DropGuard: Student Retention Analysis

The proposed system begins with the design and deployment of

[1]Dr.Ajay Paithane , [2]Pratik Shinde , [3]Aaryan Rankhamb , [4]Ojas Thakare
[1234]D. Y. Patil Institute of Engineering Management and Research
[1]ajay.paithane@dypiemr.ac.in[2]pratikshinde0501@gmail.com[3]pyranpy@gmail.com[4]thakareojas24072003@gamil.com

***Abstract*—Student dropout rates are concerning problem as it affecting educational institutions globally, leading to both academic and financial challenges. This proposed system introduces an AI-driven software that leverages machine learning calculations to anticipate understudy dropout probabilities early in the academic journey. By analyzing an extensive range of student data, including demographic, academic performance, attendance, and socioeconomic factors, the system identifies patterns and risk factors associated with dropout behavior. The proposed model integrates supervised learning techniques, allowing educators to take proactive interventions based on the predictive insights. The system continuously learns and improves by adapting to new data, thus enhancing its accuracy over time. Through early assessment of students at risk , the software aims to significantly cut down on dropout rates, improve academic outcomes, and foster long-term student success. This innovative solution provides educational institutions with a robust tool to make data-driven decisions, optimizing student retention strategies**

Keywords— **Educational management plat-form; Learning achievements; Machine learning applications in teaching; Predictive modeling**

I. INTRODUCTION

THIS study tackles the pressing issue of student dropout in higher education, A challenge that has profound implications not just for the students themselves as well as for academic institutions and the larger societal context.. High dropout rates contribute to reduced career opportunities for students, diminished institutional reputation, and a less skilled workforce. To address this, the proposed system investigates the usage of machine learning (ML) and deep learnings (DL) strategies to predict at-risk students early, thereby enabling timely interventions that can improve student retention and success.

Unlike traditional approaches that focus primarily on academic performance as a predictor of dropout, this study expands the scope to include a variety of psychosocial and non-academic factors. These factors include mental health status, financial challenges, communication barriers with faculty, personal confidence, and engagement with course material—elements that are often overlooked in conventional retention studies. By understanding the broader array of factors influencing student persistence, the study plans to offer a more expanscomprehensive model for predicting and for addressing dropout risk.

a comprehensive survey aimed at collecting detailed data on these factors. Students from various departments will participate in the survey, ensuring the dataset reflects a broad spectrum of experiences. After gathering the data, it will undergo cleaning and preprocessing to convert responses into a well-organized structure conducive to analysis. This procedure ensures that the data is both manageable and of high quality, making it ready to use in machine learning applications.

The core objective of the system involves developing a predictive model that uses the cleaned data to detect students who are at risk of leaving school. By applying a combination of supervised machine learning methods (including classification algorithms) and deep learning models, the system aims to build a robust model capable of making accurate predictions. Exploratory data analysis (EDA) will be performed to identify trends and associations within the dataset, ensuring it is reliable and complete before model training begins.

The predictive model will be integrated into the institution's student portal, providing both students and faculty with early warnings about dropout risk. This proactive system will allow for timely interventions, such as personalized counseling, academic support, or financial assistance, which can help at-risk students overcome challenges and stay on track. The integration of this tool into the student portal will facilitate easy access to dropout predictions, making it convenient for students and faculty to monitor risk levels and act accordingly.

A critical part of the study is the assessment of the model's effectiveness. Crucial evaluation metrics such as precision, recall, F1-score, and accuracy will be used to determine how well the model identifies at-risk students. This evaluation ensures the reliability of the model and its effectiveness in predicting dropout risk. The ultimate goal is to create a predictive tool that is both accurate and actionable, providing meaningful insights that can be used for personalized student interventions.

The study's objectives are multifaceted. They include identifying the key factors contributing to student dropout, developing a robust and reliable predictive model, ensuring data quality, and integrating the model into a user-friendly interface. The study will also generate actionable recommendations for institutions to implement targeted interventions, such as improved academic advising, mental health support, or financial aid programs. Additionally, the study aims to contribute to the academic literature by exploring factors that

Doi:10.48047/ijiee.2025.15.5.64

are often overlooked in traditional dropout research, providing a new perspective on student retention.

By achieving these objectives, the study aims to offer a data-driven solution to improve student retention in higher education. The proposed system also seeks to inform institutional policies, helping schools allocate resources more effectively and design interventions that address the underlying causes of dropout. In the broader context, the research adds to the growing domain of educational data analytics, delivering insights that can be applied to other institutions and regions.

Ultimately, this proposed system aims not just to minimize student attrition, but also to foster a more supportive and inclusive educational environment. By identifying students facing challenges early and offering them the necessary support, this framework has the potential to improve educational outcomes, enhance institutional effectiveness, and develop a more skilled and capable workforce that can contribute positively to society.

This study addresses the critical issue of student dropout in higher education by proposing an AI-driven system that integrates machine learning and deep learning. By analyzing diverse academic and non-academic factors, the study aims to predict at-risk students early, enabling timely interventions to enhance student retention and foster a supportive educational environment.

II. Literature Review

Related Work

A study conducted by Abdelhamid Tayebi et al. [1] explores the economic impact of students dropping out. They highlight several factors influencing the motivation of students, particularly in the electrical and computing engineering disciplines in Spain. They surveyed 624 students from 8 different Spanish universities to identify the major factors contributing to students' motivation, which leads them to halt their pursuit of an engineering education. According to their findings, 23 out of the 40 factors they analyzed were found to correlate with low motivation, and about 46% of the surveyed students admitted to having considered dropping out. The main issues identified in their research included poor relationships with professors, followed by underperformance academically. Most of the data was gathered via questionnaires, with very few personal interviews conducted.

Another study conducted by Qabil et al. [2], at the University of Prishtina in Kosovo, Serbia, delves deeply into student dropout in the Electrical and Computer Engineering (ECE) department. This research spanned 14 years, from 2001 to 2015, and discusses its impact on the industry as well as potential solutions. The problem is described as complex and multifaceted. The study surveyed 6,442 students in their first year, employing questionnaires and interviews. Among them, 1,980 students graduated, while 4,408 dropped out, indicating a significant dropout rate. They categorized the students into four groups: The first group consisted of students who enrolled in the program but did not pass any exams. The second category included those who enrolled but withdrew before completing their first year. The third group covered students who passed all first-year exams but left the program before advancing to the second year. The final category included students who completed their second year but left the faculty without passing the exams for their third year.

They attribute the reasons to the war-torn history of Serbia and poor foundation in mathematics.

Another study by Smith et al. [3] at the University of Melbourne examined dropout rates among engineering students in the mechanical and civil departments. Conducted between 2008 and 2016, the research analyzed the progress of 4,567 students, revealing that 1,980 students completed the program, while 2,587 dropped out at various stages. The researchers classified students into three categories: those who left without completing any courses, those who completed their first year but dropped out after that, and those who persisted until the third year but eventually left without graduating.

Smith et al. highlighted academic pressure, lack of practical training, and financial constraints as the primary factors contributing to student attrition. They recommended enhanced mentorship programs and industry-linked internships as effective methods to improve retention. The study suggested that by aligning academic coursework more closely with industry expectations, universities could reduce dropout rates significantly.

A study conducted by Johnson et al. [4] at the University of California explored student dropout patterns in the electrical engineering department between 2010 and 2020. This study tracked 5,200 students and found that 2,900 did not complete their degrees. The researchers divided the students into three categories: those who withdrew in the first year without passing any courses, those who left after completing the first year but before the second, and those who dropped out after reaching the third year but failing to complete their degree.

Johnson et al. identified several critical factors contributing to the high dropout rate, including heavy academic workloads, insufficient support systems, and the mismatch between course content and student expectations. They suggested implementing adaptive learning technologies and early intervention strategies to help students vulnerable to dropping out. Their research emphasized the importance of offering more hands-on proposed systems to keep students engaged and better prepared for industry challenges.

A thorough review of related studies reveals the multifaceted and interconnected nature of student dropout, influenced by academic performance, financial limitations, mentorship deficiencies, and broader societal challenges. Prior research highlights key predictors of attrition but often overlooks the potential of emerging technologies. Building upon these insights, this study integrates advanced ML and DL techniques to develop a predictive framework capable of identifying dropout risks with greater accuracy. By addressing the gaps in traditional approaches, it provides educational institutions with actionable insights to implement proactive, personalized interventions that significantly reduce dropout rates

Doi:10.48047/ijiee.2025.15.5.64

III. Proposed Methodology

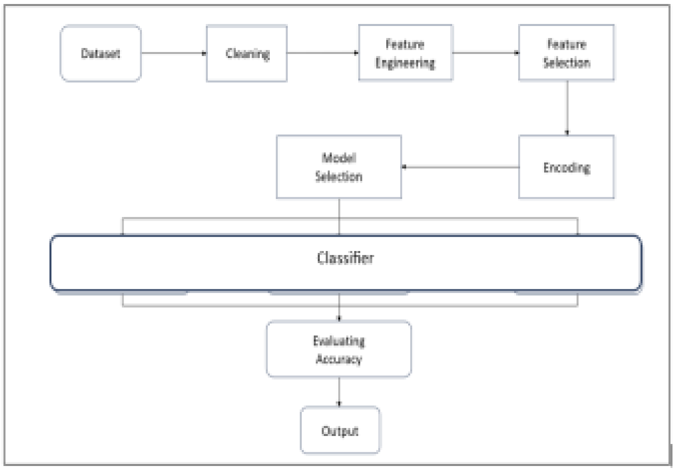


Fig. 1 : Diagram of Proposed Methodology

A. Dataset

Before process the dataset let's describe it briefly:
This dataset focuses on the academic performance of secondary school students from two Portuguese schools. It contains 395 observations (students) and 31 variables, including a variety of demographic, social, and school-related attributes. The data is complete, meaning there are no missing values, and therefore, no additional data cleaning is required.The last column indicates whether a student passed the final exam or not.

B. Detailed Description of the Column

Table 1 : Attributes used in Dataset

Feature Name	Description	Possible Values
school	School attended by the student	GP: Gabriel Pereira, MS: Mousinho da Silveira
sex	Gender of the student	F: Female, M: Male
age	Age of the student	15 to 22
address	Residential address type	U: Urban, R: Rural
famsize	Family size	LE3: ≤3 members, GT3: >3 members
Pstatus	Parental cohabitation status	T: Together, A: Apart
Medu	Mother's education level	0: None, 4: Higher education
Fedu	Father's education level	0: None, 4: Higher education
Mjob	Mother's occupation	Teacher, Health, Services, At_home, Other
Fjob	Father's occupation	Teacher, Health, Services, At_home, Other
reason	Reason for school selection	Home, Reputation, Course, Other
guardian	Primary guardian	Mother, Father, Other
traveltime	Travel time to school	1: <15 mins, 4: >1 hour
studytime	Weekly study hours	1: <2 hours, 4: >10 hours
failures	Number of past failures	0 to 4
schoolsup	Extra educational support	Yes, No
famsup	Family educational support	Yes, No
paid	Extra paid classes	Yes, No
activities	Extracurricular activities	Yes, No
nursery	Attended nursery school	Yes, No
higher	Aspiration for higher education	Yes, No
internet	Internet access at home	Yes, No
romantic	Romantic relationship	Yes, No
famrel	Family relationship quality	1: Very bad, 5: Excellent
freetime	Free time after school	1: Very low, 5: Very high
goout	Frequency of going out with friends	1: Very low, 5: Very high
Dalc	Alcohol consumption on weekdays	1: Very low, 5: Very high
Walc	Alcohol consumption on weekends	1: Very low, 5: Very high
health	Current health status	1: Very bad, 5: Very good
absences	School absences	0 to 93

C. Data Analysis

We conducted a preliminary exploratory data analysis (EDA) using Python 3, leveraging **Pandas** (version 2.0.2), **Scikit-learn** (version 1.2.2), and **Seaborn** (version 0.12.2) for visualization tasks.
Tables 3.2.1 through 3.2.4 present key summary statistics for all attributes in the dataset. These tables include visualizations such as histograms to display the distribution of attribute values. Additionally, the tables show the central tendency of each attribute (mode for categorical variables and mean for numerical variables), the median of each attribute, measures of dispersion, as well as the minimum and maximum values for the numerical attributes.

Table 2 :Statistics info about Demographic Data

Doi:10.48047/ijiee.2025.15.5.64

Attribute	Distrib.	Mean	Median	Dispersion	Min.	Max.
Marital status		1.180	1	0.510	1	6
Nationality		1.250	1	1.390	1	21
Displaced		0.548	1	0.907	0	1
Gender		0.352	0	1.358	0	1
Age at enrollment		23.130	20	0.320	17	70
International		0.025	0	6.262	0	1

Table 3 :Statistics info about Socioeconomic Data

Attribute	Distrib.	Mean	Median	Dispersion	Min.	Max.
Father's qualification		16.460	14	0.670	1	34
Mother's qualification		12.320	13	0.730	1	29
Father's occupation		7.820	8	0.620	1	46
Mother's occupation		7.320	6	0.550	1	32
Educational special needs		0.012	0	9.260	0	1
Debtor		0.114	0	2.792	0	1
Tuition fees up to date		0.881	1	0.368	0	1
Scholarship holder		0.248	0	1.739	0	1

Table 4 :Statistics info about Marcoeconomic Data

Attribute	Distrib.	Mean	Median	Dispersion	Min.	Max.
Unemployment rate		11.566	11.100	0.230	7.600	16.200
Inflation rate		1.228	1.400	1.126	−0.800	3.700
GDP		0.002	0.320	1152.820	−4.100	3.500

Table 5 :Statistics info about Academic Data at Enrollment

Attribute	Distrib.	Mean	Median	Dispersion	Min.	Max.
Application mode		6.890	8	0.770	1	18
Application order		1.730	1	0.760	1	9
Course		9.900	10	0.440	1	17
Daytime/evening attendance		0.891	1	0.350	0	1
Previous qualification		2.530	1	1.570	1	17

The heatmap analysis, based on the Pearson correlation coefficient, reveals strong correlations among several feature pairs, indicating multicollinearity in the dataset. This occurs when features are highly interrelated, such as demographic attributes (e.g., age and parental education) or academic metrics (e.g., attendance and grades), leading to redundancy. Additionally, correlations between distinct feature groups, like socioeconomic factors and academic performance, highlight complex interdependencies. Multicollinearity can affect model stability and predictive accuracy, necessitating techniques like feature selection or dimensionality reduction to address redundancy and enhance model performance.

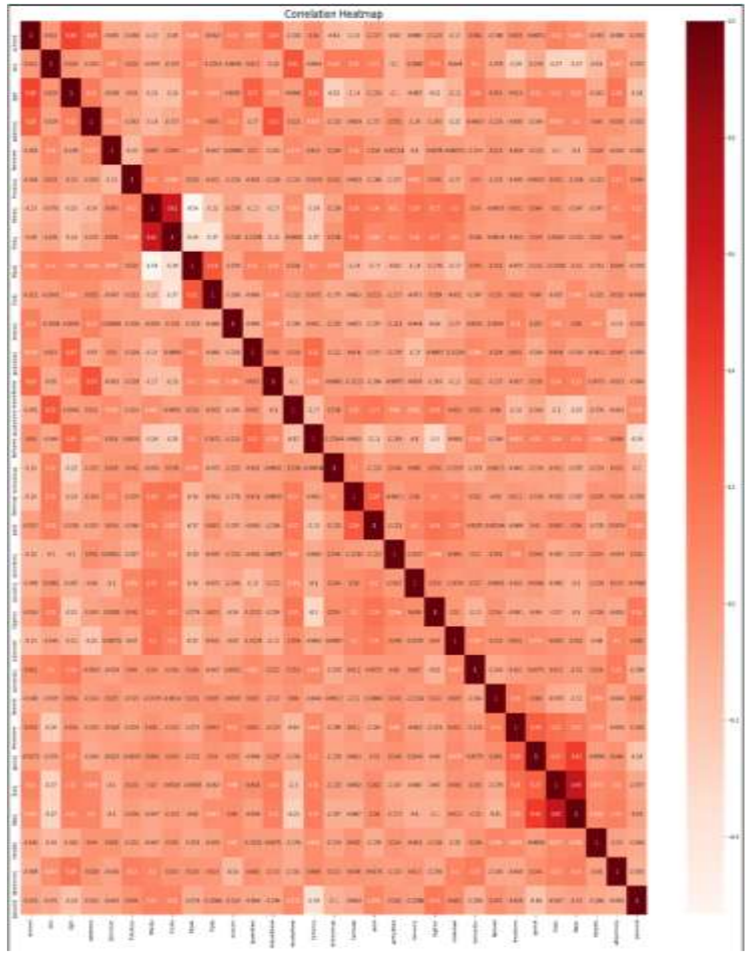


Fig. 2 : Correlation Heatmap of Data

D. Data Preprocessing

Before training the machine learning models, several preprocessing steps were applied to the dataset:

1) Handling Missing Values:

Missing values in the dataset were initially checked and addressed using appropriate techniques, such as mean imputation, mode imputation, or removal of instances, depending on the type and importance of the missing data. However, in this particular dataset, there were no missing values, so no specific handling for missing data was required

2) Feature Selection:

To enhance model performance and reduce computational complexity, feature selection techniques like correlation analysis, information gain, or stepwise regression were used to select the most relevant attributesfor training the models. In this particular dataset all the attributes either have significant corelation with the target or among each other as in section 2.2 data analysis so, all the features are taken for the model training. The target column contain three classes Graduate, Dropout and enrolled. As per our assertion we are predicting whether a student will dropout or not so, the number of "Enrolled" student

Doi:10.48047/ijiee.2025.15.5.64

is irrelevant because it won't give us any new information as all the Graduate and Dropout are also enrolled. We only need to know whether a student graduated or dropedout. So, we are dropping the "Enrolled" values and going forward with "Graduate" & "Dropout" values.

3) Encoding:

Categorical variables are encoded into numerical representations to enable machine learning algorithms to process them effectively. Techniques like one-hot encoding or label encoding are employed. The Target column which contain Text values are encoded using the label encoder in the scikit learn library which randomly assign numerical values to all the unique values.

All the other attributes are already encoded in the data set as per the following table:

Table 6 : Marital Status Values

Attribute	Values
Marital status	1—Single
	2—Married
	3—Widower
	4—Divorced
	5—Facto union
	6—Legally separated

Table 7 : Nationality Values

Attribute	Values
Nationality	1—Portuguese
	2—German
	3—Spanish
	4—Italian
	5—Dutch
	6—English
	7—Lithuanian
	8—Angolan
	9—Cape Verdean
	10—Guinean
	11—Mozambican
	12—Santomean
	13—Turkish
	14—Brazilian
	15—Romanian
	16—Moldova (Republic of)
	17—Mexican
	18—Ukrainian
	19—Russian
	20—Cuban
	21—Colombian

Table 8 : Application Mode Values

Attribute	Values
Application mode	1—1st phase—general contingent
	2—Ordinance No. 612/93
	3—1st phase—special contingent (Azores Island)
	4—Holders of other higher courses
	5—Ordinance No. 854-B/99
	6—International student (bachelor)
	7—1st phase—special contingent (Madeira Island)
	8—2nd phase—general contingent
	9—3rd phase—general contingent
	10—Ordinance No. 533-A/99, item b2) (Different Plan)
	11—Ordinance No. 533-A/99, item b3 (Other Institution)
	12—Over 23 years old
	13—Transfer
	14—Change in course
	15—Technological specialization diploma holders
	16—Change in institution/course
	17—Short cycle diploma holders
	18—Change in institution/course (International)

Table 9 : Course Values

Attribute	Values
Course	1—Biofuel Production Technologies
	2—Animation and Multimedia Design
	3—Social Service (evening attendance)
	4—Agronomy
	5—Communication Design
	6—Veterinary Nursing
	7—Informatics Engineering
	8—Equiniculture
	9—Management
	10—Social Service
	11—Tourism
	12—Nursing
	13—Oral Hygiene
	14—Advertising and Marketing Management
	15—Journalism and Communication
	16—Basic Education
	17—Management (evening attendance)

Table 10 : Previous Qualification Values

Attribute	Values
Previous qualification	1—Secondary education
	2—Higher education—bachelor's degree
	3—Higher education—degree
	4—Higher education—master's degree
	5—Higher education—doctorate
	6—Frequency of higher education
	7—12th year of schooling—not completed
	8—11th year of schooling—not completed
	9—Other—11th year of schooling
	10—10th year of schooling
	11—10th year of schooling—not completed
	12—Basic education 3rd cycle (9th/10th/11th year) or equivalent
	13—Basic education 2nd cycle (6th/7th/8th year) or equivalent
	14—Technological specialization course
	15—Higher education—degree (1st cycle)
	16—Professional higher technical course
	17—Higher education—master's degree (2nd cycle)

Table 11 :Gender Values

Attribute	Values
Gender	1—male
	0—female

Table 12 :Attendence Regime Values

Attribute	Values
Daytime/evening attendance	1—daytime
	0—evening

Table 13 : Yes/No Attributes

Attribute	Values
Displaced	
Educational special needs	
Debtor	1—yes
Tuition fees up to date	0—no
Scholarship holder	
International	

Table 14 : Mother's and Father's Values

Doi:10.48047/ijiee.2025.15.5.64

Attribute	Values
Mother's qualification Father's qualification	1—Secondary Education—12th Year of Schooling or Equivalent 2—Higher Education—bachelor's degree 3—Higher Education—degree 4—Higher Education—master's degree 5—Higher Education—doctorate 6—Frequency of Higher Education 7—12th Year of Schooling—not completed 8—11th Year of Schooling—not completed 9—7th Year (Old) 10—Other—11th Year of Schooling 11—2nd year complementary high school course 12—10th Year of Schooling 13—General commerce course 14—Basic Education 3rd Cycle (9th/10th/11th Year) or Equivalent 15—Complementary High School Course 16—Technical-professional course 17—Complementary High School Course—not concluded 18—7th year of schooling 19—2nd cycle of the general high school course 20—9th Year of Schooling—not completed 21—8th year of schooling 22—General Course of Administration and Commerce 23—Supplementary Accounting and Administration 24—Unknown 25—Cannot read or write 26—Can read without having a 4th year of schooling 27—Basic education 1st cycle (4th/5th year) or equivalent 28—Basic Education 2nd Cycle (6th/7th/8th Year) or equivalent 29—Technological specialization course 30—Higher education—degree (1st cycle) 31—Specialized higher studies course 32—Professional higher technical course 33—Higher Education—master's degree (2nd cycle) 34—Higher Education—doctorate (3rd cycle)

4) Feature Scaling:

To ensure that all features are on a comparable scale, feature scaling techniques like normalization or standardization are used. In this dataset, we apply the StandardScaler from the Sklearn library to standardize the features. This method transforms the features by subtracting the mean and scaling them to have unit variance. The standard score of a sample x is calculated as:

$$z = x - \mu / \sigma \quad (1)$$

where μ is the mean of the training data and σ is the standard deviation of the training data.

5) Data Splitting:

The processed dataset is split into training and testing subsets, with 80% of the data used for model training and the remaining 20% for performance evaluation. This division is performed using the train_test_split function from the Sklearn library.
Data preprocessing was a pivotal step in preparing the dataset for machine learning applications. Techniques such as feature selection, encoding, and scaling enhanced the dataset's quality and ensured consistency for model training. By addressing issues like class imbalances and standardizing features, the study maintained the integrity of the data while optimizing it for high-performance predictive modeling. This meticulous approach guarantees that the subsequent models operate on a clean, well-structured foundation, maximizing their predictive accuracy and relevance.

E. Model Training and Evaluation

Three Machine Learning models, such as logistic regression, KNN Classifier, supportvector machines (SVM) are trained using the training dataset. Each model is trained with the goal of accurately predicting dropout or academic success.

The performance of the trained models is assessed using various evaluation metrics, including accuracy, precision, recall, and F1-score. These metrics help evaluate the models' effectiveness and provide a comparison of their ability to predict student outcomes.

1) Logistic Regression

Logistic regression is a commonly used algorithm for binary classification tasks. It models the relationship between a set of independent variables (features) and a binary dependent variable (target) by applying the logistic function. Given that this is a binary classification problem, logistic regression is particularly well-suited for predicting the two possible outcomes of the target variable.
The Evaluation of model is as below:

```
****************************************************************
      Classification Report of Logistic Regression
****************************************************************
```

	precision	recall	f1-score	support
0	0.75	0.51	0.61	35
1	0.82	0.93	0.87	84
accuracy			0.81	119
macro avg	0.79	0.72	0.74	119
weighted avg	0.80	0.81	0.79	119

```
****************************************************************
```

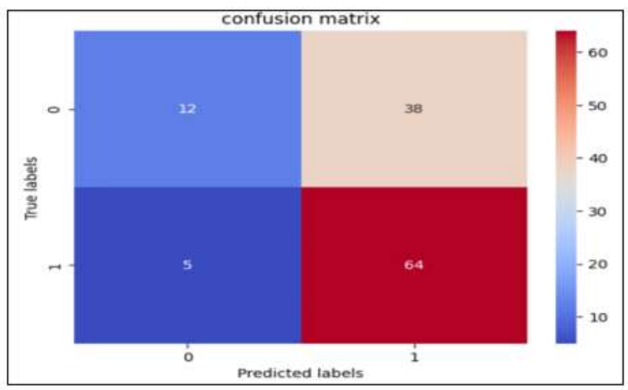


Fig. 3 : confusion Matrix of Logistic Regression with report

2) Support vector Classifier

The Support Vector Classifier (SVC), also referred to as Support Vector Machine (SVM), is a widely used machine learning algorithm for both binary and multi-class classification tasks. It is particularly effective in handling complex decision boundaries and high-dimensional feature spaces. In this case, we apply the SVC with a linear kernel to classify the data, aiming to create a hyperplane that best separates the different classes.
The Evaluation of model is as below:

Doi:10.48047/ijiee.2025.15.5.64

Classification Report of Support vector Classifier

	precision	recall	f1-score	support
0	0.88	0.65	0.75	23
1	0.83	0.95	0.89	41
accuracy			0.84	64
macro avg	0.86	0.80	0.82	64
weighted avg	0.85	0.84	0.82	64

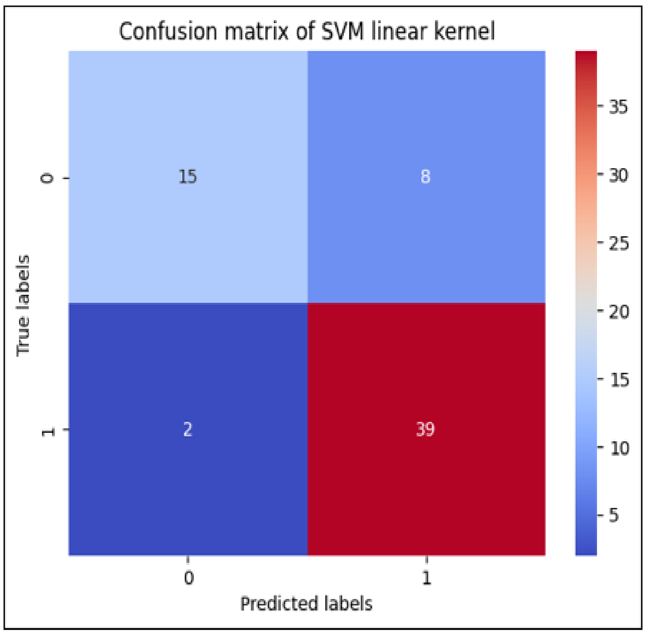


Fig. 4 : confusion Matrix of SVM with report

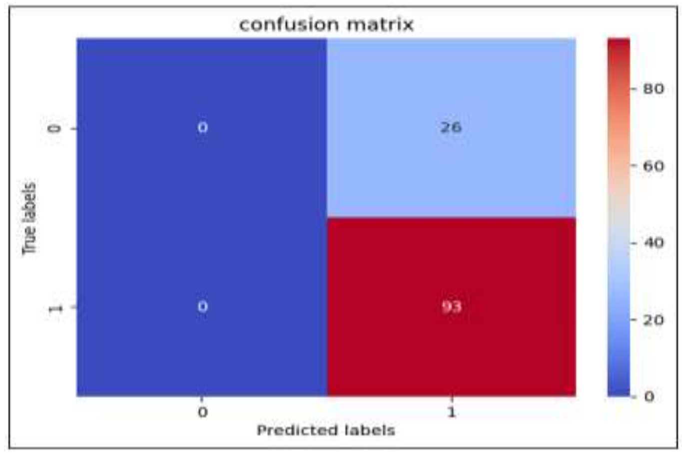
3) KNN Classifier

The K-Nearest Neighbors (KNN) classifier is a flexible machine learning algorithm suitable for both classification and regression problems. As a non-parametric method, it makes predictions based on the similarity between the input instance and its closest data points. To determine the optimal value of (K), we use the elbow method, which helps identify the most effective number of neighbors for the model. In this case, the elbow method suggests that (K = 3) yields the best performance, so the model was trained using this hyperparameter.

The Evaluation of model is as below:

Classification Report of KNN Classifier

	precision	recall	f1-score	support
0	0.00	0.00	0.00	26
1	0.78	1.00	0.88	93
accuracy			0.78	119
macro avg	0.39	0.50	0.44	119
weighted avg	0.61	0.78	0.69	119



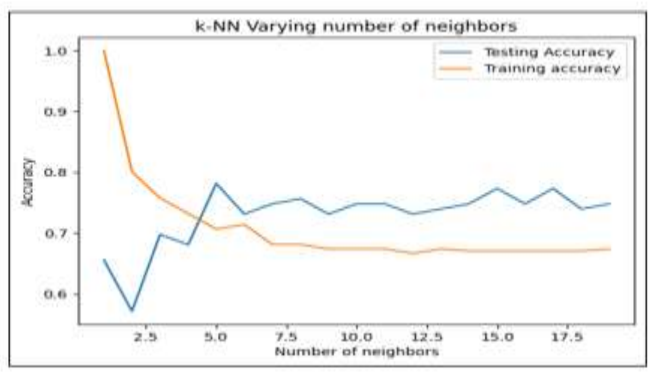


Fig. 4 : confusion Matrix of KNN Classifier with report

The study employed three machine learning models—Logistic Regression, Support Vector Classifier (SVC), and K-Nearest Neighbors (KNN)—to predict student outcomes. Each model was rigorously trained and evaluated using metrics such as accuracy, precision, recall, and F1-score. Logistic Regression demonstrated balanced performance across metrics, while SVC excelled in recall, effectively identifying students at risk of dropping out. Although KNN showed limitations in handling imbalanced datasets, it provided useful insights into proximity-based classifications. The comparative evaluation highlights the potential and challenges of each model, offering a roadmap for selecting the most suitable approach based on institutional priorities.

Doi:10.48047/ijiee.2025.15.5.64

IV. Result and Discussion

During the evaluation phase, the performance of the trained machine learning models in predicting student dropout and academic success is analyzed. The evaluation is based on metrics such as accuracy, precision, recall, and F1 score, which offer insights into how well the models perform on classification tasks.

The evaluation results provide a clear picture of how each model performs on the dataset. By comparing these metrics, we can identify the most effective model for the specific task. For instance, the logistic regression model may show higher accuracy but lower recall, indicating that it is good at predicting non-dropout students but less effective at identifying actual dropouts. In contrast, the decision tree model might exhibit higher recall but lower precision, suggesting it successfully detects more dropout cases but also generates more false positives.

Additionally, the real-world implications of the models' performance are considered, with an emphasis on how accurate predictions of dropout and academic success can lead to timely interventions and support for students at risk. The limitations of each model are acknowledged, and potential areas for improvement are explored, setting the stage for future advancements in predictive modeling for educational outcomes. The results of the predictive modeling underline the transformative potential of machine learning in addressing student dropout. By accurately identifying at-risk students, the models enable institutions to implement timely and targeted interventions, such as counseling, academic support, and financial assistance. However, the findings also highlight areas for improvement, such as reducing false positives and refining feature selection. These results not only validate the efficacy of the proposed system but also emphasize the importance of combining technological innovation with institutional strategies to create a sustainable impact on student retention.

Conclusion and Future Work

In conclusion, this proposed system successfully develops and evaluates machine learning models for predicting student dropout and academic success using the provided dataset. The findings emphasize the potential of machine learning in the education sector. By accurately predicting dropout and academic success, institutions can take proactive steps to support at-risk students, ultimately improving student retention and outcomes.

Looking ahead, several opportunities exist to further enhance the accuracy and robustness of the predictive models. One key area for future work is expanding the dataset by including additional relevant features. This could involve capturing more detailed aspects of students' backgrounds and experiences, such as socio-economic status, involvement in extracurricular activities, or the level of family support. Incorporating these factors could offer a more nuanced understanding of the influences on academic success and dropout.

Additionally, exploring alternative machine learning algorithms or ensemble methods could lead to more accurate predictions. Techniques such as gradient boosting, deep learning, or recurrent neural networks might uncover complex patterns in the data that other methods could overlook. Testing and optimizing a variety of models would help identify the best approach for predicting student outcomes.

Another promising direction is the integration of real-time data into the models. Implementing a feedback loop system would allow continuous model updates as student dynamics change over time. This real-time adaptability would improve prediction accuracy and enable timely interventions to support students who may be at risk of dropping out.

By focusing on these improvements, the models could be made more accurate, practical, and adaptable, ultimately contributing to more effective strategies for preventing student dropout and fostering academic success in educational environments..

This research presents a pioneering system that combines the power of machine learning and deep learning to address the persistent issue of student dropout. By incorporating diverse academic, demographic, and psychosocial factors, the study provides a comprehensive and actionable framework for early intervention. The results demonstrate the feasibility and potential of predictive analytics in fostering a more inclusive and supportive academic environment.

Looking ahead, future work will focus on expanding the dataset to include richer features, such as real-time student engagement metrics, and exploring advanced algorithms like ensemble methods and neural networks for enhanced predictive accuracy. The integration of real-time feedback loops will ensure the system remains adaptable to evolving student dynamics. Ultimately, this research aspires to empower educational institutions with robust tools to make informed decisions, optimize resource allocation, and significantly improve student retention and success rates, contributing to a more skilled and resilient global workforce.

V. References

1] Tayebi, Abdelhamid & Gómez, Josefa & Delgado, Carlos. (2021). Analysis on the Lack of Motivation and Dropout in Engineering Students in Spain. IEEE Access. 9. 66253-66265. 10.1109/ACCESS.2021.3076751.

2] Kabashi, Qamil & Shabani, Isak & Caka, Nebi. (2022). Analysis of the Student Dropout Rate at the Faculty of Electrical and Computer Engineering of the University of Prishtina, Kosovo, From 2001 to 2015. IEEE Access. 10. 1-1. 10.1109/ACCESS.2022.3185620.

Doi:10.48047/ijiee.2025.15.5.64

3] Andrade-Girón, Daniel & Sandivar-Rosas, Juana & Marin-Rodriguez, William J. & Susanibar-Ramirez, Edgar & Toro-Dextre, Eliseo & Ausejo Sánchez, José & Villarreal-Torres, Henry & Angeles-Morales, Julio. (2023). Predicting Student Dropout based on Machine Learning and Deep Learning: A Systematic Review. ICST Transactions on Scalable Information Systems. 10.4108/eetsis.3586.

4] Garrison, D. & Akyol, Zehra. (2009). Role of instructional technology in the transformation of higher education. J. Computing in Higher Education. 21. 19-30. 10.1007/s12528-009-9014-7.

5] Romero, Cristóbal & Ventura, Sebastian. (2010). Educational Data Mining: A Review of the State of the Art. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on. 40. 601 - 618. 10.1109/TSMCC.2010.2053532.

6] Reason, Robert. (2009). Student Variables That Predict Retention: Recent Research and New Developments. Journal of Student Affairs Research and Practice. 46. 482-501. 10.2202/1949-6605.5022.

7] Kuh, George D., Jillian Kinzie, Jennifer Buckley, Brian K. Bridges and John C. Hayek. "What Matters to Student Success: A Review of the Literature." (2006).

8] Francisco Araque, Concepción Roldán, Alberto Salguero,Factors influencing university drop out rates,Computers & Education,Volume 53, Issue 3,2009,Pages 563-574,ISSN 03601315,https://doi.org/10.1016/j.compedu.2009.03.013.

9] Salas-Morera, L., Molina, A.C., Olmedilla, J.L., García-Hernández, L., & Palomo-Romero, J.M. (2019). Factors affecting engineering students dropout: a case study. International Journal of Engineering Education, 35, 156-167.

10] Kori, K., Pedaste, M., Altin, H., Tõnisson, E., & Palts, T. (2016). Factors That Influence Students' Motivation to Start and to Continue Studying Information Technology in Estonia. IEEE Transactions on Education, 59, 255-262.

11] J. P. Smith and R. A. Naylor, Dropping out of university: A statistical analysis of the probability of withdrawal for UK university students, J. Roy. Stat. Soc., Ser. A (Statist. Soc.), vol. 164, no. 2, pp. 389405, Jan. 2001

12] L. Paura and I. Arhipova, Student dropout rate in engineering education study program, in Proc. 15th Int. Sci. Conf. Eng. Rural Develop., Jelgava, Latvia, 2016, pp. 16. [8] A. Al-Mamun and R. A. MahbubHasan, Investigating the causes of stu dents dropout from the diploma engineering programs in the polytechnic institutes of Bangladesh, Educc. Sci., vol. 3, no. 2, p. 231, 2012.

13] D.KimandS.Kim,Sustainable education:Analyzing the determinants of university student dropout by nonlinear panel data models, Sustainability, vol. 10, no. 4, p. 954, Mar. 2018, doi: 10.3390/su10040954.

14] CRUE Universidades Españolas. La Universidad Española en Cifras: 2017/2018. Accessed: Apr. 16, 2021. [Online]. Available: https://www. crue.org/wp-content/uploads/2020/02/UEC-1718_FINAL_DIGITAL.pdf 66264

15] H. Vossensteyn and B. Stensaker. (2015). Dropout and Completion in Higher Education in Europe. Publications Of ce of the European Union, Luxembourg. [Online]. Available: http://ec.europa.eu/education/ library/study/2015/dropout-completion-he_en.pdf

16] F. Belloc, A. Maruotti, and L. Petrella, How individual characteristics affect university students drop-out: A semiparametric mixed-effects model for an Italian case study, J. Appl. Statist., vol. 38, no. 10, pp. 22252239, Oct. 2011.

17] L.S.Stratton,D.M.OToole,andJ.N.Wetzel, A multinomial logit model of college stop out and dropout behavior, Econ. Educ. Rev., vol. 27, no. 3, pp. 319331, Jun. 2008.

Doi:10.48047/ijiee.2025.15.5.64

18] A. M. S. Porto and A. B. Soares, Diferenças entre expectativas e adap tação academica de universitários de diversas áreas do conhecimento. [Differences between expectations and academic adaptation of university students from different areas of knowledge], Análise Psicológica, vol. 35, no. 1, pp. 1324, 2017.

19] L. Paura and I. Arhipova, Cause analysis of students dropout rate in higher education study program, Procedia-Social Behav. Sci., vol. 109, pp. 12821286, Jan. 2014.

20] J. R. Casanova, A. Cervero, J. C. Nunez, L. S. Almeida, and A. Bernardo, Factors that determine the persistence and dropout of university stu dents, Psicothema, vol. 30, no. 4, pp. 408414, 2018.

21] R. Chen, Institutional characteristics and college student dropout risks: A multilevel event history analysis, Res. Higher Educ., vol. 53, no. 5, pp. 487505, Aug. 2012.

22] G. Taylor, N. Lekes, H. Gagnon, L. Kwan, and R. Koestner, Need satisfaction, work school interference and school dropout: An application of self-determination theory, Brit. J. Educ. Psychol., vol. 82, no. 4, pp. 622646, 2012.

23] J. J. Vossensteyn, L. Cremonini, E. Epping, G. Laudel, and L. Leisyte, International experiences with student nancing: Tuition fees and student nancial support in perspective, Dutch Ministry Educ., Sci. Culture., Enschede, The Netherlands, Final Rep., 2013. [Online]. Available: http://www.ihem.nl/themereports/Experiences%20with%20student% 20 nancing%202013%20_ nal_March.pdf

24] E. Hovdhaugen, Working while studying: The impact of term-time employment on dropout rates, J. Educ. Work, vol. 28, no. 6, pp. 631651, 2015.

25] Eurydice. (2011). Modernisation of Higher Education in Europe: Funding and the Social Dimension. [Online]. Available: http://eacea.ec.europa.eu/ education/eurydice/documents/thematic_reports/131 EN.pdf

26] P.Kinnunen and L.Malmi, WhystudentsdropoutCS1course? inProc. Int. Workshop Comput. Educ. Res. (ICER), 2006, pp. 97108.

27] B. Divjak, M. Ostroski, and V. V. Hains, Sustainable student retention and gender issues in mathematics for ICT study, Int. J. Math. Educ. Sci. Technol., vol. 41, no. 3, pp. 293310, Apr. 2010. [24] J. Gairín, X. M. Triado, M. Feixas, P. Figuera, P. Aparicio-Chueca, and M. Torrado, Student dropout rates in catalan universities: Pro le and motives for disengagement, Qual. Higher Educ., vol. 20, no. 2, pp. 165182, May 2014.

28] F. M. F. Páramo, A. Araujo, C. Tinajero-Vacas, L. S. Almeida, and M.S.Rodriguez-Gonzalez, Predictors of students adjustment during the transition to university in Spain, Psicothema, vol. 29, no. 1, pp. 6772, 2017.

Doi:10.48047/ijiee.2025.15.5.64