

Article

Educational Sustainability through Big Data Assimilation to Quantify Academic Procrastination Using Ensemble Classifiers

Syed Muhammad Raza Abidi ¹, Wu Zhang ^{1,2}, Saqib Ali Haidery ³, Sanam Shahla Rizvi ⁴, Rabia Riaz ⁵, Hu Ding ² and Se Jin Kwon ^{6,*}

¹ School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China; razaabdi@live.com (S.M.R.A.); wzhang@shu.edu.cn (W.Z.)

² Shanghai Institute of Applied Mathematics and Mechanics, Shanghai University, Shanghai 200444, China; dinghu3@shu.edu.cn

³ School of Communications and Information Engineering, Shanghai University, Shanghai 200444, China; alisaqib@shu.edu.cn

⁴ Raptor Interactive (Pty) Ltd., Eco Boulevard, Witch Hazel Ave, Centurion 0157, South Africa; sanam.shahla@raptorinteractive.com

⁵ Department of CS IT, University of Azad Jammu and Kashmir, Muzaffarabad 13100, Pakistan; rabiaiqbal18@gmail.com

⁶ Department of Computer Engineering, Kangwon National University, Samcheok 25806, Korea

* Correspondence: sjkwon@kangwon.ac.kr; Tel.: +82-33-570-6377

Received: 27 June 2020; Accepted: 24 July 2020; Published: 28 July 2020



Abstract: Ubiquitous online learning is continuing to expand, and the factors affecting success and educational sustainability need to be quantified. Procrastination is one of the compelling characteristics that students observe as a failure to achieve the weaker outcomes. Past studies have mainly assessed the behaviors of procrastination by describing explanatory work. Throughout this research, we concentrate on predictive measures to identify and forecast procrastinator students by using ensemble machine learning models (i.e., Logistic Regression, Decision Tree, Gradient Boosting, and Forest). Our results indicate that the Gradient Boosting autotuned is a predictive champion model of high precision compared to the other default and hyper-parameterized tuned models in the pipeline. The accuracy we enumerated for the VALIDATION partition dataset is 91.77 percent, based on the Kolmogorov–Smirnov statistics. Additionally, our model allows teachers to monitor each procrastinator student who interacts with the web-based e-learning platform and take corrective action on the next day of the class. The earlier prediction of such procrastination behaviors would assist teachers in classifying students before completing the task, homework, or mastery of a skill, which is useful and a path to developing a sustainable atmosphere for education or education for sustainable development.

Keywords: academic procrastination; tree ensemble classifiers; machine learning; education for sustainable development; intelligent tutoring system; SAS Visual data mining and machine learning

1. Introduction

Education for Sustainable Development (ESD) aims at promoting and improving the quality of lifelong learning aimed at acquiring sustainability knowledge, skills and values and reorienting academic curricula (rethinking, integrating, reforming, and greening education towards sustainability), thus raising public awareness through a better understanding of the concept of Sustainable Development (SD) [1–4]. ESD has the potential to develop individuals' potential by improving their awareness, skills,

and ability to act more sustainably [5–7]. Educational institutions deliver massive online courses as a recent trend in online education, and it seems a successful learning experience for students. Ubiquitous learning, unlike any location, is easily accessible to the students anywhere [8].

The demand for online courses is speedily growing and becoming a viable feature of the educational system. It is revealed that almost every third student takes the course online while attending their college or university, and this figure has surprisingly increased over the past decades [9].

Notably, the online learning environment needs a higher gradation of self-regulation than the physical classroom environment [10]. It is different from face-to-face learning, where students and teachers collaborate and interact [11]. A time- and place-independent e-learning system [12,13], a self-regulated learning process [14], and an inter-disciplinary approach to teaching and learning are vital factors in ESD [15]. In a virtual learning environment, e-learning depends on the synchronous and asynchronous correspondence and teamwork [16].

While taking an online course, students must track and manage the learning time effectively and keep an eye on progress and learning skills to meet critical deadlines. However, it appears somehow that students can lack mastery of the abilities. For example, when using the Learning Management System (LMS), students use it from home or elsewhere through their laptops, desktop PCs, tablets, or mobile phones [17]. They do all of the things on LMS, including assignments, quizzes, and graded discussion boards (GDBs). LMS can cause students to deviate and disconnect more from the course than they would in a face-to-face setting, or grasp the ability. All this uncertainty can result in procrastination as students refrain from fulfilling their responsibilities on time and wait until the last moment before the given deadline. Identifying students' learning habits related to time management, and in particular, procrastination, is one of the critical factors for enhancing online learning [18].

Higher education institutions should emphasize the promotion of interdisciplinary thinking and analysis, which is the basis of SD, by teaching the more multifaceted interrelationships between economic, social, and environmental concepts. It is done at the tertiary or university level in many countries, in terms of the principle of SD (development of course curricula within different academic disciplines), but also in terms of the daily routine operations of the institutions (as in the whole school approach) [19]. Curricular development should also be achieved by research on pedagogical methods and their effectiveness in delivering sustainability education and training programs for educators [20].

Procrastination is the propensity to delay the initiation or completion of duties, tasks, or works. To assure the fulfillment of assignments, the learners must resolve that weakness. Therefore, it is most important to identify or predict which learners are at risk of procrastination and the processes by which online learning will affect performance [21].

Schraw and colleagues explain that academic procrastination is the intentional nature of the learner to defer or postpone work that must be completed before time [22]. In the study of Michinov and colleagues, they used the e-learning framework, Modular Object-Oriented Dynamic Learning Environment (Moodle), to calculate the rate of procrastination at the beginning and end of the course. They implemented a web-based questionnaire for all participants, who were asked to fill in the questionnaire after the final evaluation, which explores learners' changing behaviors. The questionnaire also documented the different details, such as times (when learners had to start or resume working), feeling like giving up or leaving, and feeling motivated or excited about completing their course. This research further found that learners who perform poorly or have less than average results are more likely to procrastinate [21].

In the online course or education, procrastination may also be assessed by evaluating students or learners' acts or behaviors, such as clicking on lecture videos, attempting GDBs, submitting assignments and quizzes on time, visiting and reading web pages, and constructively looking for and answering questions in discussion forums. A standard method of calculating procrastination is to measure the total time a student is engaging with or communicating with the LMS before a course term. Additionally, studies that take these kinds of measures as a sign of procrastination show that the symbols have a negative association with course results [23–25].

The research of Kazerouni and colleagues explains that the assessment of learner habits, such as additional growth and procrastination, has a direct relationship with the accuracy of solutions, the completion time of the given work, or the overall amount of time spent working on a solution. They also found that procrastination contributes to lower scores [24].

The author You discussed that the learners who fail to study consistently and in time are characterized by poor academic results, procrastination, and abandonment. These triggers have proven to be persistent problems in online learning, and some approaches need to be developed and explored to keep learners motivated, enthusiastic, supervised, and active in their courses to fulfill the task [25].

LMS mechanically records all the learning activities when the student logs in and traces the self-regulated learning as well. The usage of LMS in institutions is widespread, as it is considered to be the best source for identifying the learners' presence and academic performance [26]. Moreover, LMS also assists teachers in determining essential insights [27]. It provides instant feedback for the students who are at risk of failure, procrastination, and withdrawal from the course and allows regulating the instructional policies and procedures [28].

Self-supervision deficiency is one of the causes to provoke procrastination in education, and indeed has a higher negative impact on success, usually resulting in backing off or abandonment [29]. Many studies also elaborate on the importance of formal and appropriate learning behaviors in online education [30].

Using two found behaviors linked to academic procrastination in LMS data, the author You examined the damaging impact of procrastination on academic achievement. The study identified the students who failed to stick to the weekly scheduled assignment submission or their delays in the timely submission of the assignment, which are considered the vital signs of academic procrastination. The results of this study showed that chronic and repetitive learning behaviors related to time management and preparation must be taken into account in order to prevent erroneously expected success and achievement of the course [26].

Motivated by the previous studies, we used a data-driven approach and an intelligent tutoring system (ITS), i.e., ASSISTments, to recognize academic procrastination using skill-building data from the highest number of responses to 111 skills from students. In this paper, we discuss the use of machine learning ensemble models for evaluating and predicting students who are procrastinating. To the best of our knowledge, a limited amount of research was performed on a quantitative approach to demystifying procrastinate behavior. Our ensemble models describe the hidden patterns of learning activity in the results, where the trends among the learners are recognized as matching academic procrastinating and non-procrastinating behaviors. Additionally, we detect that students who regularly procrastinate may often show a mixture of mind planning and procrastinating actions during learning. We devise a composite score addressing these subtleties, which incorporates the overall estimate of being considered a procrastinate learner. The approach we build through machine learning allows for a fine-grained study of procrastination and its connection with learning outcomes that can help understand more successful rearrangements of online learning and sustainable reforms in education.

The rest of this article is arranged as follows: Section 2 explains the evaluation of past studies on identifying procrastination among students. In Section 3, relationship explorations and procedures used in this paper are presented. In Section 4, we introduce the experiment and results, along with the proposed methods and modeling to evaluate and analyze the predictive performance through prediction models built. In Section 5, limitations of the study are discussed, and then finally, we leave the reader with discussion, concluding thoughts, and future perspectives in Section 6.

2. Related Works

The research of [25,31] appraises the significant indicators of course success in online learning, such as scheduled planning and management of time. These behaviors exhibit the learners' interests and attitudes toward online learning. On the other hand, learners who are weak in managing the time

and postpone finishing the work until the final deadline reveal substantial evidence of procrastination behaviors, which can indeed lead to poor performance and success [18].

Artino and Jones explored the relationship between emotions and self-regulated learning behavior in the undergraduate online learning course. In this study, the authors highlighted features, such as boredom, dissatisfaction, and enjoyment related to achievement emotions, and considered elaboration and metacognition as self-regulated behavior [11].

Many past studies have focused on identifying the behavior of procrastination, which hampers students from mastering online courses [18,26,31]. The study of Dvorak and Jia compared two groups of students using online analytics data and observed their respective working habits. They found that the group which completed their assignments on time had earned better grades as compared to those who do not complete the work before deadlines [32].

Further, other studies have found similar outcomes; when learners defer their assigned coursework, they are probably performing poorly [25,26]. These findings verify the unsatisfactory description of procrastination as well as highlight the ordinary learning behaviors [18].

Another comprehensive work of Chen and colleagues has shown that students from depreciated backgrounds, such as those who come from below-average income households, or who are attending college or universities for the first time, are most likely to leave or surrender Science, Technology, Engineering, and Mathematics (STEM) degrees or majors [33].

Styck highlighted the problems which are intensifying in online learning. Many essential aspects that describe the matters related to students who are coming from diminished environments or cultures hinder the students' achievement and success. Such negative factors include monetary matters, shortage of counseling, and a sense of segregation [34]. Prior work has also shown the rapidly increasing trend that underrepresented groups are more likely to engage in procrastination than their counterparts. They determined the characteristics of habitual procrastinators from a global sample based on various demographic variables [35].

Park and colleagues in their study measured the procrastination by using interviews and questionnaires. Their ultimate aim was to explore the association and connection of students' cultural upbringing and different states of procrastination on an individual basis. They looked at time management attitudes, particularly for those students who enroll in online learning courses [18].

In an online learning environment, measuring the procrastination is very candid and identical. Most commonly, researchers calculate the finishing time of a task and determine the differences between the free time [36] and the deadline of the particular task [23,24].

Table 1 represents the past research in chronological order and highlights the methods used to demystify self-regulating learning behavior and procrastination among the learners in online courses.

Table 1. Summary of previous research.

Research Type	Authors	Features Focused	Methods Used
Explanatory	[31]	Web questionnaire, and exam score	Descriptive Statistics, Pearson correlations
	[21]	Discussion forums, and web-based questionnaire	Descriptive Statistics, Pearson correlations
	[11]	Emotions (boredom, frustration, and enjoyment), and self-regulated learning behavior	Descriptive Statistics, Pearson correlations, and Regression analysis
	[35]	Demographic variables (sex, age, marital status, family size, education, community location, and national origin)	Descriptive Statistics, ANOVA, and Multiple regression
	[26]	Absence of submission, and late submission,	Multiple regression analysis

Table 1. Cont.

Research Type	Authors	Features Focused	Methods Used
Predictive	[23]	Clickstream data (video views, academic records, grades, and surveys)	Descriptive Statistics
	[25]	Year of study, regular study, login sessions, proof of reading, Late submission, and midterm exam score	Descriptive Statistics, Pearson correlations, and Hierarchical regression analysis
	[32]	Work habits (timelines, regularity, and intensity)	Probit or Logistic regression
	[36]	Regularity features (studying hours of day/week, delay in lecture view), and time management	Linear regression modeling
	[24]	The behavior of students (time of completion of work, and total time spent working on a solution)	Descriptive Statistics, Mixed model ANCOVA
	[18]	Clickstream data and time management score	Probabilistic mixture model
	[37]	Problem-solving behavior, and level of problem-solving proficiency	X-means (Variation of K-means) clustering algorithm
	[38]	Survey scores of learning questionnaire	Hierarchical clustering with Ward's method
	[39]	Features related to effort and time spent working, such as time theory, time task, time forum, and relevant actions	Class Association Rule (CAR), a data mining technique

3. Relationship Exploration and Procedures

3.1. Participants and Context

The data used in this study collected from ASSISTments, an ITS (<https://new.assistments.org/>), developed by Worcester Polytechnic Institute (WPI), USA. It provides an e-learning platform and free feedback assistance to students and assessment data to educators. Students use it for completing their homework and classwork more effectively, and teachers have precise reports about the progress of students attempting their homework assignments with them before their class on the following day [40].

The number of registered students who use ASSISTments for their schoolwork and master the skill by using the instant feedback facility provided exceed 50,000. Students use it for various subjects such as Mathematics, Science, English, and Statistics. We focused on students from the Mathematics subject and classified the procrastinating students from the features extracted from the ASSISTments.

3.2. Data Collection and Measurement of Predictors

The data we have focused on in the current study were collected from 2009–2010 ASSISTments skill-builder data [41]. Initially, the dataset comprised of the total record of 401,756 Users, five answer types such as algebra, choose_1, choose_n, fill_in_1, and open_response, and 75 School IDs. Table 2 demonstrates the related predictor variables. Name of predictor shows the short name of the predictor with a self-explanatory description in the Description column, the number of instances means that how many times a student interacts with ITS, and values shows the minimum and maximum values each predictor has, which we selected in assisting and classifying the procrastinator among the students attempting algebra mathematics homework.

Table 2. Self-regulatory proposed features.

Name of Predictor	Description	No. of Instances	Values
STD_ID	Student ID	4094	5-digit number
ASSCOUNT	Assignment Count	406	1–1085
ORGMAINSFCF	Original Main/Scaffolding	578	0.33–1.00
CORRFSTATMPT	Correct 1st Attempt	1220	0.00–1.00
ATMPTCNTSUM	Attempt Count Sum	532	0.00–4011
FSTRSPTME	First Response Time	4045	1.18–2514.98
TOTSKILLTKN	Total Skill Taken	68	1–76
HNTCNTASKHLPPROB	Hint Count (Ask Help During Problem)	268	0–894
HNTTOTL	Hint Total (No. of Possible Hints)	1736	0.00–6.00
OVERLAPTIME	Overlap Time	4045	3.14–2523.42
NOANSGVN	No Answer Given	112	0–315
FSTACTSUM	First Action Sum	136	0–434
FSTACTATMPT	First Action Attempt	373	0–956
FSTACTHNT	First Action Hint	97	0–260
FSTACTSCAF	First Action Scaffolding	56	0–119
NOBTMHNT	No Bottom Hint	58	0–137
BTMHNT	Bottom Hint	112	0–274
BTMHNTSCAF	Bottom Hint Scaffolding	363	0–933
OPTNTY	Opportunity	2166	1.00–1316.80
OPTNTYORGMMAIN	Opportunity Original Main Problem	2065	1.00–1377.63
OPTNTYORGSCAF	Opportunity Original Scaffolding	83	0–278

3.3. Data Preprocessing and Selection of Predictors

In machine learning, data standardization, cleansing, proper structuring, and removing missingness are collectively called preprocessing. It should be well performed to become the backbone for machine learning analysis [42]. We used MS Excel 2016 and SAS Viya for Learners analytics software (https://www.sas.com/en_us/software/viya-for-learners.html). We availed the cutting-edge hi-tech SAS Visual Data Mining and Machine Learning (DMML) environment via Model Studio to prepare and transform the desired data for the experiment.

Finally, after removing the noise such as duplication, missingness, skewness, and replacing incorrect values from the study sample data, we are left with a total of 4094 unique records of students, selected answer type is algebra, and retrieved the data of 71 school's students.

Afterward, in order to assure the parsimony of the machine learning models, we used the SAS Visual DMML variable selection node to select the best features based on Statistical techniques to achieve high accuracy of learning models. Table 3 shows the most preferred predictors used in our study.

Table 3. Summary of the final selected predictors according to the combination criterion.

Name of Input Variable	Fast Supervised Selection	Linear Regression Selection	Input	Rejected	Output Role
CORRFSTATMPT	Rejected	Input	1	1	Input
HNTTOTL	Input	Input	2	0	Input
ASSCOUNT	Input	Rejected	1	1	Input
ATMTCNTSUM	Rejected	Rejected	0	2	Rejected
BTMHNT	Input	Rejected	1	1	Input
BTMHNTSCAF	Input	Rejected	1	1	Input
FSTACTATMPT	Rejected	Rejected	0	2	Rejected
FSTACTHNT	Input	Input	2	0	Input
NOANSGVN	Input	Input	2	0	Input
NOBTMHNT	Rejected	Rejected	0	2	Rejected
OPTNTY	Rejected	Rejected	0	2	Rejected
OPTNTYORGMMAIN	Rejected	Rejected	0	2	Rejected
OPTNTYORGSCAF	Input	Input	2	0	Input
ORGMAINSFCF	Rejected	Rejected	0	2	Rejected
OVERLAPTIME	Input	Input	2	0	Input
TOTSKILLTKN	Rejected	Rejected	0	2	Rejected

3.4. Feature Extraction and Label (Response or Target Variable) Collection

For selecting the predicted variable (Label), we used feature extraction by selecting variables from the initial predictors shown in Table 2. We combined them to form a new predicted variable, "Procrastination Rating," which discloses and predicts the procrastinator students. The Label variable

has two levels “0” and “1”, where “0” reveals non-procrastination, and “1” shows procrastination, respectively. Following logic fulfills the condition for categorizing the procrastinator student.

If (OR (FSTRSPTME Total average (FSTRSPTME) AND (HNTCNTASKHLPPROB) Total average (HNTCNTASKHLPPROB) AND (FSTACTSCAF) Total average (FSTACTSCAF)) then (Procrastination) else (Non-procrastination).

3.5. Investigation of Predictors' Selection for Machine Learning Models

There are various statistical techniques and criteria available for feature selection, such as Backward Elimination, Forward Selection, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Deviance Information Criterion (DIC), Bayes Factor, and Mallows' CP [40,43]. In this research, we have used the SAS Visual DMML built-in variable selection node, which has plenty of statistical and machine learning criteria available for the robust selection. Varying combinations of existing criteria such as (Unsupervised Selection, Fast Supervised Selection, Linear Regression Selection, Decision Tree Selection, Forest Selection, and Gradient Boosting Selection) are accessible [42]. We kept the combination criteria at “Selected by at least 1,” which assures that any input selected by at least one of the selection criteria chosen is passed on to subsequent nodes as inputs. Table 3 delineates the predictors verified by two criteria (Fast Supervised, and Linear Regression Selection), which has two values (i.e., Rejected and Input), if both criteria have rejected the predictor then the output will be rejected, if both have input value then the output role will be accepted, and if either of the value is rejected or input then the final output will be accepted due to the “Selected by at least 1” property in SAS Visual DMML.

Whereas, in Appendix A, Tables A1–A3 precisely explain the proportion of variance using the “VARREDUCE” procedure in SAS Visual DMML on SAS Viya. It performs both supervised and unsupervised feature selection using criteria, such as AIC, AICC (the version of AIC), and BIC. In unsupervised learning, the procedure performs feature selection by classifying a set of features that mutually explain the maximum amount of variability in the data—opposed to Principal Component Analysis (PCA), which uses feature extraction (generates new features by reducing the dimensionality). The VARREDUCE reduces dimensionality by the forward selection of the features that explain the overall data variance the best and use the subset of the original features. This technique is valuable where retaining the original features is essential, and for model interpretation and exploration [44].

On the other hand, supervised feature selection also identifies a set of features that jointly describe the maximum amount of variance contained in the response (predicted) variable. The procedure supports feature selection in both the regression setting and the classification (categorization) setting. It also reduces the dimensionality by the forward selection, and the outcome lists the features in order of their contribution to explaining the response or target variance. This output can be used directly to build models utilizing these features. Moreover, it analyses the variance in the same way as Linear Discriminant Analysis (LDA) does. For selecting the features from linear regression selection, SAS Viya used the REGSELECT procedure to highlight the best features for becoming the ingredient of machine learning models. It has numerous features, such as supporting multiple parameterizations for classification variables, care hierarchy among variables, assistance to the partition of a data into training, validation, and testing. Testing provides multiple selection methods, delivers variable selection based on a variety of selection criterion, and provides stopping rules based on model evaluation criteria. Finally, it produces output that contains predicted values, residuals, and confidence limits [44].

Tables 4 and 5 show the regression output and parameter estimates of selected effects (variables) through the linear regression selection criterion. Stepwise selection stops when adding or removing an effect does not improve the Schwarz Bayesian Criterion (SBC). The model at step 6 is selected where SBC is -5713.15351 .

Table 4. Linear regression selection summary (The REGSELECT Procedure).

Root MSE:	0.24189	AICC:	−3721.38932		
R-square:	0.67385	SBC:	−5713.15351		
Adj R-square:	0.67288	ASE (Train):	0.05831		
AIC:	−3721.46060	ASE (Validate):	0.07253		
Analysis of Variance (ANOVA)					
Source	DF	Sum of Squares	Mean Square	F Value	Pr F
Model	6	244.42801	40.73800	696.26	0.0001
Error	2022	118.30634	0.05851		
Corrected Total	2028	362.73435			

MSE: Mean Squared Error, AIC: Akaike Information Criterion, AICC (Version of AIC), SBC: Schwarz Bayesian Criterion, ASE: Average Squared Error, DF: Degrees of Freedom.

Table 5. Linear regression selection parameter estimates of variables (The REGSELECT Procedure).

Parameter	DF	Estimate	Standard Error	t Value	Pr > t	Variance Inflation
Intercept	1	−0.189439	0.037716	−5.02	0.0001	0
fstacthnt	1	−0.493223	0.017289	−28.53	0.0001	13.70908
noansgvn	1	0.655997	0.016473	39.82	0.0001	14.65309
optntyorgscf	1	0.055506	0.005913	9.39	0.0001	1.44312
overlaptime	1	0.029193	0.007344	3.97	0.0001	1.04713
corrstatmpt	1	0.091733	0.023307	3.94	0.0001	1.38060
Hnttotl	1	−0.015950	0.005403	−2.95	0.0032	1.16500

DF: Degrees of Freedom.

3.6. Relationship Prediction

Adequately determining the predictors and label (response variable), we articulate the relationship prediction task as a logistic regression analysis as a base model. We use the collected label to experiment on the effects of using ensemble models. In the end, we choose the champion predictive model to classify procrastinator students.

4. Experiment and Results

4.1. Building Ensemble Machine Learning Models

Machine learning algorithms operate on statistical principles, but before applying machine learning methods to data, it is better to perform statistical tests and models. In this study, we have used Logistic Regression (Logit Reg) as a base model and then employed ensemble machine learning models, i.e., Decision Tree (DT), Gradient Boosting (GB), and Forest and its variants, i.e., regularization, and tuning of hyper-parameters settings. We then performed models' comparisons with the base model as well as within the ensemble models to build and identify an ideal model for predicting the procrastination among the students using the ITS e-learning platform. Our study focuses on supervised learning techniques as a response or predicted variable output is known. SAS Viya is the latest high-tech software for the comprehensive set of built-in environments, modules, and functionalities for computational, data mining, and machine learning. So, we used SAS Visual DMML to execute the described machine learning models and tested them.

The following are the reasons for choosing the tree ensemble models [42]:

- Supervised learning classifiers;
- best for Mid-size to large datasets;
- provides moderate interpretability;
- allows autotuning of parameters;
- handles the interval, binary, and nominal target;

- useful in modeling nonlinear and nonlinear separable phenomena in large datasets;
- interactions considered automatically but implicitly;
- capable of handling the missing values and outliers in input variables automatically;
- tree ensembles can increase prediction accuracy and decrease overfitting, but also decrease scalability and interpretability.

4.2. Training and Evaluation of Ensemble Machine Learning Models

After building a project in the SAS Viya model studio, we set the data partitioning ratio with a stratified method into (70:30) and then trained the learning classifiers. We effectively taught machine learning algorithms to curtail the disparity between real and expected values before predicting the procrastinator students attempting and acquiring mastery skills communicating with ITS. The flow chart of the methods we used in our research to predict procrastination among the students is also shown in Figure 1.

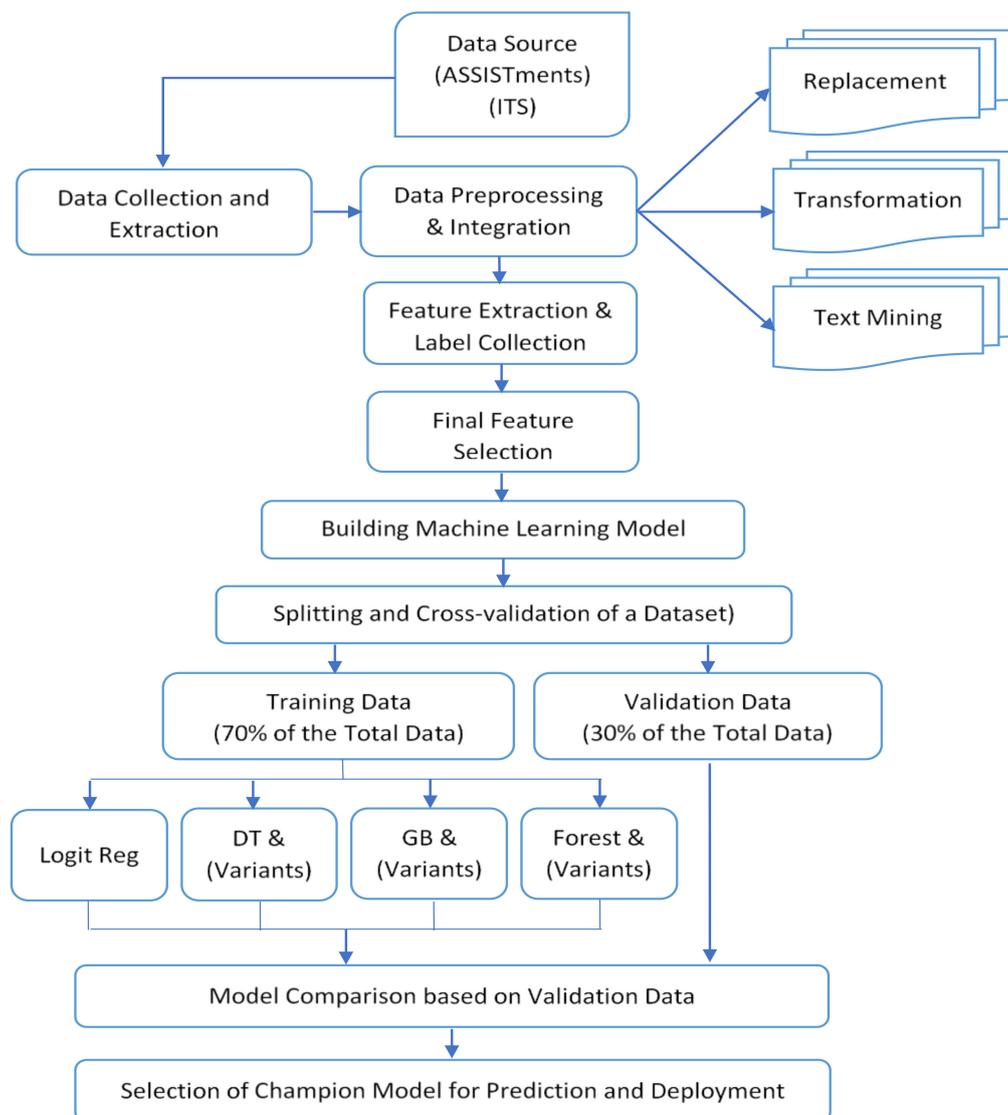


Figure 1. Flow chart of the proposed methodology.

An integral aspect of applying machine learning strategies is model evaluation or assessment. When a machine is trained on known data, then we evaluate the model on the unseen data to verify

that the model is sufficiently good, learned, and correctly classified. We also selected the best and champion model based on validation data using the Kolmogorov–Smirnov (KS) statistic as selection criteria class and Averaged Squared Error (ASE) as selection criteria interval properties.

4.3. Results and Analysis

4.3.1. Logistic Regression (Logit Reg) Model

As our predicted variable is dichotomous, thus we started with Logit Reg as a base model, as it is commonly used for nominal and binary targets. It also deals with multicollinearity and overfitting issues through regularization term and suitable for small to large datasets with high interpretability. Table 6 is the output of the Logit Reg model, which shows the selection of predictors stops when it reaches the optimal level, i.e., the local minimum, which is 798.6346 in this case. Thus, the first six variables in Table 6 has selected by this model, and Table 7 shows the parameter estimates of these selected six variables. Finally, Table 8 shows the Fit statistics summary of the logit model, which shows the details of the train and validate data set, as we have split the data into the train (70%) and validate (30%), also shown in Figure 1. Hence, based on KS value, Logit Reg is considered to be a good model.

Table 6. Logit Reg model summary.

Selection Criteria:	SBC	No. of Observations Read:	4095
Selection Method:	Forward	No. of Observations Used:	4094
Distribution:	Binary	No. of Observations Used for Training:	2866
Link Function:	Logit	No. of Observations Used for Validation:	1228
Step	Variable Entered	Number Variables In	SBC
0	Intercept	1	3134.4765
1	Noansgvn	2	1783.3284
2	Fstacthnt	3	1175.9832
3	Overlaptime	4	827.7871
4	Btmhntscf	5	801.5044
5	Optntyorgscf	6	798.6346 *
6	Btmhnt	7	802.4826
7	Asscount	8	807.5466
8	Hnttotl	9	814.7302

*: Optimal value of criterion (Selection stopped at a local minimum of the Schwarz Bayesian Criterion (SBC)).

Table 7. Parameter estimates of the Logit Reg model.

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr ChiSq
Intercept	1	−13.066089	0.956536	186.5902	0.0001
Btmhntscf	1	0.274213	0.084899	10.4321	0.0012
Fstacthnt	1	−7.985433	0.492071	263.3553	0.0001
Noansgvn	1	10.368209	0.598322	300.2878	0.0001
Optntyorgscf	1	0.289069	0.094710	9.3156	0.0023
Overlaptime	1	1.159052	0.167420	47.9281	0.0001

Table 8. Fit statistics summary of the Logit Reg model.

Data Role	Misclassification Rate	Average Squared Error	KS (Youden)	KS Cut-Off	Area Under ROC
Train	0.0499	0.0382	0.8678	0.25	0.9795
Validate	0.0529	0.0400	0.8642	0.40	0.9790

ROC: Receiver Operating Characteristic

4.3.2. Tree Ensembles (DT Model and its Variants)

DTs were typically developed to make decisions for a categorical target and provide estimates such as probabilities for categorical targets and numeric predictions for interval targets [42]. We have applied five variants of DT in our study using default settings, modified, recursive partitioning, pruning parameters, and autotuned, shown in Table 9. In the default settings, we kept all properties at their default and ran the SAS DT node. Then we modified the tree structure parameters, recursive partitioning parameters, and pruning parameters, and also built a DT model using Autotuned and compared the performance with trees already in the pipeline. Table 10 discloses that the DT Autotuned tree model is considered a robust model among all variants of DT based on ASE decreased (0.0628), and KS value (0.8158) increased on the validation data. The validation data set is referred to as a separate portion of the same data set from which the training set is derived. Validation data accuracy assures and endorses that the model is robust and useful for making predictions on the unseen data.

Table 9. List of tuning hyperparameters settings of the Decision Tree (DT) model.

Property	DT Model				
	Default Settings	Structure Parameters	Recursive Partitioning	Pruning Parameters	Autotuned
	Default Value	Modified Value	Modified Value	Modified Value	Modified Value
Grow Criterion (Class target criterion)	Information gain Ratio	Information gain ratio	Gini	Gini	Entropy, CHAID, Information gain ratio, Gini, Chi-square
Interval target criterion	Variance	Variance	Variance	Variance	Variance, F test, CHAID
Maximum no. of branches	2	2	2	2	2
Maximum depth	10	14	14	14	1–19
Minimum leaf size	5	15	15	15	5
Number of interval bins	20	100	100	100	20–200
Interval bin method	Quantile Cost	Quantile Cost	Quantile Cost	Quantile	Quantile
Subtree method	complexity	complexity	complexity	Reduced Error	Cost Complexity

Table 10. Fit statistics of the DT model and its variants.

Data Role	Model Name	Misclassification Rate	Average Squared Error	KS (Youden)	KS- Cutoff	Area Under ROC
Validate	DT (Default)	0.0774	0.0658	0.8118	0.20	0.9278
	DT (Structure Parameters)	0.0863	0.0701	0.7894	0.10	0.9306
	DT (Recursive Partitioning)	0.0757	0.0659	0.8022	0.25	0.9356
	DT (Pruning Parameters)	0.0733	0.0652	0.7988	0.35	0.9358
	DT (Autotuned)	0.0700	0.0628	0.8158	0.10	0.9394

ROC: Receiver Operating Characteristic.

4.3.3. Tree Ensembles (GB Model and Its Variants)

Further, we tested our experiment by using GB and its variants, such as default settings, modified the parameters, autotuned settings shown in Table 11, and observed the performance by noticing ASE and KS value comparing within the variants of GB. Thus, GB Autotuned is considered as being the best model shown in Table 12. GB is an improvement in boosting that can be applied to any target. This algorithm is related to boosting, apart from that at each iteration; the target is the residual from the previous DT model [45]. The base of GB is DT in SAS Visual DMML. SAS Visual DMML creates a series of trees, which form a single model. A tree in the series fits the residuals of the prediction from the previous trees in the series. Each time that the data are used to grow a tree, the correctness of the tree is computed, and the following samples adjust prior inaccuracies. Each succeeding sample is weighted per the accuracy of the previous models [42].

Table 11. Regularizing and hyperparameters settings of the Gradient Boosting (GB) model.

Property	GB Model		
	Default Settings	Modified Parameters	Autotuned
	Default Value	Modified Value	Modified Value
Number of trees	100	50	20–150
Learning rate	0.1	0.1	0.01–1
Subsample rate	0.5	0.5	0.1–1
L1 regularization	0	0	0–10
L2 regularization	1	1	0–10
Tree-splitting Options (Max Depth)	4	8	4
Minimum leaf size	5	15	5
Number of interval bins	50	100	50
Interval bin method	Quantile	Quantile	Quantile

Table 12. Fit statistics of the GB model and its variants.

Data Role	Model Name	Misclassification Rate	Average Squared Error	KS (Youden)	KS- Cutoff	Area Under ROC
Validate	GB (Default)	0.0464	0.0336	0.9074	0.20	0.9873
	GB (Modified Parameters)	0.0480	0.0344	0.9010	0.30	0.9900
	GB (Autotuned)	0.0383	0.0279	0.9177	0.15	0.9854

ROC: Receiver Operating Characteristic.

4.3.4. Tree Ensembles (Forest Model and Its Variants)

A Forest is an ensemble of DTs, and each one can predict its response to a set of input variables. The results from the individual trees are combined to provide the final prediction. For a categorical target, the Forest model's prediction is either the most popular class (as determined by a vote) or the average of the posterior probabilities of the individual trees [42]. Table 13 precisely explains the modified hyper-parameterized list. Moreover, in this experiment, we have observed that there is a slight decrease in the Forest autotuned model, and the modified Forest model has a higher value of KS (0.8781) compared to other Forest models in the pipeline shown in Table 14. Although tuning the hyperparameter gives a significant influence on the accuracy of the predictive model. However, there is no guarantee that the tuned hyperparameter or SAS autotuned option will arrive at the best model.

Table 13. Hyperparameter values of Forest model.

Property	Forest Model		
	Default Settings	Modified Parameters	Autotuned
	Default Value	Modified Value	Modified Value
Number of trees	100	50	20–150
Class target voting method	Probability	Probability	Probability
Tree-splitting Options (Class target criterion)	Information gain ratio	Entropy	Information gain ratio
Interval target criterion	Variance	Variance	Variance
Maximum Depth	20	12	1–29
Minimum leaf size	5	15	5
Number of interval bins	20	100	20
In-bag sample proportion	0.6	0.6	0.1–0.9
Number of inputs to consider per split	100	7	1–100
Interval bin method	Quantile	Quantile	Quantile

Table 14. Fit statistics of Forest model and its variants.

Data Role	Model Name	Misclassification Rate	Average Squared Error	KS (Youden)	KS- Cutoff	Area Under ROC
Validate	Forest (Default)	0.0611	0.0490	0.8432	0.30	0.9803
	Forest (Modified Parameters)	0.0619	0.0450	0.8781	0.30	0.9829
	Forest (Autotuned)	0.0619	0.0426	0.8733	0.20	0.9834

ROC: Receiver Operating Characteristic.

4.3.5. Models' Comparison

Finally, we performed ensemble machine learning and compared learning models within the tree-based pipeline through model comparison node in SAS Visual DMML, the pictorial view shown in Figure 2, and used the property settings shown in Table 15. We also included the autotuned models in our comparison.

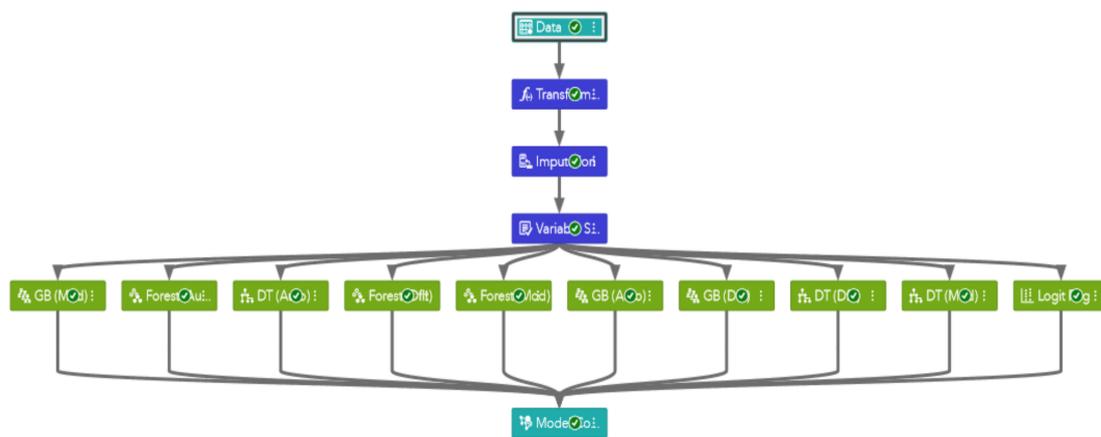


Figure 2. SAS Viya flow chart of machine learning models with a model comparison.

Table 15. SAS Visual DMML default assessment measures.

Property Name	Property Value
selectionCriteriaClass	Kolmogorov-Smirnov statistic (KS)
selectionCriteriaInterval	Average squared error
selectionTable	Validate
selectionDepth	10
Cutoff	0.5

Furthermore, Figures 3 and 4 cumulatively display train and validate data of each learning classifier, which shows the trend of how the validate data outperforms over the trained model. Thus, Tables 16 and 17 present only the output according to the validation data because model validation is carried out after model training, which authenticates the robustness of the model and determines that the models are not overfitting the data. Thus, we presented the validation data in our tables to select the best predictive model. The Lift reports plot based on Cumulative % Response, % Captured Response, Cumulative % Captured Response, Cumulative Lift, Gain, and Lift is shown in Table 16. Moreover, the cumulative lift plot based on the % Response shown in Figure 3, the Receiver Operating Characteristic (ROC) plot is based on accuracy shown in Figure 4 of all the ensemble models, and Table 17 shows the ROC Reports plot based on the F1 Score and ROC.

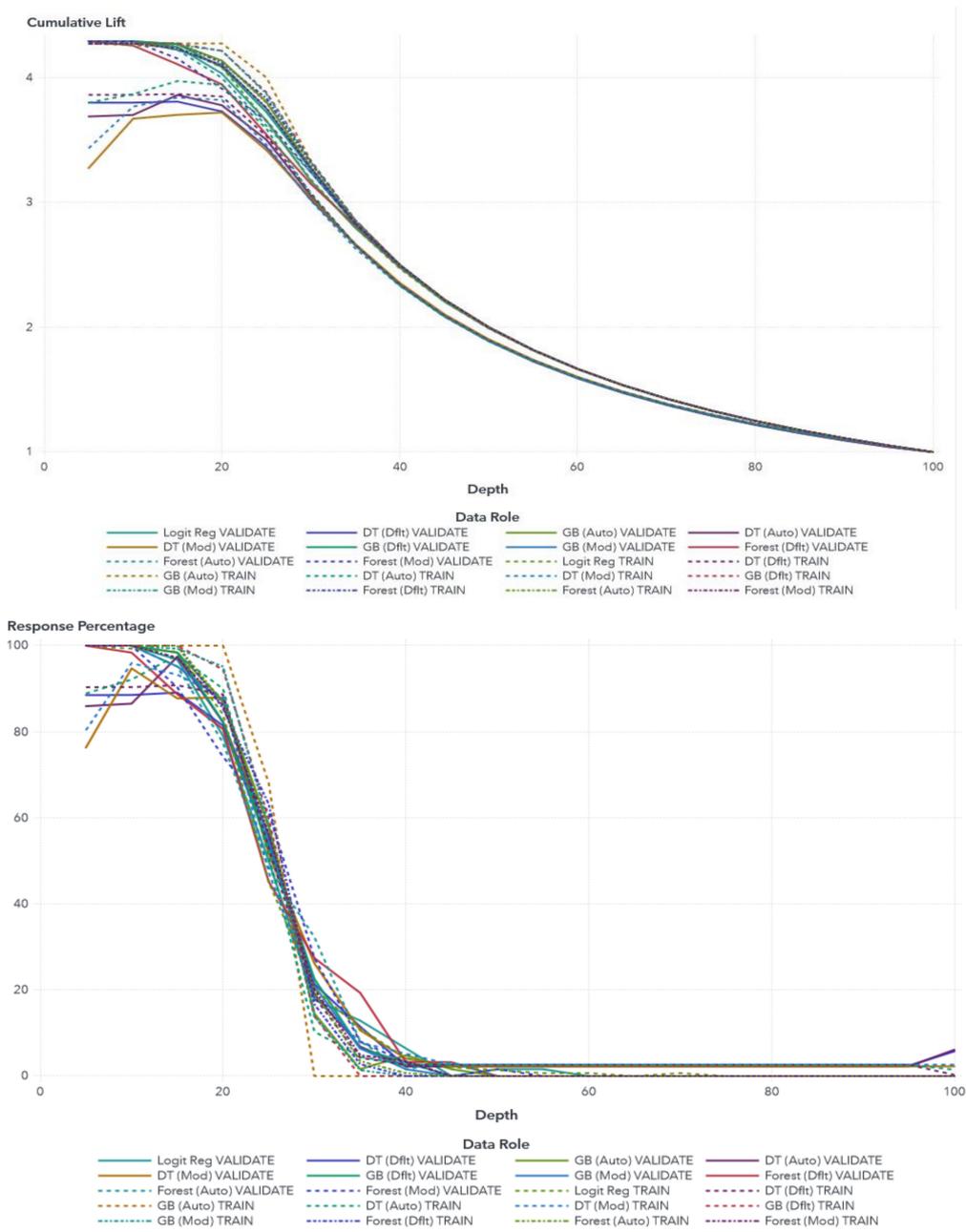


Figure 3. Visual understanding of cumulative lift and response percentage.

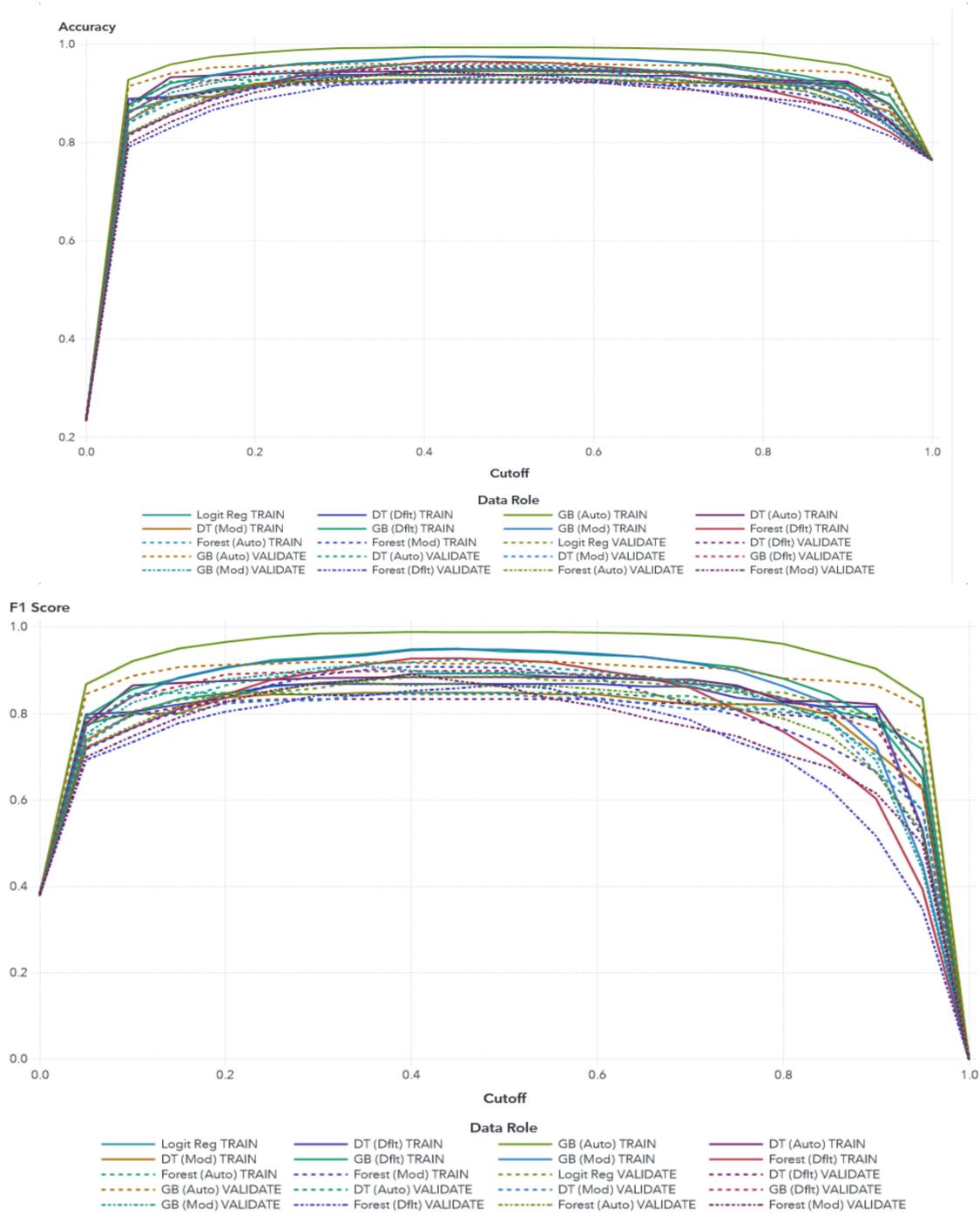


Figure 4. Pictorial view of Accuracy, F1 score, and ROC.

Table 16. Lift reports plot summary.

Data Role	Model Name	% Captured Response	Cumulative % Captured Response	Cumulative Lift	Gain	Lift
Validate	GB (Auto)	21.4533	42.9066	4.2907	3.2907	4.2907
	GB (Dflt)	21.4533	42.9066	4.2907	3.2907	4.2907
	GB (Mod)	21.4533	42.9066	4.2907	3.2907	4.2907
	Forest (Mod)	21.4533	42.9066	4.2907	3.2907	4.2907
	Forest (Auto)	21.4533	42.9066	4.2907	3.2907	4.2907
	Logit Reg	21.4533	42.9066	4.2907	3.2907	4.2907
	Forest (Dflt)	21.1073	42.5606	4.2561	3.2561	4.2215
	DT (Auto)	18.5654	37.0046	3.7005	2.7005	3.7131
	DT (Dflt)	18.9954	37.9868	3.7987	2.7987	3.7991
	DT (Mod)	20.3242	36.6964	3.6696	2.6696	4.0648

Auto: Autotuned, Dflt: Default, Mod: Modified.

Table 17. ROC reports plot summary.

Data Role	Model Name	Accuracy	F1 Score	ROC Separation	Area Under ROC
Validate	GB (Auto)	0.9617	0.9180	0.8877	0.9854
	GB (Dflt)	0.9536	0.8995	0.8579	0.9873
	GB (Mod)	0.9520	0.8952	0.8485	0.9900
	Forest (Mod)	0.9381	0.8628	0.7993	0.9829
	Forest (Auto)	0.9381	0.8618	0.7945	0.9834
	Logit Reg	0.9471	0.8870	0.8493	0.9790
	Forest (Dflt)	0.9389	0.8673	0.8147	0.9803
	DT (Auto)	0.9300	0.8481	0.7910	0.9394
	DT (Dflt)	0.9226	0.8336	0.7767	0.9278
	DT (Mod)	0.9267	0.8448	0.7988	0.9358

Auto: Autotuned, Dflt: Default, Mod: Modified, ROC: Receiver Operating Characteristic.

Table 18 succinctly explains how each model in the pipeline performs on the data partitions defined in the settings (TRAIN, VALIDATE, and TEST) for a series of fit statistics used in this study, and highlight the champion model based on increased KS value and decreased ASE on the VALIDATION data. The autotuned model is the proprietary of SAS Visual DMML; it offers an automated option for hyperparameter optimization. This elegant feature prevents time-wasting of selecting hyperparameters manually (for example, based on experience, hit, trial and error, and heuristics). Generally speaking, autotuning searches for the best combination of hyperparameters specific to each modeling algorithm. Thus, autotuning in SAS saves lots of time and provides a base and starting place to explore more and improve the learning models. We also tested our experiment without using the autotuning model, and the results are shown in Table 19.

Table 18. Fit statistics of ensemble machine learning models with Autotuned.

Data Role	Model Name	Sum of Frequencies	Misclassification at Cutoff	Root Average Squared Error	Gini Coefficient	False Positive Rate	Misclassification Rate	False Discovery Rate	Average Squared Error	KS (Youden)
Validate	GB (Auto) *	1228	0.0383	0.1672	0.9709	0.0224	0.0383	0.0739	0.0279	0.9177
	GB (Dflt)	1228	0.0464	0.1832	0.9745	0.0245	0.0464	0.0827	0.0336	0.9074
	GB (Mod)	1228	0.0480	0.1854	0.9801	0.234	0.0480	0.0803	0.0344	0.9010
	Forest (Mod)	1228	0.0619	0.2120	0.9659	0.0277	0.0619	0.0981	0.0450	0.8781
	Forest (Auto)	1228	0.0619	0.2063	0.9668	0.0256	0.0619	0.0920	0.0426	0.8733
	Logit Reg	1228	0.0529	0.2000	0.9580	0.0330	0.0529	0.1084	0.0400	0.8642
	Forest (Dflt)	1228	0.0611	0.2214	0.9607	0.0330	0.0611	0.1123	0.0490	0.8432
	DT (Auto)	1228	0.0700	0.2506	0.8788	0.0394	0.0700	0.1336	0.0628	0.8158
	DT (Dflt)	1228	0.0774	0.2564	0.8556	0.0469	0.0774	0.1560	0.0658	0.8118
	DT (Mod)	1228	0.0733	0.2554	0.8717	0.0490	0.0733	0.1581	0.0652	0.7988

*: Champion Model, Auto: Autotuned, Dflt: Default, Mod: Modified.

Table 19. Fit statistics of ensemble machine learning models without Autotuned.

Data Role	Model Name	Sum of Frequencies	Misclassification at Cutoff	Root Average Squared Error	Gini Coefficient	False Positive Rate	Misclassification Rate	False Discovery Rate	Average Squared Error	KS (Youden)
Validate	GB (Dflt) *	1228	0.0464	0.1832	0.9745	0.0245	0.0464	0.0827	0.0336	0.9074
	GB (Mod)	1228	0.0480	0.1854	0.9801	0.0234	0.0480	0.0803	0.0344	0.9010
	Forest (Mod)	1228	0.0619	0.2120	0.9659	0.0277	0.0619	0.0981	0.0450	0.8781
	Logit Reg	1228	0.0529	0.2000	0.9580	0.0330	0.0529	0.1084	0.0400	0.8642
	Forest (Dflt)	1228	0.0611	0.2214	0.9607	0.330	0.0611	0.1123	0.0490	0.8432
	DT (Dflt)	1228	0.0774	0.2564	0.8556	0.0469	0.0774	0.1560	0.0658	0.8118
	DT (Mod)	1228	0.0733	0.2554	0.8717	0.0490	0.0733	0.1581	0.0652	0.7988

*: Champion Model, Dflt: Default, Mod: Modified.

5. Limitations of the Study

Like many studies, the present research has some deficiencies related to theory and methodology, which may reduce its scope. The first shortcoming is that we have focused on the limited number of features, as there are so many attributes available (such as postponement of a task, level of interaction and participation, and time spent viewing practice videos), which could be taken into consideration. The second limitation is that by taking the autotuned model, which involves building models based on the selected hyperparameter values. Nevertheless, there is no surety that the autotuned model will arrive at the best model. However, it provides the right starting place to use further domain knowledge and expertise in improving the model.

Furthermore, in this research, we did not track the mental and physical health problems among the students that may be one of the reasons for procrastination of the assigned work. Nevertheless, it is a complete study, since in case of learning and attempting to solve the assigned task through a web-based tutoring system, the past medical record can also be considered as one of the predictors to evaluate and quantify the procrastinator students. Accumulation of these features in the data may enhance the correctness and accuracy of the prediction.

6. Discussion, Conclusions, and Perspectives

As far as the ESD perspective is concerned, the method used in this study assists teachers to pinpoint which student has procrastination behavior or has a risk of being a procrastinator. Teachers can provide attention to particular skills and counseling to those students in class the next day to avoid procrastination. Thus, the early prediction of procrastinating behavior could aid in preserving educational sustainability, as sustainable green education prioritizes the present needs rather than future needs. Online education using Massive Open Online Courses (MOOCs), and Intelligent Tutoring System (ITS) all devote to sustainable educational development. Advancement is based on the fulfillment of present-day prerequisites without negotiating future desires. Moreover, the prime objective of ESD is to balance the environmental, economic, and societal demands [40,46].

The evolution of ESD in higher education could be seen in numerous dimensions: (1) Sustainability in policy, planning, administration, and control, (2) courses and curricula, (3) research, (4) campus operation, (5) evaluation and reporting [15,47]. Several universities are already actively motivated to incorporate ESD into their educational activities. Such programs are directed towards: (i) Positive student learning outcomes; (ii) curricula and evaluation methods; (iii) elimination of barriers; (iv) changing teaching paradigms; (v) improvement of social skills; (vi) communication skills and community relations; and (viii) expansion of their participation in local and national projects (see this section for several examples of such projects) [48]. “One reason behind the latest initiatives is the UN Decade on Sustainable Development Education (DESD, 2005e2014),” undertaken by UNESCO, which aims to combine concepts, values, and practices for sustainability [49].

Sustainability competencies are, therefore, linked to the acquisition of knowledge, skills, and attitudes that allow excellent task performance and problem-solving concerning real-world sustainability issues, challenges, and opportunities [50–52]. ESD must, therefore, translate these competencies into an educational perspective, so that they can contribute fully to SD and sustainability [52–55].

According to the ‘The Brundtland Commission Report’, sustainable development (SD) in education is an integrative approach that covers the connected environmental, societal, and economic facet of the formal and informal educational curriculum. Thus, this educational approach can help students to grow their tendency, knowledge, know-how, and practical understanding to show a compelling role in eco-friendly SD for education and become a liable representative of society. Further, sharing teaching and learning approaches are also required to boost and strengthen learners to revamp their performances and take remedial actions for a sustainable educational environment. Analytical and lateral thinking, visualizing the future, and the respective decision making are the capabilities that ESD promotes [56].

Although limited in scope, the present research can lead to many meaningful educational, methodological, and practical inferences. Firstly, to the best of our knowledge and information, our research, amongst the previous studies for investigating procrastination and its behavior by using machine learning techniques for educational sustainability or ESD, is one of the rare studies that has focused on these perspectives.

6.1. Conclusions

This study introduces a data-driven methodology for predicting student procrastination in web-based online homework using an Intelligent Tutoring System (i.e., ASSISTments). The study identifies some unique patterns inside the big data, as well as makes some essential findings for academia. We suggest that (1) by attempting homework activity on ITS, students undertake various actions that could be further employed to measure procrastination. (2) Early prediction of student procrastination can help the development of an environmentally friendly and sustainable educational context while fulfilling social responsibility. By using machine learning models, we have quantified the procrastinator students. We have used the Kolmogorov–Smirnov (KS) statistic property as our selection criteria class and recorded the KS value.

Additionally, we considered the other measurement metric, such as Average Squared Error (ASE), along with KS statistics. The result suggests that based on the KS and ASE values (i.e., 91.77%, and 2.79%), respectively, GB autotuned model is considered to be the best and most ideal model for deployment that significantly outperformed as compared to the other classifiers in the pipeline due to the increasing KS and decreasing ASE values; the higher value of KS is considered excellent. We have also tested the ensemble classifiers without using the autotuned model, and we found that the GB default model is considered best to classify procrastinator students. It gives KS value 90.74% and ASE of 3.66%, which is slightly higher than the GB modified classifier.

6.2. Future Perspectives

We will focus future research on comprehensive health problems such as (smoking addiction, Influenza, congenital cataract, emotions (frustration and boredom), and behavior) and determining the well-being of students who engage in the online tutoring system to complete their homework or assignments. Poor health in teenagers and adolescents may be one of the reasons for procrastinating the assigned job. In Vietnam [57], a cross-sectional study gathering the participants from ten groups from high schools and universities also revealed the potential adverse impacts of risky online interactions and the harmful effects of Shisha smoking.

Author Contributions: Conceptualization, S.M.R.A. and S.A.H.; methodology, S.M.R.A.; software, S.M.R.A. and S.A.H.; validation, S.M.R.A., S.A.H., R.R., S.S.R. and S.J.K.; formal analysis, S.M.R.A.; investigation, S.S.R.; resources, S.S.R., S.J.K. and W.Z.; writing—Original draft preparation, S.M.R.A. and S.A.H.; writing—Review and editing, R.R., S.S.R. and S.J.K.; visualization, S.M.R.A. and S.A.H.; supervision, W.Z., S.S.R. and S.J.K.; funding acquisition, S.S.R., S.J.K., H.D. and W.Z. All authors have read and agreed to the published version of the manuscript.

Funding: The effort of this research supported by the “Basic Science Research through the National Research Foundation of Republic of Korea (NRF) funded by the Ministry of Education under grant NRF-2017R1D1A3B04031440,” “National Natural Science Foundation of China, grant number 61873156,” “The National Key RD Program of China, grant number 2017YFB0701501,” and “Program of Shanghai Municipal Education Commission (No. 2019-01-07-00-09-E00018).”

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

Appendix A

Table A1. Unsupervised feature selection summary (The VARREDUCE Procedure).

Number of Observations Read: 2867 (70% Training Data) Number of Observations Used: 2867 (70% Training Data)						
Parameter	Proportion of Variance Explained	SSE	MSE	AIC	AICC	BIC
Asscount	0.465641	8.549750	0.002983	2.240077	17.240639	2.148679
Fstacthnt	0.589013	6.575798	0.002295	1.976176	15.976729	1.888950
optntyorgmain	0.672003	5.247954	0.001832	1.748525	14.749066	1.666169
orgmainscf	0.751586	3.974631	0.001388	1.467829	13.468353	1.391039
overlaptime	0.816863	2.930195	0.001024	1.159478	12.159981	1.088953
Hnttotl	0.874999	2.000018	0.00069906	0.773380	10.773858	0.709817
corrstatmpt	0.919391	1.289736	0.00045096	0.329778	9.330227	0.273875
Nobtmhnt	0.947037	0.847406	0.00029640	-0.095816	7.904600	-0.143361
Optntyorgscf	0.966952	0.528763	0.00018501	-0.573735	6.426644	-0.612225
Btmhnt	0.977305	0.363121	0.00012710	-0.956513	5.043824	-0.985250
totskilltkn	0.986126	0.221983	0.00007773	-1.456323	3.543968	-1.474610
atmptcntsum	0.991433	0.137075	0.00004801	-1.946765	2.053476	-1.953904
noansgvn	0.995633	0.069864	0.00002448	-2.629809	0.370378	-2.625102
fstactatmpt	0.998328	0.026746	9.374797×10^{-6}	-3.599734	-1.599605	-3.582484
btmhntscf	0.999378	0.009952	3.489423×10^{-6}	-4.598837	-3.598770	-4.568347
Optnty	1.000000	0	0	-lnfty	-lnfty	-lnfty

SSE: Sum of Squared Error, MSE: Mean Squared Error.

Table A2. Fast supervised feature selection summary (The VARREDUCE Procedure).

Number of Observations Read: 2867 (70% Training Data) Number of Observations Used: 2866 (70% Training Data)						
Parameter	Proportion of Variance Explained	SSE	MSE	AIC	AICC	BIC
Noansgvn	0.474218	0.525782	0.00018352	-0.639379	1.360625	-0.640091
Fstacthnt	0.627816	0.372184	0.00012995	-0.983483	1.016525	-0.982813
optntyorgscf	0.647997	0.352003	0.00012295	-1.037835	0.962178	-1.035782
btmhntscf	0.650790	0.349210	0.00012202	-1.044406	0.955612	-1.040972
Asscount	0.658107	0.341893	0.00011950	-1.064184	0.935841	-1.059368
Btmhnt	0.659982	0.340018	0.00011889	-1.068289	0.931744	-1.062091
overlaptime	0.661472	0.338528	0.00011841	-1.071285	0.928757	-1.063705
Hnttotl	0.662102	0.337898	0.00011823	-1.071753	0.928298	-1.062791

SSE: Sum of Squared Error, MSE: Mean Squared Error.

Table A3. Linear regression feature selection detail (The REGSELECT Procedure).

Selection Criteria:	SBC	No. of Observations Read:	2867
Selection Method:	Stepwise	No. of Observations Used:	2866
Number of Effects:	17	No. of Observations Used for Training:	2029
Number of Parameters:	17	No. of Observations Used for Validation:	837
Step	Variable Entered	Number Variables In	SBC
0	Intercept	1	-3485.5671
1	noansgvn	2	-4824.7821
2	Fstacthnt	3	-5589.5762
3	optntyorgscf	4	-5703.2245
4	corrstatmpt	5	-5706.3205
5	overlaptime	6	-5712.0423
6	Hnttotl	7	-5713.1535 *

*: Optimal value of the criterion, SBC: Schwarz Bayesian criterion.

References

- Lozano, R. Incorporation and institutionalization of SD into universities: Breaking through barriers to change. *J. Clean. Prod.* **2006**, *14*, 787–796. [[CrossRef](#)]
- Læssøe, J.; Schnack, K.; Breiting, S.; Rolls, S. Climate change and sustainable development: The response from education. *Int. Alliance Lead. Educ. Inst.* **2009**, *33*, 257–258. [[CrossRef](#)]
- Wals, A.E.J. Review of Contexts and Structures for Education for Sustainable Development 2009. *Unesco* **2009**. [[CrossRef](#)]
- Azeiteiro, U.M.; Bacelar-Nicolau, P.; Caetano, F.J.P.; Caeiro, S. Education for sustainable development through e-learning in higher education: Experiences from Portugal. *J. Clean. Prod.* **2015**, *106*, 308–319. [[CrossRef](#)]
- Nousheen, A.; Yousuf Zai, S.A.; Waseem, M.; Khan, S.A. Education for sustainable development (ESD): Effects of sustainability education on pre-service teachers' attitude towards sustainable development (SD). *J. Clean. Prod.* **2020**, *250*, 119537. [[CrossRef](#)]
- Longhurst, J.; Bellingham, L.; Cotton, D.; Isaac, V.; Kemp, S.; Martin, S.; Peters, C.; Robertson, A.; Ryan, A.; Taylor, C.; et al. *Education for Sustainable Development: Guidance for UK Higher Education Providers*; Centre for Environmental Science: New Delhi, India, 2014.
- Merritt, E.; Hale, A.; Archambault, L. Changes in Pre-Service Teachers' Values, Sense of Agency, Motivation and Consumption Practices: A Case Study of an Education for Sustainability Course. *Sustainability* **2019**, *11*, 155. [[CrossRef](#)]
- Bettinger, E.P.; Fox, L.; Loeb, S.; Taylor, E.S. Virtual classrooms: How online college courses affect student success. *Am. Econ. Rev.* **2017**, *107*, 2855–2875. [[CrossRef](#)]
- Allen, I.E.; Seaman, J. *Changing Course: Ten Years of Tracking Online Education in the United States*; ERIC: Kern County, CA, USA, 2013; Volume 26, ISBN 978-0-9840-2883-2.
- Broadbent, J.; Poon, W.L. Self-regulated learning strategies and academic achievement in online higher education learning environments: A systematic review. *Internet High. Educ.* **2015**, *27*, 1–13. [[CrossRef](#)]
- Artino, A.R.; Jones, K.D. Exploring the complex relations between achievement emotions and self-regulated learning behaviors in online learning. *Internet High. Educ.* **2012**, *15*, 170–175. [[CrossRef](#)]
- Lee, J.K.; Lee, W.K. The relationship of e-Learner's self-regulatory efficacy and perception of e-Learning environmental quality. *Comput. Hum. Behav.* **2008**, *24*, 32–47. [[CrossRef](#)]
- Garrison, R. Theoretical challenges for distance education in the 21st century: A shift from structural to transactional issues. *Int. Rev. Res. Open Distance Learn.* **2000**, *1*, 6–21. [[CrossRef](#)]
- Narciss, S.; Proske, A.; Koerndle, H. Promoting self-regulated learning in web-based learning environments. *Comput. Hum. Behav.* **2007**, *23*, 1126–1144. [[CrossRef](#)]
- Lozano, R.; Lukman, R.; Lozano, F.J.; Huisingh, D.; Lambrechts, W. Declarations for sustainability in higher education: Becoming better leaders, through addressing the university system. *J. Clean. Prod.* **2013**, *48*, 10–19. [[CrossRef](#)]
- Ku, D.T.; Chang, C.S. The effect of academic discipline and gender difference on Taiwanese college students' learning styles and strategies in web-based learning environments. *Turk. Online J. Educ. Technol.* **2011**, *10*, 265–272.
- Zimmerman, B.J. Self-Regulated Learning and Academic Achievement: An Overview. *Educ. Psychol.* **1990**, *25*, 3–17. [[CrossRef](#)]
- Park, J.; Yu, R.; Rodriguez, F.; Baker, R.; Smyth, P.; Warschauer, M. Understanding Student Procrastination via Mixture Models. In Proceedings of the 11th International Conference on Educational Data Mining, Buffalo, NY, USA, 15–18 July 2018; pp. 187–197.
- Zervakis, P.; Wahlers, M. "Education for Sustainable Development" and the Bologna Process: The Implementation of the Bologna Process in Germany. *BNE J.* **2007**, *5*, 1–5.
- Lozano, R.; Young, W. Assessing sustainability in university curricula: Exploring the influence of student numbers and course credits. *J. Clean. Prod.* **2013**, *49*, 134–141. [[CrossRef](#)]
- Michinov, N.; Brunot, S.; Le Bohec, O.; Juhel, J.; Delaval, M. Procrastination, participation, and performance in online learning environments. *Comput. Educ.* **2011**, *56*, 243–252. [[CrossRef](#)]
- Schraw, G.; Wadkins, T.; Olafson, L. Doing the things we do: A grounded theory of academic procrastination. *J. Educ. Psychol.* **2007**, *99*, 12–25. [[CrossRef](#)]
- Hotle, S.L. Applications of Clickstream Information in Estimating Online User Behavior. Ph.D. Thesis, Georgia Institute of Technology, Atlanta, GA, USA, May 2015.

24. Kazerouni, A.M.; Edwards, S.H.; Shaffer, C.O.A. Quantifying incremental development practices and their relationship to procrastination. In Proceedings of the 2017 ACM Conference on International Computing Education Research, Tacoma, WA, USA, 18–20 August 2017; pp. 191–199. [CrossRef]
25. You, J.W. Identifying significant indicators using LMS data to predict course achievement in online learning. *Internet High. Educ.* **2016**, *29*, 23–30. [CrossRef]
26. You, J.W. Examining the effect of academic procrastination on achievement using LMS data in e-Learning. *Educ. Technol. Soc.* **2015**, *18*, 64–74.
27. Gašević, D.; Dawson, S.; Siemens, G. Let's not forget: Learning analytics are about learning. *Technol. Trends* **2015**, *59*, 64–71. [CrossRef]
28. Dietz-Uhler, B.; Hurn, J.E. Using learning analytics to predict (and improve) student success: A faculty perspective. *J. Interact. Online Learn.* **2013**, *12*, 17–26.
29. Klingsieck, K.B.; Fries, S.; Horz, C.; Hofer, M. Procrastination in a distance university setting. *Distance Educ.* **2012**, *33*, 295–310. [CrossRef]
30. Jo, I.-H.; Kim, Y. Impact of Learner's Time Management Strategies on Achievement in an e-learning Environment: A Learning Analytics Approach. *J. Educ. Inf. Media* **2013**, *19*, 83–107.
31. Elvers, G.C.; Polzella, D.J.; Graetz, K. Procrastination in Online Courses: Performance and Attitudinal Differences. *Teach. Psychol.* **2003**, *30*, 159–162. [CrossRef]
32. Dvorak, T.; Jia, M. Online Work Habits and Academic Performance. *J. Learn. Anal.* **2016**, *3*, 318–330. [CrossRef]
33. Chen, X.; Carroll, C.D. First-Generation Students in Postsecondary Education: A Look at Their College Transcripts: Postsecondary Education Descriptive Analysis Report. *Natl. Cent. Educ. Stat.* **2005**, 1–103.
34. Styck, K.M. Best practices for supporting upward economic and social mobility for first-generation college students. *Sch. Psychol.* **2018**, *72*, 50–57.
35. Steel, P.; Ferrari, J. Sex, Education and Procrastination: An Epidemiological Study of Procrastinators' Characteristics from a Global Sample. *Eur. J. Pers.* **2013**, *27*, 51–58. [CrossRef]
36. Boroujeni, M.S.; Sharma, K.; Kidziński, Ł.; Lucignano, L.; Dillenbourg, P. How to quantify student's regularity? *Lect. Notes Comput. Sci. (Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinform.)* **2016**, *9891 LNCS*, 277–291. [CrossRef]
37. Tóth, K.; Greiff, S.; Kalergi, C.; Wüstenberg, S. Discovering Students' Complex Problem Solving Strategies in Educational Assessment. In Proceedings of the 7th International Conference on Educational Data Mining, London, UK, 4–7 July 2014; pp. 225–228.
38. Ng, B.L.L.; Liu, W.C.; Wang, J.C.K. Student Motivation and Learning in Mathematics and Science: A Cluster Analysis. *Int. J. Sci. Math. Educ.* **2016**, *14*, 1359–1376. [CrossRef]
39. Cerezo, R.; Esteban, M.; Sánchez-Santillán, M.; Núñez, J.C. Procrastinating behavior in computer-based learning environments to predict performance: A case study in Moodle. *Front. Psychol.* **2017**, *8*, 1–11. [CrossRef] [PubMed]
40. Abidi, S.M.R.; Hussain, M.; Xu, Y.; Zhang, W. Prediction of Confusion Attempting Algebra Homework in an Intelligent Tutoring System through Machine Learning Techniques for Educational Sustainable Development. *Sustainability* **2018**. [CrossRef]
41. Heffernan, N. ASSIST Ments Data. Available online: <https://sites.google.com/site/assistmentsdata/home/assistance-2009-2010-data/skill-builder-data-2009-2010> (accessed on 2 February 2018).
42. Jeff, T.; Truxillo, C. Machine Learning Using SAS Viya. Available online: <https://www.coursera.org/learn/machine-learning-sas/> (accessed on 2 July 2019).
43. Rundel, M.C. Linear Regression and Modeling. Available online: <https://www.coursera.org/learn/linear-regression-model> (accessed on 18 June 2018).
44. SAS Institute. SAS Documentation. Available online: <https://documentation.sas.com/?docsetId=procdocsetTarget=n0n5mm9l2pmpevn1lsjqccmgp8tx.htm&docsetVersion=9.4&locale=en> (accessed on 1 August 2019).
45. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
46. Higher Education Academy. Education for Sustainable Development. Available online: <https://www.heacademy.ac.uk/knowledge-hub/education-sustainable-development-0> (accessed on 16 November 2018).
47. Saadatian, O.; Salleh, E.I.; Tahir, O.M.; Dola, K. Observations of sustainability practices in Malaysian research universities: Highlighting particular strengths. *Pertanika J. Soc. Sci. Humanit.* **2012**, *17*, 293–312.

48. Karatzoglou, B. An in-depth literature review of the evolving roles and contributions of universities to Education for Sustainable Development. *J. Clean. Prod.* **2013**, *49*, 44–53. [[CrossRef](#)]
49. Segalàs, J.; Ferrer-Balas, D.; Svanström, M.; Lundqvist, U.; Mulder, K.F. What has to be learnt for sustainability? A comparison of bachelor engineering education competences at three European universities. *Sustain. Sci.* **2009**, *4*, 17–27. [[CrossRef](#)]
50. Dale, A.; Newman, L. Sustainable development, education and literacy. *Int. J. Sustain. High. Educ.* **2005**, *6*, 351–362. [[CrossRef](#)]
51. Rowe, D. Education for a Sustainable Future. *Science* **2007**, *317*, 323–324. [[CrossRef](#)] [[PubMed](#)]
52. Barth, M.; Godemann, J.; Rieckmann, M.; Stoltenberg, U.; Barth, M.; Godemann, J.; Rieckmann, M. Developing key competencies for sustainable development in higher education. *Int. J. Sustain. High. Educ.* **2007**, *8*, 416–430. [[CrossRef](#)]
53. Mochizuki, Y.; Fadeeva, Z. Competences for sustainable development and sustainability: Significance and challenges for ESD. *Int. J. Sustain. High. Educ.* **2010**, *11*, 391–403. [[CrossRef](#)]
54. Parker, J. Competencies for interdisciplinarity in higher education. *Int. J. Sustain. High. Educ.* **2010**, *11*, 325–338. [[CrossRef](#)]
55. Wals, A.E.J. Mirroring, Gestaltswitching and transformative social learning Stepping stones for developing sustainability competence. *Int. J. Sustain. High. Educ.* **2010**, *11*, 389–390. [[CrossRef](#)]
56. Ortolano, L. The Brundtland Commission Report. Available online: <https://www.sustainabledevelopment2015.org/AdvocacyToolkit/index.php/earth-summit-history/past-earth-summits/58-the-brundtland-commission> (accessed on 7 June 2019).
57. Tran, B.X.; Nguyen, L.H.; Vu, G.T.; Le, H.T.; Nguyen, H.D.; Hoang, V.Q.; La, P.V.; Hoang, D.A.; Van Dam, N.; Vuong, T.T.; et al. Online peer influences are associated with receptiveness of youths: The case of Shisha in Vietnam. *Child. Youth Serv. Rev.* **2019**, *99*, 18–22. [[CrossRef](#)]



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).