IJAREEIE

e-ISSN: 2278-8875
p-ISSN: 2320-3765

International Journal of Advanced Research

in Electrical, Electronics and Instrumentation Engineering

Volume 14, Issue 5, May 2025

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.807

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering (IJAREEIE)

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.807 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 14, Issue 5, May 2025 ||

| DOI:10.15662/IJAREEIE.2025.1404011 |

Prediction of Student's Performance using Machine Learning

P. Nisandth, Dr. K. Poornapriya, Dr. G. Aarthi, P. Babu

Department of Master of Computer Applications, Vidyaa Vikas College of Engineering and Technology, Tiruchengode, Tamil Nadu, India

Principal, Vidyaa Vikas College of Engineering and Technology, Tiruchengode, Tamil Nadu, India

Head of the department, Department of Master of Computer Applications, Vidyaa Vikas College of Engineering and Technology, Tiruchengode, Tamil Nadu, India

Assistant Professor, Department of Master of Computer Applications, Vidyaa Vikas College of Engineering and Technology, Tiruchengode, Tamil Nadu, India

ABSTRACT: Thousands or even millions of students may learn at their own pace and according to their interests thanks to online learning platforms like Massive Open Online Courses (MOOC), Virtual Learning Environments (VLEs), and Learning Management Systems (LMS). In addition to offering several benefits, online learning environments also provide a number of difficulties, including low student engagement, disinterest, and self-regulated behavior, as well as the requirement that students define their own objectives. In this work, we suggested Rf, Svm, KNN, And Extra Tree Classifier as predictive models to examine the issues experienced by students who are considered to be at-risk. This allows teachers to intervene in a timely manner, encouraging students to become more engaged in their studies and perform better. Various machine learning (ML) and deep learning (DL) methods are used to train and evaluate the prediction model, which then characterizes students' learning behavior based on their study data. f-score, accuracy, precision, and support are used to compare the performance of different machine learning algorithms. When developing the prediction model at various percentages of the course duration, the machine learning method that yields the best results in terms of accuracy, precision, recall, support, and f-score metric is finally chosen. By identifying at-risk students early in the course and providing prompt intervention, the predictive model can assist teachers in preventing student performance prediction.

KEYWORDS: Student Performance Prediction, Machine Learning, Predictive Modeling, Education, Academic Outcomes, Dropout Prevention

I. INTRODUCTION

The incorporation of modern technology in education is a result of the ongoing pursuit of improving learning experiences and optimizing academic outcomes. The use of machine learning algorithms to forecast and evaluate student performance is one such exciting frontier. This emerging subject aims to develop prediction models that can predict a student's academic progress by utilizing the large quantity of educational data that is already available. Through the utilization of machine learning, educators and administrators can acquire significant insights into the variables impacting student performance. This can facilitate the development of tailored support strategies and prompt interventions. By enabling a more flexible and efficient educational system and enabling personalized learning, the nexus of education and technology has the potential to completely transform the way we approach learning.

1.1 STUDENT PERFORMANCE PREDICTION

The ever-changing environment of education has led to the integration of cutting-edge technology, especially when viewed through the prisms of machine learning and predictive analytics, in an effort to get insights into student performance. Large-scale datasets that have been integrated into educational systems have created new avenues for studying and predicting student results. A growing area at the nexus of technology and education is called "student performance prediction," which uses sophisticated algorithms to evaluate past data, spot trends, and build prediction models. With the help of these models, educators and administrators can proactively identify potential academic difficulties and provide tailored tactics and timely interventions to support each student's unique learning path.

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.807 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 14, Issue 5, May 2025 ||

| DOI:10.15662/IJAREEIE.2025.1404011 |

1.2 MACHINE LEARNING

A revolutionary branch of artificial intelligence called machine learning has become a key technology that might completely change a wide range of industries. Fundamentally, machine learning finds patterns and insights in data to enable systems to learn from experience and get better without explicit programming. Machine learning is at the vanguard of technological innovation thanks to its dynamic capability, which is having an impact on a wide range of industries including healthcare, finance, and education. Machine learning makes it possible to create predictive models, categorization schemes, and decision-making procedures that are superior to those made possible by conventional computational techniques by utilizing algorithms that can evaluate large datasets. Its versatility and ability to uncover obscure connections make it a priceless tool for improving decision-making and resolving challenging issues. This study investigates the use of machine learning in the particular situation of educational dropout prediction.

II. LITERATURE SURVEY

In this work, Hana Abdullah Monash [1] makes the basic point that picking candidates who are probably going to succeed scholastically in advanced education organizations requires a confirmations framework based on authentic and dependable affirmations models. This study centers around how information mining methods can be utilized to anticipate candidates' scholarly accomplishment at college to help colleges in going with affirmations choices. The recommended strategy was approved utilizing an informational index of 2,039 understudies selected at a Saudi state funded college's Software engineering and Data School somewhere in the range of 2016 and 2019. The discoveries demonstrate that certain pre-confirmation factors (secondary school grade normal, score on the Academic Accomplishment Affirmation Test, and score on the Overall Inclination Test) can be utilized to foresee a candidate's initial college accomplishment before confirmation. The results also show that the Scholastic Achievement Admission Test score is the pre-admission criterion that best predicts future student success. Along these lines, in affirmations processes, this score should be given more weight. Moreover, we found that the Counterfeit Brain Organization strategy beats the other three arrangement techniques that were inspected (Choice Trees, Backing Vector Machines, and Guileless Bayes) with an exactness pace of more than 79%. Pre-affirmation necessities, understudy execution, execution forecast, instructive information mining, and information mining draws near.

In this paper, Elman Alayna [2] poses the case that understudy accomplishment is significant in light of the fact that it's habitually used as a presentation marker for instructive establishments. Preventive methodologies related to early ID of in danger understudies can fundamentally build their accomplishment. As of late, there has been a great deal of utilization of AI strategies for expectation. Albeit the writing is loaded with examples of overcoming adversity, these techniques are for the most part accessible to teachers who are capable in "software engineering," or all the more precisely, "man-made consciousness." The facts confirm that making various decisions is vital for the powerful and proficient execution of information mining methods. Determining which machine learning approach is best suited for a given circumstance, defining student success, and choosing which student characteristics to emphasize are examples of these decisions. The objective of this task is to give educators who need to utilize information mining strategies to gauge understudy achievement a nitty gritty arrangement of rules. To do this, the writing has been inspected, and the cutting edge has been collected into a purposeful technique wherein possible decisions and boundaries are entirely examined, upheld by reasons, and made sense of.

Lassa ad K. Smyrna [3] et al. in this study propose that the world is seeing an unrivaled flood in the development of living structures because of the headway of data and correspondence innovation (ICT). The social, financial, and social aspects of life are continually affected by ICT. Various countries are endeavoring to execute the data society effectively. through strategies and foundation to propel information obtaining and make astute social orders. Because of the speedy headway of shrewd gadgets and specialized networks, various Learning The Board Framework (LMS) arrangements have been accessible in the scholarly space. Various researchers have zeroed in their endeavors on making new ways to deal with manage these issues in this climate. The primary influx of scientists zeroed in on versatile learning and proposed frameworks as replies to these issues, as. They assert that adaptive education systems may significantly aid students' academic progress through the utilization of individualized environments, such as teaching strategies, resources, and evaluations. Nonetheless, these frameworks keep on confronting a lot of trouble with respect to the type of the educational techniques and point matter. To further develop understudy grouping execution and result expectation, a Stacked Speculation for Disappointment Expectation (SGFP) is utilized in this review.

at this exploration, Faizaan Rouser [4] suggests that recognizing in danger understudies at scholarly organizations at the earliest opportunity is a huge trouble. Any educational establishment's goal is to create a learning environment that maximizes student performance and quickly identifies at-risk students. Toward the finish of a semester, disappointment rates will decline because of early distinguishing proof of in danger understudies. The motivation behind this review is

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.807 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 14, Issue 5, May 2025 ||

| DOI:10.15662/IJAREEIE.2025.1404011 |

to analyze a few characterization procedures, for example, capability based, tree-based, managed based, languid based, and Bayes-based calculations, to recognize in danger understudies from the get-go in the semester in light of engaging information. Before the semester closes, understudies who are recognized as being in danger of fizzling may get specific advising and mentoring. In the current review, grouping calculations are additionally contrasted with deference with their TP rate, FP rate, precision, review, Measure, exactness, kappa measurements, and model development time. A veritable dataset was utilized for the investigations, and the outcomes exhibit the cutthroat presentation of different calculations.

In this paper, Dipta Das [5] puts forward the viewpoint that understudy execution is urgent in delivering the greatest alumni who will be accountable for the country's social and financial turn of events. The work market is moreover worried about understudy accomplishment since the employing of late alumni depends on their scholastic standing. Deciding the reason for an understudy's exhibition fluctuation, hence, offers significant data for creating strategies and projects. Various specialists in different countries endeavor to decide the reason utilizing different information mining procedures. They didn't, be that as it may, all work with students from Bangladesh. A model was presented in this study to anticipate outcomes and identify the most important variables affecting Bangladeshi students' academic performance. In this review, a strategy that can perceive which understudies need additional support is proposed. To pick significant highlights, an assortment of element determination methods was applied, including co-connection, chi-square, and Euclidean distance. Furthermore, showing the consequences of element choice examinations utilizing the classifier calculations of fake brain organizations, choice trees, Credulous Bayes, and K-Closest Neighbor.

III. EXISTING SYSTEM

One of the most important study areas in the field of Educational Data Mining (EDM) is predicting students' academic success early in the semester. Undergraduate programs' "Programming" and "Data Structures" courses present a variety of challenges for students, which explains why failure and dropout rates are high in these subjects. In order to assist students, perform better in their next classes, EDM is used to evaluate student data collected from various educational settings in order to anticipate students' academic achievement. This study aims to investigate the effectiveness of deep learning in the field of early detection of marriage (EDM), particularly in forecasting students' academic performance and identifying at-risk pupils. Using a deep neural network (DNN), decision tree, random forest, gradient boosting, logistic regression, support vector classifier, and K-nearest neighbor, the researchers developed predictive models to predict students' academic performance of upcoming courses given their grades in the previous courses of the first academic year. The dataset was collected from a public 4-year university. Furthermore, we compared different resampling techniques, including SMOTE, ADASYN, ROS, AND SMOTE-ENN, to address the unbalanced dataset issue. Based on the results of the experiment, it is observed that the proposed DNN model, which is more accurate than models like decision trees, logistic regression, support vector classifiers, and K-nearest neighbor, can predict students' performance in a data structure course and can also identify students at risk of failing early in the semester.

IV. PROPOSED SYSTEM

By providing timely intervention to solve student problems, the suggested system improves online learning environments through a multidimensional approach. The training and testing datasets, which include important research variables, engagement level, time-dependent variables, and assessment results, are loaded first. The system creates a prediction model to examine at-risk pupils' behavior by utilizing machine learning methods including Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Extra Tree Classifier. The classification and prediction process is aided by the RF, SVM, KNN, and Extra Tree modules, which enable teachers to recognize at-risk pupils early in the course. This proactive intervention uses precise predictions and insights from many machine learning approaches to enhance study engagement and overall performance. Metrics including accuracy, precision, recall, support, and f-score are used to evaluate the system's efficacy. The most effective algorithm is then chosen to provide a prediction model that may be adjusted to different percentages of the course duration.

A. Load Training Dataset

This module involves the initial step of loading the training dataset. The dataset likely contains information on students' study variables, engagement intensity, time-dependent variables, and assessment scores. Loading this dataset is crucial for training the machine learning algorithms to build the predictive model. The first step in this module is to load the training dataset. The dataset probably includes data on the study characteristics, level of involvement, time-dependent factors, and test results of the students. In order to train the machine learning algorithms and create the prediction model, loading this dataset is essential.

B. Load Testing Dataset

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering (IJAREEIE)

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.807 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 14, Issue 5, May 2025 ||

| DOI:10.15662/IJAREEIE.2025.1404011 |

This module focuses on loading the testing dataset in a manner akin to that of the training dataset. To assess how well the trained prediction model performs, the testing dataset is necessary. It provides the foundation for evaluating the model's predictive and intervening capabilities in real-world situations and comprises data on pupils who are considered to be at-risk.

C. RF Classification and Prediction

To identify and categorize pupils who are at risk, this module applies the Random Forest (RF) classification technique. Known for its ensemble learning methodology, Random Forest combines many decision trees to improve accuracy. Using the training dataset, the RF model is trained in this module, and it is then used to forecast using the testing dataset.

D. Svm Classification and Prediction

This lesson uses Support Vector Machine (SVM), a potent classification method, to forecast and categorize pupils who are at risk. The way SVM operates is by identifying the best hyperplane to divide several classes. In order to predict at-risk students in the testing dataset, the SVM model must first be trained on the training dataset.

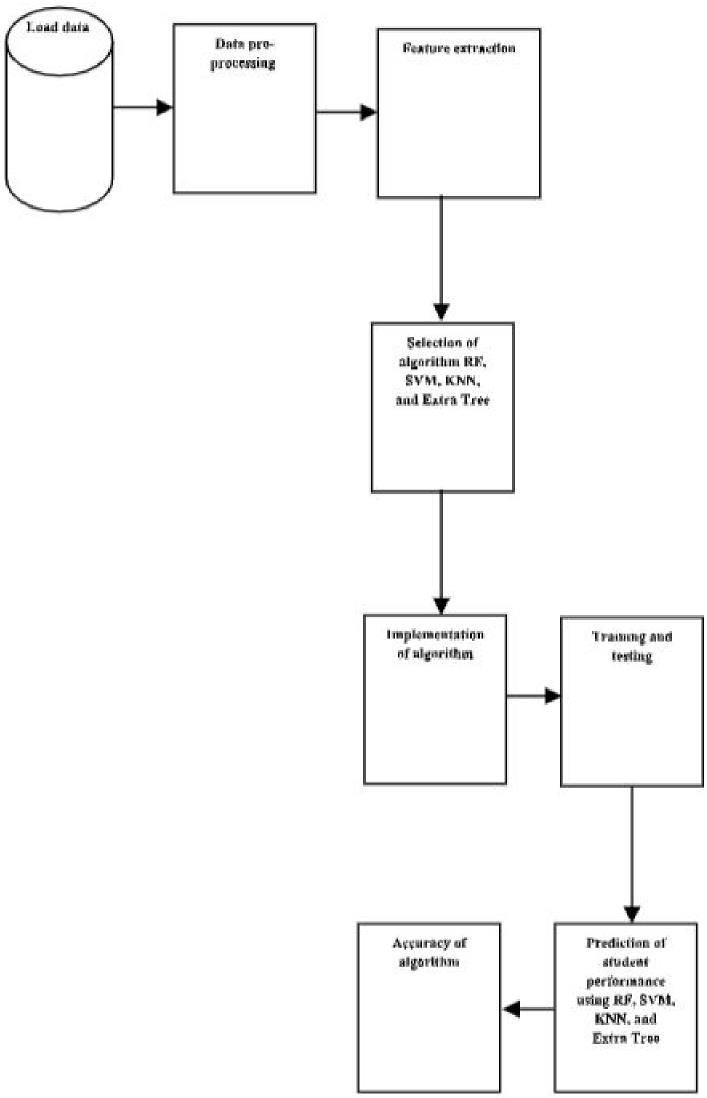


FIGURE1: BLOCK DIAGRAM

E. KNN Classification and Prediction

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering (IJAREEIE)

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.807 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 14, Issue 5, May 2025 ||

| DOI:10.15662/IJAREEIE.2025.1404011 |

The K-Nearest Neighbors (KNN) classification technique is the main topic of this module. Data points are categorized by KNN according to the majority class of their closest neighbors. In order to predict at-risk individuals in the testing dataset, the KNN model must first be trained on the training dataset.

F. Extra tree Classification and Prediction

This module uses the Extra Tree Classifier to identify and categorize pupils who are considered to be at-risk. Extra Tree is an ensemble learning technique that enhances prediction accuracy by constructing additional decision trees and combining them. Using the training dataset, the Extra Tree model is trained in this module and then used to the testing dataset to make predictions.

V. ALGORITHM DETAILS

A. Support Vector Machine

Support Vector Machine (SVM) is a popular supervised machine learning model that is used for classification and prediction of unknown data. It is asserted by several researchers that SVM is a very accurate technique for text classification. It is also widely used in sentiment classification. For instance, if we have a dataset in which data is pre-labeled into two categories: positive and negative reviews, then we can train a model to classify new data into these two categories. This is exactly how SVM works. It is the model that we train on a dataset, so it can analyze and classify unknown data into the categories that were present in the training set. SVM is a linear learning method. It finds an optimal hyper-plane to differentiate two classes. Being a supervised classification model, it tries to maximize the distance between the closest training point and either class so as to achieve better classification performance on test data. The process for classification functions is as follows:

• It takes the labeled sample of data, and draws a line separating the two classes. This line is called the decision boundary. The solution is based only on those training data points which are really close to the decision boundary. The data points are called Support Vectors. For example, if we are categorizing movie reviews (in our case), one side of the boundary will have positive reviews while the other side has negative reviews.

• Now when new data needs to be classified, it goes either into the left or right side of the decision boundary. Depending on which side the data enters, it is classified under that category. To classify our data with the best precision, we need to split the two categories such that the decision boundary separates the two classes with maximum space between them.

B. Random Forest

Random forest is a supervised learning algorithm. It is often used both for classification and regression. How Random Forest algorithm works: There are two stages in Random Forest algorithm, one is random forest creation, the opposite is to form a prediction from the random forest classifier created in the first stage. The whole process is shown below, and it's easy to understand using the figure.

1.Here the author firstly shows the Random Forest creation pseudocode: Randomly select "K" features from total "m" features where k << m.

2. Among the "K" features, calculate the node "d" using the best split point.
3. Split the node into daughter nodes using the best split.
4. Repeat the a to c steps until "l" number of nodes has been reached.
5. Build forest by repeating steps a to d for "n" number times to create "n" number of trees

C. K Nearest Neighbour

Occurrence based learning is One method for fathoming assignments of approximating discrete or genuine esteemed objective capacities and KNN is one this sort of grouping. In KNN not the same different characterizations, we store preparing Examples, when a test model is given we discover the closest neighbor.

D. Extra tree classifier

As the numbers of features are very large it is necessary to select important features from them in order to work efficiently with the dataset. In this model extra tree classifier algorithm is used in order to select necessary features from the extracted features. This algorithm is given with the input of all the features extracted from the feature extraction step. The Extra Trees Forest consists of decision tree which is constructed from the original training sample. Each decision tree is provided with a random sample of k features from the feature-set from which each decision tree must select the best feature to split the data based on some mathematical criteria.

$$\text{Entropy}(S) = \Sigma{-pi}\log2(pi)$$

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.807 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 14, Issue 5, May 2025 ||

| DOI:10.15662/IJAREEIE.2025.1404011 |

c - Number of unique class labels
i - Output Label
Formula for information gain is
Gain(S,A)=Entropy(S) $\Sigma veValues(A)|Sv|/|S|Entropy(Sv)$

VI. RESULT ANALYSIS

The examination of the data shows that the various machine learning algorithms used to detect at-risk students in online learning settings have differing degrees of accuracy. With an accuracy of 95%, the Random Forest (RF) classifier demonstrated the best level of competency in using ensemble learning for classification. The Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) models demonstrated varying degrees of success in this predictive modeling scenario, as seen by their respective accuracies of 71% and 79%. With an accuracy of 80%, the Extra Tree Classifier demonstrated competitive performance comparable to that of KNN. While accuracy gives a broad picture, additional analysis taking into account precision, recall, and F1-score could provide a more detailed understanding of the advantages and disadvantages of each algorithm, assisting in the selection of the best model for prompt intervention and support of at-risk students in online learning environments.
One of the most widely used metrics for assessing classification performance is accuracy, which is calculated as the ratio of correctly segmented samples to all samples.

Accuracy = TP/ (TP+ FN)
Precision: The number of positive class predictions that truly belong to the positive class is quantified by precision, which is estimated in the manner described below.
Precision = TP / (TP + FP)
The ratio of true positives to total (real) positives in the data is known as recall or sensitivity. Sensitivity and recall are synonymous.
Recall = TP / (TP + FN)
The ratio of genuine negatives to total negatives in the data is known as specificity. Specificity is the program's accurate designation for everyone who is actually healthy.
Specificity = TN / (TN + FP)

algorithm	precision	recall	F-score	Accuracy
Rf	0.97	0.95	0.95	0.95
KNN	0.79	0.75	0.78	0.79
Svm	0.73	0.7	0.74	0.71
Extra tree classifier	0.8	0.81	0.84	0.8

Table 1. Comparison table

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.807 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 14, Issue 5, May 2025 ||

| DOI:10.15662/IJAREEIE.2025.1404011 |

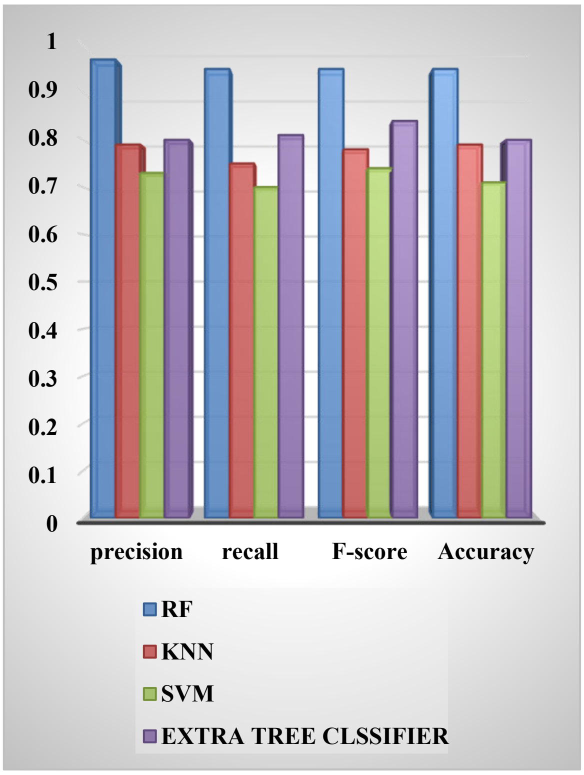


Figure 2. Comparison graph

VII. CONCLUSION

The study's findings, which emphasize the early identification of at-risk students through the use of machine learning algorithms, provide a thorough framework for resolving issues in online learning settings. By proactively identifying and assisting students who are experiencing problems, the suggested system which makes use of Random Forest, Support Vector Machine, K-Nearest Neighbors, and Extra Tree Classifier modules demonstrates the potential to improve student engagement and performance. A comprehensive evaluation of performance is ensured by the inclusion of several measures for algorithm evaluation, and the chosen algorithm creates a strong prediction model that can be adjusted to different course phases. This predictive model shows potential for teachers looking to offer focused assistance as online learning takes off, which will eventually help students' overall experience on digital learning platforms.

VIII. FUTURE WORK

Further research into the integration of continuous monitoring and real-time data streams may make it possible to create an intervention system that is more responsive and dynamic. Subsequent investigations may potentially explore the creation of customized educational interventions based on the requirements of specific students, utilizing adaptive learning technology. By examining the effects of new technology and pedagogical methods, as well as the changing landscape of online education, predictive models may be improved and made more flexible in order to better assist the academic achievement of students in virtual learning settings.

REFERENCES

1.I. R. Moonsamy, t. v. Spreitzer. As well as Korak. Mangard, "predicting student performance using data mining techniques to support decision making in university admission systems," ieee commun. Exams tuts., no. 20, no. 1, 1st quarter, pp. 465–488, 2017.
2.M. M. Guerar. f. Migliardi. Palmieri, l. A verderame, as well. Merlo, Concurrency Computing, Practice: "Predicting Academic Success in Higher Education: A Literature Review and Best Practices." Exam., no. No. 32, page 18. September 2020; E5549.

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.807 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 14, Issue 5, May 2025 ||

| DOI:10.15662/IJAREEIE.2025.1404011 |

3. Three.O Maiti. Armstrong, M. J. and Jadliwala. "Utilizing ensemble learning algorithms to forecast student failure and facilitate personalized learning pathways," he stated in the proceeding. The 11th Asia-Pacific ACMM Conference. Computer. Connect. 2016, pp. 795–806, Secur.

4.I. R. Zhao (c). Q and Yue. Han, "from students' data, early detection of failure risks," ieee trans. Def. Security in forensics, vol. 14, no. 1, January 2019, pp. 75–89.

5.M. Erini Nerini. m. and Favarelli. Chiani, A comparison of four classification systems for performance detection of university students, vol. 22, no. 13, June 2022, p. 4857.

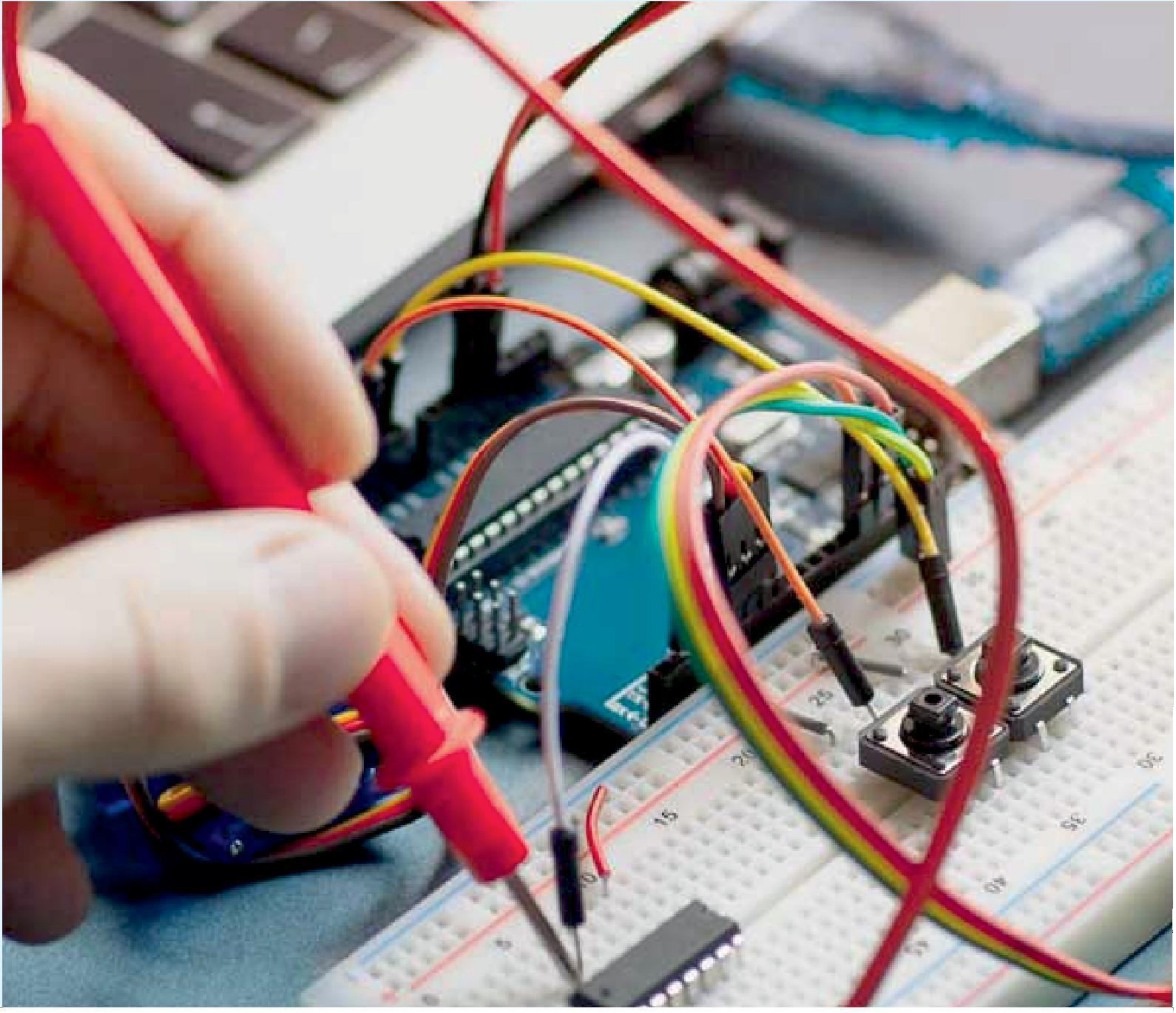
6.I. T. Van Vang, n. Both n and Sae-bae. Computer, "Comparing various resampling techniques in predicting students' performance through machine learning techniques:" Safety, no. 66, May 2017, pp. 115–128.

6.A. J. P and Kim. Kang, "Applying machine learning to predict academic success of students," Nature, vol. 12, no. 15, July 2022, p. 7590.

8.A. E. G. Ivannikova. As well as T. Hamalainen, "Deep learning algorithms' efficacy in forecasting student accomplishments," in proc. Aww, symp. Computer. Connect. July 2017, (iscc), pp. 885–889.

9.A. Yate, A. Banavavar d. As well as s. Schuckers "Feature selection based on genetic algorithms with ensemble methods for predicting academic performance of students" ieee trans. Identity science, behavior, biometrics, vol. 2, no. 4, June 2020, pp. 377–387.

10. Ten.S. You, Panda. G. Liu. A. P. You and Hancke. M. Qureshi, "K-nearest neighbor and c4.5 on smote-balanced data: a prediction of students' academic performance," sensors, vol. May 2020; 20, no. 11, p. 3015.



INNO SPACE
SJIF Scientific Journal Impact Factor

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

International Journal of Advanced Research

in Electrical, Electronics and Instrumentation Engineering