

Received October 22, 2020, accepted November 2, 2020, date of publication November 5, 2020, date of current version November 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3036336

A Machine Learning-Based Recommender System for Improving Students Learning Experiences

NACIM YANES^{1,2}, AYMAN MOHAMED MOSTAFA^{1,3}, (Member, IEEE),
MOHAMED EZZ^{1,4}, (Member, IEEE), AND SALEH NAIF ALMUAYQIL¹

¹College of Computer and Information Sciences, Jouf University, Sakaka 72314, Saudi Arabia

²RIADI Laboratory, La Manouba University, Manouba 2010, Tunisia

³Faculty of Computers and Informatics, Zagazig University, Zagazig 44519, Egypt

⁴Faculty of Engineering, Al-Azhar University, Cairo 11651, Egypt

Corresponding author: Ayman Mohamed Mostafa (amhassane@ju.edu.sa)

This work was supported by the Deanship of Scientific Research, Jouf University, Saudi Arabia, under Project 40/336.

ABSTRACT Outcome-based education (OBE) is a well-proven teaching strategy based upon a predefined set of expected outcomes. The components of OBE are Program Educational Objectives (PEOs), Program Outcomes (POs), and Course Outcomes (COs). These latter are assessed at the end of each course and several recommended actions can be proposed by faculty members' to enhance the quality of courses and therefore the overall educational program. Considering a large number of courses and the faculty members' devotion, bad actions could be recommended and therefore undesirable and inappropriate decisions may occur. In this paper, a recommender system, using different machine learning algorithms, is proposed for predicting suitable actions based on course specifications, academic records, and course learning outcomes' assessments. We formulated the problem as a multi-label multi-class binary classification problem and the dataset was translated into different problem transformation and adaptive methods such as one-vs.-all, binary relevance, label powerset, classifier chain, and ML-KNN adaptive classifier. As a case study, the proposed recommender system is applied to the college of Computer and Information Sciences, Jouf University, Kingdom of Saudi Arabia (KSA) for helping academic staff improving the quality of teaching strategies. The obtained results showed that the proposed recommender system presents more recommended actions for improving students' learning experiences.

INDEX TERMS Outcome-based education, educational data mining, recommender systems, students learning experiences, teaching strategies.

I. INTRODUCTION

Over the last century, a significant concern on the ability of education systems to equip the students with the adequate professional and career preparation needed for the 21st century has evolved. In response, OBE has been proposed as a theory that bases each part of an educational system around outcomes. As stated by [1], outcomes are the abilities that the students can acquire at the end of a learning experience. The most important aspect of an outcome is that it should be observable and measurable.

OBE is an enlightening approach of teaching strategy that is based upon a predefined set of expected outcomes [2].

Outcome-based education means clearly focusing and organizing everything in educational system which is

The associate editor coordinating the review of this manuscript and approving it for publication was Inês Domingues^{id}.

important for all students to be able to do successfully at the end of course/program/graduation as stated by many works in the literature including [2]–[5], and [6] and many others.

Teaching is then designed to engage students in learning activities that optimize their chances of achieving those outcomes, and assessment tasks are designed to enable clear judgments as to how well those outcomes have been attained [3], and [5].

Nowadays, this approach is applied in tertiary education by specifying the outcomes at three levels: (i) PEOs which are broad statements that describe the career that the program is preparing students to achieve, (ii) POs which are narrower statements that describe the knowledge, the skills and competence, those students are expected to know and be able to do by the time of graduation, (iii) COs which are statements that describe what students are expected to know, attitudes they

are expected to hold, and what they are able to do after the completion of a course [7].

In an OBE learning process, no single specified style of teaching or assessment is specified and by the end of the educational experience each student should have achieved the program educational objectives [1]. Teachers are expected to act as facilitators of learning by creating and sustaining an effective learning environment for students to develop the competencies that the program of study expects to foster [4].

No course of the study program can be dealt with in isolation. Each course is an element of a program and often related to other courses of the program. OBE does not interfere with the academic freedom of the teacher. It merely asks the teacher to follow a process in offering a course. The process consists of writing the Cos about what the students should be able to do in the context of its relationship to the other elements of the program, design assessment (how to measure the ability of students to do what they are expected to do) and instruction (how the teacher proposes to facilitate the students to acquire the ability to do what they are expected to do). All the decisions in all the three steps of the process are made by the teacher [1], [2].

Assuming the Cos, are written carefully to reflect the intended outcomes of the course, the attainment of the COs should lead to the attainment of the POs and consequently the PEOs. However, it is necessary to follow well-defined COs assessment process consistently to plan for continuous improvement in the quality of the learning. The COs assessment process is carried out based on both direct and indirect assessments [2]. The direct assessment can be determined from the performances of the students in all the relevant assessment methods. With regard to the indirect assessment, it is based on the collected feedbacks from students.

Feedbacks are taken using online forms. Teachers are asked to prepare course reports based on the results of the COs assessment process. These course reports are used at the end of each semester to improve the student learning experience. Indeed, each teacher proposes several recommended actions to enhance the quality of the courses he/she taught and therefore improving the overall educational program.

In recent years, assisting and helping academic staff in improving the teaching and learning quality is considered a major concern in both research and practice. In fact, the huge amount of data accumulated in educational institutions creates a gold mine, from where useful knowledge can be discovered. This initiates a new and fast-growing interdisciplinary research field of data mining called Education Data Mining (EDM) [8], and [9]. Several methods and applications have been proposed in EDM [10], and [11].

Indeed, various tasks and applications in the field of EDM have been studied and their purposes have been categorized [10]. Besides, [11] has described the most popular EDM methods. However, these proposed taxonomies are not exhaustive; they do not cover all the possible tasks. Indeed, there are many more specific objectives depending on the viewpoint of the final user.

These applications can follow both applied research objectives such as improving learning quality, as well as pure research objectives, which tend to improve the understanding of the learning process. The applications of EDM can target any of the stakeholders involved in educational systems, such as students, teachers, and researchers themselves [10].

Making discoveries and providing recommender systems can help the teachers by improving teaching performance and making decisions [10]. The applications of EDM have been reported in a number of academic works. A combination of neural networks and experts' prior knowledge has been applied in [12] to predict and evaluate student's learning outcomes of an academic program and as a result enhance teaching quality. A second work [13] applied K-means clustering algorithm to investigate the relationship between skills taught in Business programs and the title of the program using a dataset extracted from the program catalog.

The concept of using assessment information for course improvements is nothing new; teachers have been doing it forever. It is an integral part of being an effective educational professional. Assessment results can help teachers decide how useful their assessment strategies have been and what changes are needed to improve their effectiveness [14].

The sources of that assessment information were different; instructional decisions and actions were often based on intuition, teaching philosophy, and personal experience [15]. Indeed, there is no a systematic process (or data-driven approach) for course improvements. Teachers often like to try out different instructional strategies based on their own judgments. This trial-and-error process of choosing a strategy, and judging its efficiency is different for every teacher. How do the teacher decide which strategy to try, and how do he know whether it "worked"? The trial-and-error process is not very consistent and can lead to ambiguous results [15].

Considering the lack of knowledge and guidelines on the actions to improve students' learning [15] and the recent remarkable success of the educational data analytics, this paper proposes a recommender system, using different machine learning algorithms, for predicting suitable actions to enhance the quality of the courses and thus to improve the overall educational program. Predicting the suitable actions will be based on courses specifications, students' academic records, and course learning outcomes' assessments.

The data are fed as input to the proposed model and the recommender system will answer the following questions:

1. Which is the best method for this multi-label classification problem?
2. Which is the best algorithm to predict the most accurate courses improvement actions?
3. Which features are selected to be the most relevant predictors for courses improvement actions?

The remaining of this paper is structured as follows. Section II presents the related works. Section III discusses the methodology carried out in this study from the data collection to the model proposal. Section IV discusses the experimental results. Section V explores the analysis and discussion of

TABLE 1. Summary of Major Acronyms.

Acronym	Definition	Acronym	Definition
BR	Binary Relevance	LR	Logistic Regression
CC	Classifier Chain	ML-KNN	Multi-Label K-Nearest-Neighbors
COs	Course Outcomes	MOOC	Massive Open Online Course
CR	Course Report	MRS	Model Ranking & Selection
CRISP-DM	Cross-Industry Standard Process for Data Mining	NB	Naive Bayes
DM	Data Mining	OBE	Outcome Based Education
DT	Decision Tree	OvA	One-vs.-All
EDM	Education Data Mining	PT	Problem Transformation
FP	False Positive	PEOs	Program Educational Objectives
GBM	Gradient Boost Machine	POs	Program Outcomes
ITS	Intelligent Tutoring System	RF	Random Forest
LP	Label Powerset	SVC	Support Vector Classifier
LMS	Learning Management System	TP	True Positive

the proposed recommender system. Finally, the paper ends in section VI with a conclusion and future directions. The definitions of the major used acronyms in this paper are presented in Table 1.

II. RELATED WORKS

Nowadays, it is possible to generate a great volume and wide variety of data from educational environments such as those related to students' academic records, assessment files, courses reports and the records of interactions by e-mails between students and teachers. The generated data by educational environments play a significant role for decision support and improvement of the learning process. The improvements can be reached from the students' data analysis through of their behavior, satisfaction, and performance [16]. DM techniques are an alternative to extract knowledge from these data and therefore to enhance the learning process [17].

EDM has been a fast-growing interdisciplinary research field [8] and [11]. The researches in EDM expanded into several areas, including studies on interactions between the educational actors, planning and assessment of educational programs, teaching strategies, course contents and educational outcomes, improving students' performance, identifying students' learning styles, supporting weak students to

prevent them from failure and dropout, gaining knowledge about different educational phenomena and solving the relevant complex problems [16] and [17].

Different kinds of data can be collected in order to resolve educational problems depending on the type of the educational environment even traditional classroom education or computer-based education with the information system used, such as LMS, ITS, and MOOC [11]. Data preprocessing is considered a hard and complicated task. Educational data is available whether raw, original, or primary data to solve problems that are not in the appropriate form. Therefore, it is necessary to convert the data to an appropriate form for solving each specific educational problem [18].

The majority of traditional DM techniques including but not limited to visualization, classification and clustering techniques have been already applied successfully in the educational domain. However, educational systems have also some special characteristics with hierarchical and longitudinal data that require a specific treatment of the mining problem and preprocess of the data [11].

Nowadays, a wide array of well-known tools and frameworks can be used for the purposes of conducting EDM research. As presented in [19], 40 tools used for data mining have been reviewed in the area of education. This review is useful to researchers interested in learning about these tools not just at a theoretical level, but also in terms of practical application and use. The majority of the EDM researchers use their own data for solving their specific educational problems [11]. Gathering and preprocessing educational data is a hard and very time-consuming task [18]. That's why another option consists in using public datasets that are currently available for free download on the Web. Examples of EDM public datasets are proposed in [11].

Several works have been proposed in the literature to tackle different educational problems. These works have different objectives depending on the viewpoint of the final user (students, instructors, administrators or other stakeholders). In the Sequel, we will present some of these proposed works. A complete survey of these works is beyond the scope of the current paper and more details can be found on [10] and [11].

The first work we identified has been proposed in [20] for helping students to select the more suitable faculty based on several criteria including his/her grades in different subjects in high school, the country state where he/she is located and the gender. The proposed enrollment recommender system consists of two phases; training phase and runtime phase. The training phase takes the previous high school database and faculty database as input and generates the faculty student model, whilst the runtime takes a new student as input and produces as an output the recommendation for this student, that is, either suitable or not, to join that faculty. As shown in [20], a recommender system was applied on the enrollment process for faculty of Engineering, Al-Azhar University. They employed and compared four machine learning algorithms. Based on the obtained results, the Alternative DT

algorithm outperforms the other algorithms with an overall accuracy close to 80%.

A second work was presented in [21] where a student performance prediction system was proposed to determine students who would be expected to do well in the Faculty; precisely to succeed in studying engineering programs. They used an Artificial Neural Network model to predict the performance of a student before he/she starts his/her sophomore year in Engineering studies based on a number of factors such as high school score, results in some subjects in the freshman year including mathematics and electronics, student's gender, type of High school whether it is private or public. The proposed prediction model was tested and the overall result was 84.6%.

The third work we identified in the literature focuses on factors affecting the students and the instructor performances. Indeed, the authors in [22] investigate the factors that affect students' achievements to enhance the quality of the educational system and propose a predicting model of the instructor performance. Indeed, the main goal of the authors is to build a classification model that enables to predict factors affecting student performance. The authors used four well-known DM techniques, namely J48 DT, MLP, NB, and SMO. Based on the obtained results, authors claim that J48 DT algorithm achieves the best performance compared to the other algorithms with an accuracy of 84.8%. Another interesting issue presented in this research affirmed that the performance of an instructor is mainly affected by the number of courses that is taught.

Another work which is interested in the students' performance was proposed by [23]. The authors claim that students' performance is related to students' dynamic behaviors more strongly than teaching environment and students intrinsic characteristics. They stressed that living and study habits are greatly associated with academic success. Based on this observation, the authors proposed a framework to model students' academic performance using their behavior pattern. Discovering students with poor academic performance early is helpful to supervision and developing good study habits. The authors used the smart card records to build the students' behavior pattern. After that, they used a regularized multi-task model for the classification task to predict the students' performance of each course simultaneously. The experiments results showed a high recall of poor performance discovery and enough feasibility for early warning. As a matter of fact, the authors found that some behaviors such as shopping, meal in canteens, and leaving dorm before 8 a.m. have relative significant relation with performance. They stated that "moderate diversity on campus and a good lifestyle especially on meals and getting up are in favor of good performance".

The fifth identified work focuses also on predicting outcomes of student performance at the end of a one year school cycle [24]. Indeed, the authors employed two datasets obtained from a repository of the State Department of Education of the Federal District of Brazil. The first dataset

contains only attributes collected prior to the beginning of the school year; however the second one contains those same attributes, but included furthermore a few new variables, such as 'absences' and 'grades'. After that, the authors built a classification model based on the GBM for each dataset to predict student performance, allowing them to compare the predictive capability at two different moments in the academic cycle. The experiments results showed that, though the attributes 'grades' and 'absences' were the most relevant for predicting the end of the year academic outcomes of student performance, the analysis of demographic attributes reveals that 'neighborhood', 'school' and 'age' are also potential indicators of a student's academic success or failure.

Another educational problem has been treated by [8]. In fact, the authors have focused on a very important problem in higher education which is student dropout. As stated in this work, early detection of vulnerable students can lead to the success of any retention strategy. At-risk students would be provided with academic and administrative support to increase the chance of staying on the course. The proposed work aims to disclose interesting patterns, which could contribute to predicting students' performance and dropout, based on their pre-university characteristics, admission details, and initial academic performance at university. The authors introduced a new feature transformation method to improve the accuracy of conventional classifiers and therefore maximize the students' attention and recommending courses for them based on their progress.

In this paper, we have described some academic works proposed to solve educational problems. However, the number of possible objectives or educational problems in educational data mining is huge and there is no taxonomy that covers all the possible tasks. In fact, there are many more specific objectives depending on the viewpoint of the final user (students, teachers, scientific researchers or administrators). Indeed, to the best of our knowledge, this is the first study to deal with predicting the recommended actions to improve the students' learning experiences based on course specifications, academic records, and course learning outcomes' assessments. The current paper implements a recommender system using applied machine learning algorithms with data engineering by formulating multi-label multi-class classification problems. The proposed recommender system aims to predict the suitable actions for enhancing student learning experiences. As we mentioned in the previous section, teachers are asked at the end of each semester to prepare course reports based on the results of the COs assessment process. These course reports are used to improve the student learning experience. Indeed, each teacher proposes several recommended actions to enhance the quality of the courses he/she taught and therefore improving the overall educational program. Therefore, a recommender system that predicts automatically and efficiently the more suitable actions will be very helpful for teachers in preparing their course reports.

III. METHODOLOGY

The proposed machine learning-based recommender system is considered a simulation to the hybrid recommender approach presented in [25] where the student's or learner's average score in specific or all modules can affect the quality of course content that can lead to poor enhancements in student learning outcomes.

The dataset of the proposed recommender system is extracted based on the student course learning outcomes, academic records, and course specification for each distinct course. The student in each semester studies several courses in different scientific domains. In each studied course, a course file must be prepared at the end of the semester in order to determine the proposed improvements that can be made based on the student records and the course learning outcomes. An adaptive methodology is proposed to recommend the best suitable actions for each taught course based on the overall student academic records in the course under concern. The adaptability of the proposed methodology is executed by applying the dataset on different problem transformation and adaptive methods for dynamically predicting the suitable actions that can be taken.

The methodology is formulated based on different steps starting from dataset collection, data preprocessing and problem transformation. The methodology is illustrated in the following steps:

A. DATASET COLLECTION

To the best of our knowledge, there is no dataset addressing the main topic of the current paper. Therefore, the dataset in our research was collected from the course reports of the College of Computer and Information Sciences at Jouf University, Kingdom of Saudi Arabia (KSA). In the college of Computer and Information Sciences, the students are allowed to study in four years divided into eight semesters. The students are distributed into three scientific departments: Information Systems, Computer Science, and Computer engineering and networks.

The dataset was collected from the courses reports of 127 scientific courses taught in three departments during 4 semesters in the two academic years 2018 and 2019. In addition, there are graduation projects 1 and 2 in each department. Both graduation projects 1 and 2 are taught in separate semesters where Project 1 course is the prerequisite of Project 2 course. Each course has two sections respectively for male and female students.

At the end of each course, the teacher assesses the course outcomes in order to recommend several actions to enhance the quality of the course and therefore the overall educational program. The course outcomes assessment process is carried out based on both direct and indirect assessments. The direct assessment can be determined from the performances of the students in all the relevant assessment methods including final exams, midterm exams, assignments, labs and projects. With regard to the indirect assessment, it is based on the

TABLE 2. Teaching Mechanisms.

Course Code	Course Name	Theoretical	Exercise	Practical
CS 230	Fundamentals of Information Systems	3	1	0
CS 350	Introduction to Database Systems	3	0	2
CS 360	Modern Application Development	1	0	2
CS 410	Web Application Development	2	0	2
----	----	----	----	----

collected feedbacks from students. Since the students' feedbacks can be subjective and therefore it cannot be a valid assessment method, we used only the direct assessment in our proposal. In fact, we used only valid assessment methods including (final exams, midterm exams, assignments, projects and lab exams) to predict the suitable actions to improve each course. This dataset is considered the first dataset that has been collected to handle the recommended actions for improving and enhancing courses.

As presented in [26], different factors can affect the integrity of E-learning in Saudi Arabian universities such as obtaining solutions from other students during the examination and as a result, the success rate will be changed and as a result the recommendations for improving teaching strategies will be affected.

Another case study for studying the effect of E-learning on higher educational universities was presented in [27]. The authors of that research focused on the student satisfaction, self-efficacy, and the content of E-learning. These factors can also affect the recommended actions that must be taken to improve the course contents and teaching strategies. In order to improve the performance of the proposed recommender system, the collected dataset must be complete and accurate from both teaching and quality aspects. Each course under consideration should be identified based on the following factors:

1) NO OF STUDENTS IN EACH SECTION

The number of students in each section of the course must be identified to ideally reflect the suitable actions for improving the quality of teaching strategies. The more the students in each section, the more it will be helpful to evaluate the course learning outcomes more effectively.

2) COURSE CREDIT HOURS

Each course taught is classified based on its teaching mechanism. The teaching mechanism for each course is whether it is theoretical, exercises, and practical lab. Each course is based on at least two applied mechanisms. As shown in Table 2, a sample of several courses with different teaching mechanisms is presented.

In Table 2, the teaching mechanism in each course is completely different from other courses according to methods of teaching. For example, the course CS 230 has a teaching mechanism (3, 1, and 0). This means 3 credit hours of theoretical lectures, 1 credit hour for exercises with no practical hours. The course CS 350 which is “Introduction to Database Systems” is considered a theoretical and practical course. So, it has a teaching mechanism (3, 0, and 2) while the course CS 360 basically depends on teaching with labs more than the theoretical lectures.

As a result, it has a teaching mechanism (1, 0, and 2). The course CS 410 is based also on practical but the theoretical lectures are more than the practical lectures with a teaching mechanism (2, 0, and 2). These teaching mechanisms will be a main concept in determining the best recommended actions based on the distribution method of credit hours.

3) NO OF TOPICS IN EACH COURSE

Each course contains a number of topics that differ from any other course based on the scientific content. There is a direct relationship between the numbers of course topics and the learning outcomes. The more the number of course topics, the more the learning outcomes will be applied and the more recommendations will be taken to improve the course content. Formula (1) explains the relationship between the course topics (C_T) and the course learning outcomes (CLO) and the recommended actions (A_R):

$$\forall CLO_i \subset CLO_n \text{ where } CLO_n \in PLO \text{ and } \sum_{i=1}^n C_{Ti} \rightarrow A_R \quad (1)$$

where, the CLO_i is each course learning outcome in the intended course and CLO_n is the overall course learning outcomes of the course that is related to the program learning outcomes PLO . Increasing the sum of all course topics C_{Ti} will lead to more recommended actions A_R . For example, the information systems program, provided by the College of Computer and Information Sciences at Jouf University, encompasses 12 PLOs that describe the career of students enrolled in this program. The number of topics differs from one course to another and consequently the number of covered course outcomes; e.g., the “Information System Engineering” course which delivers more than 12 topics has 4 CLOs however the “Fundamentals of Information System” course which delivers only 8 topics covers 3 CLOs. At the end of each semester, professors propose actions to enhance the quality of the course he/she taught based on the results of the CLOs assessment process. The number of recommended actions to the “Information System Engineering” course for instance will be greater than the number of actions recommended to the “Fundamentals of Information System” course.

4) ASSESSMENT MARKS

Based on the type of course teaching mechanism presented in Table 1, different assessment marks are applied on courses

TABLE 3. Grade Distribution.

Grade Distribution	Grade Level
A+ & A	High
B+ & B	Above Average
C+ & C	Average
D+ & D	Below Average
F	Fail

based on their category whether they are theoretical or practical courses. These categories are discussed in the following sections:

- *Final Exam*: the final exam marks will be categorized based on whether the course is theoretical or lab. Theoretical courses will have a final exam marks with 60 marks out of 100 marks. The remaining 40 marks will be distributed to midterms and assignments. On practical courses, the final exam mark will be 50 marks out of 100 marks where the remaining 50 marks will be distributed to midterms, lab exams, and assignments.
- *Midterm 1 and 2 Exams*: the first and second midterms will have 15 marks each whether the course is theoretical or practical. This is considered a main issue in all courses while the remaining marks will differ according to the course type.
- *Assignments, Labs, and Projects*: in this category, the assignments such as reports, surveys, and quizzes will be applied on theoretical courses while the labs and projects will be applied on courses that depend on practical applications.

5) GRADE DISTRIBUTION

The grade distribution is based on the total marks of all assessments methods used in the course. As shown in Table 3, the grade distribution can be high, above average, average, below average, and fail.

Recommending suitable actions for improving students learning experiences is based on verifying the student category with a feature binary method. The feature binary method is applied based on the following:

- Calculate the student category feature δ by calculating the value of the mean $\mu + 10$ for each feature. As the success rate in the credit hour mechanism is based on overall percentage of success of 60%. Based on the mean value $\mu + 10$, the category feature is applied based on formula (2):

$$Student \ Category \ \delta = \begin{cases} 0, & \delta < 60 \text{ Fail Student} \\ 1, & \delta \geq 60 \text{ Pass Student} \end{cases} \quad (2)$$

In order to recommend the suitable action for improving the course, the total number of students in each grade distribution is determined.

6) COURSE LEARNING OUTCOMES (CLOs)

Learning Outcomes are clear learning results that we want students to demonstrate at the end of significant learning experiences. Learning outcomes are what learners can actually do with what they know and have learned. They are the tangible application of what has been learned [6]. As stated by [14], learning outcomes are clearly articulated statements that refer to the specific knowledge, practical skills, and areas of professional development, attitudes, or higher-order thinking skills that the instructors expect the students can develop and learn by the end of their learning.

Each course learning outcome CLO_i has different suitable actions that can be applied. The CR calculates the achievement percentage of each learning outcome and based on this value, a set of course actions are recommended to enhance the quality of teaching strategies.

7) TOTAL NUMBER OF QUESTIONS

In the final and midterm exams, each question should cover a specific learning outcome CLO_i from the overall learning outcomes CLO_n of the course and each question must also obtain a specified mark from the exam.

B. DATA PREPROCESSING

Most of the effort of applied machine learning researches are preparing and formulating the data to be in suitable format for machine learning algorithms. The preprocessing of data is highly correlated with the type of problem as in text classifications. In text classification, the problem should be formulated to generate the features from text while in image classification the problem should be formulated to extract the interested features from image. The rest of the study is based on performing different experiments using different machine learning algorithms and selecting the best algorithm suitable for the problem under consideration. The success or failure of these algorithms is based on the nature of dataset which is different from problem to another.

The preprocessing stage in many research papers are based on traditional machine learning algorithms that use single label data [28]. Different classification methods are enhanced to be adapted with more than one class label. These methods are based on two main categories and are discussed as follows:

- **Problem Transformation (PT):** the problem is formulated as a multi-label multi-class problem which is the general case for both multi-label and multi-class. In multi-label, the number of targets is more than one and the target cardinality is less than two (0 or 1) while in multi-class the number of targets equals one and the target cardinality is more than two (such as categorical). In the case of multi-label multi-class or multi-output multi-class, the number of targets is greater than one and the target cardinality is greater than 2. As presented in [29], a multi-label classification was proposed to model dependencies between data labels where a

Monte-Carlo mechanism was applied to increase accuracy. As shown in [30], the multi-label classification was used to detect connections between different multi-label methods for predicting sequential data. In this research, the multi-label multi-class data is transformed into smaller single label data classification. During the formulation, a binary relevance is applied to convert the multi-label problem into single label classification data. Each dataset has all instances of the source dataset with a single corresponding label [31], and [32]. The binary relevance uses each label as a distinct input during the learning mechanism and a binary classification process is associated to each label [33], and [34]. Once the transformation is executed, a binary classification algorithm can be used to predict suitable actions.

- **Model Ranking and Selection (MRS):** the MRS is based on selecting the best algorithm to deal with the problem under consideration. The MRS is based on ranking algorithms based on their performance and then selecting the best algorithms performance.

Based on both PT and MRS, the mechanism of the first method is executed inversely from the second method. In PT, the multi-label dataset is formulated to the proposed algorithms while in the MRS the proposed algorithm is adapted to the multi-label dataset to select the best one. In our proposed paper, advanced algorithms for manipulating multi-label classification datasets are deployed to deal with multi-label datasets.

Let S refers to the input space dataset and $L = \{l_i | i = 1 \dots M\}$ refers to the label set. Let T is the training dataset and $T = \{(I_1, O_1), (I_2, O_2), \dots, (I_n, O_n)\}$ where (I, O) is the input and output space respectively. The main objective of multi-label classification is to search for a connection function f from the input dataset S to the output space 2^L such that each single dataset $s \in S$ must be part of the output space 2^L .

1) DATA FILTERING

The process of data filtering is used to eliminate unnecessary information that will not effect on the overall performance of the recommender system.

In all educational systems, there are different types of students classified based on whether they have continued in the course or not. As shown in formula (3) the student status is defined as follows:

$$\forall ST_i \subset ST_n \text{ where } ST_i \in ST_{IN} || ST_{WD} || ST_{DN} || ST_R \quad (3)$$

where, ST_i is a single student status from the overall students ST_n . Each student ST_i can be classified based on his status on the course under consideration with the following conditions:

ST_{IN} – refers to the incomplete student who did not complete the project at the end of the semester.

ST_{WD} – is the withdrawn student who withdraw the course during the semester

TABLE 4. Action Mapping.

Mapping Value	Action	No of Samples
Action 0	Give less important information at the beginning of each lecture	22
Action 1	Add further case studies	42
Action 2	Improve the course content	23
Action 3	Explain the importance of attending a full lecture	31
Action 4	Add further assignments	39
Action 5	-1	101
Action 6	Add a new assessment method	20
Action 7	Using blended learning	8
Action 8	Recall basic concepts used in the course	18

ST_{DN} – is the denied student who was denied from attending the final examination due to exceeding the specified absence percentage.

ST_R – is the regular student who attends most theoretical and practical lectures.

Regarding the student status, the recommender system will propose different suitable actions for only regular students ST_R who have already finished the course. The data filtering process will also remove all irrelevant features such as ‘Courses ID’, ‘Section’, ‘Semester’, ‘Year’, Department’ that will not affect the prediction of suitable actions

2) HANDLING MISSING VALUES

For handling missing values, all numeric fields that have null values will be converted to 0 while the courses that have null actions will be converted to -1 to differentiate between both null value and null action.

3) MAPPING ACTIONS TO CLASSES

The recommended actions of the proposed recommender system are categorized to nine actions. As presented in Table 4, the nine actions are mapped from 0 to 8 with -1 that refers to “no action”. Each action has a number of samples with a total of 304 samples. The recommended actions were developed from the course reports prepared by professors at end of each semester. Indeed, based on the results of the COs assessment process, each teacher propose actions to enhance the quality of the course he/she taught and therefore improving the overall learning experience. So, after gathering the course reports during several semesters, we studied the proposed actions one by one. Then, we eliminated the actions that cannot be applied (such as “add more credit hours to the course”) and the redundant ones.

Based on the preprocessing step, the final input data to the recommender system will be presented in Table 5. As shown, the class actions are identified for each course learning outcome CLO_i where each class action is used to enhance the quality of teaching strategies based on the percentage

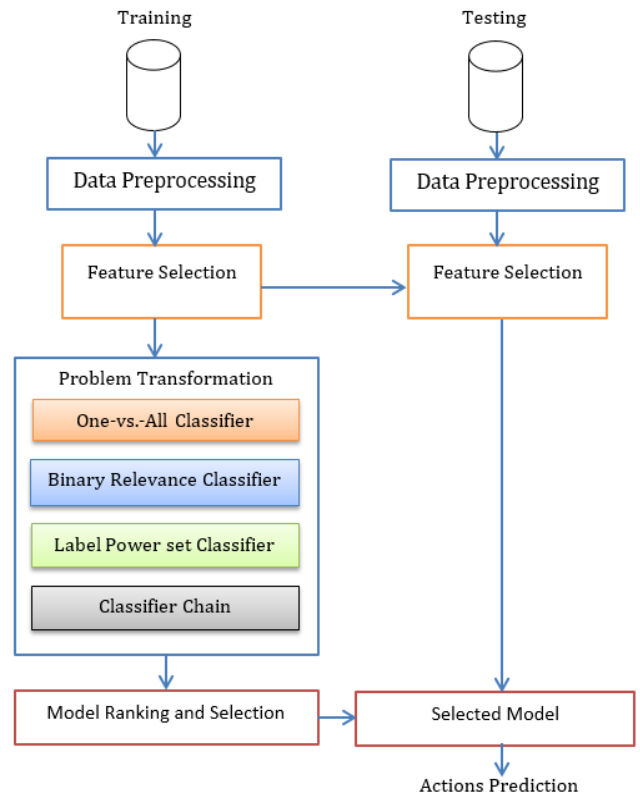


FIGURE 1. Prediction framework.

achieved in each CLO_i . As presented in $CLO_{1,1}$ no actions will be taken due to the high percentage of student achievements which is 85.67%. So, the action value will be -1.

C. PROBLEM TRANSFORMATION

In this section, a brief description of four problem transformation methods and one adaptive method that are applied on five base classifier algorithms. Multi-label learning is considered a supervised learning process into which the training dataset is associated with multi-label multi class binary classification [35] while the traditional methods are based on learning from single-label data. As shown in Figure 1, the overall framework is decomposed into training and testing mechanisms. The training dataset is used for learning by fitting the preprocessed data into the specified problem transformation methods. The testing dataset is independent from the training dataset but if the dataset is already fitted in the training dataset, the features of the testing dataset will be also fitted and selected.

As depicted in Figure 1, a multiple problem transformation is examined with different machine learning algorithms on the training dataset and the trained modeled scored on the test dataset. The MRS will select the best model performance to be deployed as a final model for predictions. In the problem transformation, the multi-label data is transformed into multiple binary classification dataset. As presented in the prediction framework, four problem transformation methods

are applied: OvA, BR, LP, and CC. The fifth method is considered the adaptive classifier that uses ML-KNN classifier. The OvA is considered a heuristic method for using binary classification algorithms [36]. It is based on decomposing the multi-label data into binary classification dataset and a binary classifier is applied to improve predictions. The CC is trained on the input data and then each next classifier is trained on the combined input and all previous classifiers output in the chain.

BR is a straightforward method for handling multi-label classification. The main strategy of BR is to reduce multi-label problem with m labels into m independent binary classification problem [37], [38]. The binary relevance is not dedicated to a specific learning mechanism. Instead, it can be applied in several binary classification algorithms and can be adapted to learn from multi-label multi-class datasets with missing labels [39]. The LP is another problem transformation method into which each label l is used as a class in multi-class datasets [40].

The multi-label is converted to multi-class problem with one classifier trained in all label combinations in the training dataset. CC is considered one of the most effective binary classification methods for handling multi-label multi-class problems. In most methods of multi-class classification, the labels are associated with small number of training datasets that is considered a challenge in most multi-label classification methods due to an imbalance of label dataset [41]. The CC can deal with negative and positive labels with low number of label datasets.

Adaptive algorithms are based on adapting the specified algorithm to the multi-label datasets. The ML – KNN is based on traditional KNN that maximizes posterior estimation to identify all labels in the dataset under consideration. The enhancement of ML – KNN is that it can use smaller training datasets with better efficiency and high performance [42].

IV. EXPERIMENTAL RESULTS

The experiments performed using a data split technique with 70% of training data and 30% for testing dataset. The results were conducted by testing different problem transformation and adaptive techniques such as OvA, BR, LP, CC, and adaptive ML – KNN with different classification algorithms such as SVC [43], and [44], LR [45], RF [46], Gaussian NB [47], and DT [48].

To measure all the metrics of the experimental results, we must include all classes of the datasets based on the following methodology:

S - is the input space dataset.

L - is the number of labels in the label set.

C - is the classifier used in the prediction process.

L_P - is the labels predicted by the classifier C .

L_I - is the true labels associated with the predicted labels L_P .

Δ - is the difference between two input datasets that is exists in either of the sets but not both. To measure the precision of the datasets, the number of the TP_i predictions is divided

by the total number of TP_i and FP_i predictions as shown in formula (4):

$$Precision = \frac{1}{S} \sum_{i=1}^S \frac{TP_i}{TP_i + FP_i} = \frac{1}{S} \sum_{i=1}^S \frac{|C(L_P) \cap L_I|}{|C(L_P)|} \quad (4)$$

The recall is measured with the ratio between the TP_i predictions divided by the total number of TP_i and FN_i predictions as shown in formula (5):

$$Recall = \frac{1}{S} \sum_{i=1}^S \frac{TP_i}{TP_i + FN_i} = \frac{1}{S} \sum_{i=1}^S \frac{|C(L_P) \cap L_I|}{|L_I|} \quad (5)$$

The F1-measure ratio is identified using the mean of both precision and recall as shown in formula (6):

$$\begin{aligned} F1 - Measure &= \frac{1}{S} \sum_{i=1}^S \frac{2 \times Precision \times Recall}{Precision + Recall} \\ &= \frac{1}{S} \sum_{i=1}^S \frac{2 \times |C(L_P) \cap L_I|}{|C(L_P) + L_I|} \end{aligned} \quad (6)$$

The input space dataset S is used in the four problem transformation classifiers: OvA, BR, LP, and CC and is applied also with the adaptive ML KNN adaptive algorithm. Each classifier is applied to five classification algorithms: SV, LR, RF, Gaussian NB, and DT. The best three classifiers for each method are explained in the following sections and the overall experimental results are obtained with the top 10 classifier results.

A. OVA CLASSIFIER

The OvA classifier is applied on different classification algorithms. When the OvA is used with SVC, the precision, recall, and F1-measures achieves good results on all actions of teaching strategies especially on action 7 which is “Using blended learning” and action 1 which is “Add further case studies”. As shown in Figure 2, the precision, recall, and F1-measure of action 7 and action 1 achieved 100%, 67%, 80% and 92%, 85%, 88% respectively.

As shown in Figure 3, the OvA is used with the LR algorithm. The precision of both action 2 and 7 achieved 100%. The recall rate of action 5 and action 1 achieved 92% and 85% respectively while F1-measure for both action 1 and 5 achieved 88% and 87% respectively.

The OvA is applied with DT classification algorithm as shown in Figure 4. The precision, recall, and F1-measure achieved high results on most actions. The precision of action 8 which is “Recall basic concepts used in the course” achieved 100%. The recall of actions 3 and 5 achieved 82% and 81% respectively while the F1-measure of actions 8 and 1 achieved 88% and 80% respectively.

B. BR CLASSIFIER

When the BR method is used on multi-class multi-label datasets with SVC, the same results are achieved like the OvA method with SVC where actions 7 and 1 will have the

TABLE 5. Preprocessing Data.

# of Students	# of Topics	Final Exam	Mid 1	Mid 2	Assessment	Lab	Project	A+	A	B+	...	CLO 1.1 %	...	CLO 4.2 %	Class 1	...
17	13	50	15	15	0	10	10	2	1	0	...	64.10	...	0	Improve the course content	...
19	8	50	15	15	0	10	10	3	0	0	...	66.78	...	90	Explain the importance of attending a full lecture	...
13	7	60	15	15	10	0	0	2	1	1	...	0.00	...	33.3	Add further case studies	...
12	8	50	15	15	0	10	10	2	1	1	...	85.67	...	0	-1	...
19	8	50	15	15	0	10	10	3	2	2	...	0.00	...	62.98	Add a new assessment method	...
...

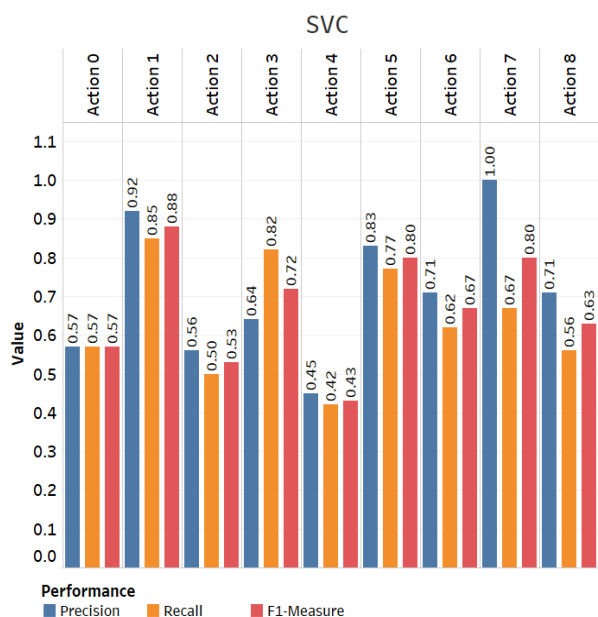


FIGURE 2. Ova - SVC.

best results on precision, recall, and F1-measure. As shown in Figure 5, the recall of actions 3 and 5 is better than the recall of action 7.

As shown in Figure 6, using BR with LR, the precision on actions 7 and 1 achieved high results with 100% and 92% respectively. Action 1 achieved the results on both recall and F1-measure with 92%.

The BR is applied with decision tree classification algorithm as shown in Figure 7. The precision, recall, and F1-measure achieved high results on most actions. The precision on actions 2, 7, and 8 obtained 100% while the recall of actions 5 achieved 81%. The F1-measure of actions 8 obtained 88% while both actions 1 and 3 achieved 80%.

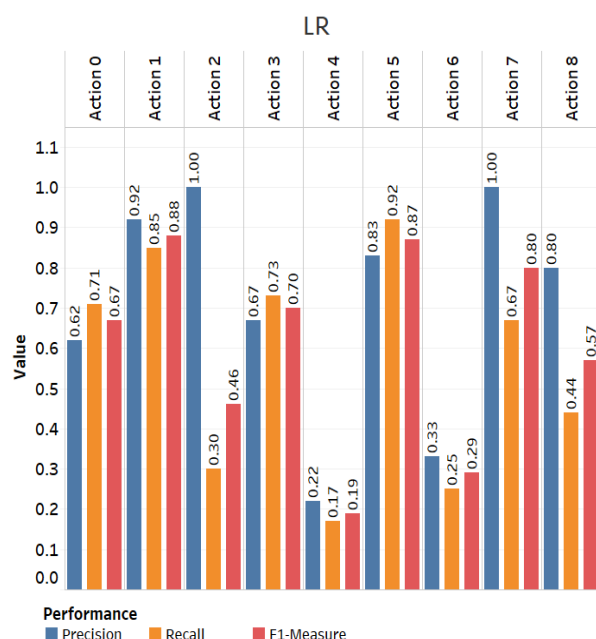


FIGURE 3. Ova - LR.

C. LP CLASSIFIER

Applying LP classifier on multi-class multi-label datasets with SVC achieved high results when compared with the Ova and BR classifiers. As shown in Figure 8, the precision on actions 2, 4, 6, and 7 achieved 100% while the recall rate achieved 96% and 91% on actions 5 and 3 respectively. The F1-measure obtained 88% and 86% on actions 5 and 0 respectively.

As presented in Figure 9, when the Gaussian NB classification algorithm was applied, high results are obtained when compared to Gaussian NB using Ova and (BR classifiers). The precision of action 2, 6, 7, and 8 achieved 100% while the recall of action 5 achieved 92%. The F1-measure achieved good results action 5 with a rate of 81%.

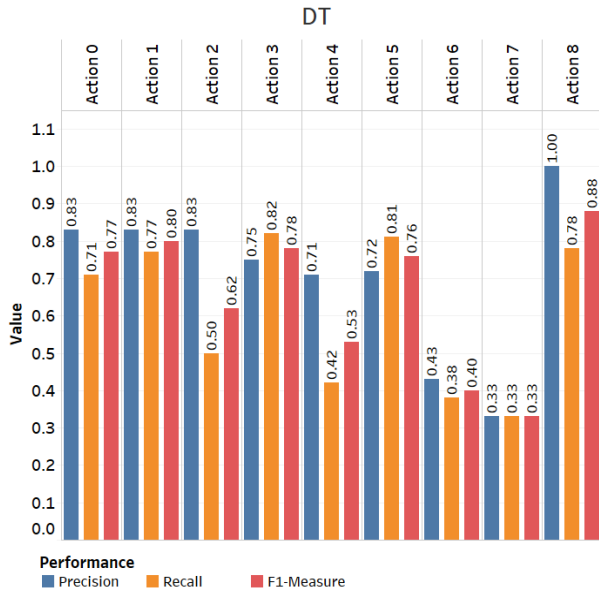


FIGURE 4. Ova - DT.

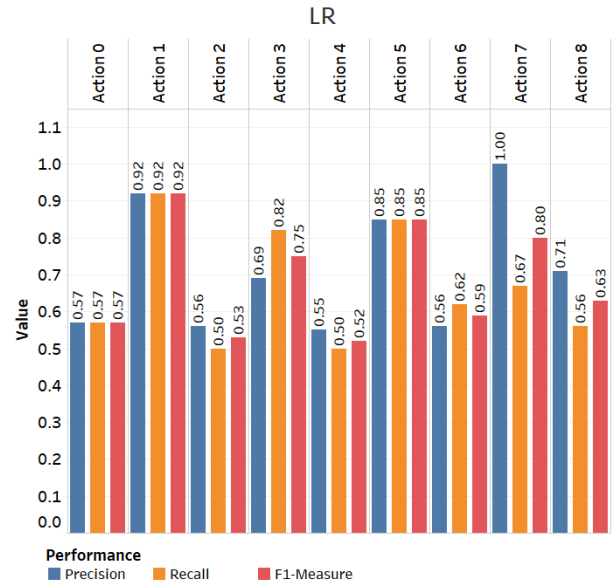


FIGURE 6. BR - LR.

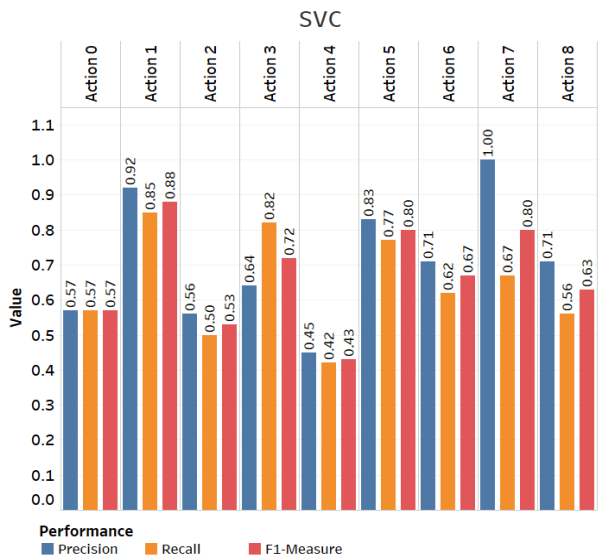


FIGURE 5. BR - SVC.

The LP is applied with DT classification algorithm as shown in Figure 10. The precision on actions 1 and 7 achieved 100% while the recall of action 5 achieved 96%. The F1-measure of actions 1 and 5 obtained 92% and 89% respectively.

D. CC CLASSIFIER

As explained in Figure 11, when the CC is applied on the same dataset, the results with SVC were somewhat similar to the OvA in that actions 7 and 1 achieved precision 100% and 92% respectively in addition to enhancements of precision on actions 5 and 6 with 83%. The recall and F1-measure of action 1 achieved 85% and 88% respectively.

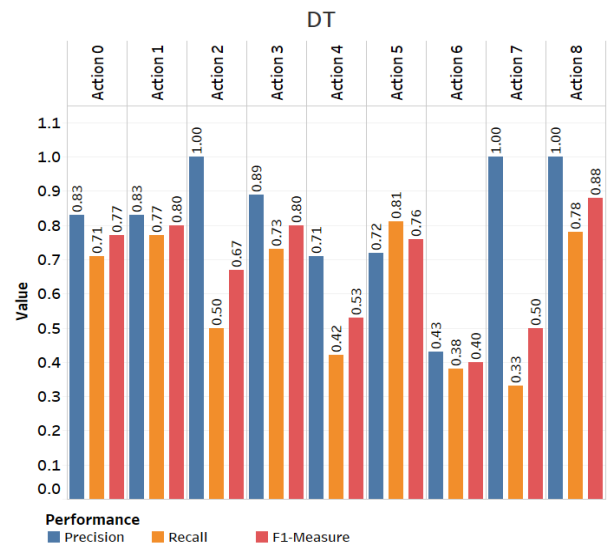


FIGURE 7. BR - DT.

As shown in Figure 12, using CC with LR, the precision on actions 7 and 2 achieved high results with 100%. Action 5 obtained 92% on recall while action 1 obtained 88% on F1-measure.

The CC is applied with DT classification algorithm as shown in Figure 13. The precision on actions 7 and 8 achieved 100% while the recall of action 5 achieved 85%. The F1-measure of both actions 1 and 3 obtained 80%.

E. ML- KNN ADAPTIVE CLASSIFIER

As explained before, adaptive methods are considered as the inverse of problem transformation methods as the algorithm is changed to be adapted to the multi-class multi-label

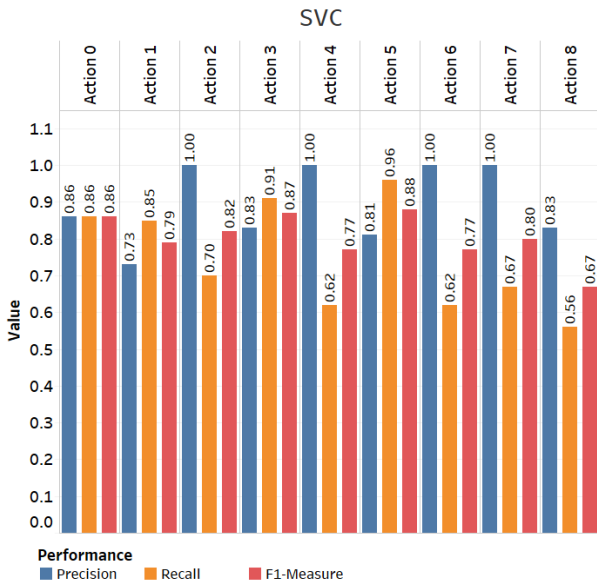


FIGURE 8. LP - SVC.

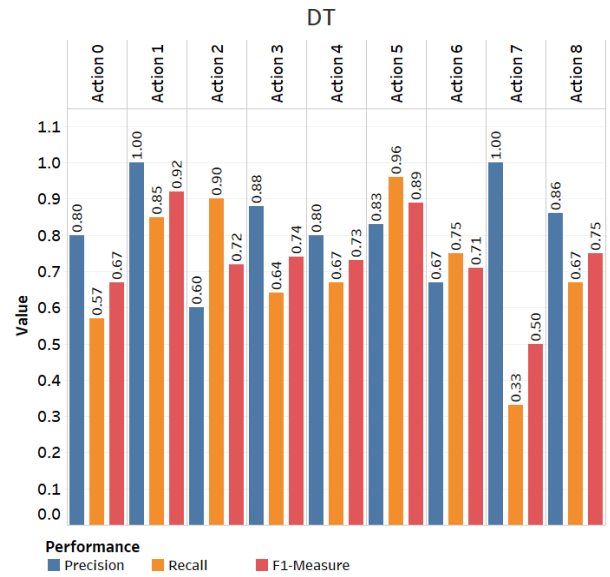


FIGURE 10. LP - DT.

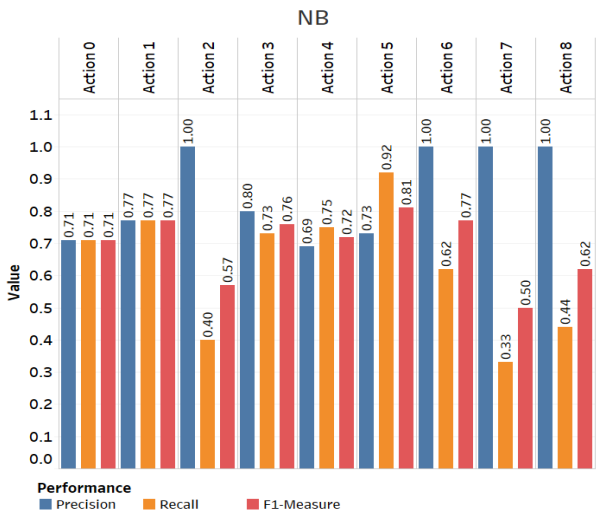


FIGURE 9. LP - Gaussian NB.

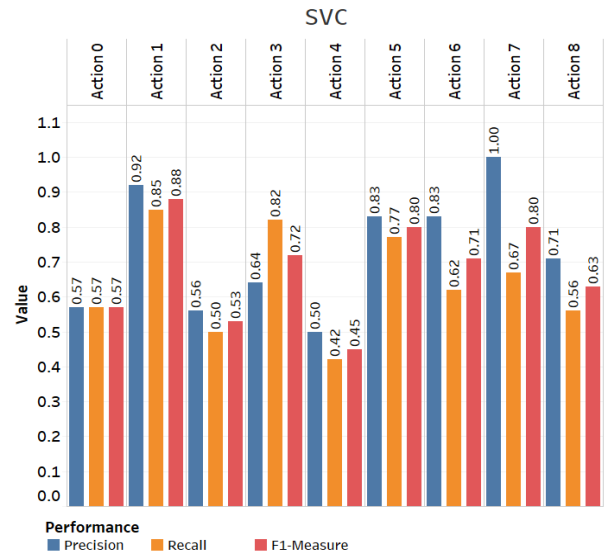


FIGURE 11. CC - SVC.

datasets. ML-KNN algorithm is one of the best adapted algorithms [26]. By applying the ML-KNN, high results are obtained on precision, recall, and F1-measure. As shown in Figure 14, actions 2, 7, and 8 obtained 100% on precision while action 5 achieved 96% on recall. The F1-measure achieved high results on actions 1 and 3 with 92% and 91% respectively.

Based on the previous performance metrics, the random forest with n estimators = 50, 100, 150, and 200 obtained poor results in precision, recall, F1-measure and hamming loss. So, they are not included in the previous figures and included only in Table 6 that obtains the overall experimental results.

F. PERFORMANCE AND HAMMING LOSS

Based on the metrics that are previously defined, the overall precision, recall, and F1-measure for all classifiers are measured. In addition, the hamming loss, macro average, and micro average are identified to verify the recommender system performance. The hamming loss is measured also to determine the relevant labels that are misclassified labels or irrelevant labels that are classified.

$$Hamming Loss = \frac{1}{S} \sum_{i=1}^S \frac{1}{L} |C(L_P) \Delta L_I| \quad (7)$$

The macro average is measured based on the average precision, recall, and F1-measure of all actions as presented in

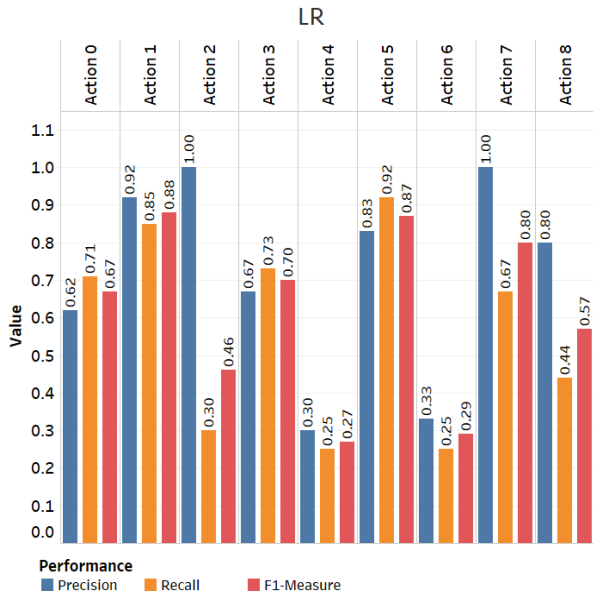


FIGURE 12. CC – LR.

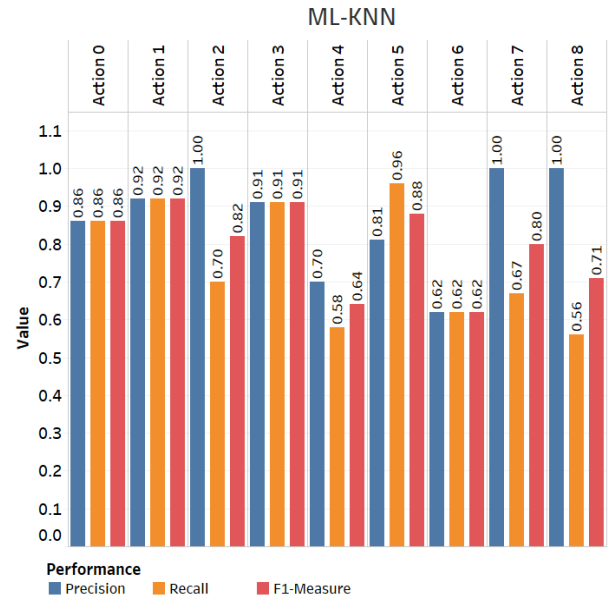


FIGURE 14. ML-KNN adaptive classifier.

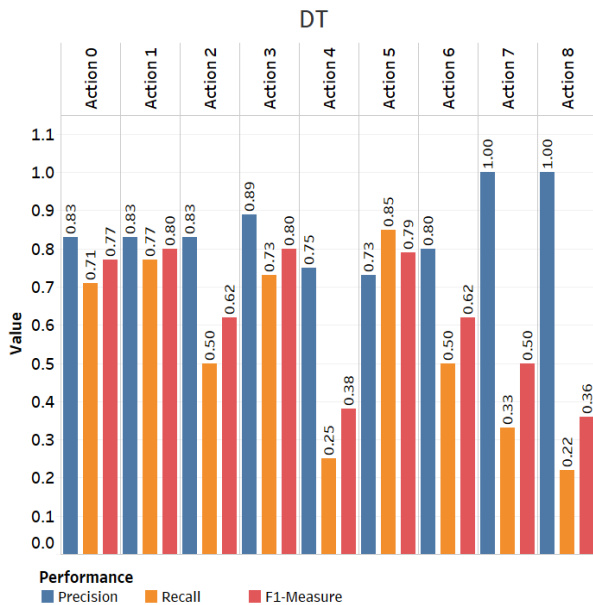


FIGURE 13. CC – DT.

formula (8):

$$Macro\ AVG = \frac{1}{L} \sum_{j=1}^L K(TP_j + FP_j + FN_j) \quad (8)$$

where, K is the calculated precision, recall, or F1-measure of all actions and L is the total number of labels.

The micro average is measured by the sum of all individual true positive, false positive and false negative and aggregating

all classes to compute the average as shown in formula (9):

$$Micro\ AVG = K \left(\sum_{j=1}^L TP_j, \sum_{j=1}^L FP_j, \sum_{j=1}^L FN_j \right) \quad (9)$$

As shown in Table 6, the overall performance of the proposed recommender system is presented.

Based on the performance metrics presented in Table 6, the hamming loss is measured by determining the relevant labels that wrongly classified or the number of irrelevant labels that are correctly classified. The lower is the value of the hamming loss; the higher is the efficiency of the predicted labels.

The adaptive ML-KNN achieved the best hamming loss with $0.099715 \approx 9.97\%$ while the LP with SVC achieved $0.102564 \approx 10.256\%$ Both LP-SVC and ML-KNN achieved F1-measure with 0.754202 and 0.753489 respectively.

Hamming loss and F1-measure are the most important metrics in our proposed recommender system. Hamming loss calculates the fraction of labels that are incorrectly predicted. So, in multi-label problem, the loss value does not reflect the relevant set of labels as the hamming loss evaluates the individual labels.

As a result, in our proposed recommender system, it is more convenient to predict one or more relevant actions for the course not to predict the full relevant actions. As the predictions in this research mainly focus on predicting one or more actions for improving courses, the predictions of the proposed recommender system will succeed when they can recommend full or partial actions. For example, when the actual actions are 3 and the recommender system predicts one or two actions correctly from the actual actions recommended in course report, it is considered as a successful prediction

TABLE 6. Overall Performance Metrics.

	Classifier	Hamming Loss	Macro AVG			Micro AVG		
			Precision	Recall	F1	Precision	Recall	F1
OvA	SVC	0.1709	0.711	0.642	0.670	0.71	0.67	0.69
	LR	0.1823	0.710	0.560	0.603	0.71	0.62	0.66
	RF 50	0.2279	0.297	0.178	0.191	0.71	0.32	0.44
	RF 100	0.2308	0.297	0.161	0.169	0.7	0.31	0.43
	RF 150	0.2279	0.408	0.172	0.189	0.71	0.32	0.44
	RF 200	0.2308	0.297	0.161	0.169	0.7	0.31	0.43
	Gaussian NB	0.3704	0.509	0.757	0.523	0.41	0.73	0.53
	DT	0.1481	0.714	0.613	0.652	0.74	0.67	0.7
BR	SVC	0.1709	0.711	0.642	0.669	<u>0.83</u>	<u>0.8</u>	<u>0.81</u>
	LR	0.1595	0.712	0.668	0.684	0.72	0.74	0.73
	RF 50	0.2279	0.297	0.178	0.191	0.66	0.4	0.5
	RF 100	0.2308	0.297	0.161	0.169	0.63	0.39	0.48
	RF 150	0.2279	0.408	0.172	0.189	0.64	0.39	0.49
	RF 200	0.2308	0.297	0.161	0.169	0.67	0.41	0.51
	Gaussian NB	0.3704	0.509	0.757	0.523	0.78	0.71	0.74
	DT	0.1624	0.823	0.603	0.679	<u>0.8</u>	<u>0.78</u>	<u>0.79</u>
LP	SVC	<u>0.1026</u>	<u>0.896</u>	0.754	0.803	0.71	0.67	0.69
	LR	0.1538	0.732	0.704	0.704	0.72	0.71	0.71
	RF 50	0.2279	0.287	0.238	0.243	0.71	0.32	0.44
	RF 100	0.2365	0.246	0.227	0.222	0.7	0.31	0.43
	RF 150	0.2336	0.272	0.228	0.229	0.71	0.32	0.44
	RF 200	0.2222	0.288	0.246	0.251	0.7	0.31	0.43
	Gaussian NB	0.1396	0.856	0.632	0.692	0.41	0.73	0.53
	DT	0.1168	0.827	0.704	0.737	0.78	0.66	0.71
CC	SVC	0.1681	0.729	0.642	0.677	0.73	0.67	0.69
	LR	0.1852	0.719	0.569	0.612	0.71	0.63	0.67
	RF 50	0.2308	0.298	0.178	0.191	0.73	0.32	0.45
	RF 100	0.2336	0.408	0.170	0.186	0.71	0.32	0.44
	RF 150	0.2336	0.297	0.161	0.169	0.7	0.31	0.43
	RF 200	0.2365	0.297	0.161	0.169	0.7	0.31	0.43
	Gaussian NB	0.3504	0.428	0.579	0.434	0.42	0.6	0.49
	DT	0.1652	0.851	0.540	0.627	0.8	0.61	0.69
Adaptive ML - KNN		<u>0.0997</u>	0.869	0.753	0.796	<u>0.84</u>	<u>0.8</u>	<u>0.82</u>

case. This is why Hamming distance is the best to measure the success of this problem because it can count the successful partial predicted actions.

The F1-measure calculates the weighted average of the precision and recall. So, it is more convenient to use F1-measure than accuracy due to the unbalanced class distribution of dataset as in our case under consideration.

Regarding the precision of the predicted labels, the LP with SVC achieved $\approx 89.6\%$ while ML-KNN achieved $\approx 86.9\%$. The LP with Gaussian NB classifier achieved $\approx 85.59\%$.

As shown in Table 6, the best classifier results with different performance metrics are presented.

The machine learning use the default parameters of Gaussian NB and DT Classifier while the RF used with different estimators from 50 to 200 and SVC algorithm used with ‘linear’ kernel and LR use Stochastic Average Gradient (sag) as solver.

G. FEATURES IMPORTANCE

An analysis of features importance for each recommended action is presented. These important features will be used

TABLE 7. Best Classifier Results.

Algorithm	Hamming Loss	F1-measure	Precision	Recall
Adaptive ML- KNN	<u>0.0997</u>	0.7956	0.8689	0.7533
LP- SVC	0.1026	<u>0.8033</u>	<u>0.8956</u>	<u>0.7542</u>
LP- DT	0.1168	0.7367	0.8267	0.7044
LP- Gaussian NB	0.1396	0.6922	0.8556	0.6318
OVA- DT	0.1481	0.6522	0.7144	0.6133
LP- LR	0.1538	0.7044	0.7322	0.7044
BR- LR	0.1595	0.6844	0.7122	0.6678
BR- DT	0.1624	0.6789	0.8233	0.6033
CC- DT	0.1652	0.6267	0.8511	0.5400
CC-SVC	0.1681	0.6767	0.7289	0.6422

by domain expert to understand important features without relying on the complexity of machine learning or problem transformation methods. Some of machine learning algorithms support extracting these important features such as LR, SVC, and RF.

Table 7 explores the best classifier results based on the parameters: hamming loss, F1-measure, precision, and recall. The adaptive ML-KNN classifier recorded the best hamming loss with $0.099715 \approx 9.97\%$. The LP-SVC recorded a hamming loss with $0.1026 \approx 10.26\%$ while LP-DT recorded a hamming loss of $0.1168 \approx 11.68\%$. As shown in Table 7, the hamming loss is increasing until it reaches $0.1681 \approx 16.81\%$ with CC-SVC.

The LP-SVC recorded the best F1-measure with $0.8033 \approx 80.33\%$ while the adaptive ML-KNN classifier achieved a close result to LP-SVC with $0.7956 \approx 79.56\%$. The remaining classifiers showed relatively low F1-measure starting from LP-DT to CC-SVC. The highest precision and recall are recorded using LP-SVC with 89.56% and 75.42% respectively while the adaptive ML-KNN recorded 86.89% and 75.33% respectively. Based on the results presented in Table 6, the best results are stated in both adaptive ML-KNN and LP-SVC on all parameters.

To extract the feature importance in this research, we selected LP- SVC algorithm which is ranked the second best algorithm in the experimental results as shown in Table 7. As the SVC result is a hyper-plane that separates action classes as best as possible. The coefficient of this algorithm represents this hyper-plane, by giving the coordinates of a support vector which is orthogonal to the separator hyper-plane. So, the absolute value of the SVC coefficient is relative to other values. This gives an indication of how important the feature for the separation. Table 8, 9, and 10 present a sample for the most important features for each recommended actions.

V. ANALYSIS AND DISCUSSION

In this research, we have demonstrated that courses specifications, academic records, and course learning outcomes' assessments can be used for predicting recommended actions

to improve students learning experiences. In this problem, the data was formulated in different formulations: OvA, BR, CC, LP, and Adaptive ML-KNN with different machine learning algorithms.

As a result, we concluded that there are two problem machine learning algorithms that have succeeded in modeling our current problem; namely, the LP-SVC and Adaptive ML-KNN algorithms that give the best F1-measure and Hamming losses as shown in Table 7.

The reason behind the success of LP-SVC to achieve the best F1-measure is that it transforms the multi-label to unique label combinations that treat the problem as a multi-class classification problem. As a result, the best F1-measure is archived while other formulations didn't consider the combinations of multiple labels. On other hand, the success of Adaptive ML-KNN to achieve the best Hamming losses is due to adapting the data to directly perform multi-label classification, rather than transforming the problem.

This can answer the first research question regarding the best method for multi-label classification problem. Regarding the second research question for identifying the best algorithm that can predict the most accurate courses improvement actions, as shown in Table 7, the adaptive ML-KNN and LP-SVC have the lowest hamming loss with 0.0997 and 0.1026 respectively while the adaptive ML-KNN and LP-SVC achieved the best F1-measure with 0.7956 and 0.8033 respectively.

Another interesting conclusion presented in this research is the relationship discovered between each action and the important features that affect these actions either positively or negatively.

As shown in Table 8, when course has "more topics" and has a "project" as one of its assessment methods, this leads to the action "Add further case studies".

This makes sense since more topics means more course complexity that needs more case studies.

While in Table 9, more number of students with "F grade" and little achievement of "CLO 1.4" (i.e. learning knowledge) lead to action "Improve the course content" which also is a logical recommendation since more students' failure

TABLE 8. Important Features for Action: "Add Further Case Studies".

Feature	Important Factor	Explanation
# of Topics	+0.623437	The number of case studies is proportional to the number of case studies used in the course. That's why the feature "# of topics" is very important when recommending "Add further case studies"
Project	+0.387187	When the "project" is used as a teaching and assessment strategy, recommending "Add further case studies" will be very helpful to enhance the course and therefore to the students in order to improve their grades.

TABLE 9. Important Features for Action: "Improve the Course Content".

Feature	Important Factor	Explanation
F	+0.320193	The "F grade" feature is an interesting feature when deciding to improve the course content. As a matter of fact, if the majority of the enrolled students fail, the most obvious recommended action is to reconsider the proposed content and improve it.
CLO 1.4	-0.143762	"The CLO 1.4" belongs to the knowledge outcomes. "Knowledge" means the outcome of the assimilation of information through learning. Knowledge is the body of facts, principles, theories and practices that is related to a field of work or study. That's why; recommending improving the course content is very interesting to the instructor when the knowledge outcome is not well achieved.

TABLE 10. Important Features for Action: "No Action Required".

Feature	Important Factor	Explanation
A+	+0.180429	The "A+ grade" feature is an interesting feature when deciding to do nothing as recommended action. Indeed, if the majority of the enrolled students succeed with A+, it means all teaching and assessment strategies used in the course are convenient and there is no need to modify them.
F	-0.200949	The "F grade" feature is also an interesting feature when deciding that no recommended actions are required. In fact, the "F grade" is inversely proportional to the action required to enhance a course. If the minority of the enrolled students fails, it most likely means all teaching and assessment strategies used in the course are convenient and there is no need to modify them.

and less achievement of knowledge mean that there is crucial needs for improving the course content to overcome these defects. This answer the third research question of which features are selected to be the most relevant predictors for courses improvement actions.

Finally, as shown in Table 10, more students with A+ grade and fewer students with F grade lead to action "No action required" to improve the course as our goal aims to achieve the best students' degrees and less student failure.

We acknowledge two important limitations of this research. First, the dataset collected from course reports is small; since this is the only available data that can be acquired at this period to the best of our knowledge. Second, the data was collected only from the College of Computer and Information Sciences, Jouf University, Kingdom of Saudi Arabia (KSA). Directions for future research are, thus, based on the identified limitations. Further data can be collected from other colleges and universities in future work.

VI. CONCLUSION AND FUTURE DIRECTIONS

This paper proposed a recommender system for predicting the suitable actions that can be proposed by faculty members' to enhance the quality of courses they teach and therefore the overall educational program. The recommended actions will be based on courses specifications, academic records, and course learning outcomes' assessments. In this proposed work, five machine learning algorithms are used for predicting suitable actions. Four methods are classified as problem transformation methods such as OvA, BR, LP, and CC while and the remaining method is the adaptive ML-KNN. The five machine learning methods are applied with different classifiers such as SVC, LR, RF, Gaussian NB, and DT. The performance metrics were measured based on the hamming loss, F1-measure, precision, and recall for all classifiers. The experimental results stated that the adaptive ML-KNN obtained the lowest hamming loss while LP-SVC obtained the best F1-measure. Further investigation is required and more work needs to be executed to enhance teaching strategies and maintaining academic integrity especially on the

college courses after applying online teaching during the epidemic of Covid 19. In addition, the dataset of the recommender system needs to be increase to include course reports from diffident universities in Kingdom of Saudi Arabia (KSA).

REFERENCES

- [1] J. Rao, "Outcome-based education: An outline," *Higher Edu. Future*, vol. 7, pp. 1–4, Jan. 2020, doi: [10.1177/2347631119886418](https://doi.org/10.1177/2347631119886418).
- [2] K. Sudheer, V. V. N. Sujit, N. V. G. Prasad, and K. Ravichand, "A novel method of learning outcome assessment in outcome based education," in *Proc. IEEE 4th Int. Conf. MOOCs, Innov. Technol. Edu. (MITE)*, Dec. 2016, pp. 328–331, doi: [10.1109/MITE.2016.071](https://doi.org/10.1109/MITE.2016.071).
- [3] M. R. Jadhav, A. B. Kakade, S. R. Jagtap, and M. S. Patil, "Impact assessment of outcome based approach in engineering education in India," *Procedia Comput. Sci.*, vol. 172, pp. 791–796, 2020.
- [4] S. Senaratne and A. N. Gunarathne, "Outcome-based education (OBE) in accounting in Sri Lanka: Insights for teacher education," in *Teaching and Teacher Education*. Cham, Switzerland: Springer, 2019, pp. 23–47.
- [5] J. Biggs, "Constructive alignment in University teaching," *HERDSA Rev. Higher Edu.*, vol. 36, no. 3, pp. 5–22, 2014.
- [6] W. G. Spady, *Outcome-Based Education: Critical Issues and Answers*. Arlington, VA, USA: American Association School Administrators, 1994.
- [7] N. Mourtos, "A systematic approach for defining and assessing program educational objectives and outcomes," in *Proc. World Congr. Comput. Sci., Eng., Technol. Educ.*, 2006, pp. 1–5.
- [8] N. Iam-On and T. Boongoen, "Improved student dropout prediction in Thai University using ensemble of mixed-type data clusterings," *Int. J. Mach. Learn. Cybern.*, vol. 8, no. 2, pp. 497–510, Apr. 2017.
- [9] Y. Altujjar, W. Altamimi, I. Al-Turaiki, and M. Al-Razgan, "Predicting critical courses affecting students performance: A case study," *Procedia Comput. Sci.*, vol. 82, pp. 65–71, 2016, doi: [10.1016/j.procs.2016.04.010](https://doi.org/10.1016/j.procs.2016.04.010).
- [10] B. Bakhshinategh, O. R. Zaiane, S. ElAtia, and D. Ipperciel, "Educational data mining applications and tasks: A survey of the last 10 years," *Edu. Inf. Technol.*, vol. 23, no. 1, pp. 537–553, Jan. 2018.
- [11] C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," *WIREs Data Mining Knowl. Discovery*, vol. 10, no. 3, p. e1355, May 2020.
- [12] S. Zhong-Mei, Q. Qiong-Fei, and F. Lu-Qi, "Educational data mining and analyzing of student learning outcome from the perspective of learning experience," in *Proc. Int. Conf. Educ. Data Mining*, 2014, pp. 359–360.
- [13] R. Crues, "Untangling the program name versus the curriculum: An investigation of titles and curriculum content," in *Proc. Int. Conf. Educ. Data Mining*, 2017, pp. 366–367.
- [14] L. Suskie, "Using assessment results to inform teaching practice and promote," in *Assessment, Learning and Judgement in Higher Education*. Dordrecht, The Netherlands: Springer, 2009, pp. 1–20.
- [15] C. A. Mertler, *The Data-Driven Classroom: How do I Use Student Data to Improve my Instruction? (ASCD Arias)*. Alexandria, VA, USA: ASCD, 2014.
- [16] M. W. Rodrigues, S. Isotani, and L. E. Zárate, "Educational data mining: A review of evaluation process in the e-learning," *Telematics Informat.*, vol. 35, no. 6, pp. 1701–1717, Sep. 2018.
- [17] M. Ezz and A. Elshenawy, "Adaptive recommendation system using machine learning algorithms for predicting student's best academic program," *Edu. Inf. Technol.*, vol. 25, no. 4, pp. 2733–2746, Jul. 2020, doi: [10.1007/s10639-019-10049-7](https://doi.org/10.1007/s10639-019-10049-7).
- [18] C. Romero, J. Romero, and S. Ventura, "A survey on pre-processing educational data," in *Educational Data Mining*. Cham, Switzerland: Springer, 2014, pp. 29–64.
- [19] S. Slater, S. Joksimovic, V. Kovanovic, R. S. Baker, and D. Gasevic, "Tools for educational data mining: A review," *J. Educ. Behav. Statist.*, vol. 42, no. 1, pp. 85–106, Feb. 2017.
- [20] M. Ezz, "Advisory system for student enrollment in University based on variety of machine learning algorithms," *Int. J. Comput.*, vol. 4, no. 2, pp. 1–12, 2015.
- [21] S. A. Naser, I. Zaqout, M. A. Ghosh, R. Atallah, and E. Alajrami, "Predicting student performance using artificial neural network: In the faculty of engineering and information technology," *Int. J. Hybrid Inf. Technol.*, vol. 8, no. 2, pp. 221–228, Feb. 2015.
- [22] M. Agaoglu, "Predicting instructor performance using data mining techniques in higher education," *IEEE Access*, vol. 4, pp. 2379–2387, 2016, doi: [10.1109/ACCESS.2016.2568756](https://doi.org/10.1109/ACCESS.2016.2568756).
- [23] X. Zhang, G. Sun, Y. Pan, H. Sun, Y. He, and J. Tan, "Students performance modeling based on behavior pattern," *J. Ambient Intell. Humanized Comput.*, vol. 9, no. 5, pp. 1659–1670, Oct. 2018.
- [24] E. Fernandes, M. Holanda, M. Victorino, V. Borges, R. Carvalho, G. V. Erven, "Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil," *J. Business Res.*, vol. 94, pp. 335–343, Jan. 2019, doi: [10.1016/j.jbusres.2018.02.012](https://doi.org/10.1016/j.jbusres.2018.02.012).
- [25] S. Wan and Z. Niu, "A hybrid E-learning recommendation approach based on learners' influence propagation," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 5, pp. 827–840, May 2020, doi: [10.1109/TKDE.2019.2895033](https://doi.org/10.1109/TKDE.2019.2895033).
- [26] A. Muhammad, A. Shaikh, Q. N. Naveed, and M. R. N. Qureshi, "Factors affecting academic integrity in E-learning of Saudi Arabian Universities. An investigation using Delphi and AHP," *IEEE Access*, vol. 8, pp. 16259–16268, 2020, doi: [10.1109/ACCESS.2020.2967499](https://doi.org/10.1109/ACCESS.2020.2967499).
- [27] W. M. Al-Rahmi, N. Alias, M. S. Othman, A. I. Alzahrani, O. Alfarraj, A. A. Saged, and N. S. Abdul Rahman, "Use of E-learning by University students in Malaysian higher educational institutions: A case in Universiti teknologi Malaysia," *IEEE Access*, vol. 6, pp. 14268–14276, 2018, doi: [10.1109/ACCESS.2018.2802325](https://doi.org/10.1109/ACCESS.2018.2802325).
- [28] X. Zheng, P. Li, Z. Chu, and X. Hu, "A survey on multi-label data stream classification," *IEEE Access*, vol. 8, pp. 1249–1275, 2020, doi: [10.1109/ACCESS.2019.2962059](https://doi.org/10.1109/ACCESS.2019.2962059).
- [29] J. Read, L. Martino, and D. Luengo, "Efficient Monte Carlo optimization for multi-label classifier chains," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 3457–3461, doi: [10.1109/ICASSP.2013.6638300](https://doi.org/10.1109/ICASSP.2013.6638300).
- [30] J. Read, L. Martino, and J. Hollmén, "Multi-label methods for prediction with sequential data," *Pattern Recognit.*, vol. 63, pp. 45–55, Mar. 2017, doi: [10.1016/j.patcog.2016.09.015](https://doi.org/10.1016/j.patcog.2016.09.015).
- [31] J. Palmer, V. Sheng, T. Atkison, and B. Chen, "Classification on grade, price, and region with multi-label and multi-target methods in wineinformatics," *Big Data Mining Analytics*, vol. 3, no. 1, pp. 1–12, Mar. 2020, doi: [10.26599/BDMA.2019.9020014](https://doi.org/10.26599/BDMA.2019.9020014).
- [32] T. T. Nguyen, M. T. Dang, A. V. Luong, A. W.-C. Liew, T. Liang, and J. McCall, "Multi-label classification via incremental clustering on an evolving data stream," *Pattern Recognit.*, vol. 95, pp. 96–113, Nov. 2019, doi: [10.1016/j.patcog.2019.06.001](https://doi.org/10.1016/j.patcog.2019.06.001).
- [33] A. Kumar, S. Vembu, A. K. Menon, and C. Elkan, "Beam search algorithms for multilabel learning," *Mach. Learn.*, vol. 92, no. 1, pp. 65–89, Jul. 2013.
- [34] J. Read, L. Martino, and D. Luengo, "Efficient Monte Carlo methods for multi-dimensional learning with classifier chains," *Pattern Recognit.*, vol. 47, no. 3, pp. 1535–1546, Mar. 2014, doi: [10.1016/j.patcog.2013.10.006](https://doi.org/10.1016/j.patcog.2013.10.006).
- [35] E. K. Y. Yapp, X. Li, W. F. Lu, and P. S. Tan, "Comparison of base classifiers for multi-label learning," *Neurocomputing*, vol. 394, pp. 51–60, Jun. 2020, doi: [10.1016/j.neucom.2020.01.102](https://doi.org/10.1016/j.neucom.2020.01.102).
- [36] J. Ramírez, J. M. Górriz, A. Ortiz, F. J. Martínez-Murcia, F. Segovia, D. Salas-Gonzalez, D. Castillo-Barnes, I. A. Illán, and C. G. Puntonet, "Ensemble of random forests one vs. Rest classifiers for MCI and AD prediction using ANOVA cortical and subcortical feature selection and partial least squares," *J. Neurosci. Methods*, vol. 302, pp. 47–57, May 2018.
- [37] E. Montañes, R. Senge, J. Barranquero, J. R. Quevedo, J. J. D. Coz, and E. Hüllermeier, "Dependent binary relevance models for multi-label classification," *Pattern Recognit.*, vol. 47, no. 3, pp. 1494–1508, Mar. 2014, doi: [10.1016/j.patcog.2013.09.029](https://doi.org/10.1016/j.patcog.2013.09.029).
- [38] G. Wu, Y. Tian, and C. Zhang, "A unified framework implementing linear binary relevance for multi-label learning," *Neurocomputing*, vol. 289, pp. 86–100, May 2018, doi: [10.1016/j.neucom.2018.02.010](https://doi.org/10.1016/j.neucom.2018.02.010).
- [39] M.-L. Zhang, Y.-K. Li, X.-Y. Liu, and X. Geng, "Binary relevance for multi-label learning: An overview," *Frontiers Comput. Sci.*, vol. 12, no. 2, pp. 191–202, Apr. 2018.
- [40] J. Read, A. Puurula, and A. Bifet, "Multi-label classification with meta-labels," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2014, pp. 941–946, doi: [10.1109/ICDM.2014.38](https://doi.org/10.1109/ICDM.2014.38).
- [41] B. Liu and G. Tsoumakas, "Dealing with class imbalance in classifier chains via random under-sampling," *Knowl.-Based Syst.*, vol. 192, Mar. 2020, Art. no. 105292, doi: [10.1016/j.knosys.2019.105292](https://doi.org/10.1016/j.knosys.2019.105292).

[42] D. Zhu, H. Zhu, X. Liu, H. Li, F. Wang, H. Li, and D. Feng, "CREDO: Efficient and privacy-preserving multi-level medical pre-diagnosis based on ML-kNN," *Inf. Sci.*, vol. 514, pp. 244–262, Apr. 2020, doi: [10.1016/j.ins.2019.11.041](https://doi.org/10.1016/j.ins.2019.11.041).

[43] S. Alam, S. Sonbhadra, S. Agrawal, and P. Nagabhushan, "One-class support vector classifiers: A survey," *Knowl.-Based Syst.*, vol. 196, May 2020, Art. no. Mar. 105754, doi: [10.1016/j.knosys.2020.105754](https://doi.org/10.1016/j.knosys.2020.105754).

[44] R. Gong, M. Chu, Y. Yang, and Y. Feng, "A multi-class classifier based on support vector hyper-spheres for steel plate surface defects," *Chemo-metric Intell. Lab. Syst.*, vol. 188, pp. 70–78, May 2019, doi: [10.1016/j.chemolab.2019.03.010](https://doi.org/10.1016/j.chemolab.2019.03.010).

[45] K. Ghazvini, M. Yousefi, F. Firoozeh, and S. Mansouri, "Predictors of tuberculosis: Application of a logistic regression model," *Gene Rep.*, vol. 17, Dec. 2019, Art. no. 100527, doi: [10.1016/j.genrep.2019.100527](https://doi.org/10.1016/j.genrep.2019.100527).

[46] Y. Li, H. Chi, X. Shao, M. Qi, and B. Xu, "A novel random forest approach for imbalance problem in crime linkage," *Knowl.-Based Syst.*, vol. 195, May 2020, Art. no. 105738, doi: [10.1016/j.knosys.2020.105738](https://doi.org/10.1016/j.knosys.2020.105738).

[47] Q. Wang, L. Pan, J. Shen, and N. Lu, "Power plant data filtering based on Gaussian naive Bayesian classification and prediction error method," in *Proc. Chin. Autom. Congr. (CAC)*, Nov. 2019, pp. 1490–1495, doi: [10.1109/CAC48633.2019.8996997](https://doi.org/10.1109/CAC48633.2019.8996997).

[48] J. Yan, Z. Zhang, K. Lin, F. Yang, and X. Luo, "A hybrid scheme-based one-vs-all decision trees for multi-class classification tasks," *Knowl.-Based Syst.*, vol. 198, Jun. 2020, Art. no. 105922, doi: [10.1016/j.knosys.2020.105922](https://doi.org/10.1016/j.knosys.2020.105922).



NACIM YANES was born in Gabes, Tunisia, in 1981. He received the master's degree in computer science applied to management from the Higher Institute of Management (ISG), Tunisia, and the Ph.D. degree in computer science from the National School of Computer Science (ENSI), Manouba University, Tunisia. He is currently an Assistant Professor with the Higher Institute of Management (ISGGB), University of Gabes, Tunisia, and also an Assistant Professor with the College of Computer and Information Sciences, Jouf University, Saudi Arabia. His current research interests include software reuse, recommenders systems in software engineering, serious games and gamification, and outcome-based education.



AYMAN MOHAMED MOSTAFA (Member, IEEE) received the M.Sc. and Ph.D. degrees in information systems from the Faculty of Computers and Informatics, Zagazig University, Egypt. He is currently an Assistant Professor in with the Faculty of Computers and Informatics, Zagazig University, Egypt, and also an Assistant Professor at with the College of Computer and Information Sciences, Jouf University, Saudi Arabia. He has published more than 25 scientific articles in various national and international journals and conferences. His current research interests include information security, cloud computing, E-business, E-commerce, big data, and data science. He is also an Oracle Certified Associate, Oracle Certified Professional, and EMC Academic Associate in Cloud Infrastructure and Services.



MOHAMED EZZ (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in systems & computers engineering from the Faculty of Engineering, Al-Azhar University. He is currently an Associate Professor with the Faculty of Engineering, Al-Azhar University, and also a Visiting Professor with the College of Computer and Information Sciences, Jouf University. He has published 20 scientific articles in various national and international journals and conferences. He has contributed in more than 16 mega software projects in electronic banking EBPP, EMV, mobile banking, and e-commerce, also CBAP Certified. His current research interests include pattern recognition, applied machine learning, application security, intrusion detection, and semantic web.



SALEH NAIF ALMUYQIL received the Ph.D. degree in computer science from the Faculty of Computing and Digital Technologies, Staffordshire University, U.K. Since 2018, he has been the Chair of the Information Systems Department. He is currently an Assistant Professor with the College of Computer and Information Sciences, Jouf University, Saudi Arabia. He has published scientific articles in various national and international journals and conferences. His current interests include health informatics, knowledge discovery, knowledge management, digital transformation, and data science.

...