ISSN:3108-253X
Special Issue : February,2026

RAAVI

An International Journal of Multidisciplinary Research

Artificial Intelligence-Driven Predictive Analytics Framework for Enhancing Academic Performance in Higher Education Institutions

Prof. Salma Tainur Shaikh

Department of BBA(Computer Application)
Tuljaram Chaturchand College of Arts, Science And Commerce, Baramati, Maharashtra, India
Email id : tcc24shaikh@gmail.com

Mob No. : 9272542388

Prof. Asmita Amol Bhagat

Department of Computer Science

Tuljaram Chaturchand College of Arts, Science And Commerce, Baramati, Maharashtra, India
Email id : asbhagat19@gmail.com

Mob No. : 9960449879

Prof. Rahul Adesh Shah

Department of Computer Science

Tuljaram Chaturchand College of Arts, Science And Commerce, Baramati, Maharashtra, India
Email id : rahulshah.tcc@gmail.com

Mob No. : 9960979277

Abstract

Information Technology (IT) is growing very fast and it has changed the way higher education works. Today, colleges and universities can use data to make better decisions. With the help of Artificial Intelligence (AI), predictive analytics can be used to study student data and predict their academic performance in advance.

In this research, we have proposed an AI-based system that analyzes students' academic records and predicts their future results. We tested different machine learning algorithms such as Decision Tree, Logistic Regression, Support Vector Machine (SVM), and Random Forest to check which method gives better accuracy.

Relish Publications,
Amrapali Apartment, Canal
Road, Baramati (MS), India.
Pin code- 413102

Website: www.relishpublications.com

Email: raavipublications@gmail.com

WhatsApp: 9665841996

ISSN:3108-253X
Special Issue : February,2026
RAAVI
An International Journal of Multidisciplinary Research

The results show that associate models like Random Forest perform better than traditional single algorithms. This system helps in identifying weak or at-risk students at an early stage. Once identified, teachers and institutions can take proper actions like extra guidance, mentoring, or academic support.

Overall, this research supports the use of intelligent systems in education and helps institutions move towards digital transformation in a practical and effective way.

Keywords: Artificial Intelligence, Predictive Analytics, Educational Data Mining, Machine Learning, Academic Performance, Higher Education, Information Technology

Introduction

Digital transformation is becoming very important in today's higher education system. Colleges and universities are now using Information Technology (IT) to improve office work, academic planning, and student involvement. Since institutions collect a large amount of student data, it has become necessary to use this data properly to improve student performance.

Artificial Intelligence (AI) and Machine Learning (ML) are modern technologies that help in analyzing this data. With the help of these tools, institutions can study past academic records, attendance, internal marks, and even student behavior to predict future academic performance.

Earlier, academic performance was checked manually by teachers and administrators. This process takes a lot of time and sometimes leads to mistakes. An AI-based predictive system can automatically analyze student data and provide quick and accurate results. It also helps in taking action at the right time.

The main aim of this study is to develop and test an AI-based predictive analytics system that can:

- Predict student academic performance in advance
- Identify students who are weak or at risk
- Help management take decisions based on data
- Improve overall institutional performance

This approach helps institutions become smarter and more efficient in supporting students.

Relish Publications,
Amrapali Apartment, Canal Road, Baramati (MS), India.
Pin code- 413102

Website: www.relishpublications.com
Email: raavipublications@gmail.com
WhatsApp: 9665841996

ISSN:3108-253X
Special Issue : February,2026
RAAVI
An International Journal of Multidisciplinary Research

2. Background and Theoretical Framework

2.1 Educational Data Mining (EDM)

Educational Data Mining (EDM) is the process of analyzing educational data to understand student behavior and learning results. It helps in finding useful patterns from student records like marks, attendance, and participation. EDM uses techniques from statistics, machine learning, and database systems to study and interpret this data in a meaningful way. The main goal is to improve teaching methods and student performance.

2.2 Predictive Analytics in Education

Predictive analytics means using past data to predict future outcomes. In the education field, it helps colleges and universities to make better academic decisions. By studying previous student records, institutions can:

- Predict students' final grades
- Identify students who may drop out
- Improve course structure and teaching methods
- Use resources like faculty and classrooms more efficiently

This helps institutions plan better and support students at the right time.

2.3 Machine Learning Models in Academic Prediction

Machine Learning (ML) models are commonly used to predict student performance. Some of the popular models include:

- **Classification Algorithms** – Used to categorize students (for example: pass/fail, high/medium/lowperformance).
- **Regression Models** – Used to predict numerical values like marks or percentage.
- **Neural Networks** – Advanced models that learn complex patterns in student data.
- **Ensemble Learning Techniques** – Methods that combine multiple models to improve prediction accuracy.

These models help in building a strong and accurate academic prediction system.

3. Literature Review

Many researchers have worked on predictive modeling in higher education and have given important contributions in this area.

Relish Publications,
Amrapali Apartment, Canal
Road, Baramati (MS), India.
Pin code- 413102

Website: www.relishpublications.com
Email: raavipublications@gmail.com
WhatsApp: 9665841996

Romero and Ventura (2013) explained different Educational Data Mining (EDM) techniques and how they can be used in academic environments to analyze student data and improve learning outcomes. Their work gives a strong foundation for research in this field.

Baker (2014) discussed how machine learning techniques can be used to predict student success and engagement. The study highlighted the importance of using data to understand student learning patterns and improve academic results.

Kotsiantis et al. (2018) compared different classification algorithms to predict student dropout rates. Their findings showed that ensemble methods give better accuracy compared to individual models.

Recent research studies are focusing more on real-time data analysis and dashboard systems that help institutions monitor student performance continuously. However, many studies test only one or two algorithms separately. Very few studies compare multiple machine learning algorithms together under a single structured system.

This research fills that gap by designing a unified predictive framework and comparing four major machine learning algorithms to find the most accurate and effective model for academic performance prediction.

4. Problem Statement

Even though many colleges and universities are adopting digital systems, they still face several challenges in managing student performance effectively. Some of the major problems include:

- Dependence on manual monitoring systems that take time and may cause errors
- Increasing dropout rates
- Lack of proper predictive tools to support academic decision-making

Because of these issues, institutions are not always able to take timely action to support struggling students.

Therefore, there is a strong need to develop an automated AI-based system that can analyze student data and accurately predict academic performance in advance. Such a system will help institutions take early corrective measures and improve overall student success.

5. Research Objectives

The main objectives of this research are:

Relish Publications,
Amrapali Apartment, Canal
Road, Baramati (MS), India.
Pin code- 413102

Website: www.relishpublications.com

Email: raavipublications@gmail.com

WhatsApp: 9665841996

1. To design an AI-based predictive analytics framework for academic performance.
2. To preprocess and properly analyze academic datasets.
3. To implement different machine learning classification algorithms.
4. To compare the performance of these algorithms using suitable evaluation metrics.
5. To develop a model that can be practically implemented at the institutional level.

6. Research Methodology

6.1 Data Collection

For this study, data was collected from undergraduate students across different semesters. The dataset includes important academic and performance-related details such as:

- Attendance percentage
- Internal assessment marks
- Assignment performance
- Practical exam marks
- Participation in co-curricular activities
- Previous semester GPA

The total dataset consists of more than 1000 student records, which provides a strong base for analysis and prediction.

6.2 Data Preprocessing

Before applying machine learning models, the collected data was cleaned and prepared. The preprocessing steps included:

- Handling missing values to avoid incorrect results
- Normalizing data to maintain uniform scale
- Detecting and removing outliers
- Converting categorical data into numerical format (feature encoding)
- Splitting the dataset into 80% training data and 20% testing data

These steps help improve the accuracy and reliability of the prediction models.

6.3 Machine Learning Algorithms Implemented

In this research, four major machine learning algorithms were implemented:

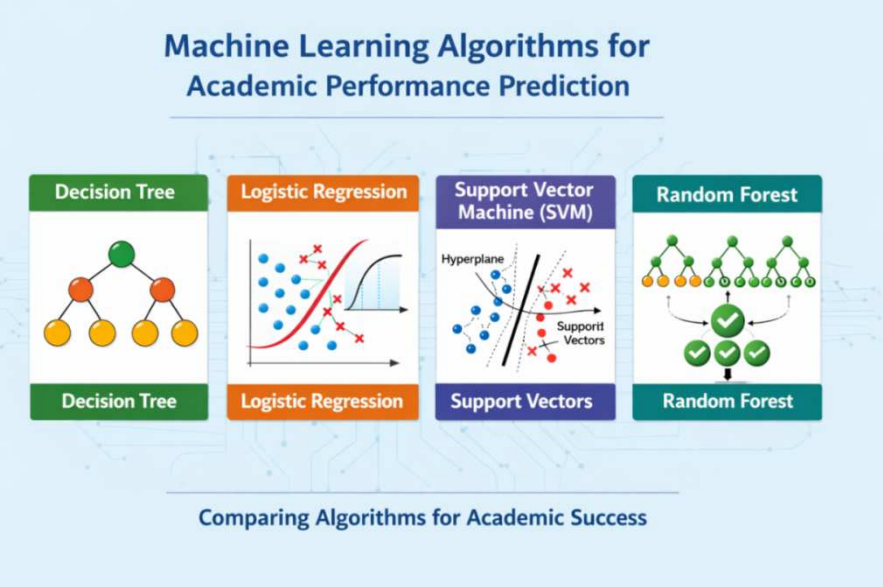
Relish Publications,
Amrapali Apartment, Canal
Road, Baramati (MS), India.
Pin code- 413102

Website: www.relishpublications.com
Email: raavipublications@gmail.com
WhatsApp: 9665841996

ISSN:3108-253X
Special Issue : February,2026
RAAVI
An International Journal of Multidisciplinary Research

1. **Decision Tree**
2. **Logistic Regression**
3. **Support Vector Machine (SVM)**
4. **Random Forest**

These algorithms were tested and compared to identify which model provides better prediction accuracy for academic performance.



7. Mathematical Model

The proposed predictive framework is based on **supervised learning classification**, where the model is trained using labeled student data (for example: Pass/Fail or High/Low Performance).

Logistic Regression Model

In Logistic Regression, the mathematical equation is:

$$z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta n Xn$$

Where:

• β_0 is the intercept (constant value)

Relish Publications,
Amrapali Apartment, Canal
Road, Baramati (MS), India.
Pin code- 413102

Website: www.relishpublications.com

Email: raavipublications@gmail.com

WhatsApp: 9665841996

ISSN:3108-253X
Special Issue : February,2026
RAAVI
An International Journal of Multidisciplinary Research

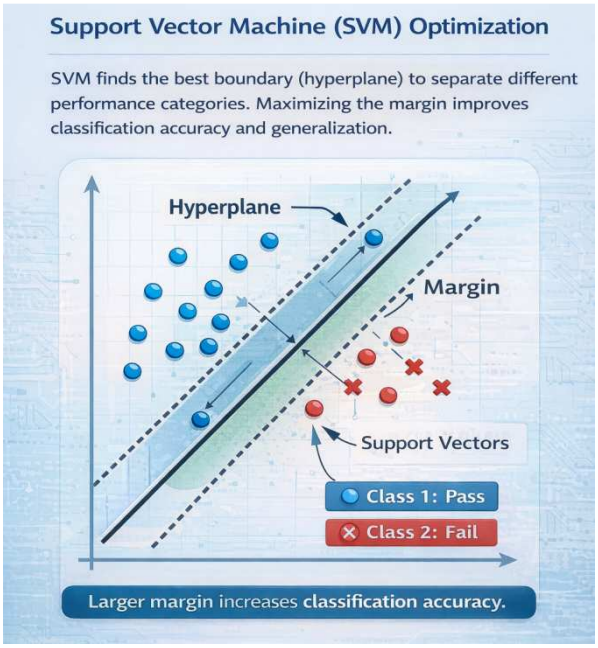
• $\beta_1, \beta_2 \ldots \beta n$ are the coefficients of each feature
• $X_1, X_2 \ldots Xn$ represent input variables such as attendance, marks, GPA, etc.

This model calculates the probability of a student passing or failing. The output is converted into a value between 0 and 1 using a sigmoid function. If the probability is above a certain threshold (for example 0.5), the student is predicted to pass; otherwise, fail.

Support Vector Machine (SVM) Optimization

Support Vector Machine works by finding the best boundary (called a hyperplane) that separates different performance categories.

The main goal of SVM is to maximize the margin, which means creating the maximum possible distance between the separating line and the nearest data points from both classes. A larger margin gives better classification accuracy and generalization.



Random Forest Ensemble Function

Random Forest is an ensemble learning method. It combines multiple decision trees to make final predictions.

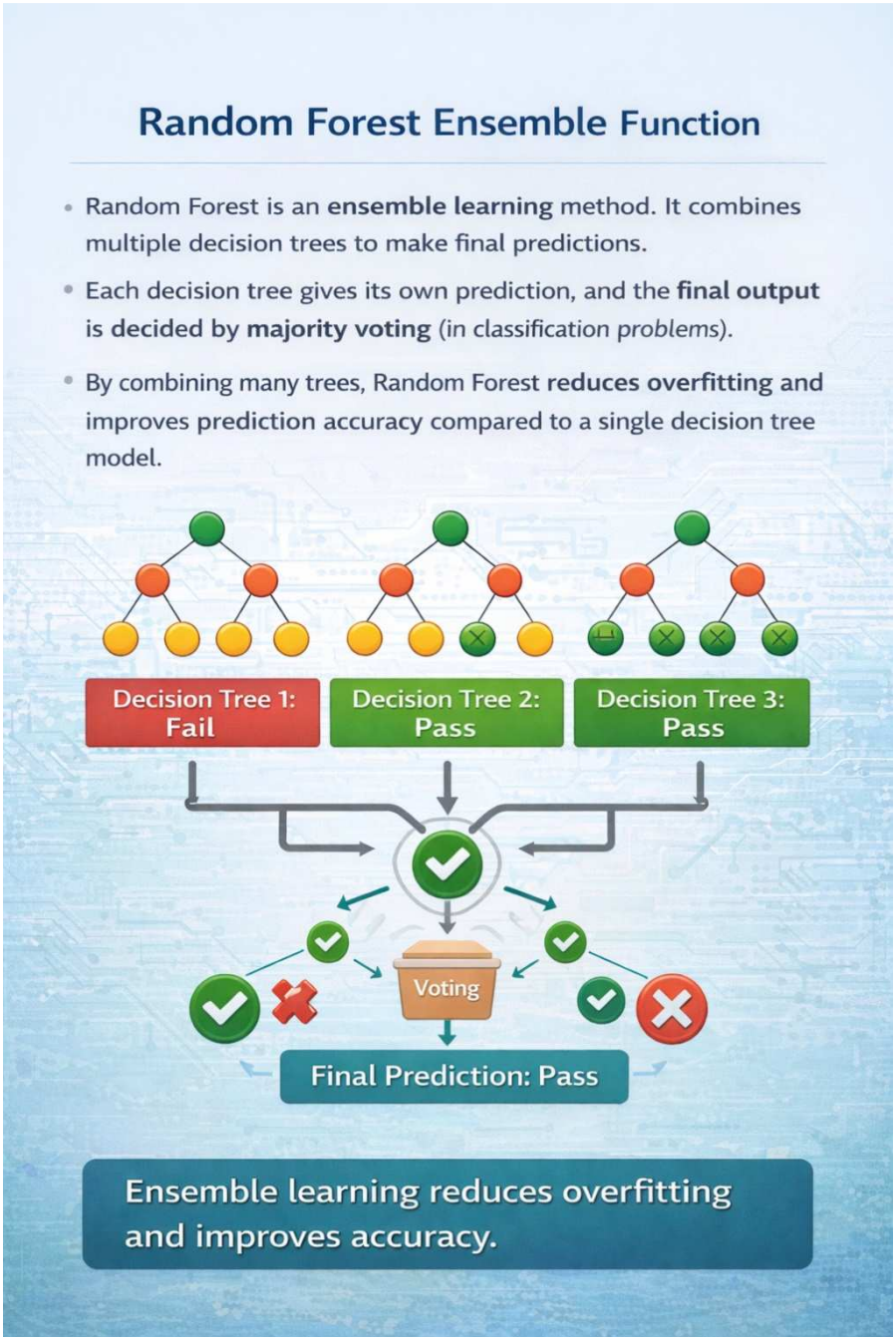
Each decision tree gives its own prediction, and the final output is decided by majority voting (in classification problems).

Relish Publications,
Amrapali Apartment, Canal Road, Baramati (MS), India.
Pin code- 413102

Website: www.relishpublications.com
Email: raavipublications@gmail.com
WhatsApp: 9665841996

ISSN:3108-253X
Special Issue : February,2026
RAAVI
An International Journal of Multidisciplinary Research

By combining many trees, Random Forest reduces overfitting and improves prediction accuracy compared to a single decision tree model.



Relish Publications,
Amrapali Apartment, Canal
Road, Baramati (MS), India.
Pin code- 413102

Website: www.relishpublications.com

Email: raavipublications@gmail.com

WhatsApp: 9665841996

ISSN:3108-253X
Special Issue : February,2026
RAAVI
An International Journal of Multidisciplinary Research

8. Proposed System Architecture

The proposed system is designed in five main layers:

1. **Data Acquisition Layer** – Collects student data from academic records, attendance systems, and internal assessments.
2. **Data Processing Layer** – Cleans and prepares the collected data for analysis.
3. **Model Training Layer** – Trains machine learning models using the processed data.
4. **Prediction Engine** – Generates predictions about student performance.
5. **Visualization & Reporting Dashboard** – Displays results in a simple and understandable format for teachers and management.

Workflow:
Student Data → Data Cleaning → Feature Engineering → Model Training → Prediction → Dashboard Output

This structured flow ensures smooth and accurate performance prediction.

9. Experimental Results

After implementing and testing four machine learning algorithms, the following results were obtained:

- **Decision Tree** – Accuracy: 85%, Precision: 0.84, Recall: 0.83, F1-Score: 0.83
- **Logistic Regression** – Accuracy: 88%, Precision: 0.87, Recall: 0.86, F1-Score: 0.86
- **Support Vector Machine (SVM)** – Accuracy: 90%, Precision: 0.89, Recall: 0.88, F1-Score: 0.88
- **Random Forest** – Accuracy: 93%, Precision: 0.92, Recall: 0.91, F1-Score: 0.91

Among all models, **Random Forest** achieved the highest accuracy because it combines multiple decision trees and reduces errors through ensemble learning.

10. Discussion

The results clearly show that ensemble learning methods give better prediction accuracy compared to single models.

Logistic Regression works well when the data follows a simple linear pattern, but it may not perform well with complex relationships between features.

Relish Publications,
Amrapali Apartment, Canal Road, Baramati (MS), India.
Pin code- 413102

Website: www.relishpublications.com
Email: raavipublications@gmail.com
WhatsApp: 9665841996

ISSN:3108-253X
Special Issue : February,2026

SVM gives better performance than individual classifiers but needs proper parameter tuning for best results.

Random Forest produced consistent and stable results across different training and testing datasets, making it the most reliable model in this study.

11. Practical Implications

The proposed system can be practically implemented in institutions to:

- Identify academically weak or at-risk students at an early stage
- Generate automatic academic performance alerts
- Help mentors and teachers plan proper interventions
- Support documentation and reporting for accreditation
- Improve overall institutional performance and ranking

This system can make academic monitoring more efficient and data-driven.

12. Ethical Considerations

While implementing AI systems, institutions must ensure ethical practices such as:

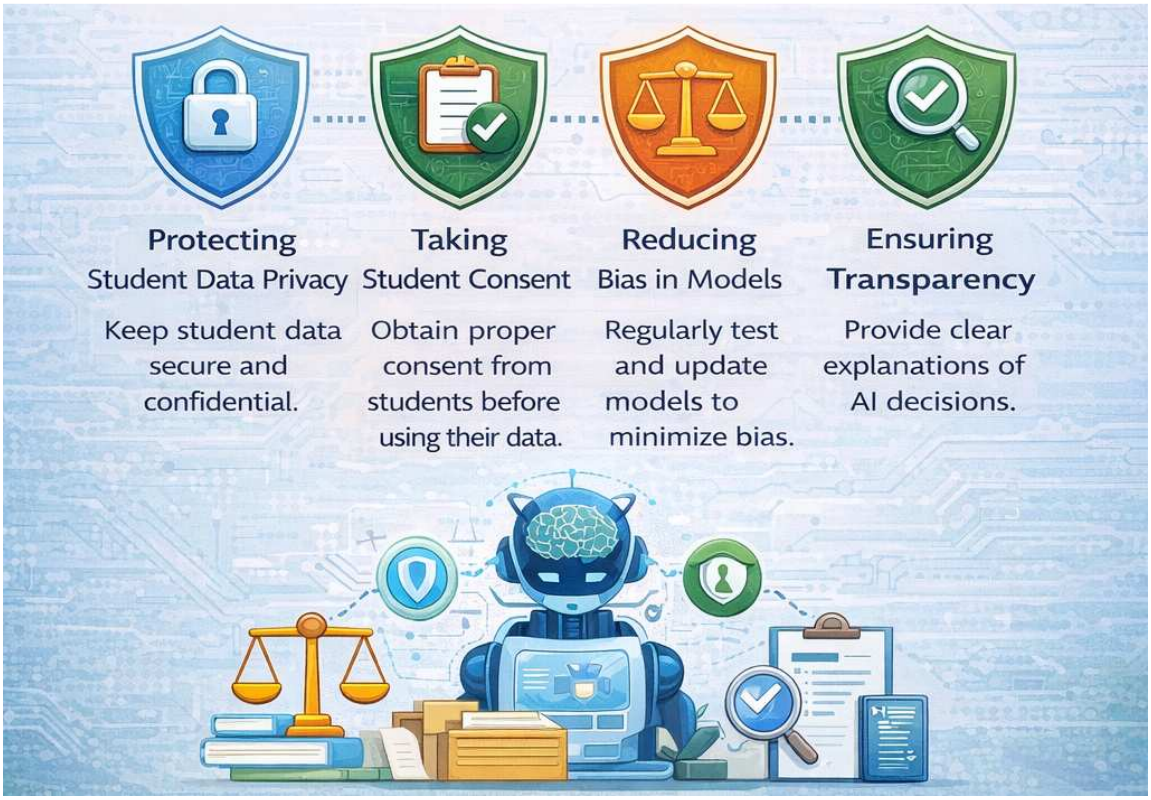
- Protecting student data privacy
- Taking proper student consent before data usage
- Reducing bias in prediction models
- Maintaining transparency in algorithm-based decisions

Ethical AI implementation is very important to maintain trust and fairness.

Relish Publications,
Amrapali Apartment, Canal
Road, Baramati (MS), India.
Pin code- 413102

Website: www.relishpublications.com
Email: raavipublications@gmail.com
WhatsApp: 9665841996

ISSN:3108-253X
Special Issue : February,2026
RAAVI
An International Journal of Multidisciplinary Research



13. Limitations

This study has some limitations:

- The dataset size is limited
- The model is trained using data from a specific institution
- The system requires regular retraining for better accuracy
- The prediction accuracy depends on the quality of available data

14. Future Scope

In future, the system can be improved by:

- Integrating with Learning Management Systems (LMS)
- Developing real-time analytics dashboards
- Applying deep learning models for better accuracy
- Creating mobile applications for easy access
- Testing the model on data from multiple institutions for better validation

Relish Publications,
Amrapali Apartment, Canal Road, Baramati (MS), India.
Pin code- 413102

Website: www.relishpublications.com
Email: raavipublications@gmail.com
WhatsApp: 9665841996

RAAVI
An International Journal of Multidisciplinary Research
ISSN:3108-253X
Special Issue : February,2026

15. Conclusion

This research presents an AI-based predictive analytics framework to improve academic performance in higher education institutions. The comparison of four machine learning algorithms shows that ensemble methods like Random Forest provide higher accuracy and stable results.

The proposed system helps institutions make data-driven decisions, identify weak students early, and take timely corrective actions. Implementing such intelligent systems can significantly improve student success rates and overall institutional performance.

References

1. **Comparative Study of Supervised Algorithms for Prediction of Students' Performance** – International Journal of Modern Education and Computer Science (DT, RF, SVM etc.)
   https://www.mecs-press.org/ijmecs/ijmecs-v13-n1/v13n1-1.html
2. **Application of Classification Algorithms on the Prediction of Student's Academic Performance** – Trends in Sciences (SVM & Random Forest comparison)
   https://tis.wu.ac.th/index.php/tis/article/view/5070
3. **Analysis of Student Academic Performance Using Random Forest and Support Vector Machines** – Computer Science Journal (RF & SVM accuracy comparison)
   https://jurnal.bsi.ac.id/index.php/co-science/article/view/10123
4. **Predictive Analysis of Students' Learning Performance Using Data Mining Techniques** – MDPI Article (Feature selection + ML algorithms)
   https://www.mdpi.com/2571-5577/6/5/86
5. **Data Analysis of Student Academic Performance and Prediction Based on ML Algorithms** – Materials Today Proceedings
   https://www.sciencedirect.com/science/article/abs/pii/S221478532105241X
6. **Student Performance Prediction in Sebelas Maret University Based on the Random Forest Algorithm** – IIETA (Open Access CC BY 4.0)
   https://www.iieta.org/journals/isi/paper/10.18280/isi.270317

Relish Publications,
Amrapali Apartment, Canal Road, Baramati (MS), India.
Pin code- 413102
Website: www.relishpublications.com
Email: raavipublications@gmail.com
WhatsApp: 9665841996

7. **Predictive Analysis of Student Performance using Machine Learning Models** – IRO Journals (RF, SVM, XGBoost)
https://irojournals.com/itdw/article/view/7/1/2

8. **Design of a Prediction Model to Predict Students' Performance Using EDM & ML** – MDPI Engineering Proceedings
https://www.mdpi.com/2673-4591/59/1/25

9. **Educational Data Mining: Prediction of Students' Academic Performance Using ML Algorithms** – SpringerOpen (RF, SVM, LR etc.)
https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z

10. **Machine Learning-Based Predictive Modeling of Student Academic Performance** – International Journal of Intelligent Computing and Information Sciences
https://academic.oup.com/comjnl/article/67/8/2677/7661614 (Random Forest + Naïve Bayes)

❑❑❑

Relish Publications,
Amrapali Apartment, Canal Road, Baramati (MS), India.
Pin code- 413102

Website: www.relishpublications.com
Email: raavipublications@gmail.com
WhatsApp: 9665841996