Arab American University - Palestine

Faculty of Graduate Studies

Predicting Student’s Performance Based on Behavior and Online Engagement Activities Using Machine Learning

Prepared by:

Noora Ismail Shawareb

Supervisor

Dr. Ahmed Ewais

This Thesis was Submitted in Partial Fulfillment of the Requirements for the Master's Degree in Computer Science

July /2023

Thesis Approval

Predicting Student's Performance Based on Behavior and Online Engagement Activities Using Machine Learning

By

Noora Ismail Shawareb

This thesis was defended successfully on 17/7/2023 and approved by:

Committee Members	Signature
1. Dr. Ahmed Ewais / Supervisor:	……................................
2. Internal Examiner: Dr. Mohammed Awad	....
3. External Examiner: Dr. Abdellatif S. Abu-Issa	....

Declaration

This is to declare that the thesis entitled " Predicting Student's Performance Based on Behavior Data and Online Engagement Activities Using Machine Learning " under the supervision of Dr. Ahmad Ewais, is my own work and does not contain any unacknowledged work or material previously published or written by another person, except where due reference is made in the text of the document, and it has not been published or submitted elsewhere for any scientific degree.

Name: Noora Ismail Shawareb

Signature: نورا شوارب ....

Student ID: 201720258

Date: 27/12/2023

Dedication

I would like to dedicate this work to my Parents, who supported me in every step of my life. I would also like to thank My Husband and our family, who are always willing to provide any support. I dedicate this work to my brothers and sisters.

I would like to thank all dear friends for their support, I would like to thank each of my instructors & colleagues represented by my university Arab American University - Jenin.

Acknowledgments

I would like to express my sincere gratitude to my supervisors, Dr. Ahmed Ewais who have always been providing support, encouragement, and guidance throughout the completion of this thesis.

Also, I extend my sincere thanks and appreciation to the staff of the Faculty of Engineering and Information Technology at the AAUP- Master's program.

Finally, I would like to thank my family, for supporting and helping me throughout this work.

Abstract

Predicting students' performance is important in improving students' outcomes in Higher Educational Institutions, especially after the spread of educational platforms as a result of Corona epidemic. Students' behavior data and engagement activities recorded in Learning Management Systems could be used in predicting students' performance to help in determining the students' needs, which help students and university. In addition, it contributes to set development plans using ML models to serve this field.

The thesis is concerned with determining the benefits of using educational platforms especially Moodle through gathering students' attributes such as behavioral data and engagement activities to predict student's performance. This is somehow problematic in many ways like extracting the data then deciding which students' features to use in the study and which students' features are more beneficial in predicting students' performance. The problem of the study also appears in designing the framework, choosing the machine learning techniques, testing the algorithms and selecting only one that suits the study.

The data were subjected to pre-processing operations in order to get rid of the random partition, empty value processors, deleting duplicate data, describing, and coding the data. After that, these data were subjected after processing operations to different machine learning techniques as well as neural network techniques, where four techniques were used for classification according to an associated set of MATLAB-supported algorithms: KNN, TREE, Ensembled Tree and SVM.

This thesis reached several results. The most important of which are: The SVM technique was superior to other classification techniques, as the classification accuracy in this technique reached 90%, while the classification value in KNN was equal to 87%. MLPNNs neural networks gave better results, the accuracy of classification was 93%, neural networks were able to classify and predict with high accuracy, this leads us to many studies on the inclusion of other techniques of neural networks in the future.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ANNs	Artificial Neural Networks
ES	Ensemble
FN	False-Negative
FP	False-Positive
KNN	K-Nearest-Neighbor
LM	Levenberg-Marquardt
LMS	Learning Management System
ML	Machine Learning
MLP	Multi-Layer Perceptron
MLT	Machine Learning Techniques
MOE	Ministry of Education
MOOCS	Massive Open Online Courses
MSE	Mean Squared Error
NNs	Neural Networks
ROC Area	Receiver Operator Characteristic Curves
SVM	Support Vector Machine
TN	True Negative
TP	True Positive

CHAPTER ONE

INTRODUCTION

Because they are the best barometers of a nation's civilization and development, and because they have the most far-reaching effects on both individuals and society, the fields of teaching and learning have consistently ranked high among humanity's most significant and fastest-growing academic pursuits [1].

Changes in teaching methods are ongoing. The COVID-19 pandemic of the past three years has caused many schools, especially universities, to seriously consider adopting LMS for use in online education. Computers, projectors, smart boards, and the Internet are just some of the IT-enabled digital devices used in today's classrooms. Using machine learning, educational platforms and LMSs are becoming increasingly important investments for HEI. Machine learning might, for instance, help educators in every possible way, especially with administrative tasks and processes [2].

The most effective strategy for enhancing the future is education, as education also enhances the present. Since evaluating students during a course relies heavily on assessment, it is one of the most crucial responsibilities in the field of education. As a result, individuals are better able to assess their own progress in learning and correct their own faults in the future [4].

Artificial intelligence (AI) is on the verge of completely transforming education, and tools such as ChatGPT are leading the way in this revolution. These technologies possess the capacity to fundamentally transform education in multiple ways. Firstly, personalized learning experiences are facilitated by customizing content to suit individual needs, learning speeds, and preferences, hence improving student engagement, and understanding. In addition, the constant availability of educational resources driven by AI caters to a wide range of schedules and places, providing flexibility to learners all over the globe. The utilization of automated grading and immediate feedback simplifies the evaluation procedure, allowing educators to allocate their attention towards teaching. Adaptive learning guarantees that students are adequately stimulated, hence promoting enhanced comprehension. Utilizing data-driven insights enables educators to make well-informed decisions on their teaching methodologies.

1.1 Motivation and Research Problem

With the increasing availability of the Internet, it becomes more synonymous with learning systems, resulting in the flourishing E-learning systems that has reached their zenith during the recent COVID-19 virus pandemic (COVID-19) and has become the only viable means of teaching and communicating with students [5].

Initially, E-learning (LMS) were just a portal (web page) for registering courses and exam grades. By enabling the creation of personal profiles, sharing teaching materials, and online communication with instructors and students, more serious LMSs has begun to emerge. LMS were primarily utilized to provide distance learning, particularly by HEIs [6].

If we are to predict students' performance in a course, we must be able to identify them and divide them into smaller categories based on their characteristics. It is possible and straightforward for a competent; professional teacher or lecturer who knows each student, but it is challenging for large groups. Therefore, LMS data is a platform that provides us with datasets containing student characteristics and course-related activities. Using the techniques of machine learning algorithms, this dataset is used to predict students' performance [7].

The difficulty of accurately predicting students' performance stems from several factors, including the availability of datasets, the suitability of algorithms, and other factors, such as the efficacy of attributes or variables. A large dataset and multiple techniques based on various algorithms are required for the study to produce conclusive results [8].

1.2 Research Questions

The study seeks to answer the following questions:

RQ1: To what extent can online engagement influence students' academic performance?

RQ2: To what extent the features in the Moodle can be considered as a reliable predictor of students' academic performance?

RQ3: What are the features with the most effect on student performance?

RQ4: Which ML Algorithms can give more accurate results to predict students' performance?

RQ5: Which ML Algorithms can give better classification accuracy i.e. better results in comparison to other algorithms?

The experiment and discussion chapter of our research will be the platform where we tackle and resolve the key research questions. Here, we will carefully plan, carry out, and evaluate experiments, using a strong methodology and insights based on data to give thorough and well-supported responses. These chapters will serve as the pivotal point where hypotheses undergo testing, data are analyzed, and our contributions to the field come to fruition, illuminating the broad objectives of our study.

1.3 Objectives and Goals

This research mainly aims to utilize machine learning approaches to forecast students' academic success by analyzing behavior data and online engagement activities obtained from Moodle. Therefore, one of the objectives is to improve student achievement, the educational system, and the quality of education. Another aim is to examine the specific components that directly influence the process of predicting (students' performance in making predictions). Utilizing hybrid feature selection approaches, namely incorporating Behavior Data and Online Engagement Activities, is crucial for enhancing the accuracy of forecasting learners' performance. This approach ensures that each feature is optimized and holds greater significance in the performance prediction process. This study examines the correlation between students' behavior and online engagement data in Moodle and their academic success [9].

The primary goal is to determine relevant features and algorithms by examining the key factors contributing to student success and performance. Furthermore, it enables us to generate valuable predictive models for forecasting performance. The second aim is to devise machine learning algorithms for monitoring students in E-learning and delivering them with pertinent information, suitable solutions, and fitting courses. Consequently, students will be able to predict their performance by comprehending the fundamental attributes, enabling them to focus on those attributes, and the instructor will experience more ease in monitoring and supporting students [10].

1.4 Contribution

Several methods and approaches were considered when attempting to predict students’ performance. Most approaches are statistical in character; they rely on statistical methods such as the mean, median, and standard deviation. Additionally, this is intended for machine learning (ML) models. The models attempt to estimate the correlation between input variables and identify patterns within the input data. These characteristics can be categorized as (1) Behavior data, which describe the time, energy, and effort expended by a pupil on a task such as a quiz or an assignment. (2) Students’ interactions with the online learning platform, such as the number of downloaded files or sent and read messages. During the investigation process, additional information about the features will become apparent.

This study contributes to investigating the use of LMS datasets and machine learning algorithms to predict pupils’ performance, resulting in an improved ability to predict this performance. In addition, it will assist educational institutions in producing more accurate results, thereby improving their use of learning management systems. furthermore, it will aid researchers in machine learning, computer science, and statistics. This study's findings shed light on the most effective machine learning algorithms for predicting students’ performance in higher education institutions.

1.5 Summary

Predicting students’ performance is essential to educational data mining and Learning Analytics. They seek to support the process of analyzing educational data that is closely related. They offer instruments and methods for collecting, processing, and analyzing educational data to improve online teaching and learning. Learning Analytics is typically founded on data mining, machine learning and statistical analysis techniques.

Old educational data mining research utilized students' past performance or non-academic factors to construct their predictive models, paying less attention to students' activity data, such as the number of platform accesses, online lectures, documents, and exams. Although disengagement has been identified as a crucial indicator that negatively impacts students' performance, it remains an issue. Numerous studies have been conducted to forecast students' performance in online courses; this will be the first to be conducted in Palestine using Moodle as my knowledge.

Predicting students' performance is essential for educational institutions seeking to enhance students' achievement and performance. Although many educational institutions and scientific researchers study the prediction of students' performance, it remains difficult due to numerous complex issues, such as the various factors that influence students' performance and course achievement.

1.6 Thesis Structure

This thesis is structured as follows:

Chapter 1: the introduction which introduces the concepts of educational platforms and its importance as a motivation for the researcher, then presents the research problem, the research questions and the research objectives and goals. After that, the researcher explains this research contribution and finishes this chapter with a summary.

Chapter 2: the background which presents a brief description of previous work and literature review related to the research about machine learning in education, then about the dataset, machine learning techniques and algorithms used in predicting student's performance. Finally, it shows a summary about the dataset and the adopted algorithms.

Chapter 3: the proposed model, which explains the data collection and preprocessing phase. After that, it describes building model phase. Finally, it shows the classification metrics selection.

Chapter 4: the experiment and results, which presents the experiment and evaluation of the results by every algorithm. Finally, it shows MLPNNs VS SVM on Classification.

Chapter 5: the conclusion and future work where the thesis work concludes and the results are discussed. Finally, it gives directions for possible future work to develop this research in particular and to develop the field of research in general.

CHAPTER TWO

BACKGROUND

Due to the demand for Information Technology (IT) and IT-related jobs, more emphasis has been placed on computer science and related disciplines in education. In recent years, IT skills and knowledge have become essential in every science and teaching domain, particularly STEM (science, technology, engineering, and mathematics), making these skills an integral part of the majority of STEM sub-disciplines, such as Artificial Intelligence, Statistics, etc.

Machine learning, a subfield of artificial intelligence, is especially well-suited to analyze and interpret this data. Through the development and training of algorithms, one can extract significant information from online learning systems. These insights may include identifying students who are at risk of performing poorly, suggesting individualized learning materials, improving the delivery of content, evaluating the effectiveness of various instructional strategies, and other related tasks. Machine learning provides a method of using data to assess and improve online learning, making it more adaptable and effective. This eventually helps to achieve the larger objective of enhancing educational results.

Online learning platforms such as Moodle provide a vast amount of data that can be utilized for assessment through the application of machine learning. Machine learning models can offer useful insights by examining student interactions, engagement, and performance. The insights may encompass the identification of students who are at risk, the suggestion of tailored learning paths, the optimization of course content, and the evaluation of the efficacy of teaching tactics. The utilization of data in this review not only improves the experience of learning online but also aids in the ongoing enhancement of educational platforms. This combination of technology and education is a potent synergy that offers significant potential for the future.

2.1 Machine Learning in Education

Artificial intelligence (AI) is considered a subsidiary of machine learning. Machine learning is as simple as granting a machine or model access to a dataset and allowing it to self-learn using intelligent algorithms and techniques to examine the dataset and identify similarities and correlations between feature values. Brilliant idea in 1959 which says that “we should not have to teach computers but instead allow them to educate themselves” [10]. He coined the term "machine learning" to characterize his theory, which is now the accepted definition for the capacity of computers to learn independently.

Currently, machine learning applications in education are limitless, and this is the primary focus of this study. It evaluates the feasibility of applying and utilizing machine learning to identify and analyze students' attributes, behavior data, and engagement activities to predict their performance and studying its current and future capabilities.

2.2 Related Work

Many LMSs are available to create, administer, and distribute digital resources for face-to-face and online instruction. it states that a learning management system (LMS) facilitates interaction between traditional teaching methods and digital learning resources while providing students with personalized e-learning opportunities [11]. Due to the COVID-19 pandemic, educational institutions have had to adjust to restrictions on physical interaction, which have precluded the majority of conventional forms of instruction and evaluation [12].

Numerous HEIs have utilized LMSs effectively and continued to investigate the efficacy of using various LMSs. 561 LMSs are currently available for academic and educational purposes [13]. Moodle, Massive Open Online Courses (MOOCs) and Google Classroom were the most extensively used and studied learning platforms from 2015 to 2020 [14]. Numerous HEIs, including STEM education, use Moodle, an open-source LMS that is cloud-based. Based on user experiences, the number of Moodle users continues to rise. It is ranked among the top 20 finest LMSs in 2018 [15]. according to 78 million in 2015 to over 368 million by 2023 [16].

Different research investigations have been conducted to determine the relationship between student learning behaviors and performance. The authors investigate a variety of learning activities, including collaborative activities and providing feedback, using data from 13 students in an experimental setting [17]. Using data from an online forum and a survey of 144 students, this study investigates the effect of the diversity of learning styles on learning scores and students' satisfaction [18]. Although these efforts encompass a vast array of learning activities, they have been conducted with small sample sizes; therefore, survey data were still necessary for the analysis.

This study systematically reviews 357 articles and identifies five categories of data sources: demographic, personal, academic, behavioral and institutional. Demographic information includes age, gender, educational level, location, health information, and family history, among other factors [19]. Personality data includes psychological and affective information (e.g., self-regulation, self-efficacy, and the big five personality traits) [20]. Academic data include course performance and past performance, for example, performance during secondary school. Time on task, number of interactions with the learning material, etc., are examples of behavioral data. Institutional data pertain to instructional methodology, high school quality, etc.

In predicting students' performance, numerous methods and approaches were considered; most are primarily statistical and designed for machine learning models. Numerous researchers have chosen to investigate the Moodle Learning Management System because it is widely utilized by Educational Platforms, particularly for online teaching and learning by HEIs. However, research on utilizing Moodle and its capabilities is dispersed throughout the published literature. The following table displays studies with various LMSs, features, and algorithms. The following tables display studies with various LMSs, features, and algorithms. It determines the objectives, characteristics, and algorithms utilized in each related article, as well as the results and precision of each study. Table 2.1 presents a comparison between a number of studies that applied machine learning algorithms on dataset extracted from Moodle.

Table 2.1: Summarize articles that used Machine Learning based on dataset from Moodle

Title	Author	Year	Publisher	Objectives	Features	Algorithms	Results	Accuracy
Predicting students' performance in e-learning using learning process and behaviour data	Feiyue Qiu1 et al	2022	nature scientific reports	Assess the Elearning procedure. Construct an Elearning performance predictor based on the process-behavior classification model (PBC model) that you propose.	Behaviors such as self-directed learning, interacting with systems, utilizing resources, and interacting with others	(SVC(R), SVC(L), naïve Bayes, KNN(U), KNN(D), and Soft max) with three different data processing methods.	Reduce the temporal complexity of the procedure and keep the predictor's predictive Performance intact with the use of a feature selection approach.	95%
Predicting students' performance in online courses using multiple data sources	M´elina Verger Hugo Jair Escalante	2021	Educational data mining	Investigates a generic strategy for predicting pupils' performance.	Learner's age and education level are examples of learner demo age. Academia: The weighted average of the student's grades in the course leading up to the final test. Actionable: The sum of all mouse clicks made in the Courses.	CART decision tree algorithm	The experimental outcomes show preliminary conclusions about which data should be considered for the task.	80%

Title	Author	Year	Publisher	Objectives	Features	Algorithms	Results	Accuracy
Predict Students' Academic Performance based on their Assessment Grades and Online Activity Data	Amal Alhassa n Bassam Zafar Ahmed Mueen	2020	International Journal of Advanced Computer Science and Applications	examine how students' academic outcomes are affected by their (LMS) assessment grades and online activity data	Time spent on Moodle, number of files seen, number of mails exchanged, number of quizzes taken, and other academic data.	Decision tree, random forest, sequential minimal optimization, multilayer perceptron, and logistic regression.	Students' evaluation grades are the single most influential factor in determining their academic success.	Random forest achieved the highest accuracy value of 99.17%
Predicting course achievement of university students based on their procrastination behaviour on Moodle	Yeong wook Yang et al	2020	Springer-Verlag GmbH Germany, part of Springer Nature 2020	Method for anticipating students' performance in a course based on their procrastination is proposed; it makes use of students' submission data and homework grade.	Free time, Doing nothing, Doing something, Homework grade, and Overall Grade	L-SVM R-SVM	kids' average homework grades are falling, and kids' free and active time seems to be shrinking, while their sedentary time is on the rise. That is, the more students procrastinate, the less likely they are to graduate.	84% 69%
Predicting Student Performance from Their Behavior in Learning Management Systems	Parisa Shayan and Menno van Zaanen	2019	International Journal of Information and Education Technology	Use a student's online activity to predict how well they will do in a course..	Information from the learning management system, details on students and courses, and details about how well they performed.	decision tree J48 and ID3	After the first half of the semester, the most relevant factors were performance data (midterm grade) and LMS data (number of views, clicks, and sessions) before the midterm and the final exam.	73%

Title	Author	Year	Publisher	Objectives	Features	Algorithms	Results	Accuracy
Predicting Student Performance from LMS Data: A Comparison of 17 Blended Courses Using Moodle LMS	Rianne Conijn, Chris Snijders , Ad Kleinge ld, and Uwe Matzat	2017	IEEE Transact-ions on Learning Technologies	see if you can forecast student performance using LMS data and isolate individual student volatility.	Quantity of mouse-overs, Quantity of used materials, Normalized Meeting Time, Disrupted study schedules, longest lull in action, Quantity of time spent online, etc.	Pearson correlation analyses, Regression analyses	The mode of prediction for final exam grade utilizing LMS variables is transferable between these classes. Is low. However, the findings also demonstrate that final grades are highly influenced by both individual and group factors among students.	overall accuracy of 67%
Analysis of Online Behavior and Prediction of Learning Performance in Blended Learning Environments	Il-Hyun JO Yeonje ong PARK* * Jeongh yun KIM Jonwoo SONG	2014	Educational Technology International	anticipate student performance in a hybrid learning environment	Measured in terms of measures like Total Login Time, Total Login Frequency, and Visit on Board (VOB). VOR Login Frequency (LIR) Measures How Often Users Visit a Repository.	Random Forest	The characteristics of both discussion-based and lecture-based classes are reflected in the crucial differences highlighted by two cases of blended learning. Based instruction	86%

MOOC is one of the most used online platform for providing online courses by well-known schools, universities, teachers. Therefore, the next table summarize some studies by studying the objectives, features and algorithms used in each study and view their results and accuracies as shown in table 2.2.

Table 2.2: Summarize articles that used Machine Learning using dataset from MOOC

Title	Author	Year	Publisher	Objectives	Features	Algorithms	Results	Accuracy
Predicting Student Dropout in Self-Paced MOOC Course Using Random Forest Model	Sheran Dass Kevin Gary James Cunning ham	2021	Mdpi journal	an RF-based model for predicting whether or not students would complete a Massive Open Online Course (MOOC) in a foundational STEM subject .	Mastered: averages, trends, skew, standard deviation, and variance	Random Forest	The only way machine learning models can boost retention and success rates is if they are actually used in practice .	94.6%
CLSA: A novel deep learning model for MOOC dropout prediction	Qian Fua Zhangha oGaob Junyi Zhoua Yafeng Zheng	2021	Science direct	The objective is to examine the possibility of course dropout based on learners' current learning behavior.	Personal information Performance Navigation Weekly	RNN CNN prediction model CLSA	DLM have achieved better prediction results on MOOC datasets.	87.6%
A Machine learning based approach to enhance MOOC users' classification	Dr. Youssef MOUR DI et al.	2020	Turkish Online Journal of Distance Education-TOJDE	Create an initial collection of features that are crucial from a pedagogical standpoint. The next step is to pick the most effective machine-learning algorithm and featureselection strategy.	Interaction with video, with the transcript, with the quiz, with effort, with personal information, with the forum, with performance, with navigation, with the weekly final exam, with prerequisites, with supplementary materials .	KNN SVM Naive Bayes Decision Tree Random Forest Ensemble Method	The complexity of a machine learning model is decreased by the selection phase, which is used by nearly all algorithms .	0,839 0,852 0,858 0,77 0,868 0,92

Title	Author	Year	Publisher	Objectives	Features	Algorithms	Results	Accuracy
Early Prediction of Success in MOOC from Video Interaction Features	Boniface Mbouza o Michel C. Desmara is Ian Shrier	2020	Springer Nature Switzerland AG	A forecast of a student's performance that can be fed into teachers' dashboards and used to modify the curriculum, activate support, and personalize interventions for each learner . groups of students	Rate of enrollment, AR; Rate of use, UR; Index of viewing, WI;	DT	After the first week and the halfway mark, the results show that these metrics are quite useful for making predictions .	78%
Predictive Learning Analytics for VideoWatching Behavior in MOOCs	Madhum itha Shridhar an et al	2018	IEEE Xplore	Create a technique using Predictive Learning Analytics (PLA) to analyze how students in MOOCs would interact with videos.	fracSpent fracComp fracPlayed fracPaused numPauses avgPBR stdPBR numRWs numFFs	The nearest neighbor	Prediction quality is enhanced in 8 out of 9 cases when the bias predictor is supplemented with either collaborative filtering or regularization.	90%
Predicting MOOC Dropout overWeeks Using Machine Learning Methods	Marius Kloft Felix Stiehler Zhilin Zheng Niels Pinkwart	2014	Conference on Empirical Methods in Natural Language Processing (EMNLP) Association for Computation al Linguistics	Show how you can use click-stream data in your analysis. Ability to track student behavior over time is a key aspect.	how many requests were made, how many days were active, Meetings Held The amount of fast-forwarding available in videos...	SVM	Prediction is stronger at the conclusion of the course than it is at the start, when there are still relatively faint signals to pick up on .	81.4%

Table 2.3 summarizes studies and articles that used data mining techniques on dataset that were extracted from other LMSs such as Blackboard and, Canvas etc.

Table 2.3: Summarize articles that used Machine Learning using dataset from Blackboard, Canvas etc.

Blackboard								
Title	**Author**	**Year**	**Publisher**	**Objectives**	**Features**	**Algorithms**	**Results**	**Accuracy**
A Novel Study of the Relation Between Students' Navigational Behavior on Blackboard and their Learning Performance in an Undergraduate Networking Course	Mr. Salman Yousaf Dr. Pouyan Ahmadi Dr. Khondkar Islam	2018	George Mason University	Find the most important characteristics that can accurately predict students' achievement.	The Full Length of the Class. Quantity of Viewed Items. Sum of All Logins. Course Material Time. Competence. Evaluations. Quizzes .	Random Forest	In order to predict which students are not likely to do well on the midterm, we have constructed a good classification model.	77.78%
Modeling and Predicting Students' Academic Performance Using Data Mining Techniques	Ahmed Mueen Bassam Zafar Umar Manzoor	2016	Internation al Journal of Modern Education and	use data mining methods to examine and forecast student success based on their forum activity	(GPA), (Quiz1, Quiz2, Average), (Submit), (Delay), (Labtest1, Labtest2, Average), (Final Exam Grade), (Total	decision trees artificial neural networks	Teachers can focus extra time and effort on certain students to help them succeed.	overall prediction 86%
			Computer Science	and academic history	Time Spent), and (Final Exam Grade).	Naïve Bayes		
The Effect of Using E-Learning Tools in Online and Campus-based Classrooms on Student Performance	Edith Galy Clara Downey Jennie Johnson	2011	Journal of Informatio n Technolog y Education	examines the variables that contribute to students' overall achievement in a course	Final Exam Midterm Discussion Project Course Grade	SVM KNN	Independent learning and course performance were predicted by students' opinions on the PU and PEU of elearning technologies.	67.4%

Canvas								
Title	**Author**	**Year**	**Publisher**	**Objectives**	**Features**	**Algorithms**	**Results**	**Accuracy**
Rank-Based Tensor Factorization for Student Performance Prediction	Thanh-Nam Doan Shaghayegh Sahebi	2019	Proceedings of The 12th International Conference on Educational Data Mining	Understanding amongst students as a loose association of unarticulated ideas. reflects learning by adding a rank-based constraint to the tensor factorization goal function	students problems attempts Avg. attempts	Rank-Based Tensor Factorization RBTF	Taking students' gradual learning as a ranking problem, we suggested a new rank-based tensor factorization algorithm (RBTF) that may forecast their performance score .	92.5% 95.24%
Student Performance Prediction by Discovering Inter-Activity Relations	Shaghayegh Sahebi Peter Brusilovsky	2008	International Conference on Educational Data Mining	Introduce a method that can use supplementary data gleaned from students' use of ungraded learning resources to predict their success on standardized tests.	Average activity records density quiz- assign.	Paired resource SVD++ Single resource SVD++	Finding these commonalities allows teachers to reorganize lessons in ways that will have the most impact on their students' education. It can also be used to mimic	94%
					exam.		students' knowledge from several sources.	
Predicting Student Performance from Online Engagement Activities	Ghassen Ben Brahim	2022	Arabian Journal for Science and Engineering	construct a categorization scheme that outperforms the current options.	(1)Characteristics of past student achievement . (2)graphical features of student demonstrations. (3) Characteristics of	Random Forest (RF) (MLP) Support Vector Machine	The best results were achieved with a 97% accurate RF classifier (F1-score = 97%) . However, NB produce the worst results becau of the	97.4% 95.7% 94.8% 82.6% 92.1%

Using Novel Statistical Features					student platform interactions (engagement) . (4) Characteristics (in order to elaborate on dedication, efficacy, etc.) (5)Academic institutions – to elaborate on the features, techniques, and methods of instruction	(SVM) Naïve Bayes (NB) Logistic Regression (LR)	model's inbuilt dependence on the giv features.	

2.3 The LMS Moodle:

Moodle is organized through plugins that represent specific activities and extra features that can be added to the fundamental feature of the Moodle courses. The most used Moodle plugins, which come with basic Moodle installation, are Assignment, Attendance, Lesson, Page, Quiz, Folder, File and Glossary, SCORM Package, Feedback, and Database.

Here are some of the plugins that can be used: Logs, Activity, Activity Completion, Live Logs, (Quiz) Statistics, (Course) Participation, Course Completion Status, Events List, Activity Results Block, (Gradebook).

Overview, Ad-hoc Database Queries, Engagement Analytics, Course Dedication, Graph Stats and Graphical Interactive Student Monitoring.

The current version of Moodle includes the plugin Inspire Analytics. One of the most exciting features of this plugin is a model that predicts students who are at risk of non-completion (dropping out) a Moodle course, based on low students' engagement [21].

2.4 Dataset

Despite the diversity of data, most of the related studies have used either one data source, academic [19] or behavioral, or at most two data sources with the combination of demographic and academic data. Researcher highlights an availability-relevance trade-off of educational data which can limit considering more data sources for prediction tasks in the field [22].

Dataset in this research was obtained with the help of the Director of E-Learning Center at Arab American University. Students' data includes assessment grades and activity data on the Moodle such as (Final score of the student obtained in a course, the last login time and the number of lessons read and/or downloaded.). All students' data will be extracted from Moodle and other files dedicated to the assessment of behavior data which is used for evaluation.

Dataset in this research was obtained from Moodle at Arab American University. Students' data include behavior data and online engagement activities. All students' data were extracted from the Learning Management System (LMS) into files. Files for students' activities on the Moodle and other files were dedicated to the assessment of behavior data. Large portion of the data was used in machine learning and developing the algorithm. The main data was then fed to the algorithm to predict students' performance, and the resulted predictions were tested statistically to decide the accuracy of the study tool (the algorithm).

2.5 Data Description and Coding

The Dataset includes general attributes that affect the students' performance. In this study, the used features are described as follows:

1. Identification number of the student: This feature will be coded by an alternative code for the student to keep his/her information confidential.
2. Identification number of the course: This feature will be coded by an alternative code for course number to identify the course.
3. Number of sessions: This feature represents the number of lectures or meetings in the course, and it's divided and coded into four values that represent the

students' attendance rate for these lectures or meetings, as shown in table 2.4. A student who misses more than 25% of lectures will be withdrawn, for example, 31 out of 42 sessions.

4. Number of assignments done: This feature represents the number of assignments or tasks taken and completed by the student in the course, and it's divided and coded into four values that represent the fulfilled assignments percentage, as shown in table 2.4. For example, the student who submits three assignments out of three for one course gets a perfect score of four.
5. Number of quizzes taken: This feature represents the number of quizzes or exams taken and completed by the student in the course, and it's divided and coded into four values that represent the fulfilled ones percentage, as shown in table 2.4. For example, some students take two out of three quizzes.
6. Average score on quizzes: This feature represents students' average of all quizzes and Exams taken and completed by the student in the course, and it's divided and coded into four values that represent the average score as a percentage, as shown in table 2.4. A quiz has a score of 15% of the course total mark, so the average score will still in the range of 0-15.
7. Number of messages sent to the forum: This feature represents the number of messages sent to the forum by the student in the course, and it's divided and coded into four values that represent the fulfilled ones percentage, as shown in table 2.4. For example 10 messages.
8. Number of messages read on the forum: This feature represents the number of messages read on the forum by the student in the course, and it's divided and

coded into four values that represent the fulfilled one's percentage, as shown in table 2.4. For example, some students read seven out of ten total sent messages.

9. Time spent on Moodle: This feature represents the Time spent on Moodle by the student in the course, and it's divided and coded into four values that represent the fulfilled ones as a percentage, as shown in table 2.4. For example, some students spend two hours for each course per semester.
10. Total time spent on assignments: This feature represents the time spent on assignments by students in each course, and it's divided and coded into four values that are represented as a percentage, as shown in table 2.4. For example, some students spend teo hours on all assignments in a course.
11. Total time spent on quizzes: This feature represents the time spent on quizzes by students on Moodle in each course, and it's divided and coded into four values that are represented as a percentage, as shown in table 2.4. For example, some students spend two hours for all quizzes in a course.
12. Total time spent on forum: This feature represents the total time spent on forum in each course, and it's divided and coded into four values that are represented as a percentage, as shown in table 2.4. For example, some students spend twenty-five hours.
13. Final score of the student obtained in a course: This feature represents students' total score in the course, which is the sum of all his/her marks in a specific course, and it's divided and coded into four values that represent the average score as a percentage, as shown in table 2.4. For example, some students score ninety-four in computer science course.

14. The last login time: This feature represents the last login time into the forum in each course with the reference to the course ending date, and it's divided and coded into four values that are represented as a percentage, as shown in table 2.4. For example, the last login time for some students is 16/05/2023.
15. The number of lessons read and/or downloaded: This feature represents the number of lessons read and/or downloaded by students in the course during the registered semester time, and it's divided and coded into four values that represent the fulfilled ones as a percentage, as shown in table 2.4.
16. The number of lectures taken and/or downloaded: This feature represents the number of lectures taken and/or downloaded by students in the course, and it's divided and coded into four values that represent the fulfilled ones as a percentage, as shown in table 2.4.
17. Largest period of inactivity: This feature represents students' inactivity period in the course like missing classes, meetings or exams or staying logged out of the forum, and it's divided and coded into four values that are represented as a percentage, as shown in table 2.4.
18. The number of files taken and/or downloaded: This feature represents the number of files taken and/or downloaded by students in the course, and it's divided and coded into four values that represent the fulfilled ones as a percentage, as shown in table 2.4.
19. The number of files uploaded: This feature represents the number of files uploaded by students to the forum in the course, and it's divided and coded into four values that represent the fulfilled ones as a percentage, as shown in table 2.4.For example, some students have ten tasks in a course.

20. Academic records: This feature represents student GPA score, which is a number that indicates how high students scored in all taken courses on average using a scale from 1.0 to 4.0, and it's coded into four values that represent the average score as a percentage, as shown in table 2.4. For example, a student with a GPA of four has a perfect score in every course s/he has taken in the university.

Table 2.4: Activities Data Derived From Online Learning Platform (AAUP-Moodle)

Number	Attribute	Code
1	Identification number of the student	An alternate code is used for the student to keep his information confidential
2	Identification number of the course	An alternative code is used for the course
3	Number of sessions	%75 =1 %82-75=2 %90-82=3 %100-90=4
4	Number of assignments done	1 : %79-60 2: %99-80 3: %100 4:less than 60%
5	Number of quizzes taken	1 : %79-60 2: %99-80 3: %100 4:less than 60%
6	Average score on quizzes	=3 %100-90=480 %89- =2 70-79-60 =1 %69 %
7	Number of messages sent to the forum	1 : %79-60 2: %99-80 3: %100 4:less than 60%
8	Number of messages read on the forum	1 : %79-60 2: %99-80 3: %100 4:less than 60%
9	Time spent on Moodle	1 : %79-60 2: %99-80 3: %100 4:less than 60%
10	Total time spent on assignments	1 : %79-60 2: %99-80 3: %100 4:less than 60%
11	Total time spent on quizzes	1 : %79-60 2: %99-80 3: %100 4:less than 60%
12	Total time spent on forum	1 : %79-60 2: %99-80 3: %100 4:less than 60%
13	Final score of the student obtained in a course	=3 %100-90=480 %89- =2 70-79-60 =1 %69 %
14	The last login time	1 : %79-60 2: %99-80 3: %100 4:less than 60%
15	The number of lessons read and/or downloaded	1 : %79-60 2: %99-80 3: %100 4:less than 60%
16	The number of Lectures taken and/or downloaded	1 : %79-60 2: %99-80 3: %100 4:less than 60%

17	Largest period of inactivity	1 : %79-60 2: %99-80 3: %100 4:less than 60%
18	The number of Files taken and/or downloaded	1 : %79-60 2: %99-80 3: %100 4:less than 60%
19	The number of Files uploaded	1 : %79-60 2: %99-80 3: %100 4:less than 60%
20	Academic records	=3 %100-90=480 %89- =2 70-79-60 =1 %69 %

Null= no records found or number is zero or less than 60 for passing grades.

2.6 Adopted Algorithms

The used data in the research consist of students' behavior data and online engagement activities data derived from online learning platform (AAUP-Moodle): the last login time, time spent on Moodle, the number of lessons read and downloaded, largest period of inactivity, the number of quizzes, academic records, and assignments completed. Such task relies on machine learning algorithms. In this research, we will use one of the algorithms that is often used to predict students' performance in E-learning like Decision Tree, KNN, etc. The methodology suggested in this study consists of four phases: data collection, data pre-processing, testing classification algorithms, and finally evaluation.

2.7 Summary

Understanding Learning Management Systems (LMS) is vital as LMS platforms have a pivotal function in the realm of education and training. They enhance the efficiency of material distribution, create personalized learning experiences, provide assessment and feedback, ensure accessibility, scalability, and cost-effectiveness. LMS systems include analytics and reporting features to track progress and ensure compliance, making them essential tools for contemporary learning and organizational growth the primary objective is to facilitate future planning by harnessing the capability of prediction, which has become increasingly viable due to the availability of large datasets and advanced machine learning algorithms.

This study categorizes students' performance using a conventional machine learning classification methodology. The students' data comprises information on their behavior and online engagement activities. The behavior data and engagement activities of all students taken from the Learning Management System component of the data are subjected to cleaning, preprocessing, description, and coding.

CHAPTER THREE

METHODOLOGY MODEL

3.1 Introduction

This chapter will discuss the mechanism used to collect data from online learning platform (AAUP-Moodle), how the collected data is processed, cleaning the data from unacceptable values, and applying data encoding operations with more than one option. The data are also arranged according to its similarity by applying the standard degree. After the data are ready for use, they are entered into the different machine learning platforms used for classification and MLP neural networks to classify the features, choose the best results, and compare the applied ML techniques.

The process starts with data collection from the AAUP. Then, data preprocessing is applied: The data are in need to be processed, altered, excluded and subjected to other steps before they are entered into a machine learning models i.e, processing the data affects the nature and accuracy of the result, after that, the feature scaling is used, and cross-validation is applied to divide the data into training and testing. Finally, ML models and artificial neural networks are applied forming the model output which is fed to the cycle again until we reached the final results.

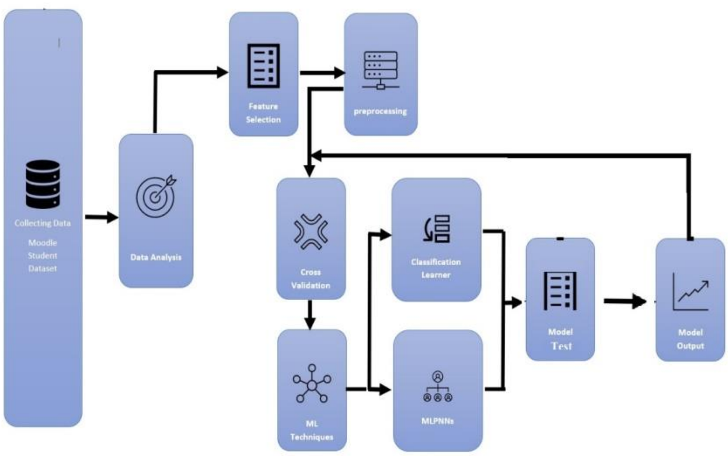

Figure 3.1 The general method procedure flow chart

3.2 Methodology

For the purpose of this research, a qualitative approach is adopted. It involves the in-depth review of literature on prediction processes that have been automated using ML, and subsequently the nature of data used; the variables used for the ML models development; and the ML algorithms applied in those existing works to logically suggest how other processes can be automated. The results from the literature review are used in proposing the framework.

This study has identified steps in the building of an ML model which include identifying data; preparing data, selecting the ML algorithm e.g. deep learning, training the algorithm (using prediction data set), evaluating different models and finally using the model to predict with new data [32].

3.2.1 Methods

This section discusses the research methods followed in this study.

Dataset: This study collects the data from students enrolled in random courses from the Moodle system of Arab American University, in Jenin -

Palestine. The types of data used for the analysis were from students' online activities such as System Log and Quiz Grades, etc.

System log: The system log consists of data of each click made in the system by a user. This dataset consists of (8220) observations and (20) attributes of (411) active users in the system.

3.2.2 Data Collection

Dataset used in the research consists of 20 features derived from online learning platform of Arab American University, in Jenin - Palestine.

Table 3.1: Activities Data Derived From Online Learning Platform

Number	Feature
1	Identification number of the student
2	Identification number of the course
3	Number of sessions
4	Number of assignments done
5	Number of quizzes taken
6	Average score on quizzes
7	Number of messages sent to the forum
8	Number of messages read on the forum
9	Time spent on Moodle
10	Total time spent on assignments
11	Total time spent on quizzes
12	Total time spent on forum
13	Final score of the student obtained in a course
14	The last login time
15	The number of lessons read and/or downloaded
16	The number of Lectures taken and/or downloaded
17	Largest period of inactivity
18	The number of Files taken and/or downloaded
19	The number of Files uploaded
20	Academic records (GPA)

a) Data pre-processing

In this step, only required features related to online activities in the log system are collected from the Moodle system for the data process in three stages.

b) Cleaning Data

Cleaning data is an essential step and the first stage of data processing, which precedes machine learning algorithms. After a thorough examination of the data taken, some samples are incomplete, and others are illogical. The data whose values are empty and illogical are deleted from the collected dataset [33].

c) Feature Scaling

This step helps the classification algorithms to work properly, so that this step normalizes the collected data and standardizes it in an organized form, as shown in the equation 3.1 [34].

$$x' = (x - \text{average}(x)) / (\text{std}(x)) \quad 3.1$$

d) Data Encoding

Neurons do not deal with values that exceed one unit hence the data encoding process is very important. For example, the student's performance for evaluation 1 is 0001, while for the second it is 0010, the third is 0100, and the fourth is 1000.This process is to ensure that neural networks work with high efficiency [35].

3.3 Framework Development

The proposed framework for developing machine learning models for prediction processes is represented in the steps followed in model development into four sections:

Section 1: Identification of the descriptive attributes or features (the independent variables). This section answers the three following questions: What decision-making processes can be significantly improved using machine learning in the prediction processes? What are the variables used in making those decisions? And what are the types and sources of the data containing those variables? The outcomes obtained from the review of literature provide the answers to these questions.

Section 2: Identification of the target attributes or labels (the various groups/clusters and the models are expected to achieve).

This section identifies the target attributes which the model is expected to use from the dataset. This section answers two questions: What decisions is the model expected to achieve? and what are the types and sources of the data to be extracted for that? The attributes have been chosen as mentioned before but it is still not decided which will make it as a decisive attribute.

Section 3: Training the machine learning models (using train data).

After identifying the descriptive and target attributes, training the machine learning models using those data becomes the next stage. For the proposed framework, the machine learning algorithms that can be employed include KNN, decision trees, SVM and Neural Networks (NN), as identified from literature. The review of the literature has also shown that the majority of prediction problems are addressed using supervised machine learning. In the proposed framework, the algorithm pool for training models for prediction processes is flexible; as novel machine learning algorithms in data analytics are developed. The used algorithms can be embraced and employed inside the framework.

Section 4: Evaluating the models (on test data) and selecting the most suitable one.

Models' evaluation and selection of the optimum machine-learning algorithm for prediction processes can be attained via comparison of the repetitive random sampling and cross-validation. Initially, a random fragmentation of the dataset is made into a training set and a test set, using a ratio and assume of, 7:3. Several duos are produced through random sampling of data points in the training set and data points in the test set (e.g. m duos). For every duo (e.g. the duoj), the training data point is engaged in training machine learning models employing every algorithm, and then a cross-validation metric, for example, k-fold cross-validation is carried Pjk)[36]. The average performance, Pj, is employed in representing the performance of the resultant algorithm. When the repetition of the training and cross-validation procedure for the respective randomly engendered data duos (m in total) are complete, the algorithms will then be examined using a suitable evaluation metric like mean squared error, mean absolute error, accuracy, etc. in determining the algorithm that outclasses the others.

Popular evaluation metrics used for machine learning models include mean absolute percentage error (MAPE), the mean squared error (MSE), the root-mean-squared error (RMSE) and the coefficient of determination (R2) or adjusted R*2. The optimum machine learning technique would vary from one particular case to another, depending on how long the period of the study is, the features are used, and the magnitude and quality achieved in the dataset. Finally, after the model has been evaluated and tested for efficiency, it will then be deployed as a part of the web-based prediction system.

3.4 Workflow

3.4.1 Building Model Phase

The used ML techniques and hybrid models in this thesis will be introduced in this section with all implementation details like the models and methods are applied to the dataset. There are two purposes of the techniques, one for classification, and another for prediction. This section shows the models, which have been applied to the study data, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Decision Tree, Ensemble Techniques, and Multi-Layer Perceptron Neural Networks (MLPNNs). MATLAB will be used to apply these models to the data set and compare the obtained results to identify the best accurate algorithm.

3.4.1.1 K-Nearest Neighbor (KNN)

The k-nearest neighbors' algorithm, aka (KNN) or (k-NN), is a nonparametric, supervised learning classification technique that depends on proximity to classify or predict the unknown value about a group of an individual dataset by using the nearest neighbor known value. It is used as a classification technique, working off the assumption that similar values can be found near each other. The KNN classifier keeps the whole training dataset during the learning process and assigns a class to a new unknown instance represented by the major vote of its nearest neighbor label in the training dataset [37].

The technique is working according to the following procedure:

- assuming the training set of the Xi observation with the associated class Yi
- making X a new observation that would be finding its classes
- calculating the distance using Euclidean distance between xi and x for i=0 … n points
- selecting D' ⊆ D. This determined query point is nearest to the k set of the training data point.

The voting technique is used to determine the predicted class which is called the weight technique as equation 3.2. [37]

$$y' = argmax_v \sum_{(xi,yi)\in D}^{v} \text{wi} * \text{I (v = yi)} \qquad 3.2$$

This equation works according to the following procedure: calculating the inverse of the distance. Hence, when the distance and the inverse equal 1\d, then calculate the sum of the inverse and divide each item by the resulting sum. After that, calculate the generated value, which belongs to its class. Finally, choose the max value to make this class the predictable class.

Advantages:

- Simple and not hard to execute: The algorithms are simple and have a very good accuracy, which is why it is considered one of the classifiers a data scientist should use.
- Easily adjusted: when adding training data samples, the algorithm can adapt to the new data because the training data is stored in the memory.
- Least demanding: KNN algorithm technique requires a value (k) and a distance metric.

3.4.1.2 Support Vector Machine (SVM)

Divide the data in this technique by finding a level capable of separating the data distributed in space, using the characteristics of this data that distinguish it from other data. The main and most important goal is not only the separation of this data, but the ability to obtain a level that separates that data with very high quality. Therefore, the aforementioned technique aims to determine the optimal level for separating the elements in space. This is done by calculating the best distance between points with similar characteristics and the proposed level, so that the level is selected with the best distances between similar points and the proposed levels [38].

The SVM technology addresses the extreme elements, as you can identify and ignore them, When an extreme value appears far from many elements, it looks at the most common features of these elements that he ignores this outlier, which gives him the power and ability to heal outliers, Also, the superior ability to separate the elements scattered in space is achieved in the best possible and comprehensive manner, see figure 3.2. This technology also has the ability to adapt to complex data through kernel technology. This technology allows dealing with linear and non-linear data and also handles complex functions [38].

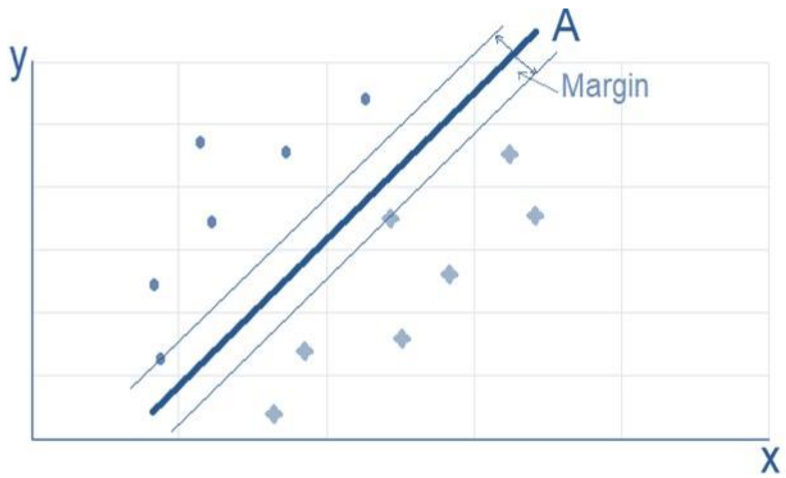

Figure 3.2 The SVM Hyperplanes, Margins, and Support Vector

The furthest possible margin is found between the assumed level and the nearest elements, and the aforementioned procedure is applied by employing the loss function c (x, y, f (x)) that establishes a relationship between the training data (X, Yj) for j = 1...N, with Xi ∈ Rd, Yj ∈ {−1, 1} and also the classifier instruction f(x), see equations 3.3,3.4 [38].

$$c(x, y, f(x)) = \begin{cases} 0 \quad , & if\ y * f(x) \geq 1 \\ 1 - y * f(x), & if\ y * f(x) < 1 \end{cases} \qquad 3.3$$

$$c(x, y, f(x)) = (1 - y * f(x))_+ \qquad 3.4$$

This technique uses the weight element in order to calculate the expected and actual values of the system. It gives a zero value to the weights of elements with similar features. If the actual expected values of the elements are not similar, the loss is calculated by weights based on that by adopting the relevant equations. The loss function updates the weight within its operations, and it uses and employs the gradient in order to achieve this. The gradient comes from partial derivatives that are directly related to the weights related to the data point, and this is evident in equation 3.5 [38].

$$\min_w \lambda \|w\|^2 + \sum_{i=1}^{n} (1 - y_i \langle x_i, w \rangle) \qquad 3.5$$

The weight updates equation, note Equation 3.7, and the update is calculated using the gradient; it calculates the partial derivatives which are related to the weights associated with the data point [38].

$$\omega = \omega - \alpha.(2\lambda\omega) \qquad 3.6$$

In the event of a false expectation, the classification of the element is lost, and in order to solve this problem use equation 3.8 which includes the loss and the regularization parameter [38].

$$\omega = \omega + \alpha.(y_i . x_i - 2\lambda\omega) \qquad 3.7$$

3.4.1.3 Decision Tree

Decision tree is a very popular technique for its ability in the prediction and classification process. This technique simulates a tree in its structure, its nodes represent a test of the traits, and the results of its prediction are displayed on its branches, each final leaf in the tree bears a class name. A decision tree learns by dividing the collected data into small subsets based on the operations and tests that the tree performs within its nodes. Operations are repeated more than once in a stage called recursive partition. The vertical division process continues until the values of the target variable and the values of the subset are equal. The vertical division process may stop if the system fails to obtain satisfactory results during several iterations. This technology is characterized by its ability and support for multidimensional data and also with its high ability to process data quickly because it divides the data into groups This gives it the ability to quickly classify data, it is also distinguished by not needing prior knowledge of the parameters, noting of this figure [39].

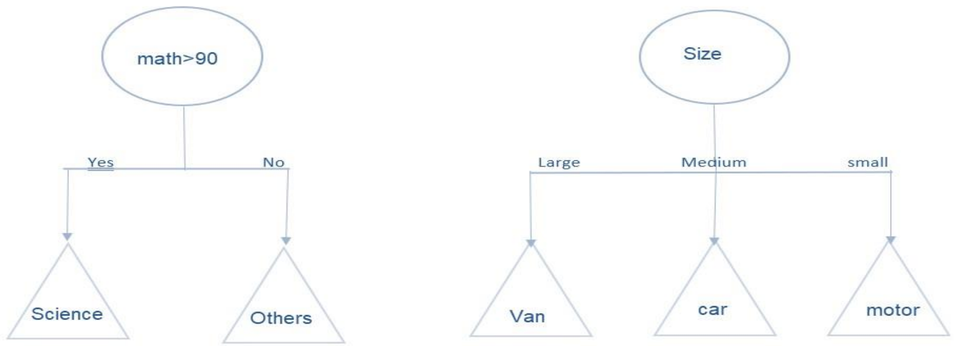

Figure 3.3 Decision Tree Two and Three Classes

And from a deep look at the structure of the tree technology, it consists of layers; each layer contains specific data. In a top-down building system, this technique is distinguished by its ability to collect similarities between data with each other. It seeks to divide the data into two or more groups to unite elements of similar characteristics with each other. The process of creating groups depends on the proper division of that data. The tree is characterized by its superior ability to divide the data in different directions, and this is absolutely useful after repeating the data in a branch of the tree except in rare cases.   The division provided by algorithm is comprehensible, and using the simplest arithmetic operations within her contract, and for you, the data must be relatively large in order to give this technology the ability to classify and predict effectively and powerfully.

The decision tree technique can be summed up in the following stages:

- The tree grows from a root node which includes all the data.
- The tree grows and is divided into two branches using the best features in the data collected at the level of this system.
- The previous steps are repeated over and over to include all sub-trees within the decision tree.
- Find the best attribute in the node used to divide the data.
- Enter the split procedure with the aim of reaching the final node or when there are no good results in future split operations.

In the upcoming sections, we will explore various advanced techniques to enhance our research, including ensemble methods and bagging. These methodologies will enable us to combine the strengths of multiple models, harnessing their collective power to address complex research questions. These approaches will be pivotal in elevating the depth and breadth of our findings.

3.4.1.4 Ensemble Techniques

This technique integrates more than one model together to achieve accurate classifications compared to using only one model. Although the procedure of merging more than one model requires complicated mathematical processes and additional calculations in the evaluation, the researchers unanimously agreed that the results are more accurate. Although other studies indicate that one model can be applied with an increase in mathematical equations and resources, the results that emerge from merging between models are better and more accurate than using one model. Most researchers who apply multiple models use fast algorithms such as decision trees with different machine learning algorithms in addition to combining some slow algorithms in some applications [40].

There are two methods to implement the multiple models: The first method is implementing the multiple algorithms in a sequential manner, so that the basic learner is used sequentially in sequential techniques. This step aims to increase the power of the relationship and dependence between the basic learners, as it gives a higher weight to the values that are not well-represented to improve the performance of the model, as shown in figure 3.4.

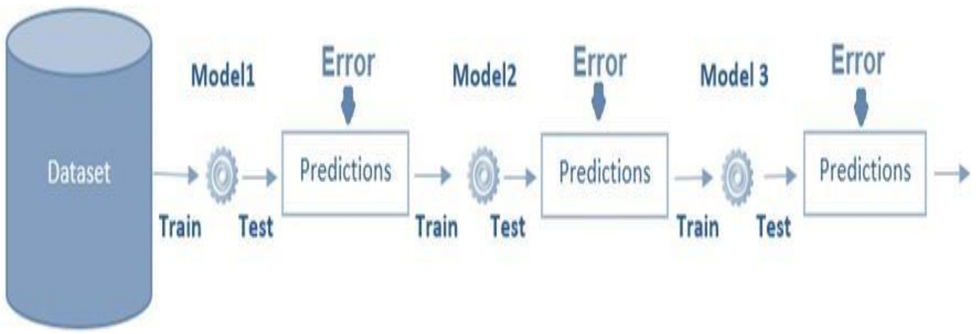

Figure 3.4 Ensemble Algorithm in a Sequential Model

Secondly, the multiple algorithms in the model are implemented in parallel. This means that the algorithms embedded in the model implement basic learning in a parallel manner such as the assembled model that combined between the decision trees algorithms to build random forest trees algorithm.

This compilation and building process encourage greater independence between algorithms in the model. Each model uses a different training dataset to find the error rate generated by the previous model, as shown in figure 3.5.

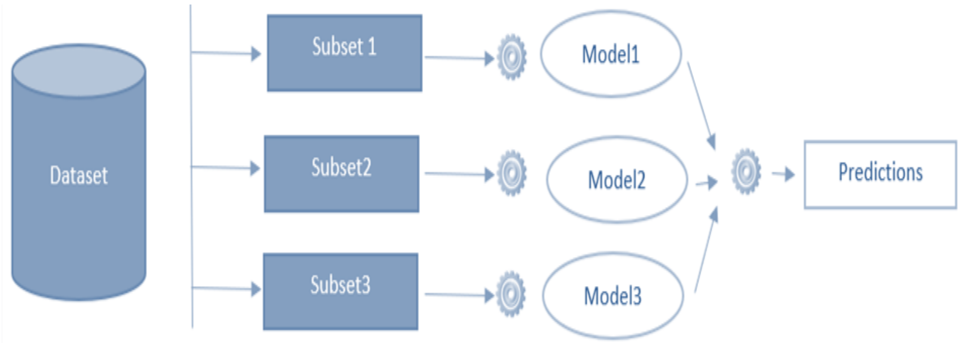

Figure 3.5 Ensemble Algorithm in a Parallel Model

Accordingly, the aforementioned techniques adopt the following scientific terminology:

a) Bagging techniques:

This word includes compilation and booting. It aims to eliminate the problem of using the same dataset to train models, which causes similar results to be obtained in each training process [41]. To overcome this problem, the bagging technique separates the dataset into sub-data sets with few changes. The bootstrap aggregating in this technique is applied to help understanding the distribution of data, as shown in the figure below:

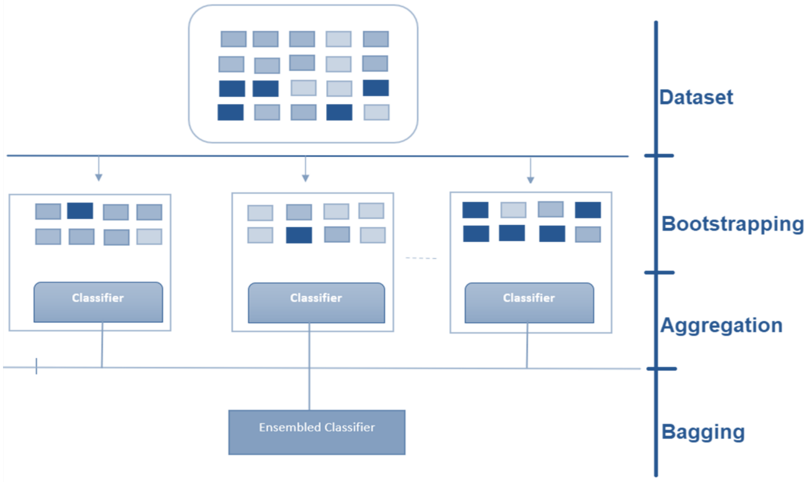

Figure 3.6 Ensemble Algorithm Bagging

The Bagging process is carried out by implementing the following steps [42]:

Step 1: Construct several subsets of the original dataset with the same size. Here it should be noted that the selected dataset must have the ability to be replaced.

Step 2: Each subset constructs a primary learning model and maps to a primary learning model.

Step 3: Each model is trained independently and in parallel manner.

Step 4: Merge the predictions obtained from all models to obtain the final predictions.

b) Random Forests Ensemble technique:

This technique performs the bagging and decision tree techniques to each generated dataset. After that, the final predictions are combined to create a final algorithm. In this process, samples and features are randomly distributed throughout the execution process. This is named the Random Forest technique. This technique is carried out by implementing the following steps [43]:

Step 1: Split the dataset into sub-data sets of the same size. Then, select a set of features to create a decision tree in order to deal with and classify them.

Step 2: Iterate the above process by using several variables covering the entire dataset, and the same variable can be used in multiple trees.

Step 3: Perform a check for each decision tree after finishing the process.

Step 4: Select the result from the obtained final results.

c) Boosting:

The boosting ensemble technique is a technique designed to deal with weak learner algorithms such as decision trees. It includes a sequential procedure aimed at minimizing errors during the classification process. To achieve this goal, the weight of misclassification data points is increased, and this procedure is repeated to get the desired classification. See figure 3.7 [44].

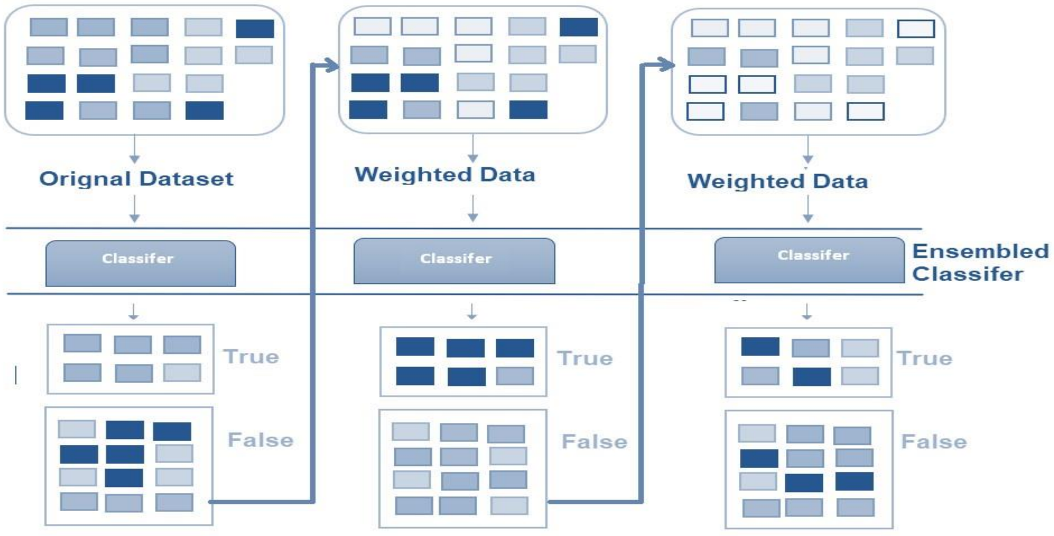

Figure 3.7 Boosting Ensemble Technique

The boosting ensemble process is carried out by implementing the following steps:

Step 1: Create a dataset and give equal weight values to all data points.

Step 2: Load weights and feed the model through the above generated dataset with corresponding weights.

Step 3: Increase the weight of some misclassified data points.

Finally: Check the result, if it is suitable, the process will be completed. But if it is not appropriate, the second step will be repeated.

3.4.1.5 Multi-Layer Perceptron Neural Networks (MLPNNs)

In recent years, the neural networks term has been frequently used in most scientific studies and research. Neural networks in machine learning mimic the functioning of neural networks in the human brain. Studies have indicated that the human brain stores all the information it obtains and how to respond to it in a small part called the human memory. Depending on this information, the brain can deal with the events that revolve around it by responding to it previously, the way others respond to it or the way to respond to it without prior knowledge [45].

Most neural networks use different machine learning methods in order to train the neural network to be able to predict correct results with the lowest possible error rate. Accordingly, there are two methods to train the neural networks; one is the supervised learning method and the other is the unsupervised learning method without the need to previous information about the event domain.

Due to the importance of neural networks, it has been applied in many areas of life such as education, medicine, agriculture, industry and many others. It also plays an important role in predicting outcomes in many areas such as disease diagnosis, academic disciplines, etc. It should be noted that neural networks are a training technique that is able to train and calibrate itself depending on the difference between the results that obtain it. This makes the neural network balanced with a stable structure capable of accurate prediction and classification.

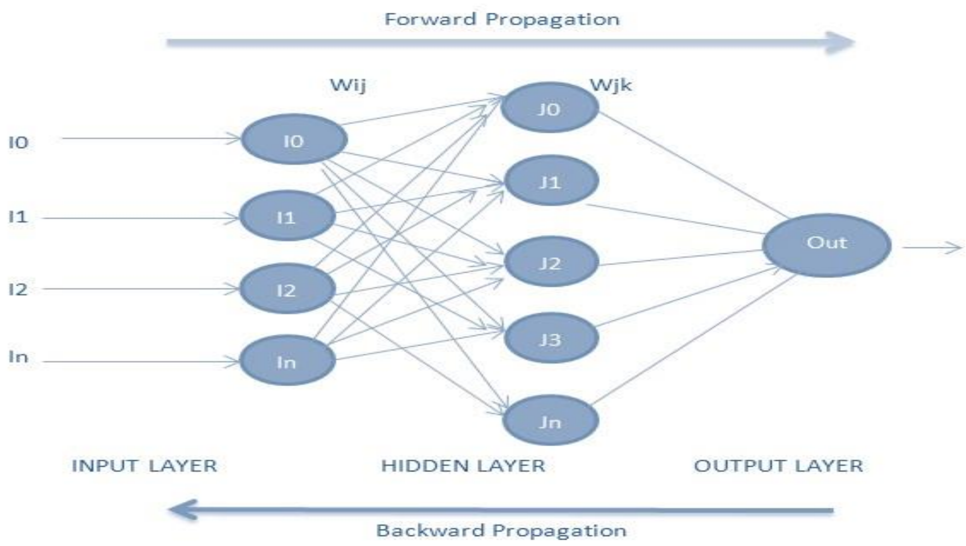

Figure 3.8 MLPNNs Techniques

The figure 3.8 shows the basic structure of neural networks. It consists of a large number of neurons that are interconnected by many connections. Each connection between neurons gains weight by applying the learning algorithms of neural networks. The value of these weights changes by knowing the error generated by the prediction. Although simple neural networks contain a single layer of neurons connected to each other, there are many complex neural networks that contain a different number of layers in which connected neurons are distributed over them. The machine learning algorithms that are applied to neural networks vary according to the application and the number of layers. Moreover, the weights are adjusted in neural networks depending on the error ratio between the expected output and the real output to obtain the least possible error and accurate results [46].

In multi-layer neural networks, there are three main layers:

First: the input layer which is the primary source of inputs for the neural networks. It works on entering data in the form of neurons in the neural network throughout the training process. These neurons represent features and characteristics that have been developed to study and test a specific application.

Second: the output layer displays the results obtained during the neural network training process. These results are compared with the actual results to calculate the error resulting from the prediction process by applying the supervised learning process. The weights of neurons are also adjusted based on the calculated error rate to obtain more accurate results.

Finally, the hidden layer is used to reset the weights in the neural network. It reaches the weights of neurons between the hidden network and the input layer. Also, it is located between the input and output layer. Accuracy of the results in neural networks depends on the accuracy of adjusting the weights between neurons where the values of the weights are mathematically multiplied by the variables values that are entered through the neural network input layer. After that, the obtained results are fed into a series of arithmetic operations inside the neural network hidden layer. In this layer, the results are processed and passed to the neural network in order to display in the form of expected results from performing neural network to them. This phase is called the forward propagation phase. This phase works to compare the expected results with the accurate results that it aims to achieve.

On the other hand, the error generated by the prediction process is used to affect the neural network leading to adjusting the weights in it. The process of adjusting the weights is done by applying a set of arithmetic operations to reach the appropriate weights between the hidden and input layer. Finally, this process is repeated to reach the lowest possible error as possible, or to stop the process if it does not reduce the error and reach the appropriate prediction. It is worth noting that the weights are randomly assigned as an initial value at the beginning of the neural network formation. Many mathematical equations are applied on neural networks to calculate the expected output for the above-mentioned operations. One of these equations is the following [47]:

$$y = \sum W_{ij} X_i + \theta_j \qquad 3.8$$

θ_j: called the bias node

Where w_{ij} is the connection weight between ith neuron in the NN input layer and the jth neuron in the NN hidden layer, and subscripted x_i: is the ith NN input features.

As mentioned above, the results obtained during the prediction process in neural networks are compared with the accurate results that it aims to reach. This comparison is applied by calculating the difference between the network output and the actual output. This is called an error, which is denoted by the symbol Δ_k and is used in the following equation 3.9. [47].

$$\Delta_k = t_k - Y_k \qquad 3.9$$

y_d is an expected output and y_i is the actual output.

On the other hand, to calculate the least square error in the neural networks, the following equation is applied [47]:

$$MSE = \frac{1}{2} \sum_i^n (y_d - y_i)^2 \qquad 3.10$$

The training process in neural networks goes into a repetition loop. The error resulting is calculated, and the weight values are adjusted to get the lowest error by using the following equation [47]:

$$\Delta w_{i+1} = \alpha.\mathrm{E}.x_i \qquad 3.11$$

In the training phase, the sum of the weights of the variables that enter into the input layer is calculated and passed to the j^th node in the hidden layer through the following equation: [47]

$$y = \sum W_{ij} X_i + \theta_j \qquad 3.12$$

Furthermore, the sigmoid function in multi-layer neural networks is known to be used. Also, the activation function in the neuron is calculated. Hence, the output of the j^th neuron is [47]:

$$Y_j = X_k = \frac{1}{1+e^{-y}} \qquad 3.13$$

δ_k: The error rate can be calculated by multiplying the Δ_k with the derivative of the sigmoid function used on the activation function [47].

$$\delta_k = \Delta_k Y_k \, (1 - Y_k) \qquad 3.14$$

The weight is reset based on the error value that is calculated. So that, it is increased or decreased by delta weight which is calculated using the following equations 3.15 [47]:

$$\Delta w_{jk} = l \, \delta_k X_k \qquad 3.15$$

Δw_{jk}: The delta weight computed is based on the weight between neuron j and k, which can be controlled by using the training rate. The weights between neurons are adjusted based on the delta weight produced in the prior equation by performing equations 3.16. [47]

$$w_{jk} = w_{jk} + \Delta w_{jk} \qquad 3.16$$

For calculating the partial error δ_j that is used in pack propagation phase, the following equation is applied: [47]

$$\delta_j = (t_k - Y_k)Y_k \sum w_{jk}\, \delta_k \qquad 3.17$$

w_{ij}: is the weights between two neurons. i,j is adjusted by using the equations 3.18 and 3.19, so: [47]

$$\Delta w_{ij} = l\, \delta_j X_j \qquad 3.18$$

$$w_{ij} = w_{ij} + \Delta w_{ij} \qquad 3.19$$

Frequently, forward and backward propagation are applied in neural networks with the aim of reducing the error rate or if there is no improvement in the prediction accuracy.

3.5 Classification Metrics Selection

Choosing classification metrics is one of the important things that must be considered in measuring the strength and efficiency of the model applied in neural networks. Accordingly, criteria are adopted to judge the feasibility of implementing the machine learning model and its capability to predict highly efficient results. There are several criteria that are applied to find out the efficiency of the model applied to the data. Also, there are other criteria that are implemented to determine the accuracy of the results. Here, the most widely used criteria that should be applied on each machine learning model will be reviewed [48].

In binary classification processes, the expected output of the prediction is either favorable or negative. So that, if the classification process is correct, then the expected output will be favorable. Otherwise, if the classification process is wrong, the expected output will be negative. However, the model sometimes makes an error in the prediction process such as it considers some correct classifications to be negative and other incorrect classifications to be favorable. The best technique used to study classifier performance is to implement a confusion matrix. The correct prediction is represented by the diagonal line. Also, the number of samples predicted correctly appears clearly. On the other hand, the other cells represent the number of samples that were incorrectly predicted, whether positive or negative.

Table 3.2 Confusion matrix

		Predicted Class	
Actual Class		Class = Yes	Class = No
	Class = Yes	TP (N)	FN(N)
	Class = No	FP(N)	TN (N)

N: represents the number of data points that are predicted.

TP: represents the correct classification for the actual positive class classified.

FP: represents the incorrect classification for the actual positive class that is not classified.

TN: represents the correct classification for the actual negative class classified.

FN: represents the incorrect classification for the actual negative class that is not classified.

In order to measure the tools of neural network learning models, the previous classifications and the amount of data in each class are used in a set of mathematical equations [49].

a) **Accuracy**: The accuracy measure is the most frequently used measure of expressing the ratio of the correct results predicted by the training model to the sum of all the results obtained. However, if the accuracy scale is very high, it does not indicate that the training model is working perfectly. Therefore, other measurement tools must be applied to ensure the efficiency and integrity of the model. To measure the accuracy scale, we use the following equation [50]:

$$Accuracy = \frac{\mathrm{TP}+TN}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}+\mathrm{TN}} \qquad 3.20$$

b) **Precision (Positive Predictive value)**: This scale is used to calculate the ratio of correct expected positive results to the ratio of total positive results. Here, the accuracy scale result is high if the number of expected false positive results is few. This scale is calculated using the following equation [51]:

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \qquad 3.21$$

c) **Recall (sensitivity or true positive rate):** This scale is used to calculate the ratio of the expected correct positive results to the ratio of the actual total positive results. This scale is calculated using the following equation:

$$Recall = \frac{T\mathrm{P}}{T\mathrm{P}+F\mathrm{N}} \qquad 3.22$$

d) **Specificity:** This scale is used to calculate the ratio of expected true negative results to the sum of true negative and false positive results. This scale is calculated using the following equation:

$$Specificity = \frac{TN}{TN+\mathrm{FP}} \tag{3. 23}$$

e) **F1 score**: This scale is used to represent the average weight of accuracy and recall scale. This requires considering all false positive and false negative results. Depending on this, the accuracy scale is high valued in cases where false positive results are equal to false negative results. Otherwise, it is appropriate to use matrices to ensure the efficiency of the learning model and the results integrity. We would like to point out here that the strength of this measure lies in the uneven distribution of predicted results [52].

$$\text{F1 Score} = \frac{2*(\text{Recall} * \text{Precision})}{(\text{Recall} + \text{Precision})} \tag{3. 24}$$

f) **Receiver-Operating Characteristic ROC:** To be able to deeply understand the applied learning model and to ensure its efficiency, the ROC scale is used. ROC represents the results obtained during the training process through graphical curves. These curves are used to analyze and understand the quality of the machine learning model that has been applied to the model. Moreover, this scale is designed to work on models consisting of multi classes. In addition, it is able to work with high efficiency on dual models and evaluate errors resulting from the prediction process.

The curve is drawn depending on the true positive rate (TPR) which is relative to the false positive rate (FPR) calculated through the following mathematical equations:

$$TPR = 1 - \text{Specificity} \tag{3.25}$$

$$FPR = FP/(FP + TN) \tag{3.26}$$

$$FDR = TP/(TP + FP) \qquad 3.27$$

In this measure, the ability of a particular model to distinguish between negative and positive outcomes resulting from the prediction process is measured by the values that appear in the area under the curve. Accordingly, if the ability of the model to distinguish between results is high, then the value that appears in the curve area AUC is high. Otherwise, if the ability of the model to distinguish between results is equal to one, this means that the efficiency of the model is ideal and it is able to distinguish between the results perfectly. On the other hand, if the ability of the model to distinguish between results is equal to zero, this means that the model is unable to distinguish between predicted results, as illustrated in figure 3.9 A. But if the values that appear in the area under the curve are greater than half and less than the ideal, this model can distinguish between results in a satisfactory way as appeared in figure 3.9 B. Finally, if the value of the area under the curve is equal to half, this means the inability of the model to distinguish between expected results. Based on that, the learning model forecasts random classes or constant classes for all expected results, as shown in figure 3.9 C [53].

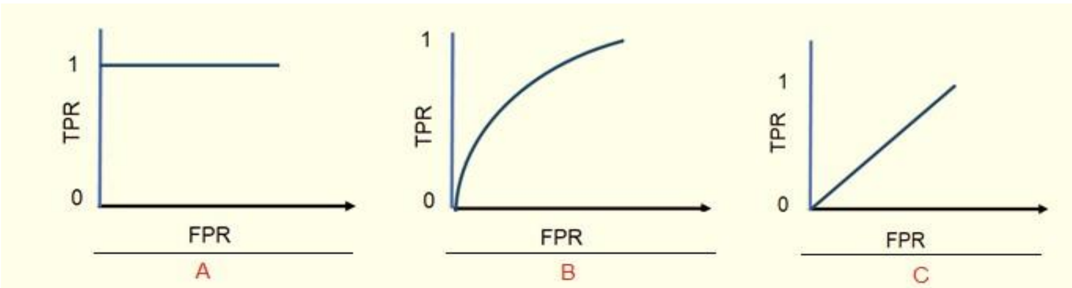

Figure 3.9 A: ROC with AUC =1, B: ROC with AUC >0.5, C: ROC with AUC =0.5

According to the above, we conclude that the model's ability to predict and distinguish between positive and negative results increases if the value of the area that is located below the curve increases. Here, multi-class measurement is performed. And the question is converted from multiple to binary class. The prediction error is one set that is complemented to the prediction correctness for all results. See table 3.3 and the equations that follow it.

Table 3.3 Confusion matrix multiclass

		Predicted Class			
Actual Class		Class 1	Class 2	Class 3	Class 4
	Class 1	TP	FN0	FN1	FN2
	Class 2	FP0	TN0	TN1	TN3
	Class 3	FP1	TN4	TN5	TN6
	Class 4	FP2	TN7	TN8	TN9

To calculate the precision of the confusion matrix, the following equation is applied:

$$\text{Precision} = \text{TP/TP} + (\text{FP0} + \text{FP1} + \text{FP2}) \qquad 3.28$$

To calculate the recall of the confusion matrix, the following equation is performed:

$$Recall = TP\ /\ TP + (FN0 + FN1 + FN2) \qquad 3.29$$

While a macro forms the mathematical mean of all results in the same matrices.

$$Macrro = (matrices1 + Matrices\ 2 + \cdots + Matrices\ n)/n \qquad 3.30$$

Moreover, the weight represents each class matrices multiplied by the number of its repetitions (class weight) divided by their number.

$$Weighted = ((C1) * Matrices1 + WC2) * Matrices2+. +W(Cn) *$$
$$Matrices)/n \qquad 3.31$$

C is class and W is weight.

This technique is called one-vs-rest which aims to transform the system from a multi-class to a binary one problem [54].

CHAPTER FOUR

EXPERIMENT AND RESULTS

4.1 Introduction

This chapter reviews the practical procedures applied to the data. After the preprocessing stage, different machine learning techniques are applied for classification. The classification learner tool in MATLAB is used for classification, the techniques are chosen depending on the category in which they are included, and the most suitable ones are chosen for classification.

MATLAB 2018b is used on a computer with HP workstation 9th generation processor, 1.9GHz, 32GB RAM and 1TB hard disk. The device gives results in a good time of less than 3 minutes. Four techniques have been applied which are SVM, KNN, Tree and Ensemble Tree. This algorithm is suitable for multi-class classification. The classification result of the four algorithms will be discussed. Also, the classification results will be compared to determine which of these algorithms is the best for our classification.

Neural network algorithm is used in the classification, and the classification is applied by using 10 neurons which is selected after many experiments on (5, 10, 15 and 20) neurons. The goal is to obtain the best results with the least number of neurons. Also, these measures aim to compare the neural networks with the classification learner algorithms, and to know which one is stronger in the classification and prediction process for our case.

4.2 Machine Learning Classification Learner Result

The results in this section are reviewed through the machine learning techniques used to classify students' performance in the portal. To judge the rating techniques, the measurement parameters mentioned in the previous chapter are used, whereas rating scales are necessary for the judgment to determine the ability of the technique. Moreover, classification matrices are very important to ensure the classification strength and integrity of machine learning techniques in all categories.

4.2.1 Machine Learning Classification Learner Result

This section reviews the different results when machine learning algorithms specialized in classification are applied.

4.2.1.1 KNN Experiment Result

KNN technology is used to rate students' performance in the portal. The input variables are 17, and the number of views is 553. 5-FOLD cross validation is performed, and number of blocks used is 10 with four output classes. The result from sunning these techniques is shown in the following table 4.1:

Table 4.1 KNN classification matrices

	Class 0	**Class 1**	**Class 2**	**Class 3**
	Predict			
Actual	103	0	28	0
	8	139	4	0
	21	1	155	4
	0	0	10	77
	Class 0	Class 1	Class 2	Class 3
TP	103	139	155	77
TN	390	398	327	459
FP	29	1	42	4

FN	28	12	26	10
(Recall)	0.79	0.92	0.86	0.89
False Negative Rate	0.21	0.08	0.14	0.11
(Precision)	0.78	0.99	0.79	0.95
False Discovery Rate	0.22	0.01	0.21	0.05
Specificity	0.93	1.00	0.89	0.99
FPR = (1-Specificity)	0.07	0.00	0.11	0.01
Accuracy	0.86			
F1 score	0.78	0.96	0.82	0.92
Macro-F1	0.87			
Macro-(Recall)	0.86			
Macro-(Precision)	0.88			
weighted-f1	0.86			
weighted-Recall	0.86			
weighted-Precision	0.87			

The table shows the classification results for the different classes, where the accuracy of the classification for the combined categories reach 0.86. When examining the results of each category alone, we see that the third class has the worst result in the classification model as shown by the F1 score. Also, Macro-F1 is 0.78, the macro matrices give the average for each category of the F1 Score, Furthermore, the number of samples in each category that affects the outcome of the F-score have to be studied. To perform this, the weighted F1 score is used. As we see, there is no difference between values in these matrices for these techniques.

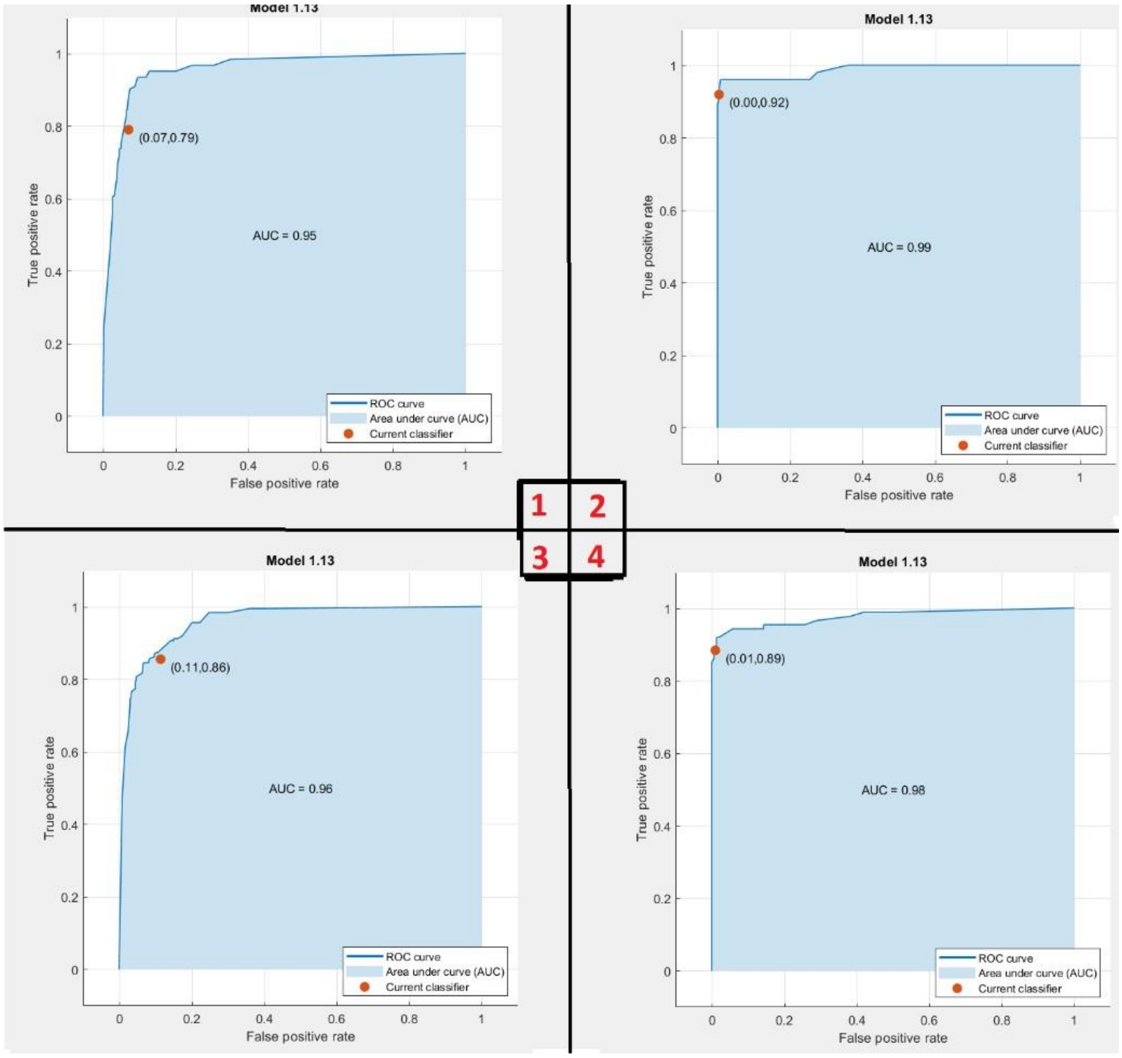

Figure 4.1 KNN ROC curve

The ROC curve is plotted between sensitivity versus specificity for each category, and each category is plotted against the other categories. The area under the curve, denoted by AUC, is calculated, and placed in the middle of each graph. This area is colored blue as shown in the graph. Looking at the graph, it is seen that the AUC value of the first class is 0.95, and this indicates that the first-class classification is excellent. Its sensitivity is 0.79, and its specificity is 0.07. The area under the curve is 0.99, and the sensitivity and specificity are 0.92 and 0.00, respectively. The AUC value is 0.96, and the sensitivity value for it is 0.86, while the specificity value is 0.11. The sensitivity and specificity value for the fourth category are 0.89 and 0.01, respectively. And the AUC value is 0.98. This shows the good result for classification in class 2.

4.2.1.2 SVM Experiment Result

Classifying students' performance on an SVM algorithm is also applied to check the quality of it. The input variables are 17, and the number of views is 553. 5 -FOLD cross validation is performed, and number of blocks used is 10 with four output classes. The result from sunning these techniques is shown in the following table 4.2:

Table 4.2 SVM classification matrices

	Class 0	**Class 1**	**Class 2**	**Class 3**
	Predict			
Actual	117	1	16	0
	5	146	0	0
	21	1	158	1
	0	0	6	81
	Class 0	Class 1	Class 2	Class 3
TP	117	146	158	81
TN	393	400	350	465
FP	26	2	22	1
FN	17	5	23	6
(Recall)	0.87	0.97	0.87	0.93
False Negative Rate	0.13	0.03	0.13	0.07
(Precision)	0.82	0.99	0.88	0.99
False Discovery Rate	0.18	0.01	0.12	0.01
Specificity	0.94	1.00	0.94	1.00
FPR = (1-Specificity)	0.06	0.00	0.06	0.00
Accuracy	0.91			
F1 score	0.84	0.98	0.88	0.96
Macro-F1	0.91			
Macro-(Recall)	0.91			
Macro-(Precision)	0.92			
weighted-f1	0.91			
weighted-Recall	0.91			
weighted-Precision	0.91			

The table shows the classification results for the different classes, where the accuracy of the classification for the combined categories reaches 0.91. When examining the results of each category alone, we see that the third class has the worst result in the classification model as shown by the F1 score. Also, Macro-F1 is 0.91, the macro matrices give the average for each category of the F1 Score, Furthermore, the number of samples in each category that affects the outcome of the F-score have to be studied, and to perform this the weighted F1 score is used. As we see, there is no difference between values in these matrices for these techniques.

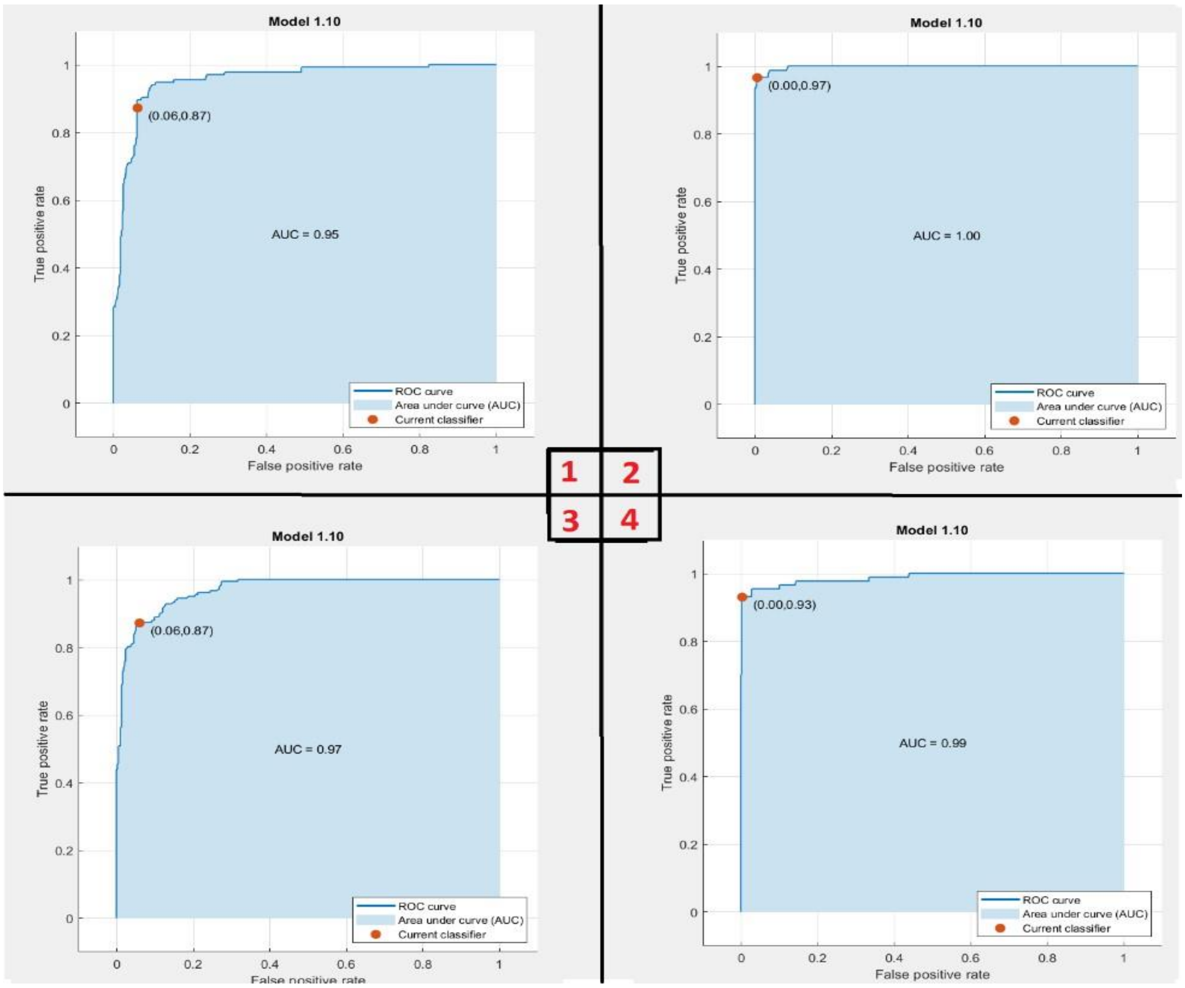

Figure 4.2 SVM ROC curve

By Looking at the graph, it is seen that the AUC value of the first class is 0.95, and this indicates that the first-class classification is excellent. Its sensitivity is 0.87, and its specificity is 0.06. The area under the curve is 1, and the sensitivity and specificity are 0.97 and 0.00, respectively. The AUC value is 0.97, and the sensitivity value for it is 0.87, while the specificity value is 0.06. The sensitivity and specificity value for the fourth category are 0.93 and 0.0, respectively. And the AUC value was 0.99. This shows the good result for classification in class 2.

4.2.1.3 Decision Tree Experiment Result

The result of this technique appears on the tree branches. This technique is similar to the tree in its construction. The number of input variables that is used in these techniques are 17, and the number of views is 553. 5 -FOLD cross validation is performed, and the number of blocks used is 10 with four output classes. The result from sunning these techniques is shown in the following table 4.3:

Table 4.3 DT classification matrices

	Class 0	Class 1	Class 2	Class 3
	Predict			
Actual	108	2	24	0
	7	143	1	0
	27	0	152	2
	0	0	6	81
	Class 0	Class 1	Class 2	Class 3
TP	108	143	152	81
TN	385	400	341	464
FP	34	2	31	2
FN	26	8	29	6
(Recall)	0.81	0.95	0.84	0.93
False Negative Rate	0.19	0.05	0.16	0.07
(Precision)	0.76	0.99	0.83	0.98
False Discovery Rate	0.24	0.01	0.17	0.02
Specificity	0.92	1.00	0.92	1.00

FPR = (1-Specificity)	0.08	0.00	0.08	0.00
Accuracy	0.88			
F1 score	0.78	0.97	0.84	0.95
Macro-F1	0.88			
Macro-(Recall)	0.88			
Macro-(Precision)	0.89			
weighted-f1	0.88			
weighted-Recall	0.88			
weighted-Precision	0.88			

The table shows the classification results for the different classes, where the accuracy of the classification for the combined categories reaches 0.88. When examining the results of each category alone, we see that the third class has the worst result in the classification model as shown by the F1 score. Also, Macro-F1 is 0.88, the macro matrices give the average for each category of the F1 Score. Furthermore, the number of samples in each category that affects the outcome of the F-score has to be studied. To perform this, the weighted F1 score is used. As we see, there is no difference between values in these matrices for these techniques.

As shown in the illustrated ROC in figure 4.3, the AUC value of the first class is 0.95, and this indicates that the first-class classification is excellent. Its sensitivity is 0.87, and its specificity is 0.06. The area under the curve is 1, and the sensitivity and specificity are 0.97 and 0.00, respectively. The AUC value is 0.97, and the sensitivity value for it is 0.87, while the specificity value is 0.06. The sensitivity and specificity value for the fourth category are 0.93 and 0.00, respectively. And the AUC value is 0.99. This shows the good result for classification in class 2.

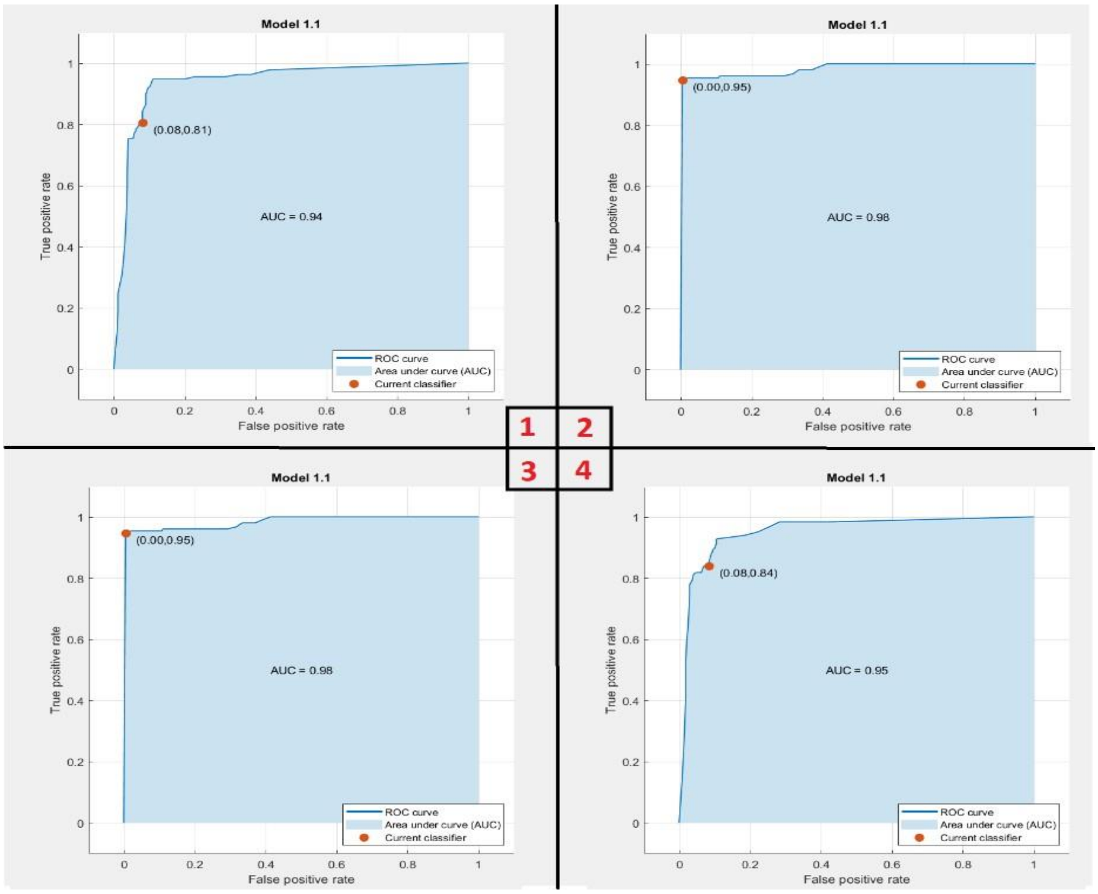

Figure 4.3 DT ROC curve

4.2.1.4 Ensembled Tree Experiment Result (ES)

Undoubtedly, this algorithm enhances the tree algorithm by using complex mathematical operations in evaluating the model. Also, ES checks the integrity of the prediction. The accuracy increases in these techniques by combining the tree algorithms and merging them together. The input variables are 17, and the number of views is 553. 5 -FOLD cross validation is performed, and the number of blocks used is 10 with four output classes. The result from sunning these techniques is shown in the following table.

Table 4.4 ES boosted tree classification matrices

	Class 0	Class 1	Class 2	Class 3
	Predict			
Actual	115	3	16	0
	6	142	2	1
	25	0	155	1
	0	1	4	82
	Class 0	Class 1	Class 2	Class 3
TP	115	142	155	82
TN	388	398	350	464
FP	31	4	22	2
FN	19	9	26	5
(Recall)	0.86	0.94	0.86	0.94
False Negative Rate	0.14	0.06	0.14	0.06
(Precision)	0.79	0.97	0.88	0.98
False Discovery Rate	0.21	0.03	0.12	0.02
Specificity	0.93	0.99	0.94	1.00
FPR = (1-Specificity)	0.07	0.01	0.06	0.00
Accuracy	0.89			
F1 score	0.82	0.96	0.87	0.96
Macro-F1	0.90			
Macro-(Recall)	0.90			
Macro-(Precision)	0.90			
weighted-f1	0.89			
weighted-Recall	0.89			
weighted-Precision	0.90			

The accuracy of the classification for the whole classes reaches 0.89. When examining the results of each category alone, we see that first class has the worst result in the classification model as shown by the F1 score. Also, Macro-F1 is 0.90, the macro matrices give the average for each category of the F1 Score. Furthermore, the number of samples in each category that affects the outcome of the F-score has to be studied. To perform this, the weighted F1 score is used. As we see, the accuracy decreases to 0.89 since the first class has a good representation in data and this affects the weight accuracy of the system.

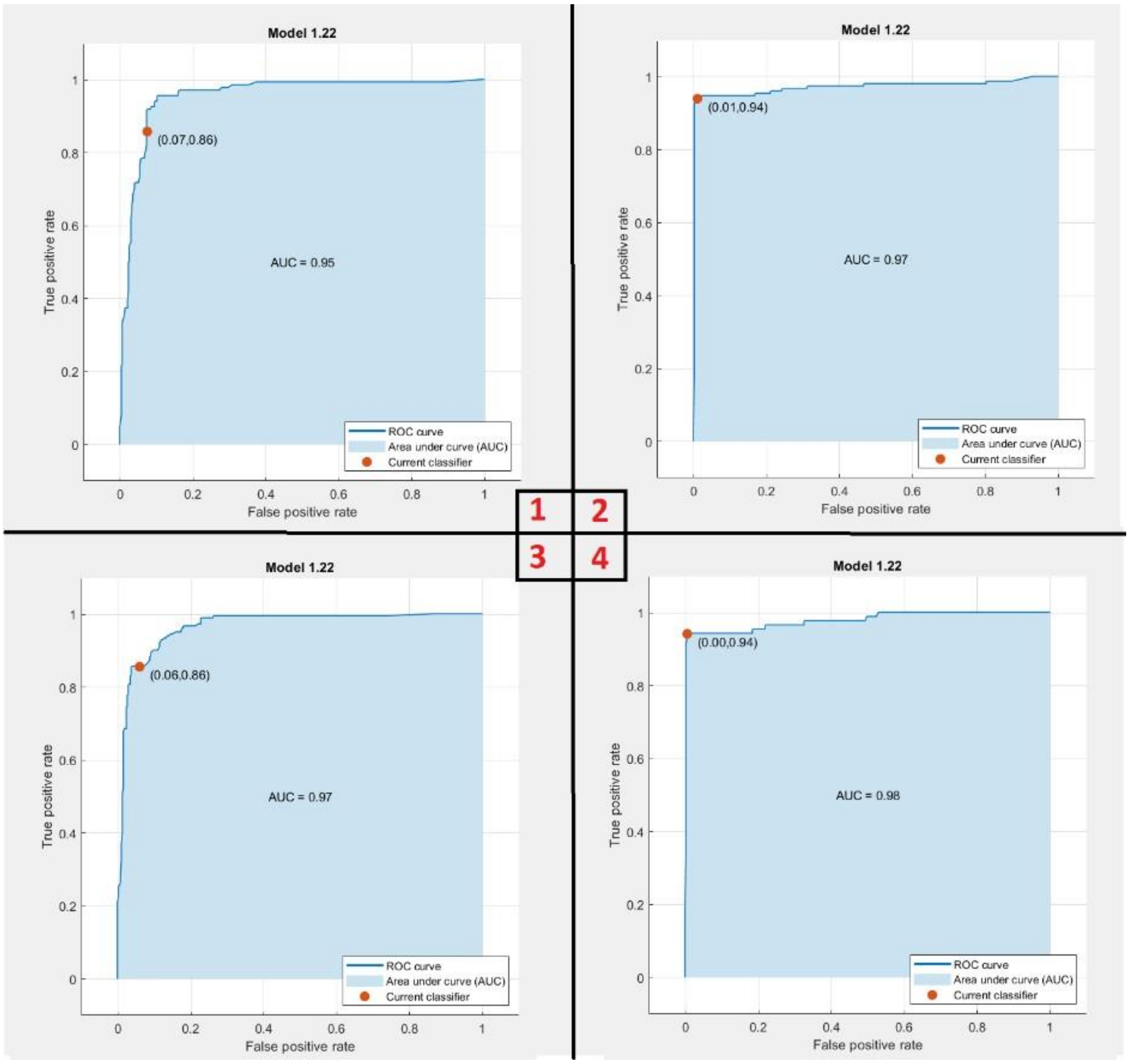

Figure 4.4 8 ES boosted tree ROC curve

The figure shows that the AUC value of the first class is 0.95, and this indicates that the first-class classification is excellent. Its sensitivity is 0.87, and its specificity is 0.06. The area under the curve is 1, and the sensitivity and specificity are 0.97 and 0.00, respectively. The AUC value is 0.97, and the sensitivity value for it is 0.87, while the specificity value is 0.06. The sensitivity and specificity value for the fourth category are 0.93 and 0.00, respectively. And the AUC value is 0.99.This shows the good result for classification in class 2.

It is very useful to make comparisons between the different classification learner algorithms to determine which one is suitable for the educational dataset which is related to the performance of students in the courses using the portal. Therefore, in this section, the four algorithms that are applied are compared and the differences between them are shown. The comparisons show that SVM outperforms the rest of the classification algorithms, followed by Ensemble Tree, and the KNN techniques are lagging in the classification.

Table 4.5 Classification learner comparison

	DT	Ensembled	KNN	SVM
Accuracy	0.88	0.89	0.86	0.91
F1 score	0.78	0.82	0.78	0.84
Macro-F1	0.88	0.90	0.87	0.91
Macro-(Recall)	0.88	0.90	0.86	0.91
Macro-(Precision)	0.89	0.90	0.88	0.92
weighted-f1	0.88	0.89	0.86	0.91
weighted-Recall	0.88	0.89	0.86	0.91
weighted-Precision	0.88	0.90	0.87	0.91

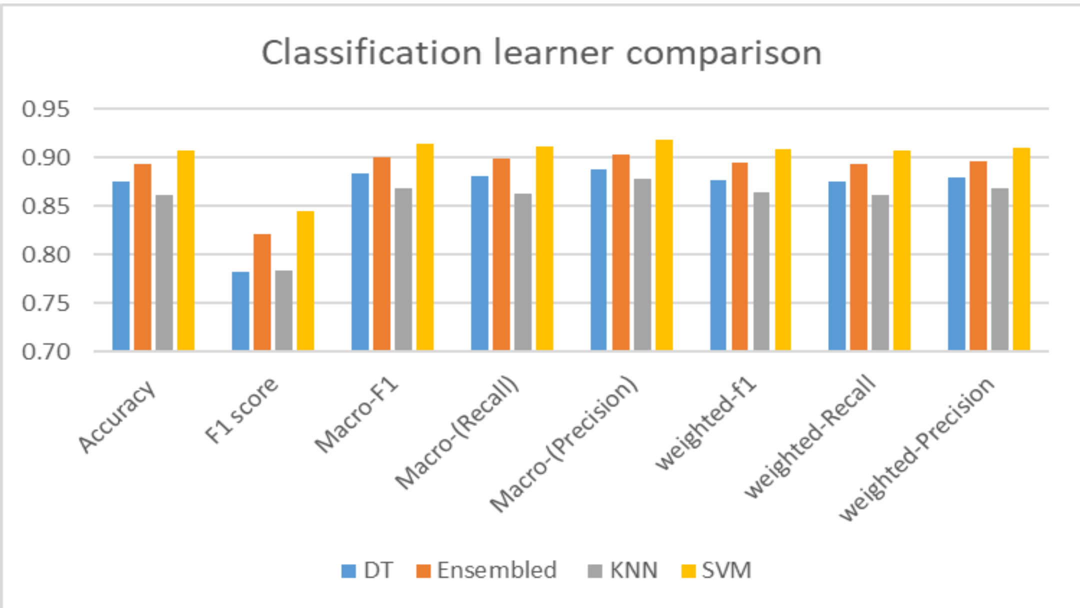

Figure 4.5 Classification learner comparison chart

4.2.1.5 Multi-Layer Perceptron Neural Networks (MLPNNs)

Neural networks are one of the machine learning techniques that have proven their ability in classification and prediction in recent years, so they are also used in classification for multiple and simple classes. In our experiments, the classification results show that they are the best among all the classification techniques that are previously used. In this section, we will present the results obtained when applying neural network techniques, so that the best results are taken after applying many tests with the lowest possible number of neurons which is 10, as reducing the number contributes to reduce computational operations and save energy.

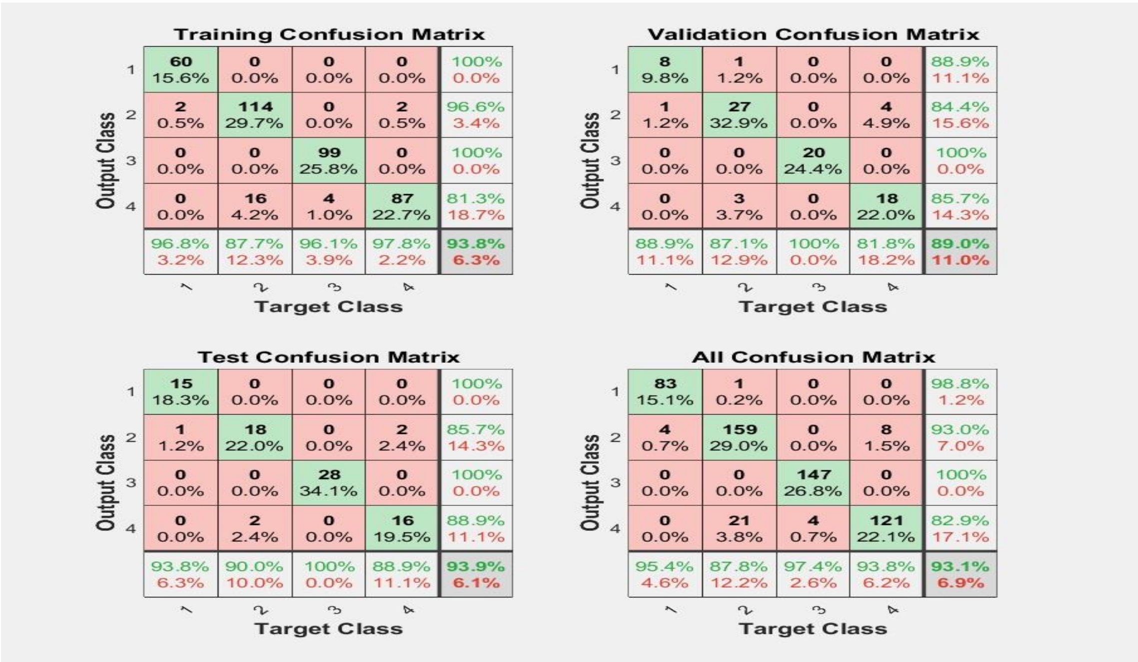

Figure 4.6 MLP 10 neuron confusion matrix

Table 4.6 MLP 10 neuron classification matrices

	Class 0	Class 1	Class 2	Class 3
TP	85	159	147	121
TN	460	357	399	396
FP	4	22	4	8
FN	1	12	0	25
(Recall)	0.99	0.93	1.00	0.83
False Negative Rate	0.01	0.07	0.00	0.17
(Precision)	0.96	0.88	0.97	0.94
False Discovery Rate	0.04	0.12	0.03	0.06
Specificity	0.99	0.94	0.99	0.98
FPR = (1-Specificity)	0.01	0.06	0.01	0.02
Accuracy	0.93			
F1 score	0.97	0.90	0.99	0.88
Macro-F1	0.94			
Macro-(Recall)	0.94			
Macro-(Precision)	0.94			
weighted-f1	0.93			
weighted-Recall	0.93			
weighted-Precision	0.93			

Note that the classification accuracy is 93%, the Macro-F1 value is 0.94, and the result of weighted-f1 is 0.93. It is noted that the result of classification of the second classes in neural networks techniques in the F1 scale is the lowest, and since the number of samples in this class has an effect, the weighted F1 results are lower than the macro.

The following ROC curve shows the ability of these techniques to classify the students' performance well, as the values of the AUC are more than 93 %. In contrast, the value of sensitivity and (1-specificity) for the first class is 0.96, 0.01, 0.88for the second class, 0.06 for the third class, and 0.94, 0.02 for the fourth class.

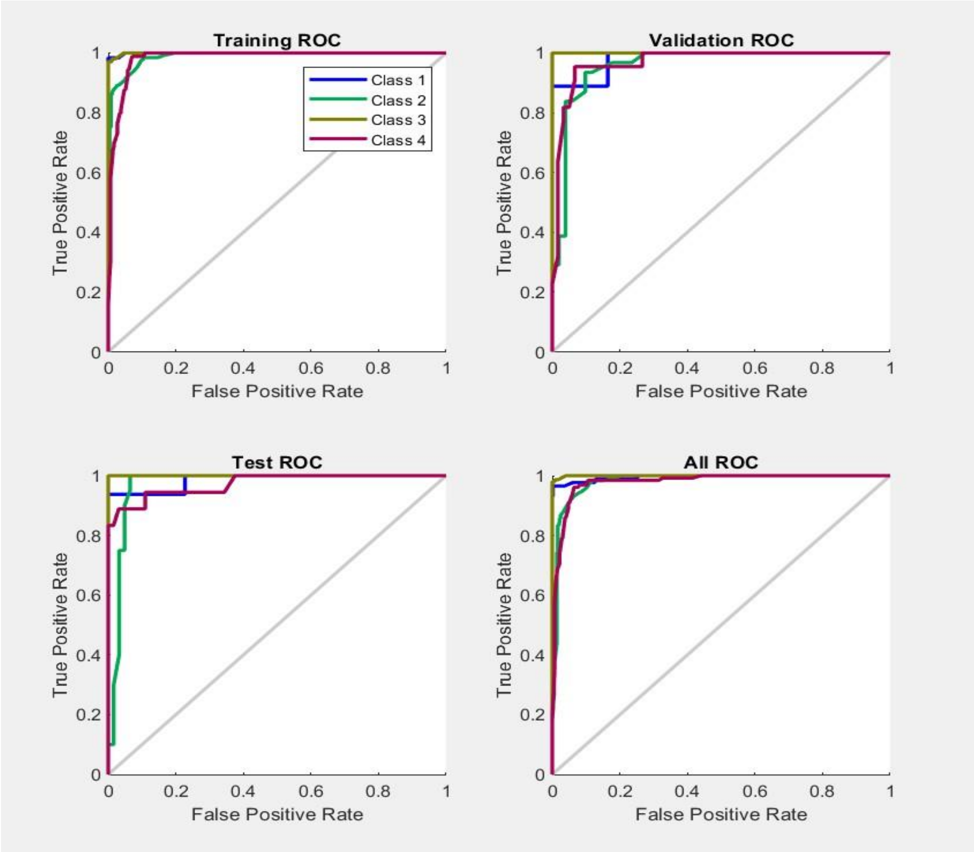

Figure 4.7 MLP 10 neuron ROC curve

The results produced by the MLPNNs are compared with the best classification learner mechanism in this section by comparing Accuracy, F1 score, Macro-F1, and weighted-f1 for each technique. The table 4.7 and figure 4.8 below show the superiority of neural network techniques, and this shows the strength of neural networks in prediction and classification operations.

Table 4.7 SVM VS MLP 10 neurons

	SVM	MLPNNs
Accuracy	0.91	0.93
F1 score	0.84	0.97
Macro-F1	0.91	0.94
Macro-(Recall)	0.91	0.94
Macro-(Precision)	0.92	0.94
weighted-f1	0.91	0.93
weighted-Recall	0.91	0.93
weighted-Precision	0.91	0.93

Furthermore, there is a noticeable difference in the result of classification coefficients appearing on Macro-F1 and weighted-F1.

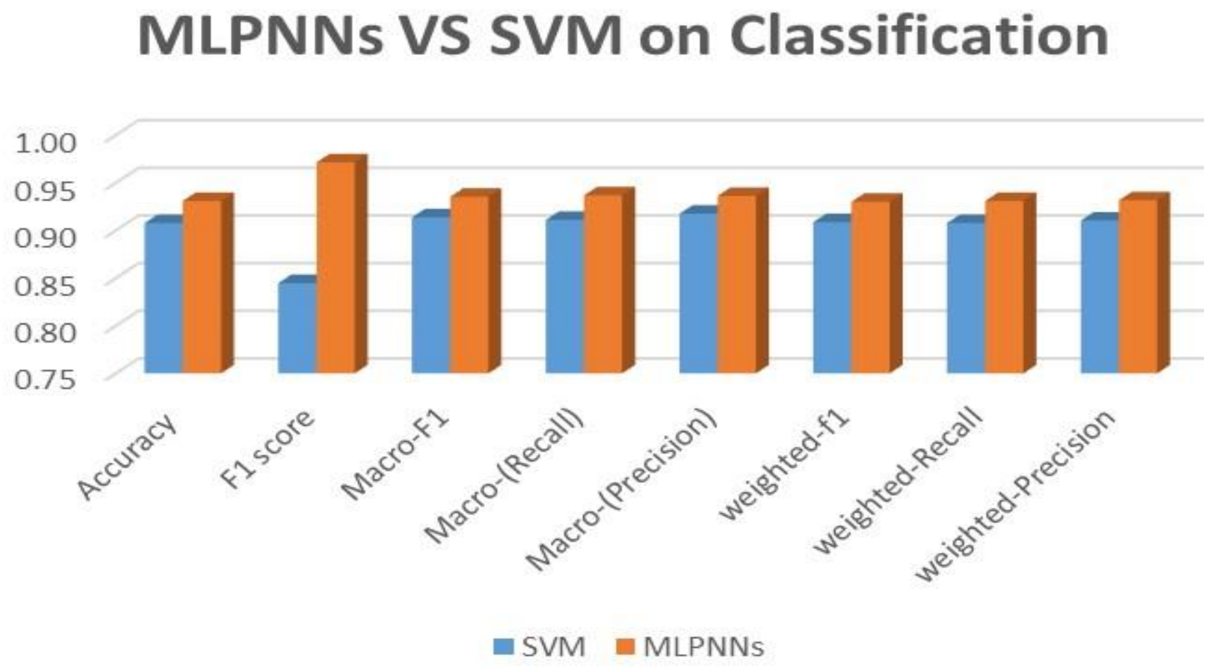

Figure 4.8 SVM vs MLPNNs

4.3 Results Discussion

To assess the effectiveness of various machine learning algorithms, such as k-nearest Neighbors (k-NN), Support Vector Machines (SVM), Ensembled Trees (e.g., Random Forest or Gradient Boosting), Decision Trees, and Multi-Layer Perceptron Neural Networks (MLPNNs), it is necessary to take into account several factors, including their capabilities and the specific problem at hand. Each algorithm possesses unique advantages and disadvantages, and the selection of the “optimal” algorithm depends on the specific circumstances and data at hand.

Below is a concise comparison of various algorithms:

The k-Nearest Neighbors (k-NN) algorithm is a straightforward and intuitive method for classification and regression tasks. It is effective when there are distinct patterns in the data and is particularly valuable for datasets of small to medium size. Nevertheless, its performance may be suboptimal in scenarios involving high-dimensional spaces or noisy data. This algorithm gives bad results for our study.

Decision Trees: Decision trees possess a high level of interpretability and can be easily visualized. They are beneficial when seeking comprehension of the cognitive process involved in making decisions. Nevertheless, deep trees are particularly susceptible to overfitting.

Ensemble Trees: Ensemble techniques such as Random Forest amalgamate numerous decision trees to enhance forecast accuracy. They possess strong durability, the ability to manage diverse data types, and the capacity to capture intricate relationships within data. They are frequently a suitable option for a diverse array of issues.

The decision to exclude Decision Trees and Ensembled Trees suggests a need to address issues like overfitting, complexity, and data scalability. Exploring alternative methods, such as neural networks or support vector machines, might better suit the study's requirements for improved model performance and relevance.

Support Vector Machines (SVM) are highly effective for classification and regression tasks, especially when the data exhibits a distinct separation boundary. Support Vector Machines (SVM) are capable of effectively processing data with a large number of dimensions and are resistant to the problem of overfitting. Nevertheless, it might be demanding regarding computational resources when dealing with extensive datasets. This justifies its superiority over classification learner algorithms for this study.

Multi-layer perceptron Neural Networks (MLPNNs) are a specific category of deep learning models. Machine learning models can comprehend intricate and non-linear connections within data, which makes them well-suited for a diverse array of tasks, such as analyzing image and text data. They possess a great degree of flexibility and can autonomously acquire feature representations.

Based on the study results, Multi-Layer Perceptron Neural Networks (MLPNNs) emerge as the best choice. MLPNNs capture complex, non-linear relationships in the data, making them suitable for the study's objectives. Their flexibility and automatic feature learning contribute to their superior performance.

Nevertheless, it is crucial to acknowledge that MLPNNs may not always be the optimal selection. So, conducting experiments with various algorithms and assessing their effectiveness on your specific dataset is crucial to identifying the most appropriate strategy because the nature of the dataset changes from one dataset to another depending on the features selected and the case chosen.

CHAPTER FIVE

CONCLUSION AND FUTURE WORK

5.1 Conclusion

Because education is fundamental in our state of Palestine, Palestinian universities, especially the Arab American University, makes the Moodle available as an electronic educational platform. This trend toward online learning and teaching begins in earnest in our academic institutions and universities after the spread of the Corona epidemic. That's why it's critical that we learn more about how students are faring on these platforms so that policymakers can better inform their decisions and aid the growth of education at colleges and universities.

The data from students enrolled in those courses over the period of three semesters are used to construct the model from which the information is collected. The tables are taken from the databases by using the SQL statements for doing so. The tables are then connected in a way that ensures pupils' confidentiality. Based on the allowed transactions by the university administration, the data is collected from approximately 500 male and female students from all three semesters who are deemed to be in a good health.

Following processing, these data are exposed to a variety of machine learning and neural network techniques with four approaches employed for categorization in accordance with their respective related group using MATLAB-supported KNN algorithms. TREE Classification accuracy is found as high as 90% when using Ensembled Tree and SVM, and as high as 93% when using MLPNNs. This indicates that it is a type of neural network that is found to be superior to other classification algorithms.

5.2 Future Work

Future considerations are directed to increase the efficiency of the system and improve the results of classification in students' performance. To reach this end, the work must be done to select more features capable of giving better results on the one hand. On the other hand, it is done to use new technologies, especially hybrid ones, and to give more accurate prediction and classification results on a more accurate level of students' performance.

Also, attention is directed towards obtaining larger quantities of data from the university itself and other universities so that the data sample is comprehensive and more accurate. The increase in the size of the dataset increases the system's power to classify the various samples entered for it in the future.

We also seek to develop mechanisms for integrating machine learning techniques with software that serves decision-makers in employing and evaluating the results that are issued by them and their satisfaction with these results in the future. So that, we have a future reference in carrying out several improvements to the system in line with the current situation of Palestinian universities.

References

1. Alhassan, B. Zafar, A. Mueen, “Predict Students’ Academic Performance based on their Assessment Grades and Online Activity Data”, *International Journal of Advanced Computer Science and Applications. May 2020*, 2020.
2. Halees, “Mining Students Data to Analyze Learning Behavior: A Case Study”, Department of Computer Science, Islamic University of Gaza P.O.Box 108 Gaza, Palestine.
3. Mueen, B. Zafar, U. Manzoor, “Modeling and Predicting Students' Academic Performance Using Data Mining Techniques”, *International Journal of Modern Education and Computer Science*, 2016.
4. Shahiria, W. Husaina, N Rashida, “A Review on Predicting Student’s Performance using Data Mining Techniques”, *The Third Information Systems International Conference Procedia Computer Science 72 (2015) 414 – 422*, 2015.
5. Albreiki, N. Zaki, H. Alashwal, “A Systematic Literature Review of Student’ Performance Prediction Using Machine Learning Techniques”, https://www.mdpi.com/journal/education, 2021.
6. Mbouzao, et al, “Early Prediction of Success in MOOC from Video Interaction Features”, Springer Nature Switzerland AG, 2020.
7. Costa, H. Alvelos, L. Teixeira, “The use of Moodle e-learning platform: a study in a Portuguese University”, *CENTERIS 2012 - Conference on* Enterprise *Information Systems Procedia Technology 5 (2012) 334 – 343*, 2012.

8. Galy, C. Downey, J. Johnson, "The Effect of Using E-Learning Tools in Online and Campus-based Classrooms on Student Performance", *Journal of Information Technology Education*, 2011.

9. Qiu1, et al, "Predicting students' performance in e-learning using learning process and behaviour data", www.nature.com/scientificreports, 2022.

10. Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. IBM Journal of research and development, 3(3), 210-229.

11. Aljawarneh, S. A. (2020). Reviewing and exploring innovative ubiquitous learning tools in higher education. Journal of computing in higher education, 32, 57-73.

12. Byrnes, K. G., Kiely, P. A., Dunne, C. P., McDermott, K. W., & Coffey, J. C. (2021). Communication, collaboration and contagion:"Virtualisation" of anatomy during COVID-19. Clinical anatomy, 34(1), 82-89.

13. Capterra (2021). LMS software. https://www.capterra.com/learning-managementsyste software/?feature=%5B38347%5D&sortOrder=sponsored Accessed 27 Nov 2021

14. Setiadi, P. M., Alia, D., Sumardi, S., Respati, R., & Nur, L. (2021, July). Synchronous or asynchronous? Various online learning platforms studied in Indonesia 2015-2020. In Journal of Physics: Conference Series (Vol. 1987, No. 1, p. 012016). IOP Publishing.

15. Henrick, B. M., Hutton, A. A., Palumbo, M. C., Casaburi, G., Mitchell, R. D., Underwood, M. A., ... & Frese, S. A. (2018). Elevated fecal pH indicates a profound change in the breastfed infant gut microbiome due to reduction of Bifidobacterium over the past century. MSphere, 3(2), e00041-18.

16. Bigler, D., & Hagel, G. (2023, June). Technical Report: Define a customized course and import it into Moodle without changes to the configuration of the Moodle

system. In Proceedings of the 5th European Conference on Software Engineering Education (pp. 180-183).

17. Hwang, W. Y., Shadiev, R., Wang, C. Y., & Huang, Z. H. (2012). A pilot study of cooperative programming learning behavior and its relationship with students' learning performance. Computers & education, 58(4), 1267-1281.
18. Shaw, R. S. (2012). A study of the relationships among learning styles, participation types, and performance in programming language learning supported by online forums. Computers & Education, 58(1), 111-120.
19. Hellas, A., Ihantola, P., Petersen, A., Ajanovski, V. V., Gutica, M., Hynninen, T., ... & Liao, S. N. (2018, July). Predicting academic performance: a systematic literature review. In Proceedings companion of the 23rd annual ACM conference on innovation and technology in computer science education (pp. 175-199).
20. Goldberg, D. E. (1990). Real-coded genetic algorithms, virtual alphabets and blocking. Champaign: University of Illinois at Urbana Champaign.
21. N. Kadoića, D. Oreški, “Analysis of Student Behavior and Success Based on Logs in Moodle”, *MIPRO 2018/CE*, 2018.
22. P. Shayan, M, van Zaanen, “Predicting Student Performance from Their Behavior in Learning Management Systems”, *International Journal of Information and Education Technology (IJIET)*, 2019.
23. Llorens-Largo, F., Gallego-Durán, F. J., Villagrá-Arnedo, C. J., Compañ-Rosique, P., Satorre-Cuerda, R., & Molina-Carmona, R. (2016). Gamification of the learning process: lessons learned. IEEE Revista Iberoamericana de tecnologías del aprendizaje, 11(4), 227-234.

24. Q. Fua, Z. Gaob, J. Zhoua, Y. Zheng, “CLSA: A novel deep learning model for MOOC dropout prediction”, https://www.sciencedirect.com, 2021.

25. R, Conijn, C, Snijders, Ad, Kleingeld, U, Matzat, “Predicting Student Performance from LMS Data: A Comparison of 17 Blended Courses Using Moodle LMS”, *IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES, VOL. 10, NO. 1, JANUARY-MARCH 2017*, 2017.

26. S. Sahebi, P. Brusilovsky , “Student Performance Prediction by Discovering Inter-Activity Relations”, *International Conference on Educational Data Mining*, 2008.

27. S. Yousaf, P. Ahmadi, K. Islam, “A Novel Study of the Relation between Students’ Navigational Behavior on Blackboard and their Learning Performance in an Undergraduate Networking Course”, *George Mason University*, 2018.

28. Sushil Shrestha, Manish Pokharel, “Educational data mining in Moodle data”, *International Journal of Informatics and Communication Technology*, Vol. 10, No. 1, April 2021.

29. T. Mai, M, Bezbradica, M. Crane, “Learning behaviours data in programming education: Community analysis and outcome prediction with cleaned data”, *Published by Elsevier B.V.* www.elsevier.com/locate/fgcs, 2021.

30. TN Doan S. Sahebi, “Rank-Based Tensor Factorization for Student Performance Prediction”, *Proceedings of the 12th International Conference on Educational Data Mining*, 2019.

31. Hurwitz, J., & Kirsch, D. (2018). Machine Learning For Dummies®, IBM Limited Edition Published by John Wiley & Sons, Inc. 111 River St. Hoboken, NJ, 07030-5774.

32. Y. Yang, et al, “Predicting course achievement of university students based on their procrastination behaviour on Moodle”, *Springer-Verlag GmbH Germany, part of Springer Nature 2020*, 2020.

33. Ilyas, I. F., & Rekatsinas, T. (2022). Machine Learning and Data Cleaning: Which Serves the Other?. ACM Journal of Data and Information Quality (JDIQ), 14(3), 1-11.

34. Saheed, Y. K., Abiodun, A. I., Misra, S., Holone, M. K., & Colomo-Palacios, R. (2022). A machine learning-based intrusion detection for detecting internet of things network attacks. Alexandria Engineering Journal, 61(12), 9395-9409.

35. Biswas, N., Uddin, K. M. M., Rikta, S. T., & Dey, S. K. (2022). A comparative analysis of machine learning classifiers for stroke prediction: A predictive analytics approach. Healthcare Analytics, 2, 100116.

36. Hu, Y., & Castro-Lacouture, D. (2019). Clash relevance prediction based on machine learning. Journal of computing in civil engineering, 33(2), 04018060.

37. Malik, A., Tikhamarine, Y., Sihag, P., Shahid, S., Jamei, M., & Karbasi, M. (2022). Predicting daily soil temperature at multiple depths using hybrid machine learning models for a semi-arid region in Punjab, India. Environmental Science and Pollution Research, 29 (47), 71270-71289..

38. Mohammadi, B., Safari, M. J. S., & Vazifehkhah, S. (2022). IHACRES, GR4J and MISD-based multi conceptual-machine learning approach for rainfall-runoff modeling. Scientific Reports, 12(1), 12096.

39. Mall, P. K., Yadav, R. K., Rai, A. K., Narayan, V., & Srivastava, S. (2022). Early Warning Signs Of Parkinson’s Disease Prediction Using Machine Learning Technique. Journal of Pharmaceutical Negative Results, 4784-4792.

40. Zhang, Y., Liu, J., & Shen, W. (2022). A review of ensemble learning algorithms used in remote sensing applications. Applied Sciences, 12(17), 8654.

41. Zehra, S. S., Magarini, M., Qureshi, R., Mustafa, S. M. N., & Farooq, F. (2022). Proactive approach for preamble detection in 5g-nr prach using supervised machine learning and ensemble model. Scientific reports, 12(1), 8378.

42. Jamei, M., Karbasi, M., Ali, M., Malik, A., Chu, X., & Yaseen, Z. M. (2023). A novel global solar exposure forecasting model based on air temperature: Designing a new multi-processing ensemble deep learning paradigm. Expert Systems with Applications, 222, 119811.

43. Begam, B., Palanivelan, M., & Preethi, S. (2023, April). An Ensemble Machine Learning Algorithm To Diagnose Alzheimer's Disease. In 2023 International Conference on Recent Advances in Electrical, Electronics, Ubiquitous Communication, and Computational Intelligence (RAEEUCCI) (pp. 1-6). IEEE.

44. Okey, O. D., Maidin, S. S., Adasme, P., Lopes Rosa, R., Saadi, M., Carrillo Melgarejo, D., & Zegarra Rodríguez, D. (2022). BoostedEnML: Efficient technique for detecting cyberattacks in IoT systems using boosted ensemble machine learning. Sensors, 22(19), 7409.

45. Chang, Z., Catani, F., Huang, F., Liu, G., Meena, S. R., Huang, J., & Zhou, C. (2023). Landslide susceptibility prediction using slope unit-based machine learning models considering the heterogeneity of conditioning factors. Journal of Rock Mechanics and Geotechnical Engineering, 15(5), 1127-1143.

46. Ghimire, S., Deo, R. C., Casillas-Pérez, D., Salcedo-Sanz, S., Sharma, E., & Ali, M. (2022). Deep learning CNN-LSTM-MLP hybrid fusion model for feature optimizations and daily solar radiation prediction. Measurement, 202, 111759.

47. Sivasankari, S. S., Surendiran, J., Yuvaraj, N., Ramkumar, M., Ravi, C. N., & Vidhya, R. G. (2022, April). Classification of diabetes using multilayer perceptron. In 2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE) (pp. 1-5). IEEE.

48. Khan, A., Khan, S. H., Saif, M., Batool, A., Sohail, A., & Waleed Khan, M. (2023). A Survey of Deep Learning Techniques for the Analysis of COVID-19 and their usability for Detecting Omicron. Journal of Experimental & Theoretical Artificial Intelligence, 1-43.

49. Ayawah, P. E., Sebbeh-Newton, S., Azure, J. W., Kaba, A. G., Anani, A., Bansah, S., & Zabidi, H. (2022). A review and case study of Artificial intelligence and Machine learning methods used for ground condition prediction ahead of tunnel boring Machines. Tunnelling and Underground Space Technology, 125, 104497.

50. Ul Hassan, I., Ali, R. H., Ul Abideen, Z., Khan, T. A., & Kouatly, R. (2022). Significance of machine learning for detection of malicious websites on an unbalanced dataset. Digital, 2(4), 501-519.

51. Niyogisubizo, J., Liao, L., Nziyumva, E., Murwanashyaka, E., & Nshimyumukiza, P. C. (2022). Predicting student's dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization. Computers and Education: Artificial Intelligence, 3, 100066.

52. Memiş, S., Enginoğlu, S., & Erkan, U. (2022). A classification method in machine learning based on soft decision-making via fuzzy parameterized fuzzy soft matrices. Soft Computing, 26(3), 1165-1180.

53. Pragasam, T. T. N., Thomas, J. V. J., Vensuslaus, M. A., & Radhakrishnan, S. (2023). CEAT: Categorising Ethereum Addresses' Transaction Behaviour with Ensemble Machine Learning Algorithms. Computation, 11(8), 156.

54. Shakya, S., Mukherjee, A., Halder, R., Maiti, A., & Chaturvedi, A. (2022, August). Smartmixmodel: machine learning-based vulnerability detection of solidity smart contracts. In 2022 IEEE international conference on blockchain (Blockchain) (pp. 37-44). IEEE.

Appendices

Appendix A:

Arab American University
Tel: 04-241-8888

الجامعـــة العربيـــة الأمريكيـــة
تلفون: 04-241-8888

الوصول للبيانات على منصة مودل

حضرة السادة في الشؤون الاكاديمية المحترمين،
تحية طيبة،

انا الطالبة نورا اسماعيل شوارب (201720258) ، تخصص ماجستير علم حاسوب ، أتقدم الى حضرتكم بطلب للحصول على بيانات الجامعة المتوفرة على منصة المودل لمساقات الثقافة العامة على الأقل ثلاثة مساقات من المساقات المذكورة ادناه في السنوات الثلاثة الأخيرة 2020-2021-2022 بما فيها الفصل الصيفي للعمل عليها في موضوع رسالة الماجستير بعنوان (Predicting Student's Performance based on Data and Online Behaviour Engagement Activities Using Moodle)، والمشرف هو الدكتور احمد عويس.
فيما يلى أسماء مساقات الثقافة العامة التي بالإمكان الحصول على البيانات المتوفرة عن ثلاثة منها على الأقل على منصة المودل :

- Introduction to psychology
- Civil society organisation
- Current world issues
- Jerusalem: Civilisation and history
- History of civilisation
- Modern Arab thought
- Critical thinking
- Physical education
- Democracy and human rights
- International relations
- Low in our life
- Arabic language
- Israel & Zionism studies I
- Islamic Culture
- Palestinian studies
- Election and political participation
- Fundamentals of research methods
- The Palestinian prisoners
- Movement

وفيما يلى بعض الخصائص التى ارغب بالوصول اليها للعمل عليها في موضوع الرسالة:

- activities data derived from
- online learning platform (AAUP-Moodle
- the last login time

Arab American University
Tel: 04-241-8888

الجامعـــة العربيـــة الأمريكيـــة
تلفون: 04-241-8888

- time spent on Moodle
- the number of lessons read and downloaded, lectures, files and others
- largest period of inactivity
- the number of quizzes
- academic records
- assignments completed

Appendix B:

اسم الطالبة: نورا اسماعيل شوارب.

التخصص: ماجستير علم حاسوب.

الكلية: الهندسة وتكنولوجيا المعلومات.

الرقم الجامعي: 201720258.

الموضوع: الحصول على كتاب تسهيل مهمة بحثية.

تحية طيبة وبعد.

اتقدم من حضرتكم بطلب للحصول على بيانات الجامعة المتوفرة على منصة المودل المذكورة في ملحق رقم (1)، لمساقات الثقافة العامة المرفق في ملحق رقم (2)، في الفصول الدراسية لسنوات 2021 / 2022 بما فيها الفصل الصيفي للعمل عليها في موضوع رسالتي وهو:

Predicting Student's Performance based on Behaviour Data and Online Engagement Activities Using Moodle

سيتم استخدام البيانات لأغراض البحث العلمي فقط وحسب الأصول المتبعة والقوانين السارية في الجامعة العربية الأمريكية.

آملاً من حضرتكم التكرم بالقبول وتزويدي بالبيانات اللازمة.

دمتم منارة للعلم

ملحق رقم: (1): السمات التي ارغب بالوصول اليها للعمل عليها في موضوع الرسالة:

Activities Data Derived From Online Learning Platform (AAUP-Moodle)

اسم البيان (السمة)	الرقم
Identification number of the student	1
Identification number of the course	2
Number of sessions	3
Number of assignments done	4
Number of quizzes taken	5
Average score on quizzes	6
Number of messages sent to the forum	7
Number of messages read on the forum	8
Time spent on Moodle	9
Total time spent on assignments	10
Total time spent on quizzes	11
Total time spent on forum	12
Final score of the student obtained in a course	13
The last login time	14
The number of lessons read and/or downloaded	15
The number of Lectures taken and/or downloaded	16
Largest period of inactivity	17
The number of Files taken and/or downloaded	18
The number of Files uploaded	19
Academic records	20

ملحق رقم (2): أسماء مساقات الثقافة العامة التي اريد الحصول على البيانات المتوفرة على منصة مودل في الجامعة لثلاث منها على الاقل:

اسم المساق	الرقم
Introduction to psychology	1
Civil society organisation	2
Current world issues	3
Jerusalem: Civilisation and history	4
History of civilisation	5
Modern Arab thought	6
Critical thinking	7
Physical education	8
Democracy and human rights	9
International relations	10
Low in our life	11
Arabic language	12
Israel & Zionism studies I	13
Islamic Culture	14
Palestinian studies	15
Election and political participation	16
Fundamentals of research methods	17
The Palestinian prisoners	18
Movement	19

الملخص

توقع أداء الطلاب مهم جدا في تطوير أداء الجامعات الفلسطينية خاصة بعد انتشار المنصات يساهم التنبؤ المبكر في تطوير النظام التعليمي وتطوير برامج التعليمية نتيجة انتشار وباء كورونا. لمعالجة الضعف الأكاديمي لدى الطلاب، كما يسهم في تحديد احتياجات الموارد البشرية للجامعات كل ما سبق الفلسطينية والمساهمة في وضع خطط التوسع للمنظومة التكنولوجية لخدمة هذا المجال، هو نتيجة للتنبؤ المبكر بأداء الطلاب الأكاديمي عنيت هذه الرسالة بتحديد الأداء المستقبلي لطلاب الجامعة العربية الأمريكية بناءً على نموذج تم يعتمد تصنيف الأداء على تقنيات التعلم الآلي وقدرتها على التنبؤ والتصنيف. المنصة التعليمية.

تمت استشارة المتخصصين تحديد الميزات التي يجب الاعتماد عليها في جمع البيانات بدقة متناهية. في هذا المجال، كما تمت الإشارة إلى العديد من الدارسات السابقة التي تتناول أداء الطلاب وتدعم هذا المجال البحثي.

خضعت البيانات لعمليات معالجة مسبقة للتخلص من القسم العشوائي ومعالجات القيمة الفارغة بعد ذلك خضعت هذه البيانات لعمليات معالجة وحذف البيانات المكررة والعديد من العمليات الهامة. من خلال تقنيات مختلفة للتعلم الآلي وكذلك لتقنيات الشبكة العصبية، حيث تم استخدام أربع تقنيات وفقًا لمجموعة مرتبطة من الخوارزميات المدعومة من Ensembled

SVM،Tree، تفوقت على تقنيات التصنيف الأخرى، حيث وصلت دقة التصنيف في SVM

لوحظ أن تقنية تساوي 87٪ وأيضًا عند استخدام الشبكات KNN هذه التقنية إلى 90٪، بينما

كانت قيمة التصنيف في، MLPNNs العصبية وجد أن دقة التصنيف أعطت أفضل النتائج في الشبكات العصبية من النوع هنا، نلاحظ بوضوح قدرة الشبكات العصبية على التصنيف والتنبؤ

بدقة .كانت دقة التصنيف 93 ٪ .عالية وهذا يقودنا إلى العديد من الدارسات حول إد ارج

تقنيات أخرى للشبكات العصبية في المستقبل.