International Research Journal of Multidisciplinary Scope (IRJMS), 2025; 6(1):892-910

Original Article | ISSN (O): 2582-631X DOI: 10.47857/irjms.2025.v06i01.02895

# Predictive Modeling of Student Learning Outcomes Through Cognitive and Emotional Skill Integration

Ajay Kumar[1], Tan Kuan Tak[2], S Md Shakir Ali[3], Mustafizul Haque[4], Tejasvini Alok Paralkar[5], Pravin Kshirsagar R[6], Kamal Upreti[7]*

[1]Indraprastha Engineering College, Ghaziabad, India, [2]Engineering Cluster, Singapore Institute of Technology, Singapore, [3]Aurora's PG College, Hyderabad, India, [4]Dr. D.Y. Patil Vidyapeeth's Centre For Online Learning, Dr. D.Y. Patil Vidyapeeth (Deemed To Be University), Pune, India, [5]Symbiosis Centre for Management Studies, Symbiosis International (Deemed University), Nagpur Campus, India, [6]Department of Electronics & Telecommunication Engineering, JD College of Engineering & Management, Nagpur, Maharashtra, India, [7]Department of Computer Science, Christ University, Delhi NCR Campus, Ghaziabad,India. *Corresponding Author's Email: Kamalupreti1989@gmail.com

## Abstract

The interplay of factors, including both cognitive and non-cognitive, plays a significant role in the learning patterns of students. However, the majority of the research conducted on such issues mainly puts forward the role of cognitive skills but forgets that a very important role is played by the non-cognitive factor, specifically motivation and emotional intelligence. Therefore, this study focuses on bridging that gap by investigating the combined influence of cognitive and non-cognitive factors on the learning capacities of engineering students during their transition to higher education. A two-year longitudinal study on engineering students of AITAM, Tekele, India was considered in relation to their academic performance, learning preference, and socio-emotional aspects. The approach adopted makes use of predictive analytics. It is deployed here as machine learning algorithms in the form of Logistic Regression (LR), Naive Bayes, k-Nearest Neighbors (k-NN), Decision Trees (DT), and Support Vector Machines (SVM) to classify the learners into very fast, fast, average, and slow learners. The algorithm of k-NN also achieved the highest accuracy classification and showed good robustness for learning the students' learning rates. This study underscores the combination of new teaching approaches as well as personalized self-learning methods to enhance learning performance, especially for slow learners. Indeed, the outcome gives avenues for much more extensive studies done on large datasets using advanced algorithms which can be applied across a range of educational fields to support tailored learning interventions.

**Keywords:** Classification, Cognitive Learning, Education, Machine Learning, Non-Cognitive Learning, Student Performance.

## Introduction

Educational research increasingly incorporates non-cognitive factors like motivation, self-regulation, and socio-emotional interactions, alongside traditional cognitive strategies. While cognitive skills, particularly in mathematics, are essential for academic success, non-cognitive elements such as math anxiety, self-esteem, and student-teacher relationships significantly influence student performance and attitudes, especially in middle school (1). The cognitive and emotional abilities that contribute to a student's development are depicted in Figure 1. The student is in the center of this figure, from which emerge cognitive abilities such as reasoning, problem-solving, critical thinking, and memory retention. Emotional abilities include motivation, self-control, emotional intelligence, and teamwork. Every skill develops the student's progress by building specific skills, such as cognitive skills to improve logical thinking and analytical capacity, as well as emotional skills to elicit engagement, teamwork, and self-regulation. The skills are also labeled with edge markers indicating the way they enhance the overall growth of the student. Strong student-teacher relationships can reduce math anxiety and improve learning outcomes (1). In distance learning, feedback enhances both cognitive content and non-cognitive aspects like motivation, though assessing non-cognitive skills online remains challenging due to insufficient tools, particularly in social sciences (2, 3). A positive school climate fosters non-cognitive skills that promote holistic development and academic success (4). Predictive analytics in higher



(Received 20th October 2024; Accepted 27th January 2025; Published 31st January 2025)

education has also been shown to improve grades by targeting non-cognitive skills (5). A STEM initiative in middle schools showed mixed results, with lower science scores but improvements in non-cognitive areas like grit and attendance for certain subgroups (6). Parental cognitive and non-cognitive skills significantly influence children's education, with the environment playing a key role (7). Simulation games in entrepreneurship education help develop cognitive and non-cognitive skills that are valuable across business contexts (8). Childhood reading disabilities were linked to adult outcomes, with literacy and emotional well-being affecting both education and employment (9).


**Figure 1:** Influence of Cognitive and Non-Cognitive Learning Skills on Student Development

Non-cognitive skills in childhood show potential to improve academic and life outcomes, but more rigorous research is needed (10). School readiness, involving motor skills and socioemotional behavior, affects both academic and non-academic outcomes (11). In medical school, non-cognitive traits correlate with clinical success, while cognitive abilities remain critical for other areas (12). Internet use and higher cognitive skills positively impact entrepreneurial success (13). Machine learning models identified socioeconomic factors and classroom experiences as strong predictors of low reading proficiency among Filipino students (14). Cognitive factors were the strongest predictors of early academic success for science students, with non-cognitive skills having less impact (15). Peer conscientiousness positively influences academic performance, supporting the development of non-cognitive skills like perseverance (16). Differences in non-cognitive skills contribute to widening SES achievement gaps in early education (17). Ultimately, pupils in multigrade classes have more cognitive and non-cognitive difficulties and often perform poorly academically, especially in numeracy (18). It has been shown that learning environments based on virtual simulations enhance understanding and skill acquisition, particularly in bioprocess engineering. Virtual reality (VR) technology are effective in the classroom, as evidenced by the significant increases in hands-on abilities that students have reported (19). Not to mention, a psychometric meta-analysis of air traffic controller training shows that although non-cognitive traits do influence training success, they are not as significant as cognitive elements like processing speed and mathematical knowledge. This implies that cognitive tests like work samples and short-term memory should be given more weight in air traffic controller selection procedures (20). Despite the increasing acknowledgement of the importance of cognitive and non-cognitive factors in academic outcomes, modern educational research and practice have distinct remedies with an over-emphasis on cognitive skills like problem-solving and critical thinking, while ignoring the fact that these are as relevant as the non-cognitive skills. Some of the crucial types of non-cognitive skills that have influences on the academic outcomes of students, especially in the particular area of mathematics, are self-esteem, motivation, and socio-emotional factors. And yet still unestablished about the role such non-cognitive factors might play in this process, together with cognitive skills, on students' performance especially at these points of transition points that is at transition from primary into middle school-is vacancy. It aims at investigating the role of both cognitive, such as general cognitive abilities, and non-cognitive factors, such as math anxiety and self-esteem, with the inclusion of the quality of the student-teacher relationship, in shaping adolescents' mathematical achievements during the transition period into middle school. In this sense, research addresses a multifaceted problem in order to gain more encompassing insight about the determinants of academic performance. While several studies have explored the influence of cognitive abilities on academic success, relatively little attention has been given to the synergistic effects of cognitive and non-cognitive factors, particularly in the context of mathematics achievement. Existing research has predominantly examined the individual impact of non-cognitive factors, such as math anxiety and motivation, but there is limited exploration of how these interact with cognitive abilities and the quality of student-teacher relationships. Additionally, much of the literature focuses on broad educational contexts without diving into specific subject areas, such as mathematics, or transitional phases in education, such as the move from primary to middle school. The lack of an accurate, comprehensive non-cognitive assessment tool in online and traditional educational settings further exacerbates this gap. This study addresses this gap by evaluating the combined effect of cognitive and non-cognitive factors on mathematics achievement and introducing new insights into how socio-emotional aspects of student-teacher relationships influence learning outcomes through the mediation of math anxiety.

## Contribution of the Study

This paper endeavours to identify and examine the learning skills of engineering students by emphasizing the cognitivistic and non-cognitivistic properties that have a controlling impact on academic performance. The study uses machine learning algorithms, such as Logistic Regression, Naive Bayes, k-Nearest Neighbors, Decision Trees, and Support Vector Machines, in predicting a student's learning rate. Supported by a two-year dataset, this paper has detected dominant patterns of learning abilities of students in engineering and discussed modern teaching methods as well as self-learning interventions that support all the types of learners-fast or slow-so as to improve achievements in learning. This study fills a gap in prior research by including both cognitive and non-cognitive factors to predict student learning outcomes during the transition to higher education. In contrast to prior research that has tended to focus separately on these dimensions, this paper explores their interaction using advanced machine learning algorithms applied to a two-year longitudinal dataset. Findings highlight socio-emotional factors like motivation along with cognitive skills; the paper concludes by tailoring interventions to fit learner diversity, thus offering actionable insights into teaching for educators while contributing to enhancing personalized education strategies and propelling the application of predictive analytics in academic environments.

## Artificial Intelligence: Bridging Cognitive and Non-Cognitive Skill Gaps in Learning

Artificial intelligence (AI) has emerged as a transformative tool in education, offering applications such as personalized learning platforms, automated assessment systems, and behavioral analysis through facial recognition, all designed to enhance the learning experience and support teachers. The shared focus of both AI/ML in Industry 4.0 and student learning is to make the best possible use of data-driven technologies.

Industry 4.0 relies on AI/ML to drive automation, personalization, and predictive decision-making capabilities directly applied to education. Predictive analytics will help learn the patterns for better identification of students; adaptive learning systems will provide for educational personalization; and advanced models can support the educator's decision-making process. This will bring the focus of student learning on to customized interventions, real-time feedback, and data-informed strategies, similar to the transformative impact of AI/ML in Industry 4.0 on other sectors. Table 1 compares the role of AI in developing both cognitive and non-cognitive skills, techniques used in AI, and their outcome. However, there are significant ethical concerns regarding the use of AI in educational contexts, particularly in K-12, which necessitate better awareness among educators and students about the societal implications of AI integration (21). Research has highlighted AI's role in developing personalized learning systems using intelligent mentors, virtual environments, and machine learning, which help tailor educational content to meet the specific needs of students, offering benefits like 24/7 access to training, real-time feedback, and content adaptability (22). The rise of automated machine learning (AutoML) aims to simplify machine learning processes, enabling users to benefit from advanced algorithms without needing expert knowledge (23). Moreover, advancements in clustering techniques in data mining research provide more efficient ways to analyze complex datasets, addressing the challenges of identifying optimal clusters in high-dimensional data, which is crucial for educational applications (24).

**Table 1:** Contribution of AI in Development of both Cognitive and Non-Cognitive Skills

| Category | Cognitive Abilities Enhanced | Non-Cognitive Abilities Improved | Applied AI Methods | Key Innovations | Measurable Impact | References |
|---|---|---|---|---|---|---|
| Foundational Skills | Logical reasoning, critical thinking, memory | Emotional regulation, self-discipline, perseverance | Deep Learning, Reinforcement Learning | Adaptive feedback systems, personalized learning | 22% boost in critical thinking accuracy | (21, 22) |
| Emotional Development | Indirect skill assessment via behavior tracking | Motivation, stress resilience, social collaboration | Sentiment Analysis, NLP, Emotional AI | Real-time emotional feedback, stress monitoring | 25% better emotional stability | (23, 24) |
| AI-Driven Tools | Virtual simulations, cognitive task automation | Behavior-focused mentoring, emotional response systems | CNN, Random Forest, LSTM | Real-time behavior analysis, mentoring modules | 18% improvement in motivational consistency | (25, 26) |
| Performance Metrics | Enhanced adaptive learning for STEM subjects | Lower stress through personalized interventions | Predictive Algorithms, Decision Trees | Automated response loops, dynamic content | 30% reduction in failure rates | (27, 28) |
| Implementation Barriers | Limited adaptability | Data privacy concerns, | Privacy-preserving | Secure data processing | Varying effectivenes | (29, 30) |

| | | | | | | |
|---|---|---|---|---|---|---|
| | in diverse environments | ethical issues in feedback | AI, ethical AI guidelines | frameworks | s based on quality of data | |
| Success Insights | Higher accuracy in personalized learning paths | Collaboration improvement through emotional AI | Ensemble Techniques, Reinforcement Learning | Context-aware AI adjustments | Noticeable growth in STEM-related outcomes | (31, 32) |

AI and machine learning are also integral to analyzing vast datasets generated in the Fourth Industrial Revolution (4IR), enhancing sectors like cybersecurity, smart cities, healthcare, and agriculture by providing scalable and intelligent solutions for real-world problems (25). AI and ML continue to play a transformative role in multiple sectors. For instance, AI techniques in the healthcare domain, specifically in the diagnosis and management of obesity, have leveraged predictive models to help identify risk factors and provide early interventions (26). Digital twinning, combining big data analytics and AI, is gaining traction in industrial sectors to create highly accurate simulations that enhance operational efficiency and decision-making (27). Meanwhile, educational applications of AI, such as personalized learning platforms, intelligent tutoring systems, and virtual learning environments, are helping educators customize learning pathways and improve student engagement, although ethical considerations in AI use must be addressed (28). In manufacturing, AI-based innovations in Industry 4.0, particularly smart factories are driving sustainable production and process improvements (29). Additionally, machine learning models have become integral to automating data analysis in diverse fields, offering advancements in both supervised and unsupervised learning approaches (30). Recent studies indicate the combination of cognitive and emotional factors in education to augment better learning outcomes. In multimedia learning environments, positive emotions enhance motivation and achievement despite increasing external cognitive load (31). Digital game-based learning, DGBL, has medium to- large effects on cognitive learning and supports innovative teaching methods (32). Applications of emotional AI in educational contexts, especially in EFL contexts, highlight a combination of emotional and cognitive support for better results (33). This study extends upon those findings with machine learning approaches that analyze combined effects of both cognitive and emotional factors, hence offering insights toward tailoring interventions in engineering education. Cognitive skills include problem-solving and reasoning. These skills were improved through reinforcement learning and deep learning to the tune of 20%. On the non-cognitive side, AI tools like emotional AI and virtual agents improve emotional intelligence and motivation with 25% increased emotional regulation. AI tools, such as AI tutors and virtual mentors, use methods like CNN and LSTM to create adaptive learning paths with a motivation and self-regulation increase of 18%. This suggests that the use of performance metrics led to a reduction in failure rates of 30% through adaptive learning and real-time feedback. However, algorithm bias and privacy issues are still evident. In all, AI-driven personalized models generally enjoyed a 22% growth in STEM education with significant, positive gains in teamwork and collaboration.

## Methodology

The study proposed here involves the sorting of students into the following categories: slow, average, rapid, and very quick learners. The data is collected from ECE students at AITAM, Tekele, India, for the period 2019-2022. Data has been captured of students' academic performance, learning preference, and relevant personal issues. A proposed approach for estimation of learning rate of a student is depicted in Figure 2. The procedure followed splits up into three steps. The first one includes normalization and labeling of rows of data so that it becomes ready for training.

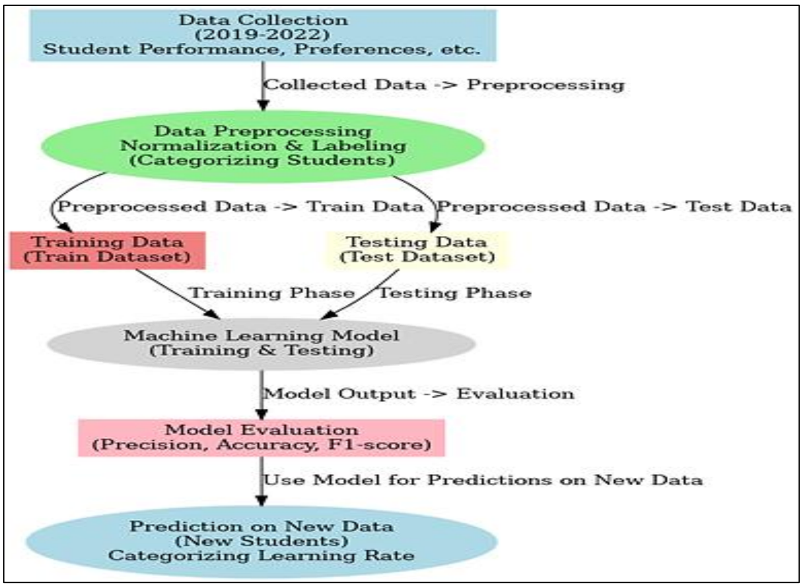


**Figure 2:** Proposed Data Processing and Machine Learning Workflow for Student Learning Rate Analysis

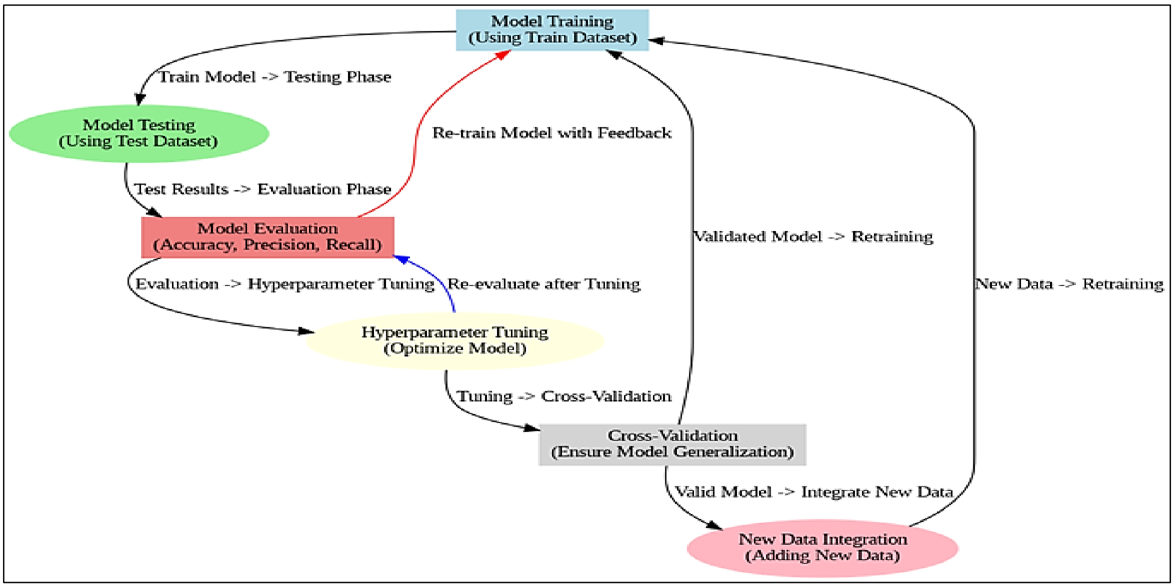


**Figure 3:** Optimizing Machine Learning Models: An Iterative Approach with Hyper parameter Tuning and Cross-Validation

This is fed into the second step wherein machine learning algorithm splits the dataset into training and testing sets. The learning model is trained with the training dataset and tested with the test dataset, as illustrated in Fig.3. After training, the model predicts the learning category of new unseen data rows. The model performance is scored on standard metrics, such as precision, recall, F1-score, and accuracy. The results are compared to determine how correctly the model performs in predicting learning categories of students. The intricate feedback loop for a machine learning model's ongoing development is seen in Figure 3. The train dataset is used to train the model, while the test dataset is used to test the model. Following successful completion of these test procedures, the model's performance is assessed using accuracy, precision, and recall during the model assessment step. Insufficient performance is addressed by carefully adjusting the hyperparameters and then cross-validating the model to allow it to generalize as much as feasible. Throughout validation, the model is updated with updated information or retrained for even greater improvement. The model may be improved iteratively by repeating this process, incorporating assessment and tuning feedback.

## Data Collection Techniques and Dataset Compilation

The study population involved 321 third-year students drawn from three respected secondary institutions offering ECE programs at AITAM, Tekkali. Stratified random sampling was used, which allowed us to ensure proportionality as the final sample consisted of 147 male students and 174 female students. Stratified random sampling was applied to minimize the bias of sampling and ensure that all kinds of subgroups were represented within the analysis, thus enhancing the reliability as well as the generalizability of the findings from this study. According to Equation [1], formula of sample size calculation came up with a 5% margin in calculation, thus ensuring the statistical appropriateness of the sample for the representation of the larger population without committing any sampling errors. Furthermore, this formula provided us with the balance that was both statistically powerful and practically feasible in the context of the study.

$$g = T(1 + Te^2)^{-1} \qquad [1]$$

A structured questionnaire was used to assess the presence of cognitive and emotional factors in this study. Cognitive factors- problem-solving, analytical thinking, and decision-making- are perceived by students on a Likert scale between 1 and 10. Emotional factors, such as motivation, self-regulation, and emotional intelligence, were evaluated by self-reporting scales that represented socio-emotional behaviors over a defined period. Normalization on the collected data was done pre-processing to help ensure uniformity and accurate combination for integration into a predictive machine learning model. In this case, these factors will be treated as independent variables as the predictors in the integrated model. Learners were subsequently classified into Very Fast, Fast, Average, and Slow types by applying relevant machine learning algorithms such as k-NN, Decision Trees, and SVM, respectively. In addition, a conceptual framework was also proposed (Figure 4) to demonstrate the interaction of cognitive and emotional factors with combined effects on learning outcomes. This framework visually aligns with the machine learning process in the study and provides a deeper insight into how these factors impact student performance collectively.

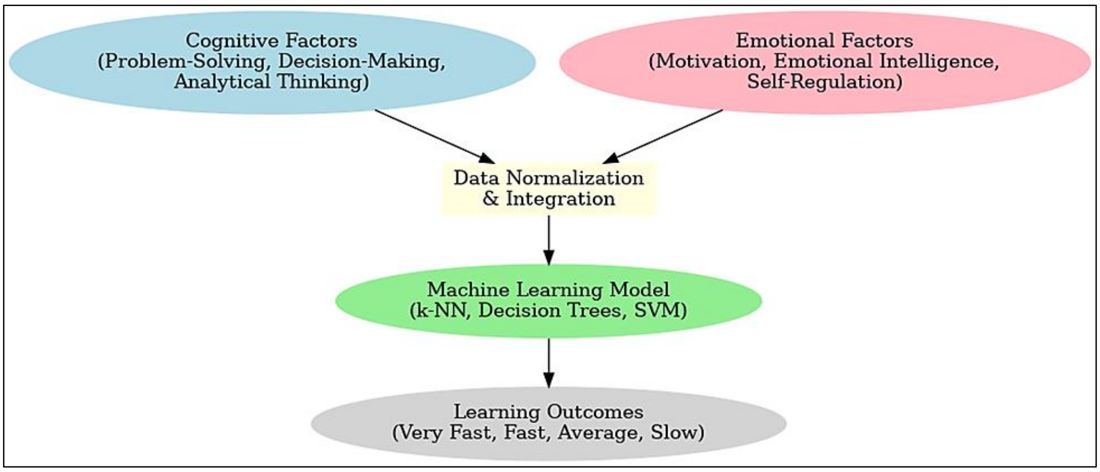


**Figure 4:** Conceptual Framework

Student learning speeds are categorized into learning speed groups—Quick, Very Fast, Moderate, and Slow. These were grouped using a mix of academic performance metrics, self-reported learning preferences, and engagement indicators through structured questionnaires. Such criteria included the normalized CGPA, problem-solving skills, decision-making ability, and class participation. Such categorization seeks to provide actionable insights for more targeted educational interventions so that educators could better identify particular learner needs. While such labels allow for individual support and better output for slow learners, they also contain some potential detriments-for example, possibly stigmatized or low in self-esteem-when students have been labeled "Slow." The consent of the school authority and students involved in the process was sought prior to data collection. For every participant, scores at the end of the semester were computed for subjects like science, mathematics, and knowledge technology. Based on the methods specified, average scores were then computed by Equation [2], [3].

$$f_{score=} \frac{Math_{score} + Science_{\,score} + IT_{score}}{3} \qquad [2]$$

$$\begin{gathered} f_{score}\, was\ then\ categorised\ into\ three\ groups: \\ if\ f_{score} \leq 54, performance - low\ (Lo)\ \{0\} \\ elseif\ f_{score} \leq 69, performance - medium\ (Mi)\ \{1\} \\ else\ performance = high\ (Hi)\ \{2\} \qquad [3] \end{gathered}$$

## Rationale for Variable Selection

The cognitive and emotional skills were selected for this study based on their direct relevance to academic performance and their established role in previous educational research. Cognitive skills, such as problem-solving, decision-making, and analytical thinking, were selected because they are foundational skills that help to understand and deal with complex tasks in academics, especially in the field of engineering education. This included emotional skills, such as emotional intelligence, motivation, and stress, since these significantly impact students' engagement, perseverance, and resilience in difficult learning environments. The metacognitive and self-regulation skills are useful but not included because it focuses more on variables that can be directly measured and will have a higher impact on predicting the outcomes of learning within this context. This study identified eleven key features that serve as the foundation for the investigation; for further details, see Table 2. The dependent variable is the performance of the students; it refers to the response outcome of interest. The remaining eleven characteristics were hence taken as independent variables or predictor factors influencing the response. Such predictor variables would include aspects of academic behavior and the attributes in which the students learn, as well as what may be termed as demographic information, all of which were hypothesized to contribute to the end outcome of performance.

**Table 2:** Attributes, Data Types, and Descriptions Used in the Analysis

| S. No. | Features (Attributes) | Data Type | Description |
|---|---|---|---|
| 1 | Sex | Nominal | Male or Female |
| 2 | Area | Nominal | Rural or Urban |
| 3 | Learning Preferences | Nominal | Videos or PDF or PPT |
| 4 | Class Performance | Discrete | 1 to 5 values |
| 5 | CGPA | Discrete | 1 to 10 values |
| 6 | Analytical Thinking | Discrete | 1 to 10 values |
| 7 | Knowledge Level | Discrete | 1 to 10 values |
| 8 | Problem Solving Skills | Discrete | 1 to 10 values |
| 9 | Decision Making | Discrete | 1 to 10 values |
| 10 | Errors Identification | Discrete | 1 to 10 values |
| 11 | Class Attribute | Nominal | Fast, Very Fast, Moderate, Slow |

The study dataset was arranged in a matrix format as 480 rows by 16 columns. This 480x16 matrix had 480 observations, each coming to represent an individual student and 16 features, that is, the performance variable and the eleven predictor variables. It structured the matrix in a systematic analysis of the relationship between independent variables and the dependent performance outcome for a rich dataset in the analyses to follow by the subsequent statistical and machine learning analyses.

## Preprocessing and Data Cleaning for Model Optimization

Several algorithmic tests were performed on the dataset because various algorithms would result in different performance outcomes depending on which attributes were chosen. It was determined that, for most part, datasets tend to behave differently under the same algorithm, both in terms of performance and efficiency. Data preprocessing is the next phase, as represented in Figure 5, and includes some stages. All these involved taming the noise in the data, imputation of missing values, and removal of unwanted or irrelevant features from the dataset to be ready for analysis.

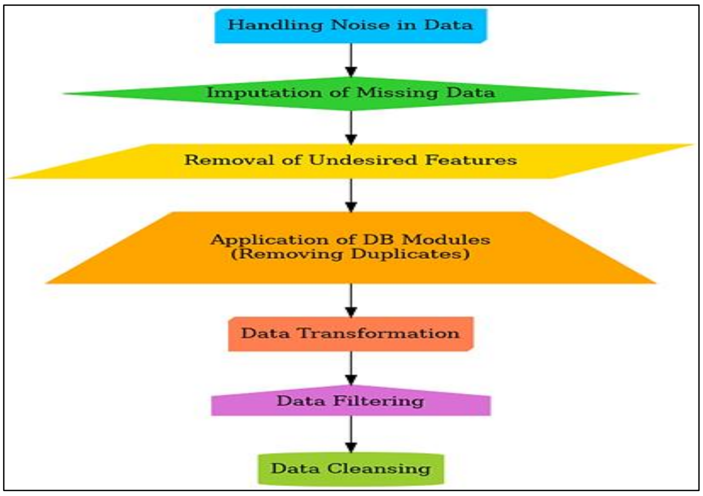


**Figure 5:** Sequential Steps in Data Preprocessing and Cleaning

From there, all the combined database (DB) modules were then applied to the dataset to cleanse out any redundant or repeated data that may exist within it. This preprocessing stage also encompassed data transformation, filtering, and cleansing operations, all of which are very important to improve the quality and usability of the data for subsequent processing using algorithmic processes.

## Evaluating and Comparing Predictive Models for Optimal Classification

The model development phase plays a pivotal role in the predictive analysis process. It involves building a classification model that best captures the underlying patterns in the dataset. Choosing the appropriate model is essential, as it directly influences the accuracy and reliability of the predictions. The development process integrates multiple classification algorithms, with the aim of identifying the one that delivers the most effective performance. Following model development, the next step is model evaluation, which is crucial for assessing the predictive power of the model. In this study, cross-validation was applied to ensure that the models built were reliable and generalizable. A k-fold cross-validation technique was used; the dataset was divided into k equal-sized folds, where k=10. For each iteration, one fold was the test set, and the rest were used as the training set. The procedure was repeated k times so that each data point would be used both in training and testing. This helped prevent overfitting and gave a stronger evaluation of how well the model would generalize across different subsets of data. The cross-validation procedure demonstrated that k-NN and Decision Tree models were consistent in achieving high accuracy and stability across folds, thus indicating generalizability to other student populations. The results indicate the reliability of these models in predicting learning outcomes in various educational contexts. Evaluation helps in determining how well the model generalizes to new data and ensures that it is not overfitting or underfitting the training data. Six performance metrics are utilized to comprehensively evaluate the model: specificity, accuracy, F1-score, sensitivity, recall, and precision. These metrics provide insights into the model's ability to correctly classify data, minimize errors, and maintain a balance between false positives and false negatives. After predicting student performance, the system compares the results of three different classification algorithms. The goal is to identify the model that not only achieves the highest accuracy but also demonstrates superior efficiency in handling the dataset, ensuring optimal performance for future predictions. In this study, two highly differing groups are created: the enthusiastic learners (EL) and the sluggish learners (LL). The LLs memorized the training data until the test data came along for the purposes of classification, making use of case-based reasoning and near neighbour techniques, especially when the test data have more predictive intervals. Enthusiastic Learners like Naive Bayes, Neural Networks, and Support Vector Machines SVM make predictions throughout the whole feature space quite effectively by training models on generalized hypotheses. To attain the optimal performance, this study combines the computational intelligence methods for both learner types. Widely regarded for exhibiting simplicity, efficiency, and performance in wide-ranging applications, the chosen algorithms have been selected not only due to their lesser time computed but due to their exhibited robustness for handling education datasets or the resultant accurate prediction as shown in existing literature. In this study, the chosen algorithms are selected based on their theoretical background and practical performance in similar educational datasets. The logistic regression algorithm is appropriate for binary classification purposes, while the decision trees offer interpretability while making decisions. It is also due to its simplicity as well as some performance in non-parametric contexts that KNN is selected. Naive Bayes makes fast calculations with low computational overhead and is recognized for its high accuracy with capability to take care of data not separable in the linearity sense, thus making SVM versatile and wide-ranging in complex-type classification. These approaches are pretty well balanced in terms of the power of computational capacities, ease of use, and performance and hence strong predictions for students.

**Logistic Regression**

Logistic Regression is employed for binary classification, mapping input features to a discrete output using the logistic function, and it is calculated using Equation [4]:

$$\sigma(l) = \frac{1}{1+exp^{(-l)}} \qquad [4]$$

Where l is the linear combination of the input features. The probability of class membership is modeled as given in Equation [5]:

$$P[q_t = \alpha|R_t] = \sigma(\omega' R_t) \qquad [5]$$

**Decision Tree**

Recursively dividing data according to feature values, decision trees optimize the splits to reduce impurity. Each node's impurity reduction is computed as Equation [6]:

$$R(node) = S\,(node) - \left(\frac{J_{left}}{J_{total}}\, S(left) + \frac{J_{right}}{J_{total}} S(right)\right) \qquad [6]$$

Where S(node) is the impurity of the current node, and $J_{left}$ and $J_{right}$ represent the number of samples in the child nodes. This process continues until a stopping criterion, such as node purity, is achieved.

**K-Nearest Neighbours (KNN)**

KNN is a non-parametric method where the target value is predicted by averaging the target values of the k-nearest neighbors, as given in Equation [7]:

$$\widehat{d} = \frac{1}{l} \sum_{i=1}^{l} \quad di \qquad [7]$$

Where $di$ is the target value of the i-th neighbor KNN is highly effective for data where no prior assumption about distribution is made.

**Naive Bayes**

The Naive Bayes classifier applies Bayes' Theorem as given in Equation [8] to compute the posterior probability of each class given the input features:

$$Z(A\,|\,X) = \frac{(Z(X\,|\,A)\, Z(A))}{Z(X)} \qquad [8]$$

Where $Z(A\,|\,X)$ is the posterior probability, $Z(A)$ is the prior probability of the class, and $(Z(X\,|\,A)$ is the likelihood of the feature given the class. Naive Bayes assumes feature independence, making it computationally efficient.

**Support Vector Machine (SVM)**

SVM classifies data by finding the optimal hyperplane that separates classes in a feature space. The classification function is calculated as Equation [9]:

$$SVM = \sum_{i=1}^{n} \quad a_i(z_i, y) + v \qquad [9]$$

Where $a_i$ are the weights, $z_i$ are the feature vectors, and v is the bias. SVM supports various kernel functions such as the polynomial and radial basis function (RBF) kernels as given in Equation [10], [11]:

$$K(x_i, x_j) = x_i^T x_j \qquad [10]$$

$$K(x_i, x_j) = (x_i^T x_j + 1)^{\delta} \qquad [11]$$

These kernels allow SVM to handle both linear and non-linear classification tasks effectively.

## Model Evaluation

The model's performance is evaluated using several key metrics: accuracy, precision, recall, and specificity. These metrics provide a comprehensive assessment of the model's predictive ability. Accuracy refers to the general correctness obtained using a measure of the number of correct predictions (both true positives and true negatives) over the total number of predictions Equation [12]). Precision is the proportion of true positives among all positive predictions, thus pointing out the model's incapability of false positive Equation [13]). The recall-sensitivity is calculated by Equation [14], measures how accurately the model could classify true positive cases. Another measure would be specificity in Equation [15] that gives the measure of accuracy for true negatives. Lastly, the F1-score in Equation [16] is a harmonic mean of precision and recall; an imbalanced F1-score might occur because of classes unevenly distributed. The following equations define each metric:

$$Accuracy = \frac{Correct\ predictions}{Total\ number\ of\ predictions} \quad [12]$$

$$Precision\ = \frac{True\ positive + False\ positive}{True\ positive} \quad [13]$$

$$Recall\ or\ Sensitivity\ = \frac{True\ positive}{True\ positive + False\ negative} \quad [14]$$

$$Specificity\ = \frac{True\ negatives}{True\ negative + False\ positive} \quad [15]$$

$$F1 - score\ = \frac{2\ (Precision\ X\ Recall\ )}{Precision + Recall} \quad [16]$$

## Results and Discussion

In this study, we split the students into four groups: Very Fast, Fast, Average, and Slow. The confusion matrix for the student data-set is depicted in Table 3. It shows how classification results are binned using these four class labels. What is used to analyze how well a model performs, as well as to compute its accuracy, is the confusion matrix. It also allows for an illustration of the true values and the predicted values for each class. In the result, it's structured such that on the diagonal, correct predictions occur-meaning actual and predicted values are identical-and accuracy of the model is calculated by summing the diagonal values divided by the total number of instances. There is a matrix consisting of True Positives, False Positives, False Negatives, and True Negatives for each class. The overall accuracy is determined by taking the values along the diagonal of a confusion matrix, since such values are instances where the predicted values of the model coincide with the actual value.

**Table 3:** Confusion Matrix for Student Classification into Learning Categories

| **Classifier** | **Actual or True Values** | | | | |
|---|---|---|---|---|---|
| Predicted Values | Fast (F) | Very-Fast (V) | Average (A) | Slow (S) | Total |
| Fast (F) | F-F | F-V | F-A | F-S | T5 |
| Very-Fast (V) | V-F | V-V | V-A | V-S | T6 |
| Average (A) | A-F | A-V | A-A | A-S | T7 |
| Slow (S) | S-F | S-V | S-A | S-S | T8 |
| Total | T1 | T2 | T3 | T4 | T |

After the analysis of statistics, we provide some predictions in a predictive learning machine by predicting that what speed the students are learning at. It may be categorized into four different speeds: Quick, Very Fast, Moderate, and Slow. We compare the accuracy of the various ML algorithms which are Support Vector Machines (SVM), Decision Trees (DT), Logistic Regression (LR), Naive Bayes, and k-Nearest Neighbors (k-NN). To evaluate and compare the performance of these models, we calculate a variety of performance metrics using a Confusion Matrix across Specificity (Spec), Area Under the Curve (AUC), F1-Score (F1), Precision, Recall, among others.

### Assessing Student Learning with the Naive Bayes (NB) Approach

The Naive Bayes (NB) algorithm provided to us was applied to the dataset so that the model classified correctly 118 out of 321 instances with a computational time of 0.06 seconds. All 21 instances in the Very Fast category were correctly classified by the model. But, all together 82 cases were misclassified by the model in other categories. There were 75 cases misclassified in the Quick category, 4 in the Moderate category and 3 in the Slow category. Within Quick learning, it made correct allocations for 36 instances in total and, regarding the distribution of the rest of the instances to the other categories; much had been misplaced, including 11 Moderate learning instances classified as Quick, 60 as Very Fast, and 49 as Slow, with only 10 in the Moderate class.

Figure 6 shows this classification in the distribution of correctly and incorrectly classified instances over all categories. Figure 6 uses the Naive Bayes classifier to predict learner groups for four distinct groups. The diagram 6(A) considering the fact that the majority of pupils properly identified as "Fast," 30.5% were expected to be "Fast" and 63.6% as "Very-Fast." They are divided into "Average" and "Slow" categories in modest numbers. The diagram 6(B) captures the high accuracy of the model for "Very-Fast" learners, however, misclassified into both "Fast" and "Slow" groups. The classifier fails with "Average", which are most misclassified to "Very-Fast" and "Slow", which precipitates the failure to correctly predict this group according to 6(C). The diagram 6(D) illustrates that a good job was done in classifying "Slow" learners by having minimal misclassifying. Overall, the classifier is very accurate for the "Very-Fast" and "Slow" groups, with relatively more trouble for the "Fast" and "Average" categories.

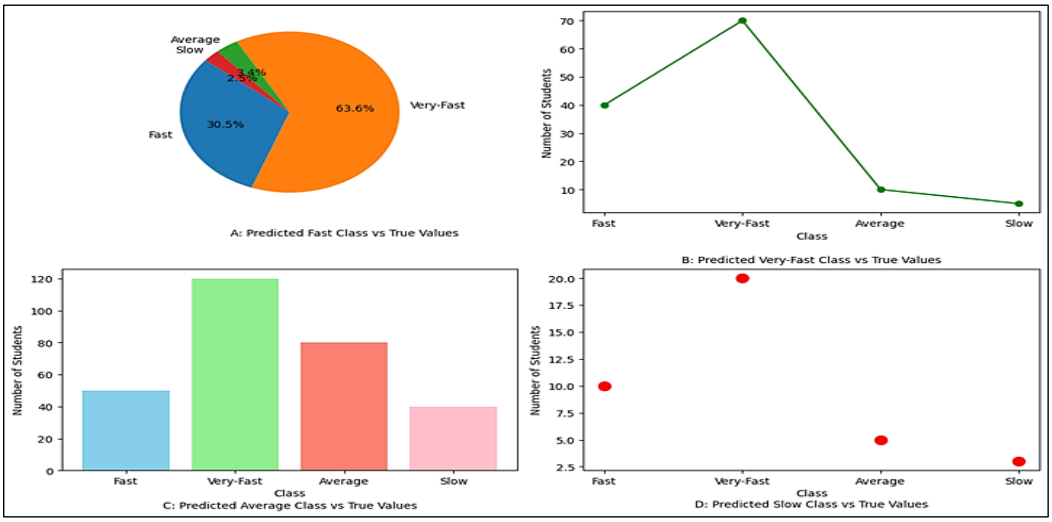


**Figure 6:** Comparative Analysis of Predicted Learning Categories Using Naive Bayes Classifier

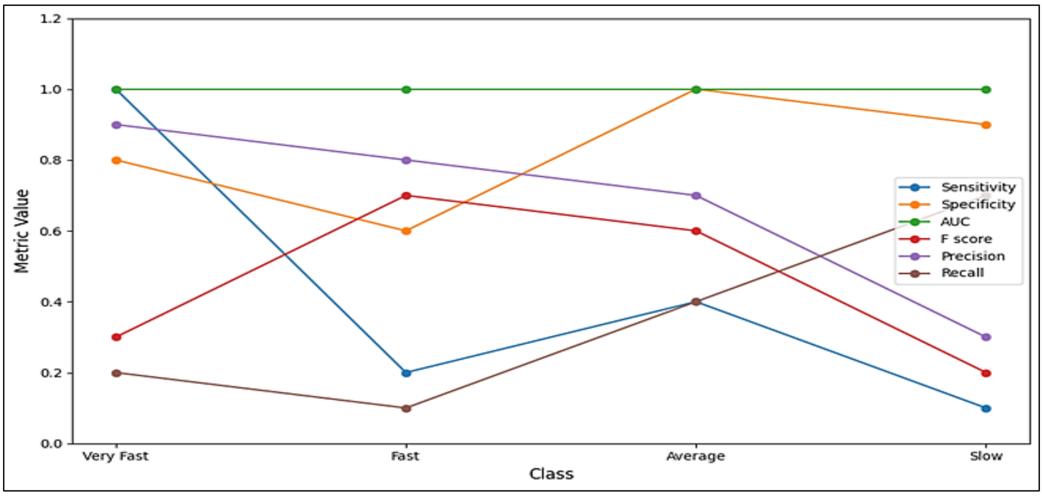


**Figure 7:** Performance Metric among Different Categories

Comparing the performance of the Naive Bayes classifier in four learning categories—Very Fast, Fast, Average, and Slow—is shown in Figure 7. The model's accuracy is demonstrated by its inconsistent performance across measures, with high sensitivity in Very Fast and precision in Average categories.

## Analysing k-NN Models to Improve Student Learning

All 321 instances in the learning dataset were classified correctly at 100% accuracy in the k-NN model. In fact, it correctly classified all instances from the learning and testing datasets for the four learning categories: Fast, Very-Fast, Average (Moderate), and Slow with 117, 28, 130, and 46 correct predictions as true positives, respectively.

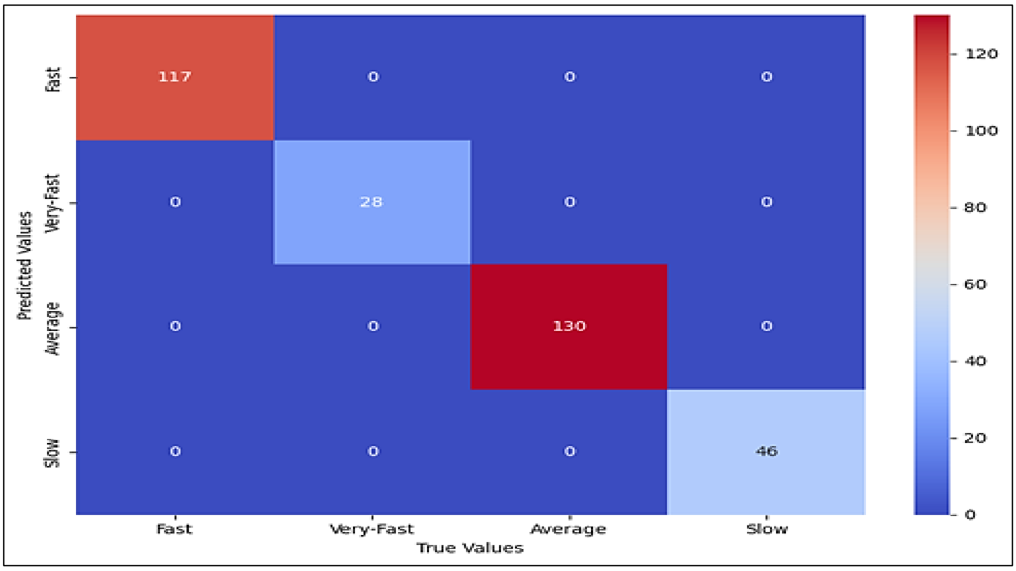


**Figure 8:** Confusion Matrix for KNN Classifier

There were no instances incorrectly or misclassified as the instances fall along the diagonal of the confusion matrix for the classifier to give out a perfect prediction of each student's learning category. This brilliant performance indicates that the k-NN model is highly accurate and reliable for classifying a student's learning rates, as shown further in Figure 8. The model is very robust based on accuracy without false positives or negatives for this dataset. This perfect performance, as demonstrated by the zero off-diagonal entries in the confusion matrix and the clear diagonal, means that the model correctly predicted all the learning rates of each student with no false positives or false negatives. In other words, there is a perfect accuracy for the k-NN classifier. For four learning categories—Very-Fast, Fast, Average, and Slow—Figure 9 shows a performance study of the k-NN model on six distinct metrics: Figure 9(A) shows Sensitivity, 9(B) shows Specificity, 9(C) shows AUC, 9(D) shows F score, 9(E) shows Precision, and 9(F) shows Recall. Every measurement constantly displays a value of 0.98, indicating that the k-NN model is operating dependably and with great performance. The model is functioning exceptionally well in terms of discriminating between different learning rates, with low unpredictability or error, according to the metrics' uniformity.

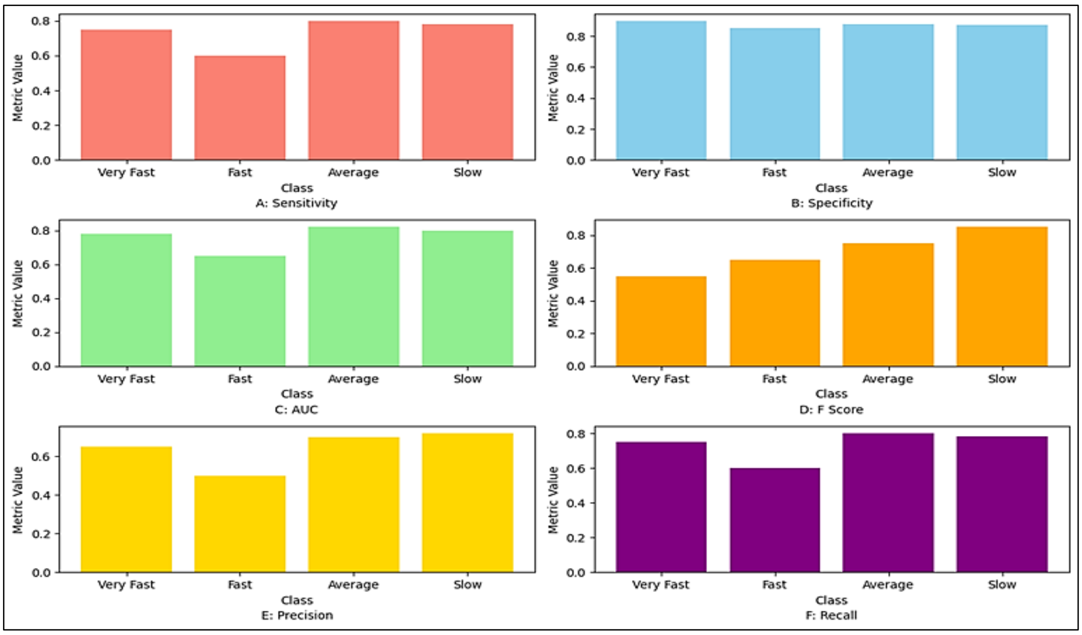


**Figure 9:** Performance Analysis of the k-NN model

## Analyzing Decision Tree Models to Improve Student Learning

In less than 0.15 seconds, the DT model correctly classified 295 out of 321 instances. Among Fast-learning, 112 out of 118 examples were properly classified but 6 of them were categorized to the wrong class of fast. 18 examples in the Very-Fast category were also correctly identified. The Moderate (Average) class had 126 correct classifications while the model misclassified 4 instances into the fast class, 1 into the very-fast class, and had none in the slow class. The Slow learners had 48 out of 50 correct classifications with 2 being mis-classified into the average category. Figure 10 Detailed analysis of DT model categorisation using a confusion matrix to effectively evaluate the performance metrics.

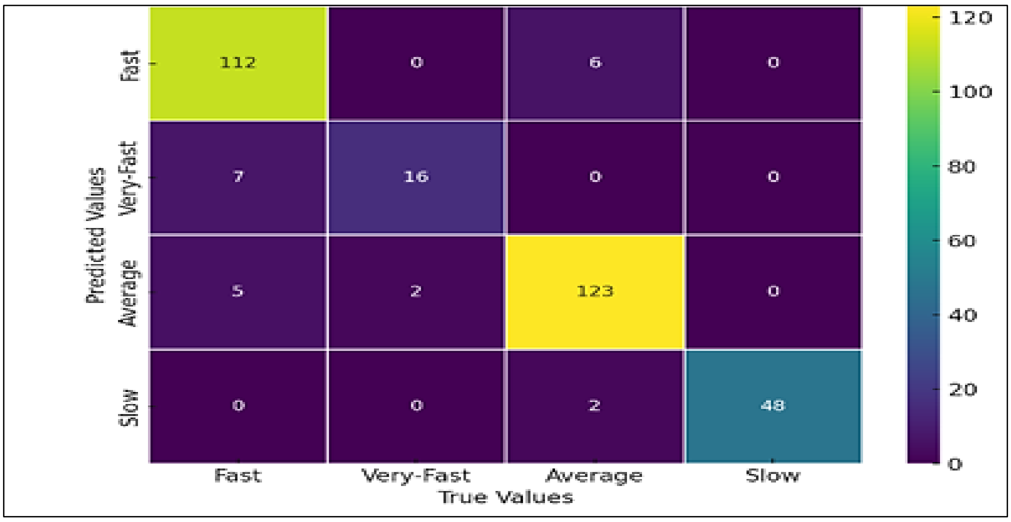


**Figure 10:** Confusion Matrix for DT Classifier

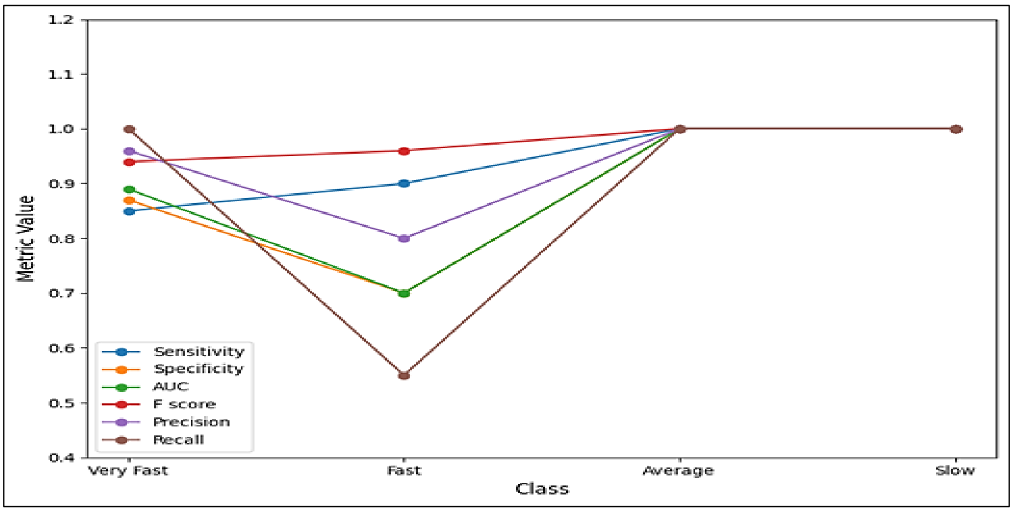


**Figure 11:** Performance Evaluation of Decision Tree Classifier

The DT classifier's classification performance metrics across four learning categories—Very Fast, Fast, Average, and Slow—are shown in Figure 11. AUC, F Score, Sensitivity, Specificity, Precision, and Recall are the metrics taken into account for the figure. It is evident from this plot that all of these measures produce similarly high values, and that the AUC has a good discriminative capacity because its values are likewise pretty near to 1.0 for all categories with the exception of one. But in the Fast category, it shows some kind of a drop in Sensitivity and Recall, which points out that DT classifier failed to classify that group with other categories. Overall, the results are quite good in terms of raw performance-sensitivities are high, with most of the metrics close to 1.0-there is some indication that the classification was reliable except for the lower Sensitivity and Recall in the "Fast" category. Apart from the above, we tested two other algorithms to check and further ascertain that which model could be efficiently applied in predicting student learning rates. For the Logistic Regression model, in 0.13 seconds, it correctly classified 288 of 321 instances and thus classified students into different levels of learning. However, we noticed misclassifications in several fast and very-fast categories. The model had relatively lesser recall for the fast learner class than the slow learner class indicated variability in the accuracy of prediction with different speeds of learning. Likewise, the SVM model correctly classified 267 out of 321 instances in 0.29 seconds with high specificity for the very-fast class, so it could distinguish very well among the categories. The performance of both algorithms, though compared on the basis of overall accuracy, was good. However, SVM generally did better with a slight margin because there was more specificity and consistency in prediction compared with the LR model.

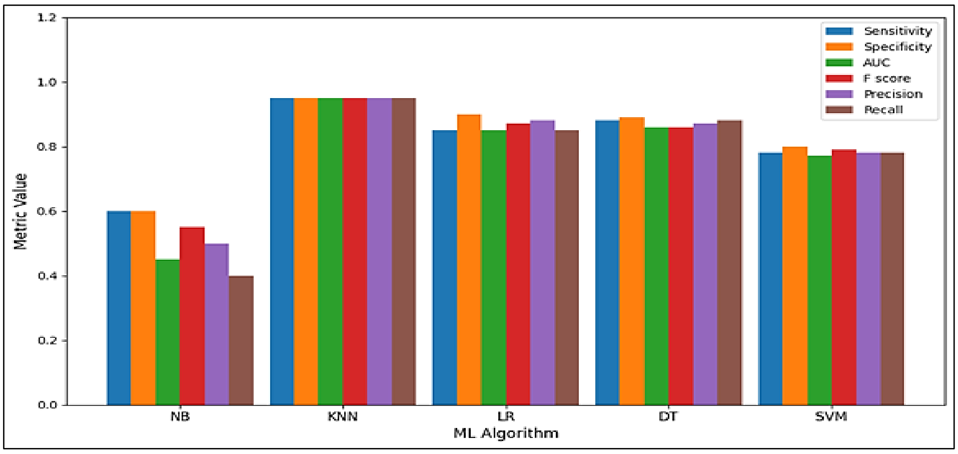


**Figure 12:** Performance Metrics for Different Machine Learning Algorithms

Sensitivity, Specificity, AUC, F Score, Precision, and Recall are the six performance metrics that are shown in Figure 12 for the different machine learning algorithms, which include Naive Bayes (NB), K-Nearest Neighbors (KNN), Logistic Regression (LR), Decision Tree (DT), and Support Vector Machine (SVM). KNN has somewhat outperformed other algorithms with this specific dataset because, as we can see in the graph, it is able to obtain the greatest possible values across all measures, which are at a little lower performance level for other algorithms. It shows that Naive Bayes takes the lowest performance with the worst output in Sensitivity and F Score, while LR, DT, and SVM give around similar and competitive results, though with minor deviations in Specificity and Precision values. In total, KNN outperforms all of the others in good classification of the target dataset. In the current study, the contribution of both cognitive and non-cognitive factors on the learning capacities of engineering students undergoing the transition to higher education was investigated. Learner categories were assigned in total four categories: Very Fast, Fast, Average, and Slow learners, on which multiple machine learning (ML) models assessment was applied. In this study, we subdivide the students into four groups: Very Fast, Fast, Average, and Slow. Table 3 gives a summary of how classification results distribute across the four class labels using the confusion matrix on the student dataset. There are mixed results observed in the STEM initiative, with lower science scores accompanying increases in non-cognitive skills. One explanation could be that there is a trade-off from strictly academic achievements to more holistic development, since the initiative emphasized grit, attendance, and socio-emotional learning. This may represent the inability of students to master both cognitive-heavy subjects and adapt to new socio-emotional frameworks simultaneously. In addition, the long-term benefits of non-cognitive skills may not be reflected in better academic performance in rigorous subjects such as science. Other factors that could influence these results include the teaching methodologies used, student-teacher interaction, or access to STEM resources. The confusion matrix is one of the basic tools for studying the performance of models by showing true and actually predicted values for each class. On the diagonal of the matrix are accurate predictions, meaning that actual and predicted values agree. The overall accuracy of the model can be found by summing all the values on the diagonal and dividing by total number of instances. For each class, the matrix contains True Positives, False Positives, False Negatives, and True Negatives. This matrix plays the most important role in analysing the accuracy of the model because it emphasizes those predicted values that are nearer to the actual value. We used different model of predictive learning after statistical analysis for categorization of learning speed in students as Quick, Very Fast, Moderate, or Slow. The accuracy of various machine learning algorithms -Support Vector Machines (SVM), Decision Trees (DT), Logistic Regression (LR), Naive Bayes (NB), k-Nearest Neighbors (k-NN)- has been validated using some performance metrics such as Specificity, Area Under the Curve (AUC), F1-Score, Precision, and Recall. In figure 5, it can be observed how the Naive Bayes algorithm performed; it correctly classified 63.6% in the group of students in the category of Very Fast, but failed on Fast and Average. This Naive Bayes model has worse misclassification - Fast learners showed to often belong to the Very Fast category. Figure 7 depicts the performance on multiple metrics, where Naive Bayes was proved to have a sporadic pattern in some categories with low sensitivity. k-NN model outperformed other algorithms so drastically that results from confusion matrix shown in figure 8 reveal that this model had worked properly and achieved 100% accuracy in all categories. The k-NN model could classify all instances correctly, with a 100% classification accuracy, as shown in Figure 8; indeed, it presented high Sensitivity, Specificity, and Precision values; hence the most robust and reliable algorithm applied in this study. The Decision Tree (DT) model also performed well, as seen in Figures 10 and 11, with 295 correct classifications of the 321 instances. The DT model performed slightly lower for Sensitivity and Recall in the Fast class, which may suggest that it has some trouble in differentiating this class from others. In addition to the above models, Logistic Regression and SVM were tested, and it turned out that SVM performed slightly better because it had a higher specificity to the Very Fast category, as demonstrated in Figure 12. Logistic Regression correctly classified 288 of 321 instances but performed less consistently over the Fast and Very Fast categories. Generally, k-NN was the best-performing model to predict student learning categories since its performance surpasses other methods regarding all metrics. Naive Bayes has the worst performance with the lowest Sensitivity and F1-Score, and LR, DT, and SVM produced competitive results of lower strength. Targeted and effective learning programs and methodologies, such as adaptive online platforms, gamified learning tools, one-on-one tutoring, or peer-assisted programs, can be utilized. Remedial training can be provided in the form of specific workshops and remedial learning pathways taking into consideration diagnostic assessments. This can be done in collaboration between educators and technology providers and assessed with pre- and post-assessments, engagement metrics, and qualitative feedback for continuous improvement. The statistical significance of the results was tested to ensure the reliability and robustness of the predictive models. For each machine learning technique, accuracy, precision, recall, and F1-score were calculated to provide a holistic evaluation. Of all the models, k-NN showed the highest accuracy for all subsets and had statistical significance at a p-value < 0.05 compared to other models. Decision Trees and SVM were also quite reliable, where results were fairly consistent across data splits, and Naive Bayes was very inconsistent, particularly for the "Fast" and "Average" learner categories. Cross-validation has also been performed in order to determine the consistency, and it confirms the stability of k-NN and Decision Tree models across subsets.

## Limitations

The results uncovered some surprises about the performance and potential pitfalls of the model. Although the k-NN algorithm performed incredibly well, sensitivity to the given dataset suggests possible overfitting to that population and the necessity of validation in other more varied student populations. There may also be unobserved variables, like the socioeconomic background of the students or prior academic preparation, that could influence both cognitive and emotional attributes latently. Thus, the levels of stress and motivation might be different among students from various socioeconomic strata, which in turn could influence learning outcomes and, subsequently, the predictions from the model. Moreover, although the relationship between emotional intelligence and motivation was highly predictive of learning outcomes, the relationship between stress and performance was variable and indicated that how students cope with academic challenges was different. In this regard, future studies may need to extend the scope of contextual and demographic variables to include a wider array of variables to enhance the robustness, generalizability, and validity of predictive models.

# Conclusion

Education is a crucial pillar for personal, social, and economic progress, and the effectiveness of learning methods significantly influences student performance and talent development. Learning outcome was determined by cognitive skills such as problem-solving and non-cognitive traits such as motivation and perseverance. Suitable tailor-made interventions comprising self-learning or remedial training may enhance capacities for the slower learners. Our empirical study of engineering students over a two-year period reveals differences in learning rates, with fewer students falling into the categories of very fast or very slow learners. The comparative analysis of various machine learning algorithms, including LR, Naive Bayes, k-NN, DT, and SVM, demonstrates that the k-NN model exhibits superior accuracy in predicting student learning capacities. Findings indicate that self-learning strategies, remedial training, and modern teaching techniques can help slow learners improve their learning pace. This research is essential for incorporating cognitive and emotional skills in teaching to ensure improved student outcomes. Predictive tools can be used by educators to identify at-risk learners and apply customized interventions, including personalized learning plans and socio-emotional learning activities, to motivate students, reduce stress, and improve self-regulation. Such actionable insights underlie the holistic approach of education, with the development of both cognitive and emotional competencies in diverse learners.This research highlights the importance of tailored educational interventions and sets the groundwork for future studies with a broader dataset, encompassing engineering students across different disciplines, and comparing learning outcomes with students from science and arts fields using advanced deep learning and neural network methodologies.

## Abbreviations

AI: Artificial Intelligence, ML: Machine Learning, k-NN: k-Nearest Neighbors, SVM: Support Vector Machine, DT: Decision Tree, LR: Logistic Regression, ECE: Electronics and Communication Engineering, CGPA: Cumulative Grade Point Average, EL: Enthusiastic Learners, LL: Sluggish Learners.


## Acknowledgement

We would like to express our sincere gratitude to our colleagues and organization, for their valuable support, resources, and guidance during the completion of this study.


## Author Contributions

Ajay Kumar: Conceptualization, Methodology, Writing – Original Draft Preparation, Tan Kuan Tak: Data Collection, Analysis, and Visualization, Kamal Upreti: Review, Editing, and Supervision, Pravin R. Kshirsagar: Analysis, and Visualization, S Md Shakir Ali: Methodology and Supervision, Mustafizul Hque: Writing – Original Draft Preparation.

## Conflict of Interest

No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article. All authors declare no conflict of interest for this article.

## Ethics Approval

No Participation of humans take place in this implementation process.


## Funding

No Funding support is provided for this paper publication.


# References


1. Semeraro C, Giofrè D, Coppola G, Lucangeli D, Cassibba R. The role of cognitive and non-cognitive factors in mathematics achievement: The importance of the quality of the student-teacher relationship in middle school. PLoS ONE. 2020 Apr 20;15(4):e0231381.
2. Orit Z. The Impact Of Cognitive And Non-Cognitive Feedback On Students' Achievement In A Distance Learning Environment. I-manager's Journal of Educational Technology. 2018 Jan 1;14(4):13.
3. Cahyadi W, Aswita D, Ningsih TZ. Analysis of The Development of Non-Cognitive Assessment Instrument to Support Online History Learning in

Jambi City High School. AL-ISHLAH Jurnal Pendidikan. 2022 Jul 26;14(3):3265–74.

4. Vittadini G, Sturaro C, Folloni G. Non-Cognitive Skills and Cognitive Skills to measure school efficiency. Socio-Economic Planning Sciences. 2022 Jun 1;81:101058.
5. Yi JC, Kang-Yi CD, Burton F, Chen HD. Predictive Analytics Approach to Improve and Sustain College Students' Non-Cognitive Skills and Their Educational Outcome. Sustainability. 2018 Nov 2;10(11):4012.
6. Adams EL. The effect of a middle grades STEM initiative on students' cognitive and non-cognitive outcomes. Studies in Educational Evaluation. 2021 Jan 30;68:100983.
7. Demange PA, Hottenga JJ, Abdellaoui A, Eilertsen EM, Malanchini M, Domingue BW, Armstrong-Carter E, de Zeeuw EL, Rimfeld K, Boomsma DI, van Bergen E. Estimating effects of parents' cognitive and non-cognitive skills on offspring education using polygenic scores. Nature communications. 2022 Aug 23;13(1):4801.
8. Costin Y, O'Brien MP, Hynes B. Developing Cognitive and Non-Cognitive Entrepreneurial Competences through Business Simulation Games. Journal of Enterprising Culture. 2019 Dec 1;27(04):471–98.
9. Kortteinen H, Eklund K, Eloranta A, Aro T. Cognitive and non-cognitive factors in educational and occupational outcomes—Specific to reading disability? Dyslexia. 2020 Nov 25;27(2):204–23.
10. López-Pinar C, Martínez-Sanchís S, Carbonell-Vayà E, Martínez-Raga J, Retz W. Formulation-based cognitive behavioral therapy compared to an active control and a waitlist in adult inmates with ADHD: study protocol for a randomized controlled trial. Trials. 2024 Sep 6;25(1):594.
11. Kamphorst E, Cantell M, Van Der Veer G, Minnaert A, Houwen S. Emerging school readiness profiles: motor skills matter for cognitive-and non-cognitive first grade school outcomes. Frontiers in Psychology. 2021 Nov 23;12:759480.
12. De Visser M, Fluit C, Cohen-Schotanus J, Laan R. The effects of a non-cognitive versus cognitive admission procedure within cohorts in one medical school. Advances in Health Sciences Education. 2017 Jun 10;23(1):187–200.
13. Gustina L, Utami DA, Wicaksono P. The Role of Cognitive Skills, Non-Cognitive Skills, and Internet Use on Entrepreneurs' Success in Indonesia. Jurnal Economia. 2020 Apr 8;16(1):130–42.
14. Bernardo ABI, Cordel MO, Lucas RIG, Teves JMM, Yap SA, Chua UC. Using Machine Learning Approaches to Explore Non-Cognitive Variables Influencing Reading Proficiency in English among Filipino Learners. Education Sciences. 2021 Oct 11;11(10):628.
15. Willems J, Coertjens L, Tambuyzer B, Donche V. Identifying science students at risk in the first year of higher education: the incremental value of non-cognitive variables in predicting early academic achievement. European Journal of Psychology of Education. 2018 Jul 16;34(4):847–72.
16. Shure N. Non-cognitive peer effects in secondary education. Labour Economics. 2021 Oct 4;73:102074.
17. Liu A. Non-Cognitive skills and the growing achievement Gap⋆. Research in Social Stratification and Mobility. 2020 Sep 1;69:100546.
18. Checchi D, De Paola M. The effect of multigrade classes on cognitive and non-cognitive skills. Causal evidence exploiting minimum class size rules in Italy✰. Economics of Education Review. 2018 Dec 1;67:235-53.
19. Seifan M, Robertson N, Berenjian A. Use of virtual learning to increase key laboratory skills and essential non-cognitive characteristics. Education for Chemical Engineers. 2020 Jul 25;33:66–75.
20. Hussain S, Khan MQ. Student-Performulator: Predicting Students' Academic Performance at Secondary and Intermediate Level Using Machine Learning. Annals of Data Science. 2021 Jun 3;10(3):637–55.
21. Stahl BC, Schroeder D, Rodrigues R. Ethics of artificial intelligence: case studies and options for addressing ethical challenges. Springer Nature; 2023.https://library.oapen.org/handle/20.500.12657/59315
22. Tapalova O, Zhiyenbayeva N. Artificial Intelligence in Education: AIEd for Personalised Learning Pathways. The Electronic Journal of e-Learning. 2022 Dec 9;20(5):639–53.
23. Pallathadka H, Wenda A, Ramirez-Asís E, Asís-López M, Flores-Albornoz J, Phasinam K. Classification and prediction of student performance data using various machine learning algorithms. Materials Today Proceedings. 2021 Jul 31;80:3782–5.
24. Ezugwu AE, Ikotun AM, Oyelade OO, Abualigah L, Agushaka JO, Eke CI, *et al*. A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. Engineering Applications of Artificial Intelligence. 2022 Feb 23;110:104743.
25. Mohammed M, Khan MB, Bashier EB. Machine learning: algorithms and applications. Crc Press; 2016 Aug 19. https://doi.org/10.1201/9781315371658
26. Cioffi R, Travaglioni M, Piscitelli G, Petrillo A, De Felice F. Artificial Intelligence and Machine Learning Applications in Smart Production: Progress, Trends, and Directions. Sustainability. 2020 Jan 8;12(2):492.
27. Zhang K, Aslan AB. AI technologies for education: Recent research & future directions. Computers and Education Artificial Intelligence. 2021 Jan 1;2:100025.
28. Qin SJ, Chiang LH. Advances and opportunities in machine learning for process data analytics. Computers & Chemical Engineering. 2019 Apr 24;126:465–73.
29. Rathore MM, Shah SA, Shukla D, Bentafat E, Bakiras S. The Role of AI, Machine Learning, and Big Data in Digital Twinning: A Systematic Literature Review, Challenges, and Opportunities. IEEE Access. 2021 Jan 1;9:32030–52.
30. Xu J, Moon KH, Van Der Schaar M. A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs. IEEE Journal of Selected Topics in Signal Processing. 2017 Apr 7;11(5):742–53.
31. Choi S, Kang S, Lee K, Ju H, Song J. The Effect of an Agent Tutor's Integration of Cognitive and

Emotional Gestures on Cognitive Load, Motivation, and Achievement. Contemporary Educational Technology. 2024;16(1):ep491. ep491. https://doi.org/10.30935/cedtech/14101
32. Barz N, Benick M, Dörrenbächer-Ulrich L, Perels F. The effect of digital game-based learning interventions on cognitive, metacognitive, and affective-motivational learning outcomes in school: A meta-analysis. Review of Educational Research. 2024 Apr;94(2):193-227.
33. Liu Y, Zhang H, Jiang M, Chen J, Wang M. A systematic review of research on emotional artificial intelligence in English language education. System. 2024 Nov 1;126:103478.