

## Article

# Machine Learning-Based Hybrid Ensemble Model Achieving Precision Education for Online Education Amid the Lockdown Period of COVID-19 Pandemic in Pakistan

Rimsha Asad <sup>1,\*</sup>, Saud Altaf <sup>1,\*</sup>, Shafiq Ahmad <sup>2</sup>, Haitham Mahmoud <sup>2</sup>, Shamsul Huda <sup>3</sup> and Sofia Iqbal <sup>4</sup>

<sup>1</sup> University Institute of Information Technology, Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi 46300, Pakistan

<sup>2</sup> Industrial Engineering Department, College of Engineering, King Saud University, Riyadh 11421, Saudi Arabia

<sup>3</sup> School of Information Technology, Deakin University, Burwood, VIC 3128, Australia

<sup>4</sup> Space and Upper Atmosphere Research Commission, Islamabad 44000, Pakistan

\* Correspondence: saud@uair.edu.pk

**Abstract:** Institutions of higher learning have made persistent efforts to provide students with a high-quality education. Educational data mining (EDM) enables academic institutions to gain insight into student data in order to extract information for making predictions. COVID-19 represents the most catastrophic pandemic in human history. As a result of the global pandemic, all educational systems were shifted to online learning (OL). Due to issues with accessing the internet, disinterest, and a lack of available tools, online education has proven challenging for many students. Acquiring accurate education has emerged as a major goal for the future of this popular medium of education. Therefore, the focus of this research was to identifying attributes that could help in students' performance prediction through a generalizable model achieving precision education in online education. The dataset used in this research was compiled from a survey taken primarily during the academic year of COVID-19, which was taken from the perspective of Pakistani university students. Five machine learning (ML) regressors were used in order to train the model, and its results were then analyzed. Comparatively, SVM has outperformed the other methods, yielding 87.5% accuracy, which was the highest of all the models tested. After that, an efficient hybrid ensemble model of machine learning was used to predict student performance using NB, KNN, SVM, decision tree, and logical regression during the COVID-19 period, yielding outclass results. Finally, the accuracy obtained through the hybrid ensemble model was obtained as 98.6%, which demonstrated that the hybrid ensemble learning model has performed better than any other model for predicting the performance of students.

**Keywords:** hybrid model; ensemble learning; online learning; machine learning; attribute selection; educational data mining; learning analytics; COVID-19; classification



check for updates

**Citation:** Asad, R.; Altaf, S.; Ahmad, S.; Mahmoud, H.; Huda, S.; Iqbal, S. Machine Learning-Based Hybrid Ensemble Model Achieving Precision Education for Online Education Amid the Lockdown Period of COVID-19 Pandemic in Pakistan. *Sustainability* **2023**, *15*, 5431. <https://doi.org/10.3390/su15065431>

Academic Editors: Ilana Paul-Binyamin and Linor L. Hadar

Received: 22 February 2023

Revised: 14 March 2023

Accepted: 16 March 2023

Published: 19 March 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

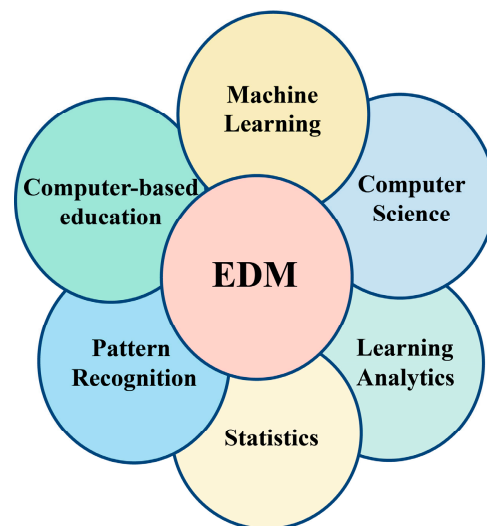
Education is the fundamental right of every citizen which leads to the development of a country [1]. In Pakistan, to provide better quality higher education to the future generation, the Higher Education Commission (HEC) was created by the government of Pakistan in 2002 [2]. Educational sectors of Pakistan have been looking forward to countering the novel challenges emerging in the way of achieving precision education [3]. Cook, Kilgus and Burns [4] pointed out that precision education is “a tactic to investigate and practice which is concerned with adapting preventive and interposition practices to individuals on the basis of best accessible evidence”.

In the achievement of precision education, platforms of digital learning play an essential role in the collection of student educational data along with various types of interactions:

their performance, learning pattern, and behavior [5], etc. As regards obtaining a higher education specifically at the university level, precision education has become a prime concern due to many reasons. Therefore, for achieving precision education, the improvement of the literacy rate has become essential.

As there is a huge amount of data on students available, there is a need to utilize these data for some valuable purpose. Data mining (DM) can justify this necessity by providing techniques to explore unseen facts and figures in students' information [6]. Two areas have been identified by Simens and Long [7] for utilizing such a bulk amount of educational data [8] gathered through digital platforms for learning as: learning analytics (LA) and educational data mining (EDM). Therefore, inspecting the subgroups of students, their attitude toward study, and their online learning pattern has drawn attention from EDM and LA-related research communities.

Educational data mining (EDM, hereafter) is delineated as the field of systematic investigation positioned over the progression of approaches aimed at the production of potential discoveries inside the unusual classes of data coming from educational settings, and later utilizing such procedures and methods to search through what means the students perform within different learning environments [9]. Some examples of specific fields where EDM is seeing widespread use are shown in Figure 1. Computer-based education, deep learning, computer science, learning analytics, statistics, and pattern recognition are the fields highlighted.

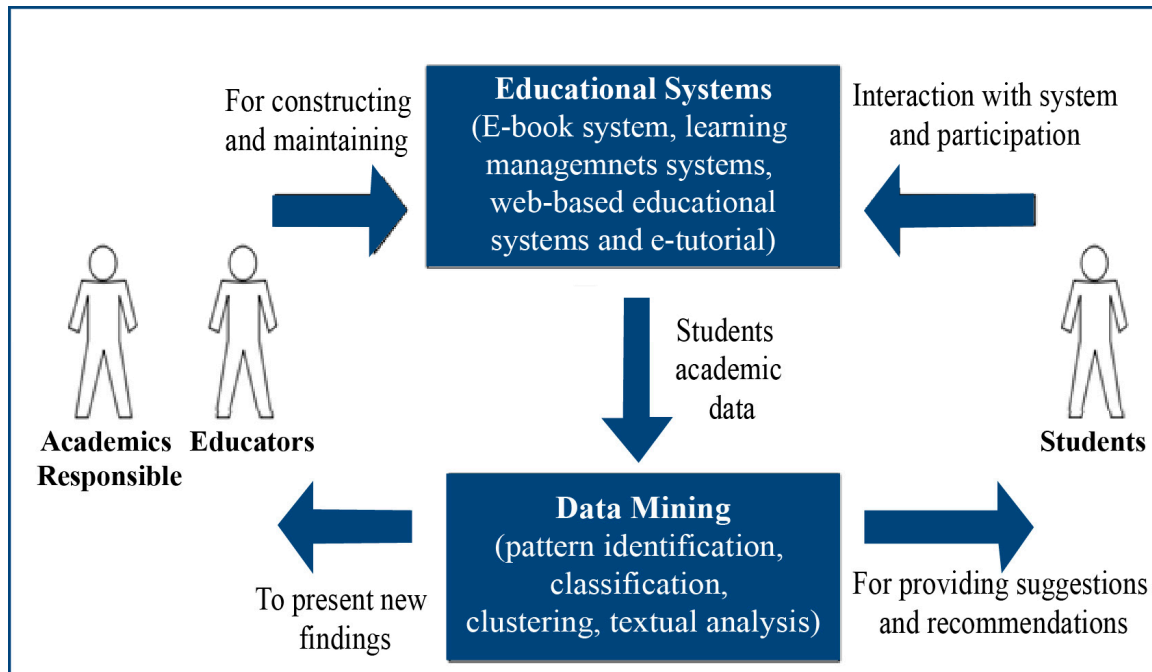


**Figure 1.** Examples of specific fields in educational data mining [10].

EDM has been utilizing DM methods for pattern mining for quite some time. Applying DM methods to the study of students' conduct yields effective results by allowing educators to foresee the likelihood of student attrition [10]. Figure 2 depicts the process through which DM is implemented in the academic setting. It explains how advisors are tasked with laying out the blueprints for the entire curriculum. Later, students interact with the plan, shared with them by their mentors. Subsequently, by applying certain DM algorithms to this educational data, unknown facts and figures are mined by giving useful recommendations about students. In traditional as well as online learning, DM techniques are being deployed for obtaining beneficial results.

Education institutions in the 21st century are inevitably going to move towards offering more courses online [11]. In the 1990s, with the advent of the Internet and the World Wide Web (WWW), online education began taking place [12]. Education delivery and learning models are changing [13] as a result of the ongoing development of information technology [14]. According to a report published by the U.S. Department of Education, which compared the results of the study conducted in a traditional classroom setting with those obtained in an online setting, the latter produced results that were either superior

to or comparable to an equal level with those obtained through the former [15]. Most of the digital learning platforms (DLP) available online that contribute to facilitating online education include Massive Open Online Course (MOOC), Google Meet, Google classroom, Small Private Online Course (SPOC), Learning Management System (LMS), and Zoom [16].



**Figure 2.** Flow of entities involved in the data mining process within the education system [10].

Online learning has enabled students to obtain quality education in any place, eliminating the time barrier and communication gap between the educator and the pupil [17]. The COVID-19 epidemic has reformed the whole world, though the influence and practice of online education milieus have significantly increased [18]. Some studies reported positive responses [19]; however, others stated negative attitudes [20] of students concerning online education [21] during the COVID-19 period. A report was published regarding the students of Pakistan, which stated that 77% of the students were having negative opinions and 84% were having reduced teacher–student communication regarding online learning during the pandemic [22]. Due to this gap, problems such as poor internet connection, and the lack of learners’ interest in studies have emerged side by side. Thus, this sudden transition of education toward online learning has posed many challenges for the learner and teacher as well [23]. In this study, we consider several research questions:

Q: Whether ML classifiers perform best individually or the hybrid model works well?

Q: Does a larger dataset help in avoiding model overfitting or not?

Some limitations of the literature that are present in the majority of current studies are outlined below:

- Recent research has focused extensively on the importance of customizing such models in relation to individual courses.
- Making models for each individual course is inefficient due to the overhead of maintaining multiple copies of each model. Therefore, a generic model is necessary.
- The scalability problem has also been identified as the smaller number of attributes considered in previous studies.
- Existing research has never used hybrid models to obtain precision education, which is essential for predicting students’ academic outcomes with superior accuracy.
- Due to a lack of data samples necessary for precise prediction, the models used in prior studies tended to overfit the data they were given.

The significant contributions made by this study are briefly summarized below.

- The proposed work has developed a model that is generic and performs well in predicting learners' outcomes in online learning for the period of COVID-19 by considering various features that are not course-dependent.
- The proposed study has used a hybrid ensemble model of machine learning considering different weak learners of supervised machine learning (SVM, logistic regression, KNN, naïve Bayes, and decision tree) for training to build a robust and efficient model.
- The large dataset was collected through a survey filled out by university students of Pakistan, primarily students of bachelor's, master's, and PhD study levels, in order to develop a portable model considering sufficient data samples.
- Three Meta-heuristic algorithms, PSO, HHO, and HGSO, for feature selection and one classifier VAE for feature extraction have been used to obtain the potential attributes that place a strong influence in making valid predictions.
- Enhanced accuracy has been accomplished using the hybrid ensemble model of machine learning, predicting the performance of students involved in advanced studies and achieving precision education as well.

## 2. Literature Review

The COVID-19 pandemic has affected the education sector worldwide. A recent study [24] proposed work for the analysis of certain factors that potentially contribute to the prediction of students' satisfaction with an electronic means of obtaining education during the COVID-19 phase. In addition, it also contributed to finding how the utilization of various DM techniques assists in finding the utmost appropriate attributes that have a certain influence on student performance. This study provided an e-learning model of classification for the in-depth examination of students. The dataset used for this study was from three schools' students in Iraq. The dataset was collected through a survey. A total of 1120 responses were collected, 1000 of which were utilized in this study after pre-processing. This questionnaire consisted of three parts: demographic information, feasibility and effectiveness of e-learning platforms, and student satisfaction with e-learning tools. In total, 35 potential attributes of the dataset were taken as a base to predict the performance of school students through the period of COVID-19. For analysis, the WEKA tool was used for this research. After the pre-processing of the data, classification algorithms were applied to train the model for student performance prediction to obtain the intended output. Later, in the second phase, the trained model was applied to make predictions on the dataset of students. Classification regressors of DT utilized for this study included random tree, decision tree, random forest, naïve Bayes, bagging, REP tree and KNN. The model successfully predicted the performance of the students. The highest accuracy achieved by the model through KNN was observed to be 96.8%.

To determine the influence of COVID-19 on the psychological well-being of learners during the lockdown period, the study has highlighted the importance and use of online tools and digital technologies during the COVID-19 period [25]. It has scrutinized the influence of physical distancing, quarantine, and seclusion on college students' psychological and mental health. The author has performed a SWOT analysis to highlight the challenges encountered by students in online teaching throughout COVID-19. This research work has used the online questionnaire to acquire data from students of Arab countries considering various attributes, i.e., their study patterns, sleep habits, psychological state, demography, etc. The total number of responses used was 1766 in number. After applying pre-processing steps to the collected data, the model training was completed. This study has utilized various classifiers of machine learning to build a model for making a prediction about students. Algorithms used for this study have predicted the real influence of online knowledge acquiring tools before and after the period of COVID-19. A 70 to 30 ratio was applied for training and testing, respectively. Chi-squared and ANOVA tests were used for validating the efficiency of the model. This study has concluded that there exists a positive relationship between online learning and student performance during it.

Studying how the student satisfaction level has been affected by online teaching through the COVID-19 period research has been conducted [26]. This paper contributed to predicting the academic performance of students to find out how the effectiveness of online learning (OL) systems can be enhanced. For the determination and extraction of the information related to student satisfaction levels and online learning during COVID-19, the study has proposed a real-time dataset. The dataset was gathered through an online questionnaire filled out by the students of seven educational institutions in Egypt for the academic study year of 2021–2022. The dataset holds the reviews of students regarding OL. The total of the responses used for building the model comprised 18,691 responses containing 20 features. The dataset was then pre-processed to eliminate erroneous data. For selecting the best attributes, 11 diverse meta-heuristic algorithms were applied to fetch the best feature out of the dataset. Later, on the dataset taken from Kafrelsheikh University and Mansoura University in Mansoura, Egypt was trained using two classifiers of machine learning: Support Vector Machine (SVM) and k-NN. For conducting the whole experiment, Python was cast off. Certain performance metrics were applied for evaluating model performance. The resultant precision observed was 100%, proving that the model is sufficiently robust.

Identification of student learning behavior in in-class learning courses during COVID-19 was performed in [27]. This study focused on tracking how various behaviors of learning affect the performance of students. This research work was directed towards a small population of students. The dataset used for this study was assembled through a survey of undergraduate students of mechanical engineering. Student response was collected via mobile app. A total of 133 responses were considered to hold the data for four different sections. The dataset was split between a ratio of 30% and 70% for testing and training, respectively. The dataset comprised student information regarding their class attendance, class participation, etc. One of the most important factors that dropped out was homework, which was not considered in this study. Later, these data collected from students were pre-processed, during which the grades of the students were converted to letters. Then, the SMOTE technique was used to balance the sampled dataset. The model was trained on various machine-learning classifiers which included support vector machine, decision tree, logistic regression, ensemble learning, random forest, and k-nearest neighbors. A small dataset was considered for the training model using 10-fold cross-validation technique for the detection of overfitting. Moreover, the grid search technique was used to optimize the performance of each used machine learning classifier. Ensemble learning showed an outclass performance with 84% accuracy as compared to other classifiers.

A study was proposed [28] for the improvement of the online learning effect on students' learning performance by providing them with timely personalized feedback to keep them safe from the risk of dropping out. The study has contributed to the prediction of students' learning performance in online education. For that purpose, the study has proposed a deep learning model known as PT-GRU. For conducting this study, two online datasets were utilized: ZJOOC and WorldUC. These datasets comprised students' data regarding online courses. The number of participants considered was 62 in total, who were enrolled in a Chinese university. Each course comprised 10 lessons. In total, 259 records were taken from ZJOOC and 7543 records from the WorldUC, splitting both datasets into 20% and 80% ratios for testing and training of the model for providing personalized feedback to students. To conduct this study on the two datasets considered, four classifiers were used. Of these, two classifiers of machine learning, decision tree and random forest, and two classifiers of deep learning model were used. Then, a quasi-experiment was conducted using the PT-GRU model. The highest accuracy achieved by the GRU in the ZJOOC dataset was 71.15% and on the other dataset, the highest accuracy was 81.44%, achieved through LSTM. The results proved that it has successfully provided personalized feedback to the university students.

Crucial factors [29] have been identified which influence the performance of university students, in addition to the effect of them using social media during the pandemic period

of COVID-19. In this study, the theory of constructivism was utilized and established with constructs that were linked with the increased use of social media for collaborative learning and the interaction of students during the pandemic for online learning. For this research, the dataset was collected through an online questionnaire from higher education students in Saudi Arabia. The questionnaire consisted mainly of 27 questions and each of the variables was graded between 1–5 on a five-point Likert scale. In total, 491 responses were received from students. After pre-processing, these responses were reduced to 480 due to the removal of erroneous data from them. Out of 27 questions, 4 questions were used to analyze online learning, 6 were used for analyzing the interaction of students with their mentors and peers, 4 were used for predicting the performance of students, and the remaining 4 were used to assess student satisfaction during the pandemic. For conducting this research, structural equation modelling (SEM) was used to analyze the dataset to discover the relationship between the dependent and independent variables considered for this research study. Later, for the model validation, three types of goodness-of-fit metrics were applied. The results of the study revealed a positive relationship between the following variables: student learning, the satisfaction of students, and the interaction of learners with mentors and peers.

An automated system was built through a recent proposed study [30] that could carry out the prediction of students' grades in online education based on the availability of the performance data of learners, all throughout the COVID-19 pandemic. To perform this study, the dataset of students was considered for the period from 2006–2017 to predict the grades of students. The total number of records used was 1000 for undergraduate and graduate students for 15 different courses. The IITR-APE dataset was used considering various parameters as a base for accurate performance prediction. Firstly, this dataset was pre-processed to remove outliers. Further, the variational auto-encoder technique was applied to obtain the most potential features out of the dataset. Later, the extracted features made some predictions about grades. The classifiers used in this research included random forest, linear regression, XGBoost, extra tree, multi-layer perceptron and KNN. Later, the model was tested using mean absolute error (MAE), R2-score, root-mean-squared error (RMSE), and mean squared error (MSE). The results revealed that deep learning models are best for making accurate predictions. Out of all the applied classifiers, the outcomes proved that the extra tree classifier achieved outclass results of 0.720 R2, 5.943 for MAE, 77.709 for MSE, and 8.781 for RMSE.

Different copying patterns were detected [31] that were faced by undergraduate and graduate students while obtaining online education through virtual classrooms that have caused various types of anxiety and stress in their student life. To perform this research study, the dataset of students was collected through a questionnaire via Qualtrics from the postsecondary institutions of the US. A total of 517 responses were used. Of these responses, 423 were filled by females, 91 were filled by males, and 3 were those who reported their gender as non-binary. This dataset was collected between May and July 2020 when COVID-19 was at its peak. A total of 25 questions were asked through a survey to the students using the Likert scale. Then, for the extraction of the best features out of the dataset, the principal axis factoring technique was used. This study has used the technique of association rule mining for the first time to transform the data into the framework of market basket analysis to mine useful patterns of students. For the implementation of the model used in this study, the advanced version of the data mining Apriori algorithm was used, which is known as the "FP-growth classifier". Then, support for each item in the dataset was found. The dataset used in this study was scanned twice. After the construction of the FP-tree divide and conquer technique, the FP-growth classifier was used to mine the items. The resultant outcome produced 78 and 14 strong "association rules" for the groups of graduate and undergraduate, respectively, collected through the dataset. Thus, the study proved that undergraduate students were more consistent during online learning throughout the pandemic period.

Use of digital platforms was traced through a study [32] to identify the regulatory factors for online education throughout the COVID-19 pandemic. Furthermore, a total of four datasets were utilized taking the 589 students' data from "X-University" and software of Microsoft Teams. Dataset-I comprised mainly seven courses and six attributes were considered from the 589 instances. Dataset-II consisted of five courses, eight attributes, and a total of 259 records of students. Dataset-III consisted of 4 subjects and 12 attributes with a record number of students, 280. Similarly, for dataset IV, only 2 subjects, 10 potential attributes, and 91 records of students were taken. To perform the proposed work, the decision tree (J48) classifier via a 10-fold cross-validation approach was used considering two dominating factors which included "Mid-term" and "Final-term", from which Final-term was taken as the root node. Later, the classifier was used to make and define the set of rules for each of the considered datasets. During this process of mining hidden patterns from student data, it became evident that only three-to-four potential attributes are enough to make a valid classification of information. The results proved that the potential attributes were "Mid-term" and "Final-term" and the remaining attributes considered did not have much impact on learners in pandemic period.

Work related to the identification of learning patterns and the behavior of students via the Ebook system to mine useful patterns from the students' data making useful predictions and achieving precision education as the major goal was conducted in [33]. The dataset used for proceeding with the research was collected from undergraduate university students. The dataset comprised only one single course of "Accounting Information Systems", with 113 entities. BookRoll was used for facilitating the Ebook system used by students. To identify the behavior of students' learning, various indicators were considered from the data. Further, the collected data and the extracted indicators were normalized to a numerical value between 0 and 1. Later, for the identification of learning patterns, the agglomerative hierarchical clustering technique was applied. The diversification in the divided subgroups based on four indicators was then verified through the Mann–Whitney U and Kruskal–Wallis tests. The study revealed that the comprehensive learning approach was found successful in the prediction of students' behavior. Below, Table 1 precisely describes the work of some recent studies, their contribution, techniques used, their results, and limitations.

**Table 1.** Comparative analysis of related works.

Paper	Contribution	Technique	Results	Limitations
[24]	Predicted performance of students in E-learning	Decision Tree, Random Tree, Naive Bayes, Random Forest, REP Tree, Bagging and KNN	96.8% accuracy for KNN	Smaller dataset considered.
[25]	Predicted impact of online learning on students	Logistic Regression, Decision Tree, SVC, XGB and AdaBoost	Efficient models except AdaBoost	More computational time and overfitting.
[26]	Enhanced the effectiveness of OL	SVM and k-NN	Both models outperformed	k-NN is a slow learner and also took more running.
[27]	Effect of student learning behavior on their performance	SVM, RF, DT, Logistic Regression, KNN and Ensemble Learning	84% accuracy for Ensemble Learning	Excessively small dataset considered.
[28]	Student personalized feedback model	LSTM, GRU, DT and RF	81.44% accuracy by LSTM	Smaller dataset, model lack generalizability.
[29]	Identification of factors and use of social media on student performance	Structural Equation Modelling	Direct Positive relationships proved	Focused on quantitative data, model overfitting.
[30]	Automated student performance system	RF, LR, Extra Tree, XGBoost, MLP, KNN	Extra Tree regressor outperformed	Model overfitting.

Table 1. Cont.

Paper	Contribution	Technique	Results	Limitations
[31]	Students coping patterns detection	FP-growth algorithm	Coping patterns identified accurately	Self-report questionnaires.
[32]	Exploiting regulatory factors for online education	Decision Tree (J48)	Successfully mined potential attributes	Lacks model generalizability.
[33]	Identification of students' learning patterns	Agglomerative hierarchical clustering	Successfully identified learning patterns	Smaller dataset, model is not generic, fewer patterns identified.

### 3. Proposed Framework for Achieving Personalized Education

This section describes the planned infrastructure for analyzing Pakistani higher education students taking courses online during the COVID-19 period using a hybrid ensemble learning model of machine learning. Due to the sudden appearance of a pandemic, people are now primarily relying on online resources for their education. There are a number of factors that have affected students' grades. A survey form was used to survey students' opinions on various aspects in order to better understand the factors involved.

The steps required to achieve precision education through digital means are depicted in Figure 3. In the pre-processing phase, any outliers or incomplete data in the acquired dataset were removed. The dataset was then further pre-processed using the min-max normalization method after the mandatory resampling was completed. We used a set of four meta-heuristic classifiers to help us find and prioritize useful features. With these features in hand, a hybrid ensemble learning model of machine learning was used to fine-tune the model for use in the final step of the process. Specific validation measures were then used to assess the quality of the models.

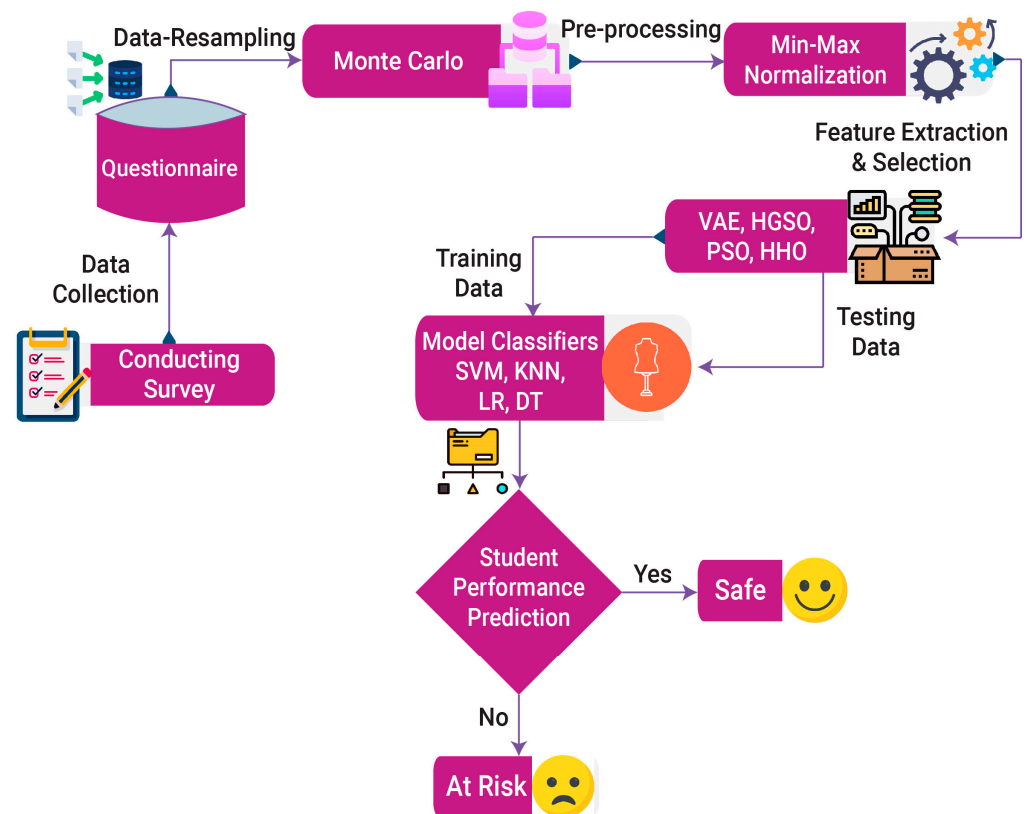


Figure 3. Proposed framework for achieving precision education in online learning.



### 3.1. Materials and Methods

Figure 3 depicts the steps that will be taken, and this section will describe each one in detail so that how they all fit together to create precision in education can be observed. Each step was completed in order (one after another).

#### 3.1.1. Data Gathering

The initial phase first starts with the acquisition of the dataset. For the collection of data from various students of higher education in Pakistan, an online Google-based questionnaire was designed and administered to the different institutes. The questionnaire inquired about all the possible attributes that were considered helpful in the analysis of learners. The questionnaire comprised 35 questions in total. The total number of responses filled in by higher education students was 11,000.

#### 3.1.2. Resampling Data

After data acquisition, the next step demands the resampling of data if required. To perform the resampling of data, the Monte Carlo technique was adopted to check its accuracy. For questionnaire data filled in late on the collected responses, Monte Carlo was applied to make the estimation for the possibility of an arbitrary variable's number. The following equation derived defines the Monte Carlo method:

$$F(G) \approx \frac{1}{M} \sum_{m=1}^M g_n \quad (1)$$

In Equation (1), the mathematical sign “ $\approx$ ” represents that on its right side, it holds the “estimation” of what the arbitrary variable  $G$  assumes the output of function  $F(G)$  to be.

#### 3.1.3. Data Pre-Processing Phase

The very next phase, which is preliminary and important, is the pre-processing of collected data to eradicate the unusual data, redundant data, and outliers. For the pre-processing, the technique of feature scaling, known as min–max normalization, was utilized. Through this technique, a linear alteration is applied to the acquired data [34]. This approach transforms the data within the range of 0 and 1 and maintains the relationship between the acquired values of data. The purpose of utilizing this confined range is that it will finish with very small standard deviations that overwhelm the consequence of outliers. Following is the formula for min–max normalization.

$$y' = \frac{y - y_{\min}}{y_{\max} - y_{\min}} \quad (2)$$

where  $y'$  in Equation (2) shows the resulted value,  $y_{\min}$ , and  $y_{\max}$  correspond to the minimum and maximum values, respectively, of the given dataset. Below, Figure 4 shows some of the most commonly applied methods opted for preprocessing the dataset.

#### 3.1.4. Feature Selection

The value of utilizing wrapper-based meta-heuristic feature selection classifiers is that these can classify the perilous attributes from within the massive data. The current research study has utilized three meta-heuristic classifiers which are described below.

##### Particle Swarm Optimization (PSO)

The most simple and vigorous optimization classifier demonstrated afterward the societal behavior [35] of animals, e.g., birds and fish. PSO was first coined in 1995 and was introduced by Kennedy and Eberhart [36]. It has been applied in many diverse fields of engineering and scientific applications including image processing, DM, ML, and robotics as well. The swarm mode utilized by PSO makes it capable of searching large sections within the solution region of an augmented objective function. Hence, it has been applied

in a variety of fields and industries to resolve problems regarding optimization. PSO is given by:

$$r = \frac{\sum(a_i - \bar{a})(e_i - \bar{e})}{\sum \sqrt{(a_i - \bar{a})^2 (e_i - \bar{e})^2}} \quad (3)$$

In Equation (3),  $u$  is the velocity,  $y$  is the size of the population and  $p_{\text{best}}$  is the position of the fittest individual and  $g_{\text{best}}$  is the fittest attribute out of the whole population.

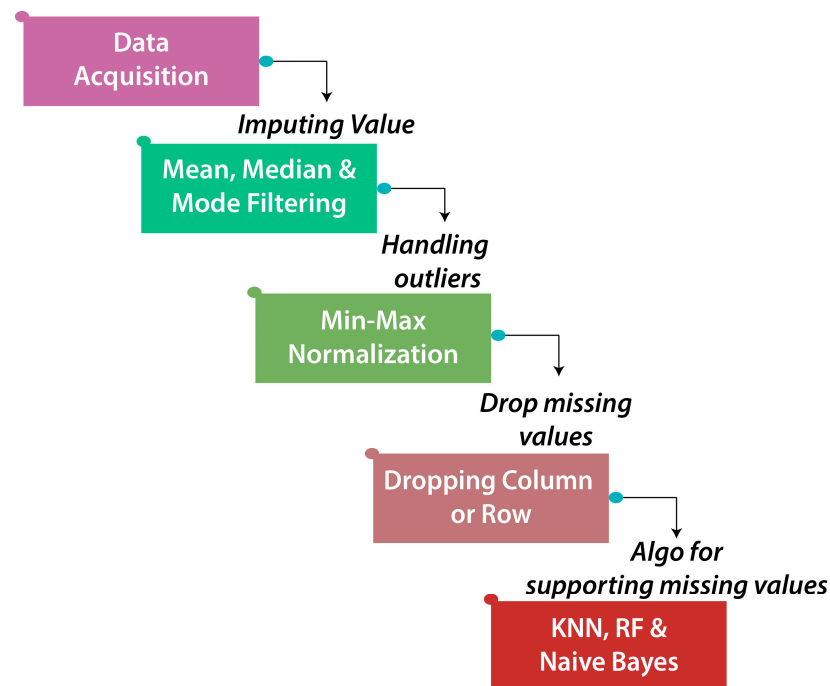


Figure 4. Methods of data pre-processing.

#### Harris Hawks Optimization (HHO)

HHO was developed by Heidari et al. [37]. It is the population-based type of meta-heuristic classifier used for optimization purposes. The main rationale for this classifier was encouraged by the supportive behaviors of one of the smartest birds, Harris' Hawks, in chasing on the run feed (most probably rabbits) [38]. Hawks follows actions representing agents in the hunt space, though feed denotes the finest position. Therefore, HHO is applied for optimization problems. Moreover, it is applied to handle the unfamiliar types of hunt space and resolve problems concerning continuous and discrete areas, giving higher quality solutions, and extracting optimum parameters out of it with high accuracy [39]. Two strategies are mainly used by HHO, which are represented as follows:

$$Y(u+1) = \begin{cases} Y_{\text{rand}}(u) - r_1 Y_{\text{rand}}(u) - 2r_2 Y(u) & q \geq 0.5 \\ (Y_{\text{rabbit}}(u) - Y_m(u)) - r_3(qb + r_4(\mu b - qb)) & q < 0.5 \end{cases} \quad (4)$$

In Equation (4), the present iteration and upcoming iterations are represented by  $t$  and  $Y(u+1)$ , respectively.  $Y_{\text{rabbit}}(u)$  denotes the location of the rabbit,  $Y(u)$  is the present position trajectory of hawks,  $r_1$ ,  $r_2$ ,  $r_3$ , and  $r_4$  represents the random numbers between 0 and 1,  $Y_{\text{rand}}(u)$  is the randomly chosen hawk in the present population,  $Y_m(u)$  is the regular locations of the hawks in the present repetition,  $qb$ , and  $\mu b$  are the lower and upper limits for the variables, respectively.

#### Henry Gas Solubility Optimization (HGSO)

The HGSO classifier was proposed by Hashim et al. [40] in 2019. Henry's law was the actual motivation behind the idea of the HGSO algorithm [41]. HGSO is considered the

global optimization problem solver because it includes the exploitation and examination phases. Additionally, it has been easier to deploy HGSO because in it, fewer operators need to be adjusted. It follows the behavior of gas particles; over a certain pressure, solubility occurs. Relationships between these variables are given by the following formula:

$$\text{Sol}_{\text{gas}} = H * \text{Pre}_{\text{gas}} \quad (5)$$

where Equation (5) denotes the Henry constant, pressure, and solubility of gas by H, P<sub>gas</sub>, and Sol<sub>gas</sub>, respectively, over a certain temperature, which is expressed as follows:

$$H(T) = H' * \exp\left(\frac{-\nabla_{\text{sol}}E}{C} \left(\frac{1}{T} - \frac{1}{T'}\right)\right) \quad (6)$$

where, in Equation (6),  $\nabla_{\text{sol}}E$  denotes the total heat suspension, T' is the temperature, C signifies the gas constant, and Henry's constant is denoted by H'.

### 3.1.5. Feature Extraction

After feature selection comes the feature extraction and the classifier used for this study for attribute extraction is as follows.

#### Variational Auto Encoder (VAE)

The frequently applied technique for the task of feature extraction to renovate closest related input [42] is variational auto encoder. VAE learns the Z variable of latent space in the form of a former dispersal, usually the Gaussian distribution. With the help of presumed distribution, latent space is dispersed in the form of "logarithmic variance"  $\log \sigma$ . KL divergence is applied for enforcing its distribution which is defined as follows:

$$A_{KL}(M(\mu_x, \sigma_x), M(\mu_y, \sigma_y)) = \log \frac{\sigma_y}{\sigma_x} + \frac{\sigma_x^2 + (\mu_x - \mu_y)^2}{2\sigma_y^2} - \frac{1}{2} \quad (7)$$

In Equation (7), two mapping are shown for distribution.  $M(\mu_y, \sigma_y)$  has enforced the distribution and  $M(\mu_x, \sigma_x)$  is the one on which distribution is made.

### 3.1.6. Machine Learning Classifiers

The subsequent sections defined the classifiers of ML utilized for the proposed study.

#### Decision Tree (DT)

The best classifier under supervised machine learning used for classification is decision tree (DT). It has the capability to make valid predictions based on a pre-defined set of rules. It resembles the structure of a tree, starting from the top root node and going down to the leaf node while expanding. Decisions are performed at the top node of the tree and based on those decisions, further actions are carried out. Data attributes, rules for decisions, and outcomes are presented through internal nodes, branches, and leaf nodes, respectively. The tree ends at the leaf node and branching stops there [43]. It is given by the formula:

$$\text{Ent}(U, Y) = \sum_{b=Y} P(b) \text{Ent}(b) \quad (8)$$

where, in Equation (8), Ent denotes the entropy calculated for deciding about root nodes, while U is the present state and Y is the attribute that is selected.

#### Naïve Bayes

The most common technique which makes the supposition of independence between the predictors, based on Bayes Theorem for doing classification, is known as naïve Bayes. It calculates the probability among the classes to make a statistical analysis of given data. It

is easy to evaluate and can quickly train the model [44]. Mostly used for clustering and classification problems, it is presented through the following formula:

$$\text{Prob}\left(\frac{b}{y}\right) = \frac{\text{Prob}\left(\frac{y}{b}\right)\text{Prob}(b)}{P(y)} \quad (9)$$

where, in Equation (9), Prob is the posterior probability,  $y$  denotes the likelihood of occurrence, and  $a$  is the class prior probability.

#### Support Vector Machine (SVM)

It most probably resolves the regression and classification problems [45] as it is a machine learning regressor. Its main application is to deal with issues related to classification. Mostly, it is applied by the DM communal because it can produce highly accurate results with a smaller number of computational resources. Its major goal is to search for the finest hyperplane that can classify the data into two classes. However, this approach opposes two substantial key challenges: parameter modification and suitable primary function selection [46]. It is given by:

$$u \cdot y + a = 0 \quad (10)$$

where  $u$  in Equation (10) denotes the normal trajectory to the hyperplane and  $a$  is a counterpoise (offset).

#### K-Nearest Neighbors (KNN)

Another commonly utilized classifier for resolving regression and classification associated problems is known as KNN [47]. The reason for its common use is that it is easily implemented compared to other ML regressors [48,49]. For pattern recognition, KNN has been applied in many fields, including finance [50], healthcare [51], forestry [52], image recognition [53], etc. KNN classifier is thought of as a packaging approach as it produces rules of classification from training samples. It falls under the category of supervised learning. It is easily understood and implemented but it has major pitfalls; in addition, its performance slows down with the increase in the size of data. It calculates the distance for classification into the categories through Euclidean distance, which is given by Equation (11) as follows:

$$d = \sqrt{\sum_{x=1}^k (a_x - b_x)^2} \quad (11)$$

#### Logistic Regression Model

Another commonly applied classifier of supervised learning is logistic regression. Usually, it is used for a large dataset for training the model with higher accuracy. It is used to solve classification problems that have binary solutions, i.e., either 0 or 1. LR is given by the formula presented below:

$$l\Theta(y) = \frac{1}{1 + u - (\alpha_0 + \alpha_1 Y)} \quad (12)$$

where, in Equation (12),  $l\Theta$  is the resulting output of the logistic regression function, the  $y$ -intercept is denoted by  $\alpha_0$ , the slope is shown by  $\alpha_1$  and  $Y$  is used for the independent variable.

#### Hybrid Ensemble Learning Model

The process through which various models are deliberately produced and joined to resolve any certain computational intelligence problem, e.g., experts or classifiers, is known as ensemble learning. It is mainly castoff to enhance the model performance or to diminish the probability of the occurrence of an unfortunate selection of a poor one. Hybrid

ensemble learning made valid decisions by combining the multiple weak regressors and giving accurate predictions with higher accuracy.

### 3.1.7. Performance Validation of Model

For the proposed research, the confusion matrix has been applied for the precise analysis of model. It consists of a  $2 \times 2$  matrix with one side showing the actual values of the dataset and the other showing the value predicted by the trained model. Some performance metrics explained below are accuracy, precision, recall, and F1 score. Accuracy is given by the formula below:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (13)$$

In Equation (13), accuracy is defined as the measure for total quantity of accurate predictions made out of the total input values in the data. Precision works for finding how accurately the model has performed, which is given by Equation (14).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (14)$$

Similarly, recall measures the validity of the model by measuring the accurate prediction of a certain class made out of total input values passed to that model. The recall is given by the formula shown in Equation (15).

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (15)$$

However, F-measure is cast-off to avoid any ambiguous assessment that may happen because of data disbalance. F-measure is delineated as the harmonic mean for precision and recall given by Equation (16).

$$\text{F-measure} = \frac{2 * (\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (16)$$

In the above-given Equations (13)–(16), TN signifies true negative, TP means true positive, FN represents false negative, and FP designates false positive.

## 4. Experiment and Results

### *Description of Dataset*

Google Forms were used to design a questionnaire to collect information from learners in order to assess their effectiveness during the COVID-19 period. The questionnaire was adaptable in its structure and included all the necessary questions to learn how the lockdown affected students' academic performance. After that, a questionnaire was distributed to a number of universities in Pakistan in order to collect information from students regarding the posed questions.

The questionnaire was composed of 35 questions in total. Of these, 25 questions were picked and the remaining were dropped. The selected questions were dealt with as potential attributes for making predictions. The questions that were dropped hold general information such as age, name, gender, and locality, which does not need to be added for conducting research—that is why they were omitted. The total number of responses collected was 12,000. After pre-processing, the total number of responses considered was reduced to 10,000. These responses were from those students who were obtaining an education at the time of COVID-19 in degrees that include computer science, information technology, software engineering, and management science. The ratio of students in the dataset was 65% bachelor students, 25% master's students, and 10% doctoral. Table 2 below shows the dataset questions along with the responses collected for each answer. Additionally, Table 3 shows the output label key used for the dataset.

**Table 2.** Dataset description.

Features	Students' Response				
	5	4	3	2	1
University, college, and degree name.	Categorical/nominal features				
Mentors were committed to course content.	1540	3990	2055	1750	665
Lectures uploaded on time by teachers.	2115	6700	645	280	260
Teacher encouraged students to ask questions.	2320	4550	1260	995	875
Freedom to prompt your point of view.	1690	6895	875	225	315
Mentor dealt with the topic in depth.	3620	5670	45	455	210
Lecturer well prepared.	2345	3710	1180	1400	1365
Lectures were informative.	2380	4410	1250	770	1190
Lecture presentation was in attractive style.	1540	3990	2055	1750	665
Enough information delivered.	2100	6020	177	918	785
You actively participated	2660	5145	620	560	1015
Practical cases included in lectures.	1430	7234	425	672	239
Assignment given weekly.	1872	7350	148	350	280
Enjoyed experience of online education.	945	7460	440	1085	70
Experimental quiz prepared.	872	8229	35	523	341
Problems solved during exams by mentors.	1388	6456	476	1260	420
Exam from within the course.	3689	4970	522	476	343
Exams taken on appropriate time.	2314	5689	444	1232	321
Marking of exams appropriate.	3359	5390	866	140	245
Decision Label	SA	A	N	D	SD

**Table 3.** Output label key for dataset.

Value	Label
1	Strongly Disagree (SD)
2	Disagree (D)
3	Normal (N)
4	Agree (A)
5	Strongly Agree (SA)

Figure 5 above shown presents a pictorial representation of the responses collected from students through a survey for each data attribute. Each color in the graph represents a different response. The dark blue color refers to strongly agree, the orange color denotes Agree, the grey color represents the neutral response, the yellow color shows disagreement, and light blue belongs to strong disagreement. Figure 6 below represents the data visualization showing the age of the students who filled in the dataset.

Figure 7 below depicts the total ratio of male and female students for each level of study. The blue in the picture corresponds to the male and female students of the bachelor level. Brown and grey represent the masters and doctoral levels, respectively.

For the resampling of data, the Monte Carlo technique was applied. After data acquisition, the next step performed to remove outliers from the data was preprocessing. The proposed study has utilized the technique of min–max normalization to remove erroneous data. Later, three meta-heuristic algorithms were applied to the refined dataset, which include PSO, HGSO, and HHO. Through these aforementioned classifiers, the fittest attributes were found. Below, Figures 8–10 show the graphical representation of bachelor, master's, and doctoral students, respectively, concerning their response regarding exam marking, free time availability, problems solved, quiz preparation, and weekly reports sent by mentors.

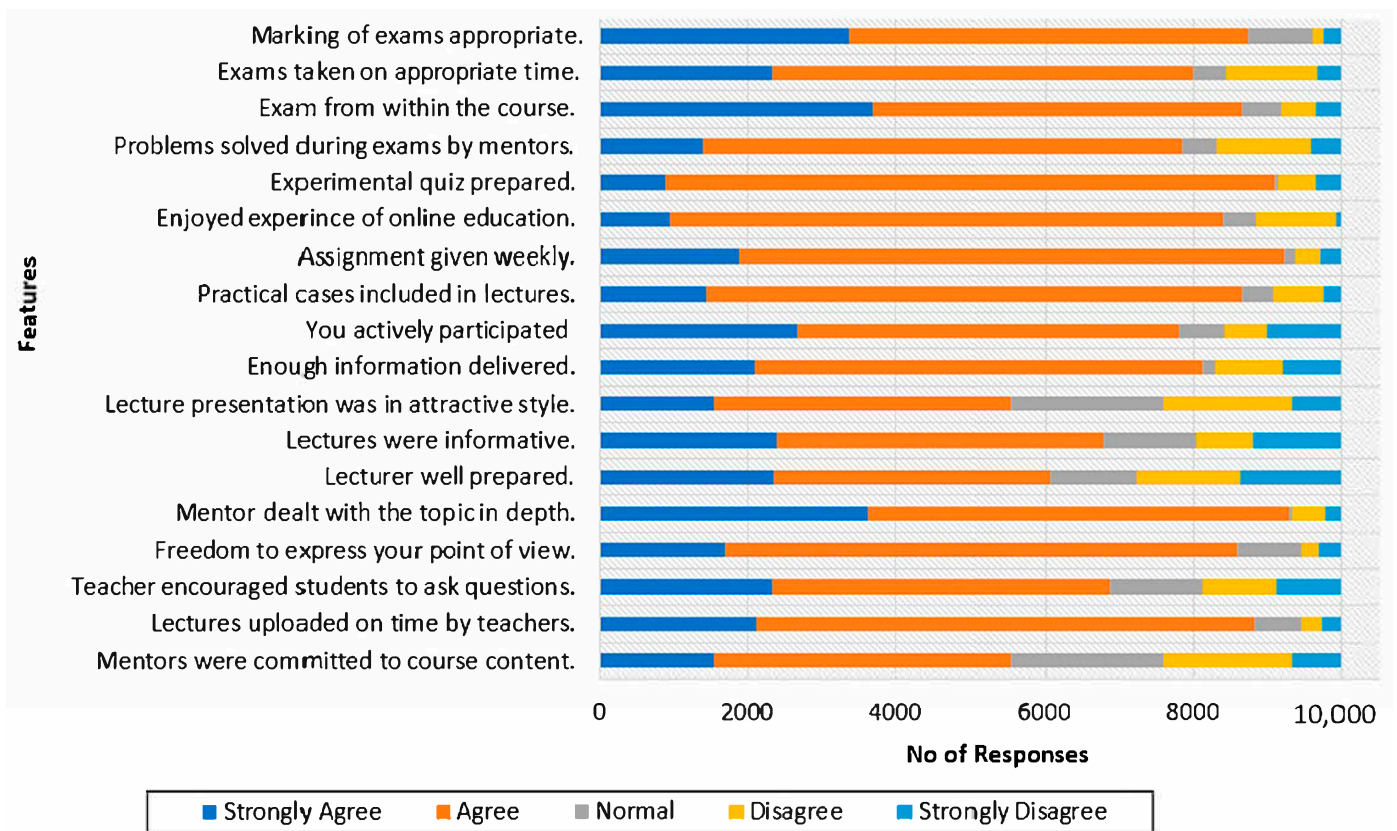


Figure 5. Frequency of each data attribute.

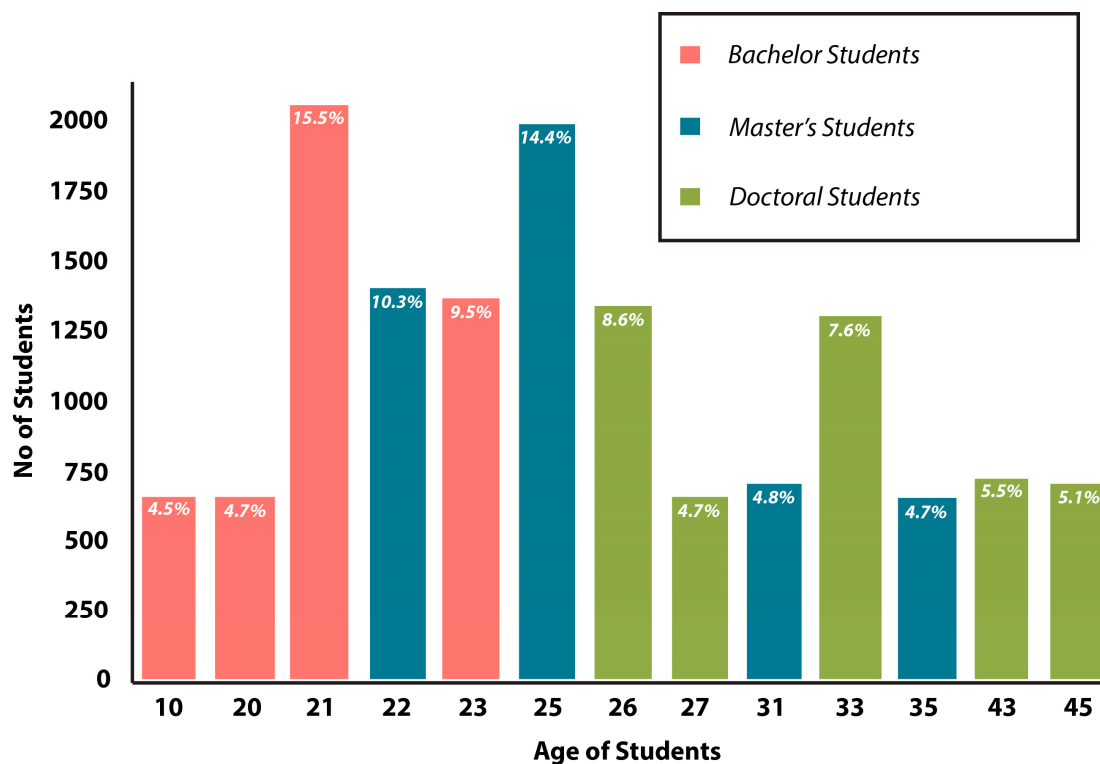


Figure 6. Data visualization of students age-wise.

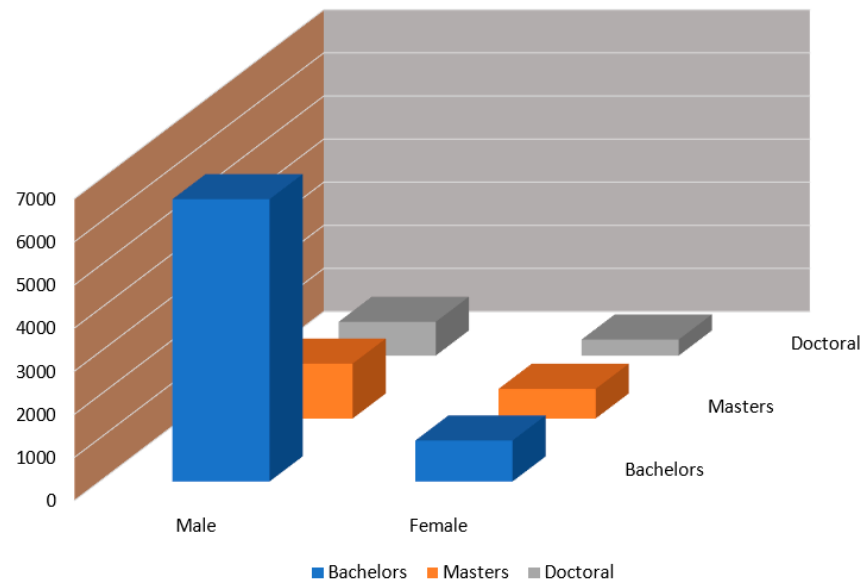


Figure 7. Gender distribution according to the levels of study.

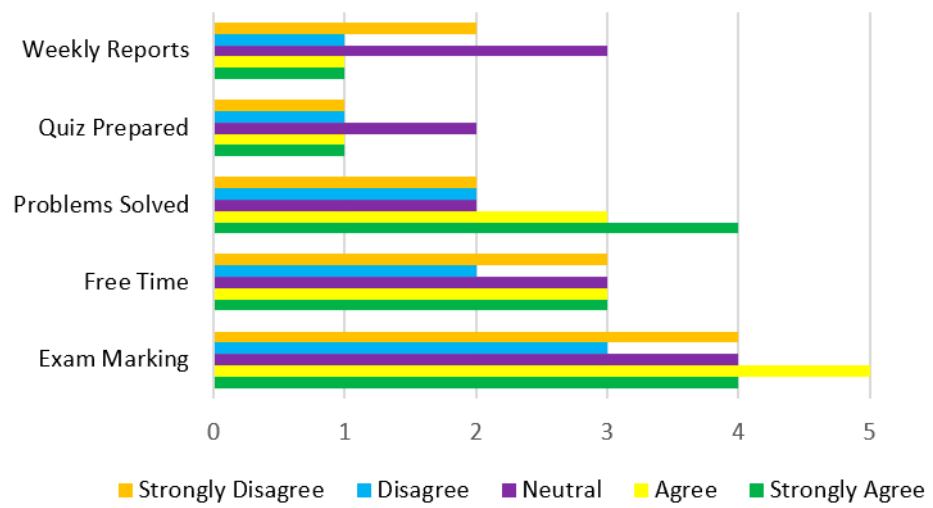


Figure 8. Response of bachelor students.

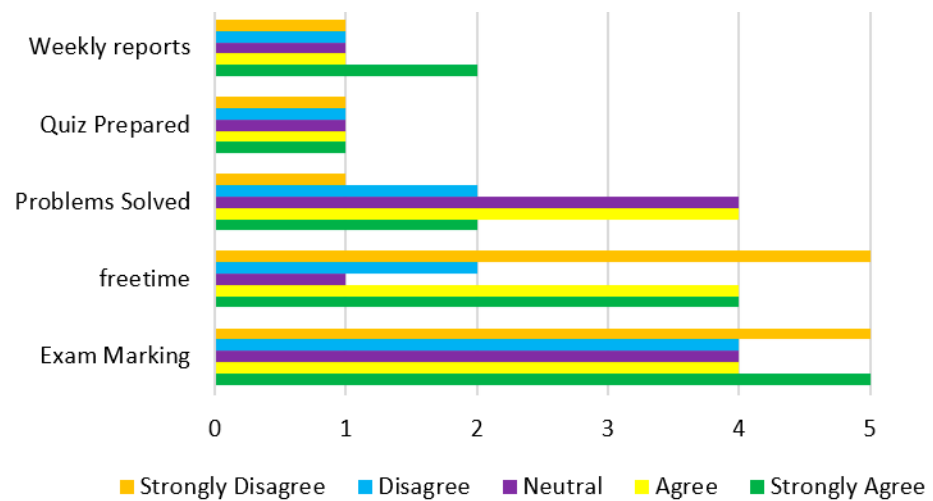


Figure 9. Response of master's students.



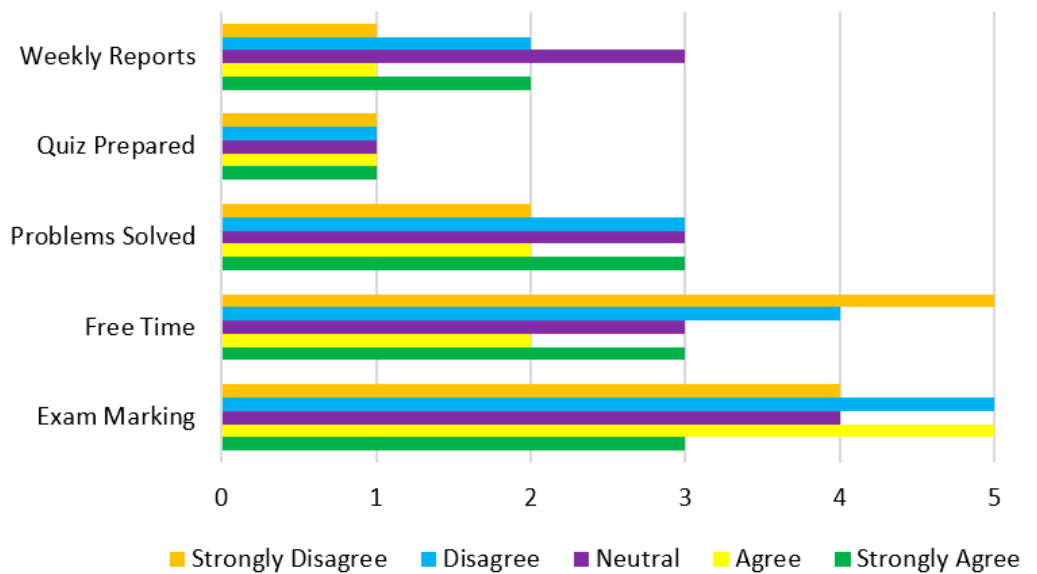


Figure 10. Response of doctoral students.

Hybrid ensemble learning has utilized five weak learners of machine learning in the proposed study. These weak learners include decision tree, naïve Bayes, support vector machine, k-nearest neighbor, and logistic regression. The dataset is passed to each of the weak learners and the individual performance is evaluated then to check the model’s efficiency in correctly predicting the performance of students. Thus, through individual training, the results given by DT, NB, SVM, KNN, and LR were observed as 85.1%, 84.3%, 87.5%, 84.9%, and 83.1%, respectively. Figure 11 below depicts the graphical representation of ML classifiers along with the resulting accuracy obtained by each model.

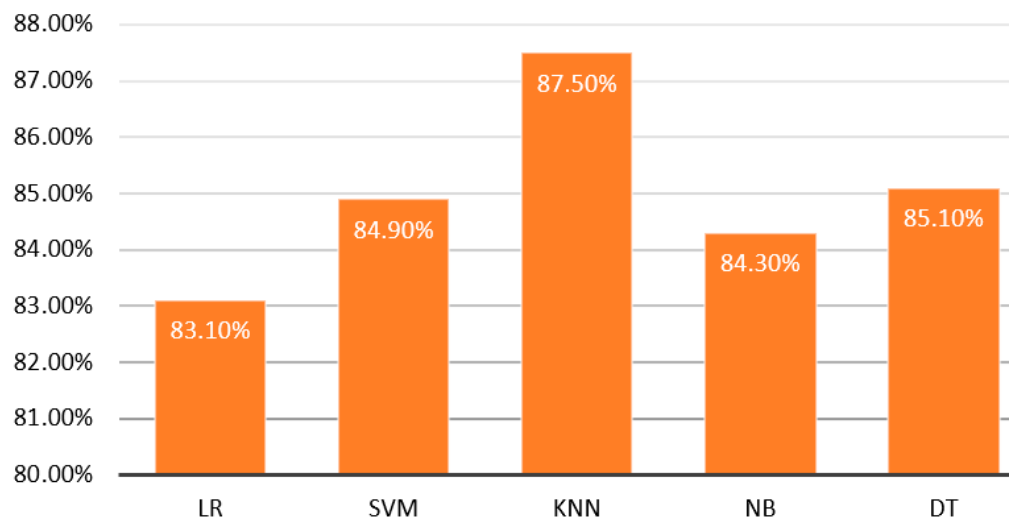
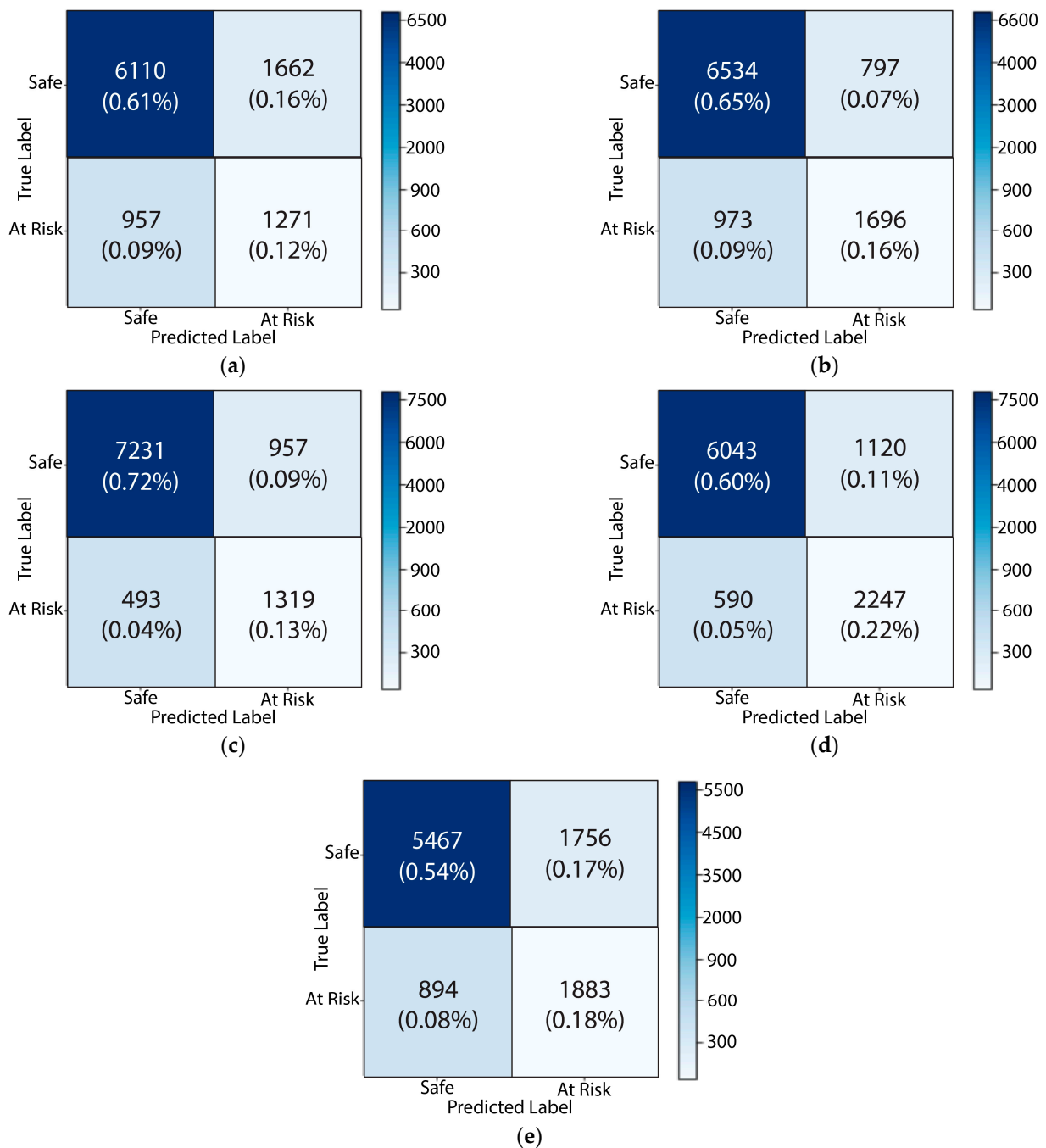


Figure 11. Accuracy representation of ML classifiers.

Table 4 below presents the elaborated statistics for the performance of ML classifiers for achieving precision education. Individually, the performance of SVM has dominated other regressors of machine learning. Figure 12a–e presents the confusion matrix for each of the classifiers.

**Table 4.** Performance evaluation of ML regressors.

ML Regressor	Precision	Accuracy	Recall	F1-Score
DT	85.8%	85.1%	83.4%	84.3%
NB	85%	84.3%	82.7%	83.5%
SVM	88.4%	87.5%	85.9%	86.7%
KNN	85.9%	84.9%	83.5%	84.2%
LR	83.8%	83.1%	81.9%	82.4%

**Figure 12.** Confusion matrix for (a) DT; (b) NB; (c) SVM; (d) KNN; (e) LR.

To further improve model efficiency, this paper proposes a hybrid ensemble learning model by combining widely used regressors from the field of machine learning. This model can be used to educate students with the utmost precision. For that purpose, a hybrid

ensemble learning model was trained using five models of each weak regressor, which include the decision tree, naïve Bayes, support vector machine, K-nearest neighbor, and logistic regression, as shown in Figure 13. Later, after training the hybrid ensemble learning model, it is validated for ensuring model efficiency.

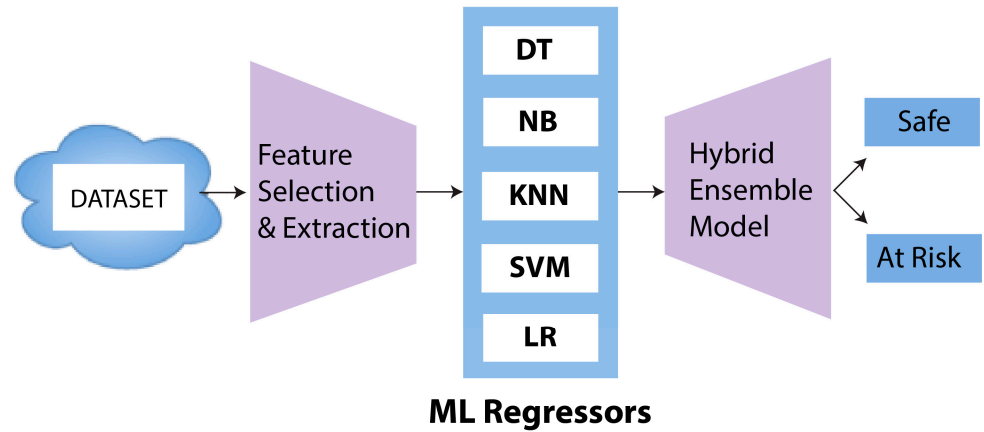


Figure 13. Hybrid ensemble learning model.

Results that have been observed through validation measures predicted that the hybrid ensemble learning model has outperformed the performance of each ML regressor. The dataset used for this model was spliced into 70% and 30% for training and testing purposes. The accuracy given by the hybrid ensemble model was far more improved than the accuracy given by machine learning classifiers. The accuracy achieved was 98.6% for the successful prediction of students’ performance in achieving precision education with an error rate of 1.4%. Figure 14 below portrays the total ratio of predicted at risk and safe students obtained by the model as 12.91% and 87.09%, respectively.

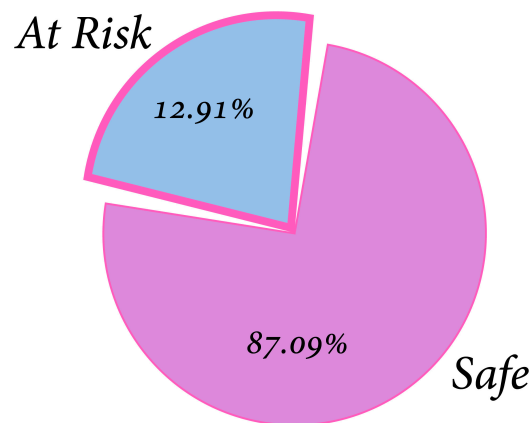


Figure 14. Data cataloging into at-risk and safe students.

Validation measures adopted to measure the correctness of the model’s results yielded 99.2%, 97.4%, and 97.9%, precision, recall and F-measure for safe label, respectively. For the at-risk classification, the achieved precision, recall, and F-measure were 97.9%, 97.2%, and 96.8%, respectively, as shown below in Table 5:

Table 5. Detailed accuracy obtained using the model.

Labels	Precision	Recall	F-Measure	ROC Area
Safe	99.2%	98.6%	97.4%	97.9%
At Risk	97.9%	96.7%	97.2%	96.8%

Figure 15 plots the  $2 \times 2$  confusion matrix that presents the predicted outcome through a hybrid ensemble learning model that has an influence on the performance of students. In this matrix, the results are plotted amongst the actual value and predicted value of the model.

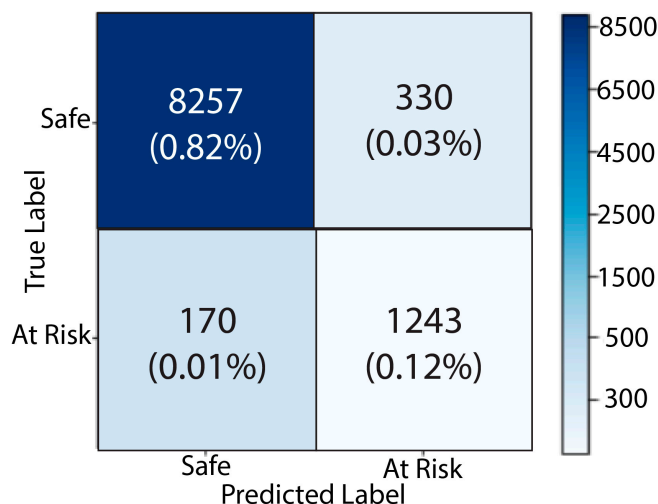


Figure 15. Confusion matrix for hybrid ensemble model.

As the model was observed to generate outclass performance, it proved that the hybrid ensemble learning model is an efficient approach to be applied to the prediction of student performance for achieving precision education. Table 6 below shows the comparison of some previously conducted potential studies with the proposed study. The comparison has been made in terms of the regressor used, the feature selection classifier applied, the number of potential attributes extracted for training of model, and the observed accuracy percentage value.

Table 6. Comparing hybrid ensemble model with other studies.

Paper	Technique	FS Classifier	Selected Features	Accuracy
[54]	ANN, DT, RF, Booting, Bagging and Voting	GA	6	81.18%
[55]	DT, KNN, NB, SVM and ANN	GA	10	91.12%
[56]	KNN, NB, DISC and DT	SFS, DE and SBS	6	83.09%
[57]	NB, NBU, BN, MLP, SMO, SL, DT, DS, J48, RT, RepT and RF	Relief, ChiSquared, CfsSubsetEval and GainRatio	24	76.39%
[58]	ANN, AdaBoost and SVM	CFS and WFS	9	91%
[59]	SVM, NB, GLM and DT	CorrelationFE, InfoGainFE, RefilefFLE, GainRFE and ClassFE	26	87.69%
[Our work]	Hybrid Ensemble Model (DT, KNN, NB, SVM and LR)	PSO, HHO and HGSO	25	98.6%

## 5. Conclusions and Future Recommendations

Acquiring a higher graduation rate through the provision of a superior education has necessitated the adoption of more precise methods of teaching. To accomplish this, certain precautions are required to be taken to ensure that students are performing well in their studies. Multiple data mining techniques can be used to uncover instructive patterns that can be implemented in the classroom. Additionally, ML regressors contribute to the success of precise education.

The proposed study has used the methods of machine learning to achieve precision education in online knowledge acquisition, specifically throughout the COVID-19 period, within Pakistan. Many recent studies have developed models for the task of student performance prediction; however, in some ways, these models lack generalizability, and the number of selected features is too small, which makes the model overfit due to a smaller number of data considered.

To resolve the aforementioned issues, the current study has used an online questionnaire to collect data from Pakistani students. A total of 12,000 responses were collected and 10,000 were utilized for training the model. Using these data, feature selection and extraction were performed. Three meta-heuristic classifiers of selection and one classifier for data extraction were applied. Of the total data, 25 data attributes were selected which have some potential influence on the performance of students. This dataset was split into the ratio of 70% and 30% for training and testing purposes. Firstly, this research work has utilized a learners' dataset for training, testing, and validation phases for building model on the individual classifiers. Classifiers include DT, NB, SVM, KNN, and LR. SVM has given the highest accuracy of 87.5%. Later, a hybrid ensemble learning model consisting of ML regressors was applied. After training, the output accuracy increased to 98.6% with a minimal error rate of 1.4% achieving precision education. Therefore, this study has aided the advancement of a generalized model proficient in envisaging the learner's performance in academia during online education gaining process in COVID-19, which helped provide early interventions to weak students, maximizing the pass-out ratio. Some limitations of the present study include the fact that, although this study has tried to consider a large dataset, it still lacks many other diverse fields which could be considered to make the model more generalizable. Moreover, the study has applied ML classifiers only. It could use other classifiers of DM as well to check whether they could improve the efficiency of model or not.

In the future, the proposed work can be extended as follows:

- To achieve better accuracy rate in precision education in higher education for the post-COVID-19 period.
- To increase the dataset and the number of attributes, as well as to improve model performance.
- Extending the proposed work by applying several other classifiers in the hybrid ensemble learning model.
- Considering vast academic fields for training models to make it more general in providing diverse feedback to students.
- To compare students from developed and developing countries after the COVID-19 pandemic, this research could be enhanced to evaluate the students' performance.
- This research can be extended to several other countries as well by considering their datasets and applying deep learning models for analysis.
- Future development should also focus on both the asynchronous and synchronous pedagogical approaches across a wide range of educational disciplines.

**Author Contributions:** Conceptualization, R.A., S.A. (Shafiq Ahmad) and S.A. (Saud Altaf); methodology, R.A., S.A. (Saud Altaf) and H.M.; software, R.A.; validation, R.A., S.A. (Shafiq Ahmad) and S.I.; formal analysis, R.A., S.A. (Shafiq Ahmad) and S.A. (Saud Altaf); investigation, R.A.; resources, S.A. (Saud Altaf); data curation, S.A. (Saud Altaf), H.M. and S.A. (Shafiq Ahmad); writing—original draft preparation, R.A., S.H. and S.I.; writing—review and editing, S.A. (Saud Altaf) and S.A. (Shafiq Ahmad); visualization, R.A., S.A. (Saud Altaf); supervision, S.A. (Saud Altaf), S.H. and H.M.; project administration, S.I.; funding acquisition, S.A. (Shafiq Ahmad) and H.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research has received funding from the King Saud University through the Researchers Supporting Project (number RSP2023R387), King Saud University, Riyadh, Saudi Arabia.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi, Pakistan Research Ethics Committee (PMAS-AAUR/R.Eth/63; 13 January 2023).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The datasets generated during and/or analyzed during the current research is available from the corresponding author on reasonable request.

**Acknowledgments:** The authors extend their appreciation to King Saud University for funding this work through the Researchers Supporting Project (number RSP2023R387), King Saud University, Riyadh, Saudi Arabia.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Gomede, E.; Gaffo, F.H.; Brigano, G.U.; Barros, R.M.D.; Mendes, L.D.S. Application of computational intelligence to improve education in smart cities. *Sensors* **2018**, *267*, 267. [[CrossRef](#)]
- Roohi Tauqir, S.; Shahid Hussain, S.; Azhar, S.M. The Role of Vice Chancellors to Promote Higher Education in Pakistan: A Critical Review of Higher Education Commission (HEC) Pakistan's Reforms, 2002. *South Asian J. Manag. Sci.* **2014**, *8*, 2074–2967.
- Yang, S.J.H. Precision education: New challenges for AI in education [conference keynote]. In Proceedings of the 27th International Conference on Computers in Education (ICCE), Kenting, Taiwan, 2–6 December 2019; pp. 27–28.
- Cook, C.R.; Kilgus, S.P.; Burns, M.K. Advancing the science and practice of precision education to enhance student outcomes. *J. Sch. Psychol.* **2018**, *66*, 4–10. [[CrossRef](#)]
- Maldonado-Mahauad, J.; Pérez-Sanagustín, M.; Kizilcec, R.F.; Morales, N.; Munoz-Gama, J. Mining theory-based patterns from Big data: Identifying self-regulated learning strategies in Massive Open Online Courses. *Comput. Hum. Behav.* **2018**, *80*, 179–196. [[CrossRef](#)]
- Baker, E. (Ed.) *International Encyclopedia of Education*, 3rd ed.; Elsevier: Oxford, UK, 2010.
- Siemens, G.; Long, P. Penetrating the fog: Analytics in learning and education. *Educ. Rev.* **2011**, *46*, 30.
- Alsuwaiket, M.; Blasi, A.H.; Al-Msie'deen, R.F. Formulating module assessment for Improved academic performance predictability in higher education. *Eng. Technol. Appl. Sci. Res.* **2019**, *9*, 4287–4291. [[CrossRef](#)]
- Alshareef, F.; Alhakami, H.; Alsubait, T.; Baz, A. Educational Data Mining Applications and Techniques. *Int. J. Adv. Comput. Sci. Appl.* **2020**, *11*, 729–734. [[CrossRef](#)]
- Asad, R.; Arooj, S.; Rehman, S.U. Study of Educational Data Mining Approaches for Student Performance Analysis. *Tech. J.* **2022**, *27*, 68–81.
- Paulsen, M.F.; Nipper, S.; Holmberg, C. *Online Education: Learning Management Systems: Global E-Learning in a Scandinavian Perspective*; NKI Gørlaget: Oslo, Norway, 2003.
- Palvia, S.; Aeron, P.; Gupta, P.; Mahapatra, D.; Parida, R.; Rosner, R.; Sindhi, S. Online education: Worldwide status, challenges, trends, and implications. *J. Glob. Inf. Technol. Manag.* **2018**, *21*, 233–241. [[CrossRef](#)]
- Bates, R.; Khasawneh, S. Self-efficacy and college students' perceptions and use of online learning systems. *Comput. Hum. Behav.* **2007**, *23*, 175–191. [[CrossRef](#)]
- Jordan, M.I.; Mitchell, T.M. Machine learning: Trends, perspectives, and prospects. *Science* **2015**, *349*, 255–260. [[CrossRef](#)] [[PubMed](#)]
- Means, B.; Toyama, Y.; Murphy, R.; Bakia, M.; Jones, K. *Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies*; U.S. Department of Education: Washington, DC, USA, 2009.
- Dascalu, M.D.; Ruseti, S.; Dascalu, M.; McNamara, D.S.; Carabas, M.; Rebedea, T.; Trausan-Matu, S. Before and during COVID-19: A Cohesion Network Analysis of students' online participation in moodle courses. *Comput. Hum. Behav.* **2021**, *121*, 106780. [[CrossRef](#)] [[PubMed](#)]
- Dias, S.B.; Hadjileontiadou, S.J.; Diniz, J.; Hadjileontiadis, L.J. DeepLMS: A deep learning predictive model for supporting online learning in the COVID-19 era. *Sci. Rep.* **2020**, *10*, 19888. [[CrossRef](#)]
- Chakraborty, P.; Mittal, P.; Gupta, M.S.; Yadav, S.; Arora, A. Opinion of students on online education during the COVID-19 pandemic. *Hum. Behav. Emerg. Technol.* **2021**, *3*, 357–365. [[CrossRef](#)]
- Bello, G.; Pennisi, M.A.; Maviglia, R.; Maggiore, S.M.; Bocci, M.G.; Montini, L.; Antonelli, M. Online vs live methods for teaching difficult airway management to anesthesiology residents. *Intensive Care Med.* **2005**, *31*, 547–552. [[CrossRef](#)]
- Al-Azzam, N.; Elsalem, L.; Gombedza, F. A cross-sectional study to determine factors affecting dental and medical students' preference for virtual learning during the COVID-19 outbreak. *Heliyon* **2020**, *6*, e05704. [[CrossRef](#)]
- Chen, E.; Kaczmarek, K.; Ohyama, H. Student perceptions of distance learning strategies during COVID-19. *J. Dent. Educ.* **2021**, *85*, 1190. [[CrossRef](#)]
- Abbasi, S.; Ayoob, T.; Malik, A.; Memon, S.I. Perceptions of students regarding E-learning during COVID-19 at a private medical college. *Pak. J. Med. Sci.* **2020**, *36*, S57. [[CrossRef](#)] [[PubMed](#)]

23. Means, B.; Bakia, M.; Murphy, R. *Learning Online: What Research Tells Us about Whether, When and How*; Routledge: London, UK, 2014.
24. Atlam, E.S.; Ewis, A.; El-Raouf, M.M.A.; Ghoneim, O.; Gad, I. A new approach in identifying the psychological impact of COVID-19 on university student's academic performance. *Alex. Eng. J.* **2022**, *61*, 5223–5233. [[CrossRef](#)]
25. Alsammak, I.L.H.; Mohammed, A.H.; Nasir, I.S. E-learning and COVID-19: Predicting Student Academic Performance Using Data Mining Algorithms. *Webology* **2022**, *19*, 3419–3432. [[CrossRef](#)]
26. Abdelkader, H.E.; Gad, A.G.; Abohany, A.A.; Sorour, S.E. An Efficient Data Mining Technique for Assessing Satisfaction Level With Online Learning for Higher Education Students during the COVID-19. *IEEE Access* **2022**, *10*, 6286–6303. [[CrossRef](#)]
27. Stadlman, M.; Salili, S.M.; Borgaonkar, A.D.; Miri, A.K. Artificial Intelligence Based Model for Prediction of Students' Performance: A Case Study of Synchronous Online Courses During the COVID-19 Pandemic. *J. STEM Educ. Innov. Res.* **2022**, *23*, 39–46.
28. Wang, X.; Zhang, L.; He, T. Learning Performance Prediction-Based Personalized Feedback in Online Learning via Machine Learning. *Sustainability* **2022**, *14*, 7654. [[CrossRef](#)]
29. Alismaiel, O.A.; Cifuentes-Faura, J.; Al-Rahmi, W.M. Social Media Technologies Used for Education: An Empirical Study on TAM Model During the COVID-19 Pandemic. *Front. Educ.* **2022**, *7*, 882831. [[CrossRef](#)]
30. Bansal, V.; Buckchash, H.; Raman, B. Computational Intelligence Enabled Student Performance Estimation in the Age of COVID-19. *SN Comput. Sci.* **2022**, *3*, 41. [[CrossRef](#)] [[PubMed](#)]
31. Zhao, Y.; Ding, Y.; Shen, Y.; Failing, S.; Hwang, J. Different Coping Patterns among US Graduate and Undergraduate Students during COVID-19 Pandemic: A Machine Learning Approach. *Int. J. Environ. Res. Public Health* **2022**, *19*, 2430. [[CrossRef](#)]
32. Al Karim, M.; Ara, M.Y.; Masnad, M.M.; Rasel, M.; Nandi, D. Student performance classification and prediction in fully online environment using Decision tree. *AIUB J. Sci. Eng.* **2021**, *20*, 70–76. [[CrossRef](#)]
33. Yang, C.C.Y.; Chen, I.Y.L.; Ogata, H. Toward Precision Education: Educational Data Mining and Learning Analytics for Identifying Students' Learning Patterns with Ebook Systems. *Educ. Technol. Soc.* **2021**, *24*, 152–163.
34. Al Shalabi, L.; Shaaban, Z.; Kasasbeh, B. Data mining: A preprocessing engine. *J. Comput. Sci.* **2006**, *2*, 735–739. [[CrossRef](#)]
35. Wang, D.; Tan, D.; Liu, L. Particle swarm optimization algorithm: An overview. *Soft Comput.* **2018**, *22*, 387–408. [[CrossRef](#)]
36. Eberhart, R.; Kennedy, J. A new optimizer using particle swarm theory. In Proceedings of the MHS'95. Sixth International Symposium on Micro Machine and Human Science, Nagoya, Japan, 4–6 October 1995; pp. 39–43.
37. Heidari, A.A.; Mirjalili, S.; Faris, H.; Aljarah, I.; Mafarja, M.; Chen, H. Harris hawks optimization: Algorithm and applications. *Future Gener. Comput. Syst.* **2019**, *97*, 849–872. [[CrossRef](#)]
38. Bednarz, J.C. Cooperative hunting Harris' hawks (*Parabuteo unicinctus*). *Science* **1988**, *239*, 1525–1527. [[CrossRef](#)] [[PubMed](#)]
39. Alabool, H.M.; Alarabiat, D.; Abualigah, L.; Heidari, A.A. Harris hawks optimization: A comprehensive review of recent variants and applications. *Neural Comput. Appl.* **2021**, *33*, 8939–8980. [[CrossRef](#)]
40. Hashim, F.A.; Houssein, E.H.; Mabrouk, M.S.; Al-Atabany, W.; Mirjalili, S. Henry gas solubility optimization: A novel physics-based algorithm. *Future Gener. Comput. Syst.* **2019**, *101*, 646–667. [[CrossRef](#)]
41. Staudinger, J.; Roberts, P.V. A critical review of Henry's law constants for environmental applications. *Crit. Rev. Environ. Sci. Technol.* **1996**, *26*, 205–297. [[CrossRef](#)]
42. Yao, R.; Liu, C.; Zhang, L.; Peng, P. Unsupervised anomaly detection using variational auto-encoder based feature extraction. In Proceedings of the 2019 IEEE International Conference on Prognostics and Health Management (ICPHM), San Francisco, CA, USA, 17–20 June 2019; pp. 1–7.
43. Kumar, A.D.; Selvam, R.P.; Kumar, K.S. Review on prediction algorithms in educational data mining. *Int. J. Pure Appl. Math.* **2018**, *118*, 531–537.
44. Kabakchieva, D. Predicting student performance by using data mining methods for classification. *Cybern. Inf. Technol.* **2013**, *13*, 61–72. [[CrossRef](#)]
45. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer Science & Business Media: New York, NY, USA, 1999.
46. Tharwat, A.; Hassanien, A.E.; Elnaghi, B.E. A BA-based algorithm for parameter optimization of support vector machine. *Pattern Recognit. Lett.* **2017**, *93*, 13–22. [[CrossRef](#)]
47. Cover, T.; Hart, P. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27. [[CrossRef](#)]
48. Altman, N.S. An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.* **1992**, *46*, 175–185.
49. Wu, Y.; Ianakiev, K.; Govindaraju, V. Improved k-nearest neighbor classification. *Pattern Recognit.* **2002**, *35*, 2311–2318. [[CrossRef](#)]
50. Chen, Y.; Hao, Y. A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction. *Expert Syst. Appl.* **2017**, *80*, 340–355. [[CrossRef](#)]
51. Li, M.; Xu, H.; Liu, X.; Lu, S. Emotion recognition from multichannel EEG signals using K-nearest neighbor classification. *Technol. Health Care* **2018**, *26*, 509–519. [[CrossRef](#)] [[PubMed](#)]
52. Chirici, G.; Mura, M.; McInerney, D.; Py, N.; Tomppo, E.O.; Waser, L.T.; Travaglini, D.; McRoberts, R.E. A meta-analysis and review of the literature on the k-Nearest Neighbors technique for forestry applications that use remotely sensed data. *Remote Sens. Environ.* **2016**, *176*, 282–294. [[CrossRef](#)]
53. Cariou, C.; Le Moan, S.; Chehdi, K. Improving K-nearest neighbor approaches for density-based pixel clustering in hyperspectral remote sensing images. *Remote Sens.* **2020**, *12*, 3745. [[CrossRef](#)]
54. Farissi, A.; Dahlan, H.M. Genetic algorithm based feature selection with ensemble methods for student academic performance prediction. *J. Phys. Conf. Ser.* **2020**, *1500*, 012110. [[CrossRef](#)]

55. Punlumjeak, W.; Rachburee, N. A comparative study of feature selection techniques for classify student performance. In Proceedings of the 2015 7th International Conference on Information Technology and Electrical Engineering (ICITEE), Chiang Mai, Thailand, 29–30 October 2015; pp. 425–429.
56. Ajibade, S.S.M.; Ahmad, N.B.; Shamsuddin, S.M. An heuristic feature selection algorithm to evaluate academic performance of students. In Proceedings of the 2019 IEEE 10th Control and System Graduate Research Colloquium (ICSGRC), Shah Alam, Malaysia, 2–3 August 2019; pp. 110–114.
57. Zaffar, M.; Hashmani, M.A.; Savita, K.S.; Rizvi, S.S.H. A study of feature selection algorithms for predicting students academic performance. *Int. J. Adv. Comput. Sci. Appl.* **2018**, *9*, 541–549. [[CrossRef](#)]
58. Jalota, C.; Agrawal, R. Feature selection algorithms and student academic performance: A study. In Proceedings of the International Conference on Innovative Computing and Communications: Proceedings of ICICC, Bhubaneswar, India, 22–23 October 2021; Springer: Singapore, 2021; Volume 1, pp. 317–332.
59. Nalić, J.; Martinović, G.; Žagar, D. New hybrid data mining model for credit scoring based on feature selection algorithm and ensemble classifiers. *Adv. Eng. Inform.* **2020**, *45*, 101130. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.