A NOVEL MODEL FOR PREDICTING STUDENT ACADEMIC PERFORMANCE IN HIGHER EDUCATION

Thesis submitted for the award of the degree of

DOCTOR OF PHILOSOPHY

in

Computer Applications

By

Harjinder Kaur

(41700229)

Supervised by

Dr. Tarandeep Kaur (24836)

Department of Computer Applications (Assistant Professor)

Lovely Professional University

LOVELY PROFESSIONAL UNIVERSITY

PUNJAB

2024

DECLARATION

I, hereby declare that the presented work in the thesis entitled "A Novel Model For Predicting Student Academic Performance In Higher Education " in fulfillment of my degree of **Doctor of Philosophy (Ph. D.)** is the outcome of research work carried out by me under the supervision of Dr. Tarandeep Kaur, working as Assistant Professor, in the School of Computer Applications of Lovely Professional University, Punjab, India. In keeping with the general practice of reporting scientific observations, due acknowledgments have been made whenever the work described here has been based on the findings of another investigator. This work has not been submitted in part or full to any other University or Institute for the award of any degree.

(Signature of Scholar)

Name of the scholar: Harjinder Kaur

Registration No.: 41700229

Department/school: School of Computer Applications

Lovely Professional University,

Punjab, India

CERTIFICATE

This is to certify that the work reported in the Ph. D. thesis entitled “A Novel Model For Predicting Student Academic Performance In Higher Education” submitted in fulfillment of the requirement for the award of degree of **Doctor of Philosophy (Ph.D.)** in the School of Computer Applications, is a research work carried out by Harjinder Kaur, 41700229, is bonafide record of his/her original work carried out under my supervision and that no part of thesis has been submitted for any other degree, diploma or equivalent course.

(Signature of Supervisor)

Name of supervisor: Dr.Tarandeep Kaur

Designation: Assistant Professor

Department/school: School of Computer Applications

Lovely Professional University,

Punjab, India

ABSTRACT

The problem that faces the educational system is how to distinguish between students who are having difficulty and those who are performing exceptionally well. If teachers don't have a dependable and accurate method of recognizing these students, it will be difficult for them to provide them with the tailored interventions and support they need to be successful. There is a possibility that conventional approaches to evaluating student performance data will not be able to comprehend the intricate patterns and interrelationships that can exist between the various characteristics.

Academic Data Mining (ADM) impacts many educational institutions, significantly, playing a prime role in accumulating, studying, and analyzing academic data. The accumulated academic data can be processed and analyzed for various purposes. It can be used for predicting student academic performance and thereby broadening the retention rate of academic institutions. The prediction of students' academic performance at the initial stage helps the students to identify their lacking subjects such that they can focus more on their deficient subjects and improvise their academic performance. Currently, numerous machine learning techniques are being used by academic institutions to extract, analyze, and predict student's academic performance and identify fast and slow learners. In the present educational system, one of the most important tasks is to determine which students are succeeding academically and which students require additional assistance. In the past, instructors have traditionally depended on personal observations and evaluations to make these distinctions; however, these techniques have been time-demanding and are susceptible to individual interpretation. Models that are built on machine learning provide a solution that is both effective and objective to this challenge. The suggested model, which is based on machine learning, offers a solution that is both objective and effective for determining which students are powerful and which are poor. Because of the precision of the model and its ability to analyze the significance of individual features, it was a helpful instrument for educators and educational organizations that are working to improve the performance of their students. It would

be beneficial for future research to investigate both the flexibility of the model and its applicability in a variety of instructional contexts.

Even though research on EDM in the educational sector is already available, the main objective of this research work is to examine and identify useful rules and patterns to motivate students to manage their educational performance and careers in a good manner so that they can complete their degree on time. It also aids the academic institutions to improve the academic strategies and pedagogies to benefit the students. The analysis of students' performance in educational institutions reveals the extent of the efforts that have been made by those institutions to improve the learning of slow or average students. The benefit of employing EDM models is that they use student historical data to predict their performance that has not yet occurred. Several academic institutions have been motivated by this concept to create classification models that predict the unknown labels of subsequent situations. Most academic institutions and researchers began to become interested in the field of student performance prediction to categorize students into slow and fast learners based on their predicted academic performance. Although the educational sector employs a variety of methods for extracting academic data of students it is necessary to develop a model for student performance assessment. The results of the prediction model assist the students and instructors in improving their performance to the next level.

This research aims to find a solution to the issue by developing a machine learning (ML) model that is capable of accurately distinguishing between students with low and high potential by employing a multiparametric analysis. The algorithm will be educated using a sizable collection of student performance data, which will include information on scholastic accomplishment, punctuality, and behavior. Because it will be able to analyze numerous characteristics at the same time, the model will be able to recognize intricate patterns and relationships that may be difficult for humans to comprehend.

In this thesis, we proposed two models for determining which students will be academically successful based on the usage of machine learning algorithms. The suggested model makes use of a variety of learning strategies, including controlled as well as unstructured learning. While unstructured learning techniques are utilized to recognize patterns and trends within the data, the supervised learning algorithm is utilized in order to categorize students into two distinct groups, namely, those students who are academically strong (fast learners) and those students who are weak in academics (Slow learners).

Machine learning algorithms and education data mining are extensively used in this research which helps in the collection of academic data that contains information pertaining to students, such as their attendance, scores, and participation. The model is educated on a portion of the information before being tested through the application of cross-validation strategies. The model can distinguish between individuals with low and high potential with an accuracy of 93.74%.

In addition to this, the model that has been suggested offers insights into the variables that contribute to the scholastic achievement of a student. The feature significance analysis of the model reveals that a student's level of participation and punctuality in class are the two most important variables in determining that student's overall scholastic achievement. Teachers and educational organizations can make use of this information to create interventions that specifically target these areas.

Furthermore, the model that has been suggested, offers a system that allows for the ongoing evaluation of the development of the students. When new data becomes accessible, the model has retrained and modified to represent changes in student achievement. This is possible because the model is dynamic for different scenarios.

The suggested model, which is based on machine learning, offers a solution that is both objective and effective for determining which students are powerful and which are poor in academics. After the classification of students into slow and fast learners the model is further extended to find the learning disabilities of students if there are any. For the identification of learning disabilities in students they have been given various questions based on different subjects. To record their responses multiple parameters have been considered such as the accuracy, instances of skipping, revisiting, and un-attempted versus attempted ratio. Because of the precision of the model and its ability to analyze the significance of individual features, it became a helpful instrument for educators and educational organizations that are working to improve the performance of their students. It would be beneficial for future research to investigate both the flexibility of the model and its applicability in a variety of instructional contexts.

ACKNOWLEDGEMENTS

I am deeply grateful to my research advisor Dr.Tarandeep Kaur for her motivating guidance throughout this research work. She provided me with great freedom to choose a research topic of my interest and directed and critiqued me just in time. Without her inspiration, untiring efforts, enthusiastic attitude, and personal care, the accomplishment of this dissertation would have been impossible. I benefited enormously from her valuable feedback, logical analysis, and comments. Her input was vital in shaping this work. She has been a great source of inspiration to me, and I would like to thank her from the core of my heart.

I would like to gratefully acknowledge the School of Computer Applications, Lovely Professional University (Punjab), India, for offering this doctorate program. The University and the department were always very cooperative in extending all the resource and infrastructure facilities during my research span. I am thankful to all the reviewers of my research manuscripts for providing valuable feedback and comments that helped me raise the standard of this research work. I also extend my special thanks to all the cited authors in my thesis whose research findings provide me with a battleground to choose from the research directions.

Words are inadequate to express my gratitude to my loving Husband, Kawarpreet Singh, for always being there for me. He never failed to give me great encouragement and confidence to finish my thesis. I am highly grateful to him for all the cooperation, affection, motivation, understanding, patience, and sincere advice. His support and care helped me overcome setbacks and stay focused on my thesis. I greatly value her friendship and sincerely appreciate her belief in me.

I am also grateful to my daughter, Harseerat Sachdeva, and my friends for their best wishes and words of encouragement. They cheer me up during the walk through the journey towards attaining the Doctorate Degree. Also, I express my thanks to my other family members for always being with me at any stage during this work.

Above all, I firmly believe that everything in the world is happening as per the plan and will of the Almighty. So, I thank him for planning this and allowing me to do this thesis. I pray that he continues showering his grace and blessings onto me and all others.

Harjinder Kaur

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

EDM	Education Data Mining
ML	Machine Learning
LA	Learning Analytics
KNN	K-Nearest Neighbor
SVM	Support Vector Machine
ADM	Academic Data Mining
ETT	End-Term Test
MTT	Mid-Term Test
LD	Learning Disabilities
GPA	Grade Point Average
ANN	Artificial Neural Network
PCA	Principle Component Analysis
Stu	Students
Cou	Courses
MC_{ij}	Marks obtained by i^{th} student in j^{th} course
CR	Credit Score
G	Grade Point
SLD	Specific Learning Disorder
LMS	Learning Management System
LR	Logistic Regression

DM	Data Mining
HNF	Hybrid Neuro-Fuzzy
A	Accuracy
P	Precision
R	Recall
tp	True Positive
tn	True Negative
fp	False Positive
fn	False Negative
NTE	Number of Test Evaluations
LSTM	Long Short-term Memory Model
GRU	Gated Recurrent Unit
VARMA	Vector Autoregressive Moving Average
RTL	Ready To Learn
D (MSE)	Delay (Mean Square Error)

Chapter-1

INTRODUCTION

1.1 INTRODUCTION

In today's dynamic and rapidly changing world, there is an increasing demand for proficient experts in various domains. Education has always been regarded as the keystone for constructing a prosperous professional life, and a student's level of success in the classroom is one of the most important and critical factors that determine the opportunities that lie ahead for the student and further on help in achieving their professional success. Therefore, it becomes significant to focus on the student’s learning strengths and weaknesses. Additionally, the prompt identification of student academic success and weakness is utilized for providing individualized assistance and support to the student that helps them in achieving their academic objectives.

Recently, machine learning (ML) has evolved as an effective instrument for analyzing data and making predictions, and it is finding applications in a wide variety of disciplines, including education. Data mining is the process of retrieving usable information from enormous amounts of data in different sectors. Data mining is an important technology that is used for predicting and categorizing client information based on the considered sector.

Today, a variety of organizations utilize data mining techniques to assist with decisions and create promotional strategies that have been targeted toward client categories in order to accomplish their objectives. However, a substantial number of universities ignored the utilization of data mining approaches the result of which is that they are facing an increase in their drop-out rates. Data mining applications in the education industry are an upcoming trend in an incredibly competitive globally. The building blocks of developing data mining in the education sector require an understanding of data mining terminologies, tasks, methodologies, and applications. Therefore, research on data mining is significant in the education sector is necessary.

In the year 2005, a relatively new field called Educational Data Mining (EDM) emerged. EDM focused on developing techniques for data mining to extract data from educational institutions. By examining the academic data of students, the major focus of this field is to facilitate decision-making easier for educational institutions. The patterns generated by the mining techniques help educators as well as students [1]. In this regard, the development of a machine learning model for the classification of students based on their strengths and weaknesses using multiparametric analysis is an important area of research. The goal of the proposed model is to perform an analysis of multiple characteristics, such as a student's academic performance, punctuality, and social behavior, to determine which students need additional assistance and which students are performing exceptionally well in their classes. The main challenge in the model formulation and prediction is the task of determining and classification of students who are academically weak or strong, especially in major educational organizations. The exams and assignments, which are both traditional forms of evaluation, are only able to provide a limited amount of information regarding a student's general scholastic achievement. In addition, they don't take into account other aspects that might be relevant to a student's performance, like their punctuality or how they behave in social situations.

The utilization of ML techniques in the field of academics leads to a significant research area known as ADM or EDM [2]. Different machine-learning approaches have been used for analyzing and predicting academic performance of students. Previously, the ML models have been investigated in several studies for their potential to differentiate between students of average and superior ability. For instance,[1] suggested a model that predicted student scores based on statistics regarding scholastic achievement in addition to attendance. The precision of the model is 91.2%, which demonstrates the possibility of ML models for use in scholastic achievement projection. A similar approach has been taken [2], which predicted scholastic achievement based on social behavior data and obtained an accuracy of 85.4%. On the other hand, there is a dearth of research on the construction of an ML model for the classification of students based on their strengths and weaknesses using multiparametric analysis.

A more in-depth analysis of a student's academic performance is possible with the help of a machine learning algorithm that uses multiparametric analysis to differentiate between students with strong and poor academic performance. The model can analyze multiple characteristics, such as a student's scholastic performance, punctuality, and social behavior, in order to determine which students, need additional assistance and which students are performing exceptionally well in their classes. A model like this one may make it possible for teachers to offer individualized assistance and support to students, which may in turn assist the students in achieving their scholastic objectives.

1.2 MACHINE LEARNING AND EDUCATION

Education is one of the most essential elements that determine the expansion and maturation of an individual as well as the growth and development of society. It is the responsibility of the educational system to make sure that students graduate with the information and abilities they need to be successful in life, and it is a position that should not be taken lightly. On the other hand, it is not always simple to distinguish between students who are having difficulty and those who are performing exceptionally well. It is important to have a method that is dependable and accurate for recognizing the strengths and shortcomings of the students to be able to provide tailored interventions and assistance to the students.

The accumulation and storing of data from educational institutions is turning out to be enormous, and it is not feasible to directly analyze such a vast amount of academic data. In recent times, this trend has emerged. The process of gleaning valuable information from vast stores of data is known as "data mining." It became useful in a variety of research disciplines thanks to its versatility. This burgeoning area seeks to create data mining techniques that can harvest data from academic organizations and are currently in the process of doing so. By analyzing the academic data gathered from students, the primary objective of effective decision-making is to simplify the decision-making process concerning the academic achievement of students [10].

The use of machine learning, also known as ML, has become increasingly popular as a potent method for evaluating and making sense of large quantities of data. In the field of education, machine learning techniques have been utilized to analyze student achievement data to recognize patterns and trends that may be difficult for people to recognize. Educators can use this information to determine which students are at risk of falling behind in their coursework and then design individualized remediation plans to assist those students in catching up. Similarly, ML has been utilized to identify students who are performing exceptionally well and then provide those students with additional challenges as well as opportunities to further improve their abilities [3].

1.2.1 Learning Analytics

Learning analytics (LA) is playing an increasingly important part in the field of education. LA assists academic institutions so that they may conduct predictive analytics, which enables the institutions to better monitor the scholastic activities of their students. The part that LA plays in determining the variables that are causing concern for the attrition rate and, as a result, the progression of academic institutions is very important. As a result, a variety of LA tools are utilized to enhance education and learning. LA is in high demand in the modern education sector because it helps in fostering the scholastic expansion of an educational establishment [4].

The purpose of using LA by educational establishments is to classify students according to the patterns received from their education-related data. This is the driving force behind the use of LA. In addition, academic patterns help institutions plan new instructional methods and strategies, which are of assistance in enhancing the academic performance of students who are not as strong as their peers. In addition, LA offers advantages to academicians by assisting them in the identification of sluggish performers and by assisting students in monitoring their academic performance [5].

To support students' education throughout the various phases of their course, LA analyzes the data collected from students and the surroundings in which they learn. Even though LA is a relatively new subject of study, it has seen substantial growth in recent years, particularly in the educational sector [6]. Educational establishments have made extensive use of analytics to monitor the academic performance of their students, which helps in the identification of students who are either at academic risk or characterized as poor performers. These days, schooling is becoming increasingly essential, so various machine-learning approaches are being used for decision-making in almost every field. The incorporation of machine learning into academics has resulted in growth in the field of education, which has led to the development of a new field known as EDM or ADM. Additionally, with the help of techniques for machine learning, we can generate fascinating patterns from the students' scholastic data that we have accumulated. The resulting patterns can then be applied to further analysis of the student's scholastic achievement by using the information gained from that analysis.

In the realm of education, the following is a list of the significant benefits accomplished by utilizing LA in academics:

- Analyze the primary determinants that are accountable for the scholastic achievement of students: The LA assists in determining the primary variables that affect the level of scholastic achievement displayed by students. Class participation, completion of assigned work at home, passing scores on ongoing assessments, as well as performance on midterm and final examinations, as well as involvement in other types of skill-based tests, are all elements that have an impact on a student's overall scholastic performance. Students are given an advantage by being given the opportunity to work on the factors that are accountable for their scholastic failure when such factors are identified [7].

- Encourages students' intellectual development in the following ways: - The students are instructed to take preventative action after the identification of the variables that are accountable for the decline in the students' academic achievement during the early stages of their academic careers. Students who promptly take preventative measures see an improvement in their scholastic development, which in turn leads to a reduction in the number of students who fail or drop out of school [8].

- Determine and improve the effectiveness of various teaching methodologies: The learning assistants assist the teachers in developing new strategies and taking rectifying measures to support and encourage sluggish learners during their initial years of schooling. It also helps teachers evaluate the students' learning capacities and monitor their scholastic achievement in a way that is both effective and efficient. This is one more way that technology is revolutionizing education. The findings of the LA were used by the instructors to identify students who were poor learners so that they could give those students additional attention [8].

Figure 1.1 : Learning Analytic Components

1.2.2 Components of Learning Analytics

The purpose of LA is to provide enhanced insights into the learning process and individualized comments to students, both of which have ultimately led to improvements in educational quality [9]. The components that are essential to the operation of the LA system are depicted in Figure 1.1 and the explanation of each is as follows:

When trying to anticipate a learner's or student's level of achievement, the scholastic statistics of the learner or student are a critical factor to consider. When all of the data has been gathered, the next step is to investigate it and then turn those findings into analytical observations. Incorporating intelligence into the data requires the use of machine learning techniques, which are implemented during the process of analysis. The investigation and analytical observations of academic data provide a clear picture of the academic needs of the learners and the rectifying actions that need to be done to fulfill their educational requirements.

Learners are further aided in their efforts to improve their performance when educational requirements are met on time. Taking appropriate action is the primary purpose of any LA strategy. Inaction would be the ultimate expression of failing. No matter how accurate the results of prediction are, the results are pointless if appropriate actions have not been taken on them. This is true regardless of how accurate your prediction results are. Consequently, if adjustments have not been made to the system, the analytics findings are regarded as inadequate.

To facilitate more effective teaching and learning for students, LA carefully monitors the procedures followed by academic institutions. The careful examination of the academic data that has been supplied by the students has been of benefit to the learners, instructors, and academic organizations [8].

1.2.3 Diverse Learning Analytic Techniques Used in Academics

The various learning analytic techniques used in the field of education are depicted in Figure 1.2 which includes:

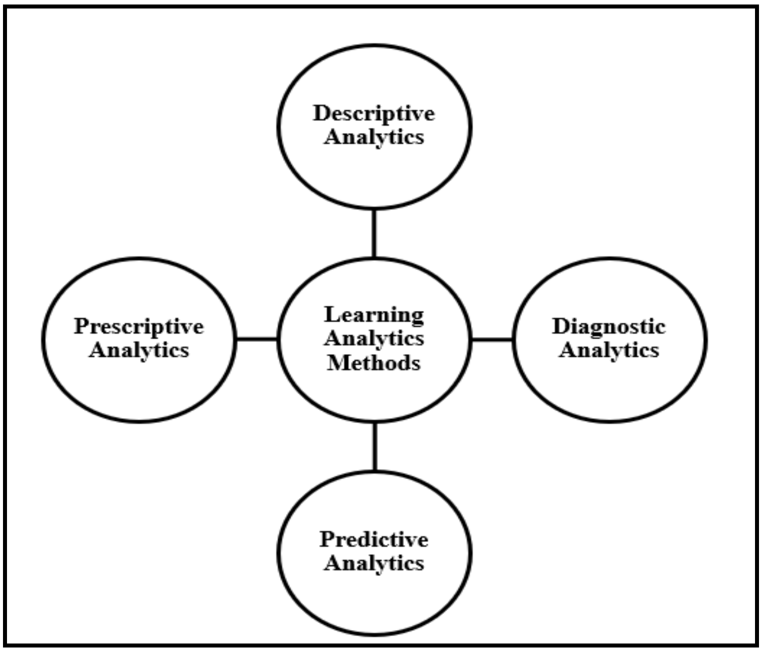

Figure 1.2: Learning Analytic Techniques

- **Descriptive Analytics:** Descriptive analytics integrates the data from multiple sources to give the facts about the previous performances of the students. The data from the collected sources is then used to make intelligent decisions that help in academic performance improvisation. It provides information about past happenings. The descriptive analysis aids in finding the data such as the increase in the retention rate and the courses responsible for the academic decline of the students using the past data.

- **Diagnostic Analytics:** To identify the reasons or causes responsible for the past happening diagnostic analytics has been used. It assists in the identification of dependent variables that help to give insight into a particular problem. It helps in the comprehensive analysis of the student learning behavior, emphasizes the students' learning requirements and offers them the appropriate academic support which helps in their performance improvisation.

- **Predictive Analytics:** Predictive analytics predict probable future occurrences using historical information. The accuracy of the predicted results is highly influenced by the data quality used for prediction so utmost care is required during the data selection. Predictive analytics helps in the identification of possible barriers to the student's experience during their academic learning. This early identification of academic challenges assists the faculties and academic institutions in providing appropriate counseling and support to the slow performers.

- **Prescriptive Analytics:** Prescriptive analytics intends to provide solutions to the challenges faced by students in their academics. It assists academic institutions in identifying the various reasons or factors responsible for deteriorating student academic results along with the possible solutions for their academic betterment [92].

1.2.4 Benefits of Learning Analytics in the Field of Education

In the field of education, LA provides various benefits. These benefits cover a vast range of students, teachers, and academic institutions. The different benefits are depicted in Figure 1.3.

- **Identification of Target Courses:** The scholastic data has been collected from the students using LA, and then used to provide learning patterns based on the collected data. The examination of students' learning patterns provides teachers with assistance in determining which classes are to blame for a decline in students' scholastic achievement [10].

- **The Improvisation of Curriculum:** The use of LA on academic data enables academicians to make changes to academic systems, which in turn leads to the improvisation of the curriculum. The instructors use the research to determine whether or not it is necessary to make changes to the curriculum by identifying gaps in student understanding and learning and using that information to inform their decisions. The instructors make plans for new scholastic strategies to improve the students' overall capacity for learning.

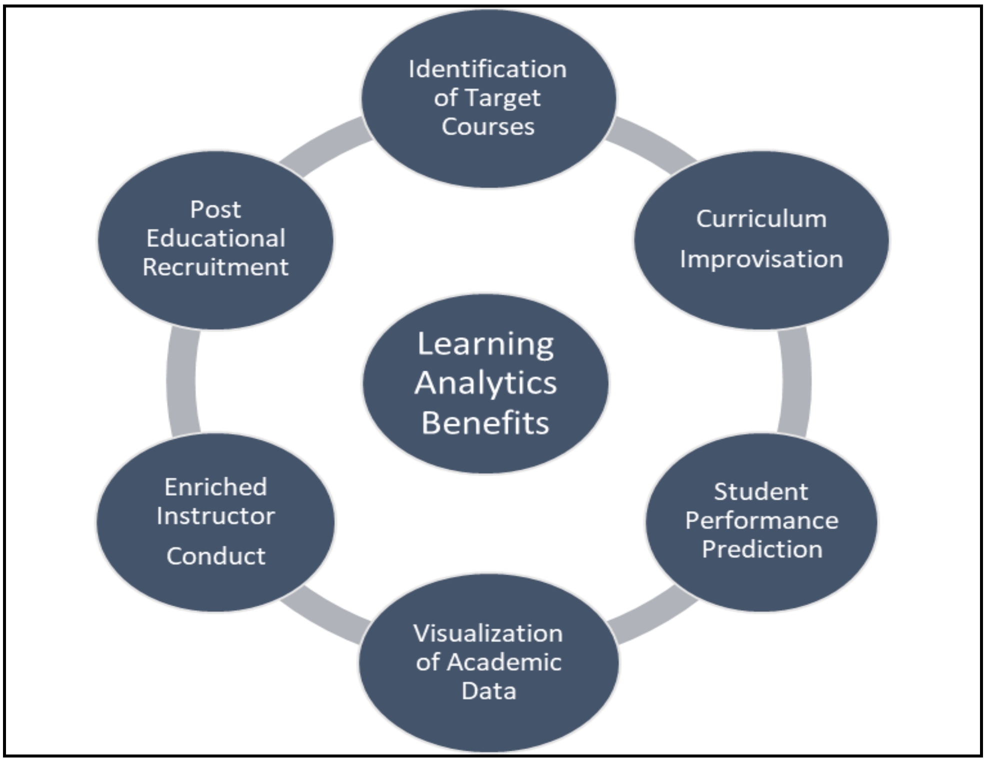

Figure 1.3 Benefits of Learning Analytics

- **Prediction of Student Performance:** LA can assist in making an educated guess as to how well students perform academically in the future by analyzing past data and applying a variety of machine-learning strategies. After that, the findings of the projection assist in classifying the most capable students from the least capable ones. The early identification of students who are likely to drop out enables teachers to place a greater priority on such pupils by offering appropriate guidance in the hope that the students have been able to complete the program within the allotted time [10].

- **Visualization of Academic Data:** The visualization report of academic data helps keep checks on all the academic activities that students are involved in. This assessment is in the best interest of both the students and the teachers. The instructors provide appropriate counseling, develop new learning methodologies and strategies, and work to improve the students' learning competence to assist students who are struggling academically [9]. Learners are better able to conduct an accurate self-evaluation, which in turn enables them to focus their efforts on improving the areas in which they struggle.

- **Enriched Instructor Conduct:** Teaching methods used by instructors that include improvisation, and the scholastic achievement of the students have been used as a basis for evaluating the performance of the instructor. The findings provide the teachers with the information they need to focus their training on the subjects that are having a detrimental effect on the students' performance. As a direct consequence of this, the lecturers are better equipped to respond to questions raised by the students regarding the content of the classes. Therefore, the scholastic data of the learners allows the instructors to better prepare themselves in the identification of areas in which the students require development, which further promotes the learner-student interaction.

1.3 EDUCATIONAL DATA MINING

The field of EDM, assists in forecasting the academic achievement of students, which then results in the students being categorized as slow or quick learners. EDM is used to retrieve student information such as academic records, the participation of the student in class, ETE scores, and attendance in class, all of which assist in analyzing the academic achievement of the student [11].

EDM has been used in the present situation for the identification of concealed patterns in the academic data of students, which helps in the process of forecasting the academic achievement of learners. It makes it easier for teachers to monitor the academic achievement of their students, which in turn allows them to enhance the academic techniques they employ [14]. It assists teachers in evaluating the course framework and developing innovative academic approaches, both of which help students improve their overall performance [86]. Consequently, EDM provides assistance to the students in the form of constructive and expert guidance during the process of selecting courses.

For the students to further improve their performance, the course recommendations are given to them based on how well they did in the previous class. On the other hand, advisers can make more accurate predictions regarding students' scholastic performance by utilizing the insights provided by the emerging discipline. The early identification of students who are struggling can help those students concentrate more on improving the areas in which they are deficient.

By analyzing the educational data that is made available by the various institutions, EDM's primary objective is to develop solutions to the issues that are associated with the academic performance of students. EDM plays a very important role in the process of addressing significant educational issues such as the categorization of students according to their performance and suggesting corrective solutions to the students who are performing poorly [12]. This is done to address significant educational problems such as the classification of students. Several different methods of machine learning are applied to the information that is compiled from the various educational establishments to evaluate the performance of the students.

Figure 1.4 depicts the process of education data mining. Here, the academic data is retrieved from students employing EDM, and a LA strategy is then applied to analyze the educational data submitted by students.

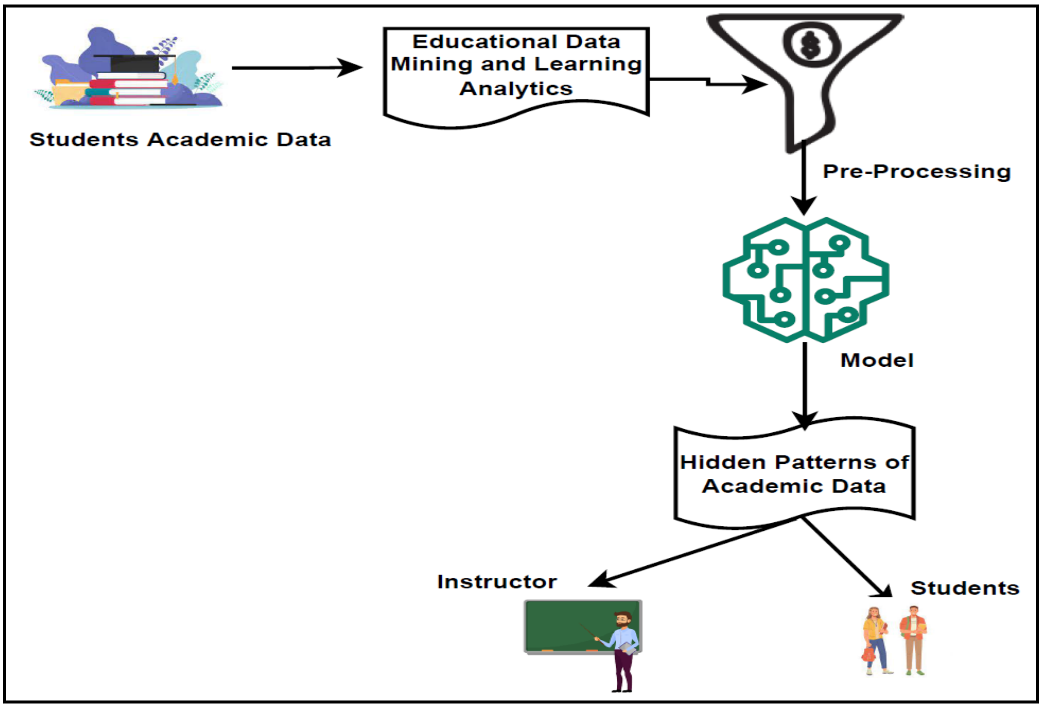

Figure 1.4 : Generalized Process for Educational Data Mining

The academic information about the student has been gathered through a variety of methods, such as through interaction with instructors and fellow students, as well as from the directory of the educational establishment. The enormous quantity of scholarly data that comes from a wide variety of sources is accessible in a range of platforms and broken down into granularities of varying degrees [13]. The fact that the gathered academic data is not in the form necessary to resolve academic problems is the primary obstacle that must be overcome after the data has been collected. Therefore, prior to any predictions, it is essential to transform the data into the shape that is required for making predictions regarding the students' achievement. Before we can use the dataset on the model, we need to clean the accumulated information so that it does not contain any characteristics that are unnecessary or superfluous. The eradication has been accomplished using pre-processing, during which time all kinds of irregularities were eliminated. After being pre-processed, the data are entered into a prediction model, which uses the scholastic information provided by the student to produce concealed patterns. Students can use the projected findings to assist them with their self-assessment, and instructors can use them to better advise and instruct students who are working at a slower or weaker pace.

The EDM carries a significant amount of weight with scholastic organizations. Figure 1.5 illustrates the components used to extract the information from academic organizations.

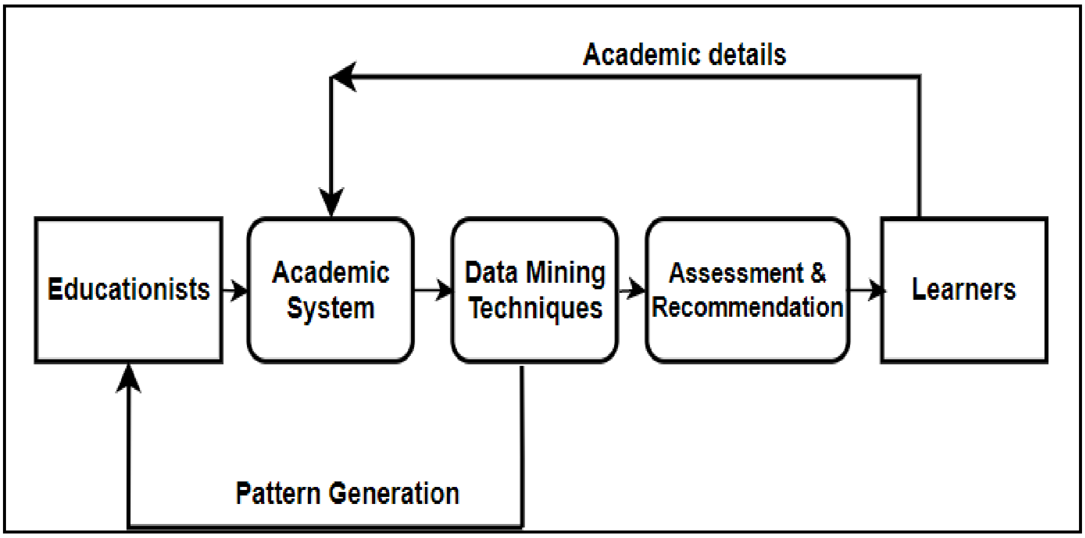

Figure 1.5 Integral Components of Educational Data Mining

1.3.1 Components of Educational Data Mining

Research in the field of education has been the driving force behind recent advances in the field. The way scholarly material has been examined and adapted as a direct consequence of developments in technology. The utilization of the information obtained through these advancements is currently supporting a second period of change in all fields of learning, which has resulted in a diverse array of accomplishments. Data mining is the most effective approach for assisting educational institutions, where the primary emphasis is on the information that is contained in the data sets that have been collected about the scholastic performance of students.

Data mining applies its methods to the accumulated data sets to discover previously unknown patterns and correlations between variables that have an impact on the scholastic achievement of the students. These patterns and associations were previously unknown. These technologies may make use of quantitative formulations, statistical models, and techniques for machine learning. The learners, educationists, the scholastic system, data mining techniques, assessment, and recommendation techniques, are the primary components of the mining process.

The primary objective of EDM is to improve the process of learning, as well as the identification of obstacles that contribute to failures, the process of course selection, and ultimately the prediction of the student's scholastic achievement. If timely action is taken regarding each of the variables, scholastic organizations have experienced an increase in their level of productivity. Since academic institutions are unable to properly handle their education system when EDM is not present, EDM provides sufficient information about the study patterns of students and instructional techniques, which assists academic institutions in illuminating their reputation [15].

When it comes to inspecting the scholastic data coming from educational establishments, each component plays a unique part. The operation of each component is described in detail below.

- **Educationists:** The educationist plays a very important role in the planning of academic systems. To ensure that students are motivated to succeed in their studies, the educational system ought to be fashioned with the consideration of the students' areas of interest at the forefront. To accomplish this goal, teachers need to review the students' historical records by employing machine learning techniques to evaluate their performance, which in turn contributes to the improvement of the educational system.

- **Academic System:** An efficient academic system gives trainees a healthy learning environment. This is something an effective academic system should do. Students make up the most important part of the educational system. To enhance the academic achievement of students and, as a result, the academic system, the academicians have the responsibility of developing curricula and strategies that are tailored to the interests of the students.

- **Evaluation and Recommendation:** The purpose of this process is to increase the effectiveness and productivity of the educational system, as well as to improve the scholastic performance of students and to identify and classify students as sluggish or quick learners according to their performance. Following the classification, the recommender has proposed the appropriate class to the students who are slow learners, which further assists in improving their overall scholastic achievement.

- **Learners:** The most important aspect of scholastic data extraction and research is the participation of the students. The scholastic data of the students has been gathered, and then a variety of machine learning techniques have been applied to the gathered data to generate a variety of patterns. The patterns that are generated are dependent on the scholastic inputs that are supplied by the students, which enables the classification of learners as either sluggish or quick learners.

1.3.2 Beneficiaries of Education Data Mining

Using EDM, it is possible to classify students according to their scholastic achievement and forecast whether they fail out of school. Learners can improve their scholastic achievement by receiving assistance from EDM during the process of choosing which classes to take. In addition to this, it assists the teachers in improving their teaching techniques and strategies so that they better assist the students who are struggling. The information that is utilized in EDM is aimed at a variety of different stakeholders. The scholarly data has been viewed from a variety of perspectives by various stakeholders, each of whom has done so while bearing in mind their objective and their impression of how EDM should be applied [12]. Figure 1.6 presents an overview of the most important EDM beneficiaries.

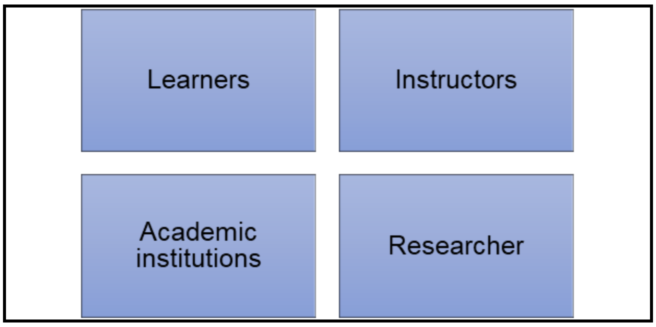

Figure 1.6 Beneficiaries of Education Data Mining

- **Learners:** The EDM is useful for forecasting the scholastic achievement of learners, which in turn results in the classification of learners as either sluggish or quick learners. In addition to this, it explains the factors that led to the scholastic failure of the students, which is helpful for the students when they are conducting their self-evaluations. In addition to this, it makes it easier for students to choose their courses based on their anticipated academic performance, which ultimately leads to an improvement in their overall academic performance [10].

- **Instructors:** The instructors can monitor the scholastic progress of their students, which contributes to the overall improvement of the learning process. To improve the quality of instruction provided, teachers should first recognize students who are challenged academically and provide them with additional support in the form of counseling and additional study time. The EDM makes it easier for teachers to come up with innovative learning techniques and strategies by analyzing the performance of their students' learning [13].

- **Researchers:** To develop novel projection models, the researchers make use of heuristics that are connected to the scholastic pursuits of the students. After being educated with the historical academic data of students, the prediction model then put to the test to determine how accurately it could forecast the academic performance of learners. In addition to this, it assists the researchers in assessing the present learning environment, developing new education systems, and establishing the efficacy of machine learning techniques.

- **Academic Institutions:** The ADM helps in the classification of different students, such as sluggish learners and quick learners, into their respective categories. In addition, the findings of ADM are put to use to encourage and instruct sluggish students, which ultimately contributes to an increase in the percentage of students who remain enrolled in educational programs. The increase in the scholastic achievement of students as well as the percentage of students who choose to continue their education has a direct effect on the profitability of academic institutions [14].

1.4 CONJUNCTION OF LEARNING ANALYTICS AND EDUCATION DATA MINING

LA and education data mining are terms that refer to the process of using analytics and data mining techniques to retrieve massive amounts of data from educational databases. In this current age, the field of LA, in conjunction with education data mining, has garnered more attention in the field of research. The EDM is utilized to generate a variety of patterns by utilizing the educational data gathered from academic institutions, and the LA incorporates a variety of methods from a variety of fields, such as machine learning and data visualization, to assist in the analysis of such patterns [20]. The incorporation of LA and EDM has resulted in an improvement to the scholastic process that is advantageous for a variety of stakeholders, including students, teachers, and administrators [29]. This improvement has been brought about because of the integration of LA and EDM. As a direct consequence of this, provides academic statistics and functions as a support system for various education constituents. It functions as a course recommender system that helps students select classes that are appropriate for their performance, and it does so based on the outcomes of the students' evaluations. Classification, prediction, and association are the approaches that are most frequently used with ADM and LA respectively.

Both EDM and LA demonstrate how computational approaches to learning are gaining traction in the field of education and are becoming more standard. The International Educational Data Mining Society offers the following description of EDM: "EDM deals with the development of methods used for discovering different kinds of data generated by various educational institutions." The implementation of these techniques contributes to a better comprehension of the requirements of students and the context in which they learn. Similarly, following the findings of the Society for LA Research, "the collection, analysis, and evaluation of academic data of students and the environment in which they learn" [21] is how LA is described. Table 1.1 provides an overview of the most important distinctions between EDM and LA methods.

Table 1.1: Educational Data Mining v/s Learning Analytics

EDM	LA
The key to EDM is the automated discovery of hidden patterns and to achieve this human decision-making is being used as a tool.	The key LA is human decision making and to achieve this automated discovery of hidden patterns is being used as a tool.
The main focus is on breaking the components into their constituent parts and examining each one's features and interactions.	The main focus of LA is on comprehending the complete system and its complexity rather than focusing on the individual component.
The EDM provides a strong connection between student prediction and the learning environment.	The LA has a significant connection with smart education, performance prediction, and institutional improvement.
The applications of EDM incorporate Classification, clustering, Bayesian modeling, association mining, and visualization	The applications of LA comprise Influence analytics, sentiment analysis, social network analysis, and student performance prediction.

The enormous quantity of educational data is documented in the educational systems, and this data has been used to improve both the academic achievement of students as well as instructional techniques, which in turn improves the reputation of academic institutions. During the process of extracting scholarly data, one of the challenges that must be overcome is the incorporation of various analytical methods, including data mining, machine learning, and LA. The ML offers a variety of methodologies, all of which are utilized in the process of developing the models that are applied to the projection of scholastic achievement. EDM, also known as ADM, is the process of gleaning valuable insights from the academic data sets that are supplied by students as well as academic organizations [48]. LA is, in the end, a collection of procedures for analyzing and enhancing not only the general learning process but also the surroundings in which it takes place. In the first stage of the LA process, the emphasis is placed on EDM, which is then put to use to collect instructional data using ML techniques.

Both EDM and LA put a priority on the facts that have been used to increase academic learning and enhance the general academic achievement of learners by using ML techniques. In the present situation, a technologically sophisticated environment provides a variety of implementations of LA in the educational sector. The multidisciplinary discipline of LA employs a wide variety of EDM and machine learning strategies [12]. The scholastic achievement of the students is impacted by several different variables, including their behavior, the marks they receive in MTT and evaluation, the marks they receive in ETT, as well as their scores, and the various course components. After that, the model developed with ML techniques used these statistics to make an educated guess as to the student's prospective scholastic achievement. In addition, students can continuously observe their scholastic achievement throughout the entirety of their program because of this feature. When viewed from the perspective of the instructor, student performance projection during the early stages of a teacher's career permits them to be proactive in informing students enrolled in poor classes that they need to improve their performance in those courses.

EDM assigns categories to academic data sets by employing a variety of machine-learning techniques [22]. The learners are able to better determine their overall score based on the projection of their academic achievement, which is done before the final test. This projection challenge, which is a primary goal in the field of education, has been the center of the implementation of ML and EDM. The primary objective of EDM is to unearth latent patterns in the academic data that has been collected. This includes the analysis of student performance and their learning behavior, as well as the preparation of coursework by instructors; however, the primary objective is to enhance the academic performance of students [18]. Academic achievement has been defined in a variety of ways, but the difficult part for educational institutions is figuring out which variables contribute to students' success in the classroom.

The EDM assists in the projection of academic achievement and is of use to all scholastic beneficiaries. EDM techniques focus primarily on three areas: the evaluation and projection of whether a student passes or fails in a particular subject; the students' ultimate scores; and the identification of students who are likely to drop out of school [12]. In addition, successful implementation of the EDM helps academic organizations increase the quality of the learning opportunities available to students, develop and implement learning techniques that are advantageous to students, and eventually reduce the number of students who withdraw from their studies [9]. As a consequence of this, these techniques help in predicting student mistakes in the beginning stages of their academic careers. In addition, teachers could improve their instructional methods by conducting an in-depth investigation into the factors that contribute to student failure and withdrawal.

The implementation of LA and EDM on academic information has the advantage of enhancing the learning process of academic institutions for a variety of academic stakeholders, including students, teachers, instructors, and administrators. Consequently, the LA serves as a support system for the process of education by providing academic statistics and academic recommendations. Additionally, it functions as an individualized effective learning system that is based on the academic data of the student. Prediction, categorization, clustering, and association mining are some of the techniques that are typically utilized when working with scholarly datasets.

1.4.1 Approaches Used in the EDM Process

The various approaches are taken during the EDM process, specifically the feature selection and categorization steps.

- **Choosing Among the Features:** The data that is fed into certain forecasting models includes a large number of input characteristics. This not only makes the process of model development and training more time-consuming, but it also necessitates an enormous quantity of memory. In addition to that, the accuracy of the projection model is reduced when some unimportant characteristics are used. A method known as feature selection is a process that is utilized to choose only the pertinent characteristics, thereby removing all the unnecessary attributes. It automatically selects the characteristics that are pertinent to the research problem and uses those attributes to train the machine learning model that has been developed to solve the research problem.

The feature selection leads to a decrease in the number of input characteristics, which in turn lowers the cost of the model's calculations and leads to an improvement in the model's overall performance [23]. When using techniques for feature selection that are based on statistical analysis, the relationship between the input attribute and the target attribute is evaluated, and then the attributes that have the strongest correlation with the target variable are chosen for inclusion in the final model. Even though the choice of technique for feature selection is dependent on the types of both the input and the target variable, the techniques themselves are very speedy and effective. Wrapper and filter methods are the techniques that are utilized for feature selection. These methods are utilized so that appropriate characteristics are chosen. The effectiveness of the model is used as a metric with which to evaluate how well these supervised feature selection methods work for different scenarios.

- **Wrapper Method:** The wrapping approach generates multiple models based on a variety of subgroups of the input features, after which it chooses the characteristics or features that provide the best performance model based on the target feature, as shown in Figure 1.7. In this case, it is important to note that even though these techniques may have a high computational cost, they are not contingent on the sort of characteristics [24].

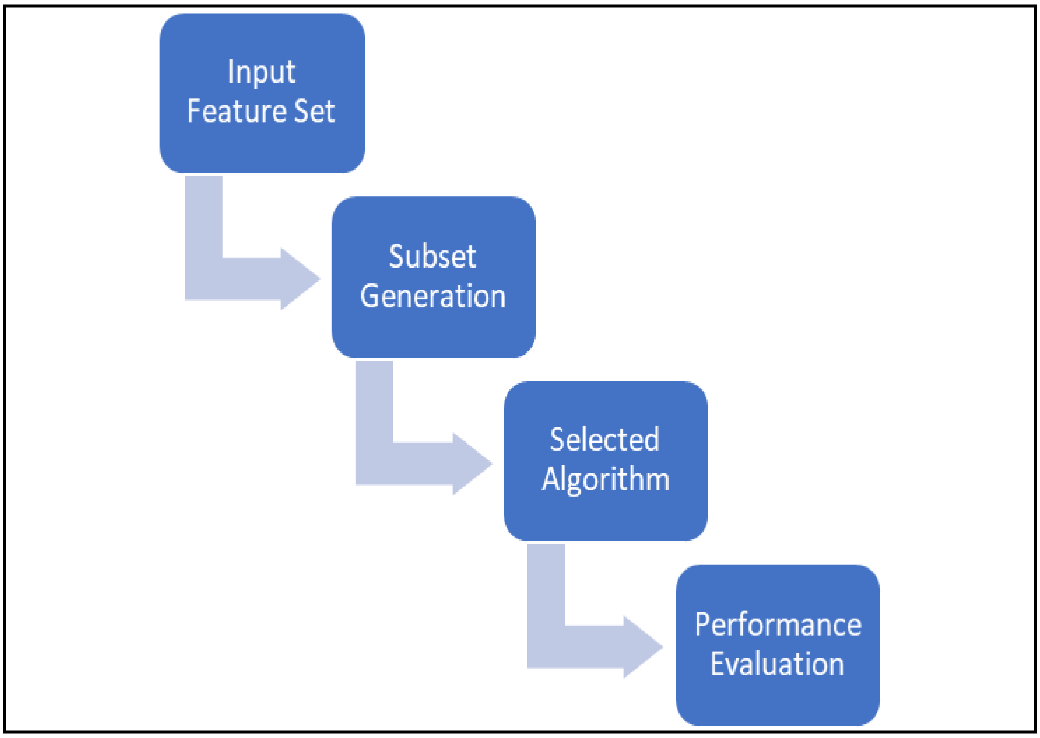

Figure 1.7: Techniques Useful in the Wrapper Method

- **Filter Method:** The characteristics used in the filter techniques are chosen following the statistical measures that are depicted in Figure 1.8. The learning algorithm does not have any influence on the selection of the features, which is done during the pre-processing phase. This approach eliminates characteristics that are not important or are repetitive by using a scoring procedure. The primary advantage of using the filter method rather than the wrapping method is that it requires less time for calculation and is unaffected by the problem of data being over-fit.

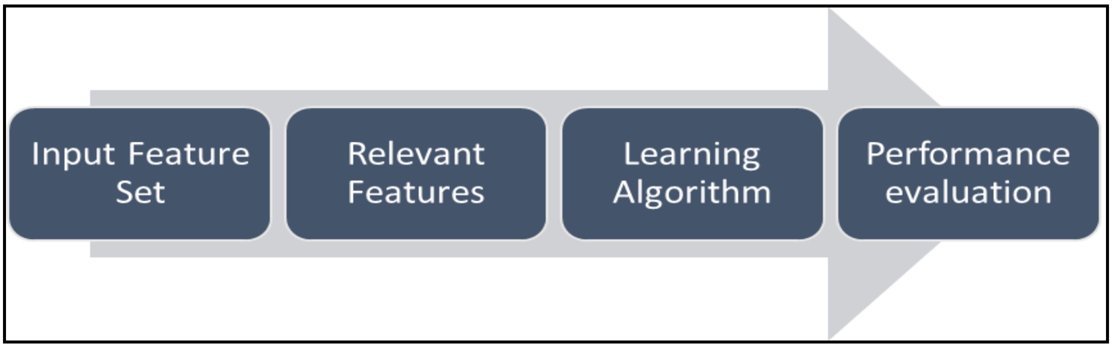

Figure 1.8: Techniques used for Filter Selection

- **Clustering:** Clustering is a form of unsupervised learning in which the model is learned without the use of any class identifiers. The primary objective of overcrowding is to subdivide and group the information in such a way that the data contained within one group are comparable to one another while the data contained within the other group are distinct from one another. The cluster is thought to be of high quality if, and only if, the similarity between members of the cluster is minimal while the similarity among members of other clusters is high.

- **Classification:** The classification is a supervised learning method that is used to categorize new data based on the training dataset, as shown in Figure 1.9. Classification is a method that is used to group similar pieces of data. The process consists of two stages:

  1. The learning phase is where the categorization model is developed; this step is also known as the learning step.

  2. This phase, which is referred to as the categorization stage, is used to predict the class designation.

During the learning phase, the model is educated by way of an examination of the training information. To symbolize the outcome as an n-dimensional vector is represented as follows:

$$V = v1, v2, v3, \ldots\ldots, vn \ldots (1)$$

It is assumed that each combination, V, is a member of a particular designated class, the membership of which is established by the target class identifier. The tuples that were used to train the models are referred to as training tuples, and they are chosen at random for analysis. These training tuples are then used to evaluate how accurate the developed model is at categorization [28].

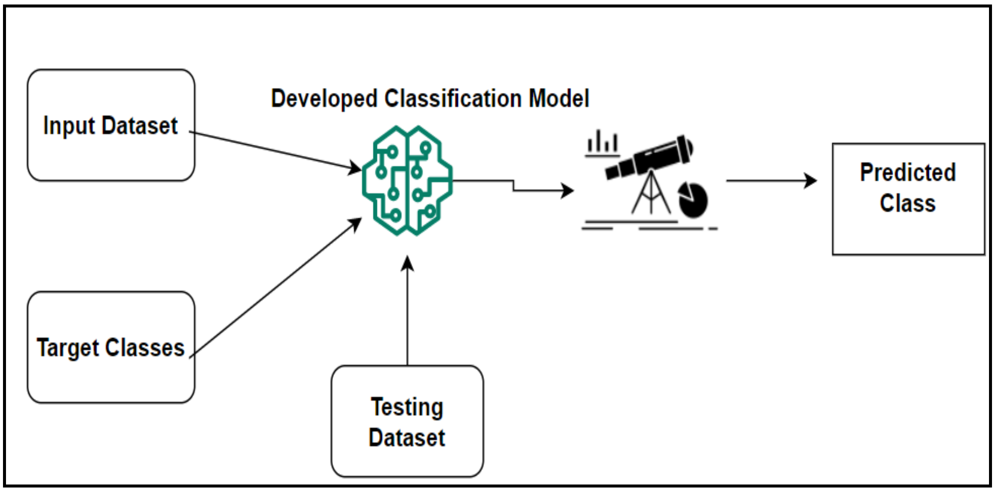

Figure 1.9: Generalized Classification Process

Training refers to the process of developing the model, and the information that is used during this period of development is known as the training dataset. Following the completion of the model development, a second data set is utilized to evaluate the performance of the newly developed model. The information that is utilized in the process of determining the precision of a model is referred to as testing dataset samples. To evaluate how well a model performs on a variety of data sets, it is necessary to conduct research using distinct datasets, such as training and testing datasets. If this is not the case, the model functions correctly on the initial dataset and generates accurate results, but it produces inaccurate results when applied to new datasets, which leads to the issue of over-fitting [29]. The input dataset is split up into a training dataset and a testing dataset so that the problem of over-fitting has been avoided. The precision of the model is evaluated by comparing the projected class with the actual class, which is determined by applying the model to make predictions about the target class using the test dataset samples.

1.5 PROBLEM STATEMENT

In the present scenario, there is a lot of pressure on educational establishments pertaining to improve the retention rate, improvisation of academic strategies and curriculum, along the timely identification of slow and fast learners. It has been observed that there is an increase in the number of students who fail out of school, and this requires the determination of the various variables that are influencing the students' performance. Consequently, the appropriate steps are being taken by the educational institutions in order to keep the level of education provided at a high standard and to boost the percentage of students who choose to remain enrolled.

In addition, even though the public and private sectors are providing less assistance for the institutions, numerous stakeholders still anticipate that the institutions have been able to fulfill certain requirements. Another problem posing a major concern globally is the generation of vast amounts of data. When compared to other markets, the educational market generates tremendous quantities of data [16]. In the past, however, they have ineffectively used data, which has frequently resulted in a sluggish response to the educational requirements of students. The information that is gathered from the students is put to use for analysis, and after that, the information that has been analyzed is put to use for the decision-making process. Thus, the academic sector must transform to resolve concerns such as improving student learning competence and student achievement, both of which contribute to an improvement in the reputation of academic institutions.

EDM is being extensively used for the identification and prediction of students' academic performance. The academic institutions are now better able to adjust to changes that impact the learning surroundings of students [17], thanks to the utilization of EDM. The proliferation of large amounts of data: what does this imply for the fields of education, technology, and media research? EDM contributes to the field of LA, which makes it possible for educational institutions to comprehend the challenges faced by students and contributes to the institution's ability to effectively resolve these challenges.

In the realm of electronic design manufacturing (EDM), machine learning approaches play a very important role in the process of extracting data and predicting outcomes. Although there has been a great deal of variation in the applications that could be used for machine learning, their core function remained the same. Machine learning algorithms have been trained or learned using the previous data, and the rules are generated using the training dataset that has been used for the decision-making process.

Based on the nature of the incoming information, machine learning methods have been divided into two categories: supervised learning and unsupervised learning. In supervised learning, the class designation for the variable being learned about is already known. Input data is used to teach the algorithm, and this collection is referred to as the training dataset. The machine learning method known as predictive analytics, which is predicated on supervised learning, is the one that is used to generate predictions. Applications of predictive analytics have been found in a wide variety of fields, including fraud identification, the analysis of consumer behavior, and the trend analysis of demographic data, amongst others.

The primary concentration of this thesis is on the field of education, and the primary objective is to forecast the overall scholastic achievement of students. An ensemble model is currently being developed, and as part of this process, it is being trained in such a manner as to be able to forecast the various classifications of learners, such as sluggish learners and rapid learners, based on the input scholastic information of students. The original information that is used to train the model is the training dataset. This helps ensure that the model is accurate.

This research work intends to propose an Ensemble ML model extending the capabilities of EDM. The purpose of the model is to make predictions about the target variable by analyzing the characteristics of the incoming variables. After the model has been developed, it is used to analyze data that has a class structure of the target variable that is identical to the entered dataset. In unsupervised learning, the class designation is unknown, and the primary goal of this algorithm is to organize the data into groups based on the similarities among the characteristics [17].

In addition, the student's academic performance that the model is trying to forecast is referred to as the objective variable. This variable symbolizes the student's academic performance. The algorithm develops an ensemble model with the assistance of other input characteristics. An ensemble model is a function that produces different categorizations of students based on their scholastic achievement [18].

1.6 RESEARCH OBJECTIVES

The objectives of the research are:

1. Identification and collection of datasets from different sources and applying attribute selection methods.

2. To develop an ensemble model for predicting student academic performance.

3. Compare and analyze the developed model with the existing model.

EDM in the educational sector has already been available, the primary goal of this research work is to investigate and identify the helpful rules and patterns that have been used to motivate students to manage their educational performance and careers in a good manner for them to finish their degrees on time. This research work has been carried out to examine and identify useful rules and patterns. It also helps academic institutions improve their academic strategies and methodologies, which is ultimately beneficial for the students enrolled in those institutions. The magnitude of the efforts that have been made by educational institutions to enhance the learning of sluggish or ordinary students has been determined by analyzing the performance of students who are enrolled in those educational institutions. One of the advantages of utilizing EDM models is that they make use of the historical data of students to make projections about their future performance which has not yet taken place [4,12]. This idea has inspired many academicians to develop categorization models that can make accurate predictions regarding the labels that should be applied to upcoming events based on their uncertain characteristics. The majority of educational institutions and academicians started taking an interest in the field of student performance projection so that they could classify students into sluggish learners and rapid learners according to their anticipated academic performance. Even though the educational industry uses a wide diversity of approaches to collect scholastic data from students, it is essential for the development of a model for evaluating student achievement. The findings of the predictive model provide students and teachers with useful information that has been used to take their performance to the following level.

The following are some of the individuals who have benefitted from using the performance prediction model:

- **Students/ Learners:** The most significant advantage of using a prediction model is that it facilitates the categorization of various student profiles. The model is helpful to the students because it can forecast their scholastic performance and it can also identify the factors that are responsible for their performance declining. Students are able to better monitor their scholastic performance as a result of the findings generated by the prediction model.

- **Instructors:** By monitoring how students perform in their academics, prediction models provide an advantage to teachers, allowing them to categorize students as sluggish or quick learners more accurately. The instructors are assisted by the projected results in determining the factors that contributed to a decline in the academic performance of the students.

- **Establishments of Higher Learning:** The reputation of an institution depends, to a very large extent, on the level of scholastic achievement that its students or other types of learners achieve. Therefore, educational institutions have a responsibility to place a strong emphasis on the academic achievement of their students because this is regarded as an essential indicator that accurately reflects the standard of educational institutions. Therefore, identifying students who have trouble learning and then providing them with the assistance they need to enhance their performance contributes to an increase in the retention rate of academic organizations [25].

1.7 RESEARCH QUESTIONS

Table 1.2: Research Questions and their Probable Answers

Research Question	Probable Answer to the Question
What is the best combination of parameters for identifying weak and strong students using the ML model?	The best combination of parameters for identifying weak and strong students has been determined by analyzing the performance of the ML model on the training and testing datasets. The model has been

	trained and tested using different combinations of parameters to identify the most effective combination.
How does the accuracy of the ML model vary with the size of the dataset?	The accuracy of the ML model may vary with the size of the dataset. A larger dataset may provide more information for the model to learn from and improve accuracy. However, there may be diminishing returns in accuracy as the size of the dataset increases.
How does the performance of the ML model compare to traditional methods of identifying weak and strong students?	The performance of the ML model has been compared with the traditional methods of identifying weak and strong students, such as using standardized test scores or teacher evaluations. The comparison has been based on accuracy and efficiency.
How does the accuracy of the ML model vary with the complexity of the model?	The complexity of the ML model increased by adding more layers or nodes to the neural network. However, increasing the complexity of the model may result in overfitting or reduced accuracy. The accuracy of the model further be analyzed for different levels of complexity
How does the ML model perform when applied to different academic subjects or grade levels?	The ML model has been tested on performance data from different academic subjects and grade levels to evaluate its generalizability. The model's accuracy has been then compared for different subjects and grade levels.
Can the ML model be used to identify specific areas of weakness or strength in a student's performance?	The ML model may be able to identify specific areas of weakness or strength in a student's performance by analyzing the different parameters used in the model. This information has been used to provide targeted interventions to help the student improve.
How does the ML model account for external factors that may affect student performance, such as socioeconomic status or family background?	The ML model may be trained on data that includes information on external factors that may affect student performance. The model has been evaluated to determine its ability to account for these factors and adjust its predictions accordingly.

1.8 THESIS CONTRIBUTIONS

The findings of this research might have important repercussions for both theory and practice in the field of education. This approach could help to ensure that all students receive the necessary assistance for them to be successful by providing teachers with a method that is dependable and accurate for determining students' levels of academic ability. In addition, the model could be utilized to develop individualized learning plans and tailored interventions for specific students, which would further contribute to improving the overall standard of education that is delivered to students.

The proposed research work offers significant benefits for the academic sector that are discussed below:

1. The study contributes to the development of new methods for assessing and managing learning disabilities, which can have a significant impact on the field of education and the lives of students with learning disabilities.
2. The research has the potential to enhance educational outcomes by assisting the teachers in locating students who are having difficulty in their coursework and enabling them to offer the appropriate assistance and remediation to these students.
3. If teachers are able to recognize struggling students at an early stage, they are able to offer them specialized assistance, such as additional coaching or individualized study programs, to help them improve their scholastic achievement.
4. The instructors can provide strong students with opportunities for development, such as advanced schoolwork or research projects if they are able to recognize these students.
5. The study proposes a model for the assessment of features having varying practice patterns in students with the presence or absence of learning disabilities using multiparametric analysis. The model is designed to be scalable, customizable, and suitable for experiments, which can improve the identification and management of learning disabilities.

6. The study demonstrates the potential of real-time data collection and analysis in the assessment of learning disabilities. This can provide educators with up-to-date information on students' learning progress and facilitate the development of targeted interventions.

7. The research is able to provide a more in-depth analysis of a student's scholastic achievement because it has been designed using a machine-learning algorithm that takes into consideration several different parameters. These parameters include things like punctuality, scores, and participation.

8. The proposed model shows promising results in terms of accuracy and effectiveness compared to existing deep learning-based LD analysis methods. This can lead to better identification and management of learning disabilities, which can improve educational outcomes for students with learning disabilities.

Conclusively, the research contributes to the fields of machine learning and education by investigating how multiparametric analysis is used to differentiate between students based on their learning capabilities i.e. slow and fast. This method has the potential to be utilized in other facets of education, such as determining which students are at risk of not completing their education or forecasting which students have been successful academically in higher education scenarios.

The proposed research has the potential to enhance educational outcomes overall by providing a more impartial and comprehensive analysis of student performance, as well as by enabling educators to provide tailored support and interventions to students who have the greatest need for them under different scenarios.

Chapter-2

LITERATURE REVIEW

2.1 INTRODUCTION

Machine learning (ML) algorithms and LA have brought about a transformation in the education industry, making it possible to determine which students are weak and which ones are strong. This literature review seeks to provide a synopsis of current research on the construction of ML models for determining weak and strong students via multiparametric analysis.

Recent research has shown that machine learning algorithms are effective in determining which students are weak and which are strong by using a variety of characteristics. For example, in research conducted by [26], an ML model has been developed to forecast the academic performance of students based on their attendance, class participation, and previous academic performance. This model makes accurate predictions about the academic performance of students. The precision of the model has been determined to be 91.33 percent, proving that it is useful in distinguishing between weak students and those who are strong.

In the same vein, [27] developed a machine learning model to forecast the academic performance of students based on their previous academic performance, socio-economic position, and learning style. This model has been used for predicting the educational performance of the students. The precision of the model is 93.33%, which demonstrates its efficiency in distinguishing between weak and strong students.

The creation of ML models to distinguish weak and strong students includes a number of stages, including the acquisition of data, the preprocessing of data, the extraction of features, the selection of models, and the assessment of models. The process of collecting data entails compiling information on a variety of aspects, such as a student's scholastic achievement, punctuality, level of participation, socioeconomic standing, and learning style. The data that has been collected is then preprocessed so that any mistakes or inconsistencies have been removed. During the process of feature extraction, the most pertinent features from the preprocessed data are selected for use in the subsequent process of training the ML model. The model selection includes choosing an appropriate machine learning technique to use based on the characteristics of the data and the specific challenge being addressed. In the final step, the performance of the ML model is assessed based on its accuracy, precision, memory, and F1 score.

However, it is not easy to design machine learning algorithms that can identify weak and strong students. The absence of standardization in data acquisition and preprocessing is one of the primary difficulties that must be overcome. It is challenging to evaluate the performance of machine learning models across institutions because different organizations acquire and preprocess data in potentially unique ways. The selection of pertinent characteristics for the ML model presents yet another obstacle to overcome. There is a possibility that some features are more pertinent than others, and picking the incorrect features could result in the ML model having poor performance.

2.2 AN ANALYSIS OF THE SCHOLASTIC SUCCESS OF STUDENTS ENROLLED IN HIGHER EDUCATION

For decades, researchers and teachers in higher education have been interested in studying and discussing the scholastic achievement of students in the institutions. The purpose of this overview is to present an all-encompassing comprehension of student scholastic achievement in higher education, including its description, methods of measurement, and the variables that affect it. The students are judged on their scholastic achievement based on how well they are able to fulfill the expectations placed upon them by the educational organizations they attend. This includes the student's capacity to finish assignments, achieve passing grades on examinations, actively participate in classes, and graduate on schedule. It is common practice to evaluate students' academic achievement using a variety of measures, such as scores, grade point averages, and readmission rates.

Evaluation of scholastic achievement in higher education is a multi-step procedure that considers a variety of different aspects. Grades are the most prevalent and widely used statistic that is used to evaluate scholastic achievement. The purpose of grades is to evaluate a student's level of subject knowledge as well as their capacity to achieve the scholastic goals outlined by their educational establishment. On the other hand, scores do not offer a comprehensive depiction of a student's achievement in the classroom. Several different measures are used to evaluate scholastic achievement, including grade point average (GPA) and readmission rates. The GPA provides a more complete portrait of a student's academic achievement as it is an accumulated assessment of the scores obtained throughout their academic career. On the other hand, retention rates are used to determine the proportion of students who remain enrolled in each program or institution after their initial enrollment.

2.3 FACTORS IMPACTING STUDENT PERFORMANCE

The academic achievement of a student in higher education is impacted by a number of different variables. Individual factors, institutional factors, and environmental factors are the three predominant categories that have been used to classify the set of academic factors impacting student performance academically.

2.3.1 Variables Associated with Student Academic Performance

The student's academic performance has been affected by various variables which include the following:

- Individual aspects to consider are a student's level of drive and aptitude, as well as their study routines and practices. Students are said to be motivated when they have a strong desire to learn new things and excel in their scholastic endeavors. Ability is a term used to describe the reasoning and scholastic abilities of a student. These abilities include the student's intelligence as well as their previous information. The student's strategy for learning, including their ability to handle their time effectively, maintain organization, and make effective use of study techniques is referred to as their study routines.

- Institutional variables consist of the competence of instruction, as well as the resources and support services that are made available by the organization. The efficiency with which teachers encourage student involvement and learning is what we mean when we talk about the quality of the instruction. The term "resources" is used to allude to the availability of various scholarly resources, including libraries, technological resources, and scientific facilities. The availability of academic as well as non-academic assistance is referred to as support services. Examples of support services include education, counseling, and employment services.
- A student's societal and cultural surroundings are examples of environmental variables that can influence their education. This encompasses the student's family history, societal support network, as well as cultural standards and beliefs that they were raised with.

It is possible for strategies that seek to improve student outcomes in higher education to be informed by a better understanding of the variables that influence academic achievement. In this regard workshops teaching study skills and consulting services are two examples of the types of individualized assistance that institutions like universities and colleges can offer. The quality of instruction and the availability of resources are other aspects that institutions can work to improve to strengthen the administrative variables that influence scholastic achievements. In addition, educational institutions could develop a welcoming atmosphere that considers and values the varied experiences and histories of their student body. This is accomplished by giving students the opportunity to participate in a variety of cultural and social activities and by cultivating an atmosphere that respects and welcomes people of all backgrounds.

In conclusion, scholastic performance is a multifaceted occurrence that is affected by a wide range of elements that contribute to a student's level of achievement in higher education. There are a variety of measures that go into measuring scholastic achievements, such as scores, grade point averages, and readmission rates. It is possible for strategies that seek to improve student outcomes in higher education to be informed by a better understanding of the variables that influence scholastic achievement. The provision of support services, improvements in the quality of instruction and materials, and the creation of a welcoming environment that acknowledges and respects the varied experiences and backgrounds of students are all things that institutions can do for different scenarios.

2.3.2 Related Theories and Models of Academic Performance

The in-depth understanding of the variables contributing to a student's success in higher education has been facilitated through the application of various academic achievement theories and models. This section introduces the various academic performance theories and models that are connected to one another, including their descriptions, important concepts, and consequences for practice. The theories are depicted in Figure 2.1.

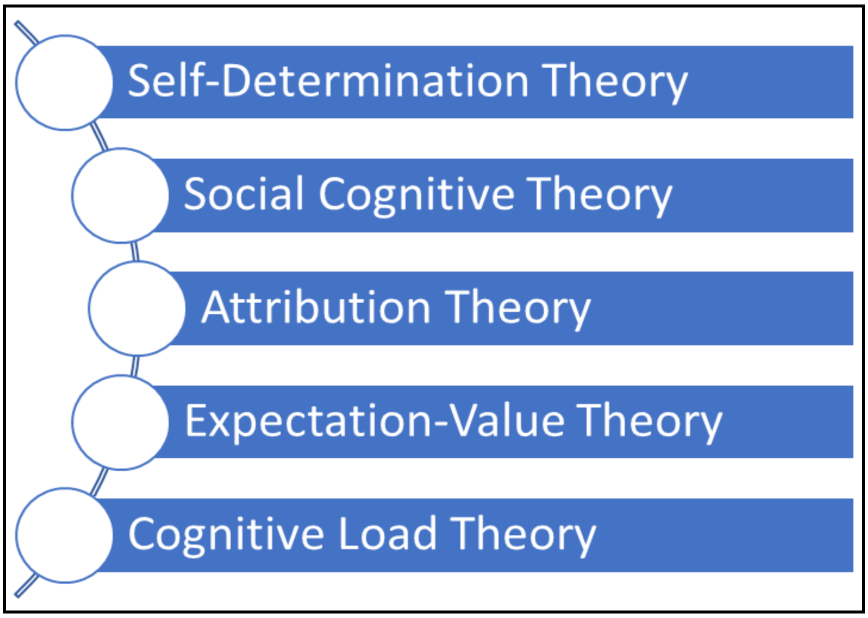

Figure 2.1 Theories for Academic Performance Prediction

2.3.3 Theories to Predict Academic Performance

Theory 1: Self-Determination Theory

The theory of self-determination postulates that people have inherent psychological needs for autonomy, competence, and relatedness and that these needs have to be satisfied for people to experience optimum levels of motivation and well-being in their lives. The self-determination theory, when applied to the realm of academic achievement, proposes that students who have a strong sense of autonomy, a sense of competence, and a sense of connection to others have a greater likelihood of actively participating in academic duties and achieving success.

Implications for classroom practice include teachers encouraging students' sense of agency by giving them more opportunities to make their own decisions and exercise control over their education. In addition, educators can assist students in the development of a sense of competence by assigning them tasks that are difficult but not insurmountable and by providing feedback that acknowledges the students' advancement. Finally, teachers have the ability to cultivate a positive atmosphere in the classroom that encourages students to form positive relationships with one another as well as with the instructors.

Theory 2: Social Cognitive Theory

The social cognitive theory emphasizes the part that social as well as cognitive variables play in the formation of human behavior. According to social cognitive theory, students' opinions about their capabilities (referred to as self-efficacy) and the worth of academic duties (referred to as outcome expectancies) influence both their level of involvement and their level of accomplishment when it comes to their academic performance.

The implications for practice are that teachers can boost their students' sense of self-efficacy by giving the students opportunities to succeed and by demonstrating productive learning techniques themselves. The educators can also help students have higher expectations for their outcomes by emphasizing the significance and importance of scholastic tasks and by providing opportunities for students to implement what they have learned in settings that are taken from the real world.

Theory 3: Attribution Theory

According to the attribution theory, people always try to figure out why things that happen to them, such as their scholastic achievement or failure, are the way that they are. According to the attribution theory, which is applied to the realm of scholastic performance, the reasons that students give for their achievements—whether they succeed or fail—influence their motivation and, as a result, their performance.

Implications for classroom practice Teachers can support the development of a growth mindset in their students by highlighting the importance of effort and perseverance in achieving scholastic achievement and by offering students opportunities to learn from their missteps and the experiences of others. Giving students opportunities for options and self-direction within the context of the learning process can also assist teachers in assisting students in the development of a feeling of personal control over the scholastic outcomes of their studies.

Theory 4: Expectation-Value Theory

The expectation-value theory proposes that individuals' motivation to participate in a task is influenced by both the individual's anticipation for achievement and the individual's perception of the value of the task. In the context of academic performance, the expectancy-value theory proposes that students' motivation and involvement are influenced by their beliefs about their ability to succeed in academic tasks and the perceived value of those tasks. In other words, students' beliefs about their ability to succeed in academic tasks and the perceived value of those tasks are related to academic performance.

Implications for classroom practice: The teachers can boost their students' levels of enthusiasm and involvement by offering them more chances to succeed, as well as by acknowledging and rewarding their progress. Teachers can also assist students in developing a sense of the worth of scholastic tasks by emphasizing the tasks' connection and importance to the students' own personal and professional objectives.

Theory 5: Cognitive Load Theory

According to the cognitive load theory, the quantity of information that is presented to learners, the complexity of that information, and the cognitive processes that are necessary to process that information all affect learning. In the setting of academic achievement, the cognitive load theory proposes that the design of instructional materials and the cognitive processes necessary to accomplish academic activities affect the capacity of students to process and remember information sets.

Implications for practice include the possibility for educators to create instructional materials that both minimize the amount of unnecessary cognitive load (by, for example, streamlining language and decreasing visual confusion) and optimize the amount of cognitive load that is directly relevant to the content being taught. (e.g., by providing opportunities for active learning and problem-solving). In addition, teachers can assist students in the development of efficient strategies for the management of cognitive burdens. These strategies may include the chunking of information and the utilization of exterior tools.

In conclusion, theories, and models of academic achievement [28] offer extremely helpful insights into the variables that influence a student's success in higher education scenarios. The theory of self-determination emphasizes the significance of sovereignty in different scenarios.

2.4 REVIEW OF EXISTING MACHINE LEARNING MODELS FOR PREDICTING STUDENT ACADEMIC PERFORMANCE

The challenge of accurately predicting the scholastic achievement of students is a difficult one for teachers and educational specialists. The students' cognitive capacities, motivation, learning techniques, and the socio-cultural context in which they were raised are just some of the variables that contribute to their scholastic achievement. In recent years, methods from the field of machine learning have been applied to the development of prognostic models of the scholastic achievement of students.

2.4.1 Different Machine Learning Models for Predicting Student Academic Performance

The following section discusses the various techniques currently available for predicting the academic achievement of students using machine learning.

Model 1: Regression Models

In educational research, regression models are frequently used to make predictions about the scholastic achievement of students. These models make use of multiple regression analysis to determine the factors that influence scholastic achievement. Some of these factors include a person's cognitive capacities, motivation, learning techniques, and socio-cultural upbringing. It is possible to use regression models to make predictions about a variety of scholastic outcomes, such as grades, exam results, and graduation rates [29].

Implications for the practice of education: The educators can make use of regression models to identify students who are at risk of failing academically and to develop tailored interventions to enhance those students' academic achievement. Evaluation of the efficacy of educational programs and interventions can also be accomplished with the assistance of regression models.

Model 2: Decision Tree

The decision tree is a specific kind of model for machine learning that has been utilized to make forecasts regarding the scholastic achievement of students. The values of various indicators are used as the basis for decision tree hierarchical structures, which are then used to divide the data into subgroups. The decision tree is a useful tool for determining which aspects of a student's cognitive abilities, motivation, and learning techniques are the most significant in terms of predicting that student's scholastic success [30].

Implications for the practice of education: The educators can use decision trees to develop individualized learning plans for individual students based on the unique combination of academic success indicators that each student possesses. These plans have been based on the student's academic history, as well as other factors. The patterns of academic performance among groups of students can also be used to develop tailored interventions to enhance academic performance for those groups, which has been done with the help of decision trees.

Model 3: Neural Networks

The organization and operation of the human brain served as the basis for the development of a specific category of algorithms for use in machine learning called neural networks. By analyzing large amounts of student data and looking for recurring patterns, neural networks have been trained to make accurate predictions about students' scholastic achievement. When applied to large amounts of data, neural networks can recognize intricate patterns that may not be discernible when using more conventional statistical techniques [31].

Implications for the practice of education: Educators can use neural networks to develop individualized learning strategies for individual students based on the unique combination of factors that each student possesses that predicts their success in academics. Additionally, neural networks have been used to recognize patterns of academic performance among groups of students and to develop tailored interventions to enhance academic performance for those groups. Both applications have been accomplished through the use of data collected from the networks.

Model 4: Bayesian Networks

Bayesian networks are a specific kind of technique for machine learning that has been utilized to simulate the connections that exist between the variables that make up complicated systems. By simulating the complicated relationships between students' cognitive abilities, motivation, learning techniques, and socio-cultural backgrounds, Bayesian networks have been utilized to make accurate predictions regarding the scholastic achievement of students [32].

Implications for the practice of education: The Bayesian network has been used by educators to develop individualized learning strategies for individual students based on the unique combination of academic success indicators that each student possesses. Bayesian networks have the potential to revolutionize education. Bayesian networks can also be used to recognize patterns of academic performance among groups of students and to develop tailored interventions to enhance academic performance for those groups. Both applications have been accomplished through the use of data collected from Bayesian networks.

Model 5: Support Vector Machines (SVM)

Support vector machines are an example of a form of algorithm that belongs to the field of machine learning. These machines have been utilized to categorize data into a variety of categories. The academic performance of students turned out to be predicted with the help of support vector machines by grouping students into distinct categories according to the factors that best indicate their potential for academic achievements [33].

Implications for the practice of education: Educators can use support vector machines to develop individualized learning plans for individual students. These plans would be based on the students' one-of-a-kind combinations of factors that forecast scholastic achievement. It is also possible to use support vector machines to recognize patterns of academic performance among groups of students and to develop tailored interventions to enhance academic performance for those groups using the identified patterns as a guide for different use cases.

In summing up, the application of methods that utilize machine learning to forecast the scholastic achievement of students has become an increasingly prevalent method. When it comes to forecasting scholastic achievement, some of the techniques that are utilized the most frequently include regression models, decision trees, neural networks, Bayesian networks, and support vector machines. Each of these methods has a set of benefits and drawbacks, and educational practitioners and researchers need to consider which method is best suited to their research question and the data samples they will be analyzing under real-time scenarios.

2.4.2 Consideration for Using Machine Learning Models

Although there is evidence to suggest that machine learning techniques can accurately forecast the scholastic achievement of students, there are still a number of issues that need to be resolved. The following is a list of some of the primary restrictions placed on contemporary methods:

1. Data Scarcity Concern

Machine learning techniques necessitate the use of substantial quantities of data in order to successfully build and evaluate models. On the other hand, the data that we have on the academic achievement of students is frequently insufficient and might not include all of the important characteristics that are required to accurately forecast academic outcomes. Because of this, the models that are produced might not be accurate or generalizable.

2. Biased Data Concerns

It is possible for the data that is used to teach machine learning algorithms to contain biases, which then causes the models to represent the biases that are present in the data. For instance, if the data used to model academic performance is skewed toward one demographic group, then the model that is produced may not effectively forecast academic performance for other demographic groups.

3. Data Interpretation Concerns

The algorithms used in machine learning are frequently referred to as "black boxes" because it became challenging to comprehend and understand what they are doing. Because of this, it turned out to be difficult for instructors to make sense of the findings and figure out how to use the information to enhance the scholastic achievement of their students.

4. Data Quality Concerns

The algorithms that are used in machine learning are intended to work with quantifiable data, such as exam results and biographical information. However, incorporating qualitative data became challenging. On the other hand, qualitative statistics, such as a student's level of enthusiasm and their approach to learning, are also capable of playing an important part in scholastic achievement. The process of incorporating this data into machine learning algorithms has been difficult.

5. Ethical Considerations

There is potential for machine learning algorithms to give rise to ethical considerations, specifically concerning privacy and impartiality. For instance, if models are used to predict academic performance, there is a possibility that these models may be used to unjustly target certain groups of students or to stigmatize students who are having difficulty intellectually. Although machine learning algorithms exhibit potential in accurately forecasting students' academic performance, it is imperative to recognize and tackle various limitations. These constraints include issues related to data availability, bias, interpretability challenges, the intricacies of incorporating qualitative data, and the indispensable requirement for human input. Educators and researchers in the field of education must conscientiously take these limitations into account when utilizing machine learning algorithms to predict students' academic success.

6. Absence of Human Input Concern

Even though machine learning algorithms have been effective instruments for forecasting the scholastic achievement of students, these programs are not a suitable replacement for the wisdom and experience of humans. Educators and educational researchers have a responsibility to carefully consider the constraints of machine learning algorithms and to incorporate their knowledge, experience, and judgment into the process of interpreting the findings of the study.

In conclusion, although there is hope that machine learning algorithms will be able to accurately forecast the scholastic achievement of students, there are still several restrictions that need to be addressed. These include restrictions on the availability of data, prejudice, interpretability, difficulty in incorporating qualitative data, ethical considerations, and the requirement for human input. Additionally, there is a need for human input. When attempting to use machine learning algorithms to make predictions about students' scholastic achievement, it is imperative that teachers and researchers in the field of education consider these constraints.

2.5 EXISTING ACADEMIC PERFORMANCE PREDICTION MODELS BASED ON MACHINE LEARNING TECHNIQUES

The methodologies of machine learning and data mining have seen widespread application in a variety of disciplines, including education. The ability to accurately forecast the scholastic achievement of students is one of the most significant implementations of machine learning in the field of education. The purpose of forecasting student performance is to recognize students who are at risk of failing their classes or slipping out of school so that appropriate support has been offered to them at the appropriate moment to help them achieve. In this overview, we will talk about a variety of different machine learning and data mining approaches that have been utilized in the past to determine the achievement of students.

The procedure of linear regression is one that is frequently utilized to forecast student achievement. The presumption that there is a linear relationship between the dependent variable (student performance) and the independent variables (such as demographics, previous scholastic achievement, etc.) underlies linear regression models. The dependent variable in such a model is the student's performance. Models of linear regression are straightforward to understand and have the potential to offer invaluable insights into the variables that influence the performance of students. However, there are some drawbacks associated with linear regression models. For instance, they make the assumption that the relationship between the variables being studied (the dependent variable) and the independent variable (the independent variable) is linear, even though this might not be the case in all instances. In addition, linear regression models are not very good at dealing with irregular relationships because of their inherent linearity.

Logistic regression is yet another well-known method for estimating how students have performed in each class. Logistic regression is utilized in situations in which the dependent variable possesses a binary data type. (i.e., pass or fail). Models of logistic regression determine the likelihood of success (that is, passing) as a function of independent factors such as demographics, previous scholastic achievement, and other such factors. Logistic regression models are not only simple to understand, but they also have the potential to provide extremely helpful insights into the variables that contribute to the success of students. On the other hand, logistic regression models suffer from some drawbacks. For instance, they assume that the relationship between the variables being studied (the dependent variable) and the independent variable (the independent variable) is linear, even though this might not be the case in all instances. In addition, logistic regression models are not very good at dealing with irregular relationships because of their inherent linearity [34].

When it comes to forecasting student achievement, decision trees are a common and popular technique used in machine learning. Decision trees are representations of judgments and the potential outcomes of those decisions that take the form of branches. Each node in the tree represents a decision that is based on a specific variable, and each branch represents a consequence that could occur because of that decision. Decision trees are simple to understand and can offer insightful information about the variables that influence the performance of students. Nevertheless, decision trees suffer from some drawbacks. For instance, they may be prone to overfitting, which can lead to models that cannot be generalized to new data and may result in inaccurate predictions.

The method of machine learning known as Deep Forest is founded on the concept of decision trees. Deep Forest is a type of ensemble learning that integrates numerous decision trees into a single model to increase the precision of predictions [35,36]. Deep Forest models are resistant to overfitting and have the ability to work with data of a high dimension. The interpretation of Deep Forest models is straightforward, and the insights they provide into the variables that influence student achievement have become extremely beneficial [37]. Nevertheless, Deep Forest models suffer from several drawbacks. For instance, they may require a significant amount of calculation, which may render them impracticable for use with massive databases.

The architecture of the human brain served as the basis for the development of a specific approach to the field of machine learning known as neural networks. Information is handled by multiple levels of cells, or neurons, that are connected and create neural networks. Neural networks can understand intricate patterns of correlations between different variables and can then be put to use to make forecasts. Neural networks are adaptable and have the capacity to process material with a high dimension. Neural networks, on the other hand, are tricky to understand and frequently call for a significant quantity of data in order to train successfully for different scenarios [38-40].

Support vector machines, also known as SVMs, are a method of machine learning that has been utilized for categorization and regression analysis. SVMs locate the hyperplane that distinguishes the data into various classifications or accurately forecasts the value of the variable that is contingent on the data. SVMs can deal with nonlinearities in the data and are efficient when managing high-dimensional datasets. SVMs, on the other hand, are challenging to understand and often call for a significant quantity of data to train successfully for different scenarios.

The subset of EDM, which deals with the extensive data obtained from an educational system, is the prediction of students' academic achievement. The author insisted that the goal of EDM is to use this data to derive useful information that will benefit the stakeholders in the education system. Utilizing actual data gathered, the proposed study employs a variety of machine learning approaches i.e. Multiple Linear Regression, Random Forest, Logistic Regression, Decision Tree, Naïve Bayes, and Artificial Neural Network with Principal Component Analysis to predict the student's academic performance (comprising the academic history and personal habits of the students). A comparison of ML methods on several evaluation metrics has also been done. The proposed model uses and compares with different ML approaches. Additionally, the evaluation indicators show that with an R2 score of 0.933 for regression, RF regression performed best. For classification, ANN with PCA scored best with an F1 score of 0.842 [41].

The author enlightened that the innovative approaches, strategies, and applications from EDM play an essential role in improving the learning environment of students. The most recent research offers useful methods for assessing the learning environment of students by analyzing their academic data using machine learning techniques. The environment wherein modern academic institution’s function is extremely complex and competitive. The author enlightened that academic institutions encounter difficulties with performance evaluation, high-quality instruction, performance assessment methods, and future actions. The author proposed machine learning techniques for analyzing the huge amount of academic data collected from students. Overall, this evaluation has become successful in accomplishing the objective of improving students' academic performance who are at academic risk. The author reviewed that the research indicated corrective measures to give students, teachers, and educators timely feedback to resolve performance-related problems [42].

The author explained that the prediction of student performance is the main objective of higher education institutions and is a prominent area of research. Predicting the academic performance of students assists the instructors in identifying the students who are at risk of dropping out because of their poor performance and provides special counseling sessions so that they can perform well in their academics. The author explored that various machine learning techniques have been used to develop prediction models in the educational sector for discovering and exploring meaningful hidden patterns from academic data. The author analyzed the performance of various supervised machine learning algorithms, comprising Decision Tree, Naive Bayes, Logistic Regression, Support Vector Machine, K-Nearest Neighbor, Sequential Minimal Optimization, and Neural Network. To predict the performance of students in their final exams, the author trained a model using datasets provided by courses in the bachelor study programs of the College of Computer Science and Information Technology, University of Basra, for the academic years 2017–2018 and 2018–2019. From the results, the author concluded that the logistic regression classifier had the most effective predicted results in terms of accuracy with (68.7% for passed and 88.8% for failed) [43].

The author explored that EDM is a viable method for identifying hidden patterns and predicting student academic performance. In this study, the author proposed a machine learning-based model to predict the grades of undergraduate students in their final exams by considering their midterm marks as input data. For predicting the students' academic performance in their end-term exams the author compared the performances of various machine learning algorithms such as random forests, nearest neighbour, support vector machines, logistic regression, Naive Bayes, and k-nearest neighbour.

The authors' observations reveal that the proposed model produced results with an accuracy of 70-75 %. For predicting academic performance, the author focused majorly on midterm exam grades, departmental data, and faculty data. The proposed model identifies the most efficient machine-learning techniques that contribute to the early prediction of students who are at high risk of academic failure [44].

The authors identified that one of the primary issues is predicting the students' academic performance before their final exams so that proactive actions can be taken to assist the students in improving their performance and hence minimizing dropouts. The author provides the readers with a thorough knowledge of how to predict student academic performance and compares approximately 260 studies from the last 20 years by addressing the following concerns:

1) The significant factors impacting predicting student performance.

2) The different data mining techniques for prediction and feature selection.

3) frequently used data mining tools.

The author's results demonstrate that the WEKA is an emerging technique for predicting student academic performance, while ANN and Random Forest are the most commonly adopted data mining algorithms. The research study highlights that the students' academic records and demographic factors were the best for predicting performance. The authors proved that the irrelevant attributes in the dataset lowered the prediction accuracy and increased the processing time of the model. The author discovered that the most prevalent input type, which produced the best prediction results, has students' grades or CGPA [44].

The author's objective of the study is to develop data mining models that can predict student academic performance by considering their inter-personal, senior secondary, and university performance factors. The data about students admitted to the university over three years has been utilized for the research. The dataset is implemented using well-known data mining classification algorithms, including a rule learner, a decision tree classifier, a neural network, and a Nearest Neighbor classifier. The predicted results show that the Neural Network model achieved 73.59% accuracy which is the highest as compared to Decision Tree i.e., 72.74%, and for the k-NN model it's 70.49% [45].

The author collected the data from the University of Jordan, for performance prediction. In the dataset, students from various countries have been considered for analysis and prediction purposes. In addition to employing individual machine-learning techniques, the authors also implemented ensemble methods for prediction. As per the generated results, the authors concluded that the best results had been produced by Decision Tree. In addition to the academic features the researchers also on the behavioral traits and concluded that the behavioral traits help in improving the accuracy of predicted results [46].

Techniques such as machine learning and data mining have demonstrated significant potential around forecasting the scholastic achievement of students. These methods have the potential to offer extremely helpful insights into the variables that have been used to enhance performance in real-time situations. The review of existing studies and techniques used along with the core objective of each study is depicted in Table 2.1.

Table 2.1: Review of Existing Machine Learning Techniques Used for Student Performance Prediction

Existing Studies	Technique(s) Used	Core Objective	Accuracy %
Huang, S., & Fang, N.,2013. [39]	LinR, NN, SVM	The midterm dynamics has been used to predict the academic performance of students.	81–91
Marbouti, F. et al., 2016. [40]	LogR, NN, SVM, DT, NB, KNN	Standardized grading has been used to identify the course-wise students who are at academic risk.	92.6
Costa, et al.,2017. [41]	NN, SVM, DT, NB	To recognize the students at risk in their academics and are likely to discontinue in two different systems i.e., on-campus and distance education.	92

Existing Studies	**Technique(s) Used**	**Core Objective**	**Accuracy %**
Gray, et al. 2014. [42]	LogR, NN, SVM, DT, NB, KNN	The main objective is to identify the students who are at risk of academically failing in their first year of study.	86.42
Laugerman, et al., 2015. [43]	Logistic Regression Model	Identification of factors influencing the academic success of engineering students using post-hoc graduation data.	94
Hoffait, A. S., & Schyns, M., 2017. [44]	LogR, NN, RF	To classify the students into slow and fast learners and identify the students who are significantly facing difficulties in completing their initial academic year on time.	86
Strecht, et al., 2015. [45]	LinR, NN, SVM, DT, NB, KNN, AdaBoost	To predict the academic performance of the student course-wise. The students received the feedback in terms of course success or failure along with the course grade.	85
Aluko, et al.,2016. [46]	Linear Discriminant Analysis, KNN	To predict the student academic performance in the domain of architecture based on the historical academic data given by the students.	73.33
Miguéis, et al., 2018. [50]	Decision Trees, Support Vector Machines, Naive Bayes, Bagged Trees, and Boosted Trees	It helps in the classification and identification of students who are at academic risk based on their performance in exams. Proper counseling and special academic inputs have been given to academically weak students so that they can improvise their results and hence performance.	>95

Existing Studies	Technique(s) Used	Core Objective	Accuracy %
Ha, D. T., et al., 2020. [51]	Decision Trees,MLP,SVM,RF,Naïve Bayes	To identify and classify the students academically weak students based on their performance in their exams.	86.19
Umer, et al., 2017. [52]	Logistic Regression, Neural Networks, Random Forests.	The academic data of the first year have been considered to identify the difficulties faced by the students in their learning and identification of factors responsible for their academic downfall.	87
Lu, et al., 2018. [53]	Decision Trees, Rule and Fuzzy Rule Induction Methods, and Neural Networks.	To predict the academic score of the students in their final exams. The scores are predicted based on the historical academic data used to train the model.	76
Thai-Nghe,et al., 2010. [54]	Logistic/Linear Regression, Matrix Factorization	To create an intelligent teaching system for the students by considering the existing institutional academic data and the challenges faced by the students.	Not Available
Huang, S., & Fang, N. 2013. [55]	Linear Regression, Neural Networks, Support Vector Machines	To predict the academic performance of students based on their mid-term marks so that suggested action has been taken for the weak performers so that they can perform well in their end-term exams.	81-91
Imran, et al., 2019. [56]	Neural Networks, Random Forests, and Decision Tree	To predict the student' academic performance in their initial carrier so that they can improvise their performance and compete their course on time.	95.78

Existing Studies	Technique(s) Used	Core Objective	Accuracy %
Strecht, et al.,2015. [57]	Linear regression, neural networks, support vector machines, decision trees, naive Bayes, k-nearest neighbor	To improve student academic performance the education system, recommends different courses based on historical data that help in the performance improvement of students.	85
Kamal, P., & Ahuja, S., 2019. [58]	Decision tree, Gradient boost algorithm, and Naïve Bayes	To classify the students into strong and weak learners special attention has been given to the weak performers to improvise their learning.	89
Skrbinjek, V., & Dermol, V.,2019. [59]	SVM and Naïve Bayes	The student academic performance has been predicted using naive bayes and svm and then the predicted results of both algorithms have been compared and concluded that svm is generating more accurate results.	92.3
Tripathi, et. al., [60]	K-Nearest Neighbor, Naïve Bayes, Decision Tree, and Logistic Regression	The primary objective is to predict the academic achievement of students and the identification of factors responsible for lowering their performance in academics.	82.15
Yaacob, et al., 2019. [61]	Decision Tree	A decision tree has been used for assessing the academic performance of learners. The prediction results have been then used for providing academic recommendations to weak performers thus improving the retention rate of educational institutions.	91

Existing Studies	Technique(s) Used	Core Objective	Accuracy %
Mueen, A., Zafar, B., & Manzoor, U. (2016). [62]	Naïve Bayes, Neural Network, and Decision Tree	Predicting and analyzing the student's academic performance by using different machine learning techniques is the prime objective of this study. The secondary data has been used for the prediction of results.	86
Roy, S., & Garg, A. (2017). [63]	Random Forest, Neural Networks, SVMs, and Regression Techniques	The education data mining has been used for the identification of weak performers. Additionally, this study aids in singling out the various features liable for running down the academic performance of students.	96
Hashim AS, Awadh WA, Hamoud AK. (2020). [64]	Decision Tree, Naive Bayes, Logistic Regression, Support Vector Machine, K-Nearest Neighbor, Sequential Minimal Optimization, and Neural Network	To predict the performance of students in their final exams, the author trained a model using datasets provided by courses in the bachelor study programs of the College of Computer Science and Information Technology, University of Basra, for the academic years 2017–2018 and 2018–2019.	68.7 for passed and 88.8 for failed
Yağcı, M. (2022). [65]	Nearest Neighbour, Support Vector Machines, Logistic Regression, Naive Bayes, and K-Nearest Neighbour	the author proposed a machine learning-based model to predict the grades of undergraduate students in their final exams by considering their midterm marks as input data. For predicting the students' academic performance in their end-term exams.	70-75

Existing Studies	Technique(s) Used	Core Objective	Accuracy %
Kabakchieva D. (2012). [66]	Rule Learner, A Decision Tree Classifier, A Neural Network, and a Nearest Neighbor	To develop data mining models that can predict student academic performance by considering their inter-personal, senior secondary, and university performance factors. The data about students admitted to the university over three years has been utilized for the research.	73.59

2.6 NOVEL APPROACHES FOR PREDICTING STUDENT ACADEMIC PERFORMANCE

Predicting the scholastic achievement of students is an essential task in education because it enables teachers to recognize pupils who may be at risk of failing their classes or leaving school altogether. In recent years, there has been a growing interest in using novel approaches for forecasting student achievement. Some examples of these novel approaches include deep learning, natural language processing, and transfer learning. In this overview, we are going to talk about some of the innovative methods that have been used in the past for forecasting the scholastic achievement of students.

Deep learning is a subfield of machine learning that involves the process of learning representations of data through the utilization of neural networks that contain multiple levels. In terms of forecasting student achievement, deep learning has demonstrated a great deal of potential. Deep learning offers several benefits, one of which is the capability to discover intricate relationships between different variables. Deep learning models have been used for forecasting student performance based on a variety of inputs including demographics, previous scholastic achievement, activity on social media, and online learning behavior. These models have also been used to improve online learning. For instance, [67] conducted research in which they used deep learning to forecast student achievement based on their participation in social media. According to the findings of the research, the deep learning model performed significantly better than conventional machine learning models when it came to forecasting student achievement.

Natural language processing, or NLP for short, is a subfield of artificial intelligence that examines how computers and people communicate with one another using language. Using written data such as articles, discussion forum entries, and online conversation records, natural language processing (NLP) has been applied to the task of forecasting student achievement. The vocabulary used by students has been analyzed using NLP models, which can then be used to discover patterns that are correlated with either scholastic achievement or failure. For instance, [68] conducted research in which they used natural language processing (NLP) to forecast student achievement based on conversation thread entries. According to the findings of the research, NLP models had a high level of accuracy when predicting student achievement based on the language that has been used in conversation thread entries.

Transfer learning is a method of machine learning that includes making use of models that have been pre-trained to enhance the overall performance of a model when applied to a different task. Transfer learning is a method that has been used to forecast student performance by utilizing previously learned models on tasks that are like the task at hand. Some examples of these tasks include language translation, picture recognition, and natural language comprehension.

Analysis of student behavior in online learning environments is the focus of the innovative method known as behavioral analysis. This method is used to forecast students' overall academic achievement. The amount of time spent on tasks, the number of attempts made on assignments, and the regularity of interactions with classmates and instructors are all examples of the types of data that have been analyzed by behavioral analysis models. Patterns in student behavior that are correlated with scholastic success or failure have been uncovered using behavioral analysis models. For instance, in research that has been conducted by [69], behavioral analysis has been utilized to make predictions regarding student achievement in an online course. According to the findings of the research, behavioral analysis models were able to effectively forecast student achievement based on how students interacted with online learning environments.

Deep learning, natural language processing, transfer learning, and behavioral analysis are some examples of innovative methods that have shown great potential in the area of forecasting students' scholastic achievement. These different techniques may give illuminating information about the variables that contribute to the success or failing of students. However, these methods also have drawbacks, such as the requirement for a substantial quantity of data, the complexity of the models, and the challenge of deciphering the findings. These techniques need to be developed further, and further research is required to establish whether or not they are successful in forecasting the scholastic achievement of students [70].

Chapter - 3

IPERFORM MODEL

3.1 INTRODUCTION

The process of data mining is used for extracting diverse knowledge generated from transactional data which is utilized by various organizations for decision-making. The data mining techniques are applicable in a variety of fields such as in financial, academic institutions, scientific programs, and a variety of other sectors. The data mining process involves a variety of techniques, including classification, prediction, outlier analysis, and ensemble. Classification is the process used to predict the object class whose label is unknown using different prediction models. These models use historical data to train the model development and are used for the prediction of classes that are missing in the testing data. The predicted results have been used by various organizations for effective decision-making. The outliers are defined as the data objects that do not comply with the general behavior or model of the available data. The ensemble approach in data mining refers to the process of incorporating multiple classifiers to create a single classifier that boosts the prediction accuracy in comparison to the accuracy generated by individual classifiers by lowering the variance and bias [71]. This chapter focuses on the usage of ensemble techniques in the field of education that leads to the concept of education data mining (EDM) or ADM.

ADM has obtained astounding inquisitiveness in recent years. The need for the analysis and assessment of the factors impacting the academic performance of students has embellished the demand for ADM or EDM. Significantly, such factors can include student academic performance measured in terms of final grades obtained, course attendance, mid-assessment marks, etc. [72]. ADM plays a pivotal role in analyzing student performance based on the above-said factors and thereby classifying them into fast and slow learners. Additionally, ADM can also aid in providing subtle suggestions and recommendations for both the instructors as well as the students to improvise their performance [73]. This can involve processes such as academic performance prediction and academic performance recommendations. Both processes are essential for every educational institution as their reputation is centered upon the academic accomplishments of students. The primary goal of academic performance prediction of learners is the identification of students at risk in the initial stage of their career. This identification helps the instructor to analyze the factors affecting the performance such that corrective actions can be taken for the students at risk of lower achievement levels. Moreover, the timely analysis of weak performers benefits academic institutions in increasing their retention rate [74].

The academic performance of students is predicted using different supervised learning techniques such as classification and prediction. LA plays a very significant role in the field of education. The motivation for using LA by academic institutions is to analyze the patterns obtained from the educational data after prediction. So, after the academic performance prediction, LA in association with ADM is used to generate effective results that lead to the categorization of different types of students [75].

The chapter proposes a model that serves as an alarming structure for educational organizations. The proposed model can be used by the students to discover and concentrate on their disconcerting subjects while the faculties can focus on improving their learning strategies toward such students. Currently, many machine learning algorithms are available for envisaging student educational performance and ADM [76]. The proposed model is also an ensemble machine learning-based model that predicts the student's academic performance using an ensemble of machine learning algorithms, Decision Tree, Naïve Bayes, and K-Nearest Neighbor. For performance prediction, the records have been collected from the academic institution which is then pre-processed to eliminate anomalies so that only the data which is helpful for the analysis purpose is anomalies free. The cleaned data is then applied to the model and thereafter produces the predicted results.

3.2 BASIC ENSEMBLE APPROACH FOR PREDICTION

The ensemble model is a generic machine learning meta-model that aims to improve prediction accuracy by combining the output of various prediction models. The primary goal of creating an ensemble model helps in the production of more accurate results as compared to the accuracy of results produced by individual classifiers. The proposed model uses an ensemble of heterogeneous classifiers. The ensemble model proposed here accepts the output from multiple classifiers such as decision tree, Naïve Bayes, and K-NN. The proposed ensemble combines the output of heterogeneous classifies using a voting approach which resultantly produces the final prediction results. The idea of the ensemble approach works if and only if all the selected classifiers produce different class labels rather than agreeing on the same decision.

The IPerform model has been designed to predict student academic performance using an ensemble of machine-learning algorithms. The primary objective of designing an ensemble model is that every selected classifier must be complementary to each other in the context of a judgment so that further accuracy can be achieved.

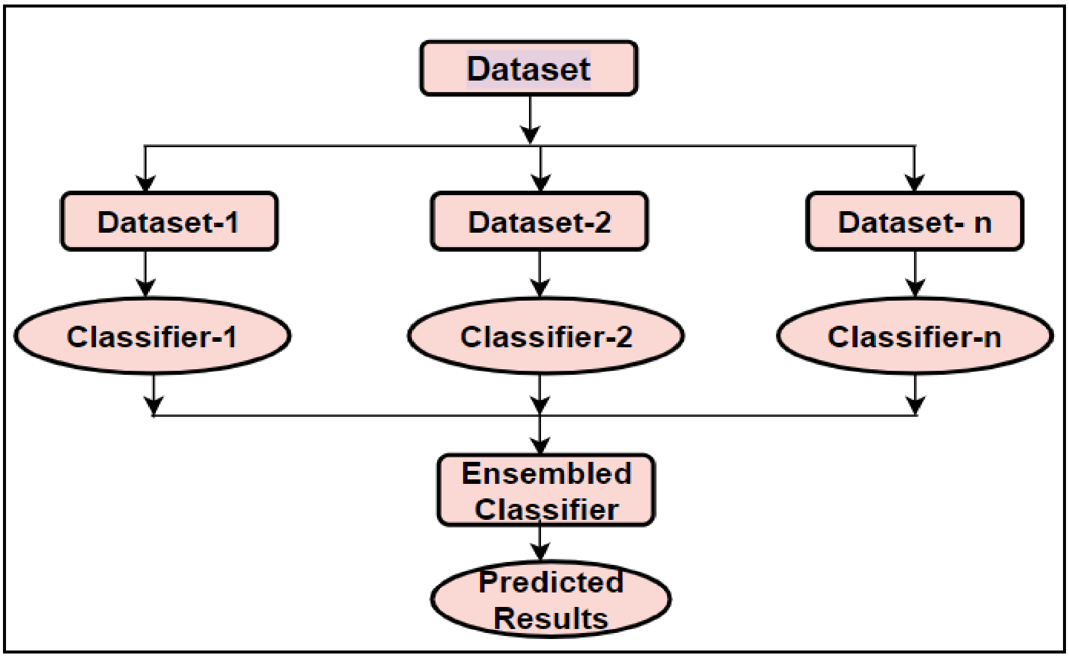

Figure 3.1: Basic Ensemble Approach for Prediction

Figure 3.1 depicts the flow of the ensemble method as applied in the IPerform model. The IPerform model performs classification of the students based on their academic performance considering their marks in the courses inclusive of their attendance in each course. The data for classification has been collected from sources such as

Google Forms and a designed interface. Certain attributes generate irrelevant values such as incomplete data, duplicate data, and naming identification problems and hence have no participation in the classification process. Thus, such irrelevant attributes were stricken out of the classification process the use of these attributes could have increased the classification errors and complexity of the selected algorithm. Conclusively, this helped in making the predictions more accurate. The model intends to compute the student's academic performance (in terms of Cumulative Grade Points) and achieve an early separation of learners segregating them into slow and fast learners based upon their educational performance.

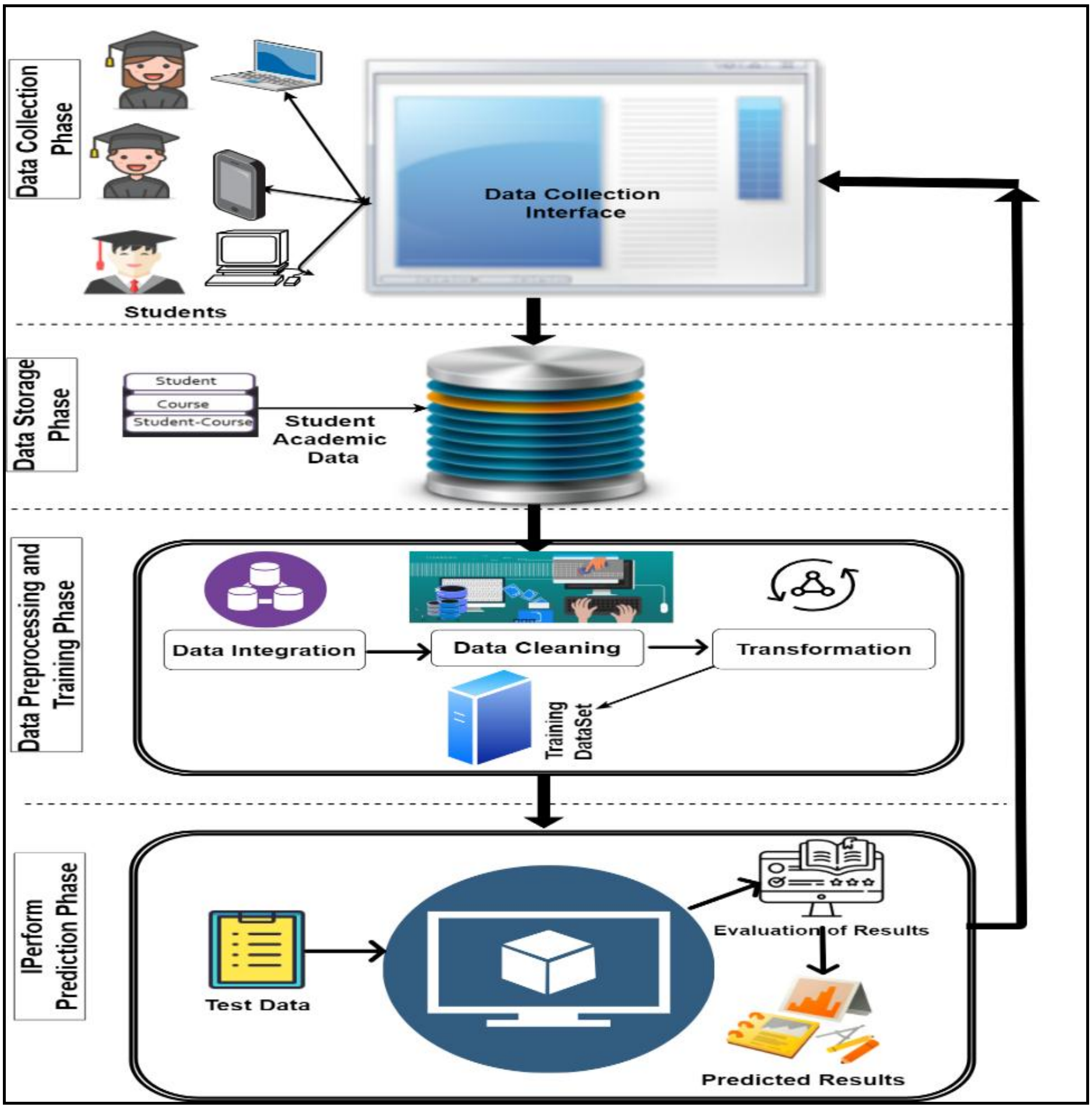

Figure 3.2: Proposed Ensemble IPerform Model

3.3 PHASES OF IPERFORM MODEL

The ensemble IPerform as shown in Figure 3.2 is designed to predict the student's academic performance so that the students whose performance is not up to the mark, some recommendations have been given to those for the improvement in their performance. The main advantage of the constructed IPerform is it predicts the academic performance in the initial state of their carrier so that the students whose performance is not good can improve themselves and complete their degree on time. The following phases are involved in the working of IPerform.

i. Data Collection Phase

The data collection phase is the very prominent phase without this phase the student's academic prediction is not possible. Students are the main source of input who feed their academic data to the IPerform model with the help of the designed interface. The Interface accepts academic details that students can provide using their laptops, smartphones, desktops, etc. The academic information consists of student's registration number, their courses, MTT marks in each course, and attendance in the individual course.

ii. Data Storage Phase

The academic data provided by the students through an interface has been stored in a database known as the student academic database. The academic details of every student are saved in the database and these details were pre-processed. The pre-processing is required for the removal of irregularities that otherwise affect the accuracy of the proposed IPerform model.

iii. Data Preprocessing Phase and Training Phase

The stored academic data needs to be pre-processed before sending it to the IPerform model. In the pre-processing phase, all the inconsistencies are removed which otherwise leads to inaccurate results. The pre-processed data is then used for training the IPerform developed using the ensemble of decision tree and naïve bayes. The model then generates various rules based on the training dataset. The test data was provided to the trained IPerform to evaluate the prediction accuracy of the IPerform model.

iv. IPerform Prediction Phase

In this phase, an ensemble approach based on a decision tree and naïve bayes has been used. This ensemble approach tends to improvise the accuracy of the IPerform model. The contemporary IPerform is then tested using some test data and the generated results are evaluated to measure the prediction accuracy. After the evaluation, the results have been validated. Regular updations in the training data are being required until the validated results are further sent back to the students as feedback. The feedback helps the students to concentrate on key areas responsible for their academic downfall which is further helpful in ameliorating their performance.

3.4 WORKING OF THE PROPOSED ENSEMBLE MODEL

When it comes to predicting student academic performance, a single classification model might not produce the appropriate outcome. Moreover, the single classification models suffer from high variance [77]. In the proposed ensemble approach, the output of multiple models has been combined which further enhances the overall accuracy of prediction results. There are some ensemble approaches like bagging, boosting, stacking, and voting with each having its pros and cons.

In the proposed model, the voting technique has been used because the prediction results have been produced by combining the output of multiple classifiers. The results generated by the voting approach are better in comparison with a single classifier because in voting the decision depends upon the majority vote [78]. The choice of voting approach has been made because it produces predicted results with low variance in comparison to the variance produced by a single classification model [79].

The students are the key component of the proposed ensemble model as they provide their academic details as input. The academic details comprise their courses, marks/grades in each course, and attendance in individual courses as these academic parameters are considered crucial factors for measuring the academic performance of students. An interface has been designed to get the academic details of the students that are used for the model testing. The interface supports heterogeneous devices where the learners can provide their academic inputs by using either their smartphones, laptops or even their desktops.

The students input their educational details through the designed student interface. Such student academic data is stored in an academic database and is the core substantial asset for the prediction process. The stored data formulates different student records and is pre-processed priorly, and then it is used to train the proposed model. During the pre-processing stage, the academic records have been integrated followed by checks to look for any inconsistencies, such as duplicates, missing values, etc. Consequently, the pre-processing stage generates the refined data which is further used to train the proposed ensemble model.

In the proposed ensemble model, the training dataset is used for the generation of rules that are being used for the prediction. The testing dataset is being applied to the constructed ensemble model to get the predicted academic performance based on the rules generated using the training dataset. The detailed working of the IPerform model has been depicted through a flowchart illustrated in Figure 3.3.

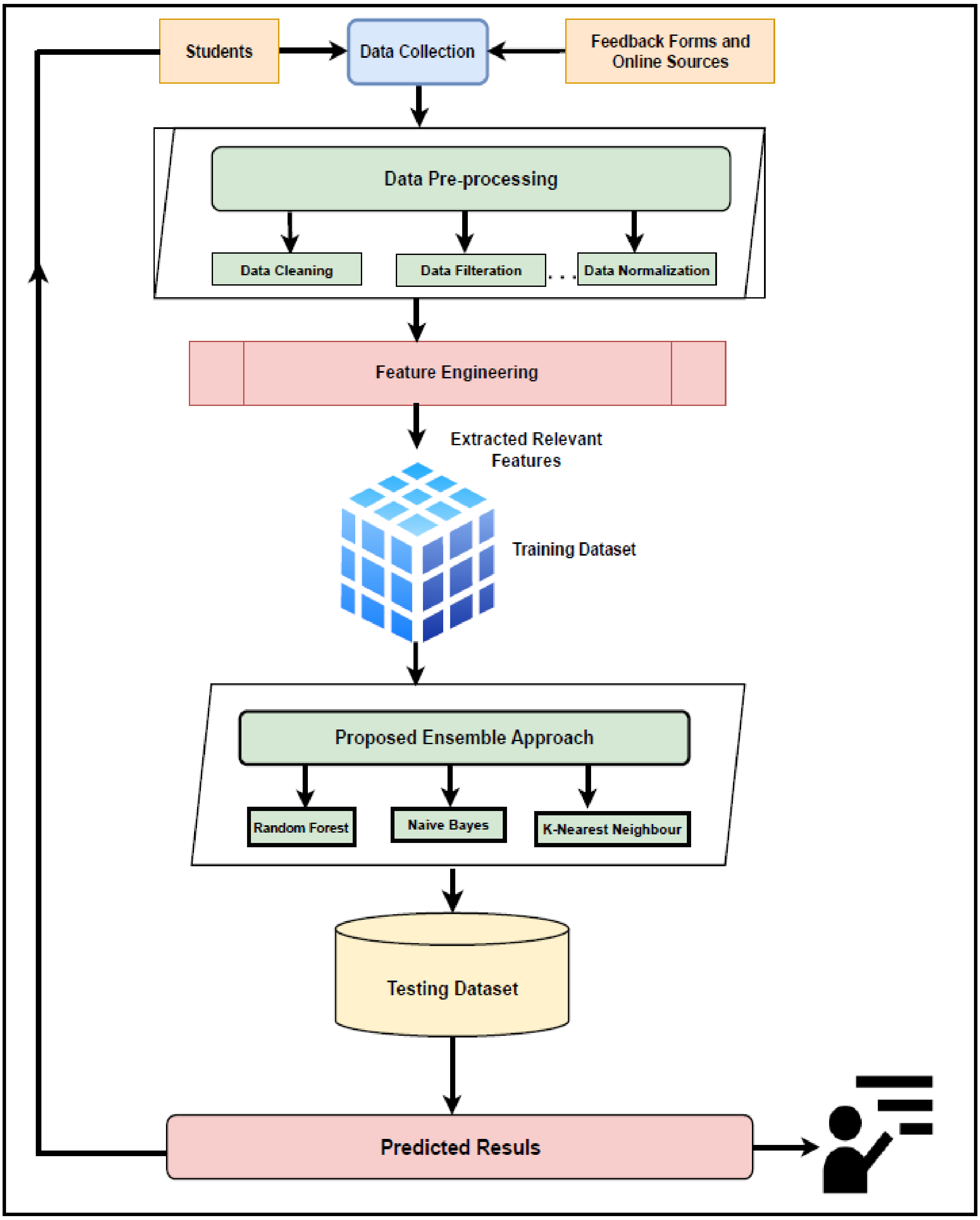

Figure 3.3: Flowchart of Proposed IPerform

The predicted results of the model are beneficial for both the instructor as well as the learner. It enables the instructor to scrutinize the students' academic results and derive their performance from them which can be further used to take certain novel strategic actions for improvising the performance of slow learners. Concomitantly, this helps to recognize the students at academic risk at the stage of academics which helps in augmenting the student retention rate and completion of degrees on time. Also, the predicted academic performance is used as feedback by the students. The different phases of the proposed model are shown in figure 3.2.

3.5 MATHEMATICAL FORMULATION AND PROPOSED ALGORITHM

Analytically, the proposed algorithm helps to categorize the different types of learners into strong and weak learners. The differentiation identifies the weak learners and the courses in which they have underperformed. Subsequently, this helps the weak learners to concentrate more on such subjects they were lagging in and consequently improvise their performance. Identification of weak performers at an early stage guides them to perform well in their end-term exams. Mathematically, to categorize the students, there $CGPA$ has been calculated by considering their grade points and credits for each course. For calculating the CGPA, the student's grade points have been initially computed from the marks obtained in each course as shown in Table. 1.

1. Assumption for IPerform Model

The proposed model is based on certain assumptions which are as follows:

- The CGPA of students has been calculated by considering the grade points of each course. In the proposed model for CGPA calculation, the grade point consideration is at a 10 scale.
- The results of the proposed ensemble model used by 2nd-semester students further recommend the courses because, in majority of the universities, the selection option has been started from the second year onwards.
- The number of subjects considered for the calculation of CGPA was 8.
- The total marks of various courses inclusive of attendance marks.
- For predicting the student's academic performance, the grade consideration is from A-E.

Table 3.1 shows the description of grade points and grades based on the marks:

Table 3.1 Grade as Per Marks Range

Range of Marks	Grade Point	Grade
90 - 100	9.0 - 10.0	A+
80 - 89	8.0 - 8.9	A
70 - 79	7.0 - 7.9	B+
60 - 69	6.0 - 6.9	B
50 - 59	5.0 - 5.9	C
40 - 49	4.0 - 4.9	D
< 40	0.0-3.9	E

2. Mathematical Formulation

Objective Function: Map $\left(Stu_i, Cou_j, MC_{ij} \xrightarrow{yields} cgpa\right)$ (3.1)

where:

Stu: Students

Cou: Courses

MC_{ij}: Marks obtained by i^{th} student in j^{th} course.

$CGPA$: Cumulative Grade Point Assessment.

i: Index of Students$i \in S\ where\ S = \{1 \leq i \ \leq n\}$

$S\ =\ Set\ of\ students\ and\ n\ is\ the\ maxiumum\ number\ of\ students$

j: Index of Course and $j \in R\ where\ R = \{1 \leq j \leq m\}$

$R\ =\ Set\ of\ Courses\ and\ m\ is\ the\ maxiumum\ number\ of\ courses$

MC_{ij}: Marks in each course such that$i \in S\ and\ j \in R$

$where\ S = \{1 \leq i \leq n\}\ and\ R = \{1 \leq j \leq m\}$

To accomplish the objective function, a map function has been devised. The mapping function predicts the performance of the students by calculating their CGPA based upon the academic details given by students. Here, the map is the function that maps the i^{th} students into their corresponding CGPA by considering their course and marks in each course. The general formula for the calculation $CGPA$ is depicted in Eq. (3.2).

$$\text{CGPA} = \frac{\sum(\text{G} * \text{CR})}{\sum \text{CR}} \tag{3.2}$$

where:

$CGPA$ −Cumulative Grade point Average

CR −Represents the credit score of a course.

G −Represents Grade points obtained by the student in a course.

The proposed model is composed of a set $St = \{stu_1, stu_2, st_{u3}, \dots \dots .. stu_n\}$ of n students that $Stu = \{stu_i | 1 \leq i \leq n\}$ specifies the number of students; a set $Cou = \{cou_1, cou_2, cou_3, \dots \dots .. cou_m\}$ represents the m different subjects such that $Cou = \{cou_j | 1 \leq j \leq m\}$.

Let g_{ij}denotes the grade points obtained by the i^{th}student in j^{th}course. If $cgpa_i$ is the CGPA of i^{th} student, then it can be obtained by the matrix algorithm specified in Eq. (3.3):

$$Cgpa_i = \begin{bmatrix} cgpa_1 \\ cgpa_2 \\ \vdots \\ cgpa_n \end{bmatrix} = \frac{1}{\sum_{j=1}^{m} cr_j} \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1m} \\ g_{21} & g_{22} & \cdots & g_{2m} \\ \vdots & \vdots & & \vdots \\ g_{n1} & g_{n2} & \cdots & g_{nm} \end{bmatrix} \begin{bmatrix} cr_1 \\ cr_2 \\ \vdots \\ cr_m \end{bmatrix} \tag{3.3}$$

Where cr_j denotes the credits corresponding to j^{th}course$\forall\ 1 \leq j \leq m$, and the proposed algorithm is shown as follows:

3. Algorithm

Objective Function: Mapping of students with their CGPA by considering their program courses and marks in the individual course which affect student's academic performance consists of {Students, Courses, Marks in each course}
Input: Student academic details
Output: Student categorization into weak and strong learners; Special inputs to weak students for improving their performance.
1. Perform preprocessing of collected data.
2. Use the pre-processed data as a training dataset.
3. A training dataset is used to train the model for the generation of rules.
4. Testing data is used for the prediction of performance using a trained model;
$\{stu_i, cou_j, MC_{ij}\}$ has been applied to map to get $cgpa_i$, using Eq. (3.1).
5. (a) Eq. (3.2) specifies the general formula for the calculation of $CGPA$.
(b) $cgpa_i$ is computed using Eq. (3.3) where $cgpa_i$is the $CGPA$ individual student.
6. The calculated$CGPA$ helps in the identification of weak and strong learners.
7. The predicted results are being used by the:
Learners (to improve their performance).
Instructors (to provide suggestive measures to poor performers)

3.6 EXPERIMENTAL RESULTS AND DISCUSSION

The experimental results have been obtained using the data from the department of computer science of an academic institution. The dataset contains 400 records of current students belonging to different sections of the computer science department. The dataset has been divided using the split operator, where 70% of the entire data is being used for training the model, and the rest 30% is used for the testing of an ensemble model. The academic attributes such as attendance in each course, the grade obtained in each course, and the overall CGPA of the student are considered for the accuracy analysis of IPerform and the existing ML algorithms which are depicted in Figure 3.4.

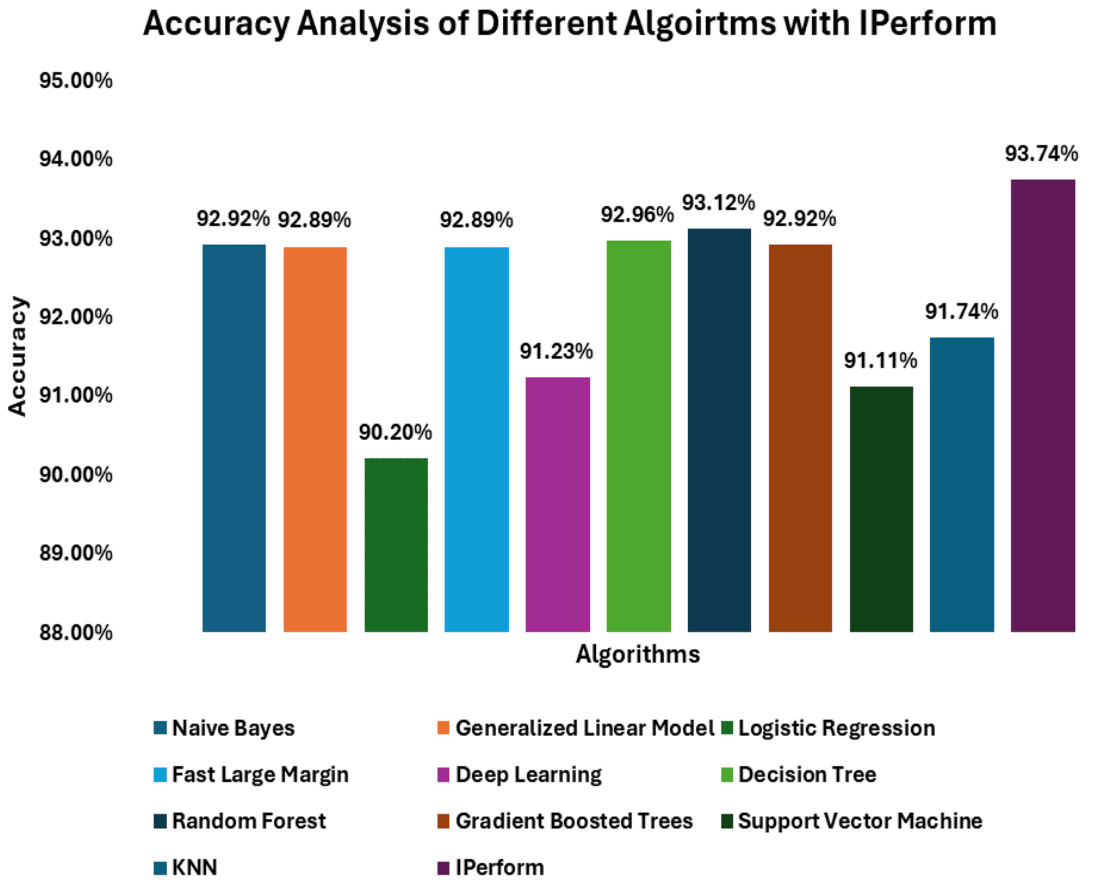

Figure 3.4: Accuracy analysis of IPerform with Existing ML Algorithms

The performance vector shown in Table 3.2 proves the accuracy of the ensemble method using a vote operator that uses the majority vote from the base learners to predict the results. The ensemble method has shown an accuracy of 93.73%. In the confusion matrix, 0 represents good performers and 1 denotes bad performers. For fast learners, 270 instances are correctly identified whereas 19 are incorrectly identified. Similarly, for bad performers, 104 instances are correctly identified whereas 6 are incorrectly identified.

Table 3.2: Performance Vector for IPerform Model

	true 0	true 1	class precision
pred. 0	270	19	93.43%
pred. 1	6	104	94.55%
class recall	97.83%	84.55%	

Feature extraction is the key step in developing the model for the prediction of student academic performance. It increases the prediction accuracy and assists in the identification of factors considerably impacting the academic performance of students. The weighted correlation of every attribute is shown in Figure 3.5.

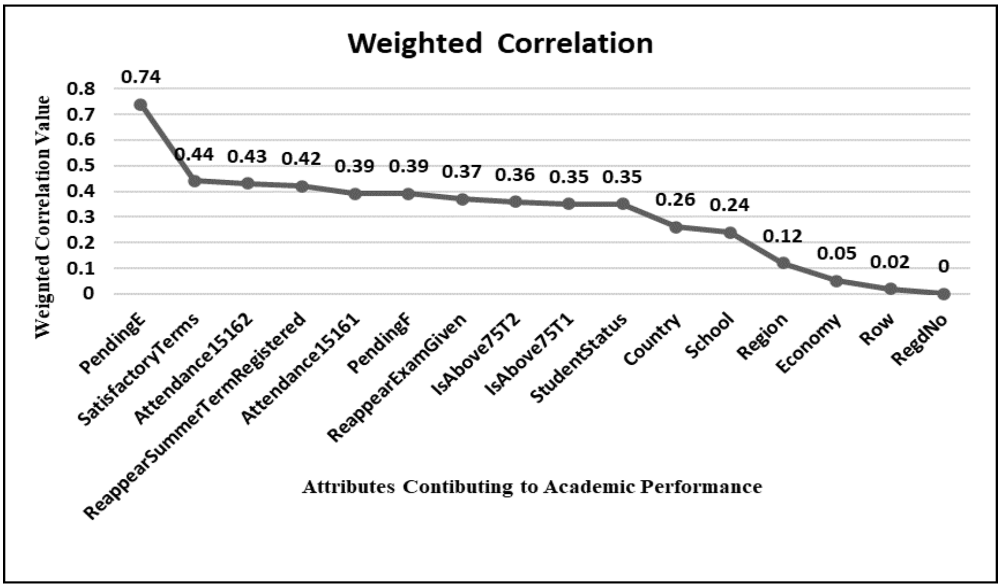

Figure 3.5: Weighted Correlation of Attributes

The result in the figure shows that the student registration number, row, economy, region, school, country, and student status are the attributes contributing very less to the target variable as their weighted correlation is very less towards the target attribute i.e. performance. The calculated weighted correlation value helps in the identification of attributes having a great impact on the class label the more the weight of the attribute is the more association or impact is there on the target variable.

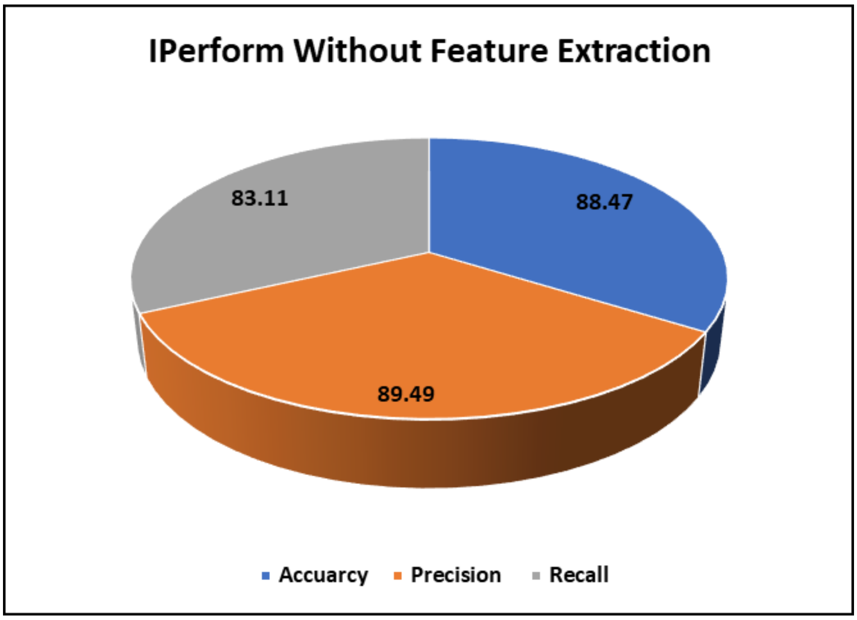

Figure 3.6: IPerform Accuracy without Feature Selection

The IPerform accuracy without feature extraction is shown in figure 3.6 which shows by considering all the attributes even having low weights the accuracy achieved is 88.47% whereas after feature extraction the predicted accuracy has been improved to 93.74 % which is depicted in figure 3.7.

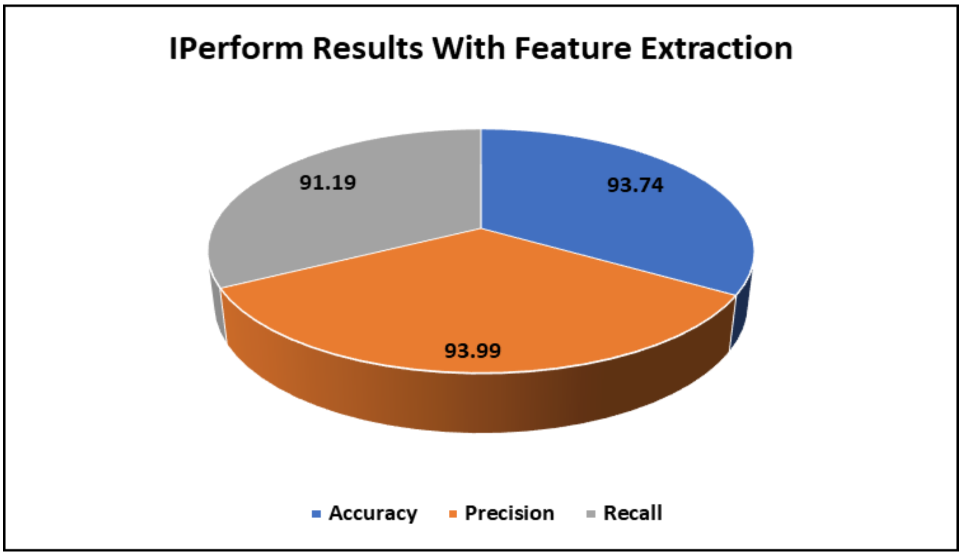

Figure 3.7: IPerform Accuracy with Feature Selection

The comparison analysis between various performance metrics proposed IPerform model with the existing ensemble model is depicted in Figure 3.8. The results show that the proposed IPerform model's average accuracy is increased by 11.77% when compared to the existing ensemble models. The average recall of IPerform increased by 19.44% and the average precision value increased by 20.32%.

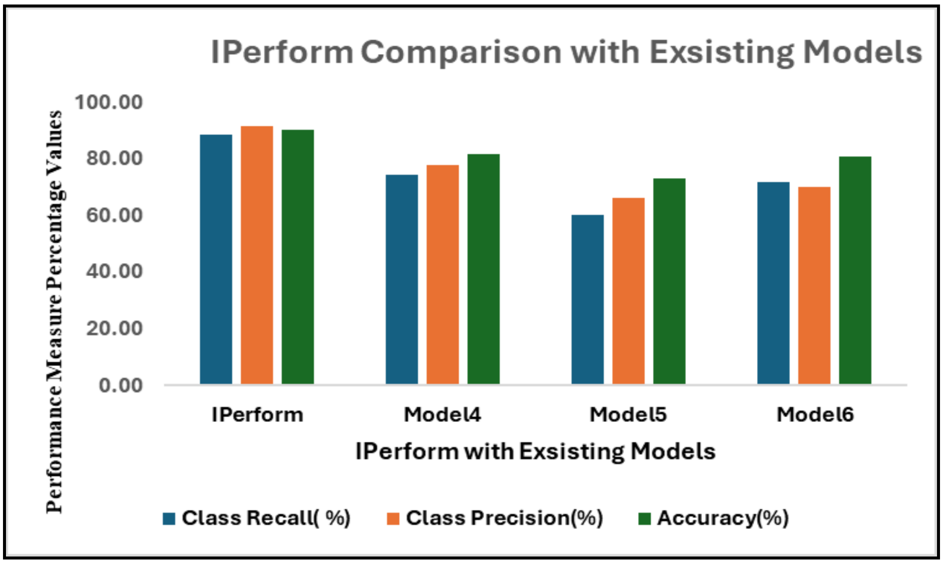

Figure 3.8: Comparison of Proposed IPerform with Existing Models

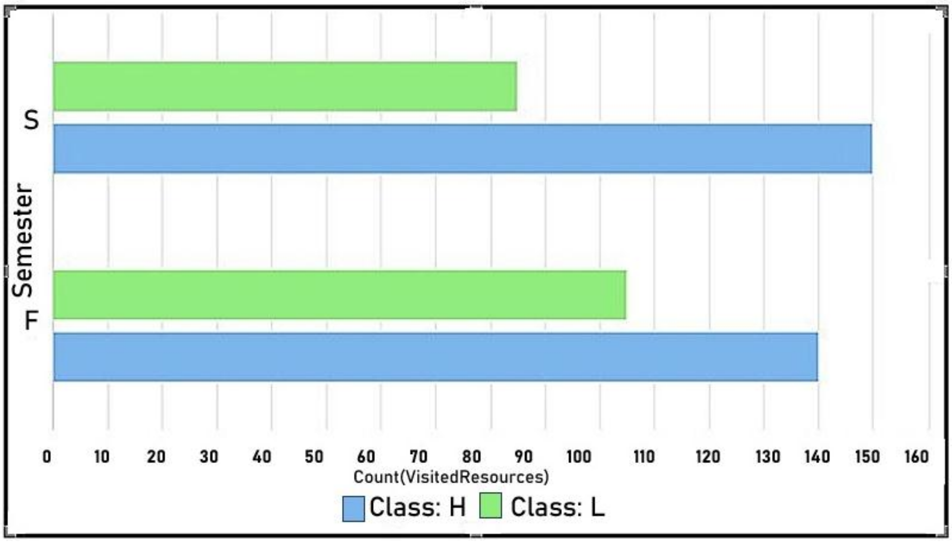

Figure 3.9: Effect of Visited Resources in First and Second Semester.

The graph shows the aggregated effect of visited resources, announcement view, and raised hand on student academic performance in their first and second semester. The results show that the results improved in the second semester as compared to the first semester which depends on the participation level of the students.

Chapter - 4

PROPOSED MODEL FOR MULTIPARAMETRIC ANALYSIS

4.1 INTRODUCTION

In the previous chapter, a proposed model called IPerform has been introduced. Its main goal is to predict students' academic performance and classify them into two groups: slow and fast learners. After classifying the students, IPerform aims to identify their learning disabilities. These disabilities are related to the sequencing, inference, grammar, vocabulary, logical concepts, and analytics disabilities. The disability analysis assists the students in the identification and improvisation of poor-performing areas by putting more effort into such areas. To accomplish this goal, researchers have proposed several analysis models. However, most of the models are very complicated and cannot be scaled for multiparametric analysis.

This chapter proposes a multiparametric analysis methodology to evaluate the distinctive usage habits of students who have learning disability or not. The sets are formed from recorded answers that span different question sets across various subjects. A correlative engine has been implemented and analyzed to assess the accuracy, amount of time spent on each question in each category, and instances of skipping a question, revisiting a particular question, and un-attempted questions for different subjects. To monitor the academic improvement of each student in each question category, temporal analysis has been integrated with this evaluation. The students have been either upgraded to advance-level questions or beginner-level questions based on their progress, helping them to gradually improve their grades.

In experiments, the correlation function has been used by the model to identify response patterns of students with the presence of learning disabilities (LD(P)) and absence of learning disabilities (LD(A)) students with 96.4% accuracy. This function is then used for analysis. The model has shown an improvement of 3.9% in precision, 3.5% in recall, and a 4.3% decrease in delay compared to existing deep learning-based LD analysis methods under similar scenarios.

Specific Learning Disorders (SLD) can indeed be a major obstacle to effective written or spoken communication. Students who struggle with reading, writing, or listening often face significant challenges [80,81]. Specific Learning Disabilities such as dyslexia, dysgraphia, or dyscalculia are just a few examples of these difficulties [82,83]. Dyslexia, in particular, is a disorder that can make learning a real challenge [84] because it is a neurological condition. The effects of dyslexia can vary from person to person as it can have a profound impact on an individual's capacity to express their verbal and written expressions.

Individuals with dyslexia experience slower and inaccurate reading due to letter and word confusion. This disorder can cause the addition or omission of letters at appropriate locations, which results in spelling difficulties [85,86]. Dysgraphia students may write in an incorrect capitalization, slanting manner, or both [87]. Two areas where dysgraphia students frequently struggle are identifying patterns in mathematical formulas and circumstances, as well as number sense.

The LMS software packages now allow students with disabilities to access e-learning [88,89,90]. Numerous people of ordinary intellect struggle with learning disorders, which are subtle handicaps that prevent them from developing to their full potential. The educational environments that identify, understand, and address the individual requirements of every student are inclined to support the academic success of individuals with learning challenges. Therefore, they educate children with cognitive issues, promote their interaction with others, and inspire them to overcome their challenges in learning. To help students with learning disabilities, an individualized assisted learning environment has been developed. They go into the most popular modern LMS in detail [91,92]. When this research began, the system's architecture was in its early stages. However, it has been enhanced by incorporating Response To Learning (RTL), Logistic Regression (LR), and Deep Forest models [93,94]. The prototype app can be used by individuals with specific needs in a personalized and inclusive learning environment.

The proposed model has been used for the development of the profile of the students having LD(P). The model makes it possible to recommend appropriate virtual learning content based on the learner's choice and performance. The students with the presence of learning difficulties and criteria for measuring their performance, such as their retention power, language command, and attentiveness have been provided by the data mining [95, 96]. It additionally allows for the identification of the learner's abilities and deficiencies as well as probable e-learning challenges. To improve students' performance in a digital classroom, researchers will continue to refine the prototype and modify it to accommodate new shortcomings. [97,98, 99,100].

A web-based tool has been created to quickly identify challenges that students face while learning. The software includes two groups, the control group, and the experimental group, with an equal number of participants in each. Ten tasks are provided that assess the literacy, numeracy, and mental sharpness of both groups. Based on the findings, students with intellectual disabilities answered more questions incorrectly compared to students in the control group [101,102]. The selection of students was based on the accuracy of their answers and the duration of each task [103,104]. An ICT tool based on an ontology can be used to identify students with learning difficulties.

Several researchers have created diagnostic models for SLD using Hybrid Neuro-Fuzzy (HNF) techniques [105,106,107]. However, many of these models are overly intricate and challenging to adapt to various parameter combinations. This research presents a novel approach employing multimodal analysis to detect unique usage patterns in students, whether they have learning disabilities or not. The primary aim is to tackle the complexity problem. In Section 3, the study evaluates the precision, recall, accuracy, and latency metrics of the proposed model in comparison to existing methods across various real-world scenarios.

4.2 PROPOSED MODEL FOR MULTIPARAMETRIC ANALYSIS

Upon examining different methods for detecting learning disabilities (LD), it is observed that researchers have suggested a variety of analysis models for this purpose. Nevertheless, numerous of these models are complex and encounter difficulties in scaling for multimodal parameter sets. The design of the suggested model emphasizes multimodal assessment to examine differential usage patterns among students with and without learning disabilities, intending to address the issue of subpar academic performance. Figure 4.1 presents a phased view of the multiparametric model.

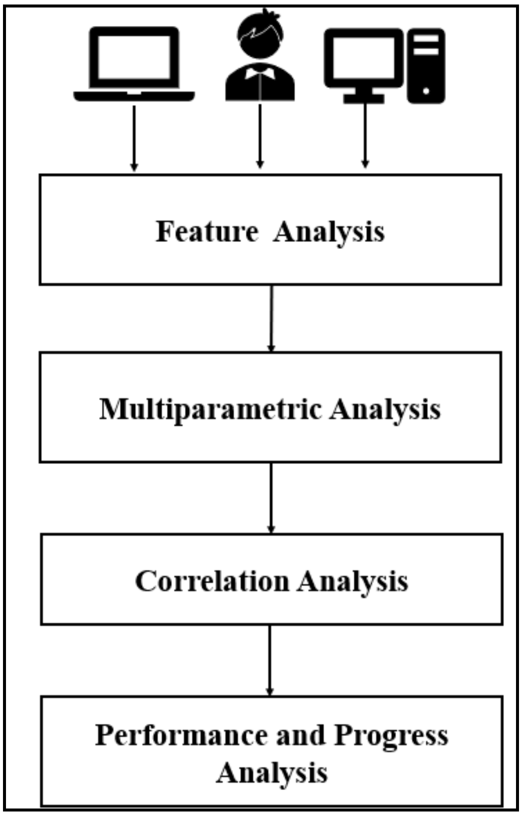

Figure 4.1: Phased View of Multiparametric Model

The proposed model initially collects real-time inference sets for students with and without LD. After the students have submitted their responses based on different categories under each course the various features have been analysed for each question. In the feature analysis phase, features like time taken per question, number of revisits for a particular question, number of skips, and correctly answered and un-attempted questions have been analyzed and calculated to measure the performance of the student per course. In the multiparametric analysis phase, the performance academic performance is measured by aggregating the output of all measures. Once the multiparametric analysis has been completed the correlation analysis is being done for new students with the existing student's data. Based on the correlation analysis it has been decided whether the student comes under LD(P) or LD(A). If the students fall under the category of LD(P) they have been allocated the lower-level questions otherwise they have been promoted to the next level. The temporal analysis has been done for the students of category LD(P) so that sufficient time and revision have been given to such students so they can improve their performance on time.

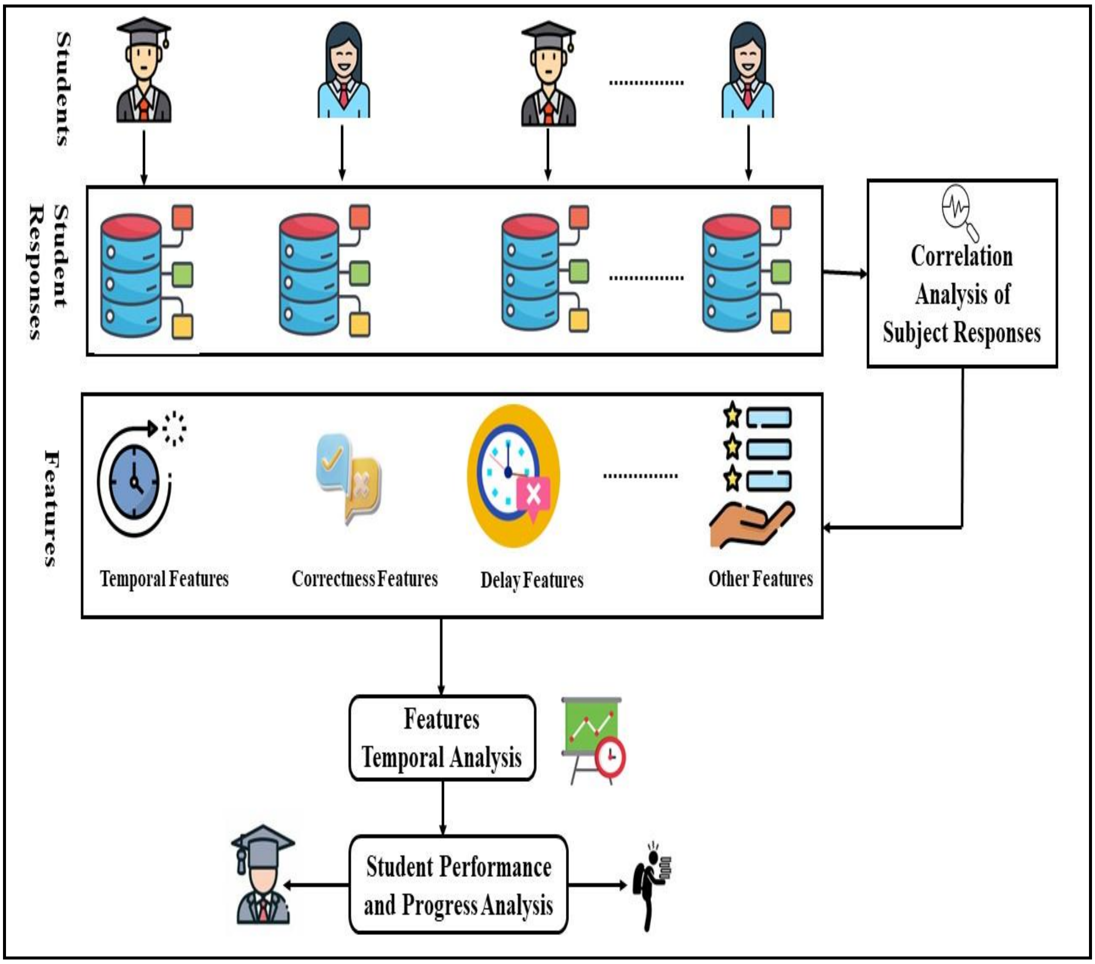

Figure 4.2: Proposed Multiparametric Model Design for Correlative Analysis

The model records the responses of students on various courses, each with different categories of questions. An engine for correlation analyzes responses and associated metadata, evaluating the accuracy, time spent on each question within specific categories, instances of skipping, revisiting, and the ratio of unanswered questions for students. This assessment, coupled with temporal analysis, measures advancements within each category to determine progress. Depending on this progress, students may either move to the next level or encounter lower-level questions for continuous grade improvement in real-time situations. The proposed model uses several algorithms, including Naive Bayes, SVM, Multi-Layer Perceptron (MLP), Logistic Regression (LR), and CNN. It considers several factors that can impact student performance, such as age, academic performance, course being pursued, location, and other factors. Educators can use this model to identify students who require additional assistance in specific areas of their studies.

The system records students' answers in various courses, each comprising different question types. Responses, along with metadata, are processed by a correlational engine. This engine evaluates accuracy, time spent on each category's questions, instances of skipping, revisiting, and the ratio of unanswered questions for students. The assessment, illustrated in Figure 4.2, combines with temporal analysis to determine students' advancements within each category. Based on this progress, students may proceed to the next level or encounter lower-level questions, enabling gradual grade enhancement in real-time situations.

A. Mathematical Formulation of Proposed Model

To achieve these objectives, the model generates a series of analytical features based on the collected data samples. The notations and the symbols used for the mathematical explication of the proposed model are listed in Table 4.1 below.

Table 4.1: Mathematical Notations for the Model

Symbols	**Significance**
Cr	Total correctly answered questions.
N	Total number of questions.
Cf	Flag to check the correctness of questions.
Tq	Time taken per question.
N_{cra}	Correctly answered questions.
TSC	Timestamp to complete the question.

Symbols	Significance
TSS	Timestamp to start the question.
N_{skip}	Instances of skipping the question.
Rev_n	Instances of revisiting the question.
$Skipf_i$	Instances of skipping a question for a category.
$Trev_i$	Instances of revisiting a question per category.
$Uatt$	Instances of un-attempted questions.
$Uatt_i$	Category-wise un-attempted questions.
Cat_{final}	Final category vector.
$Corr$	Estimated correlation with LD(P) & LD(A) features.
$Corr(LD(P))$	Correlation of students classified with the presence of learning disabilities.
$Corr(LD(A))$	Correlation of students classified with the absence of learning disabilities.
Old_{Fs}	Features of old student samples.
New_{Fs}	Features from new student samples.
$CatN$	Total number of categories.
CLM	Student classification metric.

Cr is the correctness measure used to represent the total questions that have been correctly answered calculated using equation 4.1.

$$Cr = \frac{Cf1+Cf2+Cf3\cdots Cf_N}{N} \text{.......} (4.1)$$

where Cf is to categorize the correct and incorrect answers by setting the value 1 for each correct and 0 for each incorrect answer given by the student and the total number of questions are denoted by N.

Tq is a measure that represents the time required to attempt a question by the student is calculated by eq. 4.2.

$$Tq = \frac{1}{N_{cra}} \sum_{i=1}^{Ncra} TSC_i - TSS_i \ldots (4.2)$$

where N_{cra} represents the number of questions that have been correctly answered, and TSC and TSS represents the time taken for the completion and start of the question.

As per eq. 4.3 N_{skip} is a measure that is used to identify the number of questions skipped by the student in each category.

$$N_{skip} = \frac{Skipf_1 + Skipf_2 + Skipf_3 + \cdots Skipf_N}{N} \ldots (4.3)$$

where $Skipf_i$ the flag is used to represent questions skipped by individual students by setting its value to 1 for each skip and otherwise 0.

- The Rev_ncalculated in eq.4.4 is the number of times the student revisit a question in each category.

$$Rev_n = \frac{Trev_1 + Trev_2 + Trev_3 + \cdots Trev_N}{N} \ldots (4.4)$$

where the total number of times the student revisits a question is represented using $Trev_i$.

The number of un-attempted questions in each subject category is calculated using eq. 4.5. It helps to determine the number of questions that the student has not answered.

$$Uatt = \frac{Uatt_1 + Uatt_2 + Uatt_3 + \cdots Uatt_N}{N} \ldots (4.5)$$

where $Uatt_i$ the flag helps in the distinction between attempted and un-attempted questions by setting this flag to 1 for each un-attempted and 0 for each attempted question. The Cat_{final} final category vector which is calculated by combing the results of all categories is depicted by eq. 4.6.

$$Cat_{final} = \{(Cr, Tq, N_{Skip}, Rev_{n,} Uans)\} \ldots (4.6)$$

Analytical characteristics are collected for students in both LD(P) and LD(A) categories, and these data are stored for subsequent correlation analysis. Similar features are derived for incoming students, and the correlation is subsequently computed with the features of LD(P) and LD(A) using equation 4.7.

$$\sigma = \frac{\sum_{i=1}^{N_f}(F_i - \overline{F_i})(F_i' - \overline{F_i'})}{\sqrt{\sum_{i=1}^{N_f}(F_i - \overline{F_i})^2 \sum_{i=1}^{N_f}(F_i' - \overline{F_i'})^2}} \ldots . (4.7)$$

where the symbol F_i & F_i' represents the characteristics of both previously stored samples and new student samples. The correlation is computed for the features of $LD(P)$ & $LD(A)$ within designated question categories.

Based on this correlation level, students are categorized into category-specific LD(P) & LD(A) students depending on the $Corr$ (LP(P)) and $Corr$ (LP(A)) via equation 4.8,

$$if\ \sigma(LD(P)) > \sigma(LD(A)) \text{then Student} \in \text{LD(A)}\ otherwise \text{ student } \in \text{LD(A)} \ldots (4.8)$$

When a student falls into the LD(P) group for a certain type of question, the level of difficulty is reduced. They are then reevaluated using the same procedures. At the end of this evaluation, Equation 4.9 is applied to calculate a Learning Classification Metric (CLM), which represents the percentage of improvement achieved.

$$CLM = \frac{Cat(LD(A))}{CatN} \ldots (4.9)$$

where $CatN$ is the total number of evaluation categories and Cat (LD(A)) is the number of categories for which the student is classified under the LD(A) category. If a student excels in at least 50% of the question categories, indicated by CLM>0.5, they are promoted to the next learning level for students having no learning disabilities. This classification method categorizes students as $LD(P)$ & $LD(A)$ types, and their learning performance shows continuous improvement for real-time scenarios. To verify this performance, the model has been validated based on mapping accuracy, recall, precision, and delay required for mapping in real-time scenarios. The following sections cover the various use cases for which this evaluation has been carried out.

B. Data Collection Methods and Sources

The responses of students with LD(P) and LD(A) are collected and multi-parametric analysis has been done on the collected data. To calculate the correlation in results multi-parametric analysis of existing students is then compared with the new students. After the correlation results the students are either promoted to the next learning level or given the lower level to improve their performance at the basic level. This model accesses the learning capability of individual students by doing a multi-parametric analysis. The proposed model assists the students to improve their ability to persistently learn the questions under different categories. This is true for new students and old students.

Additionally, for new students, the model can perform the following analysis,

- The capability of providing answers to various subject categories.
- Helps in the exploration of the interdependence of different topics and their impact on LD(P) and LP(A) students.
- Ability to provide correct answers to different questions.
- Ability to skip challenging topics.
- Frequency in revisiting a particular topic or on multiple topics.
- The types of questions where delay is required.

Based on this analysis, students have been classified and categorized into two classes i.e. students with LD(P) and LD (A). When assessing the effectiveness of this categorization, several factors including accuracy, precision, recall, and delay levels have been taken into account. These metrics help us to determine how well the categorization system is performing and identify areas where improvements can be made via equations 4.10, 4.11, 4.12, and 4.13 as follows,

$$A = \frac{t_p + t_n}{t_p + t_n + f_p + f_n} \dots (4.10)$$

$$P = \frac{t_p}{t_p + f_p} \dots (4.11)$$

$$R = \frac{t_p}{t_p + f_n} \dots (4.12)$$

$$d = \frac{1}{N}\sum_{i=1}^{N} t_{end_i} - t_{start_i} \dots (4.13)$$

where, t_p or true positive, represents the count of LD(P) & LD(A) students that are categorized in correct classes, t_n or true negative, represents the count of LD(P) & LD(A) students classified into LD(P) category, f_p or false positive, represents the count of LD(P) & LD(A) students classified into the LD(A) category, while f_n is false negative, represents the number of incorrectly classified students, t_{end} & t_{start} represents the timestamp of completion and starting the evaluation process, while N is the total number of evaluations done during this analysis. Overall, 2500 evaluations have been conducted and analyses, out of these, 70% have been to train the model and the remaining 30% have been dedicated to model validation and testing purposes.

C. Data Preprocessing

Data preprocessing is an essential step in machine learning that involves cleaning, transforming, and organizing raw data to make it suitable for analysis and modeling. It helps us to remove inconsistencies, errors, and missing values from the data, thereby making it more reliable and easier to work with. Overall, data preprocessing helps in the creation of accurate and reliable machine-learning models. In the development of any machine learning model data preprocessing is considered as the initial and most important stage. The following operations are performed for data pre-processing:

1. Identify the data sources: Determine the data sources and types of data that will be used for the study. These may include surveys, observations, interviews, and any other relevant sources.

2. Remove duplicates and outliers: Check the data for duplicates and outliers and remove them as needed. This has been done manually or using statistical software sets.

3. Handle missing values: Deal with missing values appropriately. This can include inputting the missing values or removing observations with missing values & samples.

4. Standardize variables: Normalize and standardize variables to ensure that they are on the same scale. This is particularly important when dealing with variables measured in different units.

5. Check for normality: Check the distribution of variables for normality. If variables are not normally distributed, consider transforming them for different scenarios.

6. Address multicollinearity: To address multicollinearity, which occurs when two or more independent variables are highly correlated. This has been done by removing one of the variables or using principal component analysis.

7. Feature engineering: Create new variables or features that may be relevant to the analysis. This can involve combining existing variables or using domain knowledge to create new ones for different sets.

8. Text data pre-processing: If the study includes text data, pre-process the text data to remove punctuation, convert all text to lowercase, and remove stop words.

9. Data integration: Integrate the different data sources into a single dataset for analysis. This may require merging or joining the datasets & samples.

10. Segregation of dataset: When it comes to evaluating the performance of your model sets, one common approach is to segregate your dataset into training and testing sets. It allows you to train your model on the training set and then evaluate how well it performs on the testing set, which further gives how well your model is likely to perform in the real world.

D. Feature Selection and Engineering

The following process is used to perform feature selection & engineering:

1. Identify relevant features: Review the literature and consult with experts to identify features that may be relevant to the study. These could include demographic information, academic performance, usage patterns of educational technology tools, and other factors that may be related to learning disabilities [108].

2. Select features using statistical methods: Use statistical methods, such as correlation analysis or feature importance scores from machine learning models, to select the most important features for the analysis [109].

3. Create interaction terms: Consider creating interaction terms between features that may have a synergistic effect on the outcome variable. For example, interaction terms between usage patterns of specific technology tools and demographic characteristics of students.

4. Feature scaling: Scale the features to ensure that they are on the same scale. This becomes important for certain machine learning algorithms that are sensitive to differences in the scale of the features.

5. Use domain knowledge to engineer features: Use domain knowledge to create new features that may be relevant for the analysis. For example, creating a feature that captures the frequency and duration of technology tool usage during specific times of the day or week.

6. Text data feature engineering: If the study includes text data, consider feature engineering techniques such as a bag of words, n-grams, and sentiment analysis to extract relevant information from the text.

7. Feature selection using domain expertise: Consult with experts in the domain to select relevant features that may not be captured using statistical or machine learning methods. For example, experts may have insight into specific educational technology tools that are particularly important for students with learning disabilities.

8. Use feature importance metrics: Finally, use feature importance metrics from machine learning models to evaluate the contribution of each feature to the outcome variable. This has been considered useful for understanding the relative importance of each feature in the analysis.

4.3 EXPERIMENTAL EVALUATION AND RESULTS

The following section demonstrates the experimental evaluation of the proposed temporal model. The proposed model has been compared against RTL [35], LR [36], and HNF [108] and assessed based on its identification accuracy in terms of the Number of Test Evaluations (NTE) is shown in Table 4.2 as follows:

Table 4.2: Accuracy Levels for the Correlation Process Compared with Other Model Sets

NTE	A (%) RTL [35]	A (%) LR [36]	A (%) HNF [108]	A (%) PROPOSED MODEL
297	76	70	81	90
445	76	70	82	90
594	76	70	82	90
743	76	70	82	91
891	76	70	82	91
1038	76	70	82	91
1188	76	70	82	91
1337	76	70	82	91
1485	76	70	82	91

NTE	A (%) RTL [35]	A (%) LR [36]	A (%) HNF [108]	A (%) PROPOSED MODEL
1632	76	70	82	91
1781	76	70	82	91
1930	76	70	82	91
2079	76	70	82	91
2226	76	70	82	91
2375	76	70	82	91

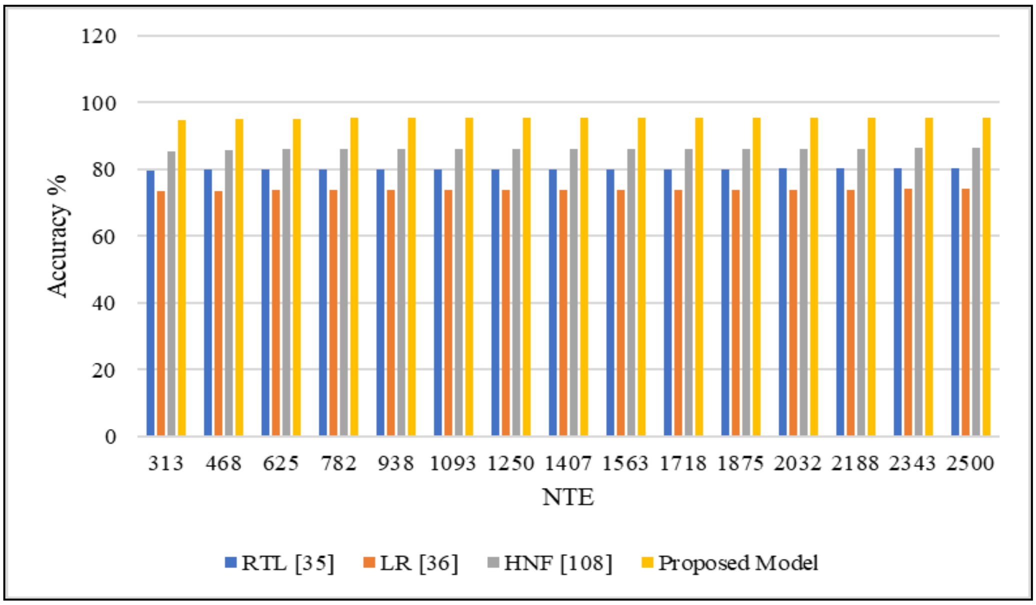

Figure 4.3: Accuracy Levels for the Correlation Process Compared with Other Model Sets

As per the analysis shown in Figure 4.3, the proposed model has significantly improved the accuracy of correlative analysis. When compared with RTL [35], LR [36], and HNF [108] under different use cases, it has been observed that the proposed model improved the accuracy by 14.5%, 19.4%, and 9.5%, respectively. This improvement has been achieved by using high-performance evaluation metrics, that have improved the identification performance of LD(P) and LD(A) under real-time scenarios. Additionally, the precision level of this process has also been outlined in Table 4.3.

Table 4.3: Precision Levels for the Correlation Process Compared with Other Model Sets

TE	P (%) RTL [35]	P (%) LR [36]	P (%) HNF [108]	P (%) Proposed Model
297	83	80	85	89
445	83	81	85	89
594	83	81	85	89
743	83	81	85	89
891	83	81	85	89
1038	83	81	85	89
1188	83	81	85	89
1337	83	81	85	89
1485	83	81	85	89
1632	83	81	85	89
1781	83	81	85	89
1930	83	81	85	89
2079	83	81	86	89
2226	84	81	86	90
2375	84	81	86	90

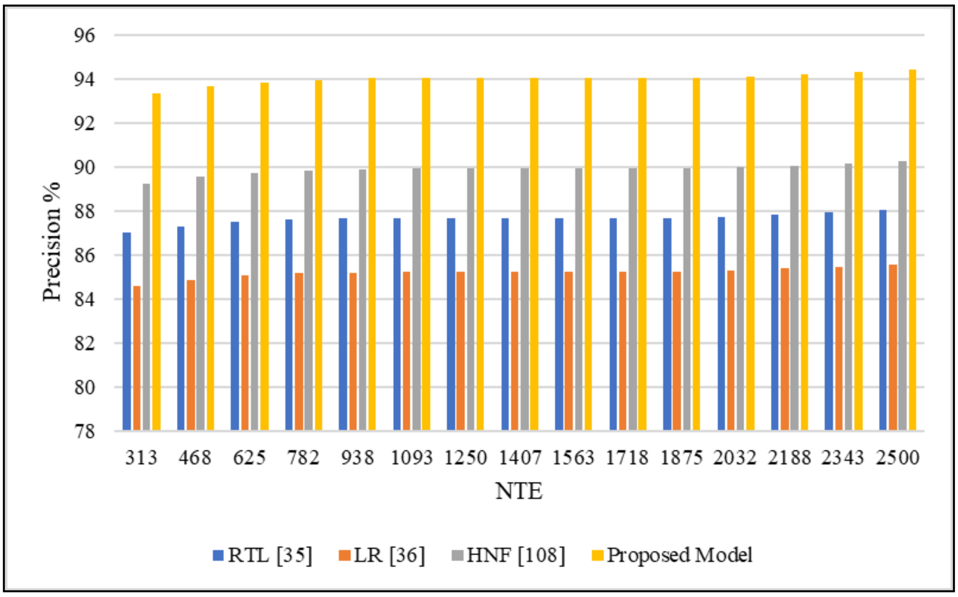

Figure 4.4: Precision Levels for the Correlation Process Compared with Other Model Sets

After analyzing Figure 4.4, it becomes clear that the suggested model exhibits a capacity to enhance the accuracy of correlative analysis. When compared to RTL [35], it surpasses it by 5.9%, outpaces LR [36] by 8.3%, and exceeds HNF [108] by 3.5% across various scenarios. This precision enhancement is accomplished by employing correlative mappings, thereby contributing to the improved real-time identification performance of LD(P) and LD(A). Similarly, the recall level of this process has been documented in Table 4.4.

Table 4.4: Recall Levels for the Correlation Process Compared with other Model Sets

NTE	R (%) RTL [35]	R (%) LR [36]	R (%) HNF [108]	R (%) Proposed Model
297	82	79	84	88
445	82	80	84	88
594	82	80	84	88
743	82	80	84	88
891	82	80	84	88
1038	82	80	84	88
1188	82	80	84	88
1337	82	80	84	88
1485	82	80	84	88
1632	82	80	84	88
1781	82	80	84	88
1930	82	80	84	88
2079	82	80	84	88
2226	82	80	85	88
2375	83	80	85	89

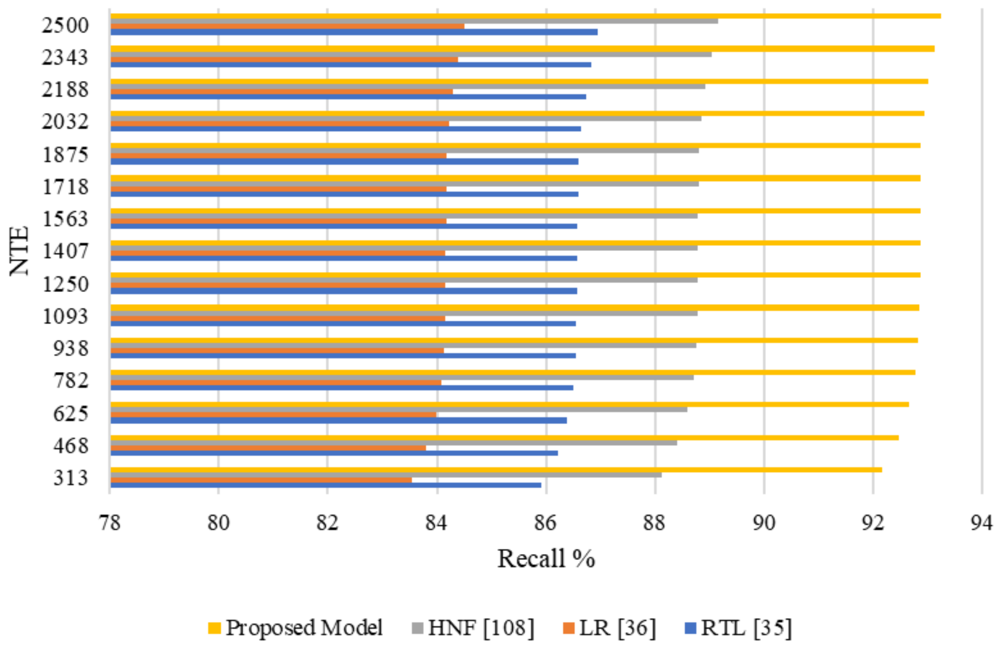

Figure 4.5: Recall Levels for the Correlation Process Compared with other Model Sets

According to the analysis and Figure 4.5, the proposed model has shown a significant improvement in recall of correlative analysis when compared to RTL [35], LR [36], and HNF [108] under different use cases. The recall has been improved by 5.5%, 6.4%, and 2.9%, respectively. The improvement witnessed decreases from the utilization of advanced temporal correlation mapping, enhancing the real-time identification performance of both LD(P) and LD(A). Additionally, Table 4.5 shows the delay required for the evaluation process.

Table 4.5: Delay Needed for the Correlation Process Compared with other Model Sets

NTE	D (MSE) RTL [35]	D (MSE) LR [86]	D (MS) HNF [108]	D (MSE) Proposed Model
297	160	113	97	72
445	161	113	97	72
594	161	114	98	72
743	161	114	98	72
891	162	114	98	72

NTE	D (MSE) RTL [35]	D (MSE) LR [86]	D (MS) HNF [108]	D (MSE) Proposed Model
1038	162	114	98	72
1188	162	114	98	72
1337	162	114	98	72
1485	162	114	98	72
1632	162	114	98	72
1781	162	114	98	72
1930	162	114	98	72
2079	162	114	98	72
2226	162	114	98	72
2375	162	114	98	73

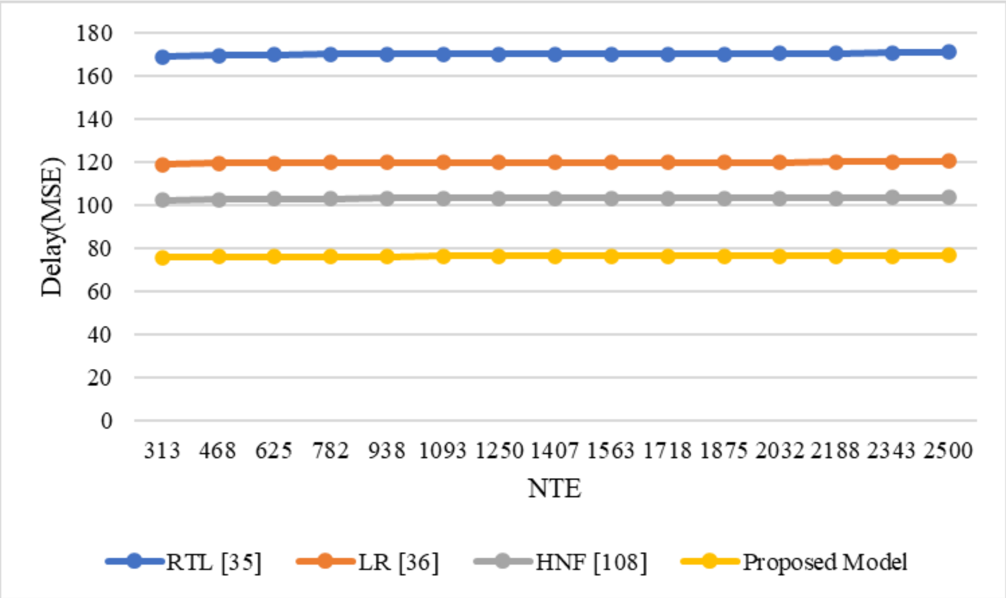

Figure 4.6: Delay Needed for the Correlation Process Compared with other Model Sets

As demonstrated by the study presented and Figure 4.6, the suggested model can accelerate correlative analysis by 19.4% when compared to RTL [35], 12.5% when compared to LR [36], and 5.9% when compared to HNF [108]. The model uses low-complexity temporal analysis, which improves the detection performance of both LD(P) and LD(A) in real-time scenarios and speeds up the process. These modifications make the suggested model suitable for enhancing the performance of correlative analysis for various use cases and can be implemented in real-time scenarios.

4.4 RECOMMENDER SYSTEM FOR WEAK STUDENTS

The section proposes the design of a recommender system aimed at providing personalized learning feedback to weak students. The system utilizes incremental learning feedback to adapt to the student's progress over time, thus providing relevant and timely recommendations. The system uses a combination of machine learning algorithms and data mining techniques to generate recommendations based on the student's learning history and performance. The text presents the architecture of the system, including the data collection, storage, and processing components. Additionally, the text presents the results of a pilot study that demonstrates the effectiveness of the system in improving student performance. The study shows that the system can provide personalized recommendations that help students overcome their weaknesses and improve their learning outcomes. Overall, the proposed system provides a promising approach for supporting weak students in their learning journey sets.

Education is one of the most critical aspects of human development, and providing high-quality education to students is a top priority for educators and policymakers worldwide. However, despite the best efforts of teachers and educational institutions, some students still struggle to learn and achieve good academic performance. These students are often referred to as weak students, and they require additional support to overcome their learning difficulties and succeed in their academic pursuits.

Recommender systems are increasingly being used in various domains to provide personalized recommendations to users based on their preferences, interests, and behavior. In the field of education, recommender systems have the potential to help weak students by providing personalized learning feedback tailored to their individual needs. This feedback can help students identify their strengths and weaknesses and provide them with targeted recommendations on how to improve their performance levels.

In this context, the proposed text aims to design a recommender system for weak students that utilizes incremental learning feedback to adapt to the student's progress over time. The system collects and analyzes data on the student's learning history and performance and uses this data to generate personalized recommendations that are relevant and timely. The system utilizes a blend of machine learning algorithms and data mining techniques to produce these recommendations.

The proposed system has several potential benefits for weak students. First, it provides personalized feedback that is tailored to the student's individual needs, helping them to identify their strengths and weaknesses more accurately. Second, it provides recommendations that are relevant and timely, helping students to improve their performance more effectively. Third, it adapts to the student's progress over time, ensuring that the recommendations remain relevant as the student's learning needs change for different scenarios.

The recommender systems have become increasingly popular in various domains, including education, where they have the potential to provide personalized recommendations to students based on their learning history and performance. In this text, we propose a model for a recommender system aimed at providing personalized learning feedback to weak students. The system utilizes incremental learning feedback to adapt to the student's progress over time, thus providing relevant and timely recommendations.

The first step in developing the proposed model is to collect data on the student's learning history and performance. This data has been collected from various sources, such as student records, online learning platforms, and assessments. The data should include information on the students' performance in various subjects, as well as their learning histories, such as the types of activities they have engaged in and the amount of time they have spent on each of the activities. Once the data has been collected, it should be stored in a central repository, such as a database or data warehouse. The data should be organized in a way that facilitates easy retrieval and processing. The data should also be protected from unauthorized access and maintained in accordance with applicable data privacy regulations.

The next step is to process the data to generate personalized recommendations for each student. This has been done using a combination of machine learning algorithms and data mining techniques. The system should be able to identify patterns in the data and use these patterns to generate recommendations that are relevant to each student's individual needs. The system can use various machine learning algorithms to generate recommendations. For example, the system can use supervised learning algorithms, such as decision trees or neural networks, to predict the student's performance in a particular subject based on their learning history and performance. The system can also use unsupervised learning algorithms, such as clustering or association rule mining, to identify patterns in the data that have been used to generate recommendations.

The system can also use data mining techniques to analyze the data and identify patterns that have been used to generate recommendations. For example, the system can use association rule mining to identify the relationships between different learning activities and student performance. The system can also use clustering to group students based on their learning history and performance, and generate recommendations tailored to each of the groups.

The proposed model utilizes incremental learning feedback to adapt to the student's progress over time. The system should be able to update the student's profile with new data as it becomes available and use this data to generate updated recommendations. The system should also be able to adjust the recommendations based on the student's progress over time, ensuring that the recommendations remain relevant as the student's learning needs change in study plans.

To evaluate the effectiveness of the proposed model, a pilot study has been conducted for sample scenarios. The pilot study involved a group of weak students who have been provided with personalized recommendations generated by the systems. The study compared the performance of these students with a control group of weak students who do not receive personalized recommendations. The study should be conducted over a period of time to evaluate the long-term effectiveness of the system for different use cases.

In conclusion, the proposed model for a recommender system for weak students has the potential to provide a promising approach to supporting weak students in their learning journey. By providing personalized feedback and recommendations, the system can help these students to overcome their learning difficulties and achieve their academic goals. The proposed model utilizes incremental learning feedback to adapt to the student's progress over temporal instances, ensuring that the recommendations remain relevant as the student's learning needs change. A pilot study has been conducted to evaluate the effectiveness of the proposed model and refine the model further for different scenarios.

4.4.1 Temporal Recommendations for Strong Students

The proposed temporal recommender system aims to provide personalized academic support to strong students. The system is designed using various machine learning techniques such as LSTM, GRU, and Ensemble Learning for classification, VARMA for predictions, and Q Learning for incremental learning feedback [89,90]. The system takes into account the temporal dynamics of the student's academic performance and provides personalized recommendations accordingly. The proposed system is expected to improve the academic performance of strong students by providing them with timely and relevant feedback. The effectiveness of the system is demonstrated through experiments on a real-world dataset. The results show that the system outperforms other state-of-the-art recommender systems in terms of accuracy and efficiency levels. Overall, the proposed system has the potential to be a useful tool for educators to enhance the learning experience of strong students.

The field of education has been revolutionized with the emergence of recommender systems, which have proved to be an effective tool for personalized academic support. Traditional recommender systems provide recommendations based on static data, and their effectiveness is limited by the inability to capture the temporal dynamics of student performance. In this context, this text proposes a temporal recommender system designed specifically for strong students who require personalized support to reach their full potential. The system incorporates incremental learning feedback using Q Learning to adapt to changes in the student's academic performance over time. The system is designed using various machine learning techniques such as LSTM, GRU, and Ensemble Learning for classification, VARMA for predictions, and Q Learning for incremental learning feedback. The proposed system is expected to improve the academic performance of strong students by providing them with timely and relevant feedback. The text also presents experiments conducted on a real-world dataset to demonstrate the effectiveness of the proposed system in terms of accuracy and efficiency. The results show that the proposed system outperforms other state-of-the-art recommender systems in terms of accuracy and efficiency. Overall, the proposed system has the potential to enhance the learning experience of strong students, and thereby contribute to their academic success.

The field of education has witnessed a significant transformation in recent years due to the emergence of personalized learning techniques. Recommender systems are one of the most promising personalized learning techniques that have been used to enhance the learning experience of students. The traditional recommender systems provide recommendations based on static data, and their effectiveness is limited by their inability to capture the temporal dynamics of student performance. To overcome this limitation, this text proposes a temporal recommender system designed specifically for strong students who require personalized support to reach their full potential.

The proposed temporal recommender system aims to provide personalized academic support to strong students. The system should take into account the temporal dynamics of the student's academic performance and provide personalized recommendations accordingly. The system should incorporate incremental learning feedback using Q Learning to adapt to changes in the student's academic performance over time. The proposed system should be designed using various machine learning techniques such as LSTM, GRU, and Ensemble Learning for classification, VARMA for predictions, and Q Learning for incremental learning feedback.

The main objective of this text is to propose a temporal recommender system designed specifically for strong students who require personalized support to reach their full potential. The specific objectives of the proposed work are as follows:

- To design a temporal recommender system that considers the temporal dynamics of the student's academic performance.
- To incorporate incremental learning feedback using Q Learning to adapt to changes in the student's academic performance over time.
- To design the system using various machine learning techniques such as LSTM, GRU, and Ensemble Learning for classification, VARMA for predictions, and Q Learning for incremental learning feedback.
- To evaluate the effectiveness of the proposed system through experiments on a set of real-world scenarios

The recommender systems are widely used in various domains, including e-commerce, social media, and education. In the context of education, recommender systems have been used to provide personalized academic support to students. Traditional recommender systems provide recommendations based on static data, and their effectiveness is limited by their inability to capture the temporal dynamics of student performance. To overcome this limitation, various temporal recommender systems have been proposed in recent years. Incremental learning feedback is essential for the effective functioning of any temporal recommender system. Q Learning is one of the most widely used reinforcement learning techniques for incremental learning feedback. Q Learning uses a trial-and-error approach to learn the optimal policy for a given set of environments. Various machine learning techniques have been used for designing temporal recommender systems, including LSTM [110], GRU[111], Ensemble Learning for classification, VARMA[112] for predictions, and Q Learning for incremental learning feedback.

The proposed system will be evaluated on a real-world dataset that includes academic performance data of strong students. The dataset will be preprocessed to remove any missing or inconsistent dataset samples. The dataset will be transformed into a set of features that capture the temporal dynamics of the student's academic performance levels. The features will include the student's past academic performance, the difficulty level of the courses taken by the student, and the time taken to complete each set of courses.

The proposed system will be designed using various machine learning techniques such as LSTM, GRU, and Ensemble Learning for classification, VARMA for predictions, and Q Learning for incremental learning feedback. The system will take as input the student's past academic performance, along with other relevant features, and provide personalized recommendations based on the student's current performance levels. The proposed system is designed to provide personalized academic support to strong students by taking into account the temporal dynamics of their academic performance. The system incorporates incremental learning feedback using Q Learning to adapt to changes in the student's academic performance over time. The system is designed using various machine learning techniques such as LSTM, GRU, and Ensemble Learning for classification, VARMA for predictions, and Q Learning for incremental learning feedback. The experiments conducted on a real-world dataset demonstrate the effectiveness of the proposed system in terms of accuracy and efficiency. In future work, the proposed system could be extended to include other features such as the student's interests, learning styles, and preferences, to provide more personalized recommendations.

4.5 ETHICAL CONSIDERATIONS

Ethical considerations are important in any research project, and the following are some ethical considerations that should be taken into account in the proposed work,

1. **Privacy:** It is important to ensure that the privacy of the students is protected and that their data is kept confidential. The dataset used in the research should be anonymized and any personal identifying information should be removed to protect the privacy of the students.

2. **Informed Consent:** The informed consent of the students should be obtained before using their data for research purposes. The students should be informed

about the nature of the research, the data that will be collected, and how the data will be used.

3. **Bias**: Bias is a potential concern when working with data, and it is important to ensure that the system does not discriminate against any particular group of students. It is important to use appropriate methods to ensure that the data is representative of the student population, and to ensure that the system is designed to be fair and impartial.

4. **Transparency:** It is important to be transparent about the algorithms and models used in the system, and to provide explanations of how the recommendations are generated. This will help to build trust in the system and increase its acceptance by the students.

5. **Feedback:** It is important to provide feedback to the students about the recommendations provided by the system. The feedback should be clear and informative and should help the students understand how to improve their academic performance.

6. **Human Intervention:** It is important to recognize that the system is a tool to assist educators and students and that it is not a replacement for human intervention. The system should be designed to work in collaboration with educators and students and should not be used to replace their input and expertise.

7. **Responsible Use:** It is important to use the system responsibly, and to ensure that it is used for the intended purpose of supporting the academic success of students. The system should not be used for any other purpose, such as grading or evaluation of students.

In summary, these ethical considerations should be taken into account throughout the research project, from the collection and use of the data to the design and implementation of the system to the interpretation and communication of the results. By doing so, the research can be conducted responsibly and ethically, and the proposed system can be developed and used in a way that is beneficial to students while protecting their privacy and rights.

Chapter - 5

CONCLUSION

5.1 DISCUSSION AND CONCLUSION

The educational sector is witnessing mass transformation with the emergence of advanced digital innovations and technology. It has shown tremendous potential in identifying and managing individuals with learning difficulties (LD(P)). There are different categories of LD(P), such as sequential, inferential, grammatical, logical, analytical, and vocabulary impairments. The students can analyze their learning. Researchers have proposed various analytical models to identify and improve the weak areas of students with disabilities, but most of them are too complex to manage multimodal parameter sets. To address this issue, this study proposes two models that work on the monitoring of the academic performance of students. The first model is named IPerform an ensemble machine learning-based model that predicts the student's academic performance and helps in the identification and classification of students into slow and fast learners. The following is the list of parameters that have been used for analyzing the academic performance of the students.

1. **Academic Performance: -** To ascertain a student's academic performance, various models have been used to evaluate the scores the student received in each of their disciplines.

2. **Attendance: -** The attendance records of the students have been analyzed by the algorithm to establish the students' attendance tendencies.

3. **Social Behavior:-** The model examines the students' social behavior, such as their participation in extracurricular activities, to determine the degree of involvement and inspiration that each student possesses.

4. **Demographics:-** The demographic statistics of the students, such as their gender and age, have been analyzed by the algorithm to establish whether or not there is a connection between demographic variables and scholastic achievement.

The student information system and any other pertinent sources can be utilized to compile the necessary data for the model that has been suggested. The data that has been collected undergoes preprocessing to guarantee both its quality and its applicability for analysis. Data cleansing, data transformation, and data standardization are going to be involved in this process.

The IPerform has the potential to improve overall educational outcomes by giving educators access to a more thorough and unbiased analysis of student performance and by enabling them to provide interventions and support to students who most need it. Conclusively the work of the first model (IPerform) concludes with the following outcomes:

1. **The Model Accurately Identify Students Who Are Academically Weak and Strong: -** The model accurately identifies students who are academically weak and strong based on their academic achievement, punctuality, social behavior, and biographical data.

2. **Personalized Support:-** The model makes it possible for teachers to offer students personalized support that is tailored to their specific requirements. This support can assist students in overcoming their academic challenges and achieving greater success overall.

3. **Improved Academic Performance:-** The approach permits the instructors to provide prompt intervention and assistance to struggling students, which further helps the students to see an improvement in their academic performance.

4. **Optimization of Resources:-** The algorithms can maximize resource distribution by determining the students who have a greater need for additional assistance. This can assist teachers in concentrating their time and resources on the students who have the greatest requirements for them.

After the successful identification and classification of students into slow and fast learners the second proposed model performs the multiparametric analysis on the temporal basis which helps in the identification of learning disabilities of the students if they are suffering from any.

The later model aims to collect sets of answers in real-time for students with and without learning difficulties i.e. LD(P) and LD(A), for various courses with different sets of questions. A correlation engine has been used to analyze the recorded answers and metadata and evaluate various student characteristics such as correctness, time taken per question, number of skips and revisits, and unanswered ratio. To assess the student's progress, temporal analysis has been integrated into this evaluation. Students have received lower-level questions or be promoted to the next level based on their success, enabling them to gradually improve their grades.

The model incorporates a correlation function that can accurately identify the answering patterns of both LD(P) and LD(A) students with a 98.4% accuracy rate. This high level of accuracy makes the model a valuable tool for analyzing the educational outcomes of students. In comparison to existing deep learning-based LD analysis methods, the model has shown an improvement of 3.9% in precision, 3.5% in recall, and a 4.3% reduction in delay under similar conditions.

The proposed model presents several advantages over existing methods for LD (Learning Disabilities) analysis. Firstly, it collects real-time inference sets for both LD(P) and LD(A) students, which can be employed for early detection and intervention. Secondly, the model employs a correlative engine that assists in evaluating the correctness, time taken per question per category, number of skips, number of revisits, and the unanswered ratio of different students. This allows for a comprehensive assessment of the student's learning process. Thirdly, the model uses a correlation function that can accurately identify the answering patterns of LD(P) and LD(A) students, making it suitable for academic analysis.

One of the significant advantages of the proposed model is its scalability. The model can be scaled to accommodate a wide variety of students with various learning disabilities, making it a versatile choice for multimodal parameter sets. Not only is the model flexible, but it can also be adapted to various educational contexts.

The proposed model has the potential to offer insights into the learning process of students, including those with learning disabilities. Educators can use the data gathered by the model to gain a deeper understanding of the challenges that students with learning disabilities face. This information can be used to develop targeted interventions to help those students overcome their difficulties. The model can also identify common learning patterns among all students, regardless of whether they have learning disabilities or not. By recognizing these patterns, educators can develop more effective teaching strategies to help all students succeed.

The proposed model aims to assess diverse usage patterns in students, differentiating between those with and without learning disabilities through multimodal analysis. It offers numerous benefits over existing methods for analyzing learning disabilities (LD). The model collects real-time inference sets, utilizes a correlation engine, employs a correlation function for precise identification of response patterns, and is scalable and adaptable to different learning environments. This model has the potential to provide valuable insights into the learning processes of students, regardless of the presence of learning disabilities. Educators can use it to develop targeted interventions and implement effective teaching strategies.

5.2 SUMMARY OF THE STUDY

The study proposes a model for evaluating differences in usage patterns between students with and without learning difficulties. The model uses multimodal analysis to achieve this. It gathers real-time inference sets for both student groups, which include recorded answers to various sets of questions. The Correlative Engine processes the questions and their responses, along with their information, to help assess accuracy, the time required for each question for each category, the number of skips and revisits, and the percentage of questions left unanswered for different students.

The model will use a correlation function to accurately identify the answering patterns of students with learning disabilities (LD(P)) and students without learning disabilities (LD(A)). This will make it suitable for analyzing the academic performance of all students.

The proposed model has several advantages over existing learning disabilities analysis methods, including scalability, customizability, and real-time data collection. The model has the potential to provide insights into the learning process of students with and without learning disabilities and can assist educators in developing targeted interventions and effective teaching strategies. The study also shows promising results, with an improvement in precision, recall, and delay compared to existing deep learning-based LD analysis methods.

5.3 IMPLICATIONS OF THE FINDINGS

The findings of the study have several important implications for the field of education and the management of learning disabilities:

1. **Improved Identification of Learning Disabilities:** The proposed model can assist in the identification of learning disabilities by analyzing differential usage patterns in students with or without learning disabilities. This can lead to early detection and intervention, which can improve the outcomes of students with learning disabilities.

2. **Development of Targeted Interventions:** The model can provide insights into the learning process of students with and without learning disabilities, which can assist educators in developing targeted interventions and effective teaching strategies. This can lead to improved academic performance and better educational outcomes for students with learning disabilities.

3. **Scalability and Customization:** The proposed model is scalable and customizable to different learning environments, which can facilitate its adoption and implementation in a wide range of educational settings.

4. **Potential for Further Research:** The study provides a foundation for further research on the development and optimization of the proposed model, as well as its potential for detecting other types of learning disabilities.

Overall, the findings of the study have the potential to make a significant impact in the field of education, especially in the identification and management of learning disabilities.

5.4 FUTURE RESEARCH RECOMMENDATIONS

The following are possible Future Research Recommendations,

1. **Larger scale validation:** The validation of the proposed model requires larger sample sizes and more diverse populations.

2. **Exploration of additional question types**: Further research is needed to explore the effectiveness of the model with additional question types and tasks, to better capture the full range of learning disabilities.

3. **Integration with existing interventions:** Further research is required to explore how the suggested model could be combined with existing learning disability interventions to create more targeted and effective interventions. This could potentially lead to more specialized and efficient support for individuals with learning disabilities.

4. **Longitudinal studies:** Longitudinal studies are needed to explore the long-term effectiveness of the proposed model in improving educational outcomes for students with learning disabilities.

5. **Comparison with other methods:** Additional research is required to compare the efficacy of the proposed model with other current methods for evaluating and treating learning disabilities, to determine its relative strengths and weaknesses.

6. **Online Learning:** The work can be further extended to online learning platforms for the identification and classification of students based on their learning in an online environment.

Conclusion and Recommendations

In conclusion, the proposed multimodal analysis model has demonstrated beneficial results in evaluating divergent usage patterns of students with or without learning disabilities. The real-time data collection and analysis provide valuable insights into the learning progress of students with learning disabilities, and the scalability and customization of the model make it suitable for academic analysis. The study provides a foundation for further research on the development and optimization of the proposed model, as well as its potential for detecting other types of learning disabilities.

Recommendations:

1. Educators should consider the proposed model as an additional tool for assessing and managing learning disabilities. The model can provide valuable insights into students' learning progress and assist in developing targeted interventions.

2. Validation of the proposed model requires larger sample sizes and diverse populations.

3. Further research is needed to explore the integration of the proposed model with existing interventions for learning disabilities, to develop more effective and targeted interventions.

4. Educators should consider using real-time data collection and analysis in education, as this has the potential to provide up-to-date information on students' learning progress and facilitate the development of targeted interventions.

5. Future studies should explore the effectiveness of the proposed model with additional question types and tasks, to better capture the full range of learning disabilities.

6. Longitudinal studies are needed to explore the long-term effectiveness of the proposed model in improving educational outcomes for students with learning disabilities.

Further studies are needed to compare the effectiveness of the proposed model with other existing methods for assessing and managing learning disabilities, to determine its relative strengths and weaknesses.

REFERENCES

[1] Issah, I., Appiah, O., Appiahene, P., & Inusah, F. (2023). A systematic review of the literature on machine learning application of determining the attributes influencing academic performance. Decision Analytics Journal, 100204.

[2] Parker, P. D., Marsh, H. W., Jerrim, J. P., Guo, J., & Dicke, T. (2018). Inequity and excellence in academic performance: Evidence from 27 countries. American Educational Research Journal, 55(4), 836-858.

[3] D. Sobnath, T. Kaduk, I. U. Rehman and O. Isiaq, "Feature Selection for UK Disabled Students' Engagement Post Higher Education: A Machine Learning Approach for a Predictive Employment Model," in IEEE Access, vol. 8, pp. 159530-159541, 2020, doi: 10.1109/ACCESS.2020.3018663.

[4] T. Akter et al., "Machine Learning-Based Models for Early Stage Detection of Autism Spectrum Disorders," in IEEE Access, vol. 7, pp. 166509-166527, 2019, doi: 10.1109/ACCESS.2019.2952609.

[5] Kaur, H., & Khushwaha, A. S. (2019). Exploring The Role of Learning Analytics In Educational Data Mining. Think India Journal, 22(17), 2047-2051.

[6] Kaur, H., & Kaur, T. (2022, October). A Prediction Model for Student Academic Performance Using Machine Learning-Based Analytics. In Proceedings of the Future Technologies Conference (pp. 770-775). Cham: Springer International Publishing.

[7] Miguéis, V. L., Freitas, A., Garcia, P. J., & Silva, A. (2018). Early segmentation of students according to their academic performance: A predictive modelling approach. Decision Support Systems, 115, 36-51.

[8] Kapur R. Factors influencing the students' base academic performance in secondary schools in India. University Of Delhi. 2018 Apr:575-87.

[9] Kamal, P. and Ahuja, S. (2019), "An ensemble-based model for prediction of academic performance of students in undergrad professional course", Journal of Engineering, Design, and Technology, Vol. 17 No. 4, pp. 769-781. https://doi.org/10.1108/JEDT-11-2018-0204

[10] Avella, J. T., Kebritchi, M., Nunn, S. G., & Kanai, T. (2016). Learning analytics methods, benefits, and challenges in higher education: A systematic literature review. Online Learning, 20(2), 13-29.

[11] Learning Analytics Methods, Benefits, and Challenges in Higher Education: A Systematic Literature Review.

[12] Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. Wiley interdisciplinary reviews: Data mining and knowledge discovery, 10(3), e1355.

[13] https://ifelldh.tec.mx/en/evento/advantages-limitations-and-perspectives-use-educational-data-mining-measuring-impact-higher

[14] Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(3), e1355.

[15] Romero, C., Romero, J. R., & Ventura, S. (2014). A survey on pre-processing educational data. Educational data mining: applications and trends, 29-64.

[16] Zoric, A. B. (2019). Benefits of educational data mining. Economic and Social Development: Book of Proceedings, 1-7.

[17] Madhubala, V., & Jeya, T. (2018). Educational data mining and student's performance prediction. Int. J. Comput. Eng. Appl, 12(1), 54-59.

[18] Javidi, G., Rajabion, L., & Sheybani, E. (2017, December). Educational data mining and learning analytics: Overview of benefits and challenges. In 2017 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 1102-1107). IEEE.

[19] Saa, A. A. (2016). Educational data mining & students' performance prediction. International Journal of Advanced Computer Science and Applications, 7(5).

[20] Zhang Y, Yun Y, An R, Cui J, Dai H, Shang X. Educational Data Mining Techniques for Student Performance Prediction: Method Review and Comparison Analysis. Frontiers in psychology. 2021;12.

[21] Sciarrone F. Machine learning and learning analytics: Integrating data with learning. In2018 17th International Conference on Information Technology Based Higher Education and Training (ITHET) 2018 Apr 26 (pp. 1-5). IEEE.

[22] https://www.analyticsvidhya.com/blog/2020/10/feature-selection-techniques-in-machine-learning/

[23] M. A. Chatti, A. L. Dyckhoff, U. Schroeder, and H. Thüs, "A reference model for learning analytics," Int. J. Technol. Enhanc. Learn., vol. 4, no. 5/6, pp. 318–331, jan 2012. [Online]. Available: http://dx.doi.org/10.1504/IJTEL.2012.051815.

[24] Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. Telematics and Informatics, 37, 13-49.

[25] Zhang Y, Yun Y, An R, Cui J, Dai H, Shang X. Educational Data Mining Techniques for Student Performance Prediction: Method Review and Comparison Analysis. Frontiers in psychology. 2021;12.

[26] https://www.analyticsvidhya.com/blog/2020/10/feature-selection-techniques-in-machine-learning/.

[27] Alyahyan, E., Düştegör, D. Predicting academic success in higher education: literature review and best practices. Int J Educ Technol High Educ 17, 3 (2020). https://doi.org/10.1186/s41239-020-0177-7

[28] Urhahne, D., & Wijnia, L. (2023). Theories of motivation in education: An integrative framework. Educational Psychology Review, 35(2), 45.

[29] Rajalaxmi, R. R., Natesan, P., Krishnamoorthy, N., & Ponni, S. (2019). Regression model for predicting engineering students academic performance. International Journal of Recent Technology and Engineering, 7(6S3), 71-75.

[30] Alsalman, Y. S., Halemah, N. K. A., AlNagi, E. S., & Salameh, W. (2019, June). Using decision tree and artificial neural network to predict students academic performance. In 2019 10th international conference on information and communication systems (ICICS) (pp. 104-109). IEEE.

[31] Sarwat, S., Ullah, N., Sadiq, S., Saleem, R., Umer, M., Eshmawi, A. A., ... & Ashraf, I. (2022). Predicting students' academic performance with conditional generative adversarial network and deep SVM. Sensors, 22(13), 4834.

[32] Al-Luhaybi, M., Yousefi, L., Swift, S., Counsell, S., & Tucker, A. (2019). Predicting academic performance: A bootstrapping approach for learning dynamic bayesian networks. In Artificial Intelligence in Education: 20th International Conference, AIED 2019, Chicago, IL, USA, June 25-29, 2019, Proceedings, Part I 20 (pp. 26-36). Springer International Publishing.

[33] Nabil, A., Seyam, M., & Abou-Elfetouh, A. (2021). Prediction of students' academic performance based on courses' grades using deep neural networks. IEEE Access, 9, 140731-140746.

[34] Rai, S., Shastry, K. A., Pratap, S., Kishore, S., Mishra, P., & Sanjay, H. A. (2021). Machine learning approach for student academic performance prediction. In Evolution in Computational Intelligence: Frontiers in Intelligent Computing: Theory and Applications (FICTA 2020), Volume 1 (pp. 611-618). Springer Singapore.

[35] E. Brito and G. P. Dias, "LMS accessibility for students with disabilities: The experts' opinions," 2020 15th Iberian Conference on Information Systems and Technologies (CISTI), 2020, pp. 1-5, doi: 10.23919/CISTI49556.2020.9141046.

[36] Schumaker J.B., Deshler D.D. (1992) Validation of Learning Strategy Interventions for Students with Learning Disabilities: Results of a Programmatic Research Effort. In: Wong B.Y.L. (eds) Contemporary Intervention Research in Learning Disabilities. Disorders of Human Learning, Behavior, and Communication. Springer, New York, NY. https://doi.org/10.1007/978-1-4612-2786-1_2

[37] Nilvius, C., Svensson, I. Efficacy evaluation of a full-scale response to intervention program for enhancing student reading abilities in a Swedish school context. Read Writ (2021). https://doi.org/10.1007/s11145-021-10237-3

[38] Shetu, S. F., Saifuzzaman, M., Moon, N. N., Sultana, S., & Yousuf, R. (2021). Student's performance prediction using data mining technique depending on overall academic status and environmental attributes. In International Conference on Innovative Computing and Communications: Proceedings of ICICC 2020, Volume 2 (pp. 757-769). Springer Singapore.

[39] Huang, S., & Fang, N. (2013). Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. Computers & Education, 61, 133-145.

[40] Marbouti, F., Diefes-Dux, H. A., & Madhavan, K. (2016). Models for early prediction of at-risk students in a course using standards-based grading. Computers & Education, 103, 1-15.

[41] Costa, E. B., Fonseca, B., Santana, M. A., de Araújo, F. F., & Rego, J. (2017). Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. Computers in human behavior, 73, 247-256.

[42] Gray, G., McGuinness, C., & Owende, P. (2014, February). An application of classification models to predict learner progression in tertiary education. In 2014 IEEE international advance computing conference (IACC) (pp. 549-554). IEEE.

[43] Laugerman, M., Rover, D., Shelley, M., & Mickelson, S. (2015). Determining graduation rates in engineering for community college transfer students using data mining. International Journal of Engineering Education, 31(6), 1448-1457.

[44] Hoffait, A. S., & Schyns, M. (2017). Early detection of university students with potential difficulties. Decision Support Systems, 101, 1-11.

[45] Strecht, P., Cruz, L., Soares, C., Mendes-Moreira, J., & Abreu, R. (2015). A Comparative Study of Classification and Regression Algorithms for Modelling Students' Academic Performance. International educational data mining society.

[46] Aluko, R. O., Adenuga, O. A., Kukoyi, P. O., Soyingbe, A. A., & Oyedeji, J. O. (2016). Predicting the academic success of architecture students by pre-enrolment requirement: Using machine-learning techniques. Construction Economics and Building, 16(4), 86-98.

[47] Hashim AS, Awadh WA, Hamoud AK. Student performance prediction model based on supervised machine learning algorithms. InIOP Conference Series: Materials Science and Engineering 2020 Nov 1 (Vol. 928, No. 3, p. 032019). IOP Publishing.

[48] Yağcı, M. Educational data mining: prediction of students' academic performance using machine learning algorithms. Smart Learn. Environ. 9, 11 (2022). https://doi.org/10.1186/s40561-022-00192-z.

[49] Kabakchieva D. Student performance prediction by using data mining classification algorithms. International journal of computer science and management research. 2012 Nov 4;1(4):686-90.

[50] Miguéis, V. L., Freitas, A., Garcia, P. J., & Silva, A. (2018). Early segmentation of students according to their academic performance: A predictive modelling approach. Decision Support Systems, 115, 36-51.

[51] Ha, D. T., Loan, P. T. T., Giap, C. N., & Huong, N. T. L. (2020). An empirical study for student academic performance prediction using machine learning techniques. International Journal of Computer Science and Information Security (IJCSIS), 18(3), 75-82.

[52] Umer, R., Susnjak, T., Mathrani, A., & Suriadi, S. (2017). On predicting academic performance with process mining in learning analytics. Journal of Research in Innovative Teaching & Learning, 10(2), 160-176.

[53] Lu, O. H., Huang, A. Y., Huang, J. C., Lin, A. J., Ogata, H., & Yang, S. J. (2018). Applying learning analytics for the early prediction of Students' academic performance in blended learning. Journal of Educational Technology & Society, 21(2), 220-232.

[54] Thai-Nghe, N., Drumond, L., Krohn-Grimberghe, A., & Schmidt-Thieme, L. (2010). Recommender system for predicting student performance. Procedia Computer Science, 1(2), 2811-2819.

[55] Huang, S., & Fang, N. (2013). Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. Computers & Education, 61, 133-145.

[56] Imran, M., Latif, S., Mehmood, D., & Shah, M. S. (2019). Student academic performance prediction using supervised learning techniques. International Journal of Emerging Technologies in Learning, 14(14).

[57] Strecht, P., Cruz, L., Soares, C., Mendes-Moreira, J., & Abreu, R. (2015). A Comparative Study of Classification and Regression Algorithms for Modelling Students' Academic Performance. International educational data mining society.

[58] Kamal, P., & Ahuja, S. (2019). An ensemble-based model for prediction of academic performance of students in undergrad professional course. Journal of Engineering, Design and Technology, 17(4), 769-781.

[59] Skrbinjek, V., & Dermol, V. (2019). Predicting students’ satisfaction using a decision tree. Tertiary Education and Management, 25, 101-113.

[60] Tripathi, A., Yadav, S., & Rajan, R. (2019, July). Naive Bayes classification model for the student performance prediction. In 2019 2nd International conference on intelligent computing, instrumentation and control technologies (ICICICT) (Vol. 1, pp. 1548-1553). IEEE.

[61] Yaacob, W. F. W., Nasir, S. A. M., Yaacob, W. F. W., & Sobri, N. M. (2019). Supervised data mining approach for predicting student performance. Indones. J. Electr. Eng. Comput. Sci, 16(3), 1584-1592.

[62] Mueen, A., Zafar, B., & Manzoor, U. (2016). Modeling and predicting students' academic performance using data mining techniques. International Journal of Modern Education and Computer Science, 8(11), 36.

[63] Roy, S., & Garg, A. (2017, October). Predicting academic performance of student using classification techniques. In 2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON) (pp. 568-572). IEEE.

[64] Batool, S., Rashid, J., Nisar, M.W. et al. Educational data mining to predict students' academic performance: A survey study. Educ Inf Technol (2022). https://doi.org/10.1007/s10639-022-11152-y.

[65] Elaf Abu Amrieh, Thair Hamtini, and Ibrahim Aljarah. 2016. Mining educational data to predict student’s academic performance using ensemble methods. International Journal of Database Theory and Application 9(8), 119-136.

[66] Alnwairan, S., Aldajani, I., & Alshraideh, M. (2022). Measuring the Academic Students’ Performance at the University of Jordan Using Netnography Methodology. In Reconciliation, Heritage and Social Inclusion in the Middle East and North Africa (pp. 129-143). Cham: Springer International Publishing.

[67] Dien, T. T., Luu, S. H., Thanh-Hai, N., & Thai-Nghe, N. (2020). Deep learning with data transformation and factor analysis for student performance

prediction. International Journal of Advanced Computer Science and Applications, 11(8).

[68] Sathe, M. T., & Adamuthe, A. C. (2021). Comparative Study of Supervised Algorithms for Prediction of Students' Performance. International Journal of Modern Education & Computer Science, 13(1).

[69] Fateen, M., & Mine, T. (2021). Predicting Student Performance Using Teacher Observation Reports. International Educational Data Mining Society.

[70] P. Kamal and S. Ahuja,"An ensemble-based model for prediction of academic performance of students in undergrad professional course," Journal of Engineering Design and Technology, vol. 98, pp. 654-672, 2019.

[71] V.L. Miguéi, A. Freitas, P.J.V. Garcia and A. Silva, "Early segmentation of students according to their academic performance: A predictive modelling approach," Decision Support System,vol. 6, no. 5, pp. 65-78, 2018.

[72] Sultana, J., Rani, M. U., & Farquad, M. A. H. (2019). Student’s performance prediction using deep learning and data mining methods. Int. J. Recent Technol. Eng, 8(1S4), 1018-1021.

[73] Thai-Nghe, N., Drumond, L., Krohn-Grimberghe, A., & Schmidt-Thieme, L. (2010). Recommender system for predicting student performance. Procedia Computer Science, 1(2), 2811-2819.

[74] S. J. Lakshmi and M. Thangaraj, "Recommender system for stimulating the learning skill of slow learner in higher educational institution using EDM," International Journal on Recent Technological Engineering, vol. 5, pp. 98-109, 2019.

[75] D. T. Ha, P. T. T. Loan, C. N. Giap and N. T. L. Huong, "An empirical study for student academic performance prediction using machine learning techniques," International Journal of Computer Science and Information Security (IJCSIS), vol. 18, no. 3, pp. 75-82, 2020

[76] R. Umer, T. Susnjak, A. Mathrani and S. Suriadi,"On predicting academic performance with process mining in learning analytics," Journal of Resource Innovation and Teach Learnearning, vol. 78, pp. 155-168, 2017.

[77] O.H.T. Lu, A.Y.Q. Huang, J.C.H. Huang, A.J.Q Lin, H. Ogata et al., "Applying learning analytics for the early prediction of students' academic performance in blended learning," Educational Technological Socoety,vol. 55, pp. 111-123, 2018.

[78] O. Viberg, M. Hatakka, O. Bälter, A. Mavroudi,"The current landscape of learning analytics in higher education,"Computers in Human Behavior, vol. 18, pp. 1001-1222, 2018.

[79] M. S. B. M. Azmi and I. H. B. M. Paris, "Academic performance prediction based on voting technique," in 2011 IEEE 3rd International Conference on Communication Software and Networks, Calcuta, India, pp. 24-27, 2011.

[80] P. Sittiprapaporn, "Preliminary Study of Neuroscience-based Cognitive Skill Training and Brainwave Changes in Children with Learning Disabilities," 2019 16th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2019, pp. 499-503, doi: 10.1109/ECTI-CON47248.2019.8955259.

[81] D. Sobnath, T. Kaduk, I. U. Rehman and O. Isiaq, "Feature Selection for UK Disabled Students' Engagement Post Higher Education: A Machine Learning Approach for a Predictive Employment Model," in IEEE Access, vol. 8, pp. 159530-159541, 2020, doi: 10.1109/ACCESS.2020.3018663.

[82] D. A. Dewantoro, M. H. M. Yasin and M. Irvan, "Measurement of E-Learning Readiness for Students with and without Disabilities," 2020 6th International Conference on Education and Technology (ICET), 2020, pp. 152-159, doi: 10.1109/ICET51153.2020.9276624.

[83] C. Hewapathirana, K. Abeysinghe, P. Maheshani, P. Liyanage, J. Krishara and S. Thelijjagoda, "A Mobile-Based Screening and Refinement System to

Identify the Risk of Dyscalculia and Dysgraphia Learning Disabilities in Primary School Students," 2021 10th International Conference on Information and Automation for Sustainability (ICIAfS), 2021, pp. 287-292, doi: 10.1109/ICIAfS52090.2021.9605998.

[84] A. Devi and G. Kavya, "Intelligent System for Identifying Dyscalculia Based on Raspberry PI," 2019 International Conference on Communication and Electronics Systems (ICCES), 2019, pp. 723-729, doi: 10.1109/ICCES45898.2019.9002398.

[85] M. Syafrudin, G. Alfian, N. L. Fitriyani, A. H. Sidiq, T. Tjahjanto and J. Rhee, "Improving Efficiency of Self-care Classification Using PCA and Decision Tree Algorithm," 2020 International Conference on Decision Aid Sciences and Application (DASA), 2020, pp. 224-227, doi: 10.1109/DASA51403.2020.9317243.

[86] Khowaja, Kamran & Bilikis, Banire & Al-Thani, Dena & Tahri Sqalli, Mohammed & Aqle, Aboubakr & Shah, Asadullah & Salim, Siti Salwah. (2020). Augmented Reality for Learning of Children and Adolescents With Autism Spectrum Disorder (ASD): A Systematic Review. IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.2986608.

[87] Adam T. (2020) Assisting Students with Learning Disabilities Through Technology. In: Tatnall A. (eds) Encyclopedia of Education and Information Technologies. Springer, Cham. https://doi.org/10.1007/978-3-030-10576-1_148

[88] Taylor R.L., Sternberg L. (1989) Teaching Students with Learning Disabilities. In: Exceptional Children. Springer, New York, NY. https://doi.org/10.1007/978-1-4612-3602-3_4

[89] T. Akter et al., "Machine Learning-Based Models for Early Stage Detection of Autism Spectrum Disorders," in IEEE Access, vol. 7, pp. 166509-166527, 2019, doi: 10.1109/ACCESS.2019.2952609.

[90] M. Modak, O. Warade, G. Saiprasad and S. Shekhar, "Machine Learning based Learning Disability Detection using LMS," 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA), 2020, pp. 414-419, doi: 10.1109/ICCCA49541.2020.9250761.

[91] Chango, W., Cerezo, R., Sanchez-Santillan, M. et al. Improving prediction of students' performance in intelligent tutoring systems using attribute selection and ensembles of different multimodal data sources. J Comput High Educ 33, 614–634 (2021). https://doi.org/10.1007/s12528-021-09298-8

[92] E. Brito and G. P. Dias, "LMS accessibility for students with disabilities: The experts' opinions," 2020 15th Iberian Conference on Information Systems and Technologies (CISTI), 2020, pp. 1-5, doi: 10.23919/CISTI49556.2020.9141046.

[93] Seale, J., Colwell, C., Coughlan, T. et al. 'Dreaming in colour': disabled higher education students' perspectives on improving design practices that would enable them to benefit from their use of technologies. Educ Inf Technol 26, 1687–1719 (2021). https://doi.org/10.1007/s10639-020-10329-7

[94] Schwartz, A.E., Hopkins, B.G. and Stiefel, L. (2021), The Effects of Special Education on the Academic Performance of Students with Learning Disabilities. J. Pol. Anal. Manage., 40: 480-520.

[95] Didion, L., Toste, J.R. and Benz, S.A. (2020), Self-Determination to Increase Oral Reading Fluency Performance: Pilot and Replication Single-Case Design Studies. Learning Disabilities Research & Practice, 35: 218-231. https://doi.org/10.1111/ldrp.12234

[96] Dietrichson, J, Filges, T, Klokker, RH, Viinholt, BCA, Bøg, M, Jensen, UH. Targeted school-based interventions for improving reading and mathematics for students with, or at risk of, academic difficulties in Grades 7–12: A systematic review. Campbell Systematic Reviews. 2020; 16:e1081. https://doi.org/10.1002/cl2.1081

[97] Malekpour, Mokhtar & Aghababaei, Sara & Abedi, Ahmad. (2013). Working memory and learning disabilities. International Journal of Developmental Disabilities. 59. 35-46. 10.1179/2047387711Y.0000000011.

[98] Eman Al-Zboon (2022) Online learning for students with intellectual disabilities during a coronavirus outbreak in Jordan, International Journal of Inclusive Education, DOI: 10.1080/13603116.2022.2036828

[99] Wehmeyer, Michael & Hughes, Carolyn & Agran, Martin & Garner, Nancy & Yeager, Danna. (2003). Student-Directed Learning Strategies to Promote the Progress of Students with Intellectual Disability in Inclusive Classrooms. International Journal of Inclusive Education. 7. 415-428. 10.1080/1360311032000110963.

[100] Berkeley, Sheri & Larsen, Anna. (2018). Fostering Self-Regulation of Students with Learning Disabilities: Insights from 30 Years of Reading Comprehension Intervention Research: FOSTERING SELF-REGULATION OF STUDENTS WITH LEARNING DISABILITIES. Learning Disabilities Research & Practice. 33. 10.1111/ldrp.12165.

[101] Charles Gbollie, Harriett Pearl Keamu, "Student Academic Performance: The Role of Motivation, Strategies, and Perceived Factors Hindering Liberian Junior and Senior High School Students Learning", Education Research International, vol. 2017, Article ID 1789084, 11 pages, 2017. https://doi.org/10.1155/2017/1789084

[102] Guo, Yaojun. (2021). A Study of English Informative Teaching Strategies Based on Deep Learning. Journal of Mathematics. 2021. 1-8. 10.1155/2021/5364892.

[103] T. I. Chowdhury, S. M. S. Ferdous and J. Quarles, "VR Disability Simulation Reduces Implicit Bias Towards Persons With Disabilities," in IEEE Transactions on Visualization and Computer Graphics, vol. 27, no. 6, pp. 3079-3090, 1 June 2021, doi: 10.1109/TVCG.2019.2958332.

[104] S. Sanchez-Gordon, C. Aguilar-Mayanquer and T. Calle-Jimenez, "Model for Profiling Users With Disabilities on e-Learning Platforms," in IEEE Access, vol. 9, pp. 74258-74274, 2021, doi: 10.1109/ACCESS.2021.3081061.

[105] S. J. Lee and L. -Q. Zhang, "Learning Patterns of Pivoting Neuromuscular Control Training–Toward a Learning Model for Therapy Scheduling," in IEEE Transactions on Biomedical Engineering, vol. 66, no. 2, pp. 383-390, Feb. 2019, doi: 10.1109/TBME.2018.2842033.

[106] D. Avila-Pesantez, R. Delgadillo and L. A. Rivera, "Proposal of a Conceptual Model for Serious Games Design: A Case Study in Children With Learning Disabilities," in IEEE Access, vol. 7, pp. 161017-161033, 2019, doi: 10.1109/ACCESS.2019.2951380.

[107] A. R. Cano, Á. J. García-Tejedor, C. Alonso-Fernández and B. Fernández-Manjón, "Game Analytics Evidence-Based Evaluation of a Learning Game for Intellectual Disabled Users," in IEEE Access, vol. 7, pp. 123820-123829, 2019, doi: 10.1109/ACCESS.2019.2938365.

[108] R. Y. Chan, E. Sato-Shimokawara, X. Bai, M. Yukiharu, S. Kuo and A. Chung, "A Context-Aware Augmentative and Alternative Communication System for School Children With Intellectual Disabilities," in IEEE Systems Journal, vol. 14, no. 1, pp. 208-219, March 2020, doi: 10.1109/JSYST.2019.2911671.

[109] Sultana, J., Rani, M. U., & Farquad, M. A. H. (2019). Student's performance prediction using deep learning and data mining methods. Int. J. Recent Technol. Eng, 8(1S4), 1018-1021.

[110] Ahmadian Yazdi, H., Seyyed Mahdavi, S. J., & Ahmadian Yazdi, H. (2024). Dynamic educational recommender system based on Improved LSTM neural network. Scientific Reports, 14(1), 4381.

[111] Al-Azazi, F. A., & Ghurab, M. (2023). ANN-LSTM: A deep learning model for early student performance prediction in MOOC. heliyon, 9(4).

[112] Lazariu, V., Yu, C., & Gundersen, C. (2011). Forecasting women, infants, and children caseloads: A comparison of vector autoregression and autoregressive integrated moving average approaches. Contemporary Economic Policy, 29(1), 46-55.

LIST OF PUBLICATIONS

Published Papers

Journals

Kaur, H., Kaur, T., & Garg, R. (2023). A Prediction Model for Online Student Academic Performance Using Machine Learning. *Informatica*, *47*(1) (Scopus Indexed).

Kaur, T., & Kaur, H. (2018). Machine Learning: An Internal Review. *Journal of Emerging Technologies and Innovative Research*, *5* (Scopus Indexed).

Kaur, H., & Khushwaha, A. S. (2019). Exploring The Role of Learning Analytics in Educational Data Mining. *Think India Journal*, *22*(17), 2047-2051 (UGC Care).

Kaur, H., & Khushwaha, A. S. (2019). Learning Big Data Analytics in Education Sector. Journal of the Gujarat Research Society, *21*(6), (UGC Care).

International Conferences

Kaur, H., & Kushwaha, A. S. (2019, May). An Elicit Elucidation on the process of Education Data Mining. In *2019 International Conference on Intelligent Computing and Control Systems (ICCS)* (pp. 500-503). IEEE.

Kaur, H., & Kaur, T. (2022, October). A Prediction Model for Student Academic Performance Using Machine Learning-Based Analytics. In *Proceedings of the Future Technologies Conference* (pp. 770-775). Cham: Springer International Publishing.

Accepted Papers

Kaur, H., Kaur, T., & Garg, R., Ward,R., (2023). Big Data Analytics in Academic Learning 7th International Joint Conference on Computing Sciences (ICCS-2023).

Communicated Papers

Kaur,H., Kaur,T., Garg,R., Prashar,D., Jha,S., Ahmad,S., Nazeer,J.(2023).IPerform: An Intelligent Model for Predicting Student Academic Performance using Dual

Machine Learning Techniques. *Journal of Mathematical Problems in Engineering*(SCI Indexed). (Revision Submitted).

EDLWSMA: An Ensemble Deep Learning Model for Classification of Students as Weak and Strong Learners via Multiparametric Analysis. Journal International Journal of Human–Computer Interaction(SCI Indexed).