American Journal of Educational Research, 2022, Vol. 10, No. 4, 223-232
Available online at http://pubs.sciepub.com/education/10/4/10
Published by Science and Education Publishing
DOI:10.12691/education-10-4-10

A Gaussian-Bernoulli Mixed Naïve Bayes Approach to Predict Students' Academic Procrastination Tendencies in Online Mathematics Learning

Charles Darwin O. Godinez[1,*], Laila S. Lomibao[2]

[1]Basic Education Department, St. Rita's College of Balingasag, Misamis Oriental, Philippines
[2]College of Science and Technology Education, University of Science and Technology of Southern Philippines, Cagayan De Oro City, Philippines
*Corresponding author: charlesdarwingodinez@gmail.com

Received March 05, 2022; Revised April 08, 2022; Accepted April 15, 2022

Abstract The abrupt shift of learning venue from face-to-face to online distance learning actually instigated many pedagogical problems such as students' tendencies to procrastinate on tasks and assignments required as learners take on independent and active role. Delayed and unsubmitted tasks then result to poor academic performance. In this paper, a method for the early detection of academic procrastination tendencies in online mathematics learning was explored using educational data mining methods. Thirty-five (35) input features were retrieved from a repository of student data from St. Rita's College of Balingasag, Philippines. K-modes clustering was used for output label clustering, the Boruta Algorithm for feature selection, and a Gaussian-Bernoulli mixed Naïve Bayes algorithm for model building, all implemented in RStudio. Focus Group Discussions were also performed to triangulate results of the model as well as suggest actions for mathematics' educators based on students' experiences. The results revealed that only 14 features out of the 35 identified features are relevant in predicting students' procrastination tendencies and our Gaussian-Bernoulli Mixed Naïve Bayes model can successfully predict students' later procrastination behaviors – whether they are Non Procrastinators, Low, Moderate, or High Procrastinators – with a testing accuracy of 89% and Kappa Score of 84%. Mathematics' educators can utilize this model to forecast and take actions to prevent the adverse effects of students' procrastination.

Keywords: *academic procrastination, online learning, educational data mining*

Cite This Article: Charles Darwin O. Godinez, and Laila S. Lomibao, "A Gaussian-Bernoulli Mixed Naïve Bayes Approach to Predict Students' Academic Procrastination Tendencies in Online Mathematics Learning." *American Journal of Educational Research*, vol. 10, no. 4 (2022): 223-232. doi: 10.12691/education-10-4-10.

1. Introduction

The dramatic development in information and communications technology resulted to the advancement in the different aspects of teaching and learning in the 21st century. Various innovations specifically in the learning environment have been made to accommodate the needs of digitally-inclined learners. Online learning, a form of distance learning, is catalyzing a pedagogical shift in delivering education. However, studies have shown that learners in an online learning environment are more prone to delaying or not submitting tasks at all [1]. As an example, all schools in the Philippines suspended face-to-face teaching [2] and started to use internet platforms to deliver online learning as an urgent response to the coronavirus pandemic. This abrupt shift of learning venue incurred challenges to students that adversely impacted their completion of tasks, motivation to continue studying, and overall academic performance [3].

Procrastination has become more prevalent in the conduct of online distance learning. Based on initial investigations and interviews, 63% of students in St. Rita's College of Balingasag with online learning as their modality option had skipped submitting some or all of the tasks required of them in any subject, specifically in Mathematics. Many researchers had already concluded that procrastination is more frequent when the learning process is not face-to-face, i.e. distance or computer-based learning, as the student has to take an active role [4,5]. Moreover, procrastination behaviors can lead to poor academic results especially in online learning environments [1,6].

Since procrastination has long been identified to have a negative correlation with students' academic performance, this research will delve on predicting students' academic procrastination tendencies in the online teaching of Mathematics. To do such, Educational Data Mining (EDM) methods were utilized.

Educational Data Mining is an emerging discipline that connects Data Mining and its applications to Education.

Many predictive and classification models have been made with EDM techniques. However, most of these predictive models focus on predicting features with students' overall performance and grades. The researchers argue that instead of having students' grades as the output, it is much better to specifically predict students' procrastination tendencies so teachers can be more guided to doing intervention activities, as variables affecting grades are many. Moreover, EDM predictive models are more focused in higher education learning and not much in the basic education where student failure is much more accounted to the teacher. Further, this research will use the current mode of online learning to predict students' procrastination tendencies specifically to the teaching of Mathematics to provide this gap in the extant literature.

A Mixed Gaussian-Bernoulli Naïve Bayes algorithm was used in this research with features or the predictor variables including students' Math Anxiety Score, Online Learning Readiness Score, Overall Attitudes toward Online Learning Score, Time Management Score, Availability of Gadgets, Mode of Access to Internet, Challenges to Online Learning, Modality, and their Math grades in the previous year. The researchers hope that using the built model, it can be useful to mathematics teachers in helping them better the quality of their instruction in the conduct of online learning in mathematics, and by considering the suggested intervention activities to students identified by the model as high procrastinators or moderate procrastinators.

To prevent the adverse effects of academic procrastination, a novel method for the early detection of such academic procrastination tendencies would be of great significance. Specifically, this study seeks to answer the following questions:

1. Which of the identified features/variables contributed more to the prediction of students' academic procrastination?
2. How accurate was our model in predicting students' academic procrastination tendencies?
3. What activities can be suggested based on students' online mathematics learning experience to prevent possible negative procrastination behaviors?

1.1. Academic Procrastination in Online Learning

Lay [7] defined procrastination as a recurrent failure at doing what are supposed to be done to reach goals. When someone fails to accomplish what he or she is ought to do to reach specific desirable goals, procrastination as a behavior becomes clear. Asikhia [8] seconded as he also defined procrastination as a behavioral problem that involves delaying tasks that are needed to be accomplished.

Schraw [9] defined procrastination as "*intentionally delaying or deferring work that must be completed*". It is a negative action that produces negative result to the persons. On the other hand, Chu [10] expressed that procrastinators can be categorized into two categories: active and passive. Active procrastinators intentionally delay tasks or procrastinate as a sort of '*a positive academic strategy*'. For active procrastinators, they work best under pressure. Passive procrastinators, on the other hand, allow their negative, indecisive conduct to handicap them, resulting in unfavorable scholastic repercussions. Steel [11] also mentioned that procrastination can be a healthy activity, and that some researchers even refer to it as a *"functional delay."* However, Steel found in his meta-analysis of procrastination research that the 'positive' procrastination reference is only secondary to the conventional negative connotation.

Hooshyar [12] expressed that a number of studies have found a link between procrastination and poor academic performance, with negative outcomes including lower course accomplishment, poorer long-term learning, low goal-setting, low grades, and many others. Further, the team suggested that educators must be made aware of the prevalence of these procrastination behaviors since students with lower procrastination tendencies and practices usually achieve and perform better in comparison to those students who have high procrastination tendencies. Therefore, in this study, the term 'procrastination' refers to its original, inactive, and negative meaning.

One of the essential criteria for improving online learning, Park [13] claimed, is being able to recognize students' learning habits linked to time management, particularly procrastination. They studied students' time management attitudes, particularly those who enrolled in online learning courses, and discovered a strong association between students' cultural background and various levels of procrastination on an individual basis. Procrastination in online courses was also found to be a predictor for students' overall performance of the course.

Tuckman [14] concluded that when compared to traditional physical classroom-based settings, academic procrastination has a bigger impact in online learning settings. This type of misbehavior has also been associated with increased student dropout rates in online learning environments compared to traditional learning environments. Therefore, given the current mode of learning due to the coronavirus pandemic, an exploration of academic procrastination in online learning is vital. Cerezo [15] supports this with their conclusion that Procrastination has been shown to have detrimental consequences on learning and performance in traditional educational settings, but there is a paucity of study in online learning contexts.

An Educational Data Mining study was conducted by Akram [16] that utilized the homework submission data of students to make a prediction of therie academic procrastination in a blended learning course. Although many different types of academic and behavioral indicators, such as delaying or not submitting academic tasks, self-reports, assessment scales, and questionnaires, can be used to detect procrastination, the team noted that in online learning environments, it is much easier to get data about students task submissions [17], and such data may be utilized to efficiently assess students' procrastinating habits. Students' homework submission data were used to create a feature vector of their homework submission habits in their research. In their model, this feature vector was utilized to classify each student as a procrastinator or non-procrastinator.

Hooshyar [12] also proposed an approach to predict students' procrastination behaviors using students homework submission data. However, they included a spare time — the period of time between when a student submits an assignment and when it is due — and an inactive time — the time it takes for a student to view an assignment for the first time after it is made available — for each student to explain their procrastination behaviors. This submission data with spare and inactive time are then used to predict students' procrastination behaviors.

Following these two data mining models on procrastination, the researchers will also use students' submission data, whether they submitted on-time, delayed, or no submission. Other evidence of students' participation like attendance in synchronous sessions and entries in discussion forums were not considered since not all are required to join those activities especially to students who have limited internet connection for a video-conferencing tool.

1.3. Features Used in This Study

Due to conflicting study findings, gender variations in procrastination frequency have become a popular issue [18]. In the Studies of Dempsey [19] and Balkis & Duru [18], gender was found to be a major predictor of procrastination, with men having a greater procrastination rate. Mandap [20] on the other hand showed that males have lower perceived academic self-efficacy than females and it has significantly higher procrastination rates than those with high self-efficacy. While in a study of Ozer [21] male and females' academic procrastination scores have showed no significant difference.

Steel [11] did a meta-analysis based on 16 studies and it was found out that a strong negative association is observed between age and their procrastination habits, after correcting for range restrictions In addition, Ozer [21] agreed with the same results that significant difference among the grade levels of the students is shown in their procrastination. Undergraduate students, in particular, reported to procrastinate more than graduate or high school students. High school and college students reported to procrastinate on studying for examinations almost often or always, but graduate students postpone on completing term papers more. Rosario [22] in their study concluded that the students' academic procrastination level increases as the student advances through his or her educational process. Thus, higher grade-level students are more prone to academic procrastination than students from the lower-grade levels.

There are very few studies in the existing literature that correlated students' address to their procrastination. In this study, the place where students' reside was considered a likely factor since internet connectivity is limited to some areas in the country especially those living in the mountainside, and based on the researchers' experience as well in the conduct of online learning in the previous academic year. In a study by Casillano [23] internet connectivity issues is included as one of the challenges in implementing online learning platforms especially in an internet-struggling province in the Philippines even before the COVID-19 pandemic. Because only a significant fraction of pupils have access to the internet, students were unable to use the platform.

Rosario [22] in their study revealed that students' procrastination increases along with their number of siblings. They hypothesized that having more siblings is linked to a higher procrastination rate because these young people are exposed to more distracting and appealing non-academic tasks than studying, with their brothers and sisters acting as either distracters interrupting their study time or role models of academic low-commitment. They also pointed out that the fact that procrastination is positively connected to school grade level (and hence age) shows that older siblings may be emulating their younger siblings' procrastination habits. The researchers also considered the number of siblings as a likely factor for academic procrastination since there are concerns regarding gadget availability. The more siblings are currently studying at the same time, the less they have access to their gadgets especially if it is limited only.

The topic on the antagonistic effects of Mathematics anxiety to students' performance is already well-researched. Çağırgan [24] investigated whether students' math anxiety and their engagement in mathematics predict their responsibilities toward learning. It was then concluded that students' social, emotional, and cognitive engagement in Mathematics are positively correlated with students' responsibility towards learning whereas students' math anxiety was a significant negative predictor of students' responsibility towards learning. Thus, learning as a responsibility of the student can be affected if one has a high mathematics anxiety.

A study of Mendoza [25] said that there is an increase of the levels of Math anxiety observed especially during the COVID-19 pandemic. They came to the conclusion that detecting and analyzing virtual mathematical anxiety is critical for identifying potential educational failures in online learning. They went on to say that online education's failings have a direct influence not just on students' performance and learning, but also on their emotional well-being. Thus, students' mathematics anxiety was considered as a predictor variable in this study.

Reyes and Gonzales [26] studied the online readiness of some Filipino college students and correlated it to their engagement in online distance learning. The students used mobile data packages and had poor to mediocre internet access, according to the findings of their study. These were then regarded as contributing reasons in why students have difficulty in joining online classes. A laptop computer is required to fully participate in online learning. Furthermore, students' home learning environments were found to be unsuitable, and they concluded that traditional face-to-face classroom learning is more successful than online learning. Online readiness was also shown to very important as in the study of Çiğdem & Yıldırım [27]. Among the factors that determine the efficiency of online learning, preparedness stands out. It's crucial to look at online learning readiness and the qualities of students who influence it.

Many researchers [28,29,30] emphasized the significance of studying and reporting the students' attitudes toward online learning by educational institutions. In the study of Forsyth [31], the positive attitude towards online learning of the students are shown to be critical to their readiness and their inclusion in the distance form of learning. In the same study, it was evident that learners who have a clearly stated positive attitude towards online distance learning, which implied that these students are more likely to accept it well as a mode of education [31].

A number of studies have already concluded the correlations of procrastination tendencies to their time management [15]. Visser [32], for instance, stated that there is a relationship between the fear of failure and time management. Students who have fear of failure tend not to procrastinate and manage their time more effectively. Studies of Duru [18] and You [6] also highlighted the impact of time management on the academic achievement of learners. Self-regulated learning, perceived academic control, and academic self-efficacy was explored to be correlated as well with time management.

In the study of Baticulon [33] about the hurdles to online learning that were grouped into five categories during COVID-19 pandemic online learning, from a nationwide survey of medical students in the Philippines. The five categories were technological, individual, home, institutional, and community barriers. Concerns about a shortage of gadgets or limited access owing to gadget sharing, as well as unreliable, poor, or no internet connectivity, were recognized significant technical impediments.

Fabito [34] explored the different insights from one private university in the Philippines about the impediments and challenges of computing students in their online learning environments. They concluded that there are three major challenges and barriers facing by the students in their classes, one of which is the lack of a stable internet connectivity for their participation to classes and other online activities. As such, it should be that both the faculty as well as the students are fully ready to undergo full online learning. The study of Baticulon [33] seconded their result as medical students consider internet connectivity a barrier under the technological category.

Tus [35] conducted a research on the significance of the role of parental involvement in private school students' performance in online learning. The statistical analysis of the study found a substantial link between parental participation and kids' academic performance, and that students who received parental involvement did well in class through online learning modes. In this study, instructional support from students' parents/guardians, grandparents, siblings, other members of the family, and tutor/house helper are considered. Students may also have no instructional support or they'll specify that they can do independent learning.

In the study of Baticulon [33] about the barriers to online learning in the time of COVID-19 pandemic, a national survey of medical students in the Philippines, Technological, individual, personal, institutional, and communal impediments were all divided into five categories. Domestic impediments included a lack of study space, the necessity to fulfil tasks at home, family disputes, and financial hardship. Also, A study of Fabito [34] also agreed to their conclusion as students' lack of study space or working area to attending to their online learning activities is considered among the three major challenge that computing students face in the conduct of online learning.

2. Methods

This research made use of educational data mining (EDM) methods to discover likely hidden significant patterns from various collections of data. In summary, this study clustered students' procrastination behaviours using K-modes clustering from their tasks submission data; identified highly influential predictive variables in the procrastination of a student using Boruta Algorithm; trained the Mixed Gaussian-Bernoulli Naïve Bayes model, with 10-fold cross validation; and, performed a confusion matrix analysis to test the model's performance accuracy.

In building the predictive model, the data collected in this study was analyzed, pre-processed and synthesized to identify implicit relationships implemented in RStudio, an integrated development environment for R Programming using R. R is a programming language especially for statistical computing and graphics. Package 'Boruta' [36] was used for feature selection, package 'fastNaiveBayes' [37] is used for Naïve Bayes model training implementation, and package 'caret' [38] is used for model validation. A methodological framework in Figure 1 visually shows the flow of the processes used for building the model. Detailed discussion of the methodological framework is presented in the research procedure.

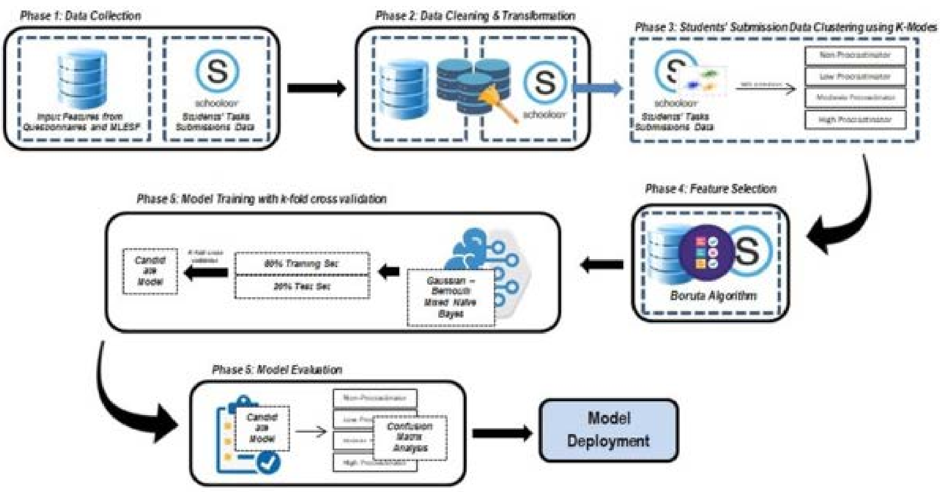

Figure 1. Methodological Framework for Model Building

In addition, Focus-Group Discussions (FGDs) were carried out after students have been classified with their procrastination tendencies to triangulate results of the model as well as plan suggestible programs or activities that Mathematics Educators or School Administrators can do to prevent procrastination to the students predicted by the model, based on students' experiences. Five students under each category of model-identified *Non Procrastinators*, *Low Procrastinators*, *Moderate Procrastinators*, and *High* Procrastinators were randomly purposively selected to a Focus-Group Discussion. The FGDs was made possible using an online video-conferencing tool. With this, an Interpretative Phenomenological Analysis (IPA) was utilized. IPA approach was strongly recommended to multiple focus groups with smaller numbers of participants [39]. Therefore, superordinate themes and themes will need to be amalgamated across all focus groups, once the analysis for each focus group has been completed.

Phase 1: Sampling and Data Collection

There were two types of data that the researchers had collected for this study – the input data and the output data. Since there was no available, pre-existing dataset related to procrastination and to the features set in this study, data collection to both the students' procrastination behaviors (output data) and to the features (input data) were needed.

The features selected in this study were data that can be easily gathered even in the first week of school. Data on students' grade level, gender, age, address, learning modality, and grades in the previous school year are considered '*institutional data'* that can be easily retrieved from the school records, while features including number of family members currently studying, availability of gadgets, mode of access to internet, instructional support, and challenges to online learning was gathered from students'/parents' answers on the Modified Learner Enrolment and Survey Form (MLESF) asked to them upon enrolment. Since the mentioned features are personal data of the students, confidentiality and proper process of acquiring these data were safeguarded.

The second type of data collected in this study was the students' procrastination behaviors, or the output data. The data to be collected came from the tasks given to them by their Math subject in the whole 4 quarters. Data on students' submissions whether on-time, delayed, and no submission on the tasks were easily collected as the school uses a Learning Management System (LMS) where task submissions were easy to identify.

Phase 2: Data Cleaning and Transformation

The data pre-processing phase of the study includes cleaning and preparing the data. Data with missing values were removed so as to ensure an unbiased estimate. Collected data were then structured by converting the categorical data into numerical using categorical variable encoding.

Phase 3: Output Data Clustering

After gathering and cleaning the '*on-time*', '*delayed*', and '*no submission*' records in the pre-processing stage, data were then employed with a Machine Learning method called K-Modes Clustering to identify which among our students can be branded as Non-procrastinators, Low Procrastinators, Moderate Procrastinators, and High Procrastinators. Real-world clustering frequently involves categorical data. Thus, the conventional approach of converting categorical data into numerical ones does not necessarily produce meaningful results. For this phase, the clustering method used is the *k*-modes clustering. K-modes Clustering is an unsupervised data analysis technique used to classify the same data into a homogeneous group, it works specifically well on categorical data. Using K-modes cluster analysis segments a large set of data into subsets called clusters. Each cluster is a collection of data objects (students in this case) that are similar to one another are placed within the same cluster but are dissimilar to objects (students) in other clusters. The submissions data were separated into 4 clusters to identify students that were considered non-procrastinators, low procrastinators, moderate procrastinators and high procrastinators. After the input features were structured and the output data were identified, the two data were merged in one dataset ready for the feature selection phase.

Phase 4: Feature Selection

In the model-building process, the Naïve Bayes algorithm permits inclusion of redundant and noisy features that would cause poor predictive performance and increased computation. Thus, feature selection was considered to avoid such. Feature selection is the process of removing irrelevant features and is extremely useful in reducing the dimensionality of the data and improving the predictive accuracy [40]. To reduce the dimension of the feature space and improve the precision of the procrastination tendencies classification, the Boruta Algorithm was used in the process of feature selection in the current work. The Boruta algorithm is a wrapper built around random forest classification algorithm which iteratively attempts to capture all important features and removing irrelevant features to the classification problem [41]. Wrapper-based feature selection works well for Naïve Bayes models [42] and have been used for educational data [43] and specifically for LESF data [44].

Phase 5: Model Training and Validation

Naive Bayes classifier is a probabilistic classifier which means that given an input, it predicts the probability of such input being classified for all the classes. Gaussian Naive Bayes supports continuous valued features and models each as conforming to a Gaussian (normal) distribution. Bernoulli Naïve Bayes is used for binary data and it works on Bernoulli distribution_Since our input features comprises of both continuous and binary data, this study will make use of the two variants of the Naïve Bayes algorithm to handle mixed data. A study of Bagui [45] compared two methods in handling mixed data for Naïve Bayes model in a MapReduce environment. The first method is to discretize the data. The first method involves discretizing the continuous values and then building the NB model with the discrete values. This option requires an extra pre-processing step of discretizing the continuous values. This extra step could be time consuming and resource intensive. The second method is using the Gaussian distribution probability density function for continuous values. This model, referred to as the Mixed NB model, handled both continuous as well as discrete values. For discrete values, the Multinomial distribution was used for probability estimation, and for continuous values, the Gaussian distribution is used for probability estimation. This method does not require a pre-processing step [45].

Hence, in this study, the second method was followed but since the data were a mix of continuous and binary data, the Gaussian distribution probability density was used for probability estimation for continuous values, and the Bernoulli distribution was used for probability estimation for the binary data. Thus, a mixed Gaussian-Bernoulli Naïve Bayes Model was employed in this study. More than two thirds (80%) of the data was utilized as the training dataset and are split randomly. The remaining set (20%) was assigned as the test set and was later used for evaluating the model. In order to create a predictive model by means of a Data Mining method, the data need to be trained. Moreover, a 10 – fold cross validation was employed to achieve the best training accuracy. In a 10-fold cross validation, the test divides the dataset into 10 equal subsamples. One subsample was kept for validating the data, while the remaining 10–1 subsamples are used for training. This process was repeated until all subsamples have been used for validation. The model with the highest training accuracy in the 10-fold cross validation was then validated using confusion matrix analysis.

A Confusion Matrix Analysis (CMA) shows actual and predicted instances by classifiers. CMA is a table generally used in supervised learning to show the performance of an algorithm. Its columns denote the number of occurrences of a projected class, and its rows denote the number of occurrences of an exact class [46].

3. Results and Discussion

Clustering Students' Procrastination using their Task Submissions Data

Out of 290 students from the Grade 7 and Grade 8 level of St. Rita's College of Balingasag in Misamis Oriental, Philippines, 267 of them were able to have a complete, necessary data for our educational data mining model. Task submissions data retrieved from the Learning Management System of these 267 students were gathered, cleaned, and was implemented with a *K-modes Clustering* to identify and categorize the procrastination tendencies of these students. Table 1 below shows the distribution of students in terms of their procrastination tendencies.

Table 1. Clustering Results of Students' Procrastination Tendencies

	Non Procrastinators	Low Procrastinators	Moderate Procrastinators	High Procrastinators
Grade 7	57	24	25	18
Grade 8	53	22	20	48
Total	110	46	45	66

The K-modes Clustering, with a maximum iteration set to 10, categorized 110 Non Procrastinators, 46 Low Procrastinators, 45 Moderate Procrastinators and 66 High Procrastinators from the 267 total students. To further validate whether the results of the clustering was agreeable, the researchers reported these results to the students' respective mathematics teacher and adviser. The mathematics teachers and advisers all agree to the results of each student's categories or procrastination tendencies.

Feature Selection Using Boruta Algorithm

The data cleaning and transformation phase of this study yielded a dataset that was comprised of 35 variables or input features. To reduce redundant data or unnecessary predictor variables, the Boruta Algorithm was utilized as the Feature Selection method, to identify only the more important input or predictor variables and thus increase the overall accuracy of our model. Out of the 35 identified input features or predictor variables, the Boruta Algorithm only considered 14 important features and rejected all the other variables.

Thus, the predictor variables that the Boruta Algorithm considered to be more important and the only input features that was considered in the model training were: The students' (1) Age, (2) Online Learning Readiness, (3) Mathematics Anxiety, (4) Overall Attitudes Towards Online Learning, (5) Time Management, (6) Parent's/Guardian's Educational Attainment, (7) Desktop Computer or Laptop, (8) Broadband Connection, (9) Student is able to do Independent Learning, (10) Unstable Mobile or Internet Connection, (11) Difficulty in Independent Learning, (12) Distractions, (13) Mathemaatics Grades in the previous School Year, and their (14) Learning Modality. *Table 2* shows the list of all the features and the decisions to whether confirm or reject such features by the Boruta Algorithm.

Table 2. The result of the feature selection process using Boruta Algorithm

Features	NormHits	Decision
1. Sex	0.019230770	Rejected
2. Age	**0.750000000**	**Confirmed**
3. 4PS	0.000000000	Rejected
4. NFM	0.019230770	Rejected
5. Address	0.000000000	Rejected
6. ORQ	**1.000000000**	**Confirmed**
7. MASA	**1.000000000**	**Confirmed**
8. OATOL	**1.000000000**	**Confirmed**
9. TMQ	**1.000000000**	**Confirmed**
10. PGEA	**0.942307690**	**Confirmed**
11. TV	0.000000000	Rejected
12. CP	0.000000000	Rejected
13. Tablet	0.000000000	Rejected
14. Radio	0.134615380	Rejected
15. DCL	**0.961538460**	**Confirmed**
16. MD	0.038461540	Rejected
17. Broadband	**1.000000000**	**Confirmed**
18. CS	0.000000000	Rejected
19. OP	0.000000000	Rejected
20. PG	0.000000000	Rejected
21. Siblings	0.000000000	Rejected
22. GP	0.000000000	Rejected
23. OMF	0.000000000	Rejected
24. THH	0.000000000	Rejected
25. SIL	**1.000000000**	**Confirmed**
26. LAGE	0.076923080	Rejected
27. ILDA	0.000000000	Rejected
28. UMIC	**1.000000000**	**Confirmed**
29. EHC	0.000000000	Rejected
30. DIL	**0.711538460**	**Confirmed**
31. HEC	0.000000000	Rejected
32. CWOA	0.019230770	Rejected
33. Dist	**1.000000000**	**Confirmed**
34. Grades	**1.000000000**	**Confirmed**
35. Modality	**1.000000000**	**Confirmed**

As it can be seen in Table 2, the Boruta Algorithm indicated the answer on the importance of features in the dataset. In this case, out of 35 features, 21 of them are rejected and only 14 are confirmed to be important. No feature was marked as tentative.

With *NormHits* of 1.0, students' Modality, Grades, Distractions, Unstable Mobile/Internet Connection, Student Able to do Independent Learning, Broadband, Math Anxiety, Online Learning Readiness, Overall Attitudes towards Online Learning, and Time Management provided the highest impact to students' procrastination tendencies. Other confirmed important features also conferred considerable high impact with 0.96 *NormHits* to students' availability of desktop or Laptop Computer, 0.94 *NormHits* to their Parent/Guardian's Highest Educational Attainment, 0.75 *NormHits* to Age, and a 0.71 *NormHits* to their Difficulty in Independent Learning.

These results supported different claims as to the relevance of these features to students' procrastination. Age as a relevant feature supported Rosario [22] who concluded that the levels of academic procrastination increased as the students advanced throughout their educational process, our data also showed more High Procrastinators to Grade 8 level compared to the lower year level. Desktop/Laptop Computers, Broadband, Difficulty in Independent Learning, Distractions, Unstable Mobile/Internet Connection also supported the conclusions of Baticulon [33] the about technological, individual, domestic, institutional, and community barriers to online learning in the time of COVID-19.

Thus, the 14 identified predictor variables considered by the Boruta Algorithm as important were the only variables used in the model-building phase of the study.

Performance of the Gaussian-Bernoulli Mixed Naïve Bayes Model

Table 3 shows the different performance results of the different models generated in terms of Accuracy and Kappa Scores. There are 4 different models compared – Model A, uses all the input features; Model B, uses only the 14 input features declared by the Boruta Algorithm as relevant; Model C, has been implemented with a 10-fold cross validation aside from using only the 14 relevant input features, but with zero Laplace Smoothing; and Model D, has been implemented with a 10-fold cross validation aside from using only the 14 relevant input features, but with Laplace Smoothing set to one. All of these models are implemented still in RStudio environment.

Table 3. Performance Results of the Different Models

Model	Training Accuracy	Testing Accuracy	Kappa	Laplace Smoother
A	0.64	0.50	0.39	0
B	0.78	0.71	0.59	0
C	0.91	0.89	0.84	0
D	0.91	0.89	0.84	1

Table 3 shows the Training Accuracies, Testing Accuracies, and Kappa scores of the different models generated. Generally, training accuracy should be higher than the testing accuracy, if otherwise is observed, we can say that 'overfitting' occurred. The models, however, do not show any 'overfitting' results. The Kappa Score, also known as Cohen's kappa coefficient, is another metric used to show results for classification models.

For comparison purposes, the researchers included Model A, the model using all the identified input features which contains redundant and unnecessary input features. True to the claim, Model A has the least accuracies and Kappa score among the models, and as can be observed from Model B that uses only the 14 input features confirmed by the Boruta Algorithm, there was a significant increase in their accuracies. Laplace Smoothing is sometimes made in a Naïve Bayes model to help tackle the problem of zero probability, which ensures that the posterior probabilities are never zero.

Among the models, Model C is our best model; it uses only the 14 important features and was implemented with a 10-fold cross validation. Model D which was implemented with a Laplace Smoothing did not make any changes to the results of the model and thus can be concluded that zero probability did not exist in our data. With 89**%** accuracy, and 84% Kappa Statistic, Model C gives us great performance results to predict students' future procrastination tendencies in online mathematics learning.

Table 4. The Confusion Matrix Analysis of the Best Model Generated

CMA		ACTUAL			
		Non Proc	Low Proc.	Moderate Proc.	High Proc
PREDICTED	Non Proc.	17	3	0	0
	Low Proc.	1	7	0	1
	Moderate Proc.	0	0	9	0
	High Proc.	0	0	1	14

Considering Model C as our best model, the Confusion Matrix Analysis for the different classes of procrastination tendencies is shown in Table 4. When the model made 53 predictions, 47 of those predictions are correct predictions. Among the different classes or procrastination tendencies, it can be seen in the confusion matrix that only few cases are misclassified by the model. If the model will be utilized specifically to identify the High Procrastinators and Moderate Procrastinators in a set of students, the model can give considerably good results as lesser misclassifications are observed in those classes. To further understand the performance of the model in each classes or procrastination tendencies, Table 5 shows the summary of performance of each class derived from the Confusion Matrix Analysis.

The figures presented in Table 5 are derived from the Confusion Matrix Analysis in Table 4. In simple terms, Sensitivity shows the competence of our model in correctly identifying students in that specific procrastination tendency. Specificity, on the other hand, shows the competence of our model in correctly identifying students that is not in that specific procrastination tendency. The Confusion Matrix Analysis shows that all Moderate Procrastinators are correctly predicted by the model. Thus, its sensitivity score is shown to be 1.0. Its specificity score, however, is 0.98 since one high procrastinator was misclassified as a moderate procrastinator.

Table 5. The Sensitivity and Specificity of the Each Class in the Best Model Generated

Class	Sensitivity	Specificity
Non Procrastinator	0.85	0.97
Low Procrastinator	0.78	0.93
Moderate Procrastinator	1.00	0.98
High Procrastinator	0.93	0.97

Overall, our model is greater in identifying the students who are expected to procrastinate more, or in identifying the moderate and high procrastinators. If a teacher uses this model as a tool for the early detection of procrastinating behaviors in his students, then this model can definitely serve its purpose.

Procrastination as experienced by the Students and recommended actions

Now that students are classified and described in their procrastination tendencies and that there is a novel model that can identify students' procrastination tendencies, another purpose of this study is to offer recommended actions to teachers once their students are identified as procrastinators. To do so, Focus Group Discussions were made to each class of procrastination tendencies to interview in terms of their experience as either Non procrastinators, Low, Moderate, or High Procrastinators.

Collaizzi's Method of Data Analysis was used. It included transcriptions, analyses of significant statements, formulated meanings, and themes. Using purposive sampling, a total of 20 students – five in each procrastination category – participated in the Focus Group Discussions. The FGDs were done remotely and separately in each class or procrastination tendency using a video-conferencing tool.

During the FGDs, the following five (5) central themes were gathered from students' experiences. Recommended actions are taken from these experiences:

Online classes have technological requirements, if these technological requirements are not met, learning would be a hassle [33]. It was expressed by the students that those who have limited resources like using only smart phone and/or mobile data to work on online classes and tasks took more effort to complete it.

To help students with technological constraints, based on students' experiences, the researchers recommend that mathematics teachers would record their synchronous sessions so students with internet connectivity issues can asynchronously watch the lesson by the time their connectivity returns. In addition, it is best for mathematics teachers to consider the kind of platform to use on performance-based tasks, that it may still be mobile-friendly to students who have lack of gadgets. In a study of Terry [47], instructor-created videos during online classes are concluded to be well-received by the students, allowing students to learn more the content, be clarified with assignment directions, and even for more personal connection to instructors during online classes.

Online learning demands more active learning on the part of the student [4,5]. Thus, students with little focus may possibly wander off non-academic-related activities during their free time.

To help students who get easily distracted, students during the FGD expressed that it would definitely help them if they get reminded from time-to-time. From this, the researchers recommend that teachers can lend a hand to easily distracted students by giving constant reminders to them about tasks and submissions that they are ought to do during the day and week and quarter. These recommended action would agree to the results of the studies of Humpfrey [48] that explored the use of text message reminders to improve student performance as more students submit assignments on-time. It was also revealed in the study of Nichol [49] that reminding students, sent through emails, lead to increased lecture viewership and further improved course outcomes.

Table 6. Themes Derived from the Analysis of the FGD Responses

Thematic Use	**Non-verbatim Response/s with frequency**	**Sample Responses**	**Generated Description of Themes**
Technological Constraints	Slow internet connection causes us to submit our tasks late (15); Using cellphone only limits tasks that we can do (8);	*I believe slow internet cultivates procrastination because if I can't join classes due to slow internet, I don't know how to do the tasks because I don't understand how to do it.* (Student 17); *Sometimes tasks require us to use Microsoft Office which we cannot do in phones* (Student 11)	Technological limitations such as slow internet connectivity and available gadget to use contributes to their procrastination
Distractions	Social Media (17) and Online Games (11) consumes a lot of their free time.	*When I have free time, I have not noticed its already been long when I open Facebook or TikTok* (Student 6)	Students are not focused on tasks and are distracted
Frequency of Overall Tasks	Some students find it hard to complete all the tasks with other subjects (7) and they are not reminded always (5)	*I do Math tasks last, but tasks of other subjects are given immediately and as well as the deadline* (Student 9)	Organized system of giving tasks must be observed
Low Motivation	Students deliberately miss the tasks (4)	*I find doing other things more interesting than listening to class* (Student 18)	Low interest in joining classes
Anxiety and Pressure	Students are afraid to answer tasks (11) especially when they were absent during discussion (8)	*It looks very hard to do it especially when I don't know the topic because I'm not around* (Student 4)	Students' anxiety in Math makes them feel like they can't do it.

During the FGDs, students expressed that some subjects gave many tasks and demands to be submitted in a short time. Students expressed, further, that some of them do math-related tasks after doing the other tasks.

Although a school-level concern, the researchers suggest that teachers can agree with the number of tasks that must be given in a week, or should be open to extending the deadlines, if possible. A news report from the Philippine Daily Inquirer [50] articulated that students are 'overwhelmed' by tasks under the new normal learning system, and are overworked with tasks.

Students have shown that sometimes they easily get tired and deliberately not attend to their tasks at all. Instead of being pressured, the pile of undone tasks gives them the impression that they can no longer do it.

Based on these students' experiences, the researchers recommend that the school may conduct webinars or activities that can motivate students in their online classes. Math teachers can also opt for interesting activities to keep students engaged during lessons. This aligns to the recommendations in study of Mamolo [51] that urged Mathematics teachers to use more engaging instructional strategies as learners tend to have low motivation in self-studying specifically in Mathematics. Students expressed being pressured and anxious as contributors to why they would tend to procrastinate on their tasks. Based on these experiences, the researchers recommend that mathematics teachers can consider doing activities and strategies that can augment students' anxiety towards the subject.

The study of Mamolo [51] in his recommendations suggests Mathematics teachers more engaging activities to decrease students' anxiety and improve their self-efficacy. To those students who needs more time understanding mathematics topics, remedial classes to low performing students where these students can raise their questions without fear of being ridiculed and for more guided discussion is suggested.

4. Conclusions and Recommendations

Based on the findings of this study, the researchers conclude that the generated Gaussian-Bernoulli Mixed Naïve Bayes Model is a great tool that mathematics educators can utilize to predict students' later procrastination behaviors in their online learning of mathematics subject. The results of the predictions of this tool can be used by mathematics teachers in taking early precautions, instant actions, or selecting a student that is fit for a certain task. Some interventions and activities are recommended according to the students' experiences during their online classes. The researchers recommends that since Naïve Bayes is just one among the many Machine Learning models available, investigating more recent and sophisticated algorithms, like Neural Networks and others, with more data or input features that may also have great predictive value to the students' procrastination in Mathematics is suggested.

References

[1] N. Michinov, S. Brunot, O. Le Bohec, J. Juhel, and M. Delaval, "Procrastination, participation, and performance in online learning environments," *Comput. Educ.*, vol. 56, no. 1, pp. 243-252, 2011.

[2] Deped Order no. 18, "No Title," 2020. [Online]. Available: https://www.deped.gov.ph/wp-content/uploads/2020/08/DO_s2020_018.pdf.

[3] J. S. Barrot, I. I. Llenares, and L. S. del Rosario, "Students' online learning challenges during the pandemic and how they cope with them: The case of the Philippines," *Educ. Inf. Technol.*, no. 0123456789, 2021.

[4] F. J. Hinojo-Lucena, I. Aznar-Díaz, and M. P. Cáceres Reche, "Percepciones del alumnado sobre el blended learning en la universidad," *Comunicar*, vol. 16, no. 33, pp. 165-174, 2009.

[5] K. Nordby, K. B. Klingsieck, and F. Svartdal, "Do procrastination-friendly environments make students delay unnecessarily?," *Soc. Psychol. Educ.*, vol. 20, no. 3, pp. 491-512, 2017.

[6] J. W. You, "Examining the Effect of Academic Procrastination on Achievement Using LMS Data in E-Learning," *Educ. Technol. Soc.*, vol. 18, no. 3, pp. 64-74, 2015.

[7] C. Lay, "At last, my research article on procrastination.Journal of Research in Personality," *J. Res. Pers.*, vol. 20, no. 4, pp. 474-495, 1986.

[8] O. A. Asikhia, "Academic Procrastination in Mathematics: Causes, Dangers and Implications of Counselling for Effective Learning," *Int. Educ. Stud.*, vol. 3, no. 3, pp. 205-210, 2010.

[9] G. Schraw, T. Wadkins, and L. Olafson, "Doing the things we do: A grounded theory of academic procrastination," *J. Educ. Psychol.*, vol. 99, no. 1, pp. 12-25, 2007.

[10] A. H. Chun Chu and J. N. Choi, "Rethinking procrastination: Positive effects of 'active' procrastination behavior on attitudes and performance," *J. Soc. Psychol.*, vol. 145, no. 3, pp. 245-264, 2005.

[11] P. Steel, "The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure," *Psychol. Bull.*, vol. 133, no. 1, pp. 65-94, 2007.

[12] D. Hooshyar, M. Pedaste, and Y. Yang, "Mining educational data to predict students' performance through procrastination behavior," *Entropy*, vol. 22, no. 1, p. 12, 2020.

[13] J. Park, R. Yu, F. Rodriguez, R. Baker, P. Smyth, and M. Warschauer, "Understanding student procrastination via mixture models," *Proc. 11th Int. Conf. Educ. Data Mining, EDM 2018*, pp. 187-197, 2018.

[14] B. W. TUCKMAN, "Relations of Academic Procrastination, Rationalizations, and Performance in a Web Course With Deadlines," *Psychol. Rep.*, vol. 96, no. 4, p. 1015, 2005.

[15] R. Cerezo, M. Esteban, M. Sánchez-Santillán, and J. C. Núñez, "Procrastinating behavior in computer-based learning environments to predict performance: A case study in Moodle," *Front. Psychol.*, vol. 8, no. AUG, pp. 1-11, 2017.

[16] A. Akram *et al.*, "Predicting Students' Academic Procrastination in Blended Learning Course Using Homework Submission Data," *IEEE Access*, vol. 7, pp. 102487-102498, 2019.

[17] A. Akram *et al.*, "Akram2019," no. 1, p. 1.

[18] M. Balkis and E. Duru, "Gender differences in the relationship between academic procrastination, satifaction with academic life and academic performance," *Electron. J. Res. Educ. Psychol.*, vol. 15, no. 1, pp. 105-125, 2017.

[19] H. L. Dempsey, "Running head: GENDER AND AGE DIFFERENCES IN PROCRASTINATION Gender and Age Differences in Procrastination Brittany Cangialosi Jacksonville State University Honors Thesis in Psychology Supervisor: Heidi L. Dempsey, Ph. D., Professor of Psychology Com," no. December, 2019.

[20] C. M. Mandap, "Examining the Differences in Procrastination Tendencies among University Students," *Int. J. Educ. Res.*, vol. 4, no. 4, pp. 431-436, 2016, [Online]. Available: www.ijern.com.

[21] B. U. Ozer, "A cross sectional study on procrastination: Who procrastinate more?," *Int. Conf. Educ. Res. Innov.*, vol. 18, pp. 34-37, 2011.

[22] P. Rosário, M. Costa, J. C. Núñez, J. González-Pienda, P. Solano, and A. Valle, "Academic procrastination: Associations with personal, school, and family variables," *Span. J. Psychol.*, vol. 12, no. 1, pp. 118-127, 2009.

[23] N. F. B. Casillano, "Challenges of Implementing an E-Learning Platform in an Internet Struggling Province in the Philippines," *Indian J. Sci. Technol.*, vol. 12, no. 10, pp. 1-4, 2019.

[24] D. Çağırgan and İ. Soytürk, “The relationship between math anxiety, student engagement in mathematics and responsibilities towards learning among middle school students,” *Elem. Educ. Online*, vol. 20, no. 1, pp. 456-467, 2021.

[25] D. Mendoza, M. Cejas, G. Rivas, and C. Varguillas, “Anxiety as a prevailing factor of performance of university mathematics students during the covid-19 pandemic,” *Obraz. i Nauk.*, vol. 23, no. 2, pp. 94-113, 2021.

[26] J. T. G. E. J. G. Reyes, “BSIT Students Readiness: An Assessment of Students’ Engagement in Online Learning At Nueva Ecija University of Science And Technology, San Isidro Campus during COVID-19 Pandemic,” *Int. J. Sci. Adv. Res. Technol.*, vol. 7, no. 2, pp. 107-115, 2021.

[27] H. Çiğdem and O. G. Yıldırım, “Effects of Students’ Characteristics on Online Learning Readiness: A Vocational College Example,” *Turkish Online J. Distance Educ.*, vol. 15, no. July, pp. 80-93, 2014.

[28] T. Cunningham, C. McDonnell, B. McIntyre, and T. McKenna, “A reflection on teachers’ experience as e-learners,” *Appl. E-Learning E-Teaching High. Educ.*, pp. 56-83, 2008.

[29] S. Çiftci, E. Güneş, and M. T. Üstündağ, “Attitudes of distance education students towards web based learning - A case study,” *Procedia - Soc. Behav. Sci.*, vol. 2, no. 2, pp. 2393-2396, 2010.

[30] A. Rhema, I. Miliszewska, and E. Sztendur, “Attitudes towards e-Learning and Satisfaction with Technology among Engineering Students and Instructors in Libya,” *Proc. 2013 InSITE Conf.*, pp. 157-171, 2013.

[31] R. Peytcheva-Forsyth, B. Yovkova, and L. Aleksieva, “Factors affecting students’ attitudes towards online learning - The case of Sofia University,” *AIP Conf. Proc.*, vol. 2048, no. December, 2018.

[32] L. Visser, *Academic Procrastination among First-year Student Teachers*. .

[33] R. E. Baticulon *et al.*, “Barriers to Online Learning in the Time of COVID-19: A National Survey of Medical Students in the Philippines,” *Med. Sci. Educ.*, vol. 31, no. 2, pp. 615-626, 2021.

[34] B. Fabito, A. Trillanes, and J. Sarmiento, “Barriers and Challenges of Computing Students in an Online Learning Environment: Insights from One Private University in the Philippines,” *Int. J. Comput. Sci. Res.*, vol. 5, no. 1, pp. 441-458, 2021.

[35] J. Tus, “Amidst the Online Learning in the Philippines: The Parental Involvement and Its Relationship to the Student’s Academic Performance x Issue 3 AMIDST THE ONLINE LEARNING IN THE PHILIPPINES: THE,” *Int. Eng. J. Res. Dev.*, vol. 6, no. 3, pp. 1-15, 2021.

[36] M. B. Kursa and W. R. Rudnicki, “Boruta: Wrapper Algorithm for All Relevant Feature Selection,” pp. 1-17, 2020.

[37] K. M. Schneider, “A comparison of event models for naive bayes anti-spam E-mail filtering,” *10th Conf. Eur. Chapter Assoc. Comput. Linguist. EACL 2003*, pp. 307-314, 2020.

[38] A. Max *et al.*, “Package ‘caret’ R topics documented,” 2022.

[39] B. Love, A. Vetere, and P. Davis, “Should Interpretative Phenomenological Analysis (IPA) be Used With Focus Groups? Navigating the Bumpy Road of ‘Iterative Loops,’ Idiographic Journeys, and ‘Phenomenological Bridges,’” *Int. J. Qual. Methods*, vol. 19, pp. 1-17, 2020.

[40] P. M. Feng, H. Ding, W. Chen, and H. Lin, “Naïve bayes classifier with feature selection to identify phage virion proteins,” *Comput. Math. Methods Med.*, vol. 2013, 2013.

[41] M. B. Kursa and W. R. Rudnicki, “Feature selection with the boruta package,” *J. Stat. Softw.*, vol. 36, no. 11, pp. 1-13, 2010.

[42] H. W. Chang, Y. H. Chiu, H. Y. Kao, C. H. Yang, and W. H. Ho, “Comparison of classification algorithms with wrapper-based feature selection for predicting osteoporosis outcome based on genetic factors in a Taiwanese women population,” *Int. J. Endocrinol.*, vol. 2013, 2013.

[43] N. A. R. S. S. A. A. Kaushik, “Feature selection on educational data using Boruta algorithm,” *Int. J. Comput. Intell. Stud.*, vol. 10, no. 1, pp. 27=35, 2021.

[44] R. Lumacad, Gernel and Namoco, “A Multilayer Perceptron Neural Network Approach to Classifying Learning Modalities Under the New Normal,” *Unpubl. MastersThesis, Univ. Sci. Technol. South. Philipp.*, 2021.

[45] S. Bagui, K. Devulapalli, and S. John, “MapReduce Implementation of a Multinomial and Mixed Naive Bayes Classifier,” *Int. J. Intell. Inf. Technol.*, vol. 16, no. 2, pp. 1-23, 2020.

[46] S. Visa, B. Ramsay, A. Ralescu, and E. Van Der Knaap, “Confusion matrix-based feature selection,” *CEUR Workshop Proc.*, vol. 710, pp. 120-127, 2011.

[47] M. W. Laura Terry, Maria Zafonte, “Benefits of Recording and Sharing Instructor-Created Videos in Online Classes,” *Proc. Soc. Inf. Technol. Teach. Educ. Int. Conf.*, pp. 1899-1904, 2020, [Online]. Available: https://www.learntechlib.org/p/215969/.

[48] W. Humphrey, D. Laverie, and A. Shields, “Exploring the Effects of Encouraging Student Performance With Text Assignment Reminders,” *J. Mark. Educ.*, vol. 43, no. 1, pp. 91-102, 2021.

[49] T. Nichol, Penny; Maruani, Esther; Brothen, “Just Click Here: The Influence of Lecture Viewing Reminders on Lecture Viewing and Course Outcomes,” *MINNESOTA SUMMIT Learn. Technol.*, 2017, [Online]. Available: https://pubs.lib.umn.edu/index.php/mslt/article/view/859.

[50] M. Adonis, “Students overwhelmed by tasks under ‘new normal’ way of learning,” *Philippine Daily Inquirer*, 2022.

[51] L. A. Mamolo, “Online Learning and Students’ Mathematics Motivation, Self-Efficacy, and Anxiety in the ‘new Normal,’” *Educ. Res. Int.*, vol. 2022, 2022.