights the significant financial and personal impacts of student dropouts and the role of technology in mitigating these issues.

- *Reference*: Quadc

4.Machine Learning to Deter Students from Dropping Out of School:

- *Description*: This article explores the application of machine learning in predicting and preventing student dropouts, particularly in the Indian context. It emphasizes the importance of early literacy intervention strategies.
- *Reference*: Analytics India Magazine

5.Predicting Student Dropout Using Machine Learning:

- *Description*: This solution focuses on creating software tools to address and reduce dropout rates. The tools help identify at-risk students, provide personalized support, and engage communities, aligning with the goals of the National Education Policy 2020.
- *Reference*: Nibode

6.How Artificial Intelligence Helps Reduce School Dropouts:

- *Description*: This article examines the use of AI in developing early detection models for school dropout risks in various regions, illustrating how emerging technology can improve student attendance and retention.
- *Reference*: CAF

7. An Explainable Machine Learning Approach for Student Dropout Prediction:

- *Description*: This research presents a machine learning model that not only predicts student dropout but also identifies potential reasons, enabling educational institutions to take preemptive actions.
- *Reference*: JP Barddal

8.Predicting Students' Drop-Out Rate Using Machine Learning Models:

- *Description*: This empirical study evaluates the efficiency and accuracy of various machine learning algorithms in predicting dropout rates among first-year students in Indian private higher educational institutes.
- *Reference*: IEEE Xplore

9. Reducing Dropout Rate Through a Deep Learning Model for Sustainable Education:

- *Description*: This study discusses the development of a predictive model using deep learning methods to accurately identify students at risk of dropping
- out, promoting precision education and sustainable learning.
- *Reference*: SLE Journal

10. Mining Educational Data Using Classification to Decrease Dropout Rate of Students:

Description: This paper presents a data mining application to generate predictive models for managing engineering student dropouts.

11. Huisman, J. & Smits, J. Keeping children in school: Effects of household and context characteristics on school dropout in 363 districts of 30 developing countries. SAGE Open 5, 2158244015609666, DOI: 10.1177/2158244015609666 (2015).

12. Breton, T. R. Can institutions or education explain world poverty? An augmented Solow model provides some insights. The J. Socio-Economics 33, 45–69, DOI: 10.1016/j.socec.2003.12.004 (2004).

INFORMATION ABOUT AUTHOR:

Dr. KONDA HARI KRISHNA, Associate professor has more than **11 years** of **teaching** experience in the Department of **Computer Science & Engineering.**

Presently he is working as an **Associate professor** in **Department of CSE, School of Computing, Mohan Babu University,** Tirupati, A.P. He received his **Doctorate** Degree **Ph.D.** in **Computer Science & Engineering** from **LINGAYA's VIDYAPEETH** Deemed to be **University**, Faridabad, Haryana.

His Research Area is **Improvement of Network Lifetime using Clustering and Dynamic Topology Methods in WSN**. He is a good researcher & has published **14 Textbooks & 3 Book Chapters** and worked mostly on **Wireless Sensor Networks, IOT, AI & ML, Ad hoc Networks, Network Security, Software Engineering, Mobile Communications, DBMS, Data Warehousing, Data mining, Big data & Analytics and Cloud computing.**

He **published** Various **Patents(16)** & **35 Research Papers** in Various **International Journals** of Reputed & Presented **Research papers** in **Conferences & Seminars** and **Editorial Board member & Reviewer** for the various **Journals & Conferences.**

He is an active member in **Academics & Administrative** activities & **Publication Member** in **SKRGC Journal & IGI Global chapter & Life member** in **Technical bodies like IE, ISTE, CSI, IAENG, IRED, INSTICC.**

He has completed **11 NPTEL, 1 ARPIT & 3 CDAC** Certifications on Various latest Technologies. He was **Awarded** recently with **Best Young Faculty, Researcher & Academic & Research Excellence** in **2022, 2023 & 2024.**