4-2023

DATA-DRIVEN MODELING OF STUDENT PERFORMANCE IN THE TIME OF DISTANCE LEARNING

Iman Saad Megdadi
United Arab Emirates University

UAEU
جامعة الإمارات العربية المتحدة
United Arab Emirates University

MASTER THESIS NO. 2023: 12

College of Information Technology

Department of Computer Science and Software Engineering

DATA-DRIVEN MODELING OF STUDENT PERFORMANCE IN THE TIME OF DISTANCE LEARNING

Iman Saad Megdadi

April 2023

United Arab Emirates University

College of Information Technology

Department of Computer Science and Software Engineering

DATA-DRIVEN MODELING OF STUDENT PERFORMANCE IN THE TIME OF DISTANCE LEARNING

Iman Saad Megdadi

This thesis is submitted in partial fulfilment of the requirements for the degree of Master of Software Engineering

April 2023

United Arab Emirates University Master Thesis
2023: 12

Cover: A representative image of Machine Learning in Education, obtained from Dataquest Magazine website

https://www.dqindia.com/ai-transforming-education-industry/

Print: University Print Service, UAEU 2023

Declaration of Original Work

I, Iman Saad Megdadi, the undersigned, a graduate student at the United Arab Emirates University (UAEU), and the author of this thesis entitled "*Data-Driven Modeling of Student Performance in the Time of Distance Learning*", hereby, solemnly declare that this is the original research work done by me under the supervision of Dr. Salah Bouktif, in the College of Information Technology at UAEU. This work has not previously formed the basis for the award of any academic degree, diploma or a similar title at this or any other university. Any materials borrowed from other sources (whether published or unpublished) and relied upon or included in my thesis have been properly cited and acknowledged in accordance with appropriate academic conventions. I further declare that there is no potential conflict of interest with respect to the research, data collection, authorship, presentation and/or publication of this thesis.

Student's Signature: ___________

Date: __01/05/2023______________

Approval of the Master Thesis

This Master Thesis is approved by the following Examining Committee Members:

1) Advisor (Committee Chair): Dr. Salah Bouktif

   Title: Associate Professor

   Department of Computer Science and Software Engineering

   College of Information Technology

   Signature ______________________ Date 14/05/2023

2) Member: Prof. Zouheir Trabelsi

   Title: Professor

   Department of Information Systems and Security

   College of Information Technology

   Signature Zouheir Date 14/05/2023

3) Member (External Examiner): Dr. Sujith Samuel Mathew

   Title: Associate Professor

   College of Interdisciplinary Studies

   Institution: Zayed University – UAE

   Signature ____________ Date 14/May/2023

This Master Thesis is accepted by:

Dean of the College of Information Technology: Professor Taieb Znati

Signature ___Taieb Znati___________________

Date 15/05/2023______________

Dean of the College of Graduate Studies: Professor Ali Al-Marzouqi

Signature Ali Hassan___________

Date ______17/05/2023______

Abstract

One of the important aspects that all academic institutions work towards improving is Student Performance. It is obviously the primary indicator of success or failure of institutions. Student performance predictions are vital to instructors and educational decision makers to help, across all levels, tailor learning according to the students' needs. Therefore, it is essential for Higher Education Institutions to predict student performance in distance learning which has been, and remains, the primary method of learning in some countries due to Corona Virus pandemic. For this reason, this research is going to predetermine a fitting definition of student performance in time of distance learning by surveying literature and collecting new effective factors affecting students' performance. New concepts and attributes are discovered and considered in the new definition. Furthermore, a primary objective of this thesis is to build a prediction model for student performance during distance learning, where the new definition and its subsequent attributes are considered. The data-driven model is empirically validated, and the obtained results show the outperformance of our proposed approach; particularly the appropriateness of the introduced student performance definition as well as the machine learning based technique from which the student performance prediction model is derived. We strongly believe that such a model will benefit the educators and guide them on how to accurately make decisions based on student performance in the distance learning settings.

Keywords: Student Performance, Distance learning, Online Learning, Machine Learning.

Abstract and Title in Arabic

النمذجة القائمة على البيانات لأداء الطلاب في وقت التعلم عن بعد

الملخص

أحد الجوانب المهمة التي تعمل جميع المؤسسات الأكاديمية على تحسينها هو أداء الطلاب. من الواضح أنه المؤشر الأساسي لنجاح أو فشل المؤسسات. تعد تنبؤات أداء الطلاب أمرًا حيويًا للمدرسين وصناع القرار التربويين للمساعدة، على جميع المستويات، في تصميم التعلم وفقًا لاحتياجات الطلاب. لذلك، من الضروري لمؤسسات التعليم العالي أن تتنبأ بأداء الطلاب في التعلم عن بعد الذي كان ولا يزال الطريقة الأساسية للتعلم في بعض البلدان بسبب جائحة فيروس كورونا. لهذا السبب، سيحدد هذا البحث مسبقًا تعريفًا مناسبًا لأداء الطلاب في وقت التعلم عن بعد من خلال مسح الأدبيات وجمع العوامل الفعالة الجديدة التي تؤثر على أداء الطلاب. يتم اكتشاف مفاهيم وسمات جديدة وأخذها في الاعتبار في التعريف الجديد. علاوة على ذلك، فإن الهدف الأساسي لهذه الأطروحة هو بناء نموذج تنبؤ لأداء الطالب أثناء التعلم عن بعد، حيث يتم النظر في التعريف الجديد والسمات اللاحقة له. تم التحقق من صحة النموذج المستند إلى البيانات تجريبياً، وتظهر النتائج التي تم الحصول عليها تفوق نهجنا المقترح؛ لا سيما مدى ملاءمة تعريف أداء الطالب المقدم بالإضافة إلى تقنية التعلم الآلي القائمة على نموذج التنبؤ بأداء الطالب. نحن نؤمن بشدة أن مثل هذا النموذج سيفيد المعلمين ويرشدهم حول كيفية اتخاذ القرارات بدقة بناءً على أداء الطالب في إعدادات التعلم عن بعد.

مفاهيم البحث الرئيسية: أداء الطلاب، التعلم عن بعد، التعلم عبر الانترنت، التعلم الآلي.

Author Profile

Iman Saad Megdadi is currently a Computer Science Teacher at the Institute of Applied technology, UAE. She has more than ten years of experience working with leading educational institutions to help in the development of the youth people in the country. Iman lives in Dubai with her family. She received her bachelor's degree in the UAE from the University of Sharjah.

Acknowledgements

I would like to thank my committee for their guidance, support, and assistance throughout my preparation of this thesis, especially my advisor Dr. Salah Bouktif for his patience and continuous support during my journey in the masters and in completing my thesis.

Dedication

To my beloved husband, parents, and family

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ANN	Artificial Neural Networks
CV	Cross Validation
DL	Distance Learning
DT	Decision Trees
ET	Extra Trees
FN	False Negative
FP	False Positive
FS	Feature Selection
HEI	High Educational Institutions
LMS	Learning Management System
LR	Logistic Regression
ML	Machine Learning
RF	Random Forest
SP	Student Performance
SVM	Support Vector Machine
TN	True Negative
TP	True Positive

Chapter 1: Introduction

Shortly after the outbreak of the Corona Virus (COVID-19), the World Health Organization (WHO) declared it a pandemic in March 2020. COVID-19 had a monumental impact on educational institutions in many ways. Universities worldwide were obligated to shift to remote teaching in order to comply with social distancing and to preserve students' health and safety during a time of an unprecedented challenge, while maintaining stability for students during the panic of the pandemic. That posed challenges for the majority of students as well as the academic staff, where everyone had to adapt to this rapid shift as there were no other options available at the time.

Student performance (SP) prediction in Higher Educational Institutions (HEI) has been recognized and continues to be one of the most significant research problems in DL as described by (Wang et al., 2018). They consider SP as the pivot around which the whole educational system revolves. Narad & Abdullah (2016) mentioned in their article that any educational institution's success or failure is determined by the academic SP. Brahim (2022) stated that it offers important insights that can help and provide guidance to institutions to be able to make timely decisions and changes, leading to improved student outcome achievement. As it is advocated by (Kumar, 2021), high achieving students who exceed expectations and are above the norm within their society, are commonly expected to contribute to the sustainable growth of the economy. Kumar (2021) also emphasized the strong connection between student academic performance and the socio-economic development of a country. This connection is explicitly visible given that the acquisition of knowledge and skills that shape the performance of the students has a tangible impact by the participation of the graduates in the socio-economic activities.

During the pandemic, most of the classes were being taught remotely. Furthermore, there are various attributes that are being used to predict and model the SP in distance learning time. This research is going to identify the attributes that can be used to propose a more appropriate definition of academic performance of students in the DL time. This definition, along with the discovered attributes of performance in DL time will give the opportunity to build a rigorous data-driven model of the students' performance to achieve the aim of this research. That, in turn, will contribute to helping universities reach and surpass their students' performance goals.

1.1 Motivation

SP is a major indicator that measures the success or failure of any educational system across the globe, as mentioned by (Narad & Abdullah, 2016). Kumar (2021) added that SP can be described as the nucleus around which a whole lot of weighty components of the educational system revolve. Therefore, it is crucial to model the student performance under all circumstances, whether the learning process was carried out conventionally or online.

COVID-19 has impacted the educational systems and made us more prepared, experienced, and equipped for distance learning (DL), which is a useful learning method and has become, and continues to be, a demanded method of learning and a preferred one by many universities around the world. DL is imposing itself more and more. However, we are still assessing and predicting the students' performance based on the characteristics and factors of the conventional mode of learning. Our ultimate goal in this thesis is to introduce a new definition and develop a new model to predict SP by considering more appropriate attributes and factors that better represent new concepts, behaviors, and activities subsequent to the distance learning mode.

1.2 Objectives

The main goal of this research is to build a data-driven model to predict student performance in DL time. In order to reach the desired goal, this research has set forth the below objectives:

1) Explore the definitions of SP mentioned in the literature review, select the definitions that are most pertinent to the online mode. Study the factors of performance and the attributes used to predict the SP. We need to find the attributes that have an impact on the SP in distance learning by doing the following:
    a. Examine the attributes that affect students' performance in conventional learning.

    b. Examine the attributes that affect the students' performance in the online mode.
    c. Compare the attributes that affect the performance in conventional attendance to the ones that contribute to the online learning and collect the common attributes.
    d. Establish a list of attributes that affect student performance in DL.
2) Develop a model that will predict the performance of students in distance learning, and to achieve that we are going to:
    a. Collect the data required for modeling students' performance.
    b. Select the most suitable ML technique to predict the performance of students during distance learning.

The outcome of this research will serve as a valuable tool for students and instructors alike, as it will allow instructors to evaluate their students, create tailored plans to better suit their students' needs. As well as improve the instructor's online teaching methodology and approach.

Chapter 2: Background

This research focuses on the student performance in distance learning, which will be modeled using different machine learning algorithms. In this chapter of the research, we introduce the definitions, concepts, and techniques that we will be using to explain our problem as well as to describe our solution.

2.1 Distance Learning

Since the time of the pandemic, there has been a growing interest in distance learning, that was for some time the only applicable method of learning. DL has been referred to sometimes as e-learning, online education, online learning, or remote learning. It is defined by (Phipps & Merisotis, 2022) as the online delivery of the course material by the instructor through different platforms whether it's live lecturing or recorded videos. Phipps & Merisotis (2022) emphasized two types of DL, one is called the synchronous where live lecturing takes place with instructor/students' interaction. The second type is the asynchronous, where students can watch recorded videos or read online material uploaded by the instructor.

2.2 Student Performance

SP was defined by Narad & Abdullah (2016) as the knowledge that the students gain, which is measured through assessments carried out by teachers. The authors looked at SP from an educational perspective, where it could be measured according to how much the student accomplishes of a certain goal during a specific period of time, where this goal could differ from institution or person to another.

2.3 Performance Modeling

SP modeling is done through building a model using a machine learning algorithm that will be able to predict the performance of students after being trained on part of the data available. The more the model predicts correctly, the higher the accuracy of the model will be, giving better results for educators and decision makers to act accordingly, and be able to improve students' performance.

2.4 Machine Learning Techniques

Machine Learning (ML) is an essential component in people's lives nowadays, where its applications are everywhere around us. ML makes use of different algorithms and models to help machines improve their performance over time. Without being programmed to make such decisions, ML algorithms are capable of constructing a mathematical model using sample data automatically, known as "training data" to make decisions, as mentioned by (Tejedor et al., 2020).

Machine learning can be classified into three types: supervised, unsupervised, and reinforcement learning. Reinforcement learning (RL) is a computational technique, according to (Tejedor et al., 2020). It is developed for challenges involving a learning agent interacting with its environment in order to attain a goal. ML is often used to map an input to an output based on the labeled dataset and is known as supervised learning. However, unsupervised is without the requirement for human intervention, it analyzes unlabeled datasets as identified by (Sarker, 2021).

2.4.1 Artificial Neural Networks

Aydoğdu (2020) explained ANN in his article and used it to model the SP, where he mentioned that it consists of 3 layers where each layer has nodes that are known as neurons. The first layer is the input layer that takes the input variables for the model, hidden layer(s), as it could be more than one layer, as the intermediate layer that is formed when the input values in the input layer are multiplied by their weights; and the last layer in this network is called the output layer, which gives the results after all calculations are completed. Figure 1 below demonstrates the layers of the ANN.

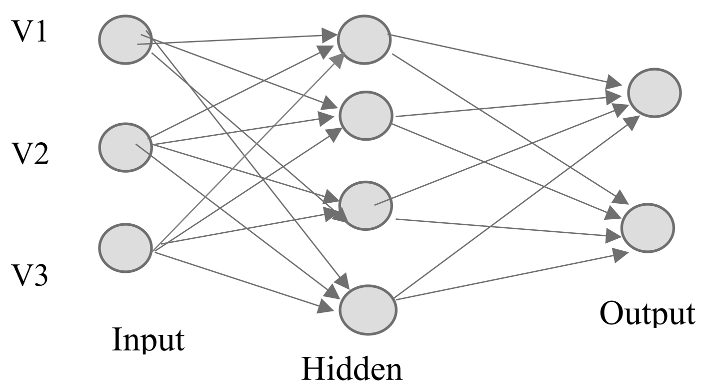

Figure 1: ANN layers

2.4.2 Random Forest

RF is a classification and regression ensemble machine learning approach that is widely used to predict SP. It works by fitting a variety of decision trees to numerous different smaller samples of the training data, and then it employs the mean to increase accuracy and limit overfitting as described by (Mubarak et al., 2022).

2.4.3 Logistic Regression

Another ML supervised technique is the Logistic Regression (LR), which was discussed by (Brahim, 2022) and is used for classification that is based on probability, it gives a binary output like 0 or 1, true or false. This technique uses a function called the Sigmoid function (mathematical function), which helps in predicting the values.

2.4.4 Support Vector Machine

This ML technique can be used with either classification or regression problems, as said by (Brahim, 2022), where he added that it actually showed good performance in linear and non-linear problems, and it works by classifying the data into classes based on the hyperplane it creates.

2.4.5 Decision Trees

Decision Tree (DT) techniques explained by (Hamoud et al., 2018). DT is a predictive modeling technique that can be used to forecast, classify, or categorize given data items. The tree has a root node, internal nodes, and leaf (terminal) nodes as part of its structure. The topmost node that has one or more outgoing edges, but no incoming nodes is known as the root node. Each internal node indicates a test on an attribute, and each edge shows the outcome of the test. An internal or middle node contains one incoming edge and one or more outgoing edges. The leaf node represents a data object's final proposed output, which is technically the predicted label.

Chapter 3: Relevant Literature

In this chapter we explore the different literature that has tried to define SP in all learning modes, as well as look closely at the models of SP that have been created in the conventional mode and the DL mode.

3.1 Definitions of SP

Throughout the reviewed literature, the definition of SP had different variations, and lacked consensus. SP definition is indeed coming from different perspectives or genres as mentioned by (Kumar, 2021), where he identified five genres of definitions starting from which various definitions of SP have been introduced. These genres are namely: knowledge centric, achievement centric, career centric, persistence centric, and skill and ability centric.

There are many reasons behind the diversity of SP definitions, such as the researchers' objectives as well as the context where the student performance is needed to be defined, assessed, or predicted. In other words, it depends on the author's perspective and the aim of the research. For instance, SP is defined by (Narad & Abdullah, 2016) as the knowledge gained by the student and assessed by a teacher via attribution of grades, where SP is derived from achievement, ability and skills genre. Narad & Abdullah (2016) also defined it as educational objectives set by students and teachers to be achieved within a specified timeframe. The latter definition is rather based on achievement, and ability and skills genre. Kumar (2021) defined the SP from knowledge genre as the outcome of two factors, which are the personal features (such as skills, motivation, etc.) and environmental features surrounding the student. While he also focused on the knowledge genre and defined SP as to what degree students are able to accomplish assigned tasks and pass related examinations. He also emphasized the career genre and defined it as the education gained by the students which offers them an opportunity to develop their abilities, find their ambitions and progress in their career to secure high satisfaction levels in jobs. The table below summarizes the definitions and adds the persistence genre definition for SP as mentioned by (Kumar, 2021).

Table 1: Summary of genre-based SP definitions

Genre	SP definition
Ability and Skills Achievement	The knowledge gained by the student and assessed by a teacher via attribution of grades and/or also defines as educational objectives set by students and teachers to be accomplished in a specific time.
Achievement	The outcome of two factors which are the personal features (such as skills, motivation, etc.) and environmental features confronting the student
Knowledge	The degree the students can achieve in the assigned tasks and related examinations
Career	The education gained by students offers them an opportunity to develop their abilities, find their ambitions and progress in career to secure high satisfaction levels regarding their career
Persistence	The academic progression toward degree completion, regardless of institution-related settings and problems.

3.2 Attributes and Modeling of SP

This section includes the reviewed articles that have modeled the performance of students in the conventional mode as well as during the distance learning mode. The differences in modeling between both modes of learning is discussed and analyzed in terms of appropriateness of the attributes of SP and the ML algorithms used to define and predict the students' performance in both modes.

3.2.1 Modeling SP in Conventional Mode

Regarding students' academic performance in conventional mode, (Dabhade et al., 2021) aimed to predict the performance for the next semester of senior year students in some courses. The attributes the authors used are from different categories, such as personal, behavioral, extra-curricular activities, and previous education. The details of the attributes considered are: Personal demographic details include hobbies, interests, household income, and time spent on social media and watching movies. Educational features consisting of past semesters' GPA. Behavioral features, which include the participation of students in academic activities, the effect of friends' circles, and interaction with faculties. And lastly, the extra-curricular features, which are the participation of students in extracurricular activities, namely sports or club activities in college. The dataset contains 85 records for 85 participants in the survey and 16 attributes used for prediction.

The authors have used several versions of the Support Vector Regression (SVR) with different types of kernels. They attempted to predict the performance for the next semester of senior year students in some courses. Their results showed that one of their SVR achieved a good accuracy score of 83.4% and added that the most dominant factor to predict the performance was the previous GPA.

Hamoud et al. (2018) predicted student performance in HEI using different decision trees. The attributes the authors used for prediction are GPA, credits (how many credit hours did the students finish), list of important notes (if the students write notes), father's occupation, and fresh food consumption. The data is obtained from surveying 151 students, and the number of attributes used in prediction is 6. In this research, three decision trees have been used to model the performance: J48, Random Tree and Rep Tree. The model derived from J48 was the one that had the highest precision and recall results, with scores of 62.9% and 63.4% for precision and recall, respectively.

Altaf et al. (2019) predicted SP by introducing a multi-layer neural network. The attributes used for prediction are provided by the campus management system. (CMS), that included: course ID, total learning sessions (number of sessions in the semester), total length of session (the length of sessions in minutes), average of all session length, total assessments in one semester, mean assessment grade, number of quizzes made, total number of emails sent, number of forum posts, and final course grade. The dataset has records of 900 students for 10 courses over the year of 2017-2017, the number of attributes considered is 9. The results showed that neural networks have promising results in predicting student performance for different courses, with accuracy scores ranging between 75%-97.1%.

Another similar study was carried out by (Ramaswami et al., 2022), where various ML algorithms have been used to predict SP, namely, catBoost, naïve bais, random forest, KNN, and logistic regression. The attributes of performance the authors used from the learning management system (LMS) include average score of prior courses, maximum score achieved in prior courses, prior course deviation score, assignment score, assignment deviation score, prior role description, LMS deviation score, and LMS engagement score. Adding to them are some attributes from the university system, such as citizenship, age, highest school qualification, gender, and English proficiency test. The dataset has 14 attributes considered for prediction. Their results showed that the highest accuracy is achieved by the CatBoost algorithm with an accuracy score of 75% and identified the prior course grades as the most significant factor in determining the performance of students.

3.2.2 Modeling SP in Distance Learning Mode

Wang et al. (2018) predicted student behaviors and performance in online learning using different versions of Decision Trees. The attributes used are based on the categories, namely basic features, behavioral and performance characteristics. Basic attributes specifically included gender, subject background, and academic qualification, while behavioral attributes included learning time span, the total learning duration, learning frequency, the average length the student remains in a study session, participating in the discussion board available on their system, taking notes on their online system, and feedback time (how many times the student received messages from his instructor and encouraging him to learn). Performance attributes included homework, practical skills, online tests, and the final written exam. The data used for modeling was collected from 2801 participating students in Shaanxi University in China using an online survey, which implies that the dataset contains 2801 records and 12 attributes used for prediction. The authors' results showed that decision trees can be very effective in helping educators adjust intervention plans for students. The decision trees created had accuracy scores ranging between 70% and 95%.

Hong et al. (2017) predicted the SP using time-related attributes. The authors proposed three novel attributes from their online system, namely video-watching frequency (the time spent watching the course videos in hours), video-watching interval (the time taken between watching one lecture and another) and learning efficiency (the time the user spent watching all videos of the course). The dataset the authors used is made up of 834 records. In this study, SVM, RF, and DT were used for predicting students' performance. The model that was selected with the best performance, according to the authors, was the DT based on its performance metric, area under the curve (AUC), which determines the overall performance of the model.

Al Karim et al. (2022) aimed to evaluate the Iterative Dichotomizer 3 (ID3) technique to predict the SP in online mode. The dataset attributes they have used for prediction are: mid-term attendance, mid-term absence, midterm quiz 1, midterm quiz 2, mid-term grade, final term attendance, final term absence, final term quiz 1, final term quiz 2, and final term grade. The ID3 model accuracy score achieved was 77.86%.

In another work, (Al Karim et al., 2021) proposed decision trees (J48) algorithms to predict SP in distance learning time. The authors built the model and tested it on four datasets; the attributes in the datasets are similar. The common attributes between all datasets are gender, absence, attendance, final term grade, and midterm grade. In the second dataset, the quiz 1 and quiz 2 grades were added. While in dataset three, different attributes were added: mid-term attendance, mid-term absence, midterm quiz 1, midterm quiz 2, mid-term grade, final term attendance, final term absence, final term quiz 1, final term quiz 2, and final term grade. In the last dataset, attributes they added were the CGPA, quiz 1, quiz 2, lab-performance, mid-term, and final-term marks. The authors used 4 datasets; the first one had a total of 6 attributes and 589 records, while the second dataset had 8 attributes and 330 records. Dataset 3 had12 attributes and 280 records, and dataset 4 had 10 attributes and 91 records only. Knowledge extracted from the decision trees shows that 'final term' and 'mid-term' are the most prominent attributes to analyze and predict students' performance, whereas other attributes i.e., quiz, gender, and attendance, had less impact.

Karalar et al. (2021) aimed to predict the performance of students at risk of failure. The authors used different ML algorithms like Artificial Neural Network, Decision Trees, Logistic Regression, Random Forest, and Extra Trees. Furthermore, they have used data collected from Moodle (the university LMS) and another software called Adobe Connect for a whole semester. In their dataset, the GPA was the column predicted by the ML algorithms as either passing or failing. Attributes used for prediction were gender, degree (2 or 4 years of study), number of lecture notes downloads, number of material downloads for the course, time spent watching the recorded session of the live lectures, the total time spent attending live lectures, and the score of the quizzes conducted. The authors used a dataset made up of 2045 records and 5 attributes to predict the performance. According to their recorded results, the most successful ensemble model to predict the performance was derived from LR, ET, and RF, with the highest specificity score of 90%.

Another similar study was conducted by (Aydoğdu, 2020), where he predicted the SP using Artificial Neural Network. The author used data from the LMS that is also synchronized with Adobe Connect. The attributes he contributed with and found that have impact on the SP are gender, content score (percentage of how many subjects the student has completed from his major), time spent on content, number of entries to reach content, homework score, number of live sessions attended, total time spent in live sessions, number of archived courses attended, and total time spent in archived courses. The total number of attributes used is 9, and the dataset is made up of 3518 records. The model output is called success status, which shows if the student will succeed or not in the course by the end of the semester. The accuracy of the author's ANN model was 80.47%. A summary of the methodologies used in the reviewed articles is listed in Table 2.

Table 2: Summary of articles and their attributes used to model SP

Distance learning mode		**Conventional learning mode**	
Predicted SP using decision trees based on the attributes from teams and university server includes gender, absenteeism students' grades midterm, quiz and final	(Al Karim et al., 2022)	Predicted SP in physical mode from the data provided from the LMS	(Altaf et al., 2019)
Predicted student behaviors and performance in online learning based on the categories of attributes: Basic features, Behavioral characteristics, and performance characteristics	(Wang et al., 2018)	Data collected from several sources: Moodle log, Student management system and enrollment system of the university	(Ramaswami et al., 2022)
The attributes used for modeling are gender, content score, time spent on content, number of entries to reach content, homework score, number of attendances to live sessions, total time spent in live sessions, number of attendances to archived courses, total time spent in archived courses	(Aydoğdu, 2020)	Students' performance comes from different aspects like personal, behavioral, extra-curricular activities and previous education	(Dabhade et al., 2021)
Author used the attributes: degree, number of lecture notes downloads, number of material downloads of the course, the time spent by watching the recorded session of the live lectures, the total time in minutes of attending live lectures and the score of the quizzes conducted	(Karalar et al., 2021)	Predicted student performance in HEI using different decision trees. The attributes the authors used for prediction are GPA, credits, list important notes), father's occupation, and fresh food consumption.	(Hamoud et al., 2018)
Predicted the SP using Time-related attributes, such as video-watching frequency, video-watching interval and learning Efficiency	(Hong et al., 2017)		

3.3 Research Gap

Looking at the definitions of SP that have been introduced in Section 3.1, there is a lack of agreement on how to define the performance in conventional learning mode. While there was no definition of SP found that fits the settings of the DL mode, in this thesis we are going to propose the SP definition that is suitable for the DL settings.

According to the modeling of performance, we found that there is no solid foundation or basic standards followed, neither in conventional nor in DL mode. There have been a lot of trials that are diversified, with no agreement on what attributes are used to model the performance. This issue is going to be addressed in this research, and we are going to contribute to validating the best model to predict the SP in DL.

3.4 Research Questions

To fill the research gap, this study will address two questions that will be answered in order to predict the performance of students in DL mode, which will help us succeed in achieving the dominant goal of the research. The questions are listed below:

1) How can the SP be defined in the distance learning context? In particular, what are the attributes that are effectively impacting the student performance in distance learning?
2) What would be the most suitable ML technique to model student performance in distance learning mode?

Chapter 4: Proposed Model for Student Performance in DL Mode

In the sections of this chapter, we are going to propose an appropriate definition of SP that suits the DL mode, introduce the selected attributes, and describe the process of modeling the SP in the DL mode.

4.1 Proposed Definition of SP in DL Mode

As stated in Chapter 3, SP had different definitions depending on the perspective of the author based on the findings and definitions explored in the literature. Considering the fact that no explicit definition of SP in DL mode has been proposed in the most relevant reviewed articles, our definition will be inspired from the SP definitions while considering the attributes that were identified as relevant to the students' performance in DL time.

In this research, our objective is to academically define the student performance during distance learning based on their achievements in the course, the knowledge they gained, and the skills obtained since that will generalize the definition of performance. Therefore, our definition is inspired by the one proposed earlier by (Narad & Abdullah, 2016) earlier (which was derived from the achievements, and skills and ability genres of the definition). They defined it as the knowledge gained by the student and assessed by a teacher via the attribution of grades. Our proposed definition of SP in distance learning time is the student's knowledge gained from effective attendance of the sessions, which is reflected on students' grades on assessments throughout the course. Our contribution to the definition adds the value of the effective attendance in DL. Effective attendance can be identified online if the student is continuously attending the live session and is being active during the lecture, or depending on whether they are watching the recorded sessions.

4.2 Selected Features to Predict SP in DL Mode

From the examined literature, we were able to identify the attributes affecting the SP and classify them into three categories, (1) the attributes that affect the SP in the conventional mode, (2) the attributes affecting the SP in the distance learning mode; and (3) the attributes in between, that are affecting the SP in both modes.

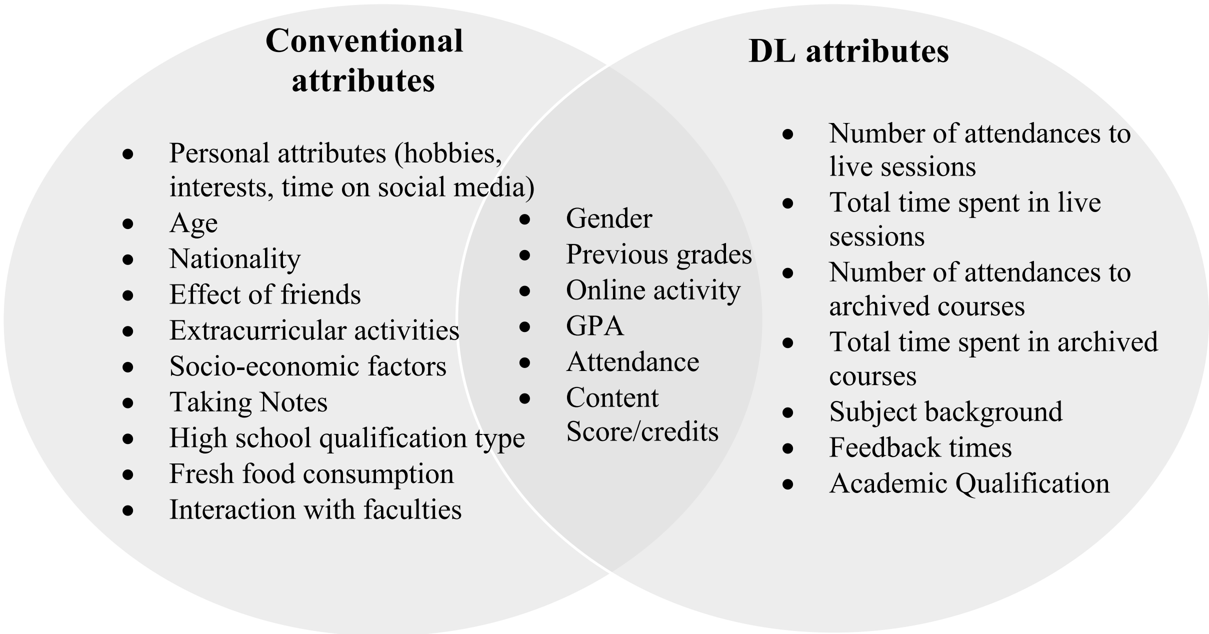

Figure 2: Categorization of identified attributes affecting SP in conventional and DL

From the above figure, we can now identify the most appropriate attributes that affect the SP in DL time. The attributes in common between the two modes, as displayed in Figure 2, will be used to predict the performance along with the new DL attributes. The attributes are further discussed in Table 3 and will be used in modeling the SP in Chapter 5.

Table 3: Attributes affecting student performance in DL

Attributes affecting performance	Paper	Definition
Previously obtained grades	(Hamoud et al., 2018), (Altaf et al., 2019), (Pallathadka et al., 2022), (Aydoğdu, 2020), (Karalar et al., 2021)	course assignments and quizzes or any other assessments made while taking the course
GPA	(Hamoud et al., 2018) (Adnan et al., 2021) (Kumar, 2021)	Students Grade Point Average
Online activity attributes	(Altaf et al., 2019) (Karalar et al., 2021) (Aydoğdu, 2020)	student interaction with LMS (attributes may vary depending on the variables taken from the LMS example, downloading material, lecture notes, number of times joined the session, number of logins to the LMS, and total email sent etc.
Gender	(Wang et al., 2018), (Misopoulos et al., 2018), (Matzavela & Alepis, 2021), (Karalar et al., 2021)	
Academic Qualification	(Wang et al., 2018)	The highest obtained qualification
Live sessions attendance	(Karalar et al., 2021), (Aydoğdu, 2020)	The attendance of the students to the live sessions
Watching recorded sessions	(Karalar et al., 2021), (Aydoğdu, 2020)	Time spent on watching the recorded or archived sessions
Time in live sessions	(Karalar et al., 2021), (Aydoğdu, 2020)	The total time spent in the live session
Content score / credits	(Aydoğdu, 2020), (Hamoud et al., 2018)	Percentage of how many courses the student has completed from his major
Feedback time	(Wang et al., 2018)	The number of messages the students receives from his instructor and how much is he encouraged in a semester

Attributes that are identified as having effect on SP are: previous grades, GPA, online activity attributes, gender, live sessions attendance, watching recorded sessions, time spent in live sessions, content score, and feedback times.

In the reviewed literature, we were able to identify the new predictors of the performance of students in the DL time. These attributes have been selected based on their relevance to the nature and practices of DL, such as the obstacles to effective online attendance, live sessions, connection issues, etc. Therefore, the adopted attributes to predict SP in the DL time in our model are content score, number of attendances to live sessions, total time spent in live sessions, total time spent in archived courses, online activity, which includes logging in to the system, lecture notes download, and material download. As well as the attributes that are common between the conventional and DL modes, which are gender, academic qualification, previously obtained grades during the term that may be quizzes, assignments, or midterms, the student's GPA, and the feedback times received from the instructor.

4.3 The Modeling Process

The modeling process starts with the data collection and preparation, followed by the feature selection to get the most important features of the dataset. Thereafter, we selected models for predicting the SP and evaluated the performance of the models. Figure 3 demonstrates the modeling process.

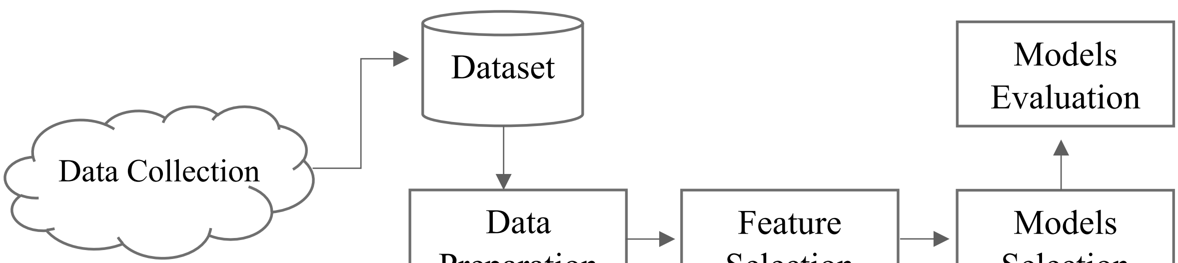

Figure 3: Modeling process

4.3.1 Data Collection and Preparation

The first step in the modeling process is collecting data. This is an essential step that requires a good amount of quality data, thus leading to better results when the prediction takes place. Data can be collected in several ways, from surveying people, extracting it from a platform, or even extracting data from photos, videos, or documents.

The process of making raw data ready for further processing and analysis is known as data preparation. The gathering, preparation, and labeling of raw data into a format appropriate for machine learning (ML), followed by data exploration and visualization, are crucial phases in the modeling process.

4.3.2 Feature Selection

Feature selection involves finding the most suitable features of the dataset that may help boost the performance of the model. There are several techniques that can be used, for instance, the Chi-square test, ANOVA (Analysis of Variance), backward feature elimination, and recursive feature elimination.

4.3.3 Prediction Model Selection

There have been plenty of machine learning algorithms used to predict the performance of students in both settings, like artificial neural networks (ANN), decision trees (DT), support vector machine (SVM), naïve bias (NB), logistic regressions (LR), clustering algorithms: K-means, K-nearest neighbor (KNN) and much more. However, the most commonly used techniques to predict the performance. according to the reviewed literature, are artificial neural networks, decision trees, support vector machine, logistic regression, and random forest. The later models will be used in our research to predict the SP in DL time.

4.3.4 Evaluation Metrics

To ensure the reliability of a model, different measures could be selected, such as accuracy, sensitivity, F1-score, precision, and specificity. The metrics explained and used in our research are listed by (Seo et al., 2021) in their article.

Starting with accuracy, which determines the percentage of correct predictions made by the model. Accuracy is calculated as follows, where True Positive is (TP), True Negative is (TN), False Positive is (FP) and False Negative is (FN).

(1) Accuracy = (TP + TN) / (TP + TN + FP + FN)

A second measure is the sensitivity, which shows how many correct predictions were made successfully from the positive class.

(2) Sensitivity = TP / (TP + FN)

While the precision shows how many positive predictions were made correctly from the whole dataset.

(3) Precision = TP / (TP + FP)

F1-score, that balances the precision and recall of a model. It is an effective measure of performance of models as it shows the harmony in the built model. It can be calculated according to the below formula.

(4) F1-score = 2 * precision * sensitivity / (precision + sensitivity)

Specificity score can compute the percentage of the correct predictions of the negative class. It can be computed using the below formula.

(5) Specificity = TN / (TN + FP)

It is important to validate the model as well, and for this purpose cross-validation can be applied. In this research, to ensure the use of the best performing model, we are going to utilize the use of accuracy, sensitivity, precision, F1-score, and specificity to measure the performance of the model.

Chapter 5: Experimental Evaluation

In this chapter, we discuss the implementation of our approach as well as its validation on a real dataset. The performance of our derived model is compared to a set of benchmarks using a rigorous validation method.

5.1 Used Dataset

The dataset used in this research was collected in a Turkish State University from two different platforms, namely, the university LMS used (Moodle), and the Adobe Connect platform (conference management system used in the university). The data covered the online learning activities spanning over 15 weeks in the Fall semester of 2020 for one of the courses called Information Technologies. The dataset contains 3223 records and 9 attributes representing different aspects and characteristics of the DL activities. Table 4 describes the attributes of the dataset that will be used in our model to predict the SP in DL time.

Table 4: Explains the attributes of the dataset

Attribute name	**Type and definition**	**Data range**
Gender	Categorical. Female/Male	0 or 1
Degree	Categorical. Specifies if collage or faculty students (2 or 4 years of study)	0 or 1
Lecture_notes	Numeric attribute. Counts how many times the learner has downloaded the lecture notes.	0 to 203
Materials	Numeric attribute. Counts how many times the learner has downloaded the material of the course	0 to 176
Video	Numeric attribute. Counts the total time in minutes the learner has spent in viewing the recorded sessions	1 to 5988
Live_Ratio	Numeric attribute. Average stay of the student in a single live session	1 to 120
Live_attendance	Numeric attribute. How many times the student has logged in to the sessions	1 to 46
Quiz	Numeric attribute. The quiz mark conducted through the term	1 to 100
Live	Numeric attribute. The total time in minutes spent in the live sessions for the learner for all lectures	1-5988

As seen in Table 4, the attributes that are used in this study to predict the performance of students in DL mode are gender, degree, lecture notes, materials, video, live attendance, quiz, and live. Some attributes that were listed in Chapter 4 that affect SP in DL, such as content score, academic qualification, and feedback times, were not listed in the dataset used and will not be used in our model as we were unable to identify values for those attributes.

The students in the dataset are classified as pass or fail. The class label "Result1" is the target attribute that has a value of either 1 or 0, for pass or fail, respectively. This dependent variable, which represents the class of students, was calculated according to the student's performance in the final and midterm exams grades. It is explicitly a weighted sum of the final (60%) and the midterm (40%) grades.

The passing grade in this dataset is different based on the degree of study; if the student is studying for a 2-year diploma, the passing grade is 40 and above, otherwise, the student's passing grade is 50 and above for a 4-year degree.

5.2 Dataset Cleansing

A pre-processing step has been applied to the dataset to get rid of the records that show abnormal behavior by students during the online learning sessions. Therefore, a number of records have been removed from the original dataset of 3223 records. The final number of records being used in the prediction is 2045 records out of 3223. Records that have a very low number of logins and a high Live_ratio have been removed from the dataset, which means that the students have logged in and have kept the session open without logging out. Threshold values were set as follows: if the number of joining (Live_attendnace<=2) and the average time spent per joining is more than 120 minutes (Live_ratio >= 120 minutes). Even students with very high number of Live_attendance but very low Live_ratio, have also been removed from the dataset. The criteria was set that if the students logged in to the sessions more than 10 times (Live_attendance >=10) and the average stay in a session is less than 20 minutes (live_ratio<20) it means that the students were just logging in and out and had technical and/or connection issues.

5.3 Feature Selection

In this section, we are going to introduce two feature selection techniques that have been used for our dataset.

Feature selection has also been applied in order to focus on the best predictors of SP among the 9 attributes available in the dataset. According to the reviewed literature, the ANOVA F-value was the most adopted technique for classification problems. Therefore, the feature selection algorithm used in our model building process is the ANOVA. It chooses the attributes that are most relevant and most likely to have an impact on the target variable. The process calculates an ANOVA F-value for each feature, which assesses the variation in the feature's mean values between classes or groups. ANOVA feature selection ranks the features according to their F-values and selects the features that appear on the highest ranking. It is available in the Sklearn library in Python. In particular, we used the SelectKBest with the f_classif (ANOVA) scoring function to select the top features for our model, as depicted in Figure 4.

```
bestfeatures = SelectKBest(score_func=f_classif, k=9)
```

Figure 4: Feature selection code of ANOVA

According to the feature selection technique, we acquired the results displayed in the graph of Figure 5. The graph shows that the quiz result is the most significant feature in the dataset, followed by the live attribute, that computes the total time the student spent in the live session in minutes. In third place comes the degree, whereas live ratio and lecture notes come in fourth and fifth place, in that order. Therefore, we can conclude that the least important features are gender, live_attendance and video, which are going to be disregarded in predicting the SP.

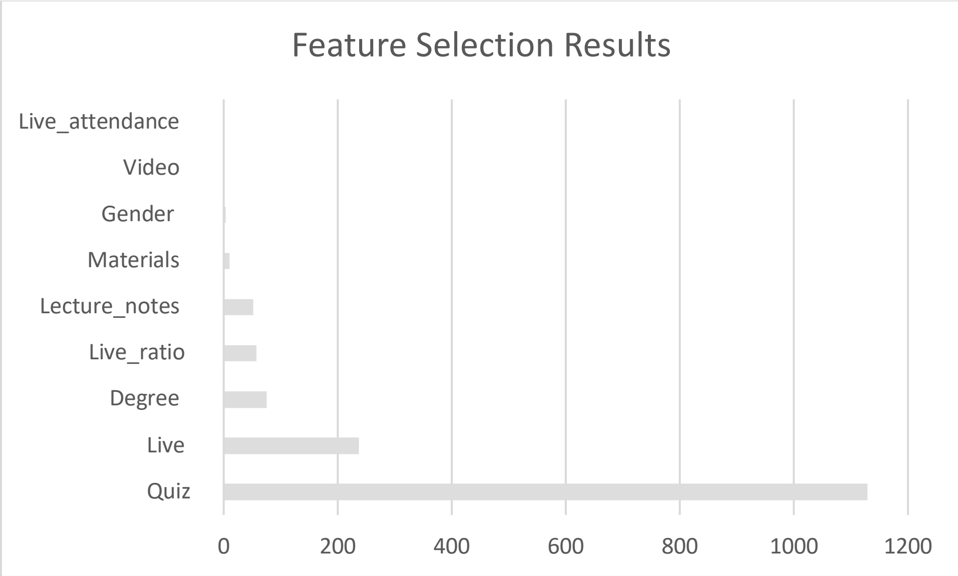

Figure 5: Feature Selection Scores

The second most commonly used feature selection technique found in the literature review is the Chi2. The strength of the link between the input features and the target variable in a dataset is assessed using the chi-square test. We have used this method as well, and Figure 6 displays the code used.

```
#apply SelectKBest class to extract top 10 best features
bestfeatures = SelectKBest(score_func=chi2, k=9)
```

Figure 6: Feature selection using Chi2 square code

According to the feature selection technique, we got the results displayed in the graph of Figure 7. The results show that the live feature had the highest score, followed by the quiz feature, third place came the video watching feature. And from there, we can tell that the least important features are the gender and live_attendance, which are going to be disregarded in predicting the SP.

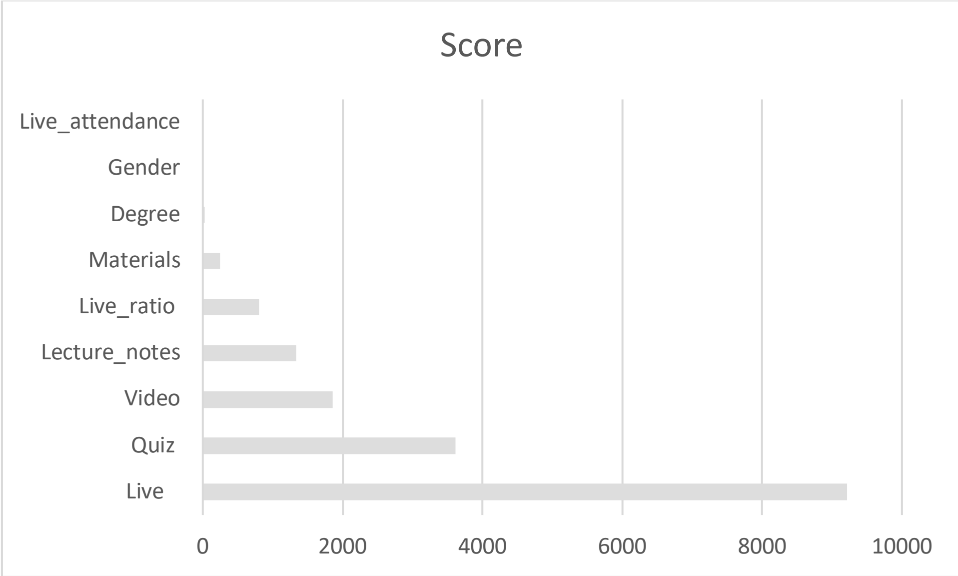

Figure 7: Feature selection results of the Chi2 square function

However, this technique of feature selection, does not fit with our dataset, since this function works best with categorical input variables, while our dataset is mixture of both, and the numeric variables are more that the categorical variables. Therefore, the adopted technique of feature selection used in this research is the ANOVA, which has been discussed earlier in this section.

5.4 Experimental Design

In this section, we are going to describe the details of our experiment. In particular, we introduce the models that we built using different ML techniques, in order to select our final model configuration. In this first part of our experiment, we target the internal validation of our proposed model. We also describe the comparison of the derived model with a set of benchmark solutions. In other terms in this second part, we are targeting the external validation of our model.

5.4.1 Model Selection and Validation

The ML techniques used in this study to derive the data-driven model in DL time are: SVM, RF, DT, LR, and ANN. These techniques are the most common ones used in predicting SP, as seen in the reviewed literature. Models derived by using these ML techniques on our suggested dataset, are explicitly discussed based on their respective performances and validated in the coming section in order to select the best model to predict the SP.

5.4.2 Benchmark Models

Benchmark1: Karalar et al. model: The first model that will be used as our benchmark model is the one built by (Karalar et al., 2021), where the authors have used the same dataset used in our research. The chosen algorithms were: logistic regression, random forest, extra trees, artificial neural networks, decision trees, quadratic discriminant analysis, and gradient boosting model. However, their goal was to create an ensemble model to detect students at risk. Karalar et al.'s final model was an ensemble model, achieving the highest performance when combined together three ML techniques, random forest, logistic regression, and extra trees.

Benchmark 2: Aydoğdu model: Our second benchmark model is the one built by (Aydoğdu, 2020), where the author used ANN to predict the SP in DL time using a different dataset. The details of this model were discussed in the Literature review (See Chapter 3). This model has used gender, content score (percentage of the subjects completed by the student from major), time spent on content, number of entries to reach content, homework score, number of attendances to live sessions, total time spent in live sessions, number of attendances to archived courses, and total time spent in archived courses, as attributes. The author's main goal was to use ANN to predict the performance of students in DL.

5.5 Model Selection and Validation

In this section we discuss the developed models, each one separately, discuss its performance, and choose the best performing model.

5.5.1 Support Vector Machine SVM

Support vector machine model has been setup with two parameters, where the following values, as shown in Figure 4 have been set: Starting with the regularization parameter C. The C parameter regulates how well the model generalizes to new data while still being able to fit the training set of data. A lower value of C results in a broader margin, and more misclassifications are permitted, whereas a higher value of C results in a tighter margin for misclassifications. The C parameter was set to 100, and lastly, the gamma parameter. The influence of each training sample on the decision boundary is defined by the gamma parameter. A smaller value of gamma indicates a broader influence of each training example and a smoother decision boundary, whereas a bigger value of gamma indicates a more localized influence of each training example and a more complex and overfitting-prone decision boundary. The gamma parameter is equal to 1, as seen in Figure 8.

```
#Support Vector Classifier
svc = SVC(C=100,gamma=1)
svc.fit(X_train,y_train)
svcpred = svc.predict(X_test)
```

Figure 8: SVM settings

The model has achieved an accuracy score of 79% before performing the feature selection, a precision score of 82%, a sensitivity score of 87% and F1-score of 84%. After performing the feature selection on the data, the model's performance has improved to achieve an accuracy score of 80%, precision score of 84%, sensitivity score of 87% and F1-score 85%.

After the feature selection for the SVM, Cross Validation (CV) has been applied to the model as well, with 10 folds, as displayed in Figure 9, the average score on the accuracy of the model recorded by the CV was 78%.

```
#cross Validartion for Support Vector Machine


k_folds = KFold(n_splits = 10)

scores = cross_val_score(svc, X, y, cv = k_folds)

print("Cross Validation Scores: ", scores)
print("Average CV Score: ", scores.mean())
print("Number of CV Scores used in Average: ", len(scores))
```

```
Cross Validation Scores:  [0.75784753 0.81165919 0.83856502 0.80717489 0.8161435  0.80269058
 0.78378378 0.73423423 0.77927928 0.68918919]
Average CV Score:  0.7820567203975276
Number of CV Scores used in Average:  10
```

Figure 9: Cross validation for SVM model

5.5.2 Decision Tree

The DT setting used in our research, consists of specifying three parameters used in modeling, max depth, min sample splits, and min sample leaf. The decision tree's maximum depth is determined by the max depth parameter. Although a deeper level of analysis depth can reveal more complicated connections in the data, it can also result in overfitting. The minimum number of samples needed to separate an internal node is specified by the min samples split option. By doing so, the model's generalization abilities will be enhanced, and overfitting will be prevented. The min samples leaf parameter determines the minimum necessary number of samples that must be present at a leaf node. By doing so, the model's generalization abilities will be enhanced, and overfitting will be prevented. The max depth parameter was set to 7, while the min sample splits was assigned the value of 4 and the min sample leaf was set to 1. Figure 10 illustrates the settings used.

```
#decision tree
dt = DecisionTreeClassifier(max_depth=7, min_samples_split=4, min_samples_leaf=1)
dt.fit(X_train,y_train)
dtpred = dt.predict(X_test)
```

Figure 10: DT settings

The model has achieved a score of 87% in terms of accuracy, a score of 91% for precision, 88% for sensitivity, 90% is the F1-score, and specificity score of 82%. After the feature selection has been applied, the model's performance has shown no changes except for the specificity score that went up by 1% to become 83%.

A 10-fold Cross validation has been applied to validate the model after feature selection, as displayed in Figure 11. It shows that the model had an average accuracy score of 85%.

```
k_folds = KFold(n_splits = 10)
scores = cross_val_score(dt, X, y, cv = k_folds)

print("Cross Validation Scores: ", scores)
print("Average CV Score: ", scores.mean())
print("Number of CV Scores used in Average: ", len(scores))
```

```
Cross Validation Scores:  [0.83408072 0.86995516 0.82959641 0.89237668 0.86547085 0.85201794
 0.82882883 0.84684685 0.86036036 0.85585586]
Average CV Score:  0.8535389649739425
Number of CV Scores used in Average:  10
```

Figure 11: Cross validation for DT model

5.5.3 Logistic Regression

The LR model setting has three parameters to be tuned. The parameter C in logistic regression defines the inverse of the regularization strength. Stronger regularization is produced by a lower value of C, which leads to a more straightforward model with fewer coefficients. A bigger value of C, on the other hand, leads to a weaker regularization, which indicates that the model will have larger coefficients and might be more complex. The second parameter is the max-iter. It determines the number of iterations necessary to converge to the optimal solution. This parameter accepts values up to 100. It was assigned the value of 30 after some trials of other values such as 10 and 20. The last parameter, the random_state was assigned a value of 1, which guarantees that each time you run the model, you will get the same results. Every time you run the model with the same random state value, the same sequence will be produced. The configuration of the LR model is illustrated in Figure 12.

```
#Logistic Regression
Lr = LogisticRegression(C=1,max_iter=30,random_state=1)
Lr.fit(X_train,y_train)
Lrpred = Lr.predict(X_test)
```

Figure 12: LR settings

LR had shown an accuracy score of 83%, which has slightly improved to 84%, after the feature selection step. A precision score of 84%, which became 87%, a sensitivity score of 88%, which became 89%. F1 score and specificity values achieved 86% and 74%, respectively, and after FS the LR achieved 88% F1-score, and no change has been noticed for the specificity score. The 10-fold CV took place, as depicted in Figure 13, to achieve an average accuracy score of 84%.

```python
k_folds = KFold(n_splits = 10)

scores = cross_val_score(Lr, X, y, cv = k_folds)

print("Cross Validation Scores: ", scores)
print("Average CV Score: ", scores.mean())
print("Number of CV Scores used in Average: ", len(scores))
```

```
Cross Validation Scores:  [0.84304933 0.83408072 0.83408072 0.8206278  0.83408072 0.86547085
 0.84684685 0.84234234 0.84684685 0.81531532]
Average CV Score:  0.8382741485880499
Number of CV Scores used in Average:  10
```

Figure 13: LR cross validation

5.5.4 Random Forest

RF has eight parameters that has been set. To begin with, n_estimators' parameter, that identifies the number of trees the algorithm has. Although, it was tested on different values, 60, 70 ,80 ,90, 100, and 110, the best performance of the model was achieved when the parameter was assigned a value of 100. The max_features parameter, the maximum number of features that can be used to split a node in a decision tree is specified by this parameter. For classification purposes, when it's set to "*auto*", the method by default employs the square root of the number of features. The max_depth parameter sets each decision tree's maximum depth. It was assigned the value of 9 after trying several values, including 3,5 and 7. The minimum number of samples that must be present at a leaf node is specified by min_samples_leaf parameter. The parameter value was set to 1 after several trials of different values, such as 5 and 10, which showed slight degradation in the model 's performance. The minimal number of samples needed to split an internal node in RF is controlled by the min_samples_split parameter. It has been tested on 2, 5, and 10, and the best performance of the model was found at 2. The random_state parameter helps in controlling the RF seed value for the random number generator. It can guarantee that the same sequence of random numbers is generated every time the model is executed by setting this option to a fixed value. This can help with results uniformity and reproducibility; the parameter was set to 42 in our model. Figure 14 demonstrates the setting of the RF model.

```
rfc = RandomForestClassifier(n_estimators=100,max_features='auto', max_depth=9,min_samples_leaf=1,
                             min_samples_split=2, random_state =42)
rfc.fit(X_train,y_train)
rfcpred = rfc.predict(X_test)
```

Figure 14: RF settings

The accuracy score of the model reached 89%, the precision score was 92%, while the sensitivity score was 89%. The model also scored 91% for the F1-score and 85% for the specificity The model performance was stable and showed a slight improvement after feature selection, which was noticed in the specificity score by an increase of 1%. After applying the 10-fold CV for the RF model shown in Figure 15, its accuracy achieved an average score of 87%.

```
k_folds = KFold(n_splits = 10)

scores = cross_val_score(rfc, X, y, cv = k_folds)

print("Cross Validation Scores: ", scores)
print("Average CV Score: ", scores.mean())
print("Number of CV Scores used in Average: ", len(scores))
```

```
Cross Validation Scores:  [0.9103139  0.86995516 0.88789238 0.87892377 0.87443946 0.86547085
 0.86036036 0.84684685 0.86936937 0.85585586]
Average CV Score:  0.87194279481274 99
Number of CV Scores used in Average:  10
```

Figure 15: Cross validation for RF

5.5.5 Artificial Neural Networks

Our ANN network has been setup to have three layers, one input layer, one hidden layer, and the output layer. Each layer has been set to have different dense layer units; the dense layer is where all neurons from that layer are fully connected to the next layer. The dense units for the input layer were set to the number of inputs in our dataset, which is 6. Whereas the dense units for the hidden layer have been set to 12 after many trials of different numbers such as 8, 9, 10 and 12, and the output layer dense was set to the value of 1. The activation function applied for both, input and hidden layers was the *ReLU*. If the input value is positive, the *ReLU* function returns that value; if not, it returns 0. This indicates that the *ReLU* function's output is consistently non-negative and that the function is simple to compute and differentiate. The output layer activation function was the

Sigmoid function, as it fits well in binary classification problems. The optimizer used for the ANN is the *Adam Optimizer*. While the loss parameter was assigned the value of binary_cross-entropy. The binary cross-entropy loss function generates a value that represents how well the model is doing at predicting the right class by calculating the difference between the predicted probability distribution and the true probability distribution. The metrics parameter is equal to accuracy, as shown in Figure 16.

```
import tensorflow as tf
ann = tf.keras.models.Sequential()
ann.add(tf.keras.layers.Dense(units =6, activation = 'relu'))
ann.add(tf.keras.layers.Dense(units =12, activation = 'relu'))

ann.add(tf.keras.layers.Dense(units = 1, activation = 'sigmoid'))
ann.compile(optimizer = 'adam', loss = 'binary_crossentropy', metrics = ['accuracy'])
ann.fit(X_train, y_train, batch_size = 32, epochs = 100)
```

Figure 16: ANN settings

The ANN model was fit with an epoch value of 100 and a batch size of 32, where different values of epoch and batch size were empirically tested until the highest accuracy score was achieved. For the batch_size, it was tested on 32, 64, 128 and 256, while the epoch was tested on 50 and 100. The achieved accuracy of the ANN model was 85%, its precision score was 87%, and its sensitivity score was 89%. The accuracy score of the model has also improved after the feature selection and reached 86%, the precision score was 88%; and sensitivity score was 90%. After a 10-fold CV the accuracy score of the ANN model has achieved 85%.

5.6 Results and Discussion of Models

When comparing the models based on their performances, the first conclusion is that all the models have achieved very good scores. This first result is obtained thanks to the successful adoption of the new definition of SP.

In order to select the best modeling technique to build an appropriate model of SP in DL time, we discuss the differences among the scores achieved by the different models. Table 5 depicts side by side all the scores achieved by different ML techniques on the same dataset before feature selection; particularly, the accuracy, precision, sensitivity, F1-score, and specificity are reported.

Table 5: Models performance before feature selection

	Accuracy%	Precision%	Sensitivity%	F1-score%	Specificity%
SVM	79	82	87	84	62
DT	87	91	88	90	82
LR	83	84	88	86	74
RF	89	92	89	90	85
ANN	85	87	89	88	78

As shown in Table 5, we can see that the top three models are ANN with a good accuracy score of 85%, precision of 87%, sensitivity of 89%, F1-score of 88%, and specificity of 78%. The second-best model is the DT, with an accuracy score of 87%, precision of 91%, sensitivity of 88%, F1-score of 90%, and specificity of 82%. The RF achieved the best results of all models with an accuracy score of 89%, precision of 92%, sensitivity of 89%, F1-score of 90%, and specificity score of 85%.

In Table 6, we demonstrate the results achieved after applying feature selection via SelectKBest using the ANOVA function to all models, where the 3 least important features were removed from the dataset (Gender, Live_attendance and video). Therefore, the improvement in the models' performances was slightly low.

Table 6: Models performance after feature selection

	Accuracy%	Precision%	Sensitivity%	F1-Score%	Specificity%
SVM	80	84	87	85	69
DT	87	91	88	90	83
LR	84	87	89	88	74
RF	89	92	89	90	86
ANN	86	88	90	89	80

From Table 6, we can see those 2 models, the RF and DT, maintained a stable and unchanged performance across the board, except for the specificity score which showed a slight increase by 1% for both, while the ANN, SVM, and LR have improved their performance slightly in all metrics. Despite the improvements in those 3 models, RF is still the one that outperforms other models in all metrics of performance, with an accuracy score of 89%, precision of 91%, sensitivity score of 93%, F1-score of 92%, and specificity score of 86%. After performing our internal validation, our results show that the best performing model to predict the SP in DL time is the Random Forest.

5.7 Comparison with Benchmark Models

In this section, we perform a comparison of our model with the two benchmarks introduced above in Section 5.4. This comparison represents an external validation of our proposed model. Our model and the results obtained with benchmarks, are compared, analyzed, and discussed.

5.7.1 Comparison with Benchmark Model1 (Karalar et al SP Model)

The first model was created by (Karalar et al., 2021), where the authors aimed to create an ensemble model to predict the performance, which is made up of Extra trees, Random Forest and Decision Trees.

The dataset used by the authors is the same one used in this research; the authors applied some filters and did not use the whole dataset, which is made up of 3223 records; only 2045 records were used for prediction. For example, records that had a high number of attendances in the live sessions, but a low total number of minutes spent in the sessions were removed from the dataset. In our research, 2045 records were considered in total, as explained in the pre-processing stage earlier, where some filtrations were also applied to the data.

One important thing to mention is the attributes considered in modeling the performance, the authors have applied feature selection by the Sklearn library using the SelectKBest and ended up considering only 5 attributes from the dataset, which are quiz results, degree, material downloads, lectures note downloads, and the video watching of the sessions, in that order. Although we have used the same dataset, we got different results when applied our feature selection. Table 7 shows a comparison between our model and Karalar et al.'s model in terms of feature selection result.

Table 7: Feature selection results of our model compared to Karalar et al.'s model

Karalar et al. model feature selection	**Our model feature selection**
Quiz	Quiz
Degree	Live
Lecture_Notes	Degree
Materials	Live_ratio
Video	Lecture_Notes
Gender	Material
Live_attendance	Video
Live	Live_attendance
Live_ratio	Gender

Comparing our results to Karalar et al.'s results for feature selection, we can see that the quiz marks are the most significant features in the dataset. However, in our feature selection, the live attribute comes in the second place, degree in the third place, live ratio, and lecture notes in the fourth and fifth place, respectively. Our least important features are gender, live attendance, and video watching of the sessions. While for Karalar et al., their least important attributes are live, attendance live, and live ratio.

Now let's have a closer look at the performance scores of the ensembled models compared to our RF model. According to the confusion matrix of the benchmark model with the values obtained from their article: TP: 359 FP:14 TN:131 FN:110, and our confusion matrix displayed in the Figure 17.

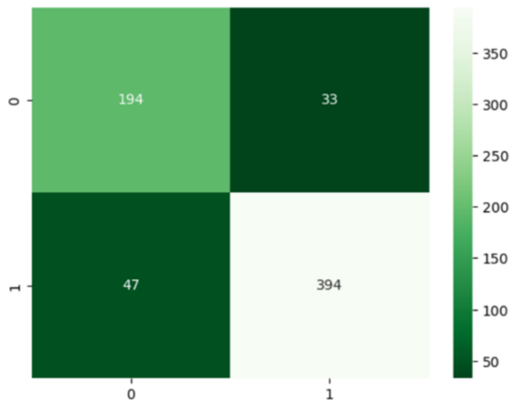

Figure 17: Confusion matrix for our RF model

We have computed the performance scores for both models; Table 8 shows the results, and compares the models in terms of accuracy, precision, sensitivity, F1-score, and specificity.

Table 8: Comparison between our model and benchmark model 1

	Accuracy%	**Precision%**	**Sensitivity%**	**F1-score%**	**Specificity%**
Karalar et al.	76	96	77	85	90
Our RF Model	89	92	89	90	86

Table 8 shows that our model has a better accuracy score with the value of 89%, compared to the benchmark's with the value of 76%. Our sensitivity score of the RF model is 93%, compared to Karalar et al.'s model with the value of 77%. On the other hand, Karalar et al.'s model has a slightly higher precision score with the value of 96%, compared to our model with the value of 92%. However, our model F1-score is higher than Karalar et al.'s model and reached 90%, in comparison to an 85% for Karalar et al.'s model. The specificity score of our model was 86%, while Karalar et al.'s model scored 90%.

To conclude the results, we will explain the models and the results achieved in depth. From the confusion matrix of Karalar et al.'s model, the false negative (FN) value obtained is 110, which indicates there is a substantial number of students who actually failed and predicted as pass. This result obtained by Karalar et al., is the reason behind their model's low sensitivity score of 76%, and in turn, reflected on the F1-score to reach 85%, since the F1-score balances sensitivity and precision to create harmony within the model. This may result in giving wrong indication to the instructor about the performance of his students. Karalar et al.'s false positive (FP) value is 14, which means that, students who actually passed and predicted as fail is as low as 14. Karalar et al.'s FP value explains their model high specificity score of 90%, and precision of 96%, as both depend on the FP value. To predict the SP accurately, there should be a harmony and balance in the model between the FP and FN, which is not evident in Karalar et al.'s model. To create a balanced, good performing model to predict the SP in DL time, the F1-score is an important measure of performance that must be considered as it balances the model.

Therefore, our model outperforms Karalar et al.'s model, as it is more balanced, and harmonic in predicting the performance in DL with F1-score of 90%, compared to 85%, for Karalar et al.'s model.

Our RF model outperforms Karalar et al.'s model in terms of accuracy, and F1-score, with values of 89% and 90%, respectively for our Rf, compared to 76%, and 85%, in that order for Karalar et al.'s model.

5.7.2 Comparison with Benchmark Model 2 (Aydoğdu's Model)

Our second benchmark model is an ANN based model created by Aydoğdu in 2019 to predict SP in DL time. This model uses 9 attributes: gender, content score, time spent on content, number of entries to reach content, homework score, number of attendances to live sessions, total time spent in live sessions, number of attendances to archived courses, and total time spent in archived courses. The model output is called "success status", which indicates whether the student will succeed in the course or not, as explained earlier in the Literature review in Section 3.2. The dataset was made up of 3518 records used in predicting the performance. The model performance was measured using the accuracy achieved.

In order to compare our results to Aydoğdu's Model, we have rebuilt the same ANN model used by Aydoğdu. The same settings that Aydoğdu used for his model were used to configure the rebuilt ANN model. We compared the models according to the accuracy score the model achieved. Aydoğdu had his ANN setting with a *batch_size* of 128, *number of epochs* = 1000, optimization function is the *adam,* and number of neurons in the hidden layer was 8. As seen in Figure 18, the same settings stated by Aydoğdu in his article were applied to our ANN.

```
import tensorflow as tf
ann = tf.keras.models.Sequential()
ann.add(tf.keras.layers.Dense(units = 9, activation = 'relu'))
ann.add(tf.keras.layers.Dense(units = 8, activation = 'relu'))

ann.add(tf.keras.layers.Dense(units = 1, activation = 'sigmoid'))
ann.compile(optimizer = 'adam', loss = 'binary_crossentropy', metrics = ['accuracy'])
ann.fit(X_train, y_train, batch_size = 128, epochs = 1000)
```

Figure 18: ANN settings with comparison to benchmark model 2

As revealed earlier, the performance measure that the model will be compared to is the accuracy. Our model achieved an accuracy score of 86%, while Aydoğdu's model achieved a score of 80%. At this point, we can conclude that our model has a better performance than Aydoğdu's Model.

5.8 Summary and Validity of Our Results

In this chapter, we have discussed the dataset used to predict the SP in DL time. The data has been analyzed and cleaned to remove irrelevant records, and feature selection took place to come up with the most relevant attributes to SP in DL time. The attributes used in modeling are live, quiz, video, lecture downloads, material downloads, and video. For further explanation of the attributes, refer to Section 5.1.

In Section 5.5, we discussed the setup of our models, DT, SVM, LR, RF, and ANN. We have performed an internal validation of the models' performance according to a number of matrices, namely: accuracy, sensitivity, precision, F1-score, and specificity. According to the results obtained, RF was the best performing model to predict the SP in DL time.

Lastly, an external validation of our model took place. We compared our model to 2 benchmarks, Karalar et al.'s model and Aydoğdu's model. When compared to Karalar et al.'s, the results showed that our model outperformed their model in accuracy and F1-score, which shows that our model is more balanced than the Karalar et al.'s model. The second benchmark was Aydoğdu's ANN model. We setup our ANN to the same settings of Aydoğdu's model. The models were compared according to the accuracy score, and our ANN outperformed the Aydoğdu's model.

Our thesis has targeted a new phenomenon, that has affected the educational institutions worldwide. The distance learning situation was never the main mode of learning before the pandemic outbreak. Therefore, the data is still rare, and we have suffered in the process of data collection to find good data with a good number of records. Specially that many universities consider its data as confidential and not for sharing for research purposes. The literature review proves that there is no sufficient data, many other studies have used datasets with very few numbers of records for modeling the SP and validating their results. We look forward to being able to find or collect data with more attributes and records for more than one course in many countries to validate our models and propose more sophisticated machine learning or deep learning models.

Chapter 6: Conclusion

The rapid worldwide shift to DL that continued for long after the pandemic remains a valid method of learning nowadays. The fact that all universities worldwide are vigilant about students' performance as it can play a major role in determining the success or failure of the institution, was the reason behind many recent research taking place, where there was a need to find the correct definition that represents student performance in DL, based on what attributes can the SP be predicted and which model is best to predict it as well.

In this thesis, a study of the literature on the student performance prediction in conventional and distance learning modes has been analyzed. The reviewed literature allowed us to identify and discover the most suitable definition of student performance that suites the distance learning time.

By using a literature review study, on one hand, we have distinguished between three categories of learning attributes impacting the SP. (1) Conventional mode attributes (e.g., extra-curricular activities) (2) Distance learning attributes (e.g., live attendance) and (3) Common attributes between both modes of learning (e.g., previous grades obtained).

On the other hand, we have suggested a new definition of the SP in the DL mode that considers the student performance as: student's knowledge gained from effective attendance of the sessions, which is reflected on students' grades on assessments throughout the course.

In order to derive the most suitable prediction model for SP in DL mode, we have conducted an empirical experiment that allowed us to compare the performance of different ML algorithms, namely, Artificial Neural Networks, Logistic Regression, Support Vector Machine, Decision Trees, and Random Forest, in predicting SP. Our results vote for RF as the most appropriate model for our prediction problem. RF achieved an accuracy score of 89%, an F1-score of 90%, and specificity score of 86%.

To externally validate our proposed model, we empirically compared its performance to a set of benchmarks used to predict SP, namely, Karalar et al. model and Aydoğdu's model.

The results showed that our RF model is more balanced and scored better in terms of accuracy and F1-score values, compared to Karalar et al. ensemble model. And our ANN had better accuracy score compared to Aydoğdu's ANN model, where the same model settings were used but different datasets.

At the end of this thesis, we believe that COVID-19 has impacted the learning environments, the students' behavior, and relationships among most of the stakeholders of education. Thus, we expect that new volumes of data reflecting this impact will be available in the near future. They were collected intensively and automatically. This data will be of critical importance to extend our work in different directions:

Discover and identify empirically new factors and attributes impacting student performance, which will contribute effectively to developing more comprehensive definition of SP in DL mode.

The characteristics of such data will better dictate the most suitable modeling technique for predicting and assessing SP.

Finally, the availability of data will be valuable in validating internally and externally different proposed models of SP.

References

Adnan, M., Habib, A., Ashraf, J., Mussadiq, S., Raza, A. A., Abid, M., Bashir, M., & Khan, S. U. (2021). Predicting at-Risk Students at Different Percentages of Course Length for Early Intervention Using Machine Learning Models. *IEEE Access*, *9*, 7519–7539. https://doi.org/10.1109/ACCESS.2021.3049446

Al Karim, M., Ara, Mst. Y., Masnad, Md. M., Rasel, M., & Nandi, D. (2021). Student performance classification and prediction in fully online environment using Decision tree. *AIUB Journal of Science and Engineering (AJSE)*, *20*(3), 70–76. https://doi.org/10.53799/ajse.v20i3.173

Al Karim, M., Tahsin, M. S., Ahmed, M. U., Ara, Mst. Y., & Hasnat Chowdhury, S. A. (2022). Evaluating the Performance of ID3 Method to Analyze and Predict Students′ Performance in Online Platforms. *2022 International Conference on Computer Science and Software Engineering (CSASE)*, 60–64. https://doi.org/10.1109/CSASE51777.2022.9759650

Altaf, S., Soomro, W., & Rawi, M. I. M. (2019). Student Performance Prediction using Multi-Layers Artificial Neural Networks: A Case Study on Educational Data Mining. *Proceedings of the 2019 3rd International Conference on Information System and Data Mining - ICISDM 2019*, 59–64. https://doi.org/10.1145/3325917.3325919

Aydoğdu, Ş. (2020). Predicting student final performance using artificial neural networks in online learning environments. *Education and Information Technologies*, *25*(3), 1913–1927. https://doi.org/10.1007/s10639-019-10053-x

Brahim, G. B. (2022). Predicting Student Performance from Online Engagement Activities Using Novel Statistical Features. *Arabian Journal for Science and Engineering*, *47*(8), 10225–10243. https://doi.org/10.1007/s13369-021-06548-w

Dabhade, P., Agarwal, R., Alameen, K. P., Fathima, A. T., Sridharan, R., & Gopakumar, G. (2021). Educational data mining for predicting students′ academic performance using machine learning algorithms. *Materials Today: Proceedings*, *47*, 5260–5267. https://doi.org/10.1016/j.matpr.2021.05.646

Hamoud, A. K., Hashim, A. S., & Awadh, W. A. (2018). Predicting Student Performance in Higher Education Institutions Using Decision Tree Analysis. *International Journal of Interactive Multimedia and Artificial Intelligence*, *5*(2), 26. https://doi.org/10.9781/ijimai.2018.02.004

Hong, B., Wei, Z., & Yang, Y. (2017). Online education performance prediction via time-related features. *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, 95–100. https://doi.org/10.1109/ICIS.2017.7959976

Karalar, H., Kapucu, C., & Gürüler, H. (2021). Predicting students at risk of academic failure using ensemble model during pandemic in a distance learning system. *International Journal of Educational Technology in Higher Education*, *18*(1), 63. https://doi.org/10.1186/s41239-021-00300-y

Kumar, S. (2021). *Defining And Measuring Academic Performance of Hei Students- A Critical Review*. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(6), 3091–3105

Matzavela, V., & Alepis, E. (2021). Decision tree learning through a Predictive Model for Student Academic Performance in Intelligent M-Learning environments. *Computers and Education: Artificial Intelligence*, *2*, 100035. https://doi.org/10.1016/j.caeai.2021.100035

Misopoulos, F., Argyropoulou, M., & Tzavara, D. (2018). Exploring the Factors Affecting Student Academic Performance in Online Programs: A Literature Review. In A. Khare & D. Hurst (Eds.), *On the Line* (pp. 235–250). Springer International Publishing. https://doi.org/10.1007/978-3-319-62776-2_18

Mubarak, A. A., Cao, H., Hezam, I. M., & Hao, F. (2022). Modeling students′ performance using graph convolutional networks. *Complex & Intelligent Systems*, *8*(3), 2183–2201. https://doi.org/10.1007/s40747-022-00647-3

Narad, A., & Abdullah, B. (2016). Academic Performance of Senior Secondary School Students: Influence of Parental Encouragement and School Environment. *Rupkatha Journal on Interdisciplinary Studies in Humanities*, *8*(2), 12–19. https://doi.org/10.21659/rupkatha.v8n2.02

Pallathadka, H., Sonia, B., Sanchez, D. T., De Vera, J. V., Godinez, J. A. T., & Pepito, M. T. (2022). Investigating the impact of artificial intelligence in education sector by predicting student performance. *Materials Today: Proceedings*, *51*, 2264–2267. https://doi.org/10.1016/j.matpr.2021.11.395

Phipps, R., & Merisotis, J. (2022). *What's the Difference: A Review of Contemporary Research on the Effectiveness of Distance Learning in Higher Education*, Eric.

Ramaswami, G., Susnjak, T., & Mathrani, A. (2022). On Developing Generic Models for Predicting Student Outcomes in Educational Data Mining. *Big Data and Cognitive Computing*, *6*(1), 6. https://doi.org/10.3390/bdcc6010006

Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, *2*(3), 160. https://doi.org/10.1007/s42979-021-00592-x

Seo, S., Kim, Y., Han, H.-J., Son, W. C., Hong, Z.-Y., Sohn, I., Shim, J., & Hwang, C. (2021). Predicting Successes and Failures of Clinical Trials With Outer Product–Based Convolutional Neural Network. *Frontiers in Pharmacology*, *12*, 670670. https://doi.org/10.3389/fphar.2021.670670

Tejedor, M., Woldaregay, A. Z., & Godtliebsen, F. (2020). Reinforcement learning application in diabetes blood glucose control: A systematic review. *Artificial Intelligence in Medicine*, *104*, 101836. https://doi.org/10.1016/j.artmed.2020.101836

Wang, G.-H., Zhang, J., & Fu, G.-S. (2018). Predicting Student Behaviors and Performance in Online Learning Using Decision Tree. *2018 Seventh International Conference of Educational Innovation through Technology (EITT)*, 214–219. https://doi.org/10.1109/EITT.2018.00050

UAEU

جامعة الإمارات العربيـة المتحدة
United Arab Emirates University

UAE UNIVERSITY MASTER THESIS NO. 2023: 12

Our ultimate goal in this thesis is to introduce a new definition and develop a new model to predict student performance by considering more appropriate attributes and factors that better represent new concepts, behaviors, and activities subsequent to the distance learning mode.

Iman Saad Megdadi received her MSc in Software Engineering from the Department of Computer Science and Software Engineering, College of Information Technology at UAE University, UAE. She received her BSc in Computer Science from the University of Sharjah, UAE.

www.uaeu.ac.ae

Online publication of thesis:
https://scholarworks.uaeu.ac.ae/etds/

UAEU عمادة المكتبات Libraries Deanship

جامعة الإمارات العربيـة المتحدة
United Arab Emirates University

Digital Library Services Section - قسم الخدمات المكتبية الرقمية