IJASTR-International Journal of Applied Science and Technical Research
Volume 3, Issue 1 | March-2026
e-ISSN: 3067-0756 Print ISSN: 3067-0748

Learning Analytics Framework for Early Prediction of Student Performance

Rendla. Ramyakrishna[1], Potu Akhila[2], R. Saritha[3]
[123]Assistant Professor, Siddhartha Institute of Technology & Sciences, Narapally, Korremula Road, Ghatkesar, Medchal-Malkajgiri District, Telangana - 500088
[1]Email: ramyanarala3@gmail.com | [2]Email: akhilapotu1998@gmail.com |
[3]Email: saritharamavath902@gmail.com
Received: 24-01-2026; Revised: 04-02-2026; Accepted: 22-02-2026; Published: 10-03-2026

Abstract
The growing use of digital learning platforms has created a vast amount of educational data that can be used to enhance student learning results by making data-driven decisions. It will be necessary to identify those students who might face academic problems as early as possible and implement appropriate interventions that will help them to improve their academic achievement. This research paper is a suggested learning analytics framework that aims at predicting student grades at a very early stage based on learning analytics data that has been gathered through educational metrics of learning management systems and institutional academic records. To examine the relationships between student engagement and learning behavior, the framework incorporates several elements such as the data preprocessing, feature engineering, predictive modeling and model evaluation. Some of the most important learning indicators which are considered as the predictive feature include attendance patterns, behaviour of submitting assignments, interaction with learning material and assessment scores. A number of machine learning models are utilized to group students into various performance groups and those learners who are at risk to perform poorly. As it has been experimentally proven, predictive analytics models are capable of analyzing the behavior patterns of students and making valid predictions of their performance. The suggested model highlights evidence-based educational decision making and allows colleges to take early intervention measures in education designed to enhance retention and learning achievement of students.

Keywords: ***Learning Analytics, Student Performance Prediction, Educational Data Mining, Machine Learning, Predictive Modeling, Academic Intervention.***

1. Introduction

1.1 Background of Learning Analytics in Digital Education

The high growth of online education has drastically changed the manner in which teaching/learning activities are carried out in contemporary learning institutions. The Internet has resulted in online learning systems, learning management systems, virtual classrooms, and digital assessment tools being an integral part of the educational infrastructure of schools and universities. These platforms do not only enable the delivery of course material, it also creates immense amounts of data concerning student interactions, participation rates and academic progress. Such educational data has brought about new possibilities to researchers and educators to analyze learning processes and formulate strategies that would make students achieve better results.

Learning analytics has become an interdisciplinary research area that aims at gathering, measuring, analyzing, and interpreting the data created in the educational setting. Using digital traces of student activity, learning analytics tries to interpret how learners will access course materials, how they will interact with instructors and peers and also how such interactions will affect academic performance. By conducting a systematic study of these data sources, teachers will be able to see how students learn and discover any pattern that could potentially influence the results of education.

Techniques of data mining, artificial intelligence and computational modeling have helped to support the development of learning analytics. Educational data mining methods result in the discovery of useful patterns in complicated data sets, whereas machine learning algorithms permit predictive designs to be built with respect to historical academic data. On the one hand, these technologies permit identifying correlations between the indicators of student engagement, including attendance, patterns of assignment submissions, and the level of online activity, and the ultimate academic achievement.

Learning analytics in digital learning environment is crucial in the enhancement of instruction design and the support systems to students. Learning data analysis will enable teachers to recognize the areas students are experiencing challenges and will modify their teaching approach to address the issue. Moreover, learning analytics tools allow institutions to track the general course performance, analyze the curriculum creation, and enhance the quality of the digital education platform.

The rise in relevance of learning analytics is indicative of a more general change to the use of data-driven educational decision-making. With the ongoing integration of digital learning technologies into the educational institutions, the capacity to read and analyze educational data will be important in enhancing the quality of teaching, students interaction and academic outcomes.

1.2 Reason Wholesomeness of Early Student Performance Prediction.

Early prediction on the academic performance of students in an institution has grown to be a central goal to the educational institutions in a bid to enhance the learning outcomes and the rate of attrition of students. A lot of students have learning problems at school and face certain issues because of their poor involvement in course materials, irregular attendance, no interaction with teachers, or poor ability to grasp complicated material. These problems, when left unnoticed in a long time, may harm the performance of students and general academic development.

The conventional approaches to measuring the student performance tend to be based on the periodic examination or the assessment related to the end of the term. Although these methods do give an account of academic achievement, they usually follow considerable amounts of the learning process. As a result, teachers do not have a lot of room to step in and offer assistance to students facing difficulties in school. Anticipatory student performance prediction is a preventive measure, which can enable an educator to decisively anticipate the possibility of academic risks in a student prior to their development into serious issues.

Predictive analytics applied in education allows educational institutions to process both historical and real-time learning data, to be able to predict future academic performance. Based on the trends in student engagement, participation, and performance in assessments, predictive models can be used to estimate the probability of a student getting to a specific level of performance. The given predictions enable teachers to divide students into high-performing, moderate-performing, or risk types.

There are multiple benefits of early intervention of at-risk students on the educational management. Specific intervention strategies that can be applied in the institutions include personalized tutoring, extra learning aids, mentoring and adaptive learning support. Interventions can be used to assist students to resolve the learning problems and enhance their academic results before they are tested.

Besides, early warning systems promote retention and performance of students and institutions. The student dropout rates at universities and other educational organizations are a problem that can have an impact on the reputation of the institutions as well as the usage of resources. Institutions will be able to establish support systems that can help learners to remain active in the learning process and achieve success by identifying students who might not be engaged in the learning process.

Predictive analytics is an effective means of improving student-centered learning systems in contemporary digital education systems. This capacity to foresee the outcome of the academic performance in relation to the data of learning behavior enables the educators to shift to proactive academic support practices instead of reactive ones.

1.3 Objectives and Scope of Predictive Analytics Framework.

The main aim of the research study is to come up with a structured learning analytics model that will predict the student academic performance at an earlier stage of the learning process. The suggested framework will operate with the use of educational data produced by digital learning settings with the aim of determining trends related to student involvement, participation, and performance. The framework will make predictions that are reliable and will support timely academic interventions by employing machine learning techniques on such datasets..

Figure 1. Architecture of the Learning Analytics Prediction Framework

The integration of several elements of educational data analysis into a single predictive system is one of the primary objectives of the framework. It involves taking data off the learning management systems and academic databases, preprocessing and coding of the data, retrieving useful learning indicators out of the data, and using machine learning algorithms to produce performance predictors. The combination of these elements makes certain that the framework will be able to handle intricate educational data sets and be able to draw significant analytical conclusions.

The other major aim of the research is to determine the learning indicators that are important in determining the academic achievement of the students. The variables that may be included in these indicators include the frequency of attendance, assignment submission, engagement in online discussion, engagement with online course material, and grades. With the help of such indicators, the predictive framework will be able to calculate the best behavior patterns related to student success or risk in their academic performance.

The applicability of the suggested framework is expanded to the educational settings based on the digital learning platform and data-driven teaching systems. Although the framework is especially applicable in the setting of institutions that rely on learning management system, the principles of the framework can also be applied to the context of learning that involves hybrid and blended learning (where online and traditional instructional approaches are intertwined). The structure is also constructed to be flexible to various education scenarios where the institutions can adjust data sources and predictive models to suit their needs.

Additionally, the research will determine the efficacy of the suggested predictive framework by employing the experimental investigation of the machine learning frameworks. Evaluating the accuracy of prediction and the performance of the model, the research aims to prove the idea that learning analytics can help to provide better academic monitoring and decision-making. Finally, the framework is also supposed to help educators, administrators, and policymakers to come up with more responsive and data-driven educational systems that will improve student success and learning outcomes.

2. Learning Indicators and Educational Data Sources on Performance Prediction.

The success of predictive analytics in the field of education highly relies on the quality and accessibility of educational data gathered in digital learning settings. Current educational systems produce various datasets via learning management systems, online assessment, student information systems and also other digital platforms adopted in the learning process. These data sources have noteworthy information regarding the actions of students, patterns of engagement, and student-performance and can be assessed to discover indicators that are pointers to a successful or a struggling student. Through the systematic gathering and examination of these data sets, learning analytics frameworks can be used to draw insightful conclusions to assist in early prediction of student performance.

There are many different types of educational data sources and learning indicators that are typically employed in predictive analytics systems:

- The attendance records will indicate the presence of the student in classes or virtual classes that they attend.
- The patterns of assignment submission, such as the time and the frequency of submission.
- Online log of interaction including participation in discussions forums and communication with the instructors.
- Evaluation grades, such as quiz, intermediate assessment, and project assessment.
- The use of learning resources, which implies the number of times the students use course materials or multimedia objects.
- Activity timestamps which indicate learning platform patterns of engagement over time.

Combination of these learning indicators facilitates predictive models to build sets of features that express student engagement, academic commitment and learning advancement. These characteristics are what underlie machine learning models that categorize students based on the amount of academic performance they are predicted to have.

2.1 Learning Management System Data and Academic Records

Learning management systems have been integrated to play key roles in the contemporary learning institutions, especially in institutions that run online and blended educational programs. These systems help in the dispensing of learning resources, submission of assignments, testing and communication between the instructors and learners. When students use these platforms, the system automatically logs in-depth information about activity, creating an endless flow of learning information that can be utilized in learning analytics.

The information that is usually stored within LMS platforms would include the enrollment in a course, how many times someone was able to log in and spend their time watching course materials and taking part in discussion board, as well as completing assignments or quizzes. Such activity logs give good information on the way students interact with course materials and the level of their consistent engagement in learning activities. Individually, when students actively engage and use course materials regularly, a significant level of engagement and a better level of academic performance is likely to be portrayed.

Along with the data produced by LMS, academic records, retained in institutional databases offer an organized information on student performance. Such records consist of the past grades, cumulative study performance, the history of course completion, and demographic data that can be applied to the educational analysis. The academic records are also useful reference points to the predictive models as they indicate the historical performance trends which could impact on the outcomes in future in terms of academic performance.

By merging LMS interaction statistics with other academic data, learning analytics systems can be used to achieve both behavioral and performance indicators of student learning. Behavioral data show the way students relate with learning materials, and the academic records show their quantifiable academic results. A combination of the datasets offers a multifaceted picture of the student learning patterns that can be useful in predictive modeling.

LMS information integration with academic records is thus critical in developing predictive models that are able to identify students who can become potential victims of academic under-achievement. The simultaneous analysis of such data sources by machine learning models enables the recognition of even the slightest associations between learning behaviour and academic performance, and this allows institutions to establish early intervention strategies.

2.2 Online Learning Behavioral and Engagement Indicators.

The student behavior and interaction in the digital learning settings have a significant role in the academic success. Online learning systems unlike traditional classroom environments enable teachers to track different types of student engagement with the learning materials and communication devices. These interactions give quantifiable signs that indicate the levels of student engagement, devotion and learning activity. Learning analytics systems can use behavioral data to determine patterns in both effective and possible academic challenges.

Indicators of engagement are activities that can be seen to be carried out by students when they are using digital learning environments. These indicators help shed some light on the number of times that students access learning materials, take part in group discussions, and do the assigned tasks. The intensity of participation usually has a positive relationship with academic achievements, and the lower the participation is, the more likely that there is a possibility of poor learning.

We do not expect to analyze behavioral and engagement indicators such as:

- Frequency of logins to the platform, which reflects the frequency with which students use the learning management system.
- Enhanced time spent on course material which denotes the amount of interaction with lectures, readings and multimedia material.
- Online discussion forums, collaborative learning and communication with peers and instructors.
- The behavior of assignment completion such as submission on time and the ratio of submission.
- Intervention with quizzes and self-marking tools, which proves the active engagement in the processes of learning evaluation.
- Navigation patterns in course modules, which indicate the way students go through organized learning materials.

able 1. Learning Indicators Used for Student Performance Prediction

Indicator Category	Learning Indicator	Description
Attendance	Class/Session Attendance	Participation in scheduled lectures or online sessions
Assignment Activity	Assignment Submission Rate	Timely completion and submission of assignments
LMS Engagement	Login Frequency	Number of times a student accesses the LMS platform
Content Interaction	Learning Material Access	Frequency of viewing lecture notes, videos, and resources
Assessment	Quiz and Test Scores	Performance in quizzes, tests, and internal assessments
Collaboration	Discussion Forum Participation	Interaction with peers and instructors in forums
Behavioral Activity	Time Spent on Platform	Duration of engagement with the learning system

The predictive models can use the engagement indicators to know whether the students are engaged in learning processes or showing disengagement patterns. Timely disengagement enables teachers to offer special attention to students in academic matters and motivate them to get back to the learning activity.

2.3 Student analytics Data Integration and Feature Representation.

Volume 3, Issue 1 | March-2026
e-ISSN: 3067-0756 Print ISSN: 3067-0748

The educational data in the learning analytics can be obtained as a result of various independent systems, such as learning management systems, institutional databases, assessment systems, and communication tools. Since the systems store data in various formats and forms, successful predictive modeling entails the pooling of such data into one analytical system. Data integration guarantees the possibility to combine and analyze educational indicators, which are relevant, and consist of numerous sources.

The first phase of the process of integration is usually the identification of the data sources of interest and the acquisition of instructional data related on a student-to-student basis. The data should then be cleaned and standardized to eliminate inconsistency, duplicate data and incompleteness. Special algorithms used in data preprocessing like normalization of data, encoding of categorical data variables, and missing values are undergone to prepare the data to be subjected to the data analysis process.

The feature representation process is the subsequent one after preprocessing since raw educational data is converted into meaningful variables which can serve as input to machine learning algorithms. Feature representation transforms multifaceted behavioral and academic data into structured number or categorical features that reflect some significant element of student learning behavior. To illustrate, a frequency of a login can be defined as a numerical variable, which can be the number of times a person accessed their account weekly, whereas assignment submission patterns can be defined as binary variables indicating the status of completion.

Additional indicators can also be derived by feature engineering methods on the available data. As an example, the engagement scores can be obtained by summing several behavioral variables, or time-related attributes can be created to observe the learning patterns in various weeks of a course. These constructed capabilities improve the predictive power of machine learning models because they include more informative depictions of student behavior.

The success of predictive learning analytics frameworks highly depends on effective data integration and representation of the features. In cases where learning information has been appropriately formatted and converted into meaningful attributes, machine learning algorithms can more readily determine the correlation between learning behavior and the results of learning.

2.4 Educational Data Quality/Availability Problems.

Although the digital data of education is becoming more and more available, a number of issues regarding the quality and access of certain data can influence the functioning of predictive learning analytics systems. The most prevalent problem is that educational datasets may contain missing or incomplete data. Students can also engage with learning platforms in ad hoc fashion, creating two loopholes in the data of activity records. Missing values may decrease the predictive model accuracy and make data analysis more difficult.

Another threat is the discrepancy in the data obtained in various systems of education. The different data formats, nomenclature, or storage formats used in learning management system, institutional database, and assessment platform may differ depending on the platform. Such discrepancies can render the process of integrating datasets challenging and can demand a significant amount of preprocessing before the process of analysis can be performed.

Educational data analytics also face significant issues in the field of data privacy and ethical concerns. Student records are sensitive data that holds personal and academic information which should be secured in order to maintain confidentiality and institutional policies. Learning institutions should put policies of proper data governance in place, which would govern the data collection, storage, and analysis of student information. Predictive analytics systems should thus have relevant security features and codes of ethics to avoid abuse of student data.

Besides privacy risks, the predictive model reliability might also be impacted by data availability constraints. Other institutions do not necessarily have extensive digital data of student behaviors especially in learning settings where conventional classroom instructional approaches continue to dominate. In this scenario, they may be limited by the lack of detailed information about behaviors, which may limit the power of predictive analytics frameworks.

To overcome these difficulties, it is necessary to introduce effective data management practices, data collection standardization, and ethical principles of educational analytics. Institutions can boost the quality of their data and make sure they use it responsibly to achieve greater reliability and effectiveness of predictive learning analytics systems to improve the success of students.

3. Design of the Proposed Learning Analytics Prediction Framework.

The proposed learning analytics prediction framework architecture is aimed at processing the educational data systematically and producing the premature prediction of the academic performance of the students. In the contemporary digital learning world, data is continuously generated in high amounts via the learning management systems, evaluation platforms and institutional databases. Nevertheless, raw educational data cannot be directly used to back up predictive analysis unless it is properly prepared and organized. As such, a systematic system is necessary in order to change the uncoded data into valuable information that can guide the teachers in finding students who might be in need of some academic help.

The suggested framework will include several interacting parts which jointly will allow predicting the performance of the students. Such components are educational data collection, data preprocessing, feature engineering, predictive modeling and model evaluation. The individual components, in turn, complete a particular task in the system and make sure that the data on the education gathered are organized to undergo an analytical processing. Combining these elements into a single structure, the system is capable of effectively evaluating trends in the student engagement and academic performance indicators.

The architecture focuses on scalability and flexibility as well, meaning that the institutions can add more sources of data and machine learning models as educational technologies change. This flexibility makes sure that the framework can apply in a variety of educational environments both in an online classroom environment, a blended learning system, and digitally mediated classroom instruction. The proposed framework enables the identification of the at-risk students at an early stage and allows taking proactive measures in the academic field by offering a systematic structure over educational data analysis and predictive modeling.

3.1 System Workflow and Framework Components.

The working process of the suggested learning analytics system comprises a logical order of steps transforming raw education data into forecasting information. All steps of the working process are directed to the final goal of recognizing the patterns of student behavior and its predictive behavioral future.

Step 1: Data Collection Educational.

Student learning activity related data are gathered through various sources like the learning management systems, academic databases, online assessment systems and institutional student information systems. These data comprise the engagement information, the attendance details, the submission of assignments, and grades.

Step 2: Analysis and Data Storage.

The gathered data of various systems are merged into an analytical environment central place. Integration of the data enables the alignment of the records of the students in different platforms and the data is stored in a standard format that can be analyzed.

Step 3: Data Preprocessing

Raw data is usually full of inconsistencies, blank fields and duplications. The preprocessing phase includes cleaning and standardizing of the data so that accuracy and reliability can be guaranteed and then analytics processing can take place.

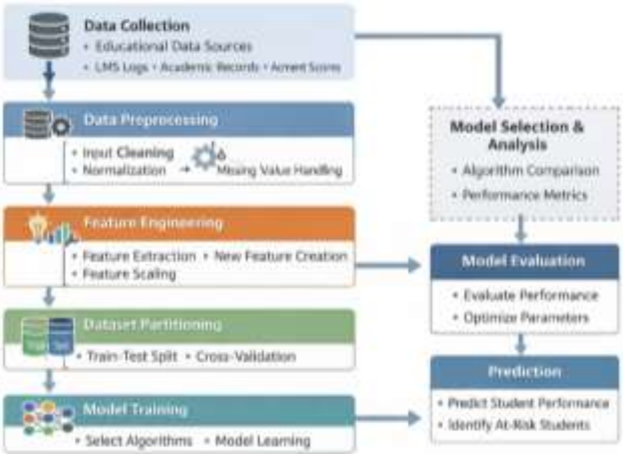


Figure 2. Workflow of Data Processing and Predictive Modeling

Step 4: Feature Extraction and Engineering.

The cleaned datasets are used to extract relevant learning indicators that would capture the student engagement and academic behavior. Other features can also be obtained by combining the existent variables to form more analytical indicators.

Step 5: Training of the machine learning model.

The dataset that has been prepared is sent to machine learning algorithms to discover learning indicators/academic performance relationships. The historical educational data is used to train the models.

Step 6: Prediction and student classification of performance.

The trained models produce predictions on the level of performance of the students and categorize them into high-performing students, average-performing students, and at-risk students.

Step 7: Model Evaluation and validation.
The predictive models are measured in terms of statistic measures so as to ascertain the accuracy and reliability of the models. The results of the evaluation can be used to assemble whether the model can be applicable in the practice of educational settings.

3.2 Preprocessing and Feature Dusting algorithms.

Educational data sets must be properly pre-processed and transformed before they can be used in predictive modeling. The framework proposed uses a number of preprocessing and feature engineering methodologies to make sure that the input data is fit to be analyzed by the machine learning.

1. **Data Cleaning and Elimination of Noise.**
   Raw data of education usually contains incomplete records, duplicate records or values which are not consistent. The data cleaning procedures are used to delete the irrelevant records, as well as to address inconsistencies in the data set.
2. **Handling Missing Values**
   The lack of data is another typical problem of educational data. Missing values are then dealt with through various methods like data imputation, mean replacement, or among others by eliminating incomplete records.
3. **Normalization and Transformation of Data.**
   The numerical data obtained in various sources can vary in terms of scales and distributions. These variables are transformed to standard ranges through normalization methods to be used in the machine learning algorithms.
4. **Categorical Data Encoding**
   Some education variables like types of courses or student composition can be displayed in form of categorical values. These categorical variables are encoded to a numerical format which can be analyzed using predicative models.
5. **Indicators Learning Feature Extractions.**
   Raw datasets are extracted to construct meaningful analytical variables, which are relevant variables representing student engagement, participation and academic performance.
6. **Derived Feature Generation**
   The additional features can be created by multiplying the variables that are already present including engagement scores, which are determined by the rate of logins and the rate of completing an assignment.

The preprocessing and feature engineering methods guarantee that the data that is used to make predictive models properly captures student learning behavior and academic performance.

3.3 Student Classification Pipeline based on Predictive Modelling.

The predictive modeling pipeline in the proposed structure is the one that analyses processed educational data and makes predictions about the academic performance of the students. The pipeline consists of a series of steps that are used to systematically prepare data to be used by machine learning algorithms and produce classification results.

Step 1: Dataset Partitioning
The processed data is further split into training and testing data. Predictive models are developed using the training dataset and the model performance is assessed using the testing dataset.

Step 2: Predictive Algorithms Extraction
The selection of the most suitable machine learning algorithms to perform the classification task is based on the appropriateness of the algorithms in analyzing the educational data sets.

Step 3: Model Training
The chosen machine learning algorithms are trained on the training dataset. In this step, the model is taught patterns through which the indicators of student learning are matched with the outcomes of academic performance.

Step 4: Model Validation
To make sure that the trained models work well on new data, validation procedures, including cross-validation, are applied.

Step 5: Prediction of performance.
The labeled model is used on the test data to provide predictions of the level of student performance.

Step 6: Classification of the students.
Students are based on their performance categories depending on the expected performance level, which includes high-performing, moderate-performing, or at-risk students.

3.4 Machine Learning Model to be incorporated in Learning Analytics Systems.

Integration of learning analytics systems based on machine learning models is a major breakthrough in the capability of learning institutions to analyze student learning behaviour and forecast their academic performance. The predictive frameworks have been made possible by the use of machine learning algorithms that determine intricate association between various learning indicators and student performance. In contrast to the classical statistical methods of analysis, machine learning models can analyze extensive and multidimensional data, and they are therefore well adapted to analyze educational data produced by digital learning environments.

Machine learning models are also commonly integrated into analytics platforms in learning analytics systems, which gather and process educational data at any given time. With the availability of new data, predictive models can be pruned and retrained so that their accuracy and relevance do not diminish. This dynamic integration enables the institutions to track student activity and academic success in almost real time giving the educators real-time insights into the student performance patterns.

The other benefit of implementing machine learning models in learning analytics systems is that it enables personalized educational interventions. As soon as predictive models can detect learners who might be exposed to low academic achievements, teachers can apply specific supportive practices including extra tutoring lessons, adaptive learning resources, or individualized feedback. These interventions have a great potential to enhance student achievement and involvement.

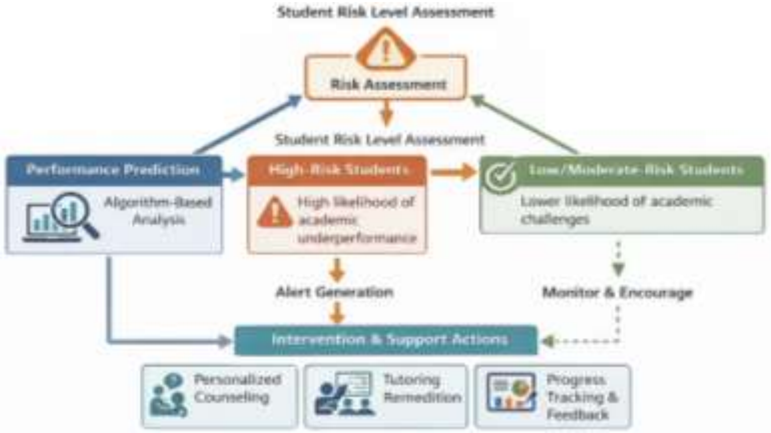


Diagram 1. Early Intervention Strategy Based on Prediction Outcomes

Moreover, learning analytics systems based on machine learning are valuable to the institutional decision-making process that can help administrators to get aggregated information about course effectiveness, student retention trends, and course performance. Through such patterns, the institutions will be able to optimize their teaching methods, better design their courses as well as formulate data-driven policies that can help improve their education results. By means of efficient employment of machine learning technologies, learning analytics systems can make learning systems responsive and student-oriented.

4. Evaluation Strategy Experimental Design and Model.

The proposed learning analytics framework is based on the experimental design, which is concerned with the effectiveness of the machine learning models to predict the academic performance of the student at the initial stage of the learning process. An effective experimental plan is necessary in predictive analytics studies so that the models developed by the researcher can come up with stable and generalized outcomes. The research design taken in the present study will analyze the educational data systematically, train predictive algorithms, and evaluate them based on adequate measures of evaluation.

The experimental design starts with the pre-preparation of the datasets based on learning management systems and academic records. Among the learning indicators present in such datasets, one can distinguish the student engagement patterns, the rates of assignment completion, attendance, and the assessment outcomes. Predictive algorithms are applied to the data after preprocessing and feature engineering of the data have been implemented so that the variables can reflect the behavior of the students in the context of learning. The pre-training step is an important step to enhance predictive accuracy of machine learning models.

Volume 3, Issue 1 | March-2026
e-ISSN: 3067-0756 Print ISSN: 3067-0748

It is in this experimental design that the proposed learning analytics framework is able to show how educational data can be systematically processed to produce meaningful predictions that can be used in academic decision-making and early intervention strategies.

4.1 Preparation of the Dataset and Experimental Design.

The proposed predictive framework has several organized steps to be followed in its experimental setup that will assure that the dataset is appropriate to be analyzed by machine learning and that the predictive models come in an effective way.

Step 1: Data Collection

Learning management systems and institutional academic databases are the sources of educational data. The data set contains the indicators of the frequency of logins, submission behaviors of assignments, attendance patterns, or score on assessments.

Step 2: Preprocessing of Data and Cleaning.

The obtained data is analyzed to determine the presence of missing data, inconsistencies, and duplicate data. Data preprocessing is used to salvage data and make it ready to be processed analytically.

Step 3, Feature Selection and Engineering.

The dataset is used to select relevant learning indicators that are used as input features to predictive models. Other derived items can be created to indicate the level of engagement and schools behavioural pattern.

Step 4: Dataset Partitioning

The ready data set is separated into training and testing sets. Machine learning algorithms are trained using training dataset and tested using testing data.

Step 5: Model Training

The training data is used to train selected algorithms in machine learning. In this phase, the models get to know about association between indicators of student engagement and academic performance.

Step 6: Model Testing and Model validation.

The trained models are used in the testing dataset to make the predictions. The validation techniques are applied to estimate the effectiveness of the models when applied to new data.

4.2 Selection of Machine Learning Algorithms for Prediction

To accurately predict student academic performance, the proposed framework employs multiple machine learning algorithms that are commonly used in classification problems. Each algorithm has unique strengths in analyzing complex datasets and identifying patterns associated with educational outcomes.

Decision Tree Classifier

Decision tree algorithms develop hierarchical information on decision that categorize students using learning indicators. They come in handy especially in the interpretation of the relationships between variables of student behavior and performance outcomes.

Random Forest Algorithm

Random forest is a method of ensemble learning, a combination of decision trees, aimed at increasing the accuracy of prediction and decreasing overfitting. It is effective in the analysis of data that has many behavioral variables.

Support Vector Machine (SVM)

The support vector machines can be useful when dealing with a high-dimensional dataset in a classification task. SVM models define the best decision boundaries which categorize students into various performance groups.

Logistic Regression

Logistic regression Logistic regression is extensively employed in binary or multi-class classification predictive analytics. It predicts the likelihood of a student that belongs to given performance category given the input variables.

K-Nearest Neighbor (KNN)

KNN algorithm performs classification of the students according to the resemblance of their patterns of learning behavior with the students in the dataset.

Through the comparison of the predictive power of these algorithms, the study determines the models that give the best predictions to early identify the at-risk students.

4.3 Evaluation Measures of student performance prediction models.

Predictive models performance will have to be assessed with the help of relevant statistical measures to identify its efficiency in student performance classification. These metrics are quantitative estimates of model accuracy, reliability and predictive power.

The typical metrics of evaluation that would be employed in the proposed framework are:

- Accuracy is a measure that determines the frequency of successful prediction of student outcomes on performance performance with regards to the number of predictions.
- Precision that measures how the model is able to correctly identify the students who fall in a given performance category without high false predictions.
- Recall (Sensitivity) This is the extent to which the model correctly identifies students who are actually at risk of poor performance in school.
- F1-Score, which is a harmonic average of both the precision and recall and offers a trade-off measure of the classification performance.
- Confusion Matrix Analysis which gives a closer comparison of the predicted and actual classification results and this allows the behavior of the model prediction to be analyzed with greater depth.

Table 2. Evaluation Metrics and Model Accuracy Results

Machine Learning Model	Accuracy	Precision	Recall	F1-Score
Decision Tree	0.70	0.67	0.69	0.68
Random Forest	0.82	0.79	0.80	0.81
Support Vector Machine	0.71	0.64	0.69	0.68
Logistic Regression	0.62	0.58	0.57	0.59
K-Nearest Neighbors	0.54	0.49	0.48	0.49

Such measures of evaluation allow scientists to compare machine learners in a structured manner and conclude on the most reliable predictive model that can perform with respect to identifying students who could need early academic support.

5. Results

The experimental analysis of the offered learning analytics framework proves the usefulness of machine learning methods in anticipating the studying performance of students at an initial point in the learning process. The predictive models used could be able to determine patterns based on student engagement and performance outcomes through the analysis of several education indicators based on learning management system and academic records. These findings indicate that machine learning algorithms can be used to categorize students by performance and identify learners that need academic support.

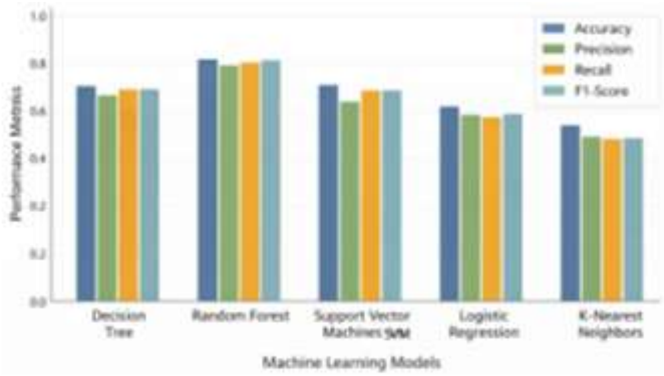
In the analysis, a number of machine learning models were trained and tested with educational datasets with behavioral and academic indicators. The evaluation of the models was based on standard classification measures to establish their predictive accuracy and reliability. The findings show that predictive analytics methods have the potential of analyzing student learning behavior and give valuable insights on the trends in academic performance.

5.1 Comparison of predictive and model performance.

The objective comparison of the various machine learning algorithms showed that there were differences in prediction and classification precision when used on the educational dataset. All the algorithms showed varying abilities in the pattern analysis of student engagement variables and academic outcomes. Ensemble methods and algorithms, which are able to deal with complex datasets, tended to make more credible forecasts.

Key notes made in the model performance comparison are:

- The predictive accuracy of the ensemble-based algorithms was better because they make use of more than one decision model and minimize the amount of prediction error.
- Algorithms which have been able to efficiently manage high-dimensional data sets provided more consistent classification outcomes on both training and testing data sets.
- More basic classification models were faster to compute but did not always have as good a predictive accuracy as the ensemble methods.
- Models that used various measures of engagement performed better compared to models that used few academic variables.



Graph 1. Comparative Performance of Machine Learning Models

This shows that the effectiveness of predictive learning analytics systems is dependent on the selection of the machine learning algorithm.

5.2 Significant Learning Indicators Analysis that has an impact on performance.

The evaluation of learning indicators in the experimental dataset showed that behaviors of student engagement are important in the prediction of the academic results. Learning management system based behavioral indicators were useful in learning how the students engage with digital learning environments. These indicators were used to make predictions on the patterns that were related with the successful and failing learners.

The patterns of assignments submission were among the most significant variables that influenced the performance of students analyzed. The students who were able to do assignments on time on a regular basis tended to show better academic results than those with inconsistent submission frequencies. This conduct shows how a student is determined to take the requirements of the course as well as the extent to which she is interested in instructional materials.

There was also a strong correlation of attendance patterns with academic performance. Learning outcomes were more favorable among students who attended scheduled learning activities as per the online session or through lectures that were recorded. Regular attendance means that the students are actively engaged in the learning process; it enables them to keep track of the course progress.

The number of times interacting with learning materials like lecture materials, reading materials as well as the multimedia resources was another important indicator. Those students that used these resources more often were more likely to show more interest in course materials that made a positive impact on their performance.

Communication in the learning collaborative environments such as discussion rooms and communication applications, also contributed to performance prediction significantly. Students who enthusiastically took part in discussions tended to enjoy the peer learning and instructor feedback which helped them to gain knowledge about course concepts.

The joint analysis of these indicators will allow concluding that academic performance depends not only on the scores in assessment but also on the trends of interaction in the digital learning environment. Predictive learning analytics systems can learn more about the behavior of student learning by the analysis of these behavioral indicators.

5.3 Predictive Modeling: Student identification of At-Risk Students.

Among the key aims of the suggested learning analytics framework, it is possible to identify the students who are prone to poor academic outcomes early. Predictive modelling methods enable the framework to examine behavioral and academic clues so as to categorise the students based on their chances of performing to certain levels. Through analysis of these patterns, the predictive models will be able to identify early warning signs that can be used to signify academic problems.

The outcomes of the conducted experiments prove that machine learning models are capable of effectively detecting students with signs of lack of engagement or irregular learning behavior. Such indicators as the lack of logins within a certain timeframe, odd submission frequency of assignments, minimal engagement with the course content, and decreasing assessment scores were often linked with students who were to face academic difficulties subsequently. When identified at an early stage, predictive systems offer educators with good chances to intervene before the academic issues get out of hand.

The early identification of the at-risk students helps education institutions to adopt specific support strategies. Such strategies can consist of special academic counseling, more tutoring time, adaptive learning materials, or direct instructor feedback intended to help a learner overcome particular learning challenges. These interventions may play a major part in the enhancement of student participation and assist the learners in gaining traction in their academic studies.

Moreover, proactive identification of high-risk students facilitates institutional intervention to enhance the retention of students and the overall performance of the education. By taking the initiative to resolve the possible obstacles to learning, the institutions can establish more conducive learning environments that can foster the growth of the students as well as low turnover rates.

The findings of the present research prove that forecasting analytics systems have high-level potential to help teachers track the progress of students and define those who need to receive academic help early. Learning analytics systems can be significant in the improvement of the quality and responsiveness of modern education settings with the help of the integration of educational data mining and machine learning techniques..

6. Conclusion

The proliferation of online learning space has provided an unprecedented chance of using educational data to enhance the student learning outcomes. With the increased use of learning management systems and the internet, there are vast volumes of student interaction data produced in the course of the learning process. This data can be useful in understanding student engagement, patterns of participation and progress, when analyzed successfully. The current paper discussed the implementation of learning analytics alongside machine learning methods to create a prescriptive model that can be used to recognize patterns in student performance at an early phase.

The study was aimed at creating an organized learning analytics prediction model based on a combination of educational data, preprocessing, machine learning algorithms, and model evaluation plans. The proposed framework provides an analysis of the use of predictive analytics in identifying students who might struggle with academic problems before final evaluations assess them. The experimental study established that machine learning models are capable of successfully categorizing students according to their engagement patterns and learning behaviors, which gives an opportunity to monitor academic performance in a proactive manner.

6.1 Overview of the Learning Analytics Framework Proposal.

The suggested learning analytics framework discussed in this paper offers a methodological framework of forecasting student academic achievement based on the educational information produced within online learning settings. The framework combines various analytical units such as the collection of educational data, preprocessing of data, feature engineering, predictive modeling, and evaluation of the model. These elements are collaborative in the way they convert raw educational information into insights that could be acted upon to help identify early student performance trends.

The framework starts by gathering information based on the learning management systems and the academic records of the institutions. These data sets include different indicators which are indicators of student engagement, learning behavior and academic achievement. When all collected, a preprocessing is done to eliminate inconsistencies, missing values and standardization of variables. The feature engineering methods are subsequently used to obtain significant indicators that reflect student contact with online learning materials.

Following the stage of data preparation, machine learning algorithms are used to examine the correlations between the learning indicators and the scholastic performance. The historical education data are used to train the predictive models which are then assessed with the help of statistical metrics in order to establish their predictive accuracy. Under this process, the framework can categorize the students into various performance groups, and those students who can be in danger of not performing well academically.

The fact that it is flexible and adaptive to various education settings is an important feature of the framework. The architecture is applicable to many digital learning settings and academic institutions since it can support different data sources and machine learning algorithms. This flexibility is necessary due to the fact that the framework is always relevant since educational technologies are constantly changing.

The general framework construction illustrates the way in which learning analytics may be utilized to convert the informational resources related to education into the predictive ones that would help teachers to track the progress of students and conduct the academic interventions in a timely manner.

6.2 Educational Implications of the Research in Terms of Decision-Making and Intervention Strategies.

The adoption of predictive learning analytics models bears considerable implication on the areas of educational decision-making and management of academic institutions. Conventional ways of assessing student achievement are usually based on examinations or end assessments, which might be administered at a very late time when the educator might be unable to provide effective support to the struggling learners. Comparatively, predictive analytics will allow institutions to recognize issues that might arise in academics much earlier into the learning process.

Early prediction of student performance will enable the educators to to come with proactive intervention measures to overcome the learning difficulties before they progress to serious levels. Upon the detection of students to be at risk of fulfilling the predictive models, the instructors and academic advisors can take specific help actions to assist learners through personalized tutoring, adaptive learning resources, extra practice activities, or mentoring. Such interventions are able to assist the students to develop their knowledge of the course concepts and their engagement in the learning activities.

Besides helping individual students, predictive analytics is also used in larger institutional decision-making. Educational administrators are able to use aggregated predictive data to determine patterns on course effectiveness, curriculum design and patterns of student retention. This knowledge can be used in strategic planning programs that will enhance the level of teaching methods, maximization of learning materials and overall quality of education.

The other significant implication of learning analytics frameworks is the advancement of data-oriented educational management. Through predictive analytics, educational institutions will be in a position to make better decisions using objective evidence gathered in case of student learning data. Such a strategy contributes to the creation of more adaptive and responsive learning contexts that are more responsive to the needs of students of different populations.

Finally, predictive learning analytics may become a key to changing the functioning of educational systems as it allows providing the early academic intervention, enhancing the student engagement, and making the educational policies more efficient. With the further development of digital learning technologies, the role of data-driven analytical frameworks integration in the future of education and improving the performance of students will gain more significance.

Acknowledgement: Nil

Conflicts of interest

The authors have no conflicts of interest to declare

References

1. Romero C, Ventura S. Educational data mining: A review of the state of the art. IEEE Trans Syst Man Cybern C. 2010;40(6):601–618.
2. Siemens G, Baker RS. Learning analytics and educational data mining: Towards communication and collaboration. Proc Int Conf Learn Anal Knowl. 2012;252–254.
3. Ferguson R. Learning analytics: Drivers, developments and challenges. Int J Technol Enhanc Learn. 2012;4(5–6):304–317.
4. Baker RSJd, Inventado PS. Educational data mining and learning analytics. Handb Learn Anal. 2014;61–75.
5. Arnold KE, Pistilli MD. Course signals at Purdue: Using learning analytics to increase student success. Proc Int Conf Learn Anal Knowl. 2012;267–270.
6. Macfadyen LP, Dawson S. Mining LMS data to develop an early warning system for educators. Comput Educ. 2010;54(2):588–599.
7. Tempelaar DT, Rienties B, Giesbers B. In search for the most informative data for feedback generation: Learning analytics in blended learning. Comput Hum Behav. 2015;47:157–167.
8. Kotsiantis SB, Pierrakeas CJ, Pintelas PE. Predicting students' performance in distance learning using machine learning techniques. Appl Artif Intell. 2004;18(5):411–426.
9. Gray G, McGuinness C, Owende P. An application of classification models to predict learner progression. Comput Educ. 2014;71:93–103.
10. Lakkaraju H, Aguiar E, Shan C, et al. A machine learning framework to identify students at risk of adverse academic outcomes. Proc ACM SIGKDD Int Conf Knowl Discov Data Min. 2015;1909–1918.
11. Al-Shabandar R, Hussain AJ, Keight R, et al. Machine learning approaches to predict learning outcomes in higher education. Expert Syst Appl. 2018;102:67–77.
12. Sweeney M, Lester J, Rangwala H. Next-term student performance prediction: A recommender systems approach. J Educ Data Min. 2016;8(1):22–51.
13. Shahiri AM, Husain W, Rashid NA. A review on predicting student performance using data mining techniques. Procedia Comput Sci. 2015;72:414–422.
14. Costa EB, Fonseca B, Santana MA, et al. Evaluating the effectiveness of educational data mining techniques for early prediction of student performance. Comput Educ. 2017;110:1–15.
15. Zhang Y, Rangwala H. Early identification of at-risk students using machine learning techniques. IEEE Trans Learn Technol. 2018;11(2):254–267.