PREDICTION OF KPIs RELATED TO CLOs/PLOs

FAHAD HASSAN ZAMAN
01-244192-003

Department of Electrical Engineering

BAHRIA UNIVERSITY ISLAMABAD

Session 2019-21

Approval for Examination

Scholar's Name: Fahad Hassan Zaman

Registration No. 63172

Programme of Study: Masters of Science [Electrical Engineering]

Thesis Title: Prediction Of KPIs Related To CLOs/PLOs

It is to certify that the above scholar's thesis has been completed to my satisfaction and, to my belief, its standard is appropriate for submission for examination.I have also conducted plagiarism test of this thesis using HEC prescribed software and found similarity index 16% that is within the permissible limit set by the HEC for the MS degrce thesis.I have also found the thesis in a format recognized by the BU for the MS thesis.

Principal Supervisor's Signature:

Date: 02-09-2021

Name: Dr.Juniad Imtiaz

Author's Declaration

I, Fahad Hassan Zaman hereby state that my MS thesis titled "Prediction Of KPIs Related To CLOs/PLOs" is my own work and has not been submitted previously by me for taking any degree from BAHRIA UNIVERSITY ISLAMABAD or anywhere else in the country/world.

At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw/cancel my MS degree.

Name of scholar: Fahad Hassan Zaman

Date: 02-09-2021

Plagiarism Undertaking

I, solemnly declare that research work presented in the thesis titled "Prediction Of KPIs Related To CLOs/PLOs" is solely my research work with no significant contribution from any other person. Small contribution / help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Bahria University towards plagiarism. Therefore I as an Author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred / cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of Ms degree, the university reserves the right to withdraw / revoke my Ms degree and that HEC and the University has the right to publish my name on the HEC / University website on which names of scholars are placed who submitted plagiarized thesis.

Scholar / Author's Sign: Fahad

Name of scholar: Fahad Hassan Zaman

Dedication

First of all, I dedicate this effort to Allah Almighty, thank you for the guidance, power of mind, protection and skills and giving me a healthy life. Then, i dedicate this study to my beloved parents who were there when i thought of giving up, who continuously provided their moral, spiritual, financial and emotional support. Finally, i dedicate this work to brothers, sisters, relatives, friends, mentor and classmates who shared their words of wisdom and encouragement to finish the study.

Acknowledgements

Foremost, I would like to express my sincere gratitude towards supervisor Dr. Junaid Imtiaz who was there for me when i needed him. He guided me through thick and thin in our practical life .He provided me with continuous support. I could not have imagined a better supervisor and mentor for this study.

Fahad Hassan Zaman

Bahria University Islamabad

SEPTEMBER 2021.

Abstract

Knowledge is one of the key elements through which a provides a familiarity, understanding and awareness of different facts, skills, and entity. Unfortunately, nowdays knowledge institutes are operational just for the sake of business and have been not able to provide the quality education and assessing methods. Recently there was a discovery in this field by proposing a system of setting program learning outcomes (PLO) and course learning outcomes (CLO) for a course. If a student earns 50% in every CLO he/she is considered passed in particular course. This is a good system but it has few flaws. This system does not provide a mechanism in which it could provide future predictions of CLOs/PLOs of different students, to analyze how many students will pass or fail a particular CLOs/PLOs. This thesis provides solution to the problem by proposing a system which is capable of predicting future results of CLO and PLO of a specific course taught by specific teacher.The system is designed in such a way that there are 6 dedicated KPIs.The data of the students of a course are entered and with the aid of machine learning algorithms the respective scores of the KPIs are calculated. If some course is getting 90% marks from past two to three years or course is getting less than 50% marks from past 2 to 3 years then Analysis can be performed by the department and further actions can be taken to enhance the enviornment of learning. The proposed system is implemented on Python is used widely for machine learning.

Contents

List of Figures

List of Tables

Abbreviations

PEOs Program Educational Outcomes

PLOs Program Learning Outcomes

CLO Course Learning Outcomes

OBE Outcome-Based Education

PDCA Plan-do-check-act

OBTL Outcomes-Based Teaching Learning

ILOs Intended Learning Outcome

TLAs Teaching and Learning Assistant

ATs Assistive Technology

SVM Support Vector Machine

RF Random Forest

LR Logistic Regression

KNN K-nearest neighbors

SLN Social Learning Networks

AUCs Area under curve

ELM Extreme Learning Machines

WS Weighted score

EX Examination score

ONL Online

ONC On-Campus

AI Artificial Intelligence

BKT Bayesian Knowledge Tracing

NN Neural Networks

MOOC Massive Open Online Courses

EDM Educational Data Mining

LA Learning-Analytics

LMS Learning Management System

KPI Key Performance Indicator

Chapter # 1

Introduction

As instructors and education enthusiasts utilize the term ’concept of knowledge’ alludes to the data that educator instructs, and understudies are required to discover in each subject or substance region. Knowledge or information is said to current realities, ideas, speculations, and rules that are instructed. It is anything but connected with perusing, composing, or exploring that student learn in educational courses. It is normal mistaken for training. Training alludes to the information gained through various schools, universities, and other instructive organizations. Schooling is a spine for the improvement of any country. In this way, hence, the advancement of any country relies upon the higher condition of its schooling framework. It creates individuals intellectually socially, and profoundly. The nations with proficient instruction frameworks also end up being the heads of the planet, both socially and financially.

These days, the world faces an extreme danger of inferiority, weakness, political unsteadiness, wretchedness, psychological warfare, sexual orientation segregation, and heaps of additional. These issues are because of the absence of mindfulness, resistance, and ignorance, created by an ineffectual educational framework. Companies have invested in building educational institutes just for the sake of business and have been fooling people by not providing the quality education and not accurately assessing children thus creating a massive chaos of unemployment in the world.

To overcome this major issue this thesis presents a solution in way how to assess students at any educational institute. As we know in the OBE systems, we have PEOs (Program Educational Outcomes), PLOs (Program Learning Outcomes) and CLO (Course Learning Outcomes) we will discuss them and how they are managed in this proposed system. Program Educational Objective (PEO) is the goal that the program is getting ready their alumni for their vocation and expert life. Program Learning outcomes (PLO) are explicit statements of what the graduates will have the option to do when they effectively complete a learning experience. Course Learning outcomes (CLO) are quantifiable explanations that solidly state what students are required to learn in a course. Whatever a student does in laboratories, classroom, assignments, and quizzes everything is mapped to its respective CLO. Similarly, a specific CLO is mapped to a particular PLO and likewise PLO is mapped to respective PEO.

There are 4 PEOs:

PEO 1:-Professional Employment

Discover work related to EE inside plan, improvement, exploration, activities, support, specialized deals, and promote and investigate business ventures and secure positions in different regions like business, law, NGOs, media, etc.

PEO 2:- Technical Competence

Show specialized ability in electrical designing by discovering answers for complex issues, planning new items, and utilizing their logical, designing, and critical thinking abilities to offer some benefit to their industry.

PEO 3:- Professional Growth

Seek after their expert development by taking up higher investigations for postgraduate educations, learn innovations as they arise, foster abilities in the utilization of new devices, embrace professional improvement courses, and keep themselves current in their picked specialization.

PEO 4:- Social Engagement

Work in multicultural groups, give the initiative in their space; be touchy to moral, good, natural, sex, and cultural issues, and leave their work on the general public and the local area.

Planned on these PEOs are 12 PLOs, which are:

PLO 1:- Engineering Knowledge

An ability to apply information on math, science, design basics, and a designing specialization to the arrangement of complex designing issues.

PLO 2:- Problem Analysis

A capacity to recognize, form, research writing, and investigate complex designing issues arriving at validated resolutions utilizing first standards of math, regular sciences, and design sciences.

PLO 3:- Design/Development of Solutions

A capacity to plan answers for complex designing issues and plan frameworks, segments, or cycles that address indicated issues with suitable thought for general wellbeing and security, social, cultural, and natural contemplations.

PLO 4:- Investigation

A capacity to systematically research complex designing issues, including writing overview, plan and direct of examinations, investigation, and understanding of exploratory information, is an amalgamation of data to determine legitimate ends.

PLO 5:- Modern Tool Usage

A capacity to make, choose and apply proper procedures, assets, and current designing and IT instruments, including expectation and demonstrating, to complex designing exercises, with a comprehension of the impediments.

PLO 6:- The Engineer and Society

A capacity to apply thinking educated by logical information to survey cultural, wellbeing, security, lawful, and social issues and the following duties pertinent to professional designing practice and answer for complex designing issues.

PLO 7:- Environment and Sustainability

A capacity to comprehend the effect of expert designing arrangements in cultural and ecological settings and show information on and need for a maintainable turn of events.

PLO 8:- Professional Ethics

Apply moral standards and focus on proficient morals and duties and standards of designing practice.

PLO 9:- Individual and Teamwork

A capacity to work viably, as an individual or in a group, in diverse or multidisciplinary settings.

PLO 10:- Communication

A capacity to impart viably, orally just as recorded as a hard copy, on complex designing exercises with the local designing area and society everywhere. For example, having the option to appreciate and compose compelling reports and plan documentation, make powerful introductions, and give and get clear directions.

PLO 11:- Project Management

A capacity to exhibit the board abilities and apply designing standards to one's work and a pioneer in a group to oversee projects in a multidisciplinary climate.

PLO 12:- Lifelong Learning

A capacity to perceive the significance of and seek after long-lasting learning in the more extensive setting of advancement and innovative turns of events.

1.1 Problem Description

As we can see Outcome Based Education (OBE) system is implemented throughout the world. The Problem statement is that there is no real time monitoring or future prediction of a PLO /CLO scores. Also there arises another problem from the faculty that if a greater number of students have failed a particular PLO/CLO then it is up to them to check if the problem is from their side by checking and revising a PLO.

1.2 Thesis Objectives

To provide an instrument to department which can identify, monitor, and notify students about their PLO score, CLO score. OBE systems is one of the best education systems to overlook at a student, but it lacks the real time tracking or future prediction of scores of a PLO/CLO. To overcome this issue this thesis provides a solution in which the PLO/CLO scores will be predicted from the start of the course using Machine Learning algorithms. So, According to result department can amend specific clos or plos.

1.3 Thesis Organization

Thesis layout includes Introduction, Literature review, Methodology, Evaluation, and conclusion and future work. Chapter 1 of the thesis is Introduction which includes background, problem description, objective and layout of thesis. Chapter 2 of thesis is Literature review which include overall research related to our thesis. Chapter 3 include Methodology which include our methodology which we use in our thesis like different classifiers. Chapter 4 of thesis is Evaluation in which my numerical results would be include and in last Chapter 5 it would include conclusion and Upcoming work.

Chapter # 2

Literature Review

This chapter discusses the literary work done on the ”PREDICTION OF KPIs RELATED TO CLOs/PLOs”.

2.1 Outcome-Based Education (OBE) Framework

Outcome-Based Education (OBE) is an illuminating theory that bases each piece of an enlightening design encompassing what is critical for all understudies to decide to do effectively near the fulfilment of their learning encounters.

For the educational system to work reasonably, the OBE structure is perceived. It guarantees that instructive plans, educating and learning strategies, and evaluation contraptions are steadily updated through appraisal communication. The framework P-D-C-A (plan-do-check-act) cycle has been applied for ideal feasibility and efficiency.

The framework gets the OBTL (Outcomes based teaching learning) execution which twirls around three huge segments:

- Portrayal of the proposed learning results (ILOs) as an activity word (learning development), its thing (the substance), and assurance of the exceptional circumstance and a standard the understudies are to accomplish.
- Setting up a learning environment using training/learning environment works out (TLAs) that address that activity word and thusly are presumably going to accomplish the normal outcome.
- Using examination tasks (ATs) that similarly contain that activity word, as such enabling the teacher to choose with the help of rubrics if and how well understudies’ displays meet the principles.

- It helps in making the goods which cannot be professionally produced by hands of men.

The execution of outcome-based guidance, which propels the demonstration of a supportive game plan between results, learning activities, and examination gadgets need an environment where all accomplices (instructors, understudies, and the associations) are busy with the pattern of phenomenal reflection and consistent movement. All of these individuals reflect regarding the others in three spaces: teacher and understudy, instructor and association, understudy and foundation that would have intrinsic quality improvement and segments for ensuring quality just as for updating quality.Building a learning neighbourhood redesigns the obligation regarding orchestrating and clever practice among its staff will set up new opportunities for critical talk among peers, and empower the total undertakings of the association is responding to the premium of duty from accreditation workplaces similarly as the public solicitation about the idea of training and learning in high-level training. OBE framework is in Fig2.1. [1]

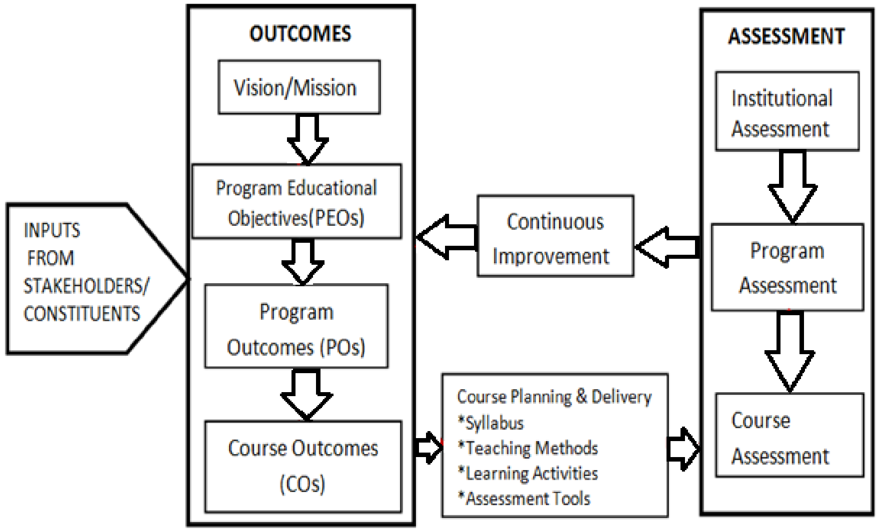

Figure 2.1: OBE Framework

2.2 Principle of OBE

There are various definitions for OBE training. The most broadly utilized one is the four-level training recommended by Spady in 1994.

An OBE instructive arrangement infers starting with an undeniable picture of what is significant for students to have the choice to do, by then planning the instructive arrangement, evaluation, and direction to guarantee this adjusting event happens. The following are the four levels as described by Spady in his study:

Clarity: -

Outcome based education system generally focuses on results. Students can comprehend what they expect, and instructors can comprehend what they must show throughout the course. Lucidity is significant as it is important to be clear in every classification or in all levels, with the goal that students can progress, and furthermore to depict all the information and capacities important to achieve this result.

Flexibility: -

Instructors can structure their classes as per the student's desires by perusing straightforwardly what should be finished. Hence, OBE determines no specific instructional procedure, educators can show any strategy of exploitation. They will even have the option to acknowledge variety among students by exploitation various instructing and appraisal strategies in OBE since it is a student-focused learning model. Educators will encourage students to get a handle on the ideas in any way (study aides, and group work, class) that encourage students learning.

Investigation: -

In OBE lecturers will examine the outcomes a student has achieved and in which region they are moved up to dissect the ability and give singular help and guiding to fulfil their needs. This helps educators and institutions. They likewise help educators to screen the turn of events and upgrade of the student over a specific amount and to assist them with accomplishing their outcomes.

Contribution: -

Student support in an institution is likewise a significant part of OBE. Swelled student contribution licenses students to believe to fault for their own learning, and that they will get familiar with a ton through this individual learning.

2.3 Components of OBE Framework

The idea of OBE takes on more significance by thinking about the different segments of OBE and their birthplaces. These include goals, results, standards referred to as estimation, dominance learning, responsibility, and competency-based schooling.

2.4 Blooms Taxonomy

In 1956, Benjamin Bloom, an American instructive clinician, driven a gathering of instructive therapists to build up a scientific categorization, or grouping framework, for learning. He recommended that learning finds a way into one of three mental spaces:

- The Cognitive area – preparing data, information, and mental abilities.
- The Affective area – Attitudes and sentiments.
- The Psychomotor area – manipulative, manual, or actual abilities.

2.5 Focus and Benefits of Adoption of OBE

Key inquiries accompanying with OBE

- What understudies(a person who learns another's role in order to be able to act at short notice in their absence) we require or what available choice we have to do?
- What can be the best options that can help understudies for achieving?
- How well we can assess whether understudies have achieved it?

- What steps can be taken to close the circle for extra improvement (Continuous quality improvement CQI)?

Benefits of OBE:

- More coordinated and intelligible educational plan.
- Graduates will be more ”applicable” to industry and different partners (all the more balanced alumni).
- Constant Quality Improvement (CQI) is set up.

2.6 Educational Data Mining and Techniques

Instructive Knowledge Discovery in Data is an arising discipline, worried about creating strategies for investigating the unique and progressively enormous scope of information from instructive settings and utilizing those techniques to get understudies and the settings more readily they learn. Regardless of whether instructive information is taken from students of different background, utilization of intuitive learning conditions, PC upheld community-oriented learning or authoritative information from schools and colleges, it frequently has various degrees of the significant pecking order, which regularly should be controlled by properties of the actual information, instead of ahead of time. Issues of time, grouping, and setting likewise assume significant parts in the investigation of instructive information.it is shown in Fig 2.2.

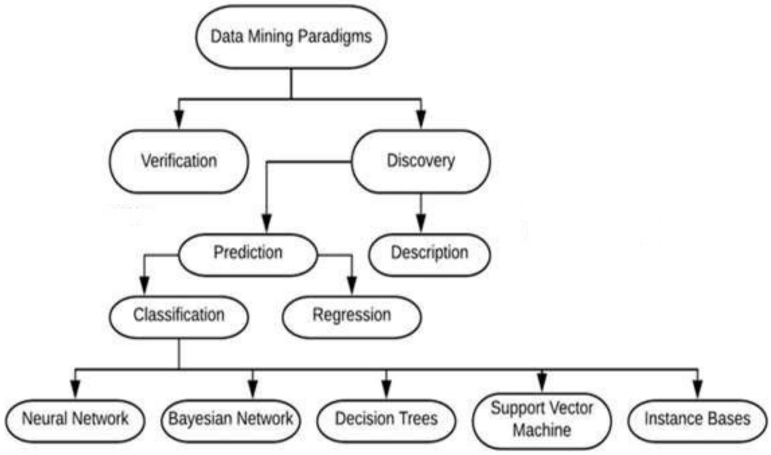

Figure 2.2: Data Mining Techniques

2.7 Brief Explanation of Proposed Prediction Framework (Applied Algos KNN, SVM, RF, LR)

We are using four different algorithms in this thesis which are KNN, SVM, RF and LR.

A **Support Vector Machine (SVM)** is an overseen AI figuring that can be used for both arrangement and relapse purposes. SVMs rely upon discovering a hyperplane that best parcels a dataset into two classes.

KNN is a non-parametric system used for regression and classification. It is conceivably the most basic ML methodology used. It is a sluggish learning model, with a local gauge.

Random forests are a group learning strategy for characterization, relapse, and different errands that work by developing a massive number of decision trees or we can say decisions required at a time, at preparing time and yielding the class that is the method of the classes or mean/average forecast of the individual trees.

It is like linear regression, **Logistic regression** is the correct calculation, to begin with, classification algorithms. Albeit the name 'Regression' comes up, it's not a Regression model, anyway an older model. It uses a determined ability to lay out a twofold yield model. Simple, quick, and basic arrangement technique and Can be utilized for multiclass classification moreover.

This thesis provides solution to the problem by proposing a system which is capable of predicting future results of CLO and PLO of a specific course taught by specific teacher.The system is designed in such a way that there are 6 dedicated KPIs.The data of the students of a course are entered and with the aid of machine learning algorithms the respective scores of the KPIs are calculated. If some course is getting 90% marks from past two to three years or course is getting less than 50% marks from past 2 to 3 years then Analysis can be performed by the department and further actions can be taken to enhance the enviornment of learning. The proposed system is implemented on Python is used widely for machine learning.

2.8 Related Research Work

In this section We will talk about literature review related to this thesis.
In [2] this paper informs us concerning Data mining that assumes a significant part in the business world and it serves the instructive organization to foresee and settle on choices identified with the understudies' scholarly status. With advanced education, presently a day exiting understudies' has been expanding, it influences the understudies' profession as well as on the standing of the establishment. The current framework is a framework that keeps up the understudy data as mathematical qualities and it simply stores

and recovers the data that it contains. So the framework has no knowledge to investigate the information. The proposed framework is an electronic application that utilizes the Naive Bayesian digging method for the extraction of helpful data. The trial is directed on 700 understudies' with 19 ascribes in Amrita Vishwa Vidyapeetham, Mysuru. Result demonstrates that Naïve Bayesian calculation gives more exactness over different techniques like Regression, Decision Tree, Neural organizations, and so on, for correlation and expectation. The framework focuses on Bayesian and the framework which keeps up all understudy confirmation subtleties, course subtleties, subject subtleties, understudy marks subtleties, participation subtleties, and so forth It accepts understudy's scholarly history as info and gives understudies' impending exhibitions based on semester.

In [3] this paper, author tell us about study learning result expectations for online courses. While earlier work has zeroed in on semester-long courses with continuous understudy appraisals, we center around short courses with single results doled out by educators toward the end. The absence of execution information and by and large little enlistments makes the conduct of students, caught as they interface with course content and with each other in Social Learning Networks (SLN), fundamental for a forecast. Our technique characterizes a few (AI) highlights dependent on handling practices gathered on (human) learning modes in a course and using them in fitting classifiers. Through assessment on information caught from three fourteen-day courses facilitated through our conveyance stages, we mention three key observable facts:

- Behavioral information contains signals prescient of learning results in short-courses (with classifiers accomplishing AUCs ≥ 0.8 after the fourteen days).

- Early identification is conceivable inside the primary week (AUCs $\geq$ 0.7 with the principal seven-day stretch of information).
- The substance highlights have a "most punctual" location ability (with higher AUC in the initial not many days), while the SLN highlights become the more prescient set over the long run as the organization develops.

We likewise talk about how our strategy can create conduct investigation for teachers.

In [4] this paper Outcome-Based Education (OBE) is the key fundamental piece of instructive associations. Result based instruction framework is a significant advance for accreditation. OBE puts centre around understudy-focused methodology. This paper is an endeavor to give the structure to robotize the achievement interaction utilizing the evaluation and planning information of Program Outcomes (POs) and Course Outcomes (COs) recovered from backhanded appraisal devices. There are different kinds of reviews like graduated class study, course inserted study, and modern study. An overview is directed to check the degree of fulfillment of understudy as to showing learning measure. The inquiries of overview are being planned with COs and in this manner with POs. At that point, various inquiries are posed by the instructor to the understudies. In this manner, information is gotten for additional counts. Hereafter, with the assistance of estimations, fulfillment levels of the Cos are determined which is additionally utilized in the figurings of the POs. At that point in the end accomplishment of various POs is acquired. Moreover, processed accomplishments are utilized for expectation utilizing Weka apparatus and relating to that various classes are related with that information. Thusly, one can improve the exhibition of the understudies by taking a gander at the various qualities, on the whole, the courses and henceforth it prompts the general advancement of the understudies. Finally, the characteristics which are gotten in the data mining table can be used as further development for different thoughts which brief the overall improvement of OBE.

In [5] with the fast headways in innovation, Massive Open Online Courses have become the most mainstream type of online instructive conveyance, to a great extent because of the evacuation of topographical and monetary obstructions for members. An enormous number of students internationally take a crack at such courses. Regardless of the adaptable openness, results demonstrate that the fulfillment rate is very low. Instructive Data Mining and Learning Analytics are arising fields of exploration that plan to upgrade the conveyance of schooling through the utilization of different measurable and AI draws near. A broad writing overview shows that no huge examination is accessible inside the space of MOOC information investigation, specifically thinking about the standards of conduct of clients. In this paper, hence, two arrangements of highlights, in light of student standards of conduct, were thought about regarding their appropriateness for anticipating the course result of students partaking in MOOCs. Our Exploratory Data Analysis shows that there is a solid connection between snap transfer activities and fruitful student results. Different Machine Learning calculations have been applied to improve the precision of classifier models. Re-enactment results from our examination have shown that Random Forest accomplished practical execution for our expectation issue, acquiring the best of the models tried. Alternately, Linear Discriminant Analysis accomplished the most minimal relative execution, however addressed just a peripheral decrease in execution comparative with the Random Forest.

In [6] this paper An instrument was developed from previous international experiences to exactly target the knowledge skills and attitude of the students. The three factors knowledge, skills and attitude were kept in mind and two summer programs at university H was introduced to check the results of the instrument. Program A and Program B were introduced in which there were a mix of different students which belonged to a different culture, respectively. The students participated in different activities including lectures, seminars, field trips, study tours under the supervision of the professors at University H. The students after going through a 15-day course were surveyed and then interviewed and from the results 63 learning outcomes were derived with 14 falling in knowledge dimension, 18 in skills dimension and 28 in attitudes dimension.

This [7] paper A computationally proficient artificial consciousness (AI) model called Extreme Learning Machines (ELM) is received to dissect designs inserted in constant appraisal to demonstrate the weighted score (WS) and the assessment (EX) score in designing science courses at an Australian territorial college. The understudy execution information assumed control over six years in different courses going from the mid-to the high level and a mixed course offering mode (i.e., nearby, ONC, and on the web, ONL) are demonstrated by ELM and further benchmarked against contending models: irregular woodland (RF) and Volterra. With the appraisals and assessment marks as key indicators of WS (prompting an evaluation in the mid-level course), ELM (as for RF and Volterra) beat its partner models for the ONC and the ONL offer. This created a relative expectation blunder in the testing stage, of just 0.74%, contrasted with about 3.12% and 1.06%, individually. At the same time, for the ONL offer, the forecast mistakes were just 0.51%, contrasted with about 3.05% and 0.70%. In demonstrating the understudy execution in cutting-edge designing science courses, ELM enlisted somewhat bigger blunders: 0.77% (versus 22.23% and 1.87%) for ONC and 0.54% (versus 4.08% and 1.31 %) for the ONL offer. This investigation advocates a pioneer execution of a robust AI strategy to uncover connections among understudy learning factors, creating instructing and learning mediation and course wellbeing checks to resolve issues identified with graduate results and understudy learning ascribes in the advanced education area.

This [8] paper centres around the issue of nonstop enhancement for instructive results and takes the designing mechanics course, as an illustration, to assist understudies with beating their learning troubles. A choice emotionally supportive network dependent on information mining and fluffy rationale is proposed to anticipate the understudy learning results. The approaches include four stages: a fluffy hypothesis to recognize the components of learning results; information mining to build impact chart; AI to set up the fluffy induction relations; and the model to anticipate the test scores toward the start obviously and in this way to help understudies improve their scores as indicated by their shortcoming.

This [9] paper tell us about the students of Master of Telecommunication engineering were assessed based on their individual personal performance, micro group and macro group. They were asked to make a certain project which had many subcategories like Transmitter creation, receiver creation and systems. The students were equally divided into 3 groups and their performance in group and individual were assessed by the TAs respectively. After the project deadline the students were asked to present their projects and then they were survey regarding the project.

This [10] paper examines about Learning result evaluation is of extraordinary importance in the field of customary nearby showing particularly on the courses of programming dialects. In this work, we exploit the information offered by our programming task judge framework and propose another item Response Theory-BKT model for assessment of learning result. This new structure: Item Response Theory model that appraises understudies' underlying information status, and goes along with it with the separation and trouble of every ability assessed to assess the likelihood of knowing an expert before preparing it. We at that point gauge boundaries learn, conjecture, and slip probabilities of the Bayesian Knowledge Tracing (BKT) Model. Utilizing genuine information, we show that the Item Response Theory-BKT model beats Item Response Theory and Bayesian Knowledge Tracing regarding expectation precision.

In [11] this paper, talks about probed creating forecast models for understudy execution in the beginning phases of mixed learning courses utilizing profound neural organization engineering and using on the web action credits as info designs. The online action credits were extricated from the action logs put away by Moodle. A sum of 885 records from college understudies taking three 3 distinct courses under 16 unique classes was used. Initially, a progression of trials was directed to decide the hyperparameters for a top-performing NN model which at that point filled in as a gauge classifier. A while later, tests were directed to test the presence of the model for anticipating understudy results (pass or fall flat) both for the midterm and finals period utilizing action information created preceding the midterm period. Results demonstrate that lone low expectation execution can be accomplished at a beginning phase, all the more explicitly during the primary month of the course.

Nonetheless, both precisions, just as ROC_AUC score, improves as more information is amassed up to the third month. This outcome upholds the discoveries from past investigations. The most noteworthy exactness accomplished for anticipating finals results for a solitary course is 91.07% with a ROC_AUC score of 0.88, while for midterm results, the most elevated is 80.36% precision with a ROC_AUC score of 0.70. This investigation is a continuous work that means fostering a device that can be applied in chosen mixed learning measurements to give a premise to programmed criticism and educator support.

In [12] this paper Massive Open Online Courses have emerged as a choice rather than the traditional informative system by the versatility in timings. Besides, it overcomes the financial and land deterrents for the customers. MOOCs moreover help understudies from arranged establishments to confer and exchange data MOOCs conversations. The quantity of understudies enrolling for such courses is high; notwithstanding the limitless accessibility, the completion rate is meager. Various factors impact the course completion by the understudies, like interest in the subject, justification evaluating the subject, and whether or not the educator can understand the understudies. EDM and LA are the fields where understudies' learning development is analyzed to procure certain fundamental information or can be used in conjecture using EDM instruments and techniques. Data examination shows a strong association between the number of events, for instance, click event, video watched conversation post, and the productive understudy's outcome. Computer-based intelligence computations are applied, and the result shows that the Decision Tree gives an optimal result with the best.

In [13] this paper Teaching is an astounding activity that anticipates educators using the best and capable training strategies to engage understudies to make strides. The essential issue in preparing instructors should consider different appearance approaches and learning strategies to suit each understudy. Today, in the PC age, electronic learning (e-learning) is by and large used basically. The headway of the World Wide Web, especially Web2.0, has incited disturbance in tutoring. Understudy collaboration with Learning the board structures - LMS achieves making colossal instructive assortments that are charming for research. LMS structures offer gadgets +following every individual understudy and quantifiable view for the more significant inspecting eventual outcome of an understudy - system correspondence. Regardless, these instruments do reject artificial thinking estimations as a supportive part of the decision. This article offers design to understudy exhibiting arranged on colossal courses of action of data using Hadoop and Mahout. This kind of structure would give information into each one understudy's development. Considering that information, instructors could shift direction materials as demonstrated by understudy revenue and data.

This [14] paper tell us about Understudy execution forecast is vital to comprehend an understudy progress rate. It is said that 'Anticipation is superior to the fix'. In this research, we are attempting to discover the understudy's flow status and anticipate his/her future outcomes. After the result, instructors can give him/her legitimate counsel to stay away from the helpless outcome and furthermore can prepare the understudy. By discovering the conditions for definite assessments. Which courses he/she ought to take in the forthcoming semester (parts of counsel/instructor). Consistently a ton of understudies linger behind due to the absence of legitimate counsel and checking. An instructor can't screen every single understudy without a moment's delay. On the off chance that a framework can help a Teacher about the understudies like which understudy needs which sort of help. At that point, it will be a lot of accommodating for the two educators and understudy. The point is assisting the understudy with evading his/her anticipated helpless outcome utilizing Artificial Intelligence. In the event that an understudy could understand what will be his/her outcome later on and inform him/her what to do to evade his/her awful outcomes by anticipating the last assessment mark. This exploration would be useful for the understudies and educators with The most noteworthy exactness of 94.88%.

This [15] paper tell us about that a huge measure of computerized information is being produced across a wide assortment of fields and Data Mining

(DM) strategies are utilized to change it into helpful data in order to recognize covered-up designs. One of the actual locales of EDM use is the movement of understudy execution check models that would expect the understudy's show in enlightening foundations. We assemble a model which can advise understudies (in basic programming course) about their plausible results at the beginning phase of the semester (when assessed for 15% evaluations). We applied 11 Machine Learning calculations (from 5 classifications) over an information source utilizing WEKA and inferred that Decision Tree (J48) is giving higher exactness regarding effectively-recognized occurrences, F-Measure rate, and genuine positive location. This investigation will serve the understudies to distinguish their likely last assessments and replace their scholarly conduct to accomplish good marks.

In [16] this paper the staff of civil-engineering at University-of-Teknologi-MARA in 2007 implemented the OBE system for its new batch because they wanted to measure the program learning outcome of their graduates. Most of the skills including cognitive, psychomotor, and effective could be measured during the student's time at the university. They implemented a system in which it contained 7 PLOs which were mapped to different courses. Like Basic solid mechanics subject was given the PLO 1 and so on. This OBE system implemented by the University changed the perspective in teachers and students. It is anything but's an unrivalled impact on the staff in improving and invigorating the tutoring structure to end up being more satisfactory in noticing understudy headway.

In [17] this paper there was a comprehensive research about the impact of coursework, examinations, and grades. They presented a research in which they concluded a strategy known as Learning Outcome Assessment Model. It states that instead of taking midterm and final exams, the students will be assessed throughout the degree with assignments and quizzes. They proposed a strategy in which there were 4 PLOs with the weightage of assignments and quizzes 50% respectively. In this way students can be monitored throughout the course which is better than the conventional system in which the student only must study or focus on exams rather than assignment and quiz.

These were the some literature review We have gone through and find out that are relatively related to my research study.

Chapter # 3

Methodology

In this chapter, I will talk about methodology used in this thesis.

3.1 Proposed System

The following picture Fig 3.1 gives a pictorial representation of our system.

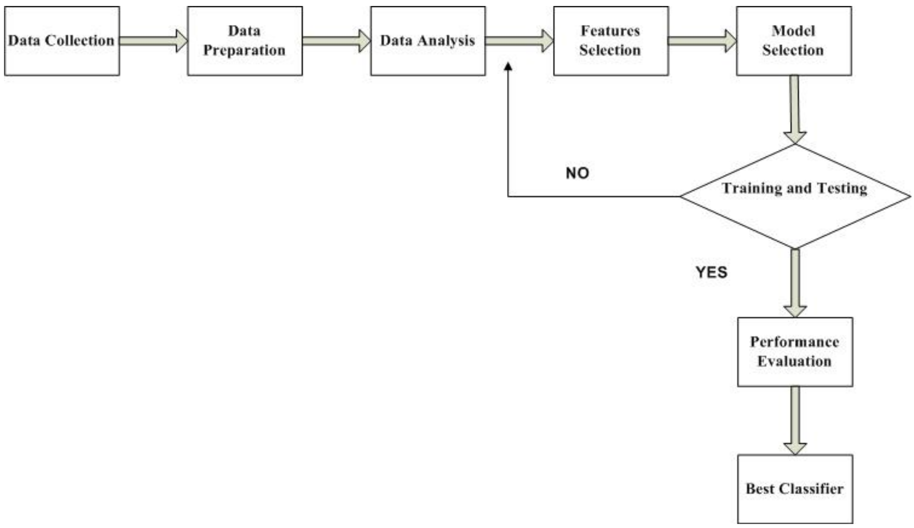

Figure 3.1: Flow Diagram

Our system will be provided to our end user. The teacher will insert data of the students. The data will be fetched to the machine learning algorithm which will then predict the future possible scores of subject in detail. Then It will be up to the teacher then to design a more competitive or more conceptual course according to the understanding of the students. All the graphs and results of the system will be shown in evaluation chapter.

Basically the KPIs mean the key performance indicators.There are 6 KPIs of the project which we are going to do in this thesis these KPIs are:

KPI 1:

It deals with how many percentages of students passed CLO1,CLO2,CLO3 with more than 50% marks.

KPI 2:

It deals with the average of CLO1,CLO2,CLO3 marks.

KPI 3:

It deals with how many percentages of students passed CLO1,CLO2,CLO3 with more than 60% marks.

KPI 4:

It deals with the average of PLO1,PLO3,PLO4 marks.

KPI 5:

It deals with how many CLOs are cleared with more than 50% marks out of the total 3 CLOs. For example: if CLO1 and CLO2 is more than 50% marks and CLO3 is less than 50% then the answer would be 2.

KPI 6:

It deals with how many CLOs cleared more than 60% marks out of the total 3 CLOs. For example: if CLO1 is more than 60% marks and CLO2 and CLO3 is less than 60% then the answer would be 1.

From now we will explain how this thesis would be done practically step by step.

Start with what basically machine learning is? Machine Learning is a logical method where the PCs figure out how to tackle an issue, without expressly program them. Profound learning is right now driving the ML race controlled by better calculations, calculation force, and enormous information. Still, ML old-style calculations have their solid situation in the field.

3.2 Data Collection and importing

Information arrangement is the way toward get-together and surveying data from unending various sources. To utilize the information we gather to make valuable artificial insight (AI) and AI blueprints, it should be collected and dealt with so much that searches useful for the business issue also. Fig 3.2 shows data of different students of my university, which fuses various features, including gender.

Year	Enrollment	Reg_No	Name	Gender	Quiz	Assignment	Mids
2019	01-133152-019	42632	Ammar Asad	Male	1.8	4.7	5
2019	01-133152-077	42681	M. Naveed Ul Hassan	Male	4.7	12.5	5
2019	01-133152-101	42699	M. Zubair Sadiq	Male	1.6	12.7	2
2019	01-133152-149	42740	Toheed Omer Paracha	Male	5.2	17.1	16
2019	01-133152-155	42746	Uzair Wali Dar	Male	4.7	10.9	4
Finals	**Total**	**CLO1**	**CLO2**	**CLO3**	**PLO1**	**PLO3**	**PLO4**
0	11.5	0.175676	0.051282	0.130435	0.175676	0.051282	0.130435
2	24.2	0.254054	0.2	0.30438	0.254054	0.2	0.30438
0	16.3	0.135135	0.187179	0.173913	0.135135	0.187179	0.173913
37	75.3	0.743243	0.743243	0.826087	0.743243	0.743243	0.826087
7	26.6	0.254054	0.254054	0.304348	0.254054	0.254054	0.304348

Figure 3.2: Data Collection and Importing

3.3 Data preparation

In AI, information planning is the way toward preparing information for the preparation, testing, and execution of a calculation. It's a multi-step measure that includes information assortment, cleaning and pre-processing, highlight designing, and naming. These means assume a significant part of your AI model, as they expand on one another to guarantee a model performs to assumptions. After data preparation we Analyse our data For that we used different kind of graphs and different relation between features.

3.4 Data Analysis Techniques

Following are the major techniques that we have used for data analysis.

3.4.1 Box Plot

A box plot is utilized for data examination, and graphical depiction of the data. It is a technique for graphically portraying get-togethers of mathematical information through their quartiles. It may also have lines extricating from the compartments showing fluctuation outside the upper and lower quartiles.As shown in fig3.3

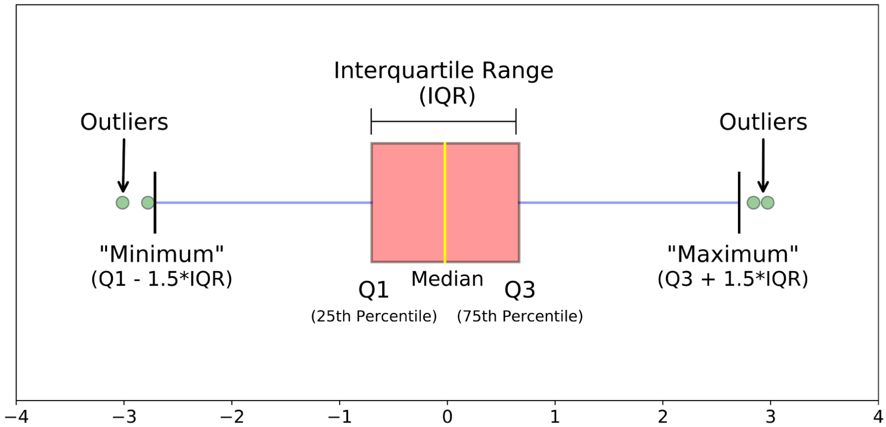

Figure 3.3: Box Plot

3.4.2 Histogram

For the estimated portrayal of the circulation of mathematical information, a Histogram is utilized. A histogram is a plot that permits you to discover, and show, the secret repeat transport (condition) of a lot of incessant data. This allows the examination of the data for its fundamental allotment (e.g., ordinary flow), exemptions, skewness, etc.

3.4.3 Scatter Plot

Scatter plot also known as scatter chart, is a plot that is used to show the cartesian coordinates for two sets of data. It basically shows us the relationship between two variables. It represent different numerical values.It show dot for every signle or individual value.

3.5 Features Selection

Feature selection techniques are expected to decrease the number of information factors to those that are accepted to be generally valuable to a model to foresee the objective variable. Feature selection is fundamentally centred around eliminating non-enlightening or excess indicators from the model.

Feature involve in this study are mid, internal and finals marks of students, Clos and Plos, and gender that whether the student is male or female. In simple words these are the input variables of the system.

3.6 Machine Learning Modelling

Create a new variable called result which says if a student is pass or fails. Which then will be used for classification problem. Assumption: If more than 50 marks then pass (1) otherwise fail (0).

3.6.1 One hot encoding

In AI, a one-hot is a get-together of pieces among which the legal mixes of characteristics are only those with a lone high (1) cycle and all the others low (0).

One hot encoding is a connection by which out and out factors are changed over into a construction given to ML estimations to make a predominant appearance in the figure. This is where another equal variable is added for each remarkable possible motivator for a tremendously worth.

- Year — Possible values are [2019, 2020]
- Gender — [MALE, FEMALE]

This is shown in fig 3.4.

Assignment	Mids	Final	Total	Clo1	Clo2	Clo3	Plo1	Plo3	Plo4	Result	Year_2019	Year_2020	Gender_FEMALE	Gender_MALE
4.7	5	0	11.5	0.175676	0.051282	0.130435	0.175676	0.051282	0.130435	0	1	0	0	1
12.5	5	2	24.2	0.254054	0.2	0.304348	0.254054	0.2	0.304348	0	1	0	0	1
12.7	2	0	16.3	0.135135	0.187179	0.173913	0.135135	0.187179	0.173913	0	1	0	0	1
17.1	16	37	75.3	0.743243	0.738462	0.826087	0.743243	0.738462	0.826087	1	1	0	0	1
10.9	4	7	26.6	0.254054	0.261538	0.304348	0.254054	0.261538	0.304348	0	1	0	0	1

Figure 3.4: One hot encoding

3.6.2 Correlation

A Pearson Correlation is a number between - 1 and +1 that shows to which degree 2 elements are straight related.

A value of 1 deduces that an immediate condition portrays the association between X and Y perfectly. All data centres lie around a line for which Y increases as X augmentations. A value of 1 derives that all data centres lie around a line for which Y lessens as X additions. A value of 0 deduces that there is no straight association between the elements Fig 3.5 shows that the feature ”Assignment” is highly correlated with other input features. This means that this feature will not cause any significant increase in the performance of our model/classifierand we can use all other features as input.

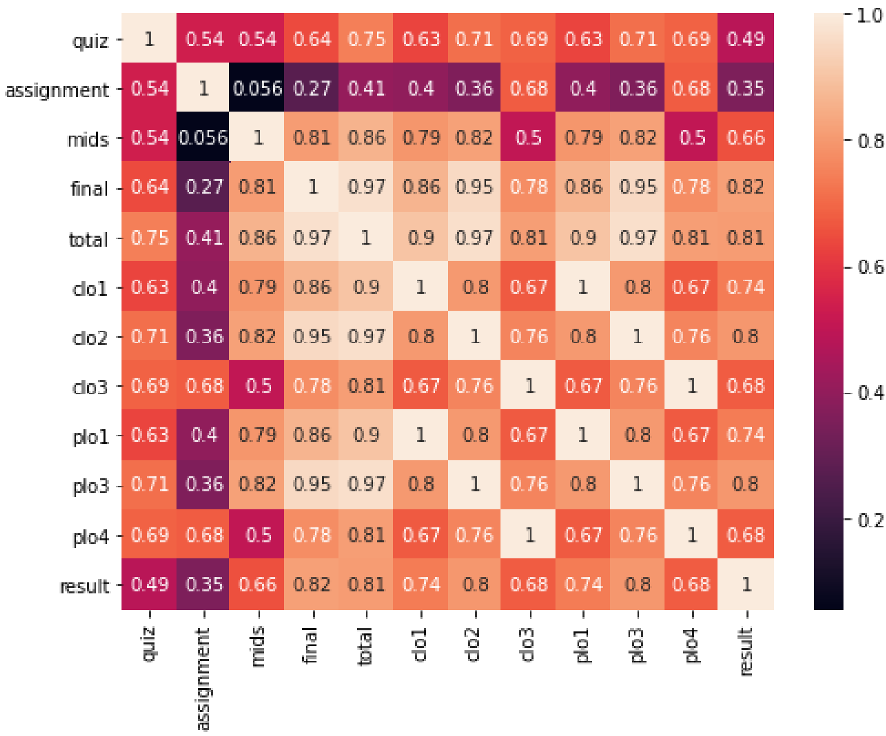

Figure 3.5: Correlation

3.7 Model Selection

After a relative report over various AI-managed strategies the procedures utilized are Linear Regression, Logistic Regression, K closest neighbours, and Random Forest and Support Vector Machine.

We are using four different algorithms in my thesis which are KNN, SVM, RF and LR.

3.7.1 Support Vector Machine

SVM or Support Vector Machine is an overseen AI figuring that can be used for both arrangement and relapse purposes. SVMs rely upon discovering a hyperplane that best parcels a dataset into two classes. As shown in fig3.6

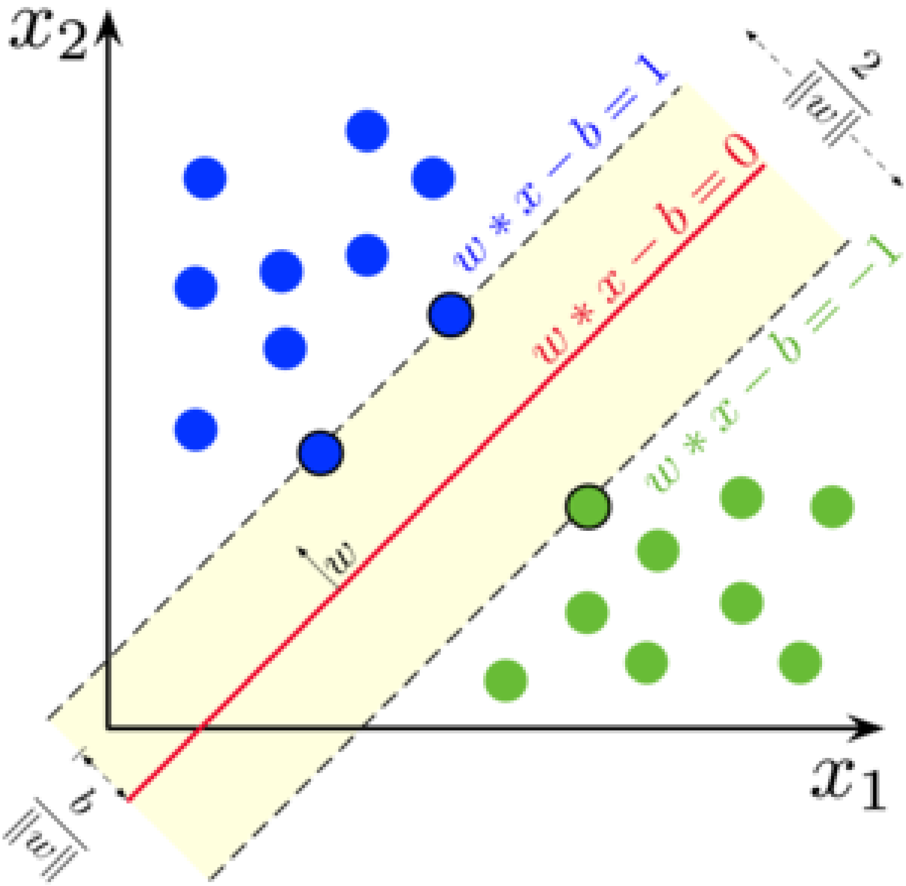

Figure 3.6: Support Vector Machine

3.7.2 KNN

KNN is a non-parametric system used for regression and classification. It is conceivably the most basic ML methodology used. It is a sluggish learning model, with a local gauge.As shown in fig3.7

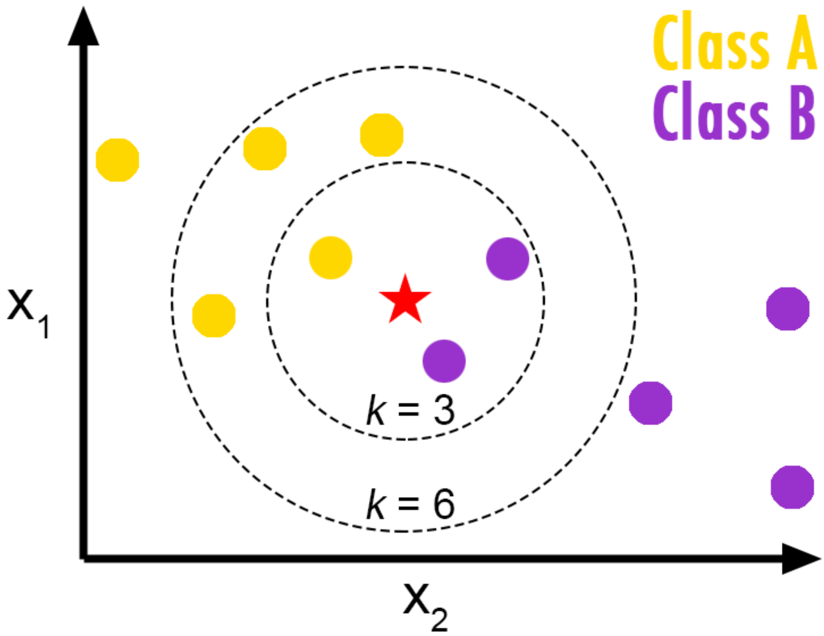

Figure 3.7: KNN

3.7.3 Random forests

Random forests are a group learning strategy for characterization, relapse, and different errands that work by developing a massive number of decision trees or we can say decisions required at a time, at preparing time and yielding the class that is the method of the classes or mean/average forecast of the individual trees.As shown in fig3.8

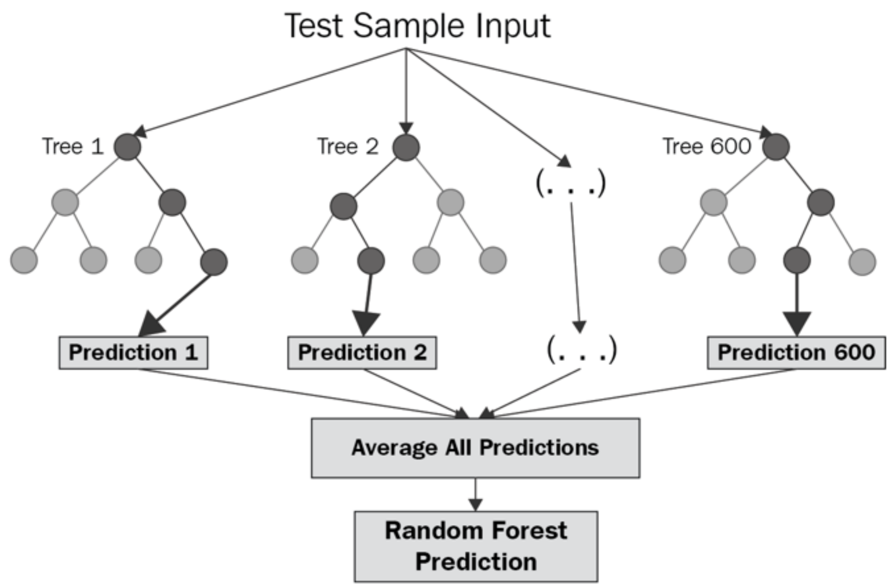

Figure 3.8: Random forests

3.7.4 Logistic regression

It is like linear regression, Logistic regression is the correct calculation, to begin with, classification algorithms. Albeit the name 'Regression' comes up, it's not a Regression model, anyway an older model. It uses a determined ability to lay out a twofold yield model. Simple, quick, and basic arrangement technique and Can be utilized for multiclass classification moreover.

3.8 Training and Testing

After that we divide our dataset in to testing and training dataset. We take 70 to 80% of data as training dataset and remaining 20 to 30% dataset as testing dataset.Mostly this comparison is used for better results. so after training our dataset we test the data to check that how much accurate our model gives answer base on that testing.

Every classifier has it's own properties we used 4 different classifier to check which classifier is best in use for us. According to our data RF gives us 100% accuracy.

Chapter # 4

Evaluation

4.1 Data Analysis

Following are the major techniques that we have used for data analysis.

4.1.1 Box Plot

Below, We have created a box plot to depict the quartiles of the score obtained by male and female students. We place gender on x-axis and total marks on y axis and also for differentiating we used two different colours of box plot to analysis our data. Fig 4.1 shows that relation between gender and no. of students.

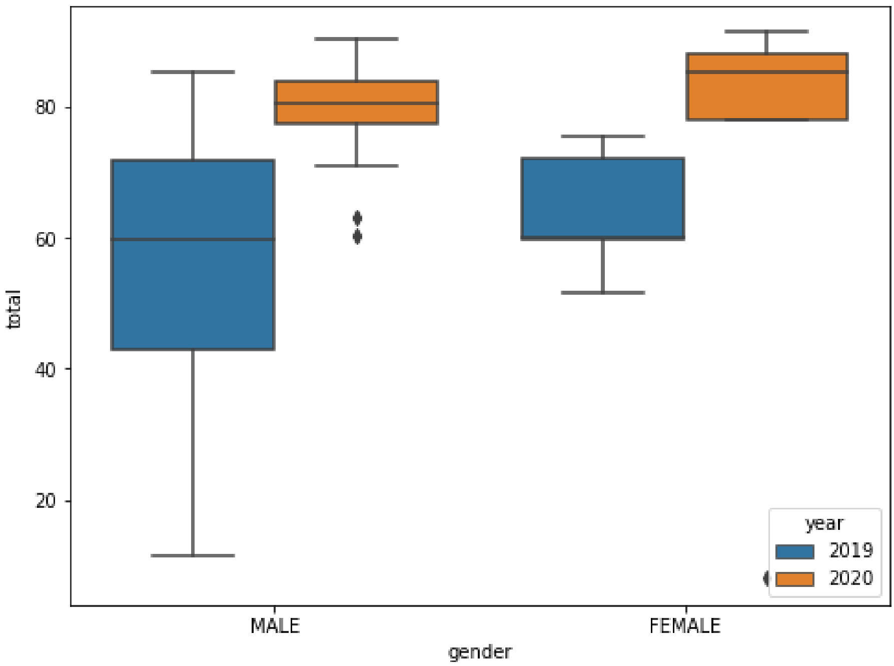

Figure 4.1: Relation between gender and no. of student

From the above graph we can safely say that average total marks of both gender students have increased.

4.1.2 Histogram

In figure 4.2 and Table 4.1 a histogram is constructed for the total marks obtained by the students in the year 2019 and 2020. As you can see from histogram below, the average marks (approx. 56) in 2019 has increased (to approx. 80) in 2020.

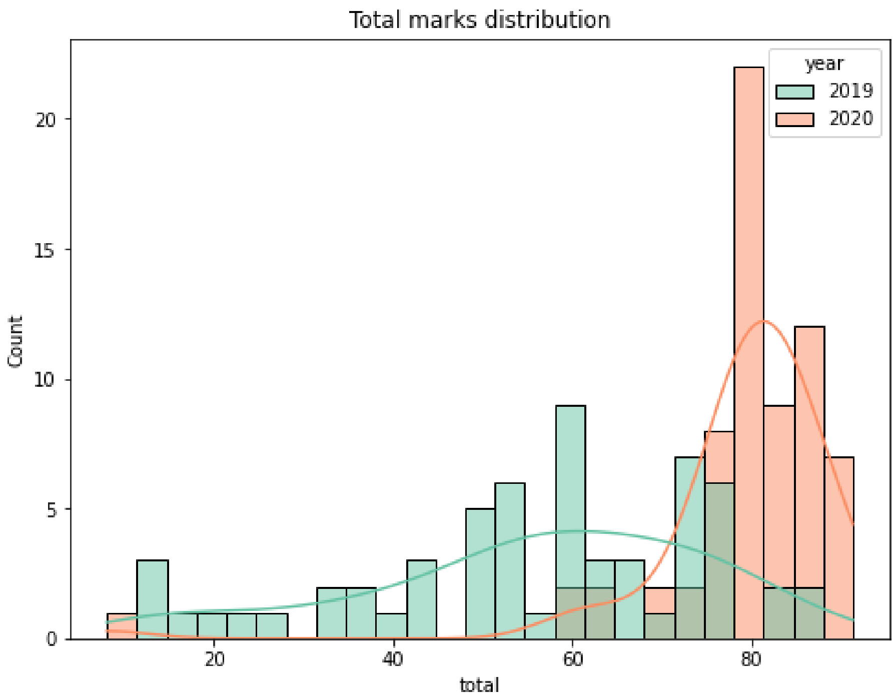

Figure 4.2: Relation between Total marks and count

year	Mean
2019	55.720833
2020	79.274478

Table 4.1: Mean value of 2019 and 2020

In Fig 4.3, I have constructed 3 histograms for each of CLOs in the year

2019 and 2020. As you can see from histogram below, the average marks in 2019 has increased in 2020 for all the CLOs.

Now to Check if it is only average has increased or all the quartiles (25%, median and 75%) are also increased? From the box plots below.

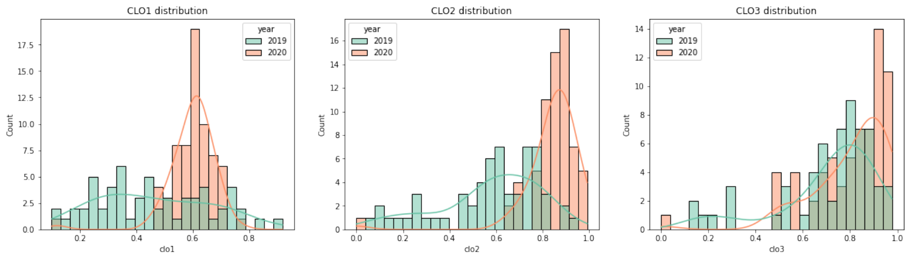

Figure 4.3: Relation between CLOs and count

According to CLO1

- Lower threshold has drastically increased to more than 0.4 in 2020 compared to 2019 which was less than O.2
- 25% quartile is increased in 2020.
- Median marks are also increased a bit in 2020 compared to 2019
- Where as the upper threshold decreased a bit in 2020.
- There is an ignorable change in 75% quartile

According to CLO2

- Lower threshold has drastically increased to more than 0.7 in 2020 compared to 2019 which was less than O.2
- 25% quartile, median marks, 75% quartile and upper threshold, all are increased in 2020.

According to CLO3

- There is an ignorable change in lower bounds and upper bounds.
- But there is an increase in 25%, median and 75% quartiles.
- We saw few outliers in 2019 but very less in 2020

These are shown in Fig 4.4.

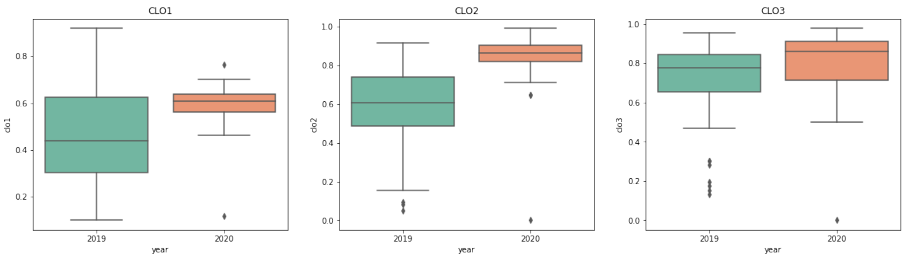

Figure 4.4: Compare 2019 and 2020 for CLOs

4.2 KPI Predicted Results

The following output shown below shows us the KPIs of past, present, and future predicted value. We can also see that in 2019 in KPI 1 38% of students secured 50% marks in CLO1 while that percentage grew to 93.3% in 2020. Keeping in view everything we can see that in 2021 100% of students are bound to secure more than 50% of marks. So we can demostrate how exactly our systems is efficient. This graph below in fig 4.5 is calculated values for our given data. From this data we predicted result for 2021 using different machine learning algorithms.

```
{'2019': {'KPI_1': [{'clo1': 0.3833}, {'clo2': 0.7333}, {'clo3': 0.8667}],
  'KPI_2': [{'clo1': 0.4612}, {'clo2': 0.5748}, {'clo3': 0.7059}],
  'KPI_3': [{'clo1': 0.2833}, {'clo2': 0.55}, {'clo3': 0.8167}],
  'KPI_4': [{'plo1': 0.4612}, {'plo3': 0.5748}, {'plo4': 0.7059}],
  'KPI_5': 2,
  'KPI_6': 1},
 '2020': {'KPI_1': [{'clo1': 0.9254}, {'clo2': 0.9851}, {'clo3': 0.9254}],
  'KPI_2': [{'clo1': 0.5984}, {'clo2': 0.8436}, {'clo3': 0.8031}],
  'KPI_3': [{'clo1': 0.5821}, {'clo2': 0.9851}, {'clo3': 0.8657}],
  'KPI_4': [{'plo1': 0.5984}, {'plo3': 0.8436}, {'plo4': 0.8031}],
  'KPI_5': 3,
  'KPI_6': 2}}
```


Figure 4.5: Calculated values from initial dataset

The predicted KPI values of 2021 can also be seen in fig 4.6.


```
{'2019': {'KPI_1': [{'clo1': 0.3833}, {'clo2': 0.7333}, {'clo3': 0.8667}],
  'KPI_2': [{'clo1': 0.4612}, {'clo2': 0.5748}, {'clo3': 0.7059}],
  'KPI_3': [{'clo1': 0.2833}, {'clo2': 0.55}, {'clo3': 0.8167}],
  'KPI_4': [{'plo1': 0.4612}, {'plo3': 0.5748}, {'plo4': 0.7059}],
  'KPI_5': 2,
  'KPI_6': 1},
 '2020': {'KPI_1': [{'clo1': 0.9254}, {'clo2': 0.9851}, {'clo3': 0.9254}],
  'KPI_2': [{'clo1': 0.5984}, {'clo2': 0.8436}, {'clo3': 0.8031}],
  'KPI_3': [{'clo1': 0.5821}, {'clo2': 0.9851}, {'clo3': 0.8657}],
  'KPI_4': [{'plo1': 0.5984}, {'plo3': 0.8436}, {'plo4': 0.8031}],
  'KPI_5': 3,
  'KPI_6': 2},
 '2021': {'KPI_1': [{'clo1': 1.0}, {'clo2': 1.0}, {'clo3': 0.9881}],
  'KPI_2': [{'clo1': 0.7764}, {'clo2': 1.0}, {'clo3': 0.9137}],
  'KPI_3': [{'clo1': 1.0}, {'clo2': 1.0}, {'clo3': 0.9176}],
  'KPI_4': [{'plo1': 0.7764}, {'plo3': 1.0}, {'plo4': 0.9137}],
  'KPI_5': 3,
  'KPI_6': 3}}
```


Figure 4.6: Predicted values of KPIs

We did not draw graph for KPIs5,6 because it has only one value and we don't need graph for that. KPI 1 and 5 are related and KPI 2 and 5 are related. We can see 3 CLO got more than 50% marks than KPI 5 would be equal to 3and same is for KPI 2 we can see 3 CLO got more than 60% marks than KPI 6would be equal to 3.

4.3 Showing trend in KPIs graph

We can see in fig 4.7 that the CLOs graphs for 2019,2020 and 2021 which show us the trend of students passing all of CLO1, CLO2 and CLO3.We can observe trend line is increasing which mean in next year probably more students get good grades.

Figure 4.7: Showing trend in KPIs graph

4.4 Comparison of all 4 classifier

The performance of all 4 classifiers are given in Table 4.2.

Evaluation metric	log_reg	KNN	SVM	RF
F1_Score	0.968610	0.981818	0.960000	1.0
Acuracy	0.944882	0.968504	0.929134	1.0
Acuracy_Pass	1.0	1.0	1.0	1.0
Acuracy_Fail	0.631579	0.789474	0.526316	1.0
True_Negative	12.0	15.0	10.0	19.0
False_Positive	7.0	4.0	9.0	0.0
False_Negative	0.0	0.0	0.0	0.0
True_Positive	108.0	108.0	108.0	108.0

Table 4.2: Comparison of all 4 classifier

As we can see that different parameters of all classifier gives the best results but when we see RF It's overall accuracy is 100% which is greater than all other classifiers and it's best use for our data.The mainn thing we focus on in this table is F1_Score and Accuracy.The F-score, also called the F1-score, is a measure of a model's accuracy on a dataset. ... The F-score is a way of combining the precision and recall of the model, and it is defined as the harmonic mean of the model's precision and recall.Accuracy tell us how accurate our results are.

Chapter # 5

Conclusion

As stated in the fig4.6 and table 4.2 proposed system is composed of 4 different machine learning classifiers (KNN, SVM, RF, Log Regression)in which we predicted the future results of KPI's which is based on CLOs and PLOs. The KPIs will identify the number of students who passed in different CLOs and other aspects and will notify the corresponding body to step in and provide aid to the students and teachers to provide an exceptional learning experience.In our model Random Forest is best classifier than other because it gives 100% efficency.Further enhancements in this domain is possible like real time monitoring and notifying system and would be readily available of providing help to anyone who would want to persue it.

References

[1] Mrs. Anuratha K.”INTERNATIONAL JOURNAL FOR INNOVATIVE RESEARCH IN MULTIDISCIPLINARY FIELD” ISSN: 2455-0620 Volume - 6, Issue - 6, June – 2020. Monthly, Peer-Reviewed, Refereed, Indexed Journal with IC Value: 86.87

[2] T. Devasia, Vinushree T P and V. Hegde, ”Prediction of students performance using Educational Data Mining,” 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE), 2016, pp. 91-95, doi: 10.1109/SAPIENCE.2016.7684167.

[3] W. Chen, C. G. Brinton, D. Cao, A. Mason-Singh, C. Lu and M. Chiang, ”Early Detection Prediction of Learning Outcomes in Online Short-Courses via Learning Behaviors,” in IEEE Transactions on Learning Technologies, vol. 12, no. 1, pp. 44-58, 1 Jan.-March 2019, doi: 10.1109/TLT.2018.2793193.

[4] M. Kaur and A. Girdhar, ”A Framework for the Indirect Assessment Tool for Outcome Based Education Using Data Mining,” 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2018, pp. 1-5, doi: 10.1109/ICCIC.2018.8782336.

[5] R. Al-Shabandar, A. Hussain, A. Laws, R. Keight, J. Lunn and N. Radi, ”Machine learning approaches to predict learning outcomes in Massive open online courses,” 2017 International Joint Conference on Neural Networks (IJCNN), 2017, pp. 713-720, doi: 10.1109/IJCNN.2017.7965922.

[6] J. Zhu et al., ”Mapping Engineering Students’ Learning Outcomes From International Experiences: Designing an Instrument to Measure Attainment of Knowledge, Skills, and Attitudes,” in IEEE Trans-

actions on Education, vol. 62, no. 2, pp. 108-118, May 2019, doi: 10.1109/TE.2018.2868721.

[7] R. C. Deo, Z. M. Yaseen, N. Al-Ansari, T. Nguyen-Huy, T. A. M. Langlands and L. Galligan, ”Modern Artificial Intelligence Model Development for Undergraduate Student Performance Prediction: An Investigation on Engineering Mathematics Courses,” in IEEE Access, vol. 8, pp. 136697-136724, 2020, doi: 10.1109/ACCESS.2020.3010938.

[8] Kevin Fong-Rey Liu and Jia-Shen Chen, ”Prediction and assessment of student learning outcomes in structural mechanics a decision support of integrating data mining and fuzzy logic,” 2010 2nd International Conference on Education Technology and Computer, 2010, pp. V3-499-V3-503, doi: 10.1109/ICETC.2010.5529492.

[9] H. Aliakbarian et al., ”Implementation of a Project-Based Telecommunications Engineering Design Course,” in IEEE Transactions on Education, vol. 57, no. 1, pp. 25-33, Feb. 2014, doi: 10.1109/TE.2013.2262800.

[10] S. Wang, Y. Han, W. Wu and Z. Hu, ”Modeling student learning outcomes in studying programming language course,” 2017 Seventh International Conference on Information Science and Technology (ICIST), 2017, pp. 263-270, doi: 10.1109/ICIST.2017.7926768.

[11] R. C. Raga and J. D. Raga, ”Early Prediction of Student Performance in Blended Learning Courses Using Deep Neural Networks,” 2019 International Symposium on Educational Technology (ISET), 2019, pp. 39-43, doi: 10.1109/ISET.2019.00018.

[12] A. Kashyap and A. Nayak, ”Different Machine Learning Models to Predict Dropouts in MOOCs,” 2018 International Conference on

Advances in Computing, Communications and Informatics (ICACCI), 2018, pp. 80-85, doi: 10.1109/ICACCI.2018.8554547.

[13] D. Vasić, M. Kundid, A. Pinjuh and L. Šerić, ”Predicting student's learning outcome from Learning management system logs,” 2015 23rd International Conference on Software, Telecommunications and Computer Networks (SoftCOM), 2015, pp. 210-214, doi: 10.1109/SOFTCOM.2015.7314114.

[14] H. M. R. Hasan, A. S. A. Rabby, M. T. Islam and S. A. Hossain, ”Machine Learning Algorithm for Student's Performance Prediction,” 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2019, pp. 1-7, doi: 10.1109/ICCCNT45670.2019.8944629.

[15] I. Khan, A. Al Sadiri, A. R. Ahmad and N. Jabeur, ”Tracking Student Performance in Introductory Programming by Means of Machine Learning,” 2019 4th MEC International Conference on Big Data and Smart City (ICBDSC), 2019, pp. 1-6, doi: 10.1109/ICBDSC.2019.8645608.

[16] A. Manaff Ismail, R. Ismail, Fariz Aswan Ahmad Zakwan and B. Nizam Ismail, ”Implementation and assessment of outcome based education (OBE) in the Faculty of Civil Engineering at Universiti Teknologi MARA (UiTM),” 2010 2nd International Congress on Engineering Education, 2010, pp. 211-214, doi: 10.1109/ICEED.2010.5940793.

[17] Dodridge, M.. “Learning outcomes and their assessment in higher education.” Engineering Science and Education Journal 8 (1999): 161-168.

Appendices

Appendix A (code)

Import python required packages

```
import json
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from matplotlib import rcParams
# figure size in inches
rcParams['figure.figsize'] = 8,6
import statsmodels.api as sm
from statsmodels.sandbox.regression.predstd import wls_prediction_std
from sklearn.ensemble import RandomForestClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.svm import SVC
from sklearn.linear_model import LinearRegression, LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score, accuracy_score, roc_auc_score,
confusion_matrix, f1_score
np.random.seed(9876789)
```

Import the data

```
df = pd.read_csv("01_data/new_data.csv", decimal=",", sep="|")
df.head()
```

Data Analysis

```
g = sns.boxplot(
    data=df,
    x="gender",
    y="total",
    hue="year"
)
g = sns.histplot(
    data=df,
    x='total', stat='count', bins=25,
    hue='year', kde=True, palette='Set2'
).set_title("Total marks distribution")
df.groupby('year').total.mean()
plt.figure(figsize=(20,5))
for i,col in enumerate(["clo1","clo2","clo3"]):
    plt.subplot(1,3,i+1)
    g = sns.histplot(
        data=df,
        x=col, stat='count', bins=25,
        hue='year', kde=True, palette='Set2'
    ).set_title(col.upper()+' distribution')
plt.show()
plt.figure(figsize=(20,5))
```

```
for i,col in enumerate(["clo1","clo2","clo3"]):
    plt.subplot(1,3,i+1)
    g = sns.boxplot(
        data=df,
        y=col,
        x='year', palette='Set2'
    ).set_title(col.upper())
plt.show()
g = sns.pairplot(
    df[["year","sno","gender","total","clo1","clo2","clo3"]],
    hue="year", diag_kind="hist", corner=True, palette='Set2'
)
fig, ax = plt.subplots(1,3,figsize=(20,6))
t = df.sort_values('total').reset_index(drop=True)
for i,col in enumerate(['clo1', 'clo2', 'clo3']):
    y1 = t[t.year == 2019][col]
    x1 = t[t.year == 2019]['sno']
    ax[i].plot(x1, y1, 'r.', label="2019-"+col)
    y2 = t[t.year == 2020][col]
    x2 = t[t.year == 2020]['sno']
    ax[i].plot(x2, y2, 'g.', label="2020-"+col)
    ax[i].hlines(
        t[t.year == 2019][col].mean(),
```

```
 min(x1.min(), x2.min()),
        max(x1.max(), x2.max()),
        label='2019-'+col+'-average ('+str(round(df[df.year == 2019][col].mean(), 4))+')',
        color='r'
    )
    ax[i].hlines(
        t[t.year == 2020][col].mean(),
        min(x1.min(), x2.min()),
        max(x1.max(), x2.max()),
        label='2020-'+col+'-average ('+str(round(df[df.year == 2020][col].mean(), 4))+')',
        color='g'
    )
    ax[i].set_xlabel('Students')
    ax[i].set_ylabel(col.upper())
    ax[i].legend(loc='best')
#plt.title("Ordinary Least Squared - Regression for %s" % col.upper())
plt.show()
```

Machine Learning Modelling

```
df.head()
df['result'] = (df.total >= 50)*1
new_df = pd.get_dummies(
    df,
    prefix=['year', 'gender'],
```

```
    columns=['year', 'gender']

)

new_df.head()

corr = df[['quiz','assignment','mids','final','total','clo1','clo2','clo3','plo1','plo3','plo4',
'result']].corr(method='pearson')

g = sns.heatmap(corr, annot=True)
```

Split the dataset into train and test

```
train, test = train_test_split(new_df, test_size=.2, shuffle=True)

train, test = train_test_split(new_df, test_size=.2, shuffle=True)

train.to_csv('01_data/train.csv', index=False)

test.to_csv('01_data/test.csv', index=False)

train.columns

X_train = train[['year_2019', 'year_2020', 'gender_FEMALE', 'gender_MALE', 'clo1', 'clo2',
'clo3', 'plo1', 'plo3', 'plo4']]

X_test  = test[['year_2019', 'year_2020', 'gender_FEMALE', 'gender_MALE', 'clo1', 'clo2',
'clo3', 'plo1', 'plo3', 'plo4']]

y_train = train.total

y_test  = test.total

lm = LinearRegression()

lm.fit(X_train, y_train)

print("Train r2 score %f" % r2_score(y_train, lm.predict(X_train)))

print("Test  r2 score %f" % r2_score(y_test, lm.predict(X_test)))

metrics = {}
```

train.columns

Linear Regression

X_train = train[['year_2019', 'year_2020', 'gender_FEMALE', 'gender_MALE', 'clo1', 'clo2', 'clo3', 'plo1', 'plo3', 'plo4']]

X_test = test[['year_2019', 'year_2020', 'gender_FEMALE', 'gender_MALE', 'clo1', 'clo2', 'clo3', 'plo1', 'plo3', 'plo4']]

y_train = train.total

y_test = test.total

lm = LinearRegression()

lm.fit(X_train, y_train)

print("Train r2 score %f" % r2_score(y_train, lm.predict(X_train)))

print("Test r2 score %f" % r2_score(y_test, lm.predict(X_test)))

metrics = {}

lm = LinearRegression()

lm.fit(X_train, y_train)

print("Train r2 score %f" % r2_score(y_train, lm.predict(X_train)))

print("Test r2 score %f" % r2_score(y_test, lm.predict(X_test)))

metrics = {}

Logistic Regression

feat = ['year_2019', 'year_2020', 'gender_FEMALE', 'gender_MALE', 'clo1', 'clo2', 'clo3', 'plo1', 'plo3', 'plo4']

X_train = train[feat]

X_test = test[feat]

```
y_train = train.result

y_test  = test.result

lr = LogisticRegression()

lr.fit(X_train, y_train)

print("Accuracy  (Train): %.4g" % accuracy_score(y_train, lr.predict(X_train)))

print("AUC Score (Train): %f" % roc_auc_score(y_train, lr.predict_proba(X_train)[:,1]))

print("Accuracy  (Test): %.4g" % accuracy_score(y_test, lr.predict(X_test)))

print("AUC Score (Test): %f" % roc_auc_score(y_test, lr.predict_proba(X_test)[:,1]))

f1_score(y_train, lr.predict(X_train))

tn, fp, fn, tp = confusion_matrix(y_train, lr.predict(X_train)).ravel()

(tn, fp, fn, tp)

tn, fp, fn, tp = confusion_matrix(y_test, lr.predict(X_test)).ravel()

(tn, fp, fn, tp)

tn, fp, fn, tp = confusion_matrix(new_df.result, lr.predict(new_df[feat])).ravel()

(tn, fp, fn, tp)

metrics["log_reg"] = [

    f1_score(new_df.result, lr.predict(new_df[feat])),

    accuracy_score(new_df.result, lr.predict(new_df[feat])),

    accuracy_score(new_df[new_df.result == 1].result, lr.predict(new_df[new_df.result ==
1][feat])),

    accuracy_score(new_df[new_df.result == 0].result, lr.predict(new_df[new_df.result ==
0][feat])),

tn, fp, fn, tp]
```

```
neigh = KNeighborsClassifier(n_neighbors=3)

neigh.fit(X_train, y_train)

print("Accuracy  (Train): %.4g" % accuracy_score(y_train, neigh.predict(X_train)))

print("AUC Score (Train): %f" % roc_auc_score(y_train,
neigh.predict_proba(X_train)[:,1]))

print("Accuracy  (Test): %.4g" % accuracy_score(y_test, neigh.predict(X_test)))

print("AUC Score (Test): %f" % roc_auc_score(y_test, neigh.predict_proba(X_test)[:,1]))

print("Accuracy  (Test): %.4g" % accuracy_score(y_test, neigh.predict(X_test)))

print("AUC Score (Test): %f" % roc_auc_score(y_test, neigh.predict_proba(X_test)[:,1]))

tn, fp, fn, tp = confusion_matrix(new_df.result, neigh.predict(new_df[feat])).ravel()

(tn, fp, fn, tp)

metrics["KNN"] = [

   f1_score(new_df.result, neigh.predict(new_df[feat])),

   accuracy_score(new_df.result, neigh.predict(new_df[feat])),

   accuracy_score(new_df[new_df.result == 1].result, neigh.predict(new_df[new_df.result
== 1][feat])),

   accuracy_score(new_df[new_df.result == 0].result, neigh.predict(new_df[new_df.result
== 0][feat])),

   tn, fp, fn, tp

]
```

Support Vector Machine classifier

```
clf = SVC(gamma='auto', probability=True)

clf.fit(X_train, y_train)
```

```
print("Accuracy  (Train): %.4g" % accuracy_score(y_train, clf.predict(X_train)))

print("AUC Score (Train): %f" % roc_auc_score(y_train, clf.predict_proba(X_train)[:,1]))

tn, fp, fn, tp = confusion_matrix(new_df.result, clf.predict(new_df[feat])).ravel()

(tn, fp, fn, tp)

metrics["SVM"] = [

    f1_score(new_df.result, clf.predict(new_df[feat])),

    accuracy_score(new_df.result, clf.predict(new_df[feat])),

    accuracy_score(new_df[new_df.result == 1].result, clf.predict(new_df[new_df.result == 1][feat])),

    accuracy_score(new_df[new_df.result == 0].result, clf.predict(new_df[new_df.result == 0][feat])),

    tn, fp, fn, tp

]
```

Random Forest Classifier

```
rf = RandomForestClassifier(max_depth=2)

rf.fit(X_train, y_train)

print("Accuracy  (Train): %.4g" % accuracy_score(y_train, rf.predict(X_train)))

print("AUC Score (Train): %f" % roc_auc_score(y_train, rf.predict_proba(X_train)[:,1]))

print("Accuracy  (Test): %.4g" % accuracy_score(y_test, rf.predict(X_test)))

print("AUC Score (Test): %f" % roc_auc_score(y_test, rf.predict_proba(X_test)[:,1]))

tn, fp, fn, tp = confusion_matrix(new_df.result, rf.predict(new_df[feat])).ravel()

(tn, fp, fn, tp)

metrics["RF"] = [
```

```
    f1_score(new_df.result, rf.predict(new_df[feat])),
    accuracy_score(new_df.result, rf.predict(new_df[feat])),
    accuracy_score(new_df[new_df.result == 1].result, rf.predict(new_df[new_df.result ==
1][feat])),
    accuracy_score(new_df[new_df.result == 0].result, rf.predict(new_df[new_df.result ==
0][feat])),
    tn, fp, fn, tp
]
```

Final result

```
res=pd.DataFrame(metrics,
index=["f1_score","accuracy","accuracy_pass","accuracy_fail","true_negative","false_posit
ive","false_negative","true_positive"])
res
```

KPI Calculation

```
result = {}
for year in df.year.unique():
    result[str(year)] = {}
    result[str(year)]["KPI_1"] = []
    for col in ["clo1", "clo2", "clo3"]:
        result[str(year)]["KPI_1"].append({
            col: round(df[(df.year == year) & (df[col] > 0.50)].shape[0] / df[df.year ==
year].shape[0], 4)
        })
    result[str(year)]["KPI_2"] = []
```

```
    for col in ["clo1", "clo2", "clo3"]:
        result[str(year)]["KPI_2"].append({
            col: round(df[(df.year == year)][col].mean(), 4)
        })
    result[str(year)]["KPI_3"] = []
    for col in ["clo1", "clo2", "clo3"]:
        result[str(year)]["KPI_3"].append({
            col: round(df[(df.year == year) & (df[col] > 0.60)].shape[0] / df[df.year ==
year].shape[0], 4)
        })
    result[str(year)]["KPI_4"] = []
    for col in ["plo1", "plo3", "plo4"]:
        result[str(year)]["KPI_4"].append({
            col: round(df[(df.year == year)][col].mean(), 4)
        })
    result[str(year)]["KPI_5"] = len([k for d in result[str(year)]["KPI_2"] for k, v in d.items()
if v > 0.50])
    result[str(year)]["KPI_6"] = len([k for d in result[str(year)]["KPI_3"] for k, v in d.items()
if v > 0.60])
```

KPI prediction for 2021

```
year = '2021'
result[year] = {}
for y1, y2 in zip(result['2019'], result['2020']):
```

```
 if isinstance(result['2019'][y1], list):
      result[year][y1] = []
      for s1,s2 in zip(result['2019'][y1], result['2020'][y2]):
          k1 = list(s1)[0]
          k2 = list(s2)[0]
          v1 = list(s1.values())[0]
          v2 = list(s2.values())[0]
          result[year][y1].append({k1:min(1.,round((1 + (v2 - v1)/v1)*v2, 4))})
  else:
      v2 = result['2020'][y2]
      v1 = result['2019'][y1]
      result[year][y1] = min(3,(1 + (v2 - v1)/v1)*v2)
```

Plot the KPIs

```
df_dict = []
for year, d in result.items():
    for kpi,d in d.items():
        if isinstance(d, list):
            for s in d:
                df_dict.append({
                    'year': year, 'KPI':kpi, 'subject':list(s.items())[0][0], 'score':list(s.items())[0][1]
                })
df_kpi = pd.DataFrame(df_dict)
```

```
plt.figure(figsize=(20,5))
for i,col in enumerate(df_kpi.KPI.unique()):
    plt.subplot(1,len(df_kpi.KPI.unique()),i+1)
    g = sns.lineplot(
        data=df_kpi[df_kpi.KPI == col],
        y='score',
        x='year', hue='subject', palette='Set2'
    ).set_title(col.upper())
plt.show()
```

Report After Defense

ORIGINALITY REPORT

16%	7%	8%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to University of Witwatersrand Student Paper	2%
2	www.maryourhelp-cebu.edu.ph Internet Source	1%
3	Rodolfo C. Raga, Jennifer D. Raga. "Early Prediction of Student Performance in Blended Learning Courses Using Deep Neural Networks", 2019 International Symposium on Educational Technology (ISET), 2019 Publication	1%
4	orcid.org Internet Source	1%
5	Weiyu Chen, Christopher G. Brinton, Da Cao, Amanda Mason-Singh, Charlton Lu, Mung Chiang. "Early Detection Prediction of Learning Outcomes in Online Short-Courses via Learning Behaviors", IEEE Transactions on Learning Technologies, 2019 Publication	1%
6	Submitted to Colorado Technical University Student Paper	1%

7	Avinash Kashyap, Ashalatha Nayak. "Machine learning models to predict the dropouts in Massive Open Online Courses", 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), 2018 Publication	1%
8	Submitted to Higher Education Commission Pakistan Student Paper	1%
9	Maninder Kaur, Akshay Girdhar. "A Framework for the Indirect Assessment Tool for Outcome Based Education Using Data Mining", 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2018 Publication	1%
10	Rijul Dhir, Anand Raj. "Movie Success Prediction using Machine Learning Algorithms and their Comparison", 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), 2018 Publication	1%
11	Submitted to Babes-Bolyai University Student Paper	1%
12	www.f1-fans.co.uk Internet Source	1%

13	Eraldo Occhetta. "Non Invasive Cardiac Output Evaluation with CO2 Rebreathing Method for CRT Patients", International Journal of Statistics in Medical Research, 2016 Publication	<1 %
14	www.bib.irb.hr Internet Source	<1 %
15	Ravinesh C. Deo, Zaher Mundher Yaseen, Nadhir Al-Ansari, Thong Nguyen-Huy, Trevor Ashley Mcpherson Langlands, Linda Galligan. "Modern Artificial Intelligence Model Development for Undergraduate Student Performance Prediction: An Investigation on Engineering Mathematics Courses", IEEE Access, 2020 Publication	<1 %
16	Submitted to Birzeit University Main Library Student Paper	<1 %
17	iosrjournals.org Internet Source	<1 %
18	ir.msu.ac.zw:8080 Internet Source	<1 %
19	Submitted to South Gloucestershire and Stroud College Student Paper	<1 %
20	Submitted to Laureate Education Inc. Student Paper	<1 %

21	Submitted to Liverpool John Moores University Student Paper	<1%
22	link.springer.com Internet Source	<1%
23	Kevin Fong-Rey Liu. "Prediction and assessment of student learning outcomes in structural mechanics a decision support of integrating data mining and fuzzy logic", 2010 2nd International Conference on Education Technology and Computer, 06/2010 Publication	<1%
24	academic.oup.com Internet Source	<1%
25	P.K. Gupta, Vipin Tyagi, S.K. Singh. "Predictive Computing and Information Security", 'Springer Science and Business Media LLC', 2017 Internet Source	<1%
26	dokumen.pub Internet Source	<1%
27	spectrum.library.concordia.ca Internet Source	<1%
28	telrp.springeropen.com Internet Source	<1%
29	Jiabin Zhu, Juebei Chen, Nathan McNeill, Tianyi Zheng, Qunqun Liu, Bing Chen, Jun	<1%

Cai. "Mapping Engineering Students' Learning Outcomes From International Experiences: Designing an Instrument to Measure Attainment of Knowledge, Skills, and Attitudes", IEEE Transactions on Education, 2018
Publication

30	d.researchbib.com Internet Source	<1 %
31	ir.lib.uwo.ca Internet Source	<1 %
32	iugspace.iugaza.edu.ps Internet Source	<1 %
33	mafiadoc.com Internet Source	<1 %
34	"Intelligent Computing Theories and Application", Springer Science and Business Media LLC, 2021 Publication	<1 %
35	Khalid Mahboob, Syed A. Ali, Umm ‐ e Laila. "Investigating learning outcomes in engineering education with data mining", Computer Applications in Engineering Education, 2020 Publication	<1 %
36	dro.dur.ac.uk Internet Source	<1 %

Report After Defense

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

PAGE 14

PAGE 15

PAGE 16

PAGE 17

PAGE 18

PAGE 19

PAGE 20

PAGE 21

PAGE 22

PAGE 23

PAGE 24

PAGE 25

PAGE 26

PAGE 27

PAGE 28

PAGE 29

PAGE 30

PAGE 31

PAGE 32

PAGE 33

PAGE 34

PAGE 35

PAGE 36

PAGE 37

PAGE 38

PAGE 39

PAGE 40

PAGE 41

PAGE 42

PAGE 43

PAGE 44

PAGE 45

PAGE 46

PAGE 47

PAGE 48

PAGE 49

PAGE 50

PAGE 51

PAGE 52

PAGE 53

PAGE 54

PAGE 55

PAGE 56

PAGE 57

PAGE 58

PAGE 59

PAGE 60

PAGE 61

PAGE 62

PAGE 63

PAGE 64

PAGE 65

PAGE 66

PAGE 67

PAGE 68

PAGE 69

PAGE 70

PAGE 71

PAGE 72

PAGE 73

PAGE 74

PAGE 75

PAGE 76

PAGE 77

PAGE 78

PAGE 79