Moodle Analytics

Prepared By
Muhammad Bilal
IS/BS (CS)/S-17/M/021
Muhammad Faheem
IS/BS (CS)/S-17/M/022

Supervised By:
Dr. Saeed Ullah
Assistant Professor
Department of Computer Science

Coordinated By:
Mr. Nasir Ayub
Lecturer
Department of Computer Science

A PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS OF THE DEGREE OF

BACHELOR OF SCIENCE IN COMPUTER SCIENCE

Session [2017-2021]

FEDERAL URDU UNIVERSITY OF ARTS, SCIENCE AND TECHNOLOGY, ISLAMABAD

Moodle Analytics

A Project Presented To

Federal Urdu University Of Arts Science & Technology

In Partial Fulfilment Of The Requirement For The

Degree of Bachelor of Computer Science

By

Muhammad Bilal
&
Muhammad Faheem

February, 2021

Final Approval

Dated

It is certified that I have read the project report submitted by Muhammad Bilal (IS/BS(CS)/S-17/M/021) and Muhammad Faheem (IS/BS(CS)/S-17/M/022) and it is our judgment that this project is of sufficient standard to warrant its acceptance by the FUUAST university, Islamabad for the Degree of BSCS in System Software.

Committee

External Examiner

Ch Anwar ul Hassan

Project Supervisor

Dr. Saeed Ullah

Department of Computer Science

Project Coordinator

Mr. Nasir Ayub

Department of Computer Science.

Dedication

"Dedicated…

To the Holy Prophet (PBUH) "The Mohsen" of whole community,

To our dearly loved parents who sanctified us with their acumen and they are at all times there when we require them the most. They guided us all through and supported us throughout the hardships of life.

To our teachers who shared with us their experience and knowledge.

And

To our friends from their good will and company we enjoyed every single moment of our Education and university life"

Abstract

A dropout before time caution system enables institutes to precautionary spot students who are at risk of dropping out of foundation, to swiftly react to them, and ultimately to facilitate would-be dropout students to carry on their learning for an enhanced future. On the other hand, the intrinsic class inequity between dropout and non-dropout students could pose complicatedness in structuring precise extrapolative modeling for a dropout early warning scheme. The current study is designed to recover the feat of a dropout early warning system.

Contents

Revision History

Version	Description	Author	Date
1.1	Final Release	Muhammad Bilal Muhammad Faheem	01-March-2021

Abstract and abbreviations

	Abbreviation
SDPS	Student Dropout Prediction System
LMS	Learning Management System
LA	Learning Analytics
MOOL	Massive Open Online Lab
DEWS	Dropout Early Warning System

1. Introduction:

The depressing fallout of understudy's dropping out of foundation are huge for the entity and the public both. The instructive inadequacies of dropout understudies could seriously restrict financial and social success in their afterward lives.

The general public additionally endures misfortunes in light of the fact that the country's gainful limit could be sabotaged by the lack of the gifted workforce, and furthermore the dropout understudies are bound to be incessant beneficiaries of government assistance and joblessness appropriations.

In light of those negative results, understudy's dropouts have for quite some time been considered as a genuine instructive issue by instructors, scientists, and politicians. A dropout early admonition framework can help schools to prematurely recognize understudies who are in danger of dropping out of school and to speedily respond to them.

The understudies in peril are presumably going to exit without circumspectly considering the negative consequences of their decisions or without getting an opportunity to chat with subject matter experts. The from the get-go intervention taught by the dropout early counsel structure can redirect potential dropout understudies onto the best approach to graduation and guide them to a predominant future.

Because of the inconceivable potential, various organizations have made dropout early caution systems. For example, the division of guidance and youth improvement in the region of Victoria in Australia developed the Student Mapping Tool (SMT) to assist schools with perceiving understudies at risk for detachment and dropout, and the area of Wisconsin in the United States developed the (DEWS) to predict understudy's dropouts.

In the United States, about part of public optional schools realized the dropout early caution structures during 2014–2015.

The Computer Science program at the Federal University of Ceara (UFC) achieved likely the most vital score in the 2014 arrival of the ENADE. A test that surveys the idea of Brazilian student programs. Despite the program's significance, its typical dropout rate some place in the scope of 2005 and 2015 was about 45%, considering understudies who were yielded by either choice test or National High School Exam (ENEM) scores. The choice test devised by UFC was applied from UFC's foundation until 2010 and the

ENEM, a public test that evaluates optional school graduates is figured by MEC and has been used as the arrangement test at UFC since 2011.

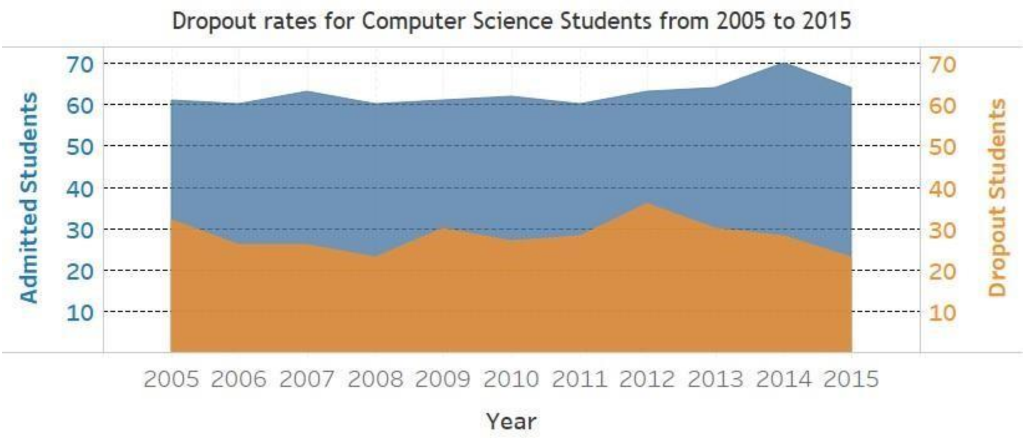

Figure 1: Dropout rate by year of Computer Science understudies at UFC

Figure 1 shows the dropout pace of the understudies in the Computer Science's 2000.1 educational plan, from 2005 to 2015. In spite of having a couple of changes, it portrays an upward pattern beginning in 2009. The class of 2012 was viewed as the most noticeably terrible with connection to sidestepped understudies, showing that over half of the understudies exited.

Educational information mining (EDM) and learning investigation are rising controls that manage the way toward examining instructive information. This evaluation is done through the mixed bag of calculable systems, strategies and mechanical assemblies, including AI and data mining. The objective of learning assessment is to give examination is of the data that starts in the informative storage facilities, similarly as in the LMS, in order to appreciate and overhaul the learning cycle and the conditions in which it occurs.

1.1 LMS Concept:

A Learning Management System (LMS) is a worker established instructional stage that permits instructive establishments to deal with an enormous number of completely on the web or blended courses utilizing a mutual interface and set of assets inside a domain that will undoubtedly existence. In the previous 20 years, LMS has seen a tremendous development in utilization for schools and colleges around the world, and gigantic incomes for specialist organizations, with a striking piece of the pie for Blackboard Inc by 2009 . The most widely recognized highlights of a LMS are satisfied formation by teachers (transferring archives, work sheet, introductions, pictures, sound and cinematic) correspondence besides collaboration among educators in addition understudies (declarations, pages, messages, conversation sheets, wikis, websites, and record sharing), evaluation devices shared by teachers and which enables them to follow understudy action (tests, articles, tests, reviews, inputs), and organization devices given to teachers to deal with the settings for the recently referenced highlights and tweak their courses (allowing admittance to understudies, selecting teachers and understudies, empowering records and courses, and following exercises).

An exploration report distributed in 2015 separates the top ventures that utilization LMS programming, with a conspicuous predominance for instruction over different fields, for example, innovation, producing, medical services, counseling, and other variation organizations and organizations all things considered as appeared in Figure 1.

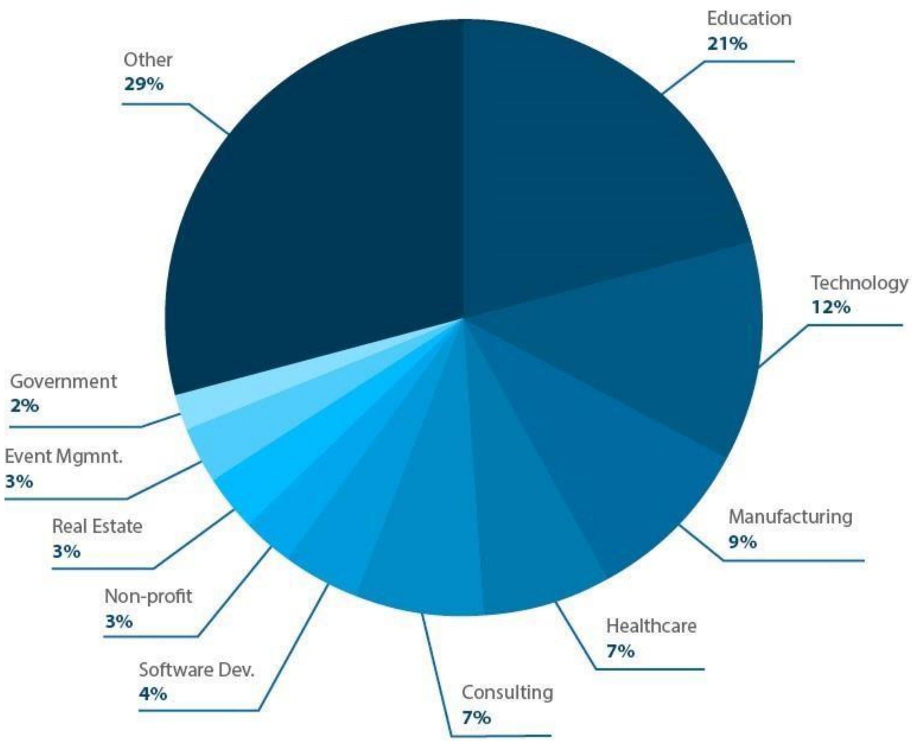

Figure 2: Top ventures that utilization LMS programming

The comparative assertion express that the most purchasers of LMS until that date were consistently organization proprietors, preparing supervisors, course creators, learning and improvement establishments, project administrators and advisors as demonstrated in Figure 2.

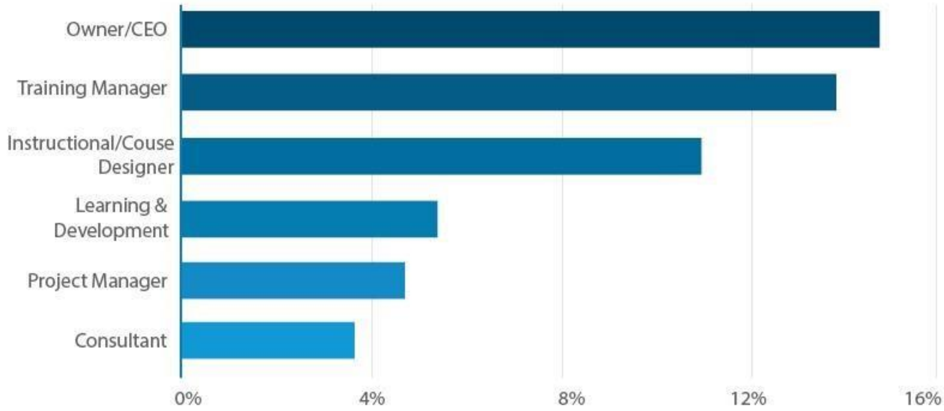

Figure 3: Most regular purchasers of LMS

Figure 3 from the comparative report shows the way that instructive establishments are utilizing a larger number of clients than business ones, which mirrors the high commitment of instructors and regulatory staff in eLearning programs.

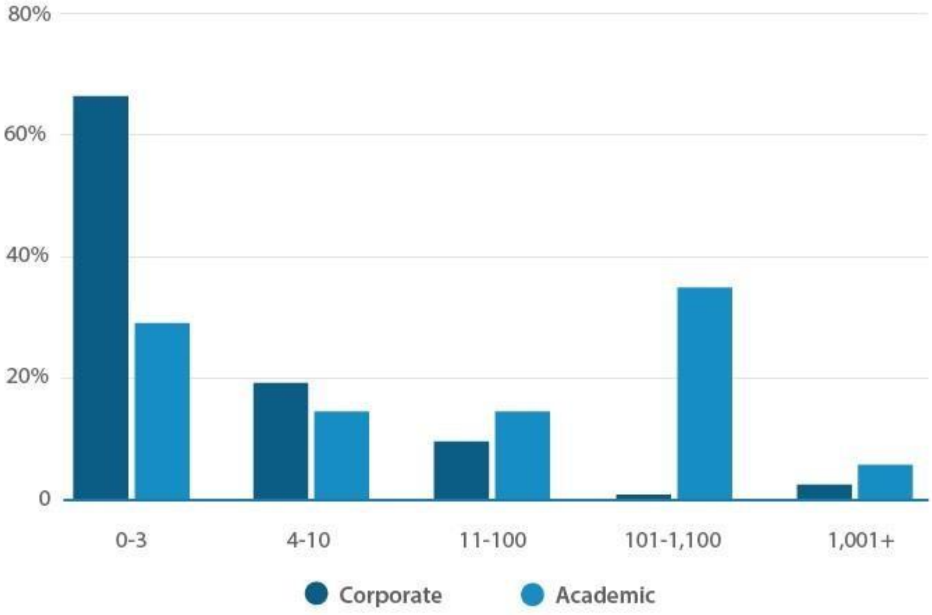

Figure 4: Number of scholastic versus corporate LMS client

Concerning the LMS programming by and by utilized by foundations, a similar report presents matchless quality for Moodle over different mediums along with Blackboard that once stood firm on the number 1 foothold in the publicize, which backs up our decision for Moodle in our postulation.

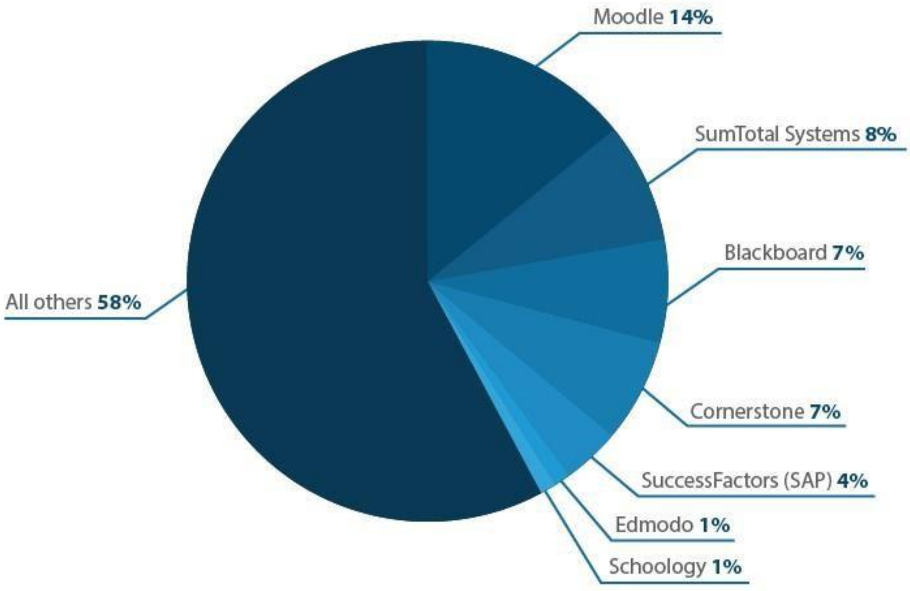

Figure 5: Top utilized LMS programming

Additionally value assumes a major part in figuring out which LMS stage to buy, usefulness stays the primary decider on which purchasers depended.

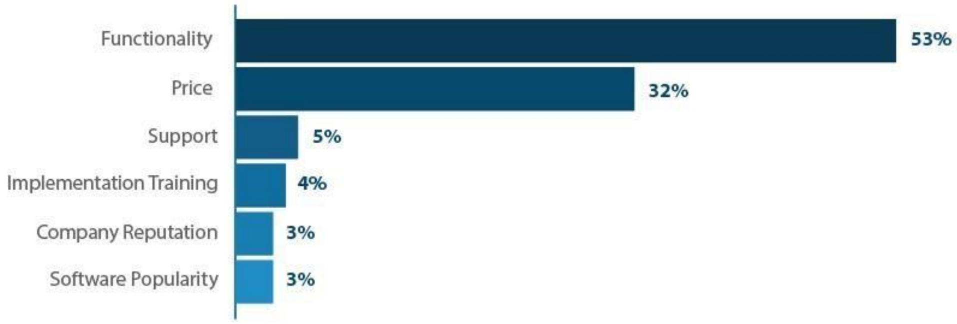

Figure 6: The main elements in LMS buy choice

These frameworks offer a great deal of advantages to the two educators and understudies. The fundamental benefits can be summed up in the accompanying:

1- Focal learning: substance is realistic for all clients consistently from a similar asset.

2- Tracking and detailing attributes: customer improvement can be checked, records can be logged and evaluated, and reports can be made to progress open courses and understudy's introduction.

3- Assessment ability: understudies can be ceaselessly evaluated during the course from one side to another verbose coursework and tests.

4- Inconvenience free overhauls: substance can be changed with no difficulty, and each and every one client will get the information simultaneously.

5- Novice's learning method: stages are typically expected to be fathomable and to help clients, educators, and heads to get the most brilliant agreement.

To upgraded uphold the importance through a LMS. It is likewise colored in explanation the most utilized LMS attributes Figure 6 and the most wanted LMS highlights Figure 7.

Figure 6: Most utilized LMS highlights

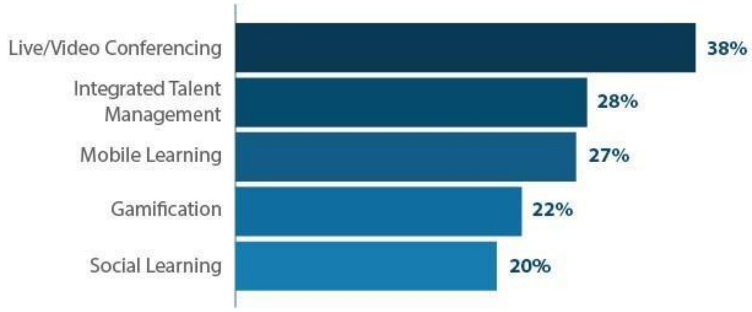

Figure 7: Most wanted LMS highlights

Regardless, there are at this point various features and viewpoints that require improvement in order to grow the level of satisfaction. For example the openness of some huge features, the accommodation of the system, particular assistance and cost as showed up in Figure 8.

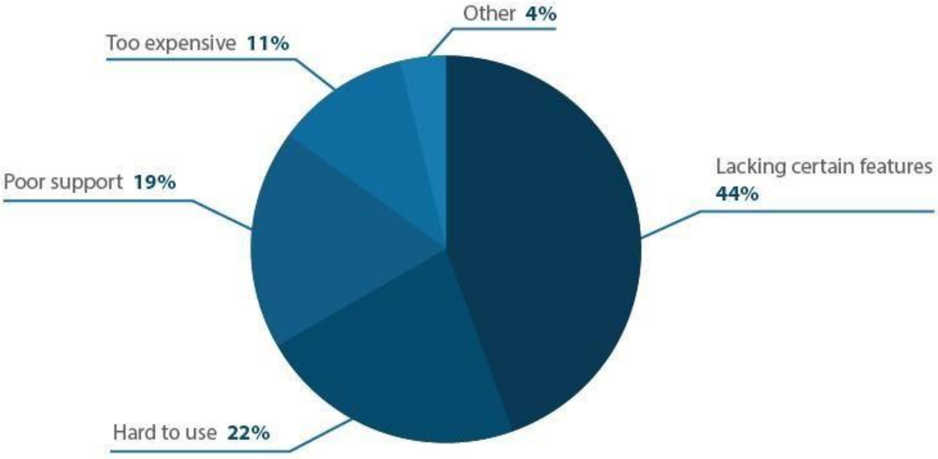

Figure 8: Why clients are not happy with their present LMS

1.2 Machine Learning:

In the book "Introduction to Machine Learning" defined machine learning (ML) as follows:

"Artificial intelligence is changing PCs to improve an introduction premise using model data or point of reference understanding. We have a structure described up to specific limits, and instruction is the usage of a PC educational program to advance the restrictions of the model utilizing the establishing data or past experience. The model may be insightful to make prospect later on or realistic to get data from data or both. It uses the speculation of estimations in building mathematical models because the middle task is making acceptance from a model."

In the course of recent years, ML strategies, for example, neural organizations and deep learning have given incredible indications of progress on both mechanical and practical levels because of the huge advancement in equipment and the high versatility gave by cloud arrangements. In any case, it is assessed that future investigations will primarily zero in on hereditary qualities and improved comprehension of the humanoid mind and its evolvement past as of now utilized techniques.

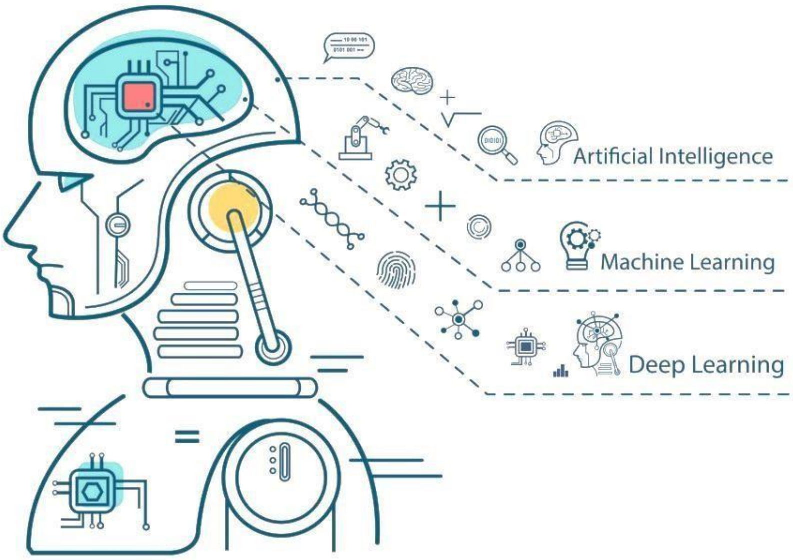

Figure 9: AI Idea

The significance of machine learning deceptions in its capacity on the way to create good guesses that can be able to prompt live and enhanced results and trial with no human conciliation by investigating large information also anticipating patterns.

Business pioneers who have executed ML ventures in their associations have distinguished and plan to pick up the accompanying advantages:

1- Identifying patterns and examples easily by social occasion and dissecting huge information.

2- Obtaining a favorable position over contenders by utilizing examined patterns to modify their deals and showcasing systems.

3- Improving the nature of examination because of the framework's capacity to improve proficiency and precision after approximately period with the emerging measures of information that become prepared.

4- Saving period and fee as the structure is robotized and doesn't bother with any human consideration to work. It could moreover achieve the ID of separating hidden drivers and to the assumption for therapeutic frameworks for upkeeps.

It considers ML just like the eventual fate of eLearning and features 5 favorable circumstances in this field:

1- Improving eLearning degree of profitability by investing more energy in upgrading preparing quality and keeping tabs on understudy's development and other learning rehearses.

2- Delivering more customized eLearning content: by utilizing design acknowledgment to foresee learning results and prescribing redid subject materials to help the understudy to accomplish information.

3- Employing visit bots as eLearning teachers: by permitting this implicit program to connect with understudies and offer exact responses to their inquiries in an intelligent and moment way.

4- Motivating students to procure information: by meeting the understudy's necessities and wiping out unessential data to spare them time.

5- Providing numerous appraisal designs: by assessing understudy's information with different test designs.

Then again, numerous disservices actually should be handled:

1- Data procurement: ML requires colossal instructive assortments to get ready on and they ought to be of adequate quality and unbiased.

2- Time and assets: It requires a lot of resources for work and a huge load of time to permit the counts to learn and develop enough to yield careful and relevant results.

3- Understanding: it's the capacity towards precisely decipher results.

4- High bungle helplessness: if lacking planning sets are used, results may not be complete and a high edge of slip-ups might be accomplished.

1.3 Motivation:

With the plentiful studies and accomplishment taking place in the most recent few years, IOT in eLearning is steadily changing from being the outlook of education to its nearby. And with the extension of IOT to exceed 50 billion devices universally, struggle between service providers is arising and ML is also taking its role in contribution end users the best-customized practice they could desire for. however, ML in eLearning in all-purpose and MOOLs in specific is not yet a well-discovered ground and holds enormous potential in augmenting the evaluation of student’s lab work since it needs illustration check and bright human-like authentication. This stimulate us to suggest and execute an LMS function that makes use of IOT and ML tools in evaluating the student’s presentation mechanically throughout research laboratory testing.

1.4. Background:

In this part we illustrate Distance Education facet, dropout notion and come up to deal with dropout prophecy using Educational Data Mining procedure.

1.4.1 Distance Education and the dropout problem:

The Distance Education licenses knowledge in a versatile way, where teachers and understudies can be isolated geographically and fleetingly. Mediation is finished with resources in different media maintain. According to KEEGA, which isolates Distance Education and self-study exercises is the opportunity of various correspondence points, where understudies and instructors can interface and confer electronically and in occasional get-togethers. Other than that, in most piece of Distance Education courses there are similarly bundle correspondence works out, related to the improvement of a couple of critical understudy's capacities. Dependent upon the educational activity can be watched a mix of self-administration, construction and trade and the adjustment of these points should be a concern in the masterminding and checking of a given course.

The term dropout licenses various interpretations. Some of the time a dropout event is seen as the course withdrawal by the understudy, regardless the proportion of classes made a difference. In various conditions the dropout is contrastingly thought to be as shown by the ordinary time span for course fulfillment. As a result, a few records measure withdrawal in a particular control of a course anyway various ones measure the withdrawal of the entire course. Thusly, these examinations consider some determined differences for the dropout. They consider it contrastingly when the dropout is related to the control, when the dropout occurs inside the course, inside the association and regardless, when the understudy pulls back from the enlightening system itself.

1.4.2 Educational Data Mining and dropout prediction in Distance Education courses:

As of now the data age limit is significantly higher than the restriction of experts to analyze the set aside data. This declaration is furthermore clear in the field of Education, in light of the fact that builds up every day the volume of data made and set aside in informational collections, on account of the wide use of motorized systems in schools and universities. Appropriately, the presence of this gigantic volume of data has engaged the use of data mining strategies, in order to help in searching for answers to express preparing questions, related to learning measures, improvement of instructional materials, checking and gauge among others. Enlightening Data mining deals with the usage of Data Mining methods with new courses of action of data got in different educational settings. The possibility of these data addresses a reasonable execution of resources that are chief to help in progress of preparing. A couple of models are time of alerts maintain for proposal structures or catch of understudy's profiles.

There are fitting approaches to manage oversee data with different perspectives concerning their capriciousness, pondering courses of action of all the more consistent data and all the more impressive data. The first is on a very basic level molded by issues related to educational history and social points of view, which are seen with more static features and addressing requested perspectives. When in doubt these game plans of data are dealt with to deliver information that can be ground-breaking in helping institutional expectation exercises with educational exercises. The ensuing one is made mainly by data recognized in the step by step association and activities. The rule exercises that such data license are those related to the improvement of activities in short time period periods, usually the ones found in the hour of one semester. The results from practices inside a request can be used to create diagnostics that show exercises inside a smaller time scope, overseeing possible brief dropout conditions. Division Education courses address an opportunity, as a result of their inclination, that incorporates a raised degree of cutting edge mediation and, thusly, can give a huge proportion of data as for understudies.

The educational practice can be used to perceive composed exertion plans that are developed among understudies with different levels of venture and with different levels of execution. For example, a couple of works shows a differentiation in the collaboration illustration of understudies with high assessments as differentiated and understudies that gave up courses. At the point when perceived, these models can be applied in future balance exercises.

Regardless, a couple of works portray a philosophy that fuses even more thoroughly the starter examination of factors to be checked by Educational Data Mining techniques, in order to make circumstances that are unsurprising with data about the general pattern of dropout. The rule point of convergence of huge piece of known works in Educational Data Mining is moved in practices looking for experimentation of arranged Data mining techniques and their variety to enlightening data setting. A significant part of the time the essential objective imparted is the endorsement of a specific system and conspicuous evidence of perspectives to be noticed.

1.3. Statement Of Problem:

Dropout in education is big loss to individuals and societies. Dropout of students is considered as a waste of human resources both for the teachers and the students (time taken to attend). It is necessary to minimize the number of dropout students, as it prevents the education system from achieving its objectives.

1.4. Proposed Methodology:

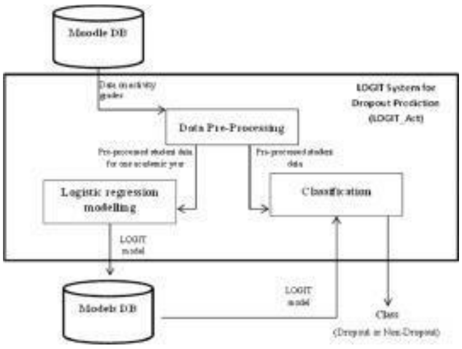

Figure 10: System Architecture

This piece communicates the strategies used to achieve the target of this paper. This examination paper looks at thirteen changed datasets (set of traits), for every one of the 4 (four) particular semesters: 2016-1, 2016-2, 2017-1 and 2017-2, of an Introductory Programming Course to calculate whether the sorts of essence, pressure in the exhibition of prescient models to early distinguish in danger understudies. The outline of the received approach is appeared in Figure 10 and the four stages strategy are appeared in Figure 11.

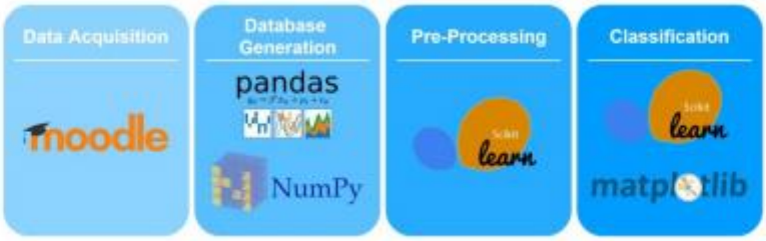

Figure 11: Outline of the embraced philosophy

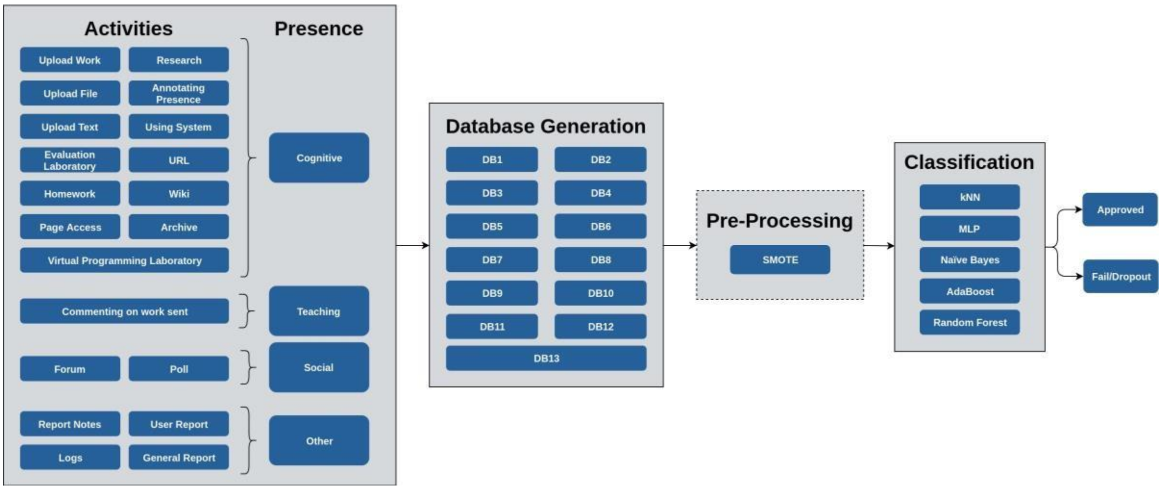

Figure 12: Steps of the implemented approach

The initial speed comprises of assemblage of information from Moodle logs, remembering that the stage records the interchanges that the understudies make in the VLE. The after that progression contain delivering the datasets holding different qualities to differentiate them and approve those which achieve the most phenomenal domino impact. In this manner, we use some pre-handling strategy, for example, oversampling, be resolved to intensify the exhibition of the form. The fourth step comprises of the creation and assessment of the arrangement models. In the closing stage, we put one next to the other accomplished outcomes to address the exploration questions. After that subsections clarified in extra detail the stages followed.

1.4.1. Generation And Evaluation Of The Models:

For classification we utilized Naive Bayes, Random Forest, AdaBoost, Multilayer Perceptron (MLP), kNearest Neighbor (kNN) and Decision Tree calculations. Be that as it may, all through the examinations, we disengaged the last one because of its over-fitting. Every one of these calculations were tried utilizing the Scikit-learn library in Python.

2. Literature Review And Related Works:

The idea of Learning Analytics has remained standing out as truly newsworthy for certain years presently, bringing a ton of enthusiasm among higher inclining foundations around the world. Embracing on the web or mixed method of conveyance infers that the greater part of student's exercises will be occurring in LMS.

Subsequently, there is a requirement for establishments to gauge the quality and power of LMS utilization. Indeed, contemplates have demonstrated that there is a relationship between’s LMS use with understudy's presentation just as understudy's fulfillment with courses offered by means of LMS.

It should be noticed that, LMS gather huge measures of information on understudy conduct that can be utilized to advise and improve understudy commitment in LMS. These data incorporate clients calls, amount of downloads, LMS devices got to, posts delivered or mailed and content sheets visited. As indicated by Whitmer, such data clarify more than multiple times the variety in conclusive evaluations contrasted with conventional understudy trademark factors, and that joining the two kinds of factors increment the nature of anticipating learning execution by over 70%.

Subsequently, completely through assessing LMS utilization by means of log information we can recognize status of understudy's learning and even predict their conceivable learning accomplishment. For this situation, we can perceive under tension understudies needing scholarly hold up. We can likewise utilize Learning Analytics apparatuses to assess the greatness of online postings and envision use practices in the framework.

Given the repayment and chances possible by log information put away in LMS data set, contemplates have built up an assortment of Learning Analytics devices to inspect such information. For instance, examined issues control understudy's scholarly achievements utilizing log information created from Moodle with 84 understudies in a college in South Korea. They build up that complete considering time, contact with peers, instantaneousness of learning span, and figure of downloads had significant impact on understudy's scholarly introduction in internet learning environmental factors.

Jo et al recognized "contender for intermediary factors" from the log informational index that could be utilized as motivation to choose understudy's exhibitions in courses open by means of Moodle. The creators utilized the perceived factors to steer them in business eLearning course with an aggregate of 200 understudies. It was uncovered that a rashness of the learning hole appeared to have relationship with learning execution.

Kotsiantis et al. utilized Learning Analytics apparatus called Moodle Parser to gather information from logs and to distinguish successful students in mixed gaining course from one finish to another understudy's exercises in Moodle. The investigation sets up that understudy's breakdown was associated with their critical methodology and understanding towards Moodle. Then again, fantastic evaluations were related with expanded Moodle utilization. The examination depended on 337 understudies enlisted in a bound together course more than three years utilizing Moodle at the University of Patras, Greece.

Modritscher and partners contemplated connection between LMS utilization examples and understudies execution utilizing information from framework log. The creators uncovered that, point sees had useful force on understudy's presentation. They likewise satisfied that in danger understudies could be recognized from well-performing understudies by their utilization execution. At long last, an examination led at Central Queensland University utilizing an example of 92,799 understudies, detailed a measurement huge connection between's the quantity of site visits and understudies last grades. As it were, the more 'perspectives' or visits to the course, the higher the last grades.

These couple of revisions and numerous others appearance that Learning Analytics instruments have appeared to empower foundations to investigate understudy's exercises and in the end helping in discovering procedures of helping the individuals who are in danger. Similarly, this examination created Learning Analytics device that empowers teachers to inspect understudy's exercises in Moodle done using information from the system record.

3. Problem Analysis:

In Problem Analysis stage we will research issues that we may look during this project. A difficult examination is an examination of the reasons for issues or disappointments. This is done to recognize enhancements to framework, cycle, methodology and plan.

The dropout early admonition framework permit foundations to preemptively distinguish understudies who are in danger of exiting organization, to speedily respond to them, and in the end to help potential dropout understudies to proceed with their learning for a superior future. Be that as it may, the inborn class dissimilarity among dropout and no dropout understudies could present trouble in building exact prescient demonstrating for a dropout early admonition framework. The current examination intended to improve the presentation of a dropout early admonition framework.

The conveyed idea of distance learning has raised new difficulties. For example, in contrast to study halls, it turns out to be a lot harder for educators in distance figuring out how to administer, control and changes the learning interaction.

In monstrous open online courses, where a large number of understudies are learning, it is hard for an educator to think about singular capacities and inclinations. Also, the appraisal obviously out comes in LMS is a difficult and requesting task for both accreditation and personnel.

Consequently, it is important to give a shrewd framework versatile capacities so it could adequately play the instructor job. Scientists proposed utilizing Learning Analytics for addressing significant data about understudies on the web.

The smart Moodle, in light of a recently evolved online LA framework:

1. Gives dashboards to educators to effectively assist them with managing their understudies on the web.
2. Predicts in danger understudies who may neglect to breeze through their end of the year tests. In particular, the utilization of some game plan components may help in anticipating understudy's

with lower execution and who can be in danger of neglecting to breeze through their last tests of the year.

3. Gives ongoing mediations as warnings by giving steady learning substance to understudies while learning.
4. Help educators plan mediations when required.
5. Backing dynamic with regards to managerial assignments.

Impacting the Future of Education:

Right now is an ideal opportunity to enter the learning investigation field and have an effect in the training of current understudies and long lasting students. To do so requires a strong essential comprehension of quantitative strategies, joined with ability in building, combining, overseeing, cleaning and breaking down information utilizing forefront programming and methods.

4. Modeling/Design:

In this research we are using logistic regression modeling and classification to predict the student at risk of dropping out at early stage and different student parameters discussed in later section.

In insights, the calculated model (or logit model) is utilized to show the likelihood of a specific class or occasion existing like pass/fizzle, win/lose, alive/dead or sound/wiped out. This can be reached out to show a few classes of occasions, for example, deciding if a picture contains a feline, canine, lion, and so forth Each item being distinguished in the picture would be relegated a likelihood somewhere in the range of 0 and 1 with an amount of one.

1. Logistic Regression Modeling:

Logistic regression is an arithmetical model that in its fundamental structure utilizes a strategic capacity to demonstrate a parallel ward variable, however numerous multifaceted extra rooms endure. In relapse examination, strategic disappointment is estimation of the constraints of a calculated model. Precisely, a paired calculated model has a ward inconsistent with two potential qualities, for example, Pass/Fail which is addressed by a pointer variable, where the two qualities are marked as "0" and "1". In the calculated model, the log-chances for the worth marked "1" is a direct gathering of at least one self-overseeing factors ("indicators"). The self-sufficient factors can each be a double lopsided (two classes, coded by a marker variable) or a constant variable (any genuine worth). The same probability of the worth marked "1" can fluctuate between 0 (absolutely the worth "0") and 1. (Absolutely the worth "1"). The capacity that modify log-chances to probability is the calculated capacity, henceforth the name. The unit of measurement for the log-chances scale is known as a logit, from strategic unit, subsequently the elective names. Comparing models with an alternate sigmoid capacity as an option of the calculated capacity can likewise be utilized, for example, the probit model. The significant element of the strategic model is that expanding one of the sovereign factors multiplicatively scales the chances of the given result at a consistent rate, with every autonomous variable having its own injury, for a parallel ward variable this sums up the chances proportion.

Calculated relapse is perhaps probably the most ideal methods of duty such grouping. Like straight debilitating, strategic disappointment delivers a model of the connection between various factors. Strategic relapse is proper when the variable being predict for is a likelihood on a double reach from 0 to 1.

2. Classification:

Moodle supports 5 classification models:

1. kNN
2. MLP
3. Naïve Bayes
4. AdaBoost
5. Random Forest

2.1. kNN:

K-Nearest Neighbor is a simple control characterization calculation. We can use to dispense a class to new information point. It very well may be utilized for decay also. KNN doesn't make any proposition on the information conveyance, consequently it is non-parametric. It keeps all the preparation information to make future estimate by registering the likeness between an info test and each preparation occurrence.

KNN can be brief as below:

1. Calculate the distance between the new information points with each preparation model.
2. For figuring the distance estimates for example, Euclidean distance, Hamming distance or Manhattan distance will be utilized.
3. Model picks K passages in the data set which are adjoining to the new information point.
4. Then it does the mass vote for example the most standard class/mark among those K sections will be the class of the new information point.

2.2. MLP:

A multilayer perceptron (MLP) is a profound reproduced neural organization. It is gathered of more than one perceptron. They are comprised of an info layer to take conveyance of the sign, a yield layer that settles on a decision or conjecture about the information, and in the middle of those two, a self-assertive number of covered up layers that are the genuine computational motor of the MLP. MLPs with one secret layer are fit for approximating any consistent capacity.

Multilayer perceptron's are frequently applied to manage learning issues. They train on a bunch of information yield combines and figure out how to show the affiliation or conditions between those data sources and yields. Preparing includes changing the boundaries, or the loads and predispositions, of the model to limit mistake. Back spread is utilized to cause those to gauge and inclination changes comparative with the blunder, and the actual mistake can be estimated in an assortment of ways, including by root mean squared mistake (RMSE).

2.3. Naïve Bayes:

Naive Bayes calculations are a bunch of managed AI calculations dependent on the Bayes likelihood hypothesis, which we'll talk about in this article. Gullible Bayes calculations surmise that there's no association between highlights in a dataset used to prepare the model.

Regardless of this improved on assumption, guileless Bayes classifiers function admirably in numerous perplexing true issues. A major benefit of guileless Bayes classifiers is that they just require a moderately modest number of preparing information tests to perform characterization proficiently, contrasted with different calculations like strategic relapse, choice trees, and backing vector machines.

Naive Bayes is a factual arrangement strategy dependent on Bayes Theorem. It is one of the least difficult regulated learning calculations. Gullible Bayes classifier is the quick, exact and trustworthy calculation. Gullible Bayes classifiers have raised accuracy and speed on huge datasets.

Naive Bayes classifier accepts that the impact of a specific element in a class is autonomous of different highlights. For instance, a credit candidate is attractive or not relying upon his/her pay, past advance and exchange history, age, and area. Regardless of whether these highlights are related, these highlights are as yet considered autonomously. This presumption improves on calculation, and that is the reason it is considered as credulous. This supposition is named class provisional individuality.

Pros:

1. It isn't just a basic methodology yet additionally a snappy and exact strategy for surmise.
2. Naive Bayes has very low down calculation price.

3. It can professionally labor on a great dataset.
4. It carries out well in case of separate rejoinder variable matched to the incessant variable.
5. It can be used with manifold class calculation problems.
6. When the postulation of autonomy holds, a Naive Bayes classifier accomplishes improved compared to other models like logistic regression.

Cons:

1. The supposition of self-governing features. In practice, it is almost not possible that model will get a set of predictors which are completely free.
2. If there is no preparation tuple of a specific class, this causes zero back likelihood. For this situation, the model can't make expectations. This issue is known as Zero Probability or Frequency Problem.

2.4. AdaBoost:

AdaBoost or Adaptive Boosting is one of gathering boosting classifier proposed by Yoav Freund and Robert Schapire in 1996. It joins different classifiers to enlarge the rightness of classifiers. AdaBoost is an iterative outfit technique. AdaBoost classifier develops a genuinely incredible classifier by joining different inadequately performing classifiers with the goal that you will get high precision solid classifier.

Adaboost is to set the loads of classifiers and preparing the information test in every emphasis with the end goal that it guarantees the precise expectations of abnormal perceptions. Any AI calculation can be utilized as base classifier on the off chance that it acknowledges loads on the preparation set.

Adaboost should meet two conditions:

1. The classifier ought to be prepared intuitively on different gauged preparing models.
2. In every emphasis, it attempts to offer an exceptional fit for these models by limiting preparing mistake.

Pros:

1. AdaBoost is not difficult to try.
2. It iteratively remedies the missteps of the feeble classifier and improves exactness by joining frail students.
3. We can utilize many base classifiers with AdaBoost.
4. AdaBoost isn't inclined to overfitting.

2.5. Random Forest:

Random forest is an adaptable, easy to utilize AI calculation that produces, even without hyper boundary tuning, an extraordinary outcome more often than not. It is likewise perhaps the most utilized calculations, due to its straightforwardness and variety (it very well may be utilized for both characterization and relapse assignments).

Random forest is a directed learning calculation which is utilized for both grouping just as relapse. Be that as it may, then again, it is for the most part utilized for classification issues. As we realize that a timberland is comprised of trees and more trees implies more powerful woodland. Moreover, arbitrary woods calculation makes choice trees on information tests and afterward gets the expectation from every one of them lastly chooses the most fitting answer by methods for casting a ballot. It is a band strategy which is superior to a solitary choice tree since it diminishes the over-fitting by averaging the outcome.

Pros:

The subsequent are the benefits of Random Forest process:

1. It defeats the issue of overfitting by averaging or consolidating the aftereffects of various choice trees.
2. Random forest function admirably for a huge scope of information things than a solitary choice tree does.
3. Random forest has less change at that point single choice tree.
4. Random forests are entirely adaptable and have high precision.
5. Scaling of information doesn't need in irregular backwoods calculation. It keeps up great exactness even in the wake of giving information without scaling.
6. Random Forest calculations keeps up great exactness even an enormous extent of the information is absent.

Cons:

The subsequent are the drawbacks of Random Forest algorithm:

1. Complication is the primary disadvantage of Random forest calculations.
2. Construction of Random forest are a lot harder and tedious than choice trees.
3. More computational capital is needed to try Random Forest calculation.
4. It is less instinctive in the event that when we have an enormous assortment of choice trees.

5. The expectation measure utilizing chance woods is tedious interestingly with different calculations.

Parameters And Indicators For Moodle:

Table 1: Moodle Parameters

	Parameters	Indicators
1.	Course at risk of not starting	Teacher Availability, Student Enrolments
2.	Students at risk of dropping out	Course accessed after end date, Course accessed before start date, Any written action in the course, Read actions amount, Assignment cognitive, Quiz cognitive etc.
3.	Students who have not accessed the course recently	Any course access
4.	Students who have not accessed the course yet	Any course access
5.	Upcoming activities due	Activities Due

5. Data Gathering:

We use dummy dataset in our project and we converted its attributes according to our project requirements. The datasets are picked from github (https://github.com/aninda052/Drop-Out-Prediction/tree/master/data). The dataset contains the information of more than 2500 students with 19 attributes and 2735 instances. The title of the dataset is ‘studentinfo’. All data provided in this dataset is nominal, numeric and float. This dataset will be used for the purpose of implementation. The features includes in our dataset is new_id, batch, sex, dobday, dobmon, dobyear, fathermonin, sscins, sscboard, sscsub, sscgrade, ssctotal, sscyear, data, hscboard, hscsub, hscgrade, hsctotal and hscyear.

new_id	batch	sex	dobday	dobmon	dobyear	fathermonin	sscins	sscboard	sscsub	sscgrade	ssctotal	sscyear	data	hscboard	hscsu
1	20162	male	15.0	7.0	1999.0	60000/-	IDEAL SCHOOL AND COLLEGE	COMILLA	SCIENCE	A	4.56	2014.0	IDEAL SCHOOL AND COLLEGE	NaN	SCIENC
2	20022	m	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Na
3	20120	male	1.0	8.0	1992.0	20000/-	TANGAIL TEC SCHOOL	BTEB	NaN	A	4.14	2009.0	TANGAIL TEC SCHOOL	BTEB	Na
4	20090	male	26.0	4.0	1990.0	30000/-	SENAPALLY HIGH SCHOOL	DHAKA	SCIENCE	NaN	4.5	2006.0	SENAPALLY HIGH SCHOOL	DHAKA	SCIENC
5	20111	male	1.0	2.0	1990.0	15000/-	JOYPURHAT R.B. GOVT. HIGH SCHOOL	RAJSHAHI	SCIENCE	A	4.25	2006.0	JOYPURHAT R.B. GOVT. HIGH SCHOOL	BTEB	SCIENC

Figure 13: Student Information Dataset

And also we used grade dataset in our project. That contains the academic record of students. This dataset contains the grade information of more than 2500 students with 7 attributes and 65,235 instances. The title of the dataset is 'csestudensgrades'. All data provided in this dataset is nominal, numeric and float. This dataset will also be used for the purpose of implementation. The features includes in this dataset is new_id, offered_id, semester_id, course_code, grade, gp and credits.

	new_id	offered_id	semester_id	course_code	grade	gp	credits
0	267.0	NaN	1	CSE111	F	0.0	3
1	1273.0	NaN	1	CSE123	F	0.0	3
2	1273.0	NaN	1	CSE231	F	0.0	3
3	1273.0	NaN	1	CSE312	F	0.0	3
4	2032.0	NaN	1	MATH124	F	0.0	3

Figure 14: Student Grade Dataset

For preprocessing the dataset, we have used Python library Sickie-learn (sklearn).

1. The dataset comprises of certain credits, which are string esteems and different ascribes are in numeric qualities. There attributes that contains strings values and NaN values. Those attributes need to be converted and values to be filled. But in our case those attributes are not of our use. So we would drop that attributes instead of converting or filling NaN values. Those attributes haven't any effect on our result or doesn't required for our prediction.
2. The attributes that our dataset includes are new_id, batch, sex, dobday, dobmon, dobyear, fathermonin, sscins, sscboard, sscsub, sscgrade, ssctotal, sscyear, data, hscboard, hscsub, hscgrade, hsctotal and hscyear.
3. Dropping the attributes names sscgrade, hscgrade, dobday, dobyear, dobmon, fathermonin, sscins, sscboard, sscsub, data, hscboard, hscsub and sex.
4. After dropping the columns that is not of our need, we added columns that contains the record of semester cgpa and semester failed subjects from first to last semester. And also added the label attribute with name drop out.
5. Checking the performance of students on the basis of their batches.

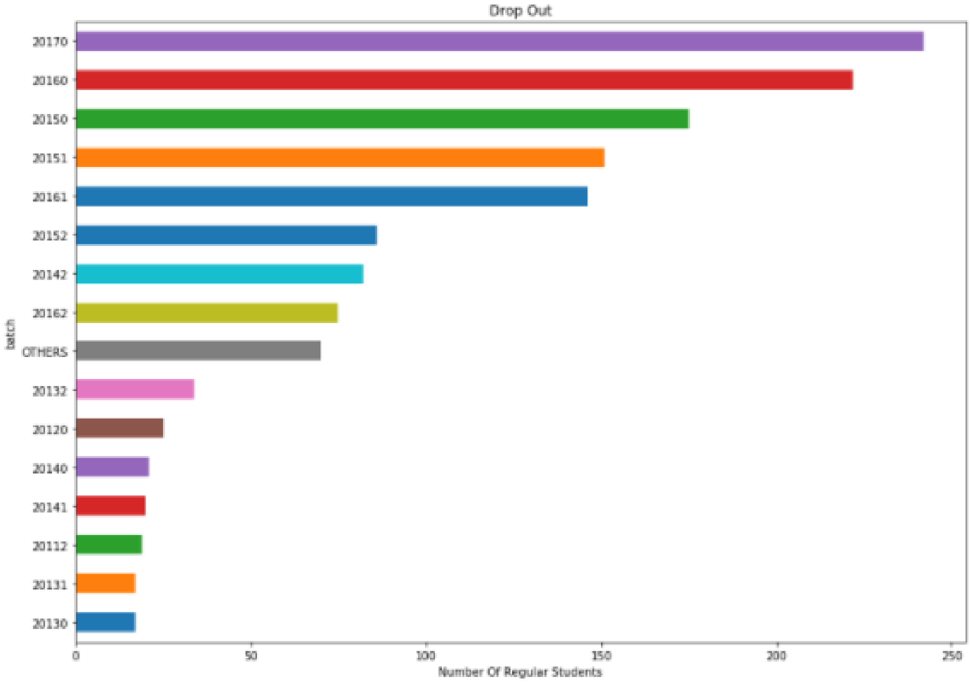

Figure 15: Student Performance On Batch Basis

6. The visualized results of overall regular and dropout students in our dataset. In our case, 26.2% of students drop out.

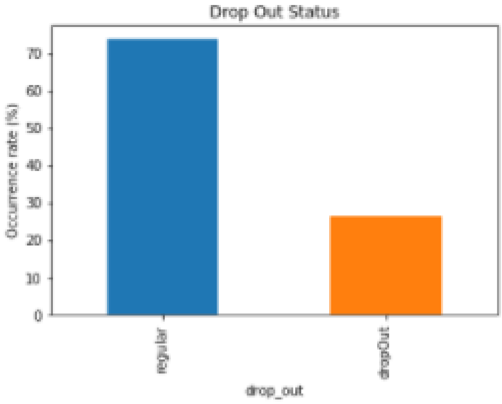

Figure 16: Dropout Status

6. Result And Discussions:

We use dataset which contains both the combination of unmitigated and mathematical qualities. In this manner, the basically centers around those calculations which can deal with the blend of both clear cut and numeric qualities. Likewise, remembering that, the calculation performs well for our characterization issue. Along these lines, a few calculations are picked to fill the need like KNN, Ada-Boost, Naïve Bayes, MLP, irregular timberland and strategic relapse. The fundamental intention of this paper is to utilize calculations on these datasets to arrange either the understudy is at the danger of exiting or not, founded on past records (for example semester gpa, semester bombed subjects). The picked calculations are applied where it gives a basic and quick method of learning a capacity. The applied calculation gives better execution for any arrangement issue and moodle underpins just these calculations. For this reason 70% of the information is utilized for the preparation reason and 30% of the information is utilized for the testing reason.

The general precision standard is utilized to quantify the absolute extent of the understudies whose last status is finished or dropout, was accurately anticipated by a method. As such, this basis gauges the quantity of effectively anticipated completers in addition to the quantity of effectively anticipated dropouts, versus the all-out number of the understudies.

KNN:

- Precision of KNN is 79.10%
- The classification description of KNN is:

Table 2: Classification Description KNN

	Precision	**Recall**	**F1-Score**	**Support**
Dropout	0.27	0.94	0.42	34
Regular	0.99	0.78	0.87	387
Macro Avg	0.63	0.86	0.65	421
Weighted Avg	0.94	0.79	0.84	421

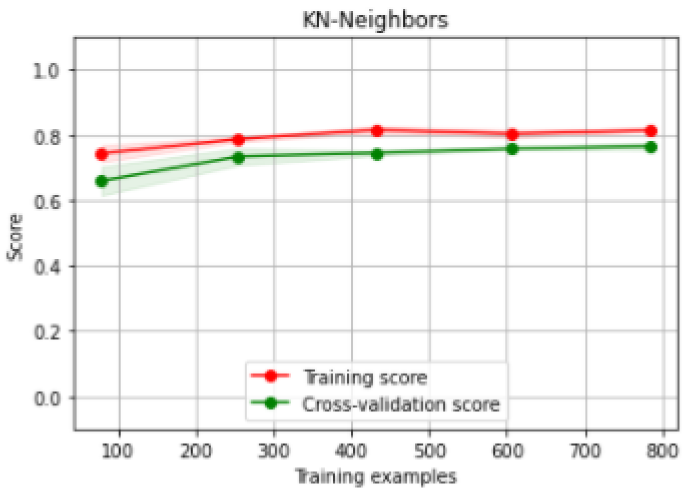

Figure 17: Learning Curve of K-Nearest neighbors

AdaBoost:

- Precision of AdaBoost is 96.19%
- The classification description of Adaboost is:

Table 3: Classification Description AdaBoost

	Precision	Recall	F1-Score	Support
Dropout	0.94	0.93	0.93	120
Regular	0.97	0.98	0.97	301
Macro Avg	0.96	0.95	0.95	421
Weighted Avg	0.96	0.96	0.96	421

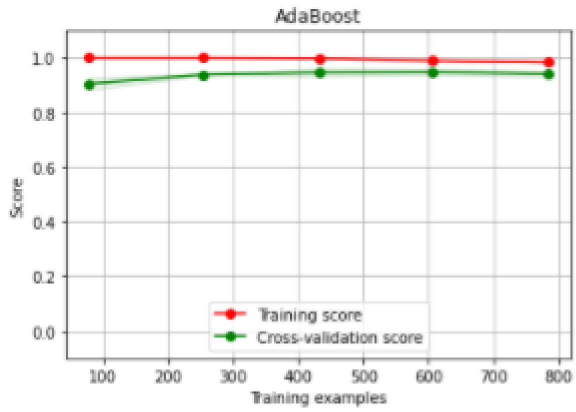

Figure 18: Learning Curve of Ada-Boost

Naïve Bayes:

- Precision of Naïve Bayes is 74.58%
- The classification description of Naïve Bayes is:

Table 4: Classification Description Naïve Bayes

	Precision	Recall	F1-Score	Support
Dropout	0.95	0.53	0.68	213
Regular	0.67	0.97	0.79	208
Macro Avg	0.81	0.75	0.73	421
Weighted Avg	0.81	0.75	0.73	421

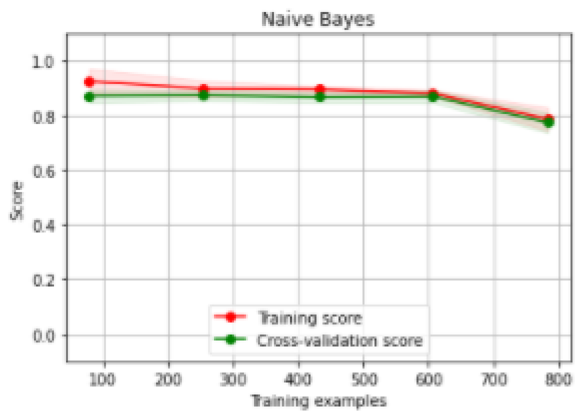

Figure 19: Learning Curve of Naïve Bayes

MLP:

- Precision of MLP is 71.97%
- The classification description of MLP is:

Table 5: Classification Description Multilayer Perceptron

	Precision	Recall	F1-Score	Support
Dropout	0.00	0.00	0.00	118
Regular	0.72	1.00	0.84	303
Macro Avg	0.36	0.50	0.42	421
Weighted Avg	0.52	0.72	0.60	421

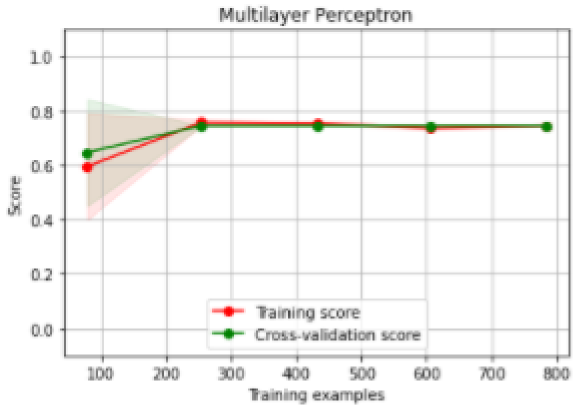

Figure 20: Learning Curve of Multilayer Perceptron

Logistic Regression:

- Precision of Logistic Regression is 87.37%
- The classification description of Logistic Regression is:

Table 6: Classification Description Logistic Regression

	Precision	Recall	F1-Score	Support
Dropout	0.81	0.67	0.74	368
Regular	0.89	0.94	0.92	1034
Macro Avg	0.85	0.81	0.83	1402
Weighted Avg	0.87	0.87	0.87	1402

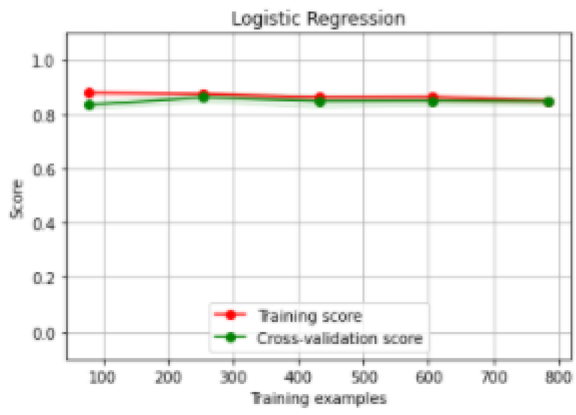

Figure 21: Learning Curve of Logistic Regression

Random Forest:

- Precision of Random Forest is 91.69%
- The classification description of Random Forest is:

Table 7: Classification Description Random Forest

	Precision	**Recall**	**F1-Score**	**Support**
Dropout	0.88	0.83	0.86	125
Regular	0.93	0.95	0.94	296
Macro Avg	0.91	0.89	0.90	421
Weighted Avg	0.92	0.92	0.92	421

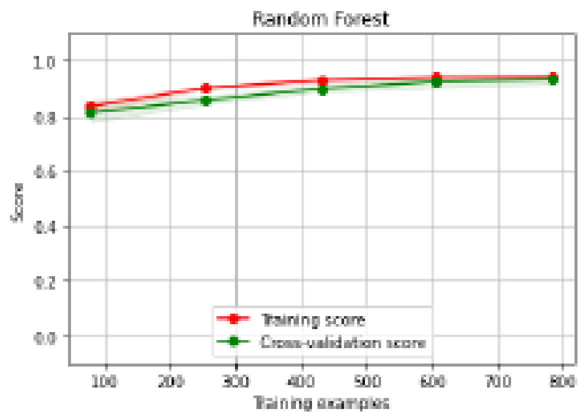

Figure 22: Learning Curve of Random Forest

The above results shows that the detection sensitivities of KNN, AdaBoost, Naïve Bayes, MLP, Logistic Regression and Random Forest are 79.10%, 96.19%, 74.58%, 71.97%, 87.37% and 91.69% respectively. Learning Curve (LC) is likewise utilized for the assessment of the grouping calculations.

In LC, model is assessed on the preparation dataset and on a hold out approval dataset after each update during preparing and plots of the deliberate presentation can made to show expectations to learn and adapt. Red lines shows the preparation score of the models where green line shows the cross-approval score of the models.

7. Conclusion:

This research make accessible a total assessment of the latest and relevant exploration exercises on AI application toward foreseeing, clarifying and taking care of the issue of understudy disappointment in Moodle. It features understudy related issues that lead to a high figure of dropouts. The paper likewise perceived a portion of the genuine faces connected with understudy dropout estimation and gives proposition and recommendations to help out analysts utilizing different AI methods in settling it very much coordinated and expertly.

The Moodle understudy dropout point is just somewhat explored and there is a wide room of exploration highlights to be inspected furthermore to develop all around characterized and more accurate prognostic answers for group, fathom and explain the reasons of dropout. A more profound comprehension of the causes why understudy dropout could help course designers and speakers to improve course substance and embrace instructive intercessions. Future work will follow on information assortment for the turn of events and examination of profound learning calculations towards settling the Moodle understudy dropout.

References:

1. Rumberger, R.W. High School Dropouts: A Review of Issues and Evidence. Rev. Educ. Res. 1987.
2. Catterall, J.S. On the social costs of dropping out of school. High School J. 1987.
3. Balfanz, R.; Herzog, L.; Maciver, D.J. Preventing Student Disengagement and Keeping Students on the Graduation Path in Urban Middle-Grades Schools: Early Identification and Effective Interventions. Educ. Psychol. 2007.
4. Dynarski, M.; Gleason, P. How Can We Help? What We Have Learned From Recent Federal Dropout Prevention Evaluations. J. Educ. Stud. Placed Risk (JESPAR) 2002.
5. Lamb, S.; Rice, S. Effective Strategies to Increase School Completion Report: Report to the Victorian Department of Education and Early Childhood Development; Communications Division, Department of Education and Early Childhood Development: Melbourne, Australia, 2008.
6. Knowles, J.E. Of needles and haystacks: Building an accurate statewide dropout early warning system in Wisconsin. JEDM J. Educ. Data Min. 2015.
7. Sullivan, R. Early Warning Signs. A Solution-Finding Report; Center on Innovations in Learning, Temple University: Philadelphia, PA, USA, 2017
8. Brasil (2015). Portaria no 564, de 18 de dezembro de 2015, anexo i. Di´ario Oficial [da] Rep´ublica Federativa do Brasil. Sec¸˜ao 1.
9. Buenaño-Fernandez, D.; Villegas-CH, W.; Luján-Mora, S. The use of tools of data mining to decision making in engineering education—A systematic mapping study. Comput. Appl. Eng. Educ. 2019.
10. [Carr, 12] Carr, S. (2012) As Distance Education Comes of Age, the Challenge Is Keeping the Students technology -The Chronicle of Higher Education.
11. [Keegan, 96] Keegan, D. (1996). Foundations of distance education. 3a. ed. London: Routledge.
12. [Moore, 93] Moore, M. (1993). Theory of transactional distance. In Theoretical principles of distance education. Oxon: Routledge.
13. [Favero, 06] Favero, R. V. M. (2006). Dialogar ou evadir: Eis a questão!: Um estudo sobre a permanência e a evasão na Educação a Distância, no Estado do Rio Grande do Sul. CINTEDUFRGS. Novas Tecnologias na Educação. 4(2).
14. [Adachi, 09] Adachi, A. A. C. T., (2009). Evasão e evadidos nos cursos de graduação da Universidade Federal de Minas Gerais, Dissertação de Mestrado. Faculdade de Educação, UFMG, 2009.
15. [Liao, 12] Liao, S. H., Chu, P. H., & Hsiao, P.-Y. (2012). Data mining techniques and applications – a decade review from 2000 to 2011. Expert Systems with Applications.
16. [Baker, 11] Baker, R., Isotani, S., Carvalho, A. (2011). Mineração de Dados Educacionais: oportunidades para o brasil. Revista Brasileira de Informática na Educação.

17. [Manhães, 11] Manhães, L. M. B., Cruz, S. M. S. da, Costa, R. J. M., Zavaleta, J., Zimbrão, G. (2011). Previsão de Estudantes com Risco de Evasão Utilizando Técnicas de Mineração de Dados. In: XXII SBIE -XVII WIE, 2011, Aracaju. Anais. . . [S.l.: s.n.], 2011.
18. [Fayyad, 96] Fayyad, U. M. Shapiro, G. P. Smyth, P. Uthurusamy. R. (1996). Advances in knowledge discovery and data mining. Menlo Park: MIT.
19. [Romero, 10] Romero, C., Ventura, S. (2010). Educational data mining: A re-view of the state of the art. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on.
20. [Romero, 12] Romero, C., Ventura, S., Pechenizkiy, M., Baker, R. S. J. (2012). Handbook of Educational Data Mining, Ed. CRC.
21. [Kampff, 09] Kampff, A., Mineração de Dados Educacionais para Geração de Alertas em Ambientes Virtuais de Aprendizagem como Apoio à Prática Docente, Tese de Doutorado, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre –RS, 2009.
22. [Durand, 11] Durand, G., Laplante, F., Kop, R. (2011). A Learning Design Recommendation System Based on Markov Decision Processes, KDD 2011 Workshop: Knowledge Discovery in Educational Data, ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2011) in San Diego, CA. August 21-24. 2011.
23. [Toscher, 10] Toscher, A., Jahrer, M. (2010). Collaborative filtering applied to educational data mining. KDD Cup 2010: Improving Cognitive Models with Educational Data Mining, 2010.
24. [Li, 11] Li, N., Cohen, W., Koedinger, K. R., Matsuda, N. (2011). A Machine Learning Approach for Automatic Student Model Discovery. EDM 2011: 31-40. Proceedings of the 4th International Conf on EducationalData Mining, Eindhoven, the Netherlands, July 6-8, 2011.
25. [Romero, 08] Romero, C., Ventura, S., Garcia, E. (2008). Data mining in course management systems: Moodle case study and tutorial. Computers & Education.
26. [Romero-Zaldivar, 12] Romero-Zaldivar V. A., Pardo A., Burgos, D, Kloos, C. A. (2012). Monitoring student progress using virtual appliances: A case study. Computers & Education.
27. [Lykourentzou, 09] Lykourentzou, I., Giannoukos, I., Nikolopoulos, V., Mpardis, G., Loumos, V. (2009). Dropout prediction in e-learning courses through the combination of machine learning techniques. Computers & Education.
28. Filippidi, A., Tselios, N., & Komis, V. (2010). Impact of Moodle Usage Practices on Students' Performance in the Context of a Blended Learning Environment. In Social Applications for Lifelong Learning (1-6). Patras, Greece.
29. Whitmer, J. (2012). *Logging On to Improve Achievement: Evaluating the Relationship between Use of the Learning Management System, Student Characteristics, and Academic Achievement in a Hybrid Large Enrollment Undergraduate Course.*

30. Macfadyen, L. P., & Dawson, S. (2010). Mining LMS Data to Develop an “Early Warning System” for Educators: A Proof of Concept. *Computers & Education*.
31. Naveh, G., Tubin, D., & Pliskin, N. (2012). Student Satisfaction with Learning Management Systems: A Lens of Critical Success Factors. *Technology, Pedagogy and Education*.
32. Beer, C., Clark, K., & Jones, D. (2010). Indicators of Engagement. In *ascilite Sydney 2010* (75–86). Sydney.
33. Yu, T., & Jo, I. (2014). Educational Technology Approach toward Learning Analytics : Relationship between Student Online Behavior and Learning Performance in Higher Education. In *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge (269-270). Indianapolis, IN, USA.*
34. Nistor, N., Trăuşan-Matud, Ş., Dascălu, M., Duttweiler, H., Chiru, C., Baltes, B., & Smeaton, G. (2015). Finding Student-Centered Open Learning Environments on the Internet: Automated Dialogue Assessment in Academic Virtual Communities of Practice. *Computers in Human Behavior*.
35. Scheffel, M., Niemann, K., Pardo, A., Leony, D., Friedrich, M., Schmidt, K., Wolpers, M. and Kloos, C. (2011). Usage Pattern Recognition in Student Activities. In C. Kloos, D. Gillet, R. Crespo García, F. Wild, & M. Wolpers (Eds.), *Towards Ubiquitous Learning (Vol. 6964, pp. 341–355). Springer Berlin Heidelberg. doi:10.1007/978-3-642-239854_27*
36. Jo, I.-H., Kim, D., & Yoon, M. (2014). Analyzing the Log Patterns of Adult Learners in LMS Using Learning Analytics. In *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge (LAK ’14)* (183-187). doi:10.1145/2567574.2567616
37. Kotsiantis, S., Tselios, N., Filippidi, A., & Komis, V. (2013). Using Learning Analytics to Identify Successful Learners in a Blended Learning Course. *International Journal of Technology Enhanced Learning*.
38. Kats, Y. (2010). *Learning Management System Technologies and Software Solutions for Online Teaching: Tools and Applications*.
39. Medved, J.P. (April 8, 2015). LMS Industry User Research Report, *Capterra*.
40. Parker, B. (December 29, 2014). Top 5 Benefits of Using a Learning Management System, *eLearning Brothers*.
41. Alpaydin, E. (2010). *Introduction to Machine Learning, 2nd Edition*, 3.
42. Samanpour, A., Ruegenberg, A., & Ahlers, R. (2018). The Future of Machine Learning and Predictive Analytics, *Digital marketplaces unleashed*.
43. Tavasoli, S. (January 10, 2020). The Importance of Machine Learning for Data Scientists, *Simplilearn*.

44. Neelakandan, N. (October 25, 2019). 5 Benefits of Machine Learning in eLearning, *eLearning Industry.*

45. Attaran, M., & Deb, P. (2018). Machine Learning: The New 'Big Thing' for Competitive Advantage, *International Journal of Knowledge Engineering and Data Mining.*

46. Agrawal, A. (September 02, 2019). Highlights the advantages and disadvantages of Machine Learning, *CIS*.

47. Garrison, D.R.; Anderson, T.; Archer, W. Critical inquiry in a text-based environment: Computer conferencing in higher education. *Internet High. Educ.* **1999**.

48. Rodríguez-del Pino, J.C.; Rubio-Royo, E.; Hernández-Figueroa, Z.J. A Virtual Programming Lab for Moodle with automatic assessment and anti-plagiarism features. In Proceedings of the International Conference on e-Learning, e-Business, Enterprise Information Systems, and e-Government (EEE), Las Vegas, ND, USA, 16– 19 July 2012; The Steering Committee of The World Congress in Computer Science, CSREA Press: Las Vegas, ND, USA, 2012.

49. Detoni, D.; Cechinel, C.; Matsumura Araújo, R. Modelagem e Predição de Reprovação de Acadêmicos de Cursos de Educação a Distância a partir da Contagem de Interações. *Revista Brasileira de Informática na Educação* **2015**.

50. Swan, K. Learning effectiveness online: What the research tells us. *Elem. Qual. Online Educ. Pract. Dir.* **2003**.

51. Moore, M.G. Three types of interaction. Am. J. Distance Educ. **1989.**

52. McKinney, W. Data structures for statistical computing in python. In Proceedings of the 9th Python in Science Conference, Austin, TX, USA, 28–30 June 2010; Volume 445.

53. Detoni, D.; Cechinel, C.; Matsumura Araújo, R. Modelagem e Predição de Reprovação de Acadêmicos de Cursos de Educação a Distância a partir da Contagem de Interações. *Revista Brasileira de Informática na Educação* **2015**.

54. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**.

55. Lemaître, G.; Nogueira, F.; Aridas, C.K. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *J. Mach. Learn. Res.* **2017**.

56. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**.

57. Bradley, A.P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit.* **1997**.

58. Gaševic´, D.; Dawson, S.; Rogers, T.; Gasevic, D. Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *Internet High. Educ.* **2016**.

59. Mann, H.B.; Whitney, D.R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* **1947**.

60. Vozniuk,A.,Govaerts,S.,&Gillet,D.(2013).Towardsportablelearninganalyticsdashboards. In 13th International Conference on Advanced Learning Technologies.

61. Yassine, S., Kadry, S., & Sicilia, M. A. (2016). A framework for learning analytics in moodle for assessing course outcomes. In Global Engineering Education Conference.