

Article

Prediction of Confusion Attempting Algebra Homework in an Intelligent Tutoring System through Machine Learning Techniques for Educational Sustainable Development

Syed Muhammad Raza Abidi ^{1,*} , Mushtaq Hussain ¹, Yonglin Xu ¹ and Wu Zhang ^{2,*}

¹ School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China; hussain0048@shu.edu.cn (M.H.); wffzxyzl@gmail.com (Y.X.)

² Shanghai Institute of Applied Mathematics and Mechanics, School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China

* Correspondence: razaabdi@live.com or abidi@i.shu.edu.cn (S.M.R.A.); wzhang@shu.edu.cn (W.Z.)

Received: 17 November 2018; Accepted: 19 December 2018; Published: 25 December 2018



Abstract: Incorporating substantial, sustainable development issues into teaching and learning is the ultimate task of Education for Sustainable Development (ESD). The purpose of our study was to identify the confused students who had failed to master the skill(s) given by the tutors as homework using the Intelligent Tutoring System (ITS). We have focused ASSISTments, an ITS in this study, and scrutinized the skill-builder data using machine learning techniques and methods. We used seven candidate models including: Naïve Bayes (NB), Generalized Linear Model (GLM), Logistic Regression (LR), Deep Learning (DL), Decision Tree (DT), Random Forest (RF), and Gradient Boosted Trees (XGBoost). We trained, validated, and tested learning algorithms, performed stratified cross-validation, and measured the performance of the models through various performance metrics, i.e., ROC (Receiver Operating Characteristic), Accuracy, Precision, Recall, F-Measure, Sensitivity, and Specificity. We found RF, GLM, XGBoost, and DL were high accuracy-achieving classifiers. However, other perceptions such as detecting unexplored features that might be related to the forecasting of outputs can also boost the accuracy of the prediction model. Through machine learning methods, we identified the group of students that were confused when attempting the homework exercise, to help foster their knowledge and talent to play a vital role in environmental development.

Keywords: education for sustainable development; confusion; intelligent tutoring system (ITS); ASSISTments; machine learning; computer-based homework; algebra mathematics technology education; sustainable development

1. Introduction

The Intelligent tutoring systems (ITSs) and Massive Open Online Courses (MOOCs) both have similar educational approaches. However, they differ in many aspects, for instance; ITS facilitates instant feedback and scaffolding practice in solving pedagogical problems. While learning through web-based interfaces, students have the opportunity to take hints and watch related topic videos. They also receive guidance on how to practice concepts and attempt the right answer. MOOC, on the other hand, provides much more interactive learning through the learning management system (LMS), in which various forms of instructive video lectures, moderated discussion boards (MDBs), and forums are available for learning with peer feedback [1].

Conversely, the students' learning enhancement is due to the most potent intelligent tutoring systems (ITSs), and these have commercial effectiveness. Although the availability of intelligent tutors is limited, their substantial cost subsidizes for content construction [2,3].

Researchers have generated numerous ITSs for various students, for instance: the Intelligent Tutoring Tools (ITTs) project of Byzantium [4], AutoTutor, Atlas and Why2 [5], Andes [6], ASSISTments [7].

In this study, our main focus will be on ASSISTments (ITS), as we collected data from Skill-builder 2009–2010 [8].

ASSISTments (ITS)

ASSISTments (<https://www.assistments.org>) is an ITS that provides a free web-based platform that facilitates the assessment of learning by both school students and teachers to assess [9]. Experimentation within this system showed that the students who received technology-assisted feedback had higher scores than those who did not receive hints or scaffolding.

ASSISTments is comprised of mathematics-related problems with hints, immediate feedback, and answers. These problems are grouped into sets, that teachers allocate to their students [10].

Essentially, ASSISTments came into being when authors were coaching, particularly math pupils, in middle school. The objective was to pursue students who had basic knowledge and recognized what they wanted to do. In previous years, each student mastered hand-based practicing skill(s), without computers. ASSISTments allows teachers to monitor the comprehension and trail the skill(s) that the students have grasped [7].

It is an environment produced by the Worcester Polytechnic Institute, USA, specially designed for middle schools. More than 50,000 students are registered and use ASSISTments for their homework, and master the skill(s) before their class on the following day. This is possible due to the instant feedback facility of ASSISTments for particular problem associated with skill(s) from the various subjects that include; Mathematics, Science, English, Statistics, etc. Similarly, teachers give two options for the students' assignments: 1. Teachers compile questions with remedies-, hints-, and problem-related videos to master the skill(s) chosen by the student for a particular subject; 2. Teachers also use built-in problem sets and broadcast the homework for students.

In the word ASSISTments, assist is associated with the teachers while assessment is related to the students. Moreover, when a student chooses the wrong answer to a given problem, the prompt feedback pings student to rectify it by taking a hint or trying another option. This allows the teachers to log the results in accurate instantaneous output, and later they can use this evidence to develop strategies for the upcoming lesson [7].

Two types of educational contents are available in ASSISTments. One is related to the mathematics textbook homework, or the problems that teachers write for their students; and the second type is designed explicitly for skill-learning practice and mastery called "skill builders." In ASSISTments, current skill-builder data consists of more than 300 topics related to middle school mathematics. The purpose of skill-builder is to master the skill by practicing the problem assigned by the teacher under defined standards or principles [9].

Skill-builder assures that the student has an expert in the topic or skill before moving forward to grasp other tasks. This is the best type of content in ASSISTments, testing and acknowledging what students learn. It is also mandatory for each student to correct three questions in a row until they understand the preliminary ability on the chosen topic area [7].

Many peer-reviewed journals publish articles in the context of prediction. ASSISTments has been broadly and widely used for data mining exploration [11], employing Bayesian networks [12], or using the platform to make classifiers of students' demonstrative state [13].

The ultimate goal of our research is to identify and predict the confusion of students while solving their mathematics homework, using mastery skill-builder learning after attempting the teacher defined criterion (e.g., correcting three consecutive answers on similar math problems). Even though instant

feedback from ITS causes student confusion. Therefore, as per education for sustainable development, we found the gap in which every student wishes to nurture their talent, knowledge, and experience to become a responsible member and citizen of the society during their ongoing development. However, the ultimate goal of sustainable development is to eliminate common issues and frustrations, which can only be achieved when we as people, students, teachers, etc. find out the remedies to resolve these issues by taking instant feedback, hints, and help from technology and humans. Confusion among the students while using ITS is a major hindrance in maintaining sustainable development, leading to an inability to meet the needs of the present.

Wang et al. [14] discourage manual classification due to its imperfection, irreproducibility, and classifications generated by the human eye. Thus, we used the digitized, computerized classification as our data were diversified. Therefore, by using machine learning classifiers allowed us to categorize students who were confused and those who were not. Machine learning algorithms work on the principle of statistics, and for this objective we used the statistical programming language R and RapidMiner 8.1 to analyze and predict the confused students in an ITS skill-builder session and showed the results.

To accomplish this task, the following was our research question, which we considered in order to bridge the gap.

- Can we categorize which machine learning algorithms are the best fit to classify mastery skill learning confusion among the students using skill-builder in an intelligent tutoring system from chosen skills?

Further, the structure of the paper is as follows: In Section 2, we present a short overview of related works and research on the particular subject; in Section 3, we define related methods used, and proposed predictive methods; in Section 4, we interpret the results and discuss prediction performance; in Section 5, we leave the reader with concluding thoughts, shortcomings, and future recommendations.

2. Related Works

ITSs are computer technology-based programs that have been developed for different subject areas (e.g., algebra, physics, science, medicine, statistics, etc.) to assist students and learners obtain domain-specific intellectual knowledge. It also models the mental, cognitive states of students and learners to provide modified instructions through instant and prompt feedback. This system provides an interface that presents and receives information to communicate with learners. For instance, by learning the concept of the subject domain (e.g., algebra), the learner can interact through interfaces to solve problems while looking for a hints or answering the questions [15].

ITSs distinguish learning for students from various abilities, understandings, and performances. Variations in learning style are taken into account to distinguish learning. Learning style can be predicted by taking independent behavior variables during tutoring discussions. Fuzzy trees have been induced to predict the learning style of individuals. Outcome classification has been observed due to the automatic behavior of learning from a collection of data [16].

Material regarding the mathematics course in ASSISTments comprises of difficulties with solutions, and in-time suggestions. Furthermore, substantial assistance readily available over the Internet to resolve the issue that students solve online. Another type of material is precisely designed for mastery-focused skills training named “Skill-builders” as discussed above. At the moment, ASSISTments covers more than 300 matters related to mathematics for middle school, and the capability is given to teachers to allocate skill-builders to pupils, allowing them to rehearse those problems while emphasizing the desired skill(s) until they receive the pre-defined standards for accuracy [9].

Nonetheless, very limited research has discovered the importance of ITS used as homework [17]. Many types of research corroborate the significance of ITS used while in a class for students at

school [18]. Hence, it was very inspiring when [19] ANDES and ITS used in this manner communicated favorable outcomes.

At present, ASSISTments is being used by massive numbers of students at middle and high school for their evening homework. The instant advice regarding homework, allows students to feel comfortable, and enables tutors to monitor the reports specifying students achievements [17]. So far, for the evening homework, multilayered tutoring systems are not suitable as on the other hand, technology-supported instructions which disseminate the same questions with a fast response about the problem are more appropriate [20].

According to Singh et al., homework on web-based tools using the tutoring system is more authentic and robust in learning and mastering student skill(s) compared to previous old-fashioned paper-based traditional style. Furthermore, this research focuses on comparing the instantaneous responses from the tutoring system against the feedback received by students from the tutor the next working day, which is time-consuming and reduces the learning ability as a whole. They further showed that 8th-grade math students who were indulged in both scenarios, expanded pointedly with an effect size of 0.40 by using technology-assisted homework [18].

As described by Fyfe, around the globe, ITS and technology-assisted homework achieved fame and pervasiveness in schools [21]. Conferring to Ma et al., personalized education, and well before advice, is the solid foundation of intelligent tutoring systems [15]. The objective of the study by Fyfe was to reveal an investigational assessment of an algebra class for middle school students who had preceding variable knowledge affected by the technology-based responses [21].

Generally, many types of research support the notion that using the in-time responses from ITSs, as usual, has useful properties on learning outputs as opposed to no response from ITSs [20,21]. Lee et al., Baker et al., and Gupta and Rose, all classified that confusion, both its roots and penalties, can be easily recognized through the performance and student actions [22–24].

Confusion causes students to halt, reproduce, and begin problem-solving to rectify their own confusion. The only way to cope with confusion is that every student must have in-depth knowledge of complicated matters, as fought with confusion is an intellectual action [25,26]. On the other hand, if a healthy learning atmosphere offers an adequate platform and timely assistance to students, and they efficiently normalize their confusion, they could achieve positive outcomes [25–28].

Moreover, many scholars used different methods like “classification or knowledge engineering” to detect the disturbance changes in students, particularly confusion [29]. Likewise, Conati and MacLaren established a detector built on logged data and survey question groupings to forecast self-described student disturbance, although this model was healthier to recognize attentive and inquisitive students but ineffective at classifying confused students [30].

Baker et al. conducted strong research particularly focused on computer software designed for education, e.g., ITSs to automatically identify confusion through affect detection where they collected this information through semantic actions of students, and labeled the existing Pittsburgh Science of Learning Center (PSLC) DataShop log files. In this research, they defined confusion as the slower pattern of students’ actions while attempting the pre-defined teacher criterion before the starting of mastery skill-builder assignment or homework. Authors focused the preliminary step and observed the percentage of clip actions [26]. Table 1 shows the summary of the illustrative methods used in the previous research in the domain of ITS.

Table 1. Summary of previous research.

Methods Used	Advantages	Disadvantages
Fuzzy Decision Trees [16]	The most popular choice for learning and reasoning systems particularly from feature based discrete values, dealing issue with uncertainty, noise, and inexact data	Does not take into consideration the connections between behavior variables and, due to the uncertainty intrinsically present in modeling learning styles, small differences in behavior can lead to incorrect predictions
Hierarchical Linear Regression Model (HLM) [9]	Simple relationship with a limited number of variables, it is the ordinary least square (OLS) regression-based analysis that takes the hierarchical structure of data into consideration	Complex form, assessed data using a fixed parameter, and thus insufficient analysis due to the neglect of the shared variance
Regression Analysis [21]	It is a statistical analysis technique used to forecast future conditions, provides the relationship between two or more related variables with the help of which we can quickly estimate or predict the unknown values of one variable from the known values of another variable	The cause and effect of the relationship between variables remain unchanged, cannot be used in a qualitative phenomenon, long and complex calculations and analysis
Mixed-effects Modeling [31]	Useful where repeated measurements are made on the same statistical units or made on clusters of related statistical units, Includes a combination of fixed and random effects, and very appropriate dealing with missing values	Increase the power in studies without sample structure
ANCOVA (Analysis of Covariance) [17]	Better power, enhanced capability to detect and evaluate interactions, and the availability of extensions to deal with measurement error in the covariates	It may not be helpful when the imbalance between the groups is large
Sensor-Free Detectors [26]	These detectors are designed to operate solely on the information available through the students' semantic actions within the interface	Not substantially better especially when subject to stringent cross-validation processes
ANOVA (Analysis of Variance) [18]	The statistical method used to compare the means of multiple groups (more than two sets of data), can also control the overall Type-I error rate	Not suitable when the samples are not independent
Discovery with a model approach [25]	Leaves clear data trials that can be re-inspected in the future, development of lifelong learning skills, supports an active engagement of the learner, use activities to focus attention, can be motivating	Inefficient, too time-consuming, possibility of confusing learner's if no initial framework is available, requires that the teacher prepares for too many corrections if discovery turns out to be wrong
Data-Driven Methodology [24]	Make up-to-date design decisions based on real user needs and prioritize issues to solve based on its relative impact for users	Data are trusted blindly without any uncertainty, and are often messy and even incorrect. Low-quality data leads to low-quality decisions
Feedback Model [22]	Reduces the discrepancy between current and desired understanding	Feedback is only built on something. It is of little use when there is no initial learning or surface information, and under particular circumstances, an instruction is more effective than feedback
Probabilistic Student Model [19]	Beneficial for making responses to help requests that are particularly relevant to domains in which there is uncertainty about the student's mental state	Unable to look at the problem of deciding what kind of response to give to the student at any given time

Table 1. Cont.

Methods Used	Advantages	Disadvantages
Academic Emotions Questionnaire (AEQ) assumptions of a cognitive-motivational model [30]	Useful for analyzing students' emotions in learning, as emotions are multifaceted and can be measured reliably by the AEQ	Due to primarily used cross-sectional or predictive designs, not allowing precise inferences of causal relations

3. Methods

This section clarifies and illustrates the effectiveness of raw data to classification via machine learning methods. Figure 1 depicts the visualization of raw data to workflow classification:

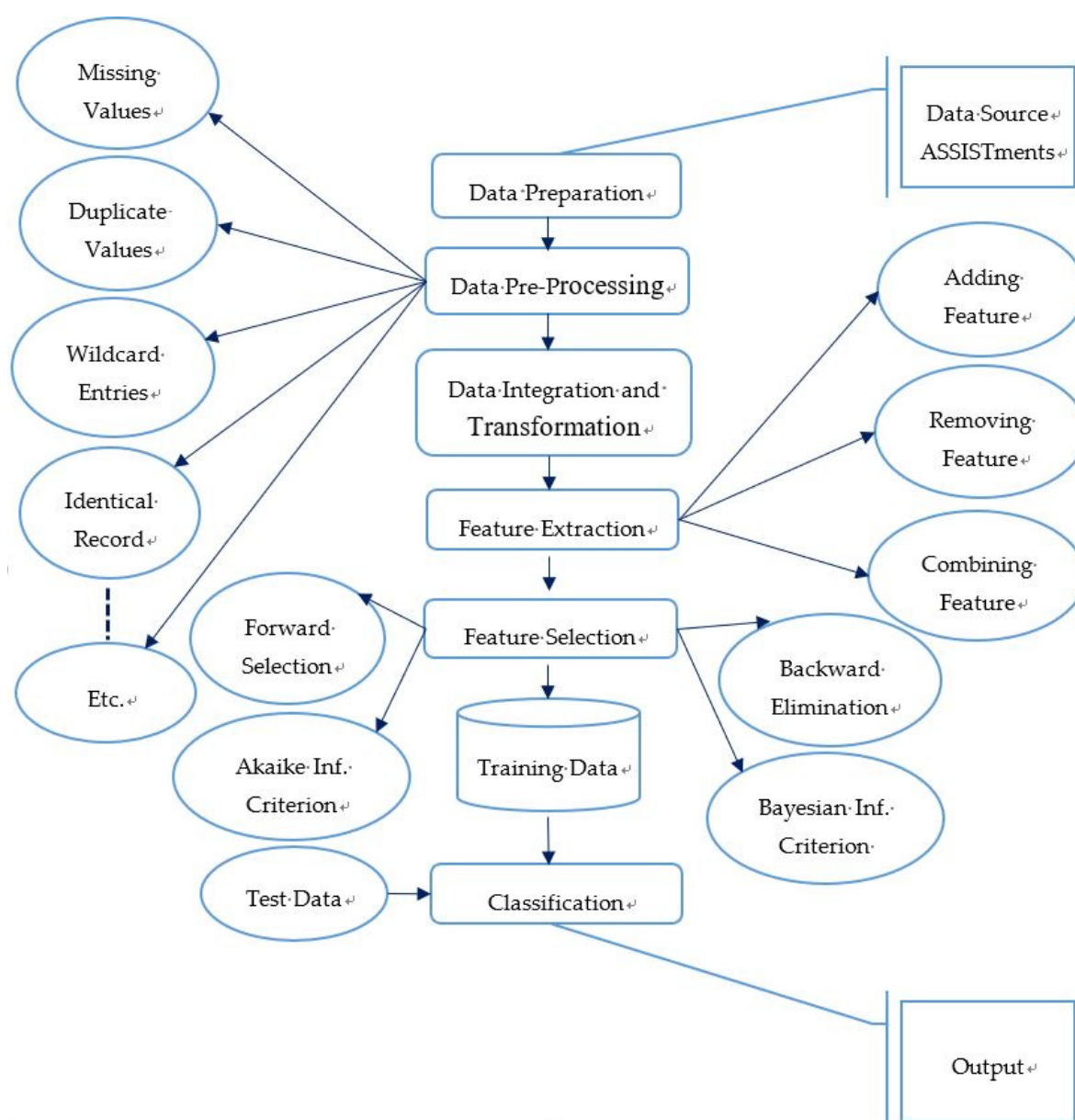


Figure 1. A pictorial view of raw data for workflow classification.

3.1. Preparation of Data

In this study, we used dataset (in the supplementary) collected from ASSISTments, Skill-builder data 2009–2010 [8].

Skill-builder problem sets have the following features:

- Questions based on one specific skill; a question can have multiple skill tagging's.
- Students must answer three questions correctly in a row to complete the assignment.
- If a student uses the tutoring ("Hint" or "Break this Problem into Steps"), the question will be marked incorrect.
- Students will know immediately if they answered the question correctly.
- If a student is unable to figure out the problem on his or her own, the last hint will answer.
- Currently, this feature is only available for math problem sets.

In this whole dataset, various features are available relating to the mastery skill-builder learning. There were almost 72 participating schools with 93 mastery skills in algebra mathematics and about 28 features. We targeted the school ID-73 because it had maximum availability of records, i.e., the total record of 5148 with 20 attributes, and four attributes had various missing values initially amongst other school IDs, and selected 10 mastery skills, i.e., Absolute Value, Addition and Subtraction of Positive Decimals, Box and Whisker, Circle Graph, Multiplication Fractions, Ordering Fractions, Percent Of, Subtraction of Whole Numbers, Venn Diagram, and Write Linear Equation from Graph, as the maximum number of students selected these chosen skills and after removing duplicate values, we got total 166 distinct student IDs remaining.

3.1.1. Measurements and Covariates

We selected the predictors (original, attempt_count, ms_first_response, correct, hint_total, overlap_time, and opportunity) from the list of features available in the dataset, following are the basic description, measured ROC, accuracy, precision, recall, F-measure, sensitivity, and specificity as performance indicators that were used by machine learning algorithms.

Original: If '0' means scaffolding problem, and other than '0' means the main problem.

Attempt_count: Number of times a student entered an answer.

Ms_first_response: Time between the start time and first student action.

Correct: If '0' means Incorrect on first attempt otherwise correct.

Hint_total: Number of possible hints on the problem.

Overlap_time: This is meant to be the time taken by student to finish the problem.

Opportunity: The number of opportunities each student has to practice on the skill.

3.1.2. Discretization of Predicted Variable

After the precise selection of predictors, we were interested in learning what features appraise the status of the confused/not confused student. In order to determine this aspect, we used a feature extraction technique to select the predicted variable. We chose and combined three variables, i.e., attempt_count, correct, and overlap_time, with the conditions mentioned below to form a new feature called "student state", and on the basis of that we categorized the status of the confused/not confused students, '1' designated confused and '0' not confused.

If (attempt_count) > Total mean (attempt_count)
 and (correct) < Total mean (correct)
 and (overlap_time) > Total mean (overlap_time) then (Confuse) otherwise (Not Confuse)

3.1.3. Experimental Manipulations or Interventions

We used the 10-fold cross-validation technique to divide our datasets into a standard (80%–20%) of training, and test datasets, respectively with stratified sampling, as our response variable was dichotomous.

3.1.4. Statistical Analysis

For statistical analysis, we used the statistical programming language R (<https://cran.r-project.org/>), and used RStudio (<https://www.rstudio.com/>) to perform basic descriptive and regression analysis.

We also checked the correlation between explanatory and response variables and identified which variables were significant, bring information to the model, and which variables do not.

3.2. Pre-Processing of Data

Data extracted either from databases, log files, or Microsoft Excel files required cleaning. Although it was in good shape, data cleaning before moving ahead was an absolute part of the pre-processing. Data could be noisy, missing, or uneven. Machine learning algorithms performed pre-processing of data up-to some extent, but these algorithms were more robust if we manually accomplished this step.

3.3. Integration and Transformation of Data

For better statistical analysis and classification, data must be integrated and transformed. For this objective, Table 2 illustrates five amongst ten mastery skills as a snapshot and each skill has four attributes (stated above in 3.1. *Preparation of Data*) for each student.

Table 2. Mastery skills and corresponding attributes.

Skill Name	Attribute-1	Attribute-2	Attribute-3	Attribute-4
Absolute Value	3.8	8.4	1.7	20.3
Box and Whisker	0.7	0.0	2.2	1.7
Circle Graph	2.5	6.3	4.8	15.5
Venn Diagram	4.0	6.8	2.2	31.2
Write Linear Equation from Graph	0.6	1.7	0.0	28.9

3.4. Feature Extraction

Feature extraction is a procedure for creating new attributes amongst existing features. Figure 1 shows a snapshot of the feature extraction step. Classification is considered to be an essential step as the performance measure of the learning process depends on significant explanatory variables. In many real-world cases, we cumulatively extract features, alter if needed, combine them, and produce one variable. The same procedure might be adopted for the selection of response variables. In this study, we selected ten mastery skills and also nominated the associated explanatory variables and clubbed them to form 40 explanatory variables for each student in each mastery skill.

3.5. Feature Selection

As revealed in Figure 1 above, there are many criterions available for feature selection, for Instance: Backward Elimination, Forward Selection, AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion), DIC (Deviance Information Criterion), Bayes factor, Mallow's Cp, etc. We performed Backward Elimination using the adjusted R^2 method with the cutoff p -value of 0.05 to construct our model because it is a common way [32]. We started with the full model and eliminated one variable at a time until the parsimonious model was reached [33].

3.6. Training of Model

Before predicting the confusion amongst the students attempting algebra mastery skills homework in ITS, we essentially trained machine learning algorithms to curtail the difference between actual and predicted values. For this objective, we split the data into (80%–20%) ratio with stratified sampling, as our response variable was nominal.

3.7. Testing and Evaluation of the Model

Model evaluation is an essential part of implementing machine learning techniques. When a machine is trained on known data, then we evaluate the model on unseen data to verify that the model is good enough, learned, and classified correctly. We used seven learning models in this study and the description is as follows:

Naïve Bayes (NB) is fast and efficient probabilistic classifier with an extensive record of research. Due to its robustness, precision, and competence this method is usually referenced [34].

Generalized Linear Model (GLM) is eventually the enhancement of old-style linear models, and the series of instructions inside these models turn to data by using the MLE technique. These models give tremendous, high-speed, and parallel computation with a small number of explanatory variables with non-zero constants [35,36].

Logistic Regression (LR) is a broadly used statistical technique for the classification of binary output. When predicting the output of the nominal response variable, usually the logistic regression algorithm is used [36,37].

Deep Learning (DL) works by a neural network that takes information that offers information about other data as input, and produces the outcome by using many layers [38]. DL can tune and select the model at an optimal level by itself, and it also achieves mining of features instinctively without involvement and interaction of individuals or humans, which spectacularly saves plenty of determination and time [39].

Decision Tree (DT) depicts tree-like building, where it has nodes (internal and leaf). It is made by training data, which consists of data rows or records. Each record is formed by a set of features and outcome labels [40].

Random Forest (RF) associates multiple tree input variables in a group. New occurrences being classified are broken down into trees, and each tree states a classification [41]. RF generates many arbitrary trees on various subsets of a data set, and the subsequent model builds on polling of these trees. Due to this variance, it is less likely to overtraining [36].

Gradient Boosted Trees (XGBoost) correlates with the Gradient Boosting Machine (GBM), which is another boosting algorithm. It produces good accuracy due to the competences of parallel computing and the effective linear model solver. It also creates decision trees which are individual understandable models [42].

We executed all the above-mentioned learning algorithms on the RapidMiner 8.1 for our experimentation. RapidMiner has a collection of machine learning/data mining algorithms, and we calculated the desired output by accuracy. Furthermore, we tested these algorithms on other performance metrics mentioned in Section 3.7.1.

3.7.1. Performance Metrics

In this study, we adopted the most common and widely used performance metrics of [43,44]. They used the ROC Area under the curve (AUC) to calculate the performance of prediction models.

- ROC Curve or AUC

ROC curve demonstrates the association between true positive and false positive rates [45]. It contains several thresholds, and each produces a 2×2 contingency table. We also used other performance metrics, i.e., accuracy, precision, recall, F-Measure, sensitivity, and specificity.

Detailed results for this present study are described in Section 4 (Results).

3.8. Classification

Plenty of machine learning/data mining tools are available. We used RapidMiner 8.1 for our investigation and testing. RapidMiner studio is well equipped with data mining/machine learning tasks with sufficient state-of-the-art collections of machine learning algorithms along with data access, pre-processing, blending, cleansing, modeling, visualization, and validation operators, which give high-tech advanced platforms to perform machine learning/data mining tasks in the most efficient and well-organized manner [46,47]. As we executed supervised learning, succeeding some linear and non-linear classifiers were used for classification, application and verification.

3.9. Statistical Analysis and Parameters

In order to discover the significance of explanatory variables for the prediction of students' confusion in algebra mastery skills in ITS, it was imperative to explore the predictors (explanatory variables), and its impact on response variables statistically. Although machine learning algorithms intrinsically perform statistical tests and analysis of variables, it is always good practice to check manually before applying any machine learning methods and techniques.

Table 3 is the weights (ranks) of the attributes, which show the universal significance of each attribute for the value of the target attribute, independent of the modeling algorithm.

Table 3. Ranks of attributes.

Attributes	Ranks (Weights)
ms_first_response_absValue	1.000
ms_first_response_percentOf	0.901
original_absValue	0.813
ms_first_response_subWholeNum	0.755
ms_first_response_multiFrac	0.627
opportunity_cirGraph	0.568
hint_total_absValue	0.561
opportunity_absValue	0.551
ms_first_response_ordFrac	0.498
original_addSubPosDec	0.490

We have used the statistical programming language R with the standard cut-off level of probability value (p -value 0.05). Table 4 displays a correlation matrix p -values with the response variable.

In this statistical summary of correlation, we found nine predictor variables that were most significant, i.e., their values were ($p < 0.05$) related to the dichotomous response variable. Correlation was used to measure the strength of the linear association between two numeric variables. Many types of correlation coefficients exist. i.e., Pearson, Kendall, and Spearman. We used Pearson's correlation coefficient as it commonly used in linear regression. It is denoted by (r or R) and its value is always in the range from -1 to $+1$, where $+1$ specifies strong positive correlation, and -1 the strong negative correlation.

Statistically, after using backward elimination techniques, we ended up with the final model which validates the significance of nine explanatory variables, Table 5 portrays descriptive statistics; Table 6 shows regression analysis including predictor's coefficients, standard errors, p -values, etc.; and Table 7 reveals regression summary.

Table 4. Correlation matrix with response variable.

Attribute	Label
user_id	−0.11
original_absValue	0.23
ms_first_response_absValue	0.28
hint_total_absValue	0.16
opportunity_absValue	0.16
original_addSubPosDec	0.15
ms_first_response_addSubPosDec	0.11
opportunity_addSubPosDec	0.10
original_box&Whis	−0.05
hint_total_box&Whis	−0.05
opportunity_box&Whis	−0.05
original_cirGraph	−0.06
ms_first_response_cirGraph	0.12
hint_total_cirGraph	−0.03
opportunity_cirGraph	−0.17
original_multiFrac	0.12
ms_first_response_multiFrac	0.18
hint_total_multiFrac	0.12
opportunity_multiFrac	0.07
original_ordFrac	0.05
ms_first_response_ordFrac	0.15
opportunity_ordFrac	−0.02
original_percentOf	0.01
ms_first_response_percentOf	0.26
hint_total_percentOf	0.08
opportunity_percentOf	−0.13
original_subWholeNum	0.02
ms_first_response_subWholeNum	0.22
opportunity_subWholeNum	−0.11
original_vennDiag	−0.07
ms_first_response_vennDiag	−0.11
hint_total_vennDiag	0.11
opportunity_vennDiag	−0.09
original_wrtLinEqGraph	−0.11
ms_first_response_wrtLinEqGraph	−0.08
opportunity_wrtLinEqGraph	−0.02

Table 5. Descriptive statistics.

Min	1Q	Median	3Q	Max
−0.94696	−0.04588	0.05360	0.13200	0.44114

Table 6. Regression analysis of predictors (explanatory variables).

Coefficients	Estimate	Std. Error	t Value	pr (> t)	
(Intercept)	0.5588550	0.0475093	11.763	$<2 \times 10^{-16}$	***
ms_first_response_absValue	0.0247446	0.0043088	5.743	4.73×10^{-8}	***
original_addSubPosDec	0.0104006	0.0058002	1.793	0.074887	.
original_box.whis	−0.1641421	0.0685545	−2.394	0.017838	*
opportunity_box.whis	0.0317376	0.0196939	1.612	0.109082	***
original_cirGraph	0.0530879	0.0123338	4.304	2.95×10^{-5}	***
opportunity_cirGraph	−0.0041942	0.0010725	−3.911	0.000137	***
hint_total_vennDiag	0.0273546	0.0059493	4.598	8.77×10^{-6}	**
original_wrtLinEqGraph	−0.0602548	0.0218751	−2.754	0.006577	*
opportunity_wrtLinEqGraph	0.0005436	0.0002309	2.355	0.019787	

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Annotation = ***, Significance level: 0.001, p -value [0, 0.001], **: Significance level: 0.01, p -value (0.001, 0.01], *: Significance level: 0.05, p -value (0.01, 0.05], .: Significance level: 0.1, p -value (0.05, 0.1], Blank space: Significance level: 1, p -value (0.1, 1].

Table 7. Regression statistics summary.

Residual Standard Error	Degrees of Freedom	Multiple R-Squared	Adjusted R-Squared	F-Statistic	p-Value
0.2936	156	0.3213	0.2821	8.205	6.123×10^{-10}

Furthermore, Table 8 reveals the efficient feature of R language, which shows what maximum adjusted R^2 value could achieve through given explanatory variables. These are the ten chosen skills with ASSISTments data attributes. The purpose of showing this Table is to validate the adjusted R^2 value in Table 7.

Table 8. Maximum adjusted R^2 representation.

Attribute	Maximum Adjusted R^2 Value
(Intercept)	0.20–0.28
user_id	0.00
original_absValue	0.00
ms_first_response_absValue	0.20–0.28
hint_total_absValue	0.00
opportunity_absValue	0.00
original_addSubPosDec	0.265–0.28
ms_first_response_addSubPosDec	0.00
opportunity_addSubPosDec	0.00
original_box&Whis	0.25–0.28
hint_total_box&Whis	0.23–0.25
opportunity_box&Whis	0.275–0.28
original_cirGraph	0.23–0.28
ms_first_response_cirGraph	0.00
hint_total_cirGraph	0.00
opportunity_cirGraph	0.23–0.28
original_multiFrac	0.00
ms_first_response_multiFrac	0.00
hint_total_multiFrac	0.00
opportunity_multiFrac	0.00
original_ordFrac	0.00
ms_first_response_ordFrac	0.00
opportunity_ordFrac	0.00
original_percentOf	0.20–0.23
ms_first_response_percentOf	0.20–0.23
hint_total_percentOf	0.00
opportunity_percentOf	0.00
original_subWholeNum	0.20–0.23
ms_first_response_subWholeNum	0.00
opportunity_subWholeNum	0.25–0.28
original_vennDiag	0.25–0.28
ms_first_response_vennDiag	0.00
hint_total_vennDiag	0.20–0.28
opportunity_vennDiag	0.00
original_wrtLinEqGraph	0.00
ms_first_response_wrtLinEqGraph	0.00
opportunity_wrtLinEqGraph	0.00

4. Results

As discussed in Section 3; Methods, ROC is Receiver Operating Characteristic and also recognized as ROC AUC or just simply ROC curve. It demonstrates the relationship between the true positive rate (TPR) and false positive rate (FPR). It also determines the cooperation between sensitivity and specificity as both are contradictory, i.e., when the sensitivity rises, specificity declines. The accuracy can be monitored if the curve is closer to the top left corner and could be considered the finest results, but if curve comes closer to the diagonal angle (45°), the results would not be accurate. Moreover, ROC AUC value >0.9 portrays excellent results; values between $0.8\text{--}0.9$ are considered good, those between $0.7\text{--}0.8$ reflects fair, and <0.6 illustrates poor [48].

Graphical representation of ROC AUC shown in Figures 2 and 3 depicts the AUC values graphs for seven machine learning algorithms, which correctly predicted the confusion amongst the students attempting algebra mastery skills in ITS. In Figure 3, vertical axis number shows the percentage value between 0% and 100%.

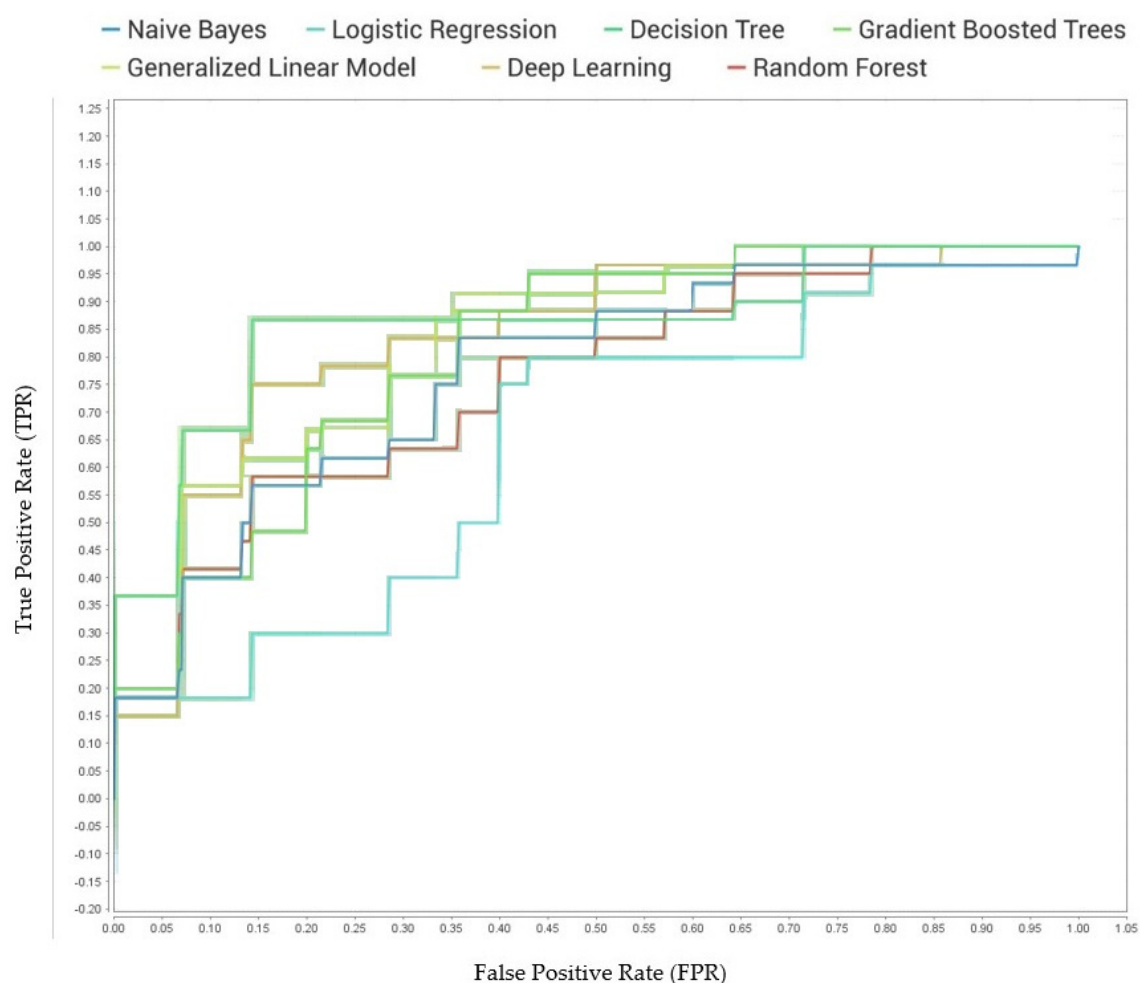


Figure 2. A symbolic view of Area under the curve (AUC) of seven selected machine learning algorithms.

We have constructed seven candidate models built on various machine learning methods. The performance achieved by each classifier is shown in Figure 4, which reveals the accuracy performance metric of each model by repetitive sampling validation technique, in which it randomly replicates division of training and test data.

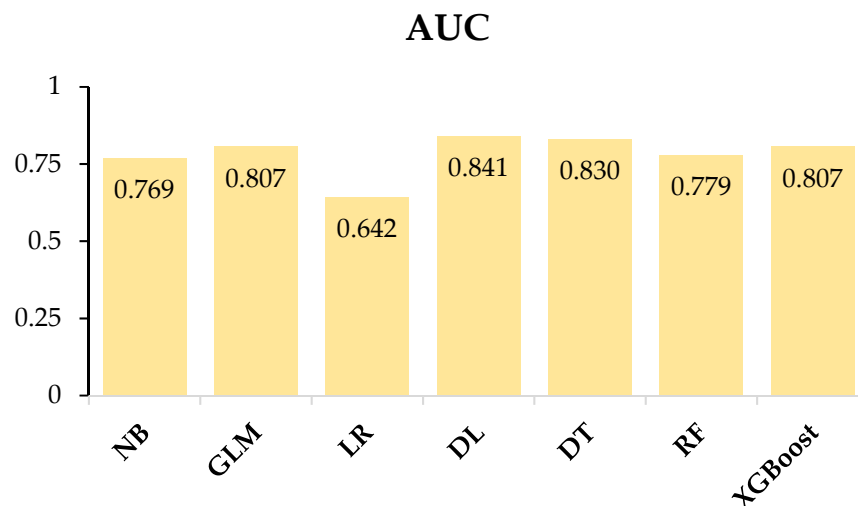


Figure 3. Graphical representation of ROC AUC of nominated machine learning models.

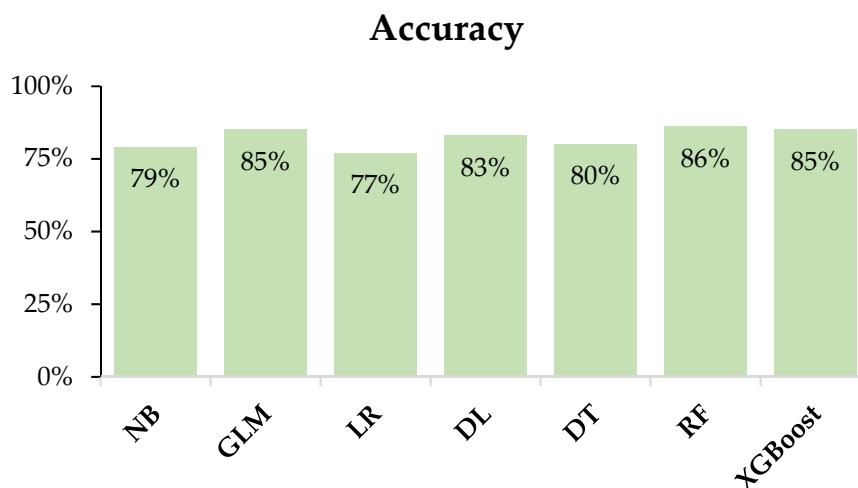


Figure 4. Candidate models' summary concerning the accuracy.

These results illustrate the ratio of time we require to acceptably predict the cases. We attained maximum accuracy with RF, GLM, XGBoost, and DL, i.e., 86.1%, 84.9%, 84.9%, and 83.1%, respectively. We also employed other classifiers, i.e., DT: 79.5%, NB: 78.9%, and LR: 77.1% as shown in Table 9.

Table 9. Complete detail of learning algorithms along with runtime.

Model	Accuracy	Precision	Recall	F-Measure	Sensitivity	Specificity	Runtime
NB	78.9%	26.9%	30.4%	28.6%	30.4%	86.7%	87 ms
GLM	84.9%	42.9%	26.1%	32.4%	26.1%	94.4%	5 s
LR	77.1%	25.8%	34.8%	29.6%	34.8%	83.9%	772 ms
DL	83.1%	39.1%	39.1%	39.1%	39.1%	90.2%	1 s
DT	79.5%	21.1%	17.4%	19.1%	17.4%	89.5%	527 ms
RF	86.1%	50.0%	13.0%	20.7%	13.0%	97.9%	3 s
XGBoost	84.9%	42.9%	26.1%	32.4%	26.1%	94.4%	1 min 33 s

ms: Millisecond; s: second; min: minute.

We checked other performance metrics, which we discussed in Section 3. Figure 5 displays the performance of seven machine learning algorithms regarding precision, recall, F-measure, sensitivity, and specificity.

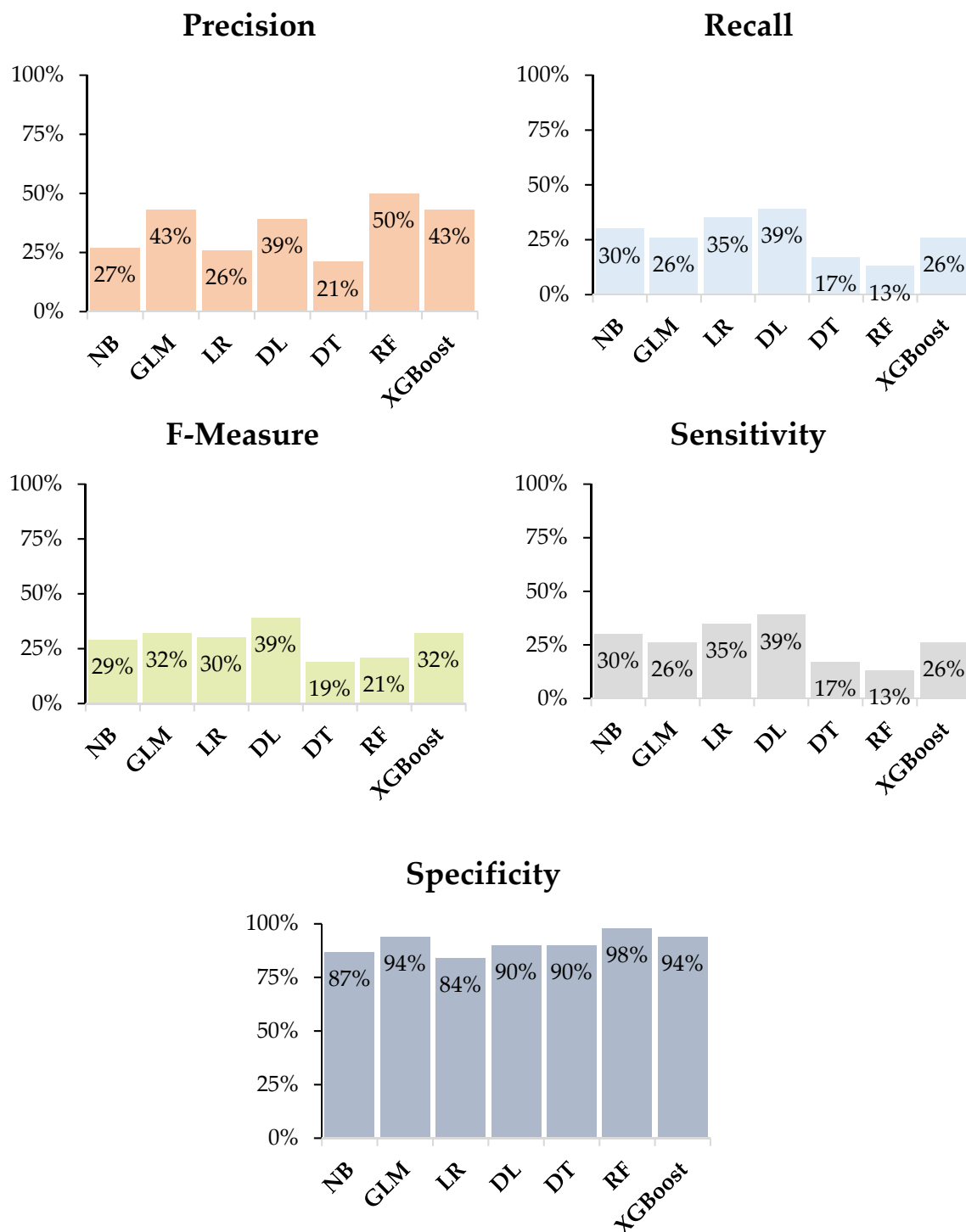
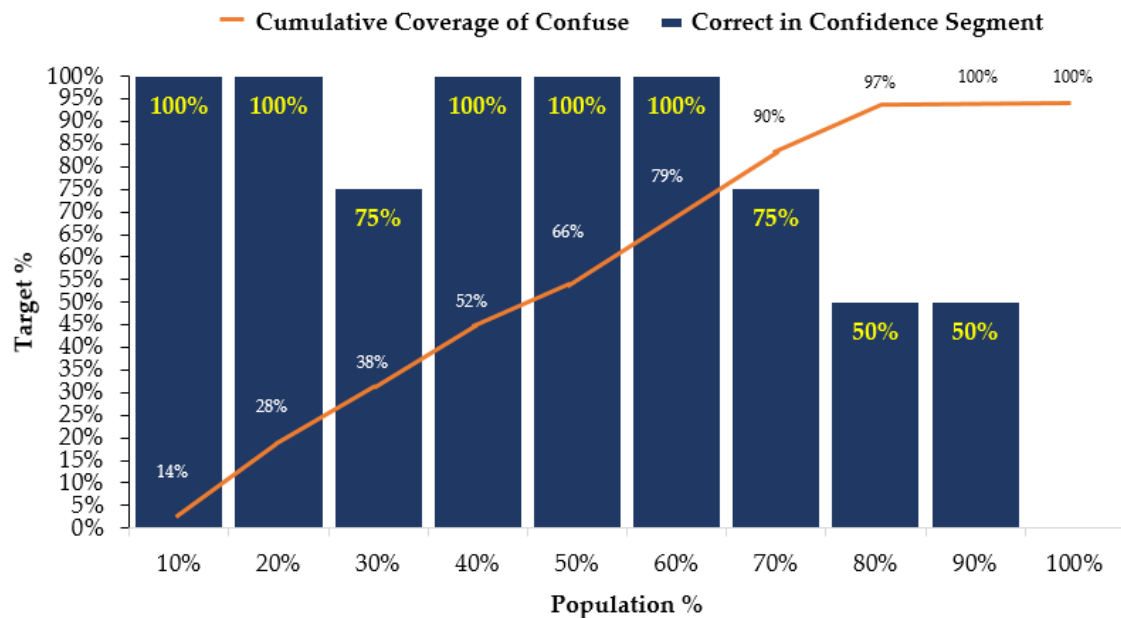


Figure 5. Performance of machine learning algorithms relating to performance measures.

Figure 6 displays the lift charts of high accuracy achieving machine learning models. A lift chart is a graphical illustration of the enhancement that a model delivers when related against a random guess [49]. It shows the efficiency of the model by measuring the ratio between the outcome obtained “with and without a model” [36].

Random Forest (RF) Lift chart:



Generalized linear model (GLM) Lift chart:

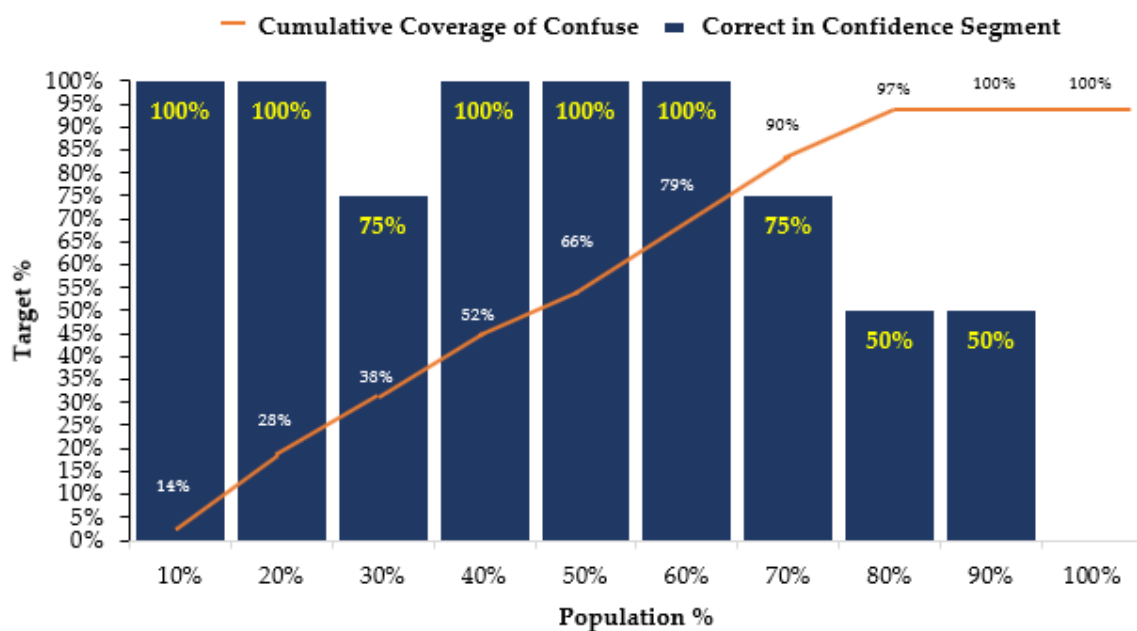
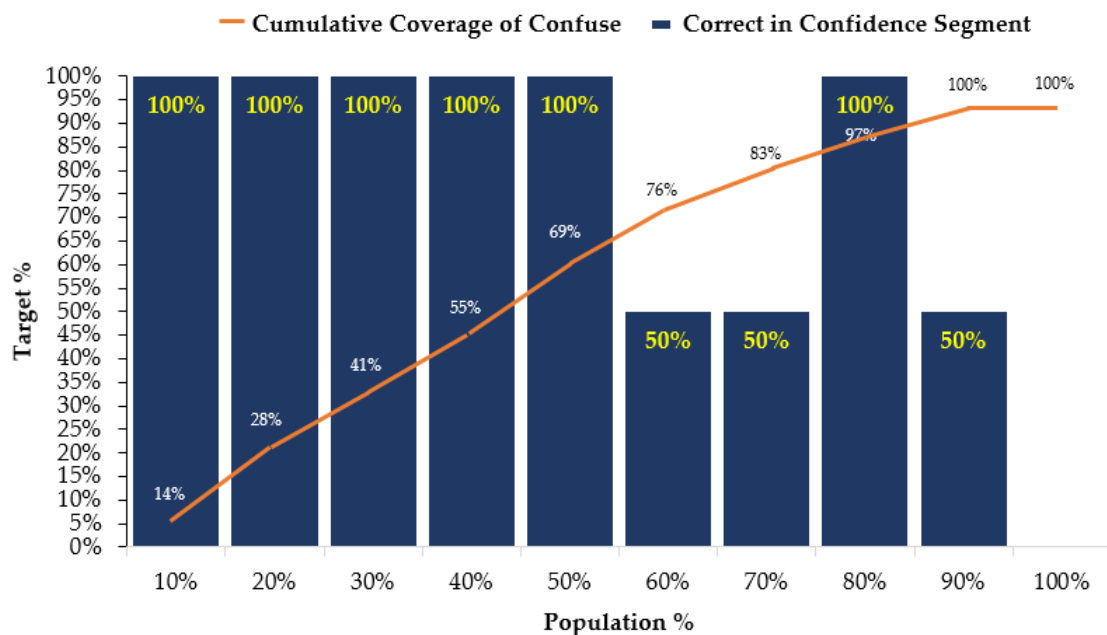


Figure 6. Cont.

Gradient Boosted Tree (XGBoost) Lift chart:



Deep Learning (DL) Lift chart:

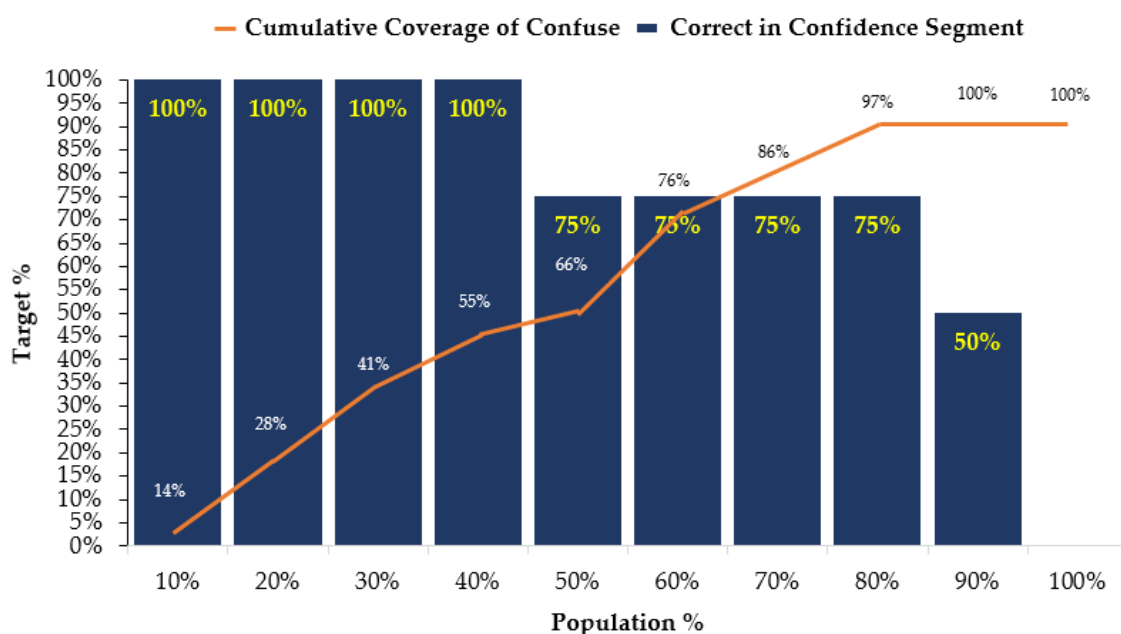


Figure 6. High achieving accuracy models' lift chart.

5. Discussion and Conclusions

This research investigates models for the prediction of confused students attempting homework using skill-builder in ITS. Analyzing confusion is a task of classification, and machine learning has plenty of robust classification algorithms. In this study, we used machine learning methods for the experiment. Performing techniques of data mining on ITS is a tough task because the multiple features are related in various extents with many of noisy data and missing fields. We then extracted explanatory variables (input features) and targeted (output) response variables from ITS. This was followed by applying machine learning models NB, GLM, LR, DL, DT, RF, and XGBoost, respectively.

The results demonstrated that RF, GLM, XGBoost, and DL models attained a high accuracy of 86.1%, 84.9%, 84.9%, and 83.1% in predicting the students' confusion in the algebra mastery skills in ITS.

Such a result can assist school tutors in next day classes, and identifying student groups which were confused attempting the homework exercise in mastery skill-builder. This tool also highlights which skill(s) need(s) more attention for further practice. Furthermore, tutors can also govern learning behaviors and student performances during various mastery skill(s), allowing them to focus only problematic skill(s) in the next day of the class, which will save a lot of time and effort for both tutors and students.

Our study has many decent inferences both educationally and practically. Firstly, to the best of our information and facts, our research, amongst the previous studies for predicting confusion by using machine learning methods for sustainable educational development, is one of the rare studies that have focused on these specific aspects. ITS contributes to sustainable development in education, as the development focuses on the necessities of the present-day without compromising the future needs. The objective of sustainable development is to stabilize our environmental, economic, and social needs [50].

Sustainable development in education is an interdisciplinary learning approach that covers the combined environmental, social, and economic aspects of the formal and informal curriculum. This educational approach can assist students develop their aptitudes, knowledge, and experience to show a significant role in ecological development, and become liable members of society. Participation and sharing teaching and learning techniques and methods are also required to encourage and empower learners to change and alter their performances and take corrective actions for sustainable development. Critical thinking, visualizing the future, and decisions making are the skills and abilities that Education for Sustainable Development (ESD) promotes [51].

5.1. Shortcomings

The shortcoming in this study is that we have used a limited number of variables as there are more attributes available, which can be used for further investigation and could be statistically stronger. Another shortcoming is that by doing rigorous optimization techniques like changing criterion, pruning, selecting a threshold of machine learning models (algorithms) we could achieve better results.

5.2. Future Recommendations

In future work, we will design and apply some strategies to further augment our model. First, a more decent optimization parameter can be used for building a more accurate model, for instance: In DT, we can change the criterion i.e., *gain_ratio*, *Information_gain*, *gini_index*, *accuracy*, *maximal depth* parameter etc.; in LR, we can set the criterion, solver, reproducible, use regularization, etc.; and in XGBoost, we can alter *maximal depth*, *min rows*, *min split improvement*, *number of bins*, etc. to optimize performance. Additionally, we will consider Convolutional Neural Network (CNN) [52] in our experimentation as this methodology assures negligible loss and best classification accuracy. Other approaches like the Dynamic Bayesian Network (DBN), Gauss Cloud Model, and Cloud Reasoning Algorithm will also be taken into consideration for best classifying accuracy and precision [53–55].

Secondly, other kinds of classification methods and techniques can be measured. Though the machine learning techniques used and applied in this study are relatively comprehensive but still, there are various unexplored methods/techniques that can be applied to the prediction problem in the domain of students in intelligent tutoring system. Thirdly, other structures and features in the data may enhance the prediction correctness and accuracy that can be added. Furthermore, as per the tutors' perspective, we can identify the benefits associated while detecting the confusion in a group of students solving mathematics homework using skill-builder in ITS.

Supplementary Materials: The dataset is available online at <http://www.mdpi.com/2071-1050/11/1/105/s1>.

Author Contributions: Conceptualization, M.H. and Y.X.; methodology, M.H. and S.M.R.A.; software, S.M.R.A.; validation, M.H., Y.X., and S.M.R.A.; formal analysis, S.M.R.A.; resources, S.M.R.A.; writing—original draft preparation, S.M.R.A.; writing—review and editing, M.H.; visualization, S.M.R.A. and Y.X.; supervision, W.Z.; funding acquisition, W.Z.

Funding: This research was funded by “National Natural Science Foundation of China, grant number 91630206, and 61572434”, and “The National Key R&D Program of China, grant number 2017YFB0701501”.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Aleven, V.; Baker, R.; Blomberg, N.; Andres, J.M.; Sewall, J.; Wang, Y.; Popescu, O. Integrating MOOCs and Intelligent Tutoring Systems: EdX, GIFT, and CTAT. In Proceedings of the 5th Annual Generalized Intelligent Framework for Tutoring Users Symposium, Orlando, FL, USA, 11 May 2017; p. 11.
2. Koedinger, K.R.; Anderson, J.R.; Hadley, W.H.; Mark, M.A. Intelligent tutoring goes to school in the big city. *Int. J. Artif. Intell. Educ.* **1997**, *8*, 30–43.
3. Aleven, V.; Sewall, J.; McLaren, B.M.; Koedinger, K.R. Rapid Authoring of Intelligent Tutors for Real-World and Experimental Use. In Proceedings of the Sixth IEEE International Conference on Advanced Learning Technologies (ICALT'06), Kerkrade, The Netherlands, 5–7 July 2006; pp. 1–5.
4. Kinshuk. Computer Aided Learning for Entry Level Accountancy Students. 1996. British Library, Imaging Services North, Boston Spa, Wetherby, West Yorkshire, UK. Available online: <https://core.ac.uk/download/pdf/77603309.pdf> (accessed on 28 April 2018).
5. Freedman, R. Atlas: A plan manager for mixed-initiative, multimodal dialogue. In Proceedings of the AAAI-99 Workshop on Mixed-initiative Intelligence, Orlando, FL, USA, 18–19 July 1999; pp. 1–8.
6. Gertner, A.; Conati, C.; VanLehn, K. *Procedural Help in Andes: Generating Hints Using a Bayesian Network Student Model*; American Association for Artificial Intelligence (AAAI): Palo Alto, CA, USA, 1988; pp. 106–111.
7. Heffernan, N.T.; Heffernan, C.L. The ASSISTments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *Int. J. Artif. Intell. Educ.* **2014**, *24*, 470–497. [CrossRef]
8. Heffernan, N. ASSISTmentsData. 2012. Available online: <https://sites.google.com/site/assistmentsdata/home/assistent-2009-2010-data/skill-builder-data-2009-2010> (accessed on 2 February 2018).
9. Roschelle, J.; Feng, M.; Murphy, R.F.; Mason, C.A. Online Mathematics Homework Increases Student Achievement. *AERA Open* **2016**, *2*. [CrossRef]
10. Roschelle, J.; Murphy, R.; Feng, M.; International, S.R.I.; Mason, C.; Fairman, J. Rigor and Relevance in an Efficacy Study of an Online Mathematics Homework Intervention. 2014. Available online: <https://www.sri.com/newsroom/press-releases/rigorous-sri-study-shows-online-mathematics-homework-program-developed> (accessed on 30 April 2018).
11. Feng, M.; Heffernan, N.; Koedinger, K. Addressing the assessment challenge with an online system that tutors as it assesses. *User Model. User Adapt. Interact.* **2009**, *19*, 243–266. [CrossRef]
12. Pardos, Z.; Heffernan, N. Tutor Modeling vs. Student Modeling. In Proceedings of the Twenty-Fifth International Florida Artificial Intelligence Research Society Conference, Marco Island, FL, USA, 23–25 May 2012.
13. Pedro, M.O.Z.S.; Baker, R.S.J.D.; Gowda, S.M.; Heffernan, N.T. Towards an understanding of affect and knowledge from student interaction with an intelligent tutoring system. In *Artificial Intelligence in Education; Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2013; Volume 7926, pp. 41–50.
14. Wang, S.; Lu, S.; Dong, Z.; Yang, J.; Yang, M.; Zhang, Y. Dual-Tree Complex Wavelet Transform and Twin Support Vector Machine for Pathological Brain Detection. *Appl. Sci.* **2016**, *6*, 169. [CrossRef]
15. Ma, W.; Adesope, O.; Nesbit, J.C.; Liu, Q. Intelligent tutoring systems and learning outcomes: A meta-analysis. *J. Educ. Psychol.* **2014**, *106*, 901–918. [CrossRef]
16. Crockett, K.; Latham, A.; Whitton, N. On predicting learning styles in conversational intelligent tutoring systems using fuzzy decision trees. *Int. J. Hum. Comput. Stud.* **2017**, *97*, 98–115. [CrossRef]

17. Kelly, K.; Heffernan, N.; Heffernan, C.; Goldman, S.; Pellegrino, J.; Soffer-goldstein, D. WEB-BASED HOMEWORK. In Proceedings of the Joint Meeting of PME 38 and PME-NA 36, Vancouver, BC, Canada, 15–20 July 2014; Volume 3, pp. 417–424.
18. Singh, R.; Saleem, M.; Pradhan, P.; Heffernan, C.; Heffernan, N.T.; Razzaq, L.; Dailey, M.D.; O'Connor, C.; Mulcahy, C. Feedback during web-based homework: The role of hints. In *Artificial Intelligence in Education; Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2011; Volume 6738, pp. 328–336.
19. Vanlehn, K.; Lynch, C.; Schulze, K. The Andes physics tutoring system: Lessons learned. *Int. J. Artif. Intell. Edu.* **2005**, *15*, 1–51.
20. Kelly, K.; Heffernan, N.; Heffernan, C.; Goldman, S.; Pellegrino, J.; Goldstein, D.S. Estimating the effect of web-based homework. In *Artificial Intelligence in Education; Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2013; Volume 7926, pp. 824–827.
21. Fyfe, E.R. Providing feedback on computer-based algebra homework in middle-school classrooms. *Comput. Hum. Behav.* **2016**, *63*, 568–574. [[CrossRef](#)]
22. Hattie, J.; Timperley, H. The Power of feedback. *Review of Educational Research.* *Rev. Educ. Res.* **2007**, *77*, 81–112. [[CrossRef](#)]
23. Alfieri, L.; Brooks, P.J.; Aldrich, N.J.; Tenenbaum, H.R. Does Discovery-Based Instruction Enhance Learning? *J. Educ. Psychol.* **2011**, *103*, 1–18. [[CrossRef](#)]
24. Gupta, N.K.; Rose, C.P. Understanding Instructional Support Needs of Emerging Internet Users for Web-based Information Seeking. *JEDM J. Educ. Data Min.* **2010**, *2*, 38–82.
25. Lee, D.M.C.; Rodrigo, M.M.T.; Baker, R.S.J.D.; Sugay, J.O.; Coronel, A. Exploring the relationship between novice programmer confusion and achievement. In *Affective Computing and Intelligent Interaction; Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2011; Volume 6974, pp. 175–184.
26. Baker, R.S.J.d.; Gowda, S.M.; Wixon, M.; Kalka, J.; Wagner, A.Z.; Salvi, A.; Aleven, V.; Kusbit, G.K.; Ocumpaugh, J.; Ocumpaugh, L. Towards Sensor-Free Affect Detection in Cognitive Tutor Algebra. In Proceedings of the 5th International Conference on Educational Data Mining, Chania, Greece, 19–21 June 2012; pp. 126–133.
27. Lehman, B.; D'Mello, S.; Graesser, A. Interventions to regulate confusion during learning. In *Intelligent Tutoring Systems; Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7315, pp. 576–578.
28. Pardos, Z.A.; Baker, R.S.J.D.; Pedro, M.O.C.Z.S.; Gowda, S.M.; Gowda, S.M. Affective States and State Tests: Investigating How Affect Throughout the School Year Predicts End of Year Learning Outcomes. In Proceedings of the Third International Conference on Learning Analytics and Knowledge (LAK'13), Leuven, Belgium, 8–12 April 2013; Volume 1, pp. 117–124.
29. Pekrun, R.; Goetz, T.; Titz, W.; Perry, R.P. Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educ. Psychol.* **2002**, *37*, 91–105. [[CrossRef](#)]
30. Conati, C.; MacLaren, H. Empirically building and evaluating a probabilistic model of user affect. *User Model. User Adapt. Interact.* **2009**, *19*, 267–303. [[CrossRef](#)]
31. D'Mello, S.; Lehman, B.; Pekrun, R.; Graesser, A. Confusion can be beneficial for learning. *Learn. Instr.* **2014**, *29*, 153–170. [[CrossRef](#)]
32. Vu, D.H.; Muttaqi, K.M.; Agalgaonkar, A.P. A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables. *Appl. Energy* **2015**, *140*, 385–394. [[CrossRef](#)]
33. Rundel, M.C. Linear Regression and Modeling. 2018. Available online: <https://www.coursera.org/learn/linear-regression-model> (accessed on 18 January 2018).
34. Smith, V.C.; Lange, A.; Huston, D.R. Predictive modeling to forecast student outcomes and drive effective interventions in online community college courses. *J. Asynchronous Learn. Netw.* **2012**, *16*, 51–61. [[CrossRef](#)]
35. Ng, V.K.Y.; Cribbie, R.A. The gamma generalized linear model, log transformation, and the robust Yuen-Welch test for analyzing group means with skewed and heteroscedastic data. *Commun. Stat. Simul. Comput.* **2018**, 1–18. [[CrossRef](#)]
36. RapidMiner. RapidMiner Documentation. 2016. Available online: <https://docs.rapidminer.com/latest/studio/operators/> (accessed on 29 April 2018).
37. Peng, C.; Lee, K.; Ingersoll, G.M. An introduction to logistic regression analysis and reporting. *J. Educ. Res.* **2002**, *96*, 3–14. [[CrossRef](#)]

38. Li, W.; Gao, M.; Li, H.; Xiong, Q.; Wen, J.; Wu, Z. Dropout prediction in MOOCs using behavior features and multi-view semi-supervised learning. In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; pp. 3130–3137.
39. Xing, W.; Du, D. Dropout Prediction in MOOCs: Using Deep Learning for Personalized Intervention. *J. Educ. Comput. Res.* **2018**. [CrossRef]
40. Kabra, R.R.; Bichkar, R.S. Performance prediction of engineering students using decision trees. *Int. J. Comput. Appl.* **2011**, *36*, 8–12.
41. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
42. Cobos, R.; Wilde, A.; Zaluska, E. Predicting attrition from massive open online courses in FutureLearn and edX. In Proceedings of the 7th International Learning Analytics and Knowledge Conference, Simon Fraser University, Vancouver, BC, Canada, 13–17 March 2017; Volume 1967, pp. 74–93.
43. Pontius, R.G.; Si, K. The total operating characteristic to measure diagnostic ability for multiple thresholds. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 570–583. [CrossRef]
44. Wang, S.; Yang, M.; Zhang, Y.; Zhang, Y.-D. Detection of left-sided and right-sided hearing loss via fractional Fourier transform. *Entropy* **2016**, *18*, 194. [CrossRef]
45. Hussain, M.; Zhu, W.; Zhang, W.; Abidi, S.M.R.; Ali, S. Using machine learning to predict student difficulties from learning session data. *Artif. Intell. Rev.* **2018**, 1–27. [CrossRef]
46. Mierswa, I.; Wurst, M.; Klinkenberg, R.; Scholz, M.; Euler, T. YALE: Rapid prototyping for complex data mining tasks. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Philadelphia, PA, USA, 20–23 August 2006; Volume 2006, pp. 935–940.
47. Godwin, K.E.; Almeda, M.V.; Petrocchia, M.; Baker, R.S.; Fisher, A.V. Classroom activities and off-task behavior in elementary school children. In Proceedings of the 35th Annual Meeting of the Cognitive Science Society, Berlin, Germany, 31 July–3 August 2013; No. 2001. pp. 2428–2433.
48. Metz, C.E. Basic Principles of ROC Analysis. 2018. Available online: <http://gim.unmc.edu/dxtests/ROC1.htm> (accessed on 4 May 2018).
49. Microsoft. Lift Chart (Analysis Services—Data Mining). 2018. Available online: <https://docs.microsoft.com/en-us/sql/analysis-services/data-mining/lift-chart-analysis-services-data-mining?view=sql-analysis-services-2017> (accessed on 7 May 2018).
50. Education for Sustainable Development | Higher Education Academy. Available online: <https://www.heacademy.ac.uk/knowledge-hub/education-sustainable-development-0> (accessed on 16 November 2018).
51. The Brundtland Commission. Available online: <https://www.sustainabledevelopment2015.org/AdvocacyToolkit/index.php/earth-summit-history/past-earth-summits/58-the-brundtland-commission> (accessed on 13 November 2018).
52. Gao, H.; Cheng, B.; Wang, J.; Li, K.; Zhao, J.; Li, D. Object Classification Using CNN-Based Fusion of Vision and LIDAR in Autonomous Vehicle Environment. *IEEE Trans. Ind. Inform.* **2018**, *14*, 4224–4230. [CrossRef]
53. Xie, G.; Gao, H.; Qian, L.; Huang, B.; Li, K.; Wang, J. Vehicle Trajectory Prediction by Integrating. *Mech. Syst. Signal Process.* **2018**, *102*, 1377–1380.
54. Li, D.; Gao, H. A Hardware Platform Framework for an Intelligent Vehicle Based on a Driving Brain. *Engineering* **2018**, *4*, 464–470. [CrossRef]
55. Gao, H.; Zhang, X.; Liu, Y.; Li, D. Cloud Model Approach for Lateral Control of Intelligent Vehicle Systems. *Sci. Program.* **2016**, *2016*, 1–12. [CrossRef]

