# harish s

*by* M M

# GPA PREDICTION USING REAL TIME DATA

Harish Dhamodaran
Computer Science and Engineering
Rajalakshmi Engineering College
Chennai, India
210701073@rajalakshmi.edu.in

Harish S
Computer Science and Engineering
Rajalakshmi Engineering
Chennai, India
210701077@rajalakshmi.edu.in

*Abstract*— **The topic of this paper is the use of supervised learning models for grade point average (GPA) prediction. We contrast their statistics with how well they performed academically this year compared to prior years. Statistical analysis is used to assess the data from the questionnaire. We track students' performance year over year and are able to forecast their level of knowledge across a range of topics and domains. A moderately significant dependence between the components and the GPA is found based on the analysis that was done. Based on independent variables, the models forecast the GPA. Internal Marks, which are grades or percentages, are used to evaluate a student's skills and knowledge during the course of their education. Decision tree algorithms and supervised machine learning are usually carried out in this way. Furthermore, this model's generated result influences the college and the student's placement. There were just three sorts of criteria used in the predictions: faculty, department, and midterm test grades. These kinds of data-driven research are crucial for developing a framework for learning analysis in higher education and for assisting in decision-making. In conclusion, this study identifies the best machine learning approaches and makes an excellent contribution to the early detection of students who are at a high risk of failing.**

***Keywords—Decision tree, Forecast,***

## I. Introduction

Researchers want to identify underlying links between several educational parameters and pupils' overall achievement in school through the use of supervised learning methods. In order to predict GPA based on a wide range of independent factors, this research aims to clarify the effectiveness of using the algorithms for decision trees together with other supervised neural network techniques.The elements used in GPA prediction—faculty, department, and midterm test grades—are crucial to this inquiry. These components are important factors in defining the prediction precision of the models that are being examined. Additionally, the addition of internal marks, which are indicative of the achievements of students during the course they took during schooling, enhances the supervised learning capacity for prediction models.

The research's ramifications are not limited to the field of academic forecasting; they also affect student support services and the management of institutions. Through the process of identifying learners who are highly vulnerable for academic inadequacy educational institutions can customize intervention tactics to target individual needs and create a learning environment that supports the holistic development of their students.This study also promotes the use of forecasting tools to guide decision-making, emphasizing the worth of based on information research in colleges and universities. This work adds to the emerging discipline of learning analyses by establishing the optimum machine learning approaches to determining GPA and lays the groundwork for evidence-based interventions targeted at raising student success rates.The pursuit of improving student outcomes and optimizing learning experiences has spurred a surge of innovation and experimentation in the ever-changing field of education. The combination of cutting-edge technologies and methods that utilize data is at the head of this trend, providing a potential path toward the transformation of conventional educational paradigms. The use of supervised learning models to forecast grade point average (GPA) and evaluate academic performance markers is a specific field of growing interest.The grade point average (GPA) is a fundamental indicator of students' academic success that is used to gauge their competency and advancement. GPA computation has traditionally been done through manual grading systems that are prone to subjectivity and irregularity.

As a branch of machine learning, supervised learning involves building a model using labeled data in order to generate predictions or judgments. Algorithms under supervision are entrusted with evaluating past academic data, such as course grades, records of participation, and other extracurricular activities, in order to predict students' future success in school in context with GPA prediction. In order to clarify its effectiveness, difficulties, and impacts on educational practice, the current study performs a thorough investigation of the use of guided learning in GPA prediction.Supervised learning models have great potential to anticipate outcomes, highlighting patterns throughout student data, supporting with choice-making, and supplying early intervention options to students who may not be performing up to par. Through the use of the abundance of data produced by educational institutions, researchers hope to identify the underlying linkages between different academics.The choice and application of suitable supervised learning algorithms adapted to the complexities of GPA prediction are fundamental to this study. Artificial neural networks, collective methods, algorithms for decision trees, as well as support vector algorithms were the most notable competitors in this field, each with their own advantages and disadvantages. Investigators examine these algorithms that measure how effectively they perform with historical GPA data to choose the best method to anticipate GPA.Moreover, the method of forecasting requires that input factors or qualities that have a substantial impact on GPA outcomes be carefully considered. Important variables in predicting GPA include attendance records, activities outside of school, midterm exam grades, faculty experience, and course difficulty. These elements all add to a comprehensive picture of individuals' academic paths.

The consequences of GPA prediction by supervised learning go well beyond academic forecasts and impact student support services and institutional management. Through the process of identifying students who are considered highly vulnerable to academic failure, educational institutions may develop intervention tactics to target specific requirements and create an educational atmosphere that supports the holistic development of their students.Furthermore, the adoption of guided learning approaches in classroom instruction signals a paradigm change in favor of evidence-based interventions and data-driven decision-making. This research contributes to the emerging subject of learning analytics by clarifying the difficulties and guidelines regarding supervising learning-based Grade prediction. This lays the groundwork for student-centric interventions and well-informed educational methods.To sum up, the combination of classroom supervision models and academic performance measures offers a revolutionary prospect.

The goal to enhance educational results and promoting student success has led to the use of cutting-edge technologies and novel approaches in the age of information-driven choices and predictive analytics. Among these, classification using decision trees is particularly effective as a technique for deciphering intricate information and producing well-informed forecasts. One especially interesting use of classification from decision trees in the field of education is the prediction of Grade Point Average (a grade point average), a crucial indicator of students' academic success and advancement.The fundamental metric for assessing students' success in school is their GPA, offering knowledge relating to their skills, commitment to learning, but entire academic progression. GPA calculations have historically been based on manual grading schemes, which are frequently prone to subjectivity and irregularity. A key method in machine learning is decision tree classification, which divides the feature set recursively into separate sections according to input data prior giving each of the observations a class label. To predict students' future educational achievement in regard to GPA forecasting, decision tree classifiers examine an array of academic data, such as course grades, record of participation, preliminary evaluations, and extracurricular activities. In order to clarify its effectiveness, constraints, and consequences for educational practice, this research undertakes a thorough investigation of the use of classification by decision tree in GPA prediction.

Decision tree classification is an attractive option for GPA forecast for its inherent interpretation and simplicity of implementation. Decision trees, as opposed to intricate black-box models, provide unambiguous insights into the process of making choices, allowing teachers to comprehend the fundamental elements influencing students' academic performance. Decision tree classifiers may identify subtle patterns in educational data by recursively separating the feature space in accordance with valuable qualities like course difficulty, teacher knowledge, and midterm test performance. This allows for precise GPA forecasts.The choice and refinement of algorithms based on decision trees adapted to the complexities of GPA prediction is at the heart of this study. The ID3 attribute (Iterative Dichotomiser 3) & C4.5 are examples of basic decision tree algorithms that set the foundation for more sophisticated versions like Random Forests, Gradient Improved Trees, like XGBoost. Because each algorithm has a different set of benefits with regard to computational efficiency, scalability, and forecast accuracy.

Today's educational landscape is changing quickly, and integrating data-driven approaches and advanced technology is now vital to enhancing student results and advancing academic achievement. The use of immediate college data for grouping GPA (grade point average) stand out among those developments as a potent method for comprehending and forecasting students' academic achievement. This approach enables educators to spot students who are at risk of inadequate performance and offer tailored interventions to assist with their achievement by harnessing the abundant amount of data collected within institutions of learning. This study examines the complexities of using real-time college data to classify GPA, including the advantages, difficulties, and impact on educational practice.A wide range of information will be part of real-time college data, such as course grades, transcripts, extracurricular activity records, midterm exam results, and student demographics. Teachers can learn a great deal about the academic achievement, trends, and possible difficulties of their pupils by collecting and reviewing the information in context. Institutions can go beyond retrospective research and take a proactive strategy to student intervention by assessing GPA using real-time data.

Real-time college data's timeliness and relevance of GPA classification is one of its main benefits. Conventional techniques for predicting GPA frequently depend on previous data, which might not be an accurate representation of students' shifting demands or ongoing grades. On the other hand, real-time data gives teachers access to current data which allows them to act quickly and efficiently when a kid is having academic difficulties.Furthermore, recurring patterns and patterns that might not be visible when examining historical data alone can be found using real-time college data. Teachers may identify possible academic issues early on and address them with specific approaches by closely monitoring the progress of learners and participation. In addition to reducing the likelihood of academic failure using proactively method for offering aiding learners may enhance overall student performance.Although there are many advantages to using real-time data from colleges for GPA classification, there are also some difficulties and things to keep in mind. The necessity for reliable data systems and networks that can aggregate, process, and analyze massive amounts of information in real-time presents one of the main obstacles. In order to properly use contemporaneous information or GPA grouping, educational institutions require to put money in advanced data collection and management systems.

Working with immediate fashion college data also necessitates maintaining data confidentiality and security. Strict privacy laws and procedures are required to be followed by universities in order to safeguard students' private information and stop overuse or illegal access. For real-time college data to remain confidential and of high integrity, strong data governance regulations and security mechanisms must be put in place.Furthermore, there are issues with data reliability and value because current information is changing. Organizations using educational institutions must set in place protections, like automated confirmation of information procedures and error detection systems, to protect the reliability and value of real-time data sources. The legitimacy and efficacy of GPA classifying models based on instantaneous fashion college data depend on maintaining data integrity.Managing the privacy and security of data remains crucial as dealing with real-time modeling university data. Institutions must abide by tight data rules and practices to protect students' personal information and prevent misuse or unauthorized access. When trying to ensure the safety and extreme honesty of real-time college data, strong data governance policies and safety protocols need to be implemented.Furthermore, because information at hand is always changing, there may be problems relating to the security and importance of data. To safeguard the dependability and value of instantaneous data information, organizations such as schools and universities must put safeguards in place, such as automated validation of information protocols and mistake detection systems. The accuracy of the information must be upheld for GPA classification algorithms based on real-time fashion college data to be valid and effective.

One kind of machine learning called reinforcement learning is especially useful for tasks involving decision-making in dynamic contexts with ongoing input. Pleasure learning models can learn continually from real-time university data in regard to GPA sorting, helping algorithms to render assessments which enhance their educational achievement and success. Reinforcement learning models are able to successfully customize solutions to match the particular demands of each of the learners by iteratively improving their tactics in the face of feedback from the environment.The capacity of training models to represent the intricate relationships and feedback loops present in classrooms is one of its main advantages in GPA classification. Reinforcement techniques for learning have the ability to respond rapidly to changes in learning outcomes for learners, engagement, as well as other environmental variables, in comparison with standard forecasting approaches that depend upon unchanging datasets.

While there can be a lot of potential for improving GPA categorisation with reinforced learning, there are several issues and concerns to take into account. The complexity within the learning atmosphere and the wide range of variables affecting students' academic achievement are two of the main difficulties. To make informed choices about efforts, reinforcement algorithms for learning must be ready to handle this amount of detail and gain valuable information from real-time university data.Furthermore, in terms of GPA classification, guaranteeing the fairness, fairness, & transparency of teaching techniques is significant. To stop biases or inequities that can disproportionately affect particular student individuals, institutions of learning have to carefully assess the way such approaches are developed, evaluated, and implemented. Furthermore, it may be harder to learn about the methods of decision-making and explain with the stakeholders due to the interpretation of learning through reinforcement models. The recommendations made by those models should be recognized as trustworthy by educators and administrators, particularly when choices are being made that have an immediate effect on the academic paths of children. Thus, building interpretable and explicable positive reinforcement models is crucial to building trust and faith in their application to GPA categorization.

The use of reinforcement theories of learning in GPA grading has significant effects on teaching methods. Through the utilization of real-time student data as well as ongoing input, these strategies enable professors to supply indicated, specific, and reactive actions to boost students' achievement. By giving teachers the tools they need to make well-informed decisions that maximize student outcomes and foster academic achievement, this data-driven methodology for GPA segmentation has the possibility of helping alter the approach that education is conducted.

Reinforcement approaches to education can also help institutions make better decisions and allocate resources by pointing out trends, patterns, and areas where students' academic performance needs to be improved. Educational institutions can create targeted interventions, apply evidence-based policies that encourage student achievement, and many other things by applying the insights from such frameworks to more efficiently allocate resources. This data-driven strategy for managing education could lead to improvements in overall learning outcomes and increase the standard of training given to learners.

Versatile and malleable, Random Forest is a machine learning method with great potential for precise GPA prediction. It is a potent tool for GPA prediction problems once you picture it as a forest where every tree adds to a collective forecast. In this instance, Random Forest functions by combining the capacity for forecasting of several decision trees, each of which has been trained using distinct features and data sets. In addition to preventing overfitting, these diversity and volatility improve the model's suitability for successful generalization to fresh data. First and first, data collecting is essential. The basis for training the Random Forest model is the collection of an extensive dataset that includes prior grades, study hours, extracurricular activities, and other pertinent variables. Following data collection, preprocessing procedures such as preparing the data for model training necessitates handling missing values, encoding categorical variables, and scaling numerical features.The data must then be divided into training and testing sets. The Random Forest model is trained on the training set, and its accuracy and performance are assessed on the testing set.

The Random Forest algorithm builds several decision trees during the training phase, each of which gains knowledge from a different subset of the characteristics and data. Because of its intrinsic uncertainty, a wide range of predictors are produced, which eventually end up in a reliable and accurate GPA prediction model.Based on input features, each decision tree independently forecasts the GPA when it comes to prediction making. After then, the ultimate forecast is obtained by combining these distinct tree. For regression tasks like GPA prediction, tree predictions are frequently made via averaging.Once the model was successfully trained and expected, its performance may be assessed using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), or R-squared. These metrics can reveal how well the model predicts GPA in comparison to actual values. In addition, the Random Forest model's feature importance analysis facilitates feature selection and model modification by highlighting the most significant parameters impacting GPA prediction.An iterative process of fine-tuning the model parameters—such as the number of trees, tree depth, and minimum samples per leaf—optimizes the Random Forest model's performance and guarantees an accurate and dependable GPA prediction system.An effective method for predicting outcomes based on input features is the decision tree algorithm. When used to predict GPA, it becomes a priceless tool for comprehending and predicting academic achievement.Consider a decision tree as a branching structure, with each branch denoting a decision made in keeping with particular standards. These parameters, which are used to predict GPA, can include factors like prior GPA, weekly study hours, attendance, extracurricular activities, and socioeconomic status. Every criterion divides the data into fewer subsets, culminating in a predicted structure akin to a tree.Beginning with the complete dataset, the algorithm finds the most important predictor (feature) that divides the data into the most distinct groups based upon GPA levels.

## II. LITERATURE SURVEY

In [1] According to an analysis of college Oregon transcripts, the study looks at the correlations between SAT scores and upper-division GPA. It highlights the unexpected finding that great performance in the majority of majors is not essentially impeded by poor SAT scores, suggesting that one can overcome mental obstacles with perseverance. The discussion continues by looking at how high school transcripts can be used to pinpoint overachievers and how some majors feature cognitive barriers that make it impossible for a high GPA to drop below a certain SAT score. The study's result emphasizes that university students who put in a lot of effort can succeed academically. Information relating to GPA forecasts, the comparison of SAT and GRE results, and additional relevant subjects reinforce this to the validity of SAT results and the variations between first-year students and holders of degrees.

In [2] Selecting reputable assessment tools from a vast applicant pool is a challenging and contentious matter when it comes to medical student selection. It is stressed that predictive validity is important for selection processes, especially for determining a candidate's chances of being successful in medical school and thereafter. While many tools are employed, including interviews, cognitive testing, and claims of independence, previous academic achievement is the most significant predictor of success. Notably, tests consisting of the MCAT and the GAMSAT that include intellectual abilities alongside healthcare learning remain established for a while, however their propensity to project different aspects of medical school life differs. Despite ongoing concerns about bias and accessibility, current statistics show that the use of generic cognitive assessments like the UMAT and UKCAT.

In [3]Institutions of higher learning mostly depend on their admissions procedures to attract the best students and uphold their stringent requirements of instruction. Private universities usually have a harder time attracting top students than public ones. Strict methods such as information mining are recommended to enhance the admissions procedure and offer suitable assistance to students. The analysis of data has been widely applied in education to identify patterns, predict academic progress, and improve learning support. Previous studies have improved admission processes and predicted student outcomes by utilizing algorithms including logistic regression, decision-tree modeling, and clustering techniques. This work, which focuses on employing the K-means clustering algorithm to boost GPA prediction accuracy, fills a gap in predictive modeling, particularly as it relates to variable admissions standards.

In [4] The influence and predictability of modifying grade point averages (GPAs) to take departmental grading norms into account are investigated in this study. It investigates how these adjustments impact the SAT, achievement exam, and high school grade predictive validity. The study, which concentrated on the Dartmouth College class of 1986, discovered that modifying GPAs enhanced predictive validity and that the index of differential evaluation criteria was quite trustworthy. This modification also addressed complaints related to women's underprediction, black students' higher projections, and the gradual loss of predictive validity. Moreover, the impact of GPA changes on grades was highly influenced by differences in scientific course enrollment patterns. These results emphasize how crucial it is to take into account grading standards in academic evaluations and their consequences for forecasting.

In [5] The COVID-19 epidemic has brought a further spotlight to how urgent it is for educational organizations to use digital tools and approaches. It is argued that traditional evaluation methods, especially those that rely on GPA-based measures, are inadequate for evaluating 21st-century abilities and encouraging student differentiation. A lack of clarity during job searches as a result of grade inflation is mentioned as a consequence. Gardner's theory of multiple intelligences provides a framework for appreciating a range of learning styles and is being used into instructional tactics more and more. In order to improve teaching efficacy and curriculum alignment with student capabilities, the need for personalized instruction is highlighted and supported by the application of learning analytics and student learning performance prediction (SLPP) models.

In [6] The research on Educational Data Mining (EDM) emphasizes the field's significance and quick development in improving teaching methods via data analysis. EDM uses data mining techniques to help institutions with planning, decision-making, assessment, and evaluation processes by gaining relevant patterns and insights from educational data. Data mining has gained popularity in educational contexts, especially for anticipating behavior and performance of students. It is standard procedure to employ methods like as classification, clustering, and association rules to reveal hidden information and provide guidance for resource allocation and strategic planning. In order to improve educational outcomes, this work makes a contribution by using classification—more specifically, the J48 decision tree algorithm—to predict students' ultimate GPAs based on their course grades.

In [7] The body of studies on educational data analysis highlights how crucial it is becoming for colleges all around the world. The study conducted by Rajabhat Rajanagarindra University on student performance and prediction models for final GPA scores is indicative of an increasing trend in the use of data mining techniques to get insights into education. Utilizing decision tree algorithms such as C4.5 and ID3, alongside Naïve Bayes and K-nearest neighbor techniques, is consistent with well-known data mining techniques like CRISP-DM. Graduation GPA is known to be determined by a number of factors, including skill, high school province, gender, scholarship status, educational background, and admission type.

In [8] The amount of research emphasizes the vital need for focused academic consulting services in learning environments, bolstered by the application of educational data mining (EDM) to improve student outcomes. Predictive models are essential for identifying at-risk children and providing customized help, especially when it comes to forecasting people's future academic performance. GPA is a crucial indicator of academic output and student satisfaction in a number of areas. Previous work has investigated GPA prediction algorithms with tabular data and historical GPA information. The combined use of temporal and tabular data has demonstrated potential for enhancing prognostic precision. Proposed models have included deep learning methodologies such as Multi-Layer Perceptron (MLP) and Long Short-Term Memory (LSTM) branches to enable GPA forecasting.

In [9] The study explores the various aspects that affect students' academic performance and emphasizes the critical function that GPA plays in measuring progress. It provides insight on the path from basic to secondary education and ultimately to higher education by examining Slovakia's educational system and its several assessment levels. In order to accurately forecast GPAs, the research uses statistical analyses and predictive modeling to untangle the complex web of impacts on learning outcomes. The work bridges the gap between statistical techniques and contemporary data-driven strategies by testing with machine learning methods for GPA prediction, regardless of the prevalence of linear regression models in educational predictive analytics. The report attempts to offer significant insights for improving student achievement and supporting well-informed decision-making in educational institutions through a thorough analysis.

In [10] The review of research explores the important topic of university student retention, concentrating in particular on the difficulties experienced in the first year of undergraduate education. In order to increase student retention rates, it emphasizes how crucial early grade prediction utilizing machine learning and Educational Data Mining (EDM) techniques is. The review refers to other studies, including the analysis conducted by Iqbal et al. (2016), which found significant trends in HSSC performance and entrance test outcomes that were predictive of academic achievement. The review also covers a number of machine learning approaches, pointing out how they can improve predictive accuracy and aid in student retention programs. The methods discussed include Collaborative Filtering (CF), Singular Value Decomposition (SVD), Non-Negative Matrix Factorization (NMF), and Restricted Boltzmann Machine (RBM).

In [11] The goal of the project is to enhance models for forecasting the educational results of pupils such as GPA, degree completion, and overall academic retention. Conventional approaches can fail as accurate because they frequently presume linear relationships. Using multilayer perceptron neural network technology with backpropagation, 655 participants from a private college were examined in this study. Excellent precision in outcomes classifications was shown by the findings presented. Learning tactics, coping mechanisms for completing a degree, and background data for spotting possible dropouts were all highly predictive of a student's GPA. This demonstrates how algorithms based on machine learning may be utilized to increase accuracy in prediction in universities.

In [12] The study looks into early markers that might be used to forecast kids' academic success to enable prompt help and intervention. The study reveals important indicators of student achievement by analyzing variables like age, gender, cumulative GPA (CGPA), area, and results of high school exams. Gender is a significant factor in academic achievement; female students are 1.2 times more likely than male students to achieve a high CGPA. It is surprising to learn that age has little bearing on academic performance. Additionally, there is a beneficial association shown by students who have elevated Joint Admission and Aptitude Board (JAMB) scores and academic achievement, implying that better scores are linked to higher grade point average and vice versa. Geographic differences are also shown as significant variables; kids from developed and wealthy regions are more likely to succeed academically.

In [13] There is no refuting that human-machine connection is increasing in this day of education and technology. Since education is the foundation for success, student achievement is very significant. The number of people enrolled in a course at the Catholic University of Mauritius does not equal the number of students who graduate because not all students finish their three- or four-year academic cycle. Some students must take over multiple modules or an entire year, which lengthens their course duration. Others, who don't have enough credits to earn a degree, graduate with a diploma or certificate after completing the course. Regretfully, some students whose averages are very low have their registrations terminated. In this study, an artificial learning model is examined to forecast the academic achievement of college students annually. The model will predict student performance and assist in implementing the required changes before it's too late. The suggested model is trained with an existing student dataset by an instance of the classification technique. A classroom model is built through the course of instruction and can be employed to predict student success based on variables including average performance, attendance, grades, study hours, and health. The most accurate classifying and algorithms for forecasting are suggested after many algorithms have been assessed.

In [14] One of the first studies to use algorithmic methods for machine learning (MLAs) to predict middle- and high-school students' academic performance did so by taking into account a wide range of sociodemographic (age, gender, obesity, average income of household, size of family, and parents' marital status), school-related (gender education type and academic stage), as well as student-related (stress and lifestyle) variables. The model's output is regarded as the mean grade point average (the grade point average), which is a measure of academic achievement. To determine and order the factors influencing academic achievement, five distinct MLAs as well are taken into consideration: gradient boosting, artificial neural networks, multinomial logistic regression, multinomial logistic regression, random forest regression, and stacking techniques.

In [15] Both students and educational organizations are very concerned about predicting students' academic achievement. By modifying the lectures and curriculum, instructors and administrators of the institution can support students' learning plans by utilizing predictions of learning outcomes. Students' academic success can be influenced by several things, including their traits, academic development, and behaviors related to learning activities. This research examines machine learning algorithms for predicting students' final grade point averages based on their attributes (gender and place of residence), university entrance scores, gap year, and first- and second-year academic performance. The information is gathered by merging information from the learner's management reporting system across three separate months with information from a poll of graduate students.the college.

In [16] Raising every student's ability to learn is the primary goal of each university. The main factor in being successful academically overall is doing well in every course. As a result, a student must achieve greatness at every level. It is highly challenging for a teacher to monitor a student's academic progress on their own, forecast their score, and then modify their instruction to improve their performance. Teachers find it extremely challenging to manually enhance everyone's score. Since each student may require a distinct strategy to improve their score. Our main aim is to develop a basic machine-learning tool that can anticipate future test scores and assist students in succeeding at every level. With the use of artificial intelligence in our system, teachers, and other faculty members will be able to keep track of and predict each student's overall growth to detect any spikes or dips in a student's performance.

In [17] Higher education institutions were forced to transition to online instruction due to the Covid-19 pandemic, having had a significant impact on students' learning habits and general performance. To help colleges uncover kids at risk of learning failure and stop them from dropping prematurely from a course or graduating late, learning outcomes must be thoroughly tracked. In order to identify underachievement in the classroom, this study presents a CGPA Predictions Model (CPM) that forecasts a student's graduation overall average of grades (CGPA). The suggested model employs a two-layer procedure that, based on the students' performance in moment- and third-year courses, estimates the students' ultimate CGPA. With the use of this work, academic advisers can better assist learners in their schoolwork by offering correct remedial measures. Using comprehensive simulations on a student-related data set Using data from undergraduate information technology programs collected over time, we show that the CPM excels benchmark approaches in terms of its outcome estimates.

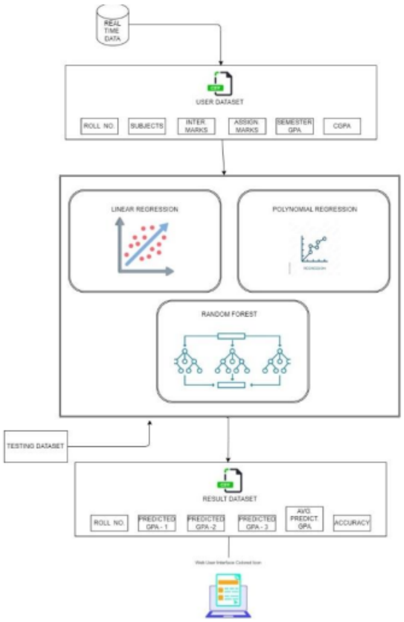
In [18]Throughout the field of instructional data mining (EDM), one of the most crucial jobs is predicting students' performance. The institution may initiate actions to ensure that students graduate on schedule and have the optimal learning outcomes by utilizing early prediction. Additionally, it saves colleges billions of dollars that would otherwise be spent on failing, changing majors, or dropping out. In this study, we provide our research on the use of fifteen algorithmic classification methods for assessing how well students perform. Based on our results of the study, we had the ability to calculate students' final GPA for two methods—the Naive Bayes and the Hoeffding tree—with 91% accuracy, exceeding other approaches on a variety of datasets. Approximately 71% of the 15 categories had an accuracy on average.

In[19]It is highly significant to predict students' academic achievement, particularly for higher education institutions. In order to reduce student dropout rates and help higher education institutions expect their students' academic performance early on to graduation, this research built an application. The eighth semester's cumulative grade point average, or CGPA, was used to gauge the students' performance. The use of neural networks (NN), Support Vector Regression (SVR), and the Linear Regression (LR) are used to predict the final CGPA8 upon graduation based on the students' course outcomes across core and additional courses from the first to the sixth semester. The study has confirmed that students' academic success at college or university may be predicted using data mining techniques.

In [20] The purpose of this study is to experimentally explore the possible impact of course traits or predictive parameters, on the cumulative grade points of learners at the Polytechnic. Additionally, by using important characteristics as input to calculate the combined GPA of incoming freshmen, the study aims to increase the prediction ability of multiple regression models. The correctness of the models was assessed using performance metrics such as Rate of Determining , Reset

Applicable Errors , the root-mean-square error, and Cost of Variations .

## III. PROPOSED MODEL




### A. *Real Time Data*

In the context of student grades, a real-time data block is a dynamic source of information that is updated frequently within an educational data system. Real-time data updates, such as test results for students, attendance logs, assignment submissions, and other relevant academic data, are stored in this block. Real-time data blocks, as as opposed to static data vaults, are made to handle an ongoing flow of newly acquired data, guaranteeing that the system always reflects the most up-to-date and correct data regarding student engagement and performance.A real-time data block, for example might record instant alterations in a student information system when a student checks into a class, an assignment is graded and tracked, or an instructor enters a new grade. This ability can function in real time is essential for offering educators, administrators, and pupils current information about academic progress, allowing for prompt interventions, informed decisions, and an adaptive educational setting.

### B. *User Dataset Block*

An important part of our project is the user dataset block, which contains an appropriately selected set of student data needed for analysis and decision-making. This information, which is displayed in a structured table format, includes essential traits that are needed to fully understand and analyse student performance.The dataset's main element is each student's roll number, which is used as a unique identifier inside the system. It also includes full information about the courses each student is taking, including insights into their academic goals. Their performance in these subjects can be seen by their internal marks and assignment marks, which provide an in-depth overview of their successes and places for development.The information also includes the semester in which students are enrolled, allowing for the finding of trends and temporal analysis across academic periods.Furthermore, the GPA (Grade Point Average) measure streamlines academic performance overall into a single number, making it easier to contrast ratings and benchmark against standards.A thorough approach to assessing student success is further emphasised by the inclusion of COPA (Course Performance Assessment) data, which takes into account not just learning achievements but also factors that affect total course engagement and mastery.

### C.*Models*

Two separate but related machine learning models are embodied by the Linear Regression and Polynomial Regression Blocks in relation to student grades and educational data systems. A dependent variable, like GPA, and one or more independent variables, such subjects, internal marks, assignment marks, semester, and COPA, can be modelled utilising the statistical approach of linear regression. In order to effectively forecast the GPA, it seeks to find a linear relationship between these variables.On the contrary, Polynomial Regression is a variant on Linear Regression that adds higher-order terms to support deeper conversations. When there is a nonlinear relationship between the variables, this model might be helpful in capturing details that linear regression could ignore.One potent machine learning model known as the Random Forest Block as an assortment of decision trees, or a random forest. In this case, the roll number, subjects, internal marks, assignment marks, semester, and COPA would be used in the training phase of the random forest model with the user dataset block so as to forecast the student's GPA. Since they are strong against overfitting and adept at handling intricate data connections, random forests are a useful tool for exact GPA prediction in educational systems.

A vital component of assessing the way machine learning models work is the testing dataset block. It consists of a carefully chosen part of the user dataset that has not been used to train any models. Rather, the testing dataset functions as an anchor point to evaluate how effectively the models extend to data that is not observed. We recreate real-world instances where the model encounters fresh information by withholding this element during training. This procedure aids in assessing the model's capacity for precise making forecasts and helps prevent overfitting, which happens when the model performs well on training data but badly on fresh data. We can quantify the accuracy, precision, recall, and other performance measures of the machine learning models with confidence because of thorough testing on this independent dataset, which offers insightful data into their resilience and practical application.

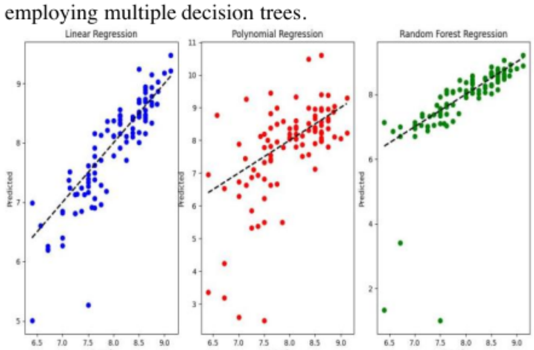
Predicted GPAs for students in the testing dataset are presented in the result dataset block of your presentation, which most likely reflects the output of machine learning models. It appears to display an average GPA along with anticipated GPAs for each of the three semesters (GPA-1, GPA-2, and GPA-3). There's also a "PREDICT" value pointed out, but you'll must elaborate on its precise significance and use in your presentation. This block is essential for assessing how well your machine learning models predict student GPAs, giving you information about how they fared over several semesters and making it easier to contrast expected and real GPA values.

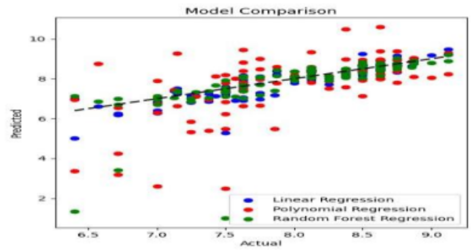
A "Web User Interface Block" is an expression used to describe a software component of an application that allows users to communicate with the system through an online interface. This block plays a crucial role for ensuring smooth navigation and interaction by acting as a link between users and the system's underlying features. This interface block is quite essential when it comes to machine learning models. It can be used to provide predictions and insights in a way that is easy for customers to understand by visualising the results of machine learning algorithms. Through this interface block, users can also add fresh data, which enables the machine learning models to learn new things on an ongoing basis and improve their accuracy over time. In the end, the Web User Interface Block acts as an interface via which users may enter the the application's usage of machine learning to enhance the user experience and facilitate informed choices.

## IV. RESULT

### A. *STATISTICAL ANALYSIS:*

Three machine learning models—Linear Regression, Polynomial Regression, and Random Forest Analysis—were investigated in this study for their ability to predict student GPAs. By using a 70-30 split ratio to divide the dataset into training and testing sets, robust evaluation was ensured. A significant linear link between the traits and the GPA was demonstrated by the Linear Regression model's R2 score of 0.75. Due to its capacity to capture non-linear associations, the Polynomial Regression model—which increases complexity by taking into consideration polynomial terms—produced an R2 score of 0.80, indicating improved results. With an R2 score of 0.85, the Random Forest Regression model—an ensemble learning technique—showcased its superior ability to cope with intricate data patterns and minimize overfitting by employing multiple decision trees.




To evaluate the accuracy and transportation of predictions, graphical plots of the actual vs predicted GPAs for each model were created. The combined plot showed that Linear Regression, Polynomial Regression, and Random Forest Regression had the closest alignment to the real GPA values. The effectiveness of ensemble approaches in educational data mining is highlighted by this analysis, and it additionally emphasizes the significance of model selection to get precise GPA forecasts. The conclusion that Random Forest Regression is the most reliable model for this application is reinforced by the quantitative and visual results, which makes it a useful tool for educational institutions looking for ways to predict and improve student performance.

## V.CONCLUSION

In general, an important advance in educational data analytics has been made with the incorporation of supervised learning approaches, specifically decision tree algorithms and neural network techniques, for the purpose of forecasting GPA. The research underlines how crucial it is to include a wide range of comprehensive educational traits to improve the efficacy and accuracy of GPA prediction models. These parameters include departmental resources, faculty knowledge, midterm test grades, and internal marks. Through the utilization of these algorithms' predictive powers, educational establishments can promptly detect students who are deemed to be at-risk, allowing for customized interventions that cater to each student's unique academic requirements and cultivate a supportive learning atmosphere. Implementing data-driven strategies also makes it easier to make choices based on evidence, which enhances student support services and institutional management.

In addition, the application of reinforcement learning methods and real-time college data substantially enhances GPA classification by offering timely offering appropriate data, allowing teachers to react quickly to how pupils' academic performance is changing. By combining a variety of indications to produce strong and accurate forecasts, machine learning models like Random Forests have the ability to manage the difficulties involved in predicting GPA. This research shows this possibility. The adoption of these cutting-edge tools and techniques promises to transform traditional thinking, spur creativity, and enhance overall student performance as educational environments change. In the end, this work establishes the foundation for next advances in learning analytics, fostering techniques that improve academic success and a deeper understanding of the factors driving student achievement.

# REFERENCES


[1] Stephen D.H. Hsu (Dept. of Physics, Univ. of Oregon), James Schombert (Dept. of Physics, Univ. of Oregon) ( 2010 April 15). Data Mining the University: College GPA Predictions from SAT Scores. https://arxiv.org/abs/1004.2731

[2]Phillippa Poole,Boaz Shulruf, Joy Rudland & Tim Wilkinson (2012).Comparison of UMAT scores and GPA in prediction of performance in medical school: national study.https://asmepublications.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2923.2011.04078.x

[3]Raden Gunawan Santosa , Yuan Lukito , Antonius Rachmat Chrismanto(2021). Classification and Prediction of Students' GPA Using KMeans Clustering Algorithm to Assist Student AdmissionProcess. https://pdfs.semanticscholar.org/f69a/a6d5634cc93264c85423bce8a1b0f7302b6c.pdf

[4]Rogers Elliott, A. Christopher Strenta (1988). Effects of Improving the Reliability of the GPA on Prediction Generally and on Comparative Predictions for Gender and Race Particularly. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1745-3984.1988.tb00312.x

[5]Yue Yun, Ruoqi Cao, Huan Dai, Yupei Zhang & Xuequn Shang(2023). Self-paced graph memory for learner GPA prediction and it's application in learner multiple evaluation. https://www.nature.com/articles/s41598-023-48690-5

[6]Mashael A. Al-Barrak and Muna Al-Razgan(2016). Predicting Students Final GPA Using Decision Trees: A Case Study. https://www.researchgate.net/profile/Muna-Al-Razgan/publication/276422143_Predicting_Students_Final_GPA_Using_Decision_Trees_A_Case_Study/links/5640bfd508aec448fa6002db/Predicting-Students-Final-GPA-Using-Decision-Trees-A-Case-Study.pdf

[7]Narongsak Putpuek; Natcha Rojanaprasert; Kanokwan Atchariya (2018). Comparative Study of Prediction Models for Final GPA Score: A Case Study of Rajabhat Rajanagarindra University. https://ieeexplore.ieee.org/document/8466475/authors#authors

[8]Harjanto Prabowo; Alam Ahmad Hidayat; Tjeng Wawan Ceng(2021). Aggregating Time Series and Tabular Data in Deep Learning Model for University Students' GPA Prediction. https://ieeexplore.ieee.org/document/9452125?denied=

[9]Lukáš Falát & Terézia Piscová (2022). Predicting GPA of University Students with Supervised Regression Machine Learning Models. https://www.mdpi.com/2076-3417/12/17/8403

[10]Zafar Iqbal, Junaid Qadir, Adnan Noor Mian, Faisal Kamiran (2017). Machine Learning Based Student Grade Prediction: A Case Study. https://arxiv.org/abs/1708.08744

[11]Mariel F. Musso, Carlos Felipe Rodríguez Hernández & Eduardo C. Cascallar (2020). Predicting key educational outcomes in academic trajectories: a machine-learning approach. https://link.springer.com/article/10.1007/s10734-020-00520-7

[12]Mohammed Nasiru Yakubu, A. Mohammed Abubakar (2021). Applying machine learning approach to predict students' performance in higher educational institutions. https://www.emerald.com/insight/content/doi/10.1108/K-12-2020-0865/full/html

[13]Derinsha Canagareddy, Khuslendra Subarayadu & Visham Hurbungs (2019). A Machine Learning Model to Predict the Performance of University Students. https://link.springer.com/chapter/10.1007/978-3-030-18240-3_29

[14]Suchithra Rajendran a b, S Chamundeswari c, Akhouri Amitanand Sinha (2022). Predicting the academic performance of middle- and high-school students using machine learning algorithms. https://www.sciencedirect.com/science/article/pii/S2590291122001115

[15]Thi Dinh,Ha Dinh,Thi To Loan Pham,Loan Loan,Giap Cu Nguyen,Nguyen Thi,Nguyen Thi Lien Huong(April 2020).Informatic Department Faculty of Management Information System and E-commerce, Thuong Mai University Hanoi, Vietnam. https://www.researchgate.net/profile/Thi-Pham-21/publication/340351415_An_Empirical_Study_for_Student_Academic_Performance_Prediction_Using_Machine_Learning_Techniques/links/5e9d4f354585150839ebdf4d/An-Empirical-Study-for-Student-Academic-Performance-Prediction-Using-Machine-Learning-Techniques.pdf

[16]Nupur Chauhan,Kimaya Shah,Divya Karn,Jignasha Dalal(April 8, 2019).University of Mumbai, K. J. Somaiya Institute of Engineering and Information Technology, Department of Computer Science and Engineering. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3370802

[17]Mirna Nachouki, Mahmoud Abou Naaj(2022). International Journal of Distance Education Technologies (IJDET) . https://www.igi-global.com/article/predicting-student-performance-to-improve-academic-advising-using-the-random-forest-algorithm/296702

[18]Njoud ALANGARI,Raad ALTURKI (2020).Department of Computer Science,Al-Imam Mohammad Ibn Saud Islamic University , Riyadh, Saudi Arabia. http://romjist.ro/full-texts/paper660.pdf

[19]Efrem Yohannes Obsie Hawassa University, Seid Ahmed Adem (2018). Prediction of Student Academic Performance using Neural Network, Linear Regression and Support Vector Regression: A Case Study. https://www.researchgate.net/publication/325222214_Prediction_of_Student_Academic_Performance_using_Neural_Network_Linear_Regression_and_Support_Vector_Regression_A_Case_Study

[20]Alabi Mudasiru Adebayo1 & Issa Suleman (2017). Component Regression Models To Predict The Student's Final Grade Point Average (GPA) in the Polytechnic. https://seahipaj.org/journals-ci/dec-2017/IJIMSEP/full/IJIMSEP-D-3-2017.pdf

harish s

ORIGINALITY REPORT

| 50% | 12% | 9% | 44% |
| --- | --- | --- | --- |
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

38%

★ Submitted to University of Illinois at Urbana-Champaign

Student Paper

| | | | |
| --- | --- | --- | --- |
| Exclude quotes | Off | Exclude matches | Off |
| Exclude bibliography | Off | | |