Journal of Information Technology Education: Innovations in Practice

An Official Publication of the Informing Science Institute
InformingScience.org

JITEiip.org

Volume 21, 2022

Categorizing Well-Written Course Learning Outcomes Using Machine Learning

Yessine Hadj Kacem*	CES Laboratory, National Engineer School of Sfax, University of Sfax, Tunisia	yessine.hadjkacem@u-gafsa.tn
Safa Alshehri	College of Computer Science, King Khalid University, Abha, Saudi Arabia	441800148@kku.edu.sa
Talal Qaid	College of Computer Science, King Khalid University, Abha, Saudi Arabia	tqaid@kku.edu.sa

* Corresponding author

Abstract

Aim/Purpose	This paper presents a machine learning approach for analyzing Course Learning Outcomes (CLOs). The aim of this study is to find a model that can check whether a CLO is well written or not.
Background	The use of machine learning algorithms has been, since many years, a prominent solution to predict learner performance in Outcome Based Education. However, the CLOs definition is still presenting a big handicap for faculties. There is a lack of supported tools and models that permit to predict whether a CLO is well written or not. Consequently, educators need an expert in quality and education to validate the outcomes of their courses.
Methodology	A novel method named CLOCML (Course Learning Outcome Classification using Machine Learning) is proposed in this paper to develop predictive models for CLOs paraphrasing. A new dataset entitled CLOC (Course Learning Outcomes Classes) for that purpose has been collected and then undergone a preprocessing phase. We compared the performance of 4 models for predicting a CLO classification. Those models are Support Vector Machine (SVM), Random Forest, Naive Bayes and XGBoost.

Accepting Editor Athanassios Jimoyiannis │ Received: April 15, 2022│ Revised: June 25, 2022 │ Accepted: June 27, 2022.
Cite as: Hadj Kacem, Y., Alshehri, S., & Qaid, T. (2022). Categorizing well-written course learning outcomes using machine learning. *Journal of Information Technology Education: Innovations in Practice, 21,* 61-75. https://doi.org/10.28945/4997

Contribution	The application of CLOCML may help faculties to make well-defined CLOs and then correct CLOs' measures in order to improve the quality of education addressed to their students.
Findings	The best classification model was SVM. It was able to detect the CLO class with an accuracy of 83%.
Recommendations for Practitioners	We would recommend both faculties' members and quality reviewers to make an informed decision about the nature of a given course outcome.
Recommendations for Researchers	We would highly endorse that the researchers apply more machine learning models for CLOs of various disciplines and compare between them. We would also recommend that future studies investigate on the importance of the definition of CLOs and its impact on the credibility of Key Performance Indicators (KPIs) values during accreditation process.
Impact on Society	The findings of this study confirm the results of several other researchers who use machine learning in outcome-based education. The definition of right CLOs will help the student to get an idea about the performances that will be measured at the end of a course. Moreover, each faculty can take appropriate actions and suggest suitable recommendations after right performance measures in order to improve the quality of his course.
Future Research	Future research can be improved by using a larger dataset. It could also be improved with deep learning models to reach more accurate results. Indeed, a strategy for checking CLOs overlaps could be integrated.

Keywords	course learning outcomes, paraphrasing, machine learning, classification, outcome based education

INTRODUCTION

Outcome Based Education (OBE) (Rao, 2020; Spady, 1994) has recently known considerable attention in higher education. It is interested in the accomplishments of targets and results. The revision of the curriculum within a specific time is determined by the exit learning outcomes the students should show at the ending of a program or a course. The accrediting bodies, such as Accreditation Board for Engineering and Technology (ABET; https://www.abet.org), and National Center for Academic Accreditation and Evaluation (NCAAE) (NCAAE; https://etec.gov.sa/en/About/Centers/Pages/Accreditation.aspx), require that all programs define the outcomes, communicate them to stakeholders particularly to students and determine their achievement level. The analysis of the outcomes is very useful to prove that the program meets the quality standards and continuous improvement.

The different learning outcomes identify enduring knowledge, competences and values. The outcomes can be defined at two main levels: (a) Program Learning Outcomes (PLOs) that describe what the graduated student from a program should be able to do and (b) Course Learning Outcomes (CLOs) which present the abilities of the students to perform behind the course completion. It should be noted that there is a strong relationship between learning outcomes at a course-level and program core competencies. A set of Key Performance Indicators (KPIs) at program level could be assessed at course level (Gogus, 2012).

While course objective illustrates what a faculty member will cover in a subject, the CLOs' statements are student centered. A CLO statement is formed by three components: an action verb which identifies the performance to be demonstrated, a learning statement presenting the demonstrated learning and a broad statement of the criterion for acceptable performance. CLOs are determined by the teacher in the course specification or syllabus. They must be concise, meaningful, measurable and achievable (Bloom et al., 1956). Many universities provide guidelines for formulating well-defined CLOs and avoiding bad writing practices. For example, in Queen Mary University of London (2014), it is mentioned that the following issues can be faced while defining a CLO:

- Confusion between course objectives and CLOs.
- The CLO sums up the syllabus or is a part of its topics.
- The CLO may be difficult to measure.
- The CLO represents only the content that the students will learn and not how well they will learn it

After measuring the CLOs through different assessments, each teacher suggests several recommended actions in the course report to improve the quality of the courses they teach and thus improve the overall educational program. Therefore, any mistake in the definition of the CLOs will present wrong measures and then will affect the educational process as well as the program quality improvement. In other words, when the design of CLOs is defective, assessments will automatically be depicted less relevant. Hence, well-written CLOs will be well aligned with grade assessments that determine how each learner has met them. Well-defined learning outcomes present obvious information about the achievements related to particular qualifications and boost the comparison of standards between qualifications.

As a result, a characterization system that automatically classifies a given CLO into six classes (*Knowledge*, *Competence*, *Value*, *Not Measurable*, *Not Clear*, and *Not Concise*) will be very helpful to faculty members in preparing the most efficient course outcomes. Actually, a lot of information related to CLOs can be collected from different universities and converted to a suitable form for a better decision-making system.

Hence, the key success factors of OBE start with the definition of relevant course outcomes, learning materials and assessment strategies. The course outcomes' measures will affect the student outcomes. In classical learning settings, department quality committees have to check the CLOs of all courses. Reviewers of accreditation bodies also need to cross check these CLOs to be sure about the credibility of KPIs for the education standards. Assume that a study plan consists of 30 courses. Each course syllabus includes an average of 6 outcomes. Thus, 180 outcomes have to be verified for a curriculum, which is a costly task in terms of time and effort.

Data Mining (DM) and Artificial Intelligence (AI) techniques are being used in a variety of areas such as engineering, medicine, manufacturing, forecast, and education. In particular, much research has already been successfully done using DM techniques for the educational field (Romero et al., 2014; Romero & Ventura, 2020). This discipline called, Educational Data Mining (EDM) aims at improving the student learning performance through the exploration of data from learning context in order to analyze student knowledge, student learning behavior and curriculum planning in a better way (Anand, 2019).

As a result, different techniques of DM and AI can deeply help in developing a framework for checking if a CLO is well written or not. Machine Learning (ML) which is one of the most advanced concepts of AI provides a strategic method for developing automated, complex and objective algorithmic techniques for data analysis. The Machine Learning process (Suthaharan, 2015) begins with data collection from a diversity of resources. Then, the next step consists in fixing the pre-processed data to adjust data-related issues and reduce space size by deleting non-valid data. The performance evaluation of the models comes in the next step. Finally, the model is optimized and improved using new dataset and rules.

In the present paper, we provide a novel method of CLOs characterization. It should be noted that there is no existing dataset containing classified CLOs. Hence, the first contribution consists in collecting data and makes its classification for the training phase. Indeed, this approach predicts if a

CLO is well written or not. As a result, it helps instructors to make real measures of CLOs as well as PLOs. The CLO Categorization will also facilitate the job of quality committee members and reviewers of accreditation bodies during the program revision process.

The remainder of this paper is organized as follows. The Literature Review gives an overview of educational data mining and the existing research works tackling OBE using ML. The developed method is presented in the CLOCML Methodology section. The dataset construction and pre-processing are described, and the methodology is then exposed. The next section shows and discusses the results of predictive models, followed by the conclusion.

LITERATURE REVIEW

This section introduces a general overview of educational data mining and machine learning and then presents some related research in this context.

OVERVIEW OF EDUCATIONAL DATA MINING AND MACHINE LEARNING

Outcome Based Education is one of the most adopted education models in various educational institutions. It focuses on the student-centered method to learning and teaching. The main emphasis is accorded to students' fulfillment level in order to improve their ability of learning and applying skills and values during the learning experience. Recently, the attainment of Outcome-Based Education has started to be analyzed using Educational Data Mining. The EDM field aims at analyzing the enormous volume of student's data to discover educational issues and then take proper actions to improve the courses delivery quality and the student's achievement (Romero & Ventura, 2013). EDM is based on methods and tools derived from Data Mining and Artificial Intelligence. In particular, the application of Machine Learning.in education field has known much more interest in last years. It collects all the methods that enable machines to learn and make suitable predictions from previous observations. ML techniques can be classified into four main types: supervised, unsupervised, semi-supervised, and reinforcement learning techniques (Suthaharan, 2015).

The supervised learning methods are algorithms which connect previous and current dataset with the aid of labeled data to predict output values. Supervised learning methods can be divided into regression and classification types. Regression consists in predicting a specific output value utilizing training data. Classification means labeling the input into two or more classes. When the number of classes is more than 2, it is called multiclass classification. With regards to the performance of different classification technique, there is no unique model that works best in all cases.

When the content of the dataset is not labelled, we talk about unsupervised learning. The latter aims at finding the clusters or patterns hidden in the non-labelled data. Semi-supervised learning techniques combine both previous methods. The reinforcement algorithms work iteratively through observations gathered from interaction as well as decision making to reduce risks and increase the performance of the model.

In the context of EDM, the data can have diverse forms: when it is in the form of words, sentences or paragraphs, the computer needs to understand it. The Natural Language Processing (NLP) is an established field that makes machines able to make sense of human language (Egger & Gokce, 2022). In fact, it is a field of Machine Learning that enables the model to make sense of a given text, extract key words and classify some topics. In sentiment analysis, for example, NLP is used to interpret what users say and then ML algorithms automatically categorize sentiment into positive or negative (Le & Nguyen, 2015).

RELATED WORKS

A large range of well-known frameworks and tools can be employed for Education Data Mining (EDM) research purposes (Khanal et al., 2020; Solak et al., 2020). In particular, several survey papers dealing with student outcomes prediction (Alturki et al., 2022; Bazelais et al., 2018; Ifenthaler & Yau, 2020; Namoun & Alshanqiti, 2021; Yağcı, 2022) for higher education exist in literature. For example, in Yogeswari and Rajermani (2022), 25 research works out of 3389 publications from the year 2015 to 2020 were reviewed. This survey indicates that there is a variety of methods applied in assessing OBE attainment using EDM. It focuses on the type of assessment and tool for measuring the achievement of Course Learning Outcomes, Program Learning Outcomes and Program Educational objectives. In Slater et al. (2017), 40 tools used for extracting data in education were reviewed.

Devine et al. (2011) implemented a software tool called Data Miner for Outcome Based Education (DMOBE). It permits the instructors to easily apply data mining methods for analyzing various key features of their pedagogy. The developed tool allows an educator to identify which outcomes of a course are pertinent to success in a subsequent course and which outcomes within a course have solid influence on the mastery of a specified outcome. DMOBE is based on supervised feature selection to extract relevant outcomes and association rule mining to discover the dependency of a specific outcome on other ones.

In Abu-Naser et al. (2015), a student performance prediction system has been proposed to identify students who are expected to perform well in college, specifically to be successful in studying engineering programs. The artificial neural network model was used for predicting students' performance before they started their sophomore year in engineering studies. The approach is based on a number of factors such as high school degree, results in some first-year subjects including mathematics and electronics, student gender, type of high school whether private or general. The proposed prediction model was tested, and the overall score is 84.6%.

The author of Agaoglu (2016) explored the factors that influence student achievement to improve the quality of the educational system and proposed a predictive model for teacher performance. In fact, the primary goal of the author is to build a rating model that enables predicting factors affecting students' performance.

In Ezz (2015), the author introduced a model that helps students to choose the most appropriate instructor based on several criteria, including obtained grades in various subjects in high school and gender. The proposed registration recommendation system consists of two stages: the training phase and the runtime phase. The training phase takes the previous high school database and the faculty database as input and creates the faculty-student model, while the runtime takes a new student as input and produces as an output a recommendation for that student, whether or not appropriate, join that college.

Gray and Perkins (2019) proposed a descriptive statistic for student attendance and applied machine learning to create a predictive tool that identifies students in need of tutor intervention. A set of experiments were conducted to select a combination of classifier to achieve an accuracy of 97%.

Mahboob et al. (2020) collected data engineering students to form three different clusters to group students according to the worst, average, and best accomplishment of CLOs/PLOs in two distinct engineering courses regularly taught in the first semesters. A data mining technique is used to determine Euclidean distances for measuring the similarities through two clustering techniques: k-means and k-medoids algorithms. The analysis of CLOs/PLOs attainments serves for classifying the students into three categories.

The authors of Alturki and Alturki (2021) use six data mining methods for predicting students' final grades and identifying honorary students during the first 4 semesters of their study journey. Random Forest performed the best accuracy of 92.6% in predicting honorary students. However, Naive Bayes was the best classifier with an accuracy of 85.8% for predicting students' final grade during the third semester.

In Hussain et al. (2022), data of Bachelor students was collected for predicting students' performance using Support Vector Machine and Decision Tree. The accuracy average of the used models was 74%. This study also determines some factors affecting students' academic achievement such as the time spent by students when using social media and the time spent while playing mobile games.

The previous cited works are unfortunately limited to the prediction of students' academic achievement. To the best of our knowledge, there is no technique for checking if a CLO is well written and predicting its classification. Our proposal differs from the mentioned related works since its aim is to develop a model which provides ML support for analyzing course outcomes. The application of such technique may help faculties to make correct measures and then improve the quality of education addressed to their students. On a larger scale, application of ML during the definition of course outcomes could become a powerful method to decrease the evaluation process by program quality committees as well as reviewer in accreditation organisms.

To address these goals, we pose the following research questions (RQ):

- **RQ1**: Can machine learning be used to determine if a course outcome is well written or not?
- **RQ2**: How would machine learning be effective for categorizing well-written course outcomes?

CLOCML METHODOLOGY

The proposed methodology named **CLOCML** (**C**ourse **L**earning **O**utcome **C**lassification using **M**achine **L**earning) on how to develop the supervised machine learning models for prediction of CLO class is depicted step by step in Figure 1. The first step of the proposed methodology consists in collecting a set of CLOs from engineering and computing programs. CLOs classification is done in the next step by quality experts. Data pre-processing is done following the classification step. After that, the training phase is performed through ML techniques. Finally, the results are analyzed, and the best technique is identified.

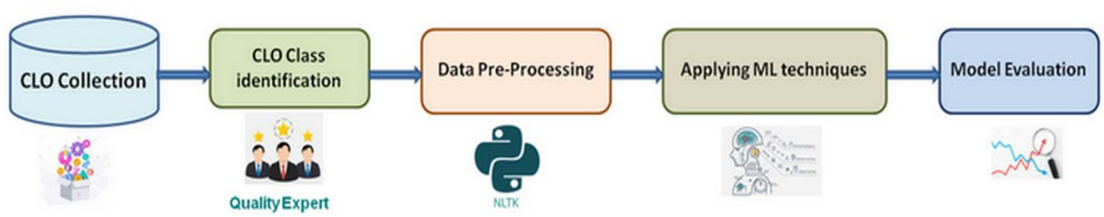

Figure 1: Proposed methodology for course learning classification: CLOCML

CLO COLLECTION AND CLASS IDENTIFICATION

Collecting and processing academic data to solve educational problems is an extremely an arduous task. In our research, with the absence of classified records, we collected the data set named **CLOC** (**C**ourse **L**earning **O**utcomes **C**lasses) from multiple universities. The CLOC data set contains about 3200 rows for engineering and computing domains. The first column represents the CLOs and it will serve as feature. The second column describes its class.

CLOC data set is collected as raw material, whereas its initial form was one column which corresponds to a CLO. In this phase, we classified our data manually by adding the second column and categorized these CLOs. The dataset was classified based on the NCAAE learning domains into *Knowledge*, *Skills*, and *Value* when it is well formulated. In case of non-well-defined CLO, the classification *Not Clear*, *Not Concise*, *Not Measurable* is used.

The annotation of the collected CLOs has been done by content quality expert team at King Khalid University (KKU, https://kku.edu.sa/en). The reason for selecting this university is that many programs in this institution have recently been accredited by ABET and NCAAE. The previously received comments regarding the CLOs paraphrasing during the first revisions in the accreditation process were useful for determining the label of non-well paraphrased CLOs. The non- well written CLOs before the accreditation served for the following data classes: Not Measurable, Not Concise, and Not Clear. Table 1 depicts concrete CLO classes extracted from our dataset.

Table 1: Example of well-defined and non-well-defined collected Course Learning Outcomes

Non-well-defined CLO		**Well-defined CLO**	
CLO	**Class**	**CLO**	**Class**
Students will understand multimedia systems	Not Measurable	Students will be able to define fundamental principles of text, audio, images, video and animation used in multimedia systems	Knowledge
Students will be able to design network	Not Clear	Students will be able to design and configure LAN Networks for standard use	Skill
Students will State, explain and apply the theories of human behavior in organizations - this will include all three levels of OB-individual, group and organizational. Oral, written, and listening Skills are developed by encouraging students to participate in class discussion, to engage in homework assignments, and to interact with classmates. The course has a significant writing component aiming at improving writing Skills	Not Concise	Students will be able to demonstrate professional and ethical skills while designing and implementing electronic circuits using semiconductor devices	Value
Illustrate different ideas of applications in data structure and algorithms	Not Measurable	Describe the general structure of an operating system and its functions.	Knowledge
Operate with the usage of Internet to work with geometrical calculations	Not Clear	Develop and run supervised and unsupervised machine learning algorithms	Skill
Activation of Students? Group Discussion sessions, on Sequential Logic Circuit Building Blocks, Latches, Flip-flops: RS, D, JK and T, Synchronous Sequential Logic Circuits etc.	Not Concise	Demonstrate the professional and ethical Value by cooperating as a team member of the project and exhibit their leadership qualities	Value

DATA PRE-PROCESSING

The most important stage is the data pre-processing because any defect in the data will inevitably lead to wrong results. Before the implementation of the classification algorithm, necessary pre-processing tasks were applied to improve data efficiency. Pre-processing includes several techniques (e.g., reduction, inconsistency, missing, and expecting values. As shown in Figure 2, out of 3200 initial CLOs, 1688 are in majority class *Skills* and 54 are in minority class *Not Clear* which clearly indicates that there is a disproportion between classes. The examples of the minor class are relatively very small while the representatives of major class appear much more frequently. This is a situation of an imbalanced (Kaur et al., 2020) dataset where there is a severe skew in the class distribution. Most CLOs were collected from departments' websites. Usually, any organization in education field checks the content that will be published on its website. Hence, most published CLOs are well-written despite their re-evaluation in our case by quality experts. The non-well written CLOs were identified by programs applying for accreditation as mentioned before.

Imbalanced classification represents a challenge in the domain of predictive modeling. The class of minority instances in such uneven classification might lead to wrong predictions (Krawczyk, 2016).

The results of the predictive performance models (specifically for the minority classes: Not concise and Not clear) can be poor since ML algorithms used for classification assume having an equal number of samples for all classes. This represents an issue in our case because classes with a fewer number of samples, being the most important ones, are more sensitive to classification errors than the majority class. To deal with the above problem, we distinguish three main methods (Japkowicz & Stephen, 2002) for increasing training data variety without gathering more data:

- Data-level approach: it is based on the generation of new objects for minority groups or/and reducing the number of samples from the majority groups
- Algorithm-level approach: it focuses on changing existing learners to lessen their bias towards majority classes
- Hybrid approach: it combines the previous two methods.

In our case, we bet our choice on the generation of new instances for the minority class since it is the simplest technique and avoids losing data (Krawczyk, 2016). Since the dataset is in text format, the production of samples for minor groups should take into account the syntax and the semantic constraints. The augmentation technique (Marivate & Sefara, 2020) offered by NLPAUG package (Umasankar, 2021) is used. Data augmentation can be performed through different methods (Li et al., 2022; Marivate & Sefara, 2020) such as:

- The paraphrasing-based technique: it focuses on proper and restrained changes to sentences using a database of synonyms, a thesaurus, semantic embeddings, back translation, or model generation
- The noising-based technique: it adds discrete or continuous faint noise without seriously affecting the semantics
- The sampling-based technique: it includes rules and trained models to generate new data

Figure 2 presents a chart of the details of the dataset records. It shows the frequency of each CLO before and after augmentation. From the chart, we can observe that the number of CLO having *Skill*s is the highest one since the used CLOs belong to engineering domain. The augmentation was applied only for the five remaining classes. As shown in Figure 2, the number of instances of each remaining class was duplicated. As a result of augmentation application, the new CLOC size is equal to 4712 records. Table 2 presents some examples of the generated CLOs using augmentation. The new CLOs belong to the 5 classes: *Knowledge, Value, Not Clear, Not Concise, Not Measurable*.

In the next pre-processing phase, we applied steps for Natural Language processing. First, we used NLP techniques to process text of the raw data. We deleted all stop words, all question marks, punctuations, and other marks. After applying stemming technique, all the words in the dataset are extracted. The sparse matrix is then constructed: the rows represent the CLOs and the columns represent words in the dataset the content of any cell is 0 or 1 depending on the occurrence of the word.

For this purpose, we used the NLTK library (Bird et al., 2009); at the end of this phase, the text became ready the deal with it.

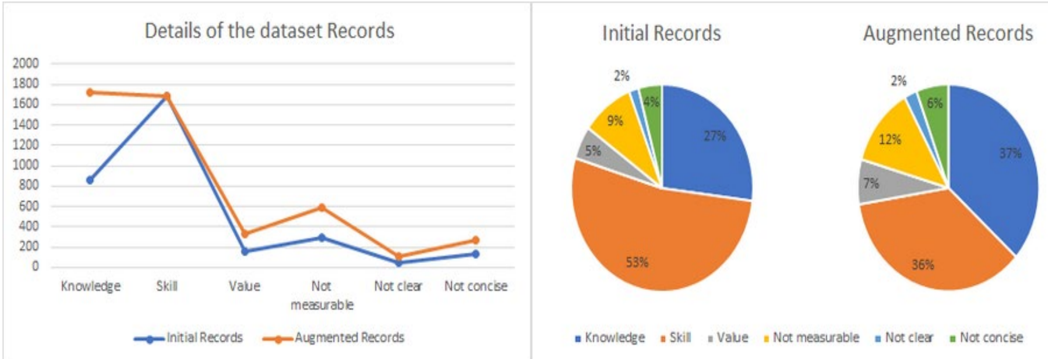

Figure 2: Chart presentation of the details of the dataset records

Table 2: Example of generated CLOs after applying augmentation

CLO before augmentation	**Class**	**CLO after augmentation**
Utilize machine learning models for solving various contemporary issues.	Knowledge	Utilize machine erudition models for solve several contemporary issues.
Evaluate ethically a secure system and assess the differences between the security technologies used.	Value	Evaluate ethically a secure system and value the differences between security technologies practice.
Read and understand Java-based software code.	Not Measurable	Read and interpret Java - found software code
Recognize the applicability of microprocessors	Not Concise	Recognize the pertinence of microprocessors
Illustrate the understood concepts in both ways of written and oral	Not Clear	Illustrate the silent concepts in both ways of write and oral

Application of Machine learning

After pre-processing and preparing the real datasets, we gained an understanding of the data set. In the next stage, we built the CLOCML model by applying the ML classification algorithms (Suthaharan, 2015). Several classifiers can be used for predicting the CLO category. Since there is no specific ML technique that delivers the best prediction in all cases, we use in our experiment the following supervised model: Support Vector Machine (SVM), Random Forest, Naive Bayes and XGBoost.

SVM (Hearst et al., 1998) is a classifier-building technique. Its goal is to establish a judgment boundary between two groups that allows labels to be predicted from one or more feature vectors. This judgment boundary, known as the hyperplane, is oriented in such a way that it is as possible from one of the classes' nearest data points as possible. Help vectors are the points that are the nearest together.

The Naive Bayes algorithm (Berry, 1995) is based on Bayes Theorem and has also a probabilistic nature. The theorem is shown in Equation 1.

$$\mathrm{P(A/C)} = {\mathrm{P(C/B)P(B)}}\big/{\mathrm{P(C)}} \qquad (1)$$

Random forest (Breiman, 2001) is a machine learning algorithm for solving classification and regression problems. It is based on ensemble learning, a technique which combines many classifiers to present solutions to large problems. It uses several decision trees. The generated forest by this algorithm is trained through bagging or bootstrap aggregating.

XGBoost (Chen & Guestrin, 2016) is an implementation of gradient boosted decision trees dedicated to solve various data science problems in an accurate and fast way. Boosting algorithms are extremely useful when dealing with bias-variance trade-off. Contrary to bagging algorithms that only control for high variance in a model; boosting controls both the aspects bias and variance.

We divided our dataset into two sets. One dataset is for training, which represents 80% of the total data. The second dataset is for testing, which represents 20% of the total data. 80/20 is very frequently occurring ratio referred to the Pareto principle (Nisonger, 2008)

Model Evaluation

The experiment involves 3200 records of the CLOC dataset to predict the category of CLOs. We evaluate the performance of our predictive models using four common model evaluation metrics for ML (LaetitiaVanCauwenberge, 2016). To measure the classifier performance, two metrics are used: Accuracy and F1-score. The test of accuracy is the proportion of the total number of correct predictions. The Equation 2 represents the formula for quantifying the accuracy:

$$\mathrm{Accuracy} = {(\mathrm{TP} + \mathrm{TN})}\big/{(\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN})} \qquad (2)$$

Where TP is the true positive, TN is the true negative, FP is the false positive and FN is the false negative.

The F1-score is represented by Equation 3. It represents the harmonic means of the values of precision (Predicted CLO class) and recall (sensitivity) for a classification problem.

$$\mathrm{F1} = 2 * {(\mathrm{Precision} * \mathrm{Recall})}\big/{(\mathrm{Precision} + \mathrm{Recall})} \qquad (3)$$

Results and Discussion

In this section, the results of CLOCML model tests are compared to find the model that gives the best solution for CLO classification. Table 3 summarizes the performance evaluation results of CLOCML models. We found that CLOCML using Random Forest and Naive Bayes models are ineffective for the problem in hand since they present low accuracy and F1-score values. However, the SVM and XGBoost models reached high scores.

Table 3: Performance evaluation results of CLOCML models

Technique	Accuracy	F1-Score
CLOCML using SVM	0.83	0.80
CLOCML using Random Forest	0.55	0.47
CLOCML using Naïve Bayes	0.56	0.57
CLOCML using XGBoost	0.73	0.70

The CLOCML using SVM model is the best classification model which achieved an accuracy of 0.83 with a F1-score of 0.8. The reason behind the good performance of SVM is indicated by (Joachims,

1998) as follows: "SVMs are very well suited for text categorization. The theoretical analysis concludes that SVMs acknowledge the particular properties of text: (a) high dimensional feature spaces, (b) few irrelevant features (dense concept vector), and (c) sparse instance vectors. The experimental results show that SVMs consistently achieve good performance on text categorization tasks, outperforming existing methods substantially and significantly"

Table 4: CLOCML-SVM Confusion matrix

	Knowledge	**Skills**	**Not clear**	**Value**	**Not concise**	**Not measurable**
Knowledge	0.911	0.074	0	0	0	0.015
Skills	0.047	0.941	0	0.006	0.003	0.003
Not clear	0.176	0.647	0	0.059	0	0.118
Value	0.043	0.170	0	0.766	0	0.021
Not concise	0.306	0.571	0	0.020	0.061	0.041
Not measurable	0.058	0.192	0	0	0	0.750

Table 4 **presents the best model (CLOCML-SVM) Confusion matrix** which illustrates the proportions of model predictions vs the actual results. The ~~used~~ green color designates the correctly classified CLO's, whereas the orange color indicates the misclassified ones. As the matrix shows, our model is able to correctly classify the CLOs as *Knowledge* or *Skills* with a high accuracy of 91% and 94%, respectively. The *Value* and *Not Measurable* classes had less but still good accuracy values reaching 76% and 75%, respectively. However, the remaining two categories, *Not Clear* and *Not Concise*, yielded low performance. This performance degradation is due to the fewer number of CLOs involved in the dataset.

As we have previously declared, our specific research area, namely the CLOs classification using ML, is still not tackled in the literature. Therefore, we compared the results of our study with a study conducted by Bazelais et al. (2018) in the general domain of Outcome Based Education but focusing in predicting the students' academic achievement. This study used two different datasets: MOOC dataset (Lemay & Doleck, 2020) and CEGEP Academic Performance dataset (Bazelais et al., 2018). The first dataset contains 6241 instances. Using this dataset, the accuracy of the different machine learning algorithms varies from 0.63 to 0.69. As for the second dataset, which includes 309 records, the performance results vary from 0.84 to 0.90 in terms of accuracy as well. Our present work is similar to this study in terms of the number of instances in the datasets. Our results are also aligned with its achieved accuracy when applying ML techniques for predicting students' academic achievement.

Referring to the research question RQ1 ("Can machine learning be used to determine if a course outcome is well written or not?"), we proved that it is possible to accurately identify a CLO category using ML. Compared to Random Forest, Naive Bayes and XGBoost, SVM performed the best in such prediction with an accuracy of 83%. As for the second research question RQ2 ("How would machine learning be effective for categorizing well-written course outcomes?"), we conclude that the automatic prediction of the CLO class is more effective than the manual classification work. It permits educators, quality committees and accreditation bodies to save time and effort during the quality management processes. It also improves the credibility of course assessments since well written CLOs will result in right measures of course and program learning outcomes.

CONCLUSION

During the accreditation processes, the CLOs measurement is required for the education quality standards. A wrong CLO paraphrasing will affect different KPI values. To the best of our knowledge, there are no existing tools for characterizing a given course learning outcome. In this paper, we developed four predictive models using ML classification algorithms. The developed models were trained and tested using a dataset collected from different universities of engineering and computer science programs. The classification of CLOs was done by quality experts. Among the four models, two of them reached high accuracy scores and the SVM model had the highest scores for all the used performance evaluation metrics. It was selected as the best classification model for CLO characterization with an accuracy of 83%.

The obtained result proves that our method serves as a reliable guidance and support tool for the teaching staff to check whether the proposed outcomes are well written or not. Well formulated CLOs result in their right achievement and consequently the PLOs right accomplishment. Actions and recommendations for a particular course can then be proposed in the appropriate way. This method will also be useful for department quality team as well as evaluators and reviewers.

The presented study is limited to 3200 CLOs from engineering and computing domains. The next steps consist in increasing the number of outcomes from more different domains and specially belonging to the *Not Clear* and *Not Concise* classes. Indeed, other augmentation techniques can be used to reach balanced dataset. Furthermore, we may compare in future studies the results of ML methods and Deep Learning techniques. We may also investigate on checking the CLOs overlaps in the same course. This study can be also extended by the proposal of a CLO suggestion in case of non-well written CLO which would be beneficial to the faculty members.

REFERENCES

Abu-Naser, S. S., Zaqout, I. S., Abu Ghosh, M., Atallah, R. R., & Alajrami, E. (2015). Predicting student performance using artificial neural network: In the faculty of engineering and information technology. *International Journal of Hybrid Information Technology*, *8*(2), 221-228. https://doi.org/10.14257/ijhit.2015.8.2.20

Agaoglu, M. (2016). Predicting instructor performance using data mining techniques in higher education. *IEEE Access*, *4*, 2379–2387. https://doi.org/10.1109/ACCESS.2016.2568756

Alturki, S., Hulpuş, I., & Stuckenschmidt, H. (2022). Predicting academic outcomes: A survey from 2007 till 2018. *Technology, Knowledge and Learning*, *27*, 275–307. https://doi.org/10.1007/s10758-020-09476-0

Alturki, S., & Alturki, N. (2021). Using educational data mining to predict students' academic performance for applying early interventions. *Journal of Information Technology Education: Innovations in Practice*, *20*, 121–137. https://doi.org/10.28945/4835

Anand, M. (2019). Advances in EDM: A state of the art. In: M. Hoda, N. Chauhan, S. Quadri, & P. Srivastava, (Eds.), *Software engineering* (pp. 193–201). Springer. https://doi.org/10.1007/978-981-10-8848-3_19

Bazelais, P., Lemay, D. J., Doleck, T., Hu, X. S., Vu, A., & Yao, J. (2018, August). Grit, mindset, and academic performance: A study of pre-university science students. *EURASIA Journal of Mathematics, Science and Technology Education*, *14*(12), em1615. https://doi.org/10.29333/ejmste/94570

Berry, D. A. (1995). *Statistics: A Bayesian perspective*. Duxbury Press.

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: Analyzing text with the natural language toolkit*. O'Reilly Media, Inc. http://117.3.71.125:8080/dspace/bitstream/DHKTDN/6460/1/Natural%20Language%20Processing%20with%20Python.4149.pdf

Bloom, B. S. (Ed.), Englehart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals*. Longmans Green.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5-32. https://doi.org/10.1023/A:1010933404324

Chen, T., & Guestrin, C. (2016, August). XGBoost: A scalable tree boosting system. In B. Krishnapuram, M. Shah, A. Smola, C. Aggarwal, D. Shen, & R. Rastogi (Éds.), *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). San Francisco, USA: ACM. https://doi.org/10.1145/2939672.2939785

Devine, T., Hossain, M., Harvey, E., & Baur, A. (2011, August 21). Improving pedagogy by analyzing relevance and dependency of course learning outcomes. In *KDD 2011 Workshop: Knowledge Discovery in Educational Data*. San Diego, CA: ACM. https://pslcdatashop.web.cmu.edu/KDD2011/papers/J-kddined2011.pdf

Egger, R., & Gokce, E. (2022). Natural language processing (NLP): An introduction. In R. Egger (Ed.), *Applied data science in tourism: Tourism on the verge* (pp. 307–334). Springer. https://doi.org/10.1007/978-3-030-88389-8_15

Ezz, M. M. (2015). Advisory system for student enrollment in university based on variety of machine learning algorithms. *International Journal of Computing Academic Research*, *4*(2), 34-45. http://www.meacse.org/ijcar/archives/55.pdf

Gogus, A. (2012). Outcomes of learning. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 2534-2539). Springer. https://doi.org/10.1007/978-1-4419-1428-6_145

Gray, C. C., & Perkins, D. (2019). Utilizing early engagement and machine learning to predict student outcomes. *Computers & Education*, *131*, 22-32. https://doi.org/10.1016/j.compedu.2018.12.006

Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and their Applications*, *13*(4), 18-28. https://doi.org/10.1109/5254.708428

Hussain, A., Khan, M., & Ullah, K. (2022). Student's performance prediction model and affecting factors using classification techniques. *Education and Information Technologies*, Online First. https://doi.org/10.1007/s10639-022-10988-8

Ifenthaler, D., & Yau, J. Y. K. (2020, June). Utilising learning analytics to support study success in higher education: A systematic review. *Educational Technology Research and Development*, *68*(4), 1961–1990. https://doi.org/10.1007/s11423-020-09788-z

Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematic study. *Intelligent Data Analysis*, *6*(5), 429–449. https://doi.org/10.3233/IDA-2002-6504

Joachims, T. (1998, April). Text categorization with support vector machines: Learning with many relevant features. *Proceedings of the 10th European Conference on Machine Learning* (pp. 137-142). Chemnitz, Germany: Springer. https://doi.org/10.1007/BFb0026683

Kaur, H., Pannu, H. S., & Malhi, A. K. (2020, July). A systematic review on imbalanced data challenges in machine learning: Applications and solutions. *ACM Computing Surveys*, *52*(4), Article No.: 79. https://doi.org/10.1145/3343440

Khanal, S. S., Prasad, P. W. C., Alsadoon, A., & Maag, A. (2020). A systematic review: Machine learning based recommendation systems for e-learning. *Education and Information Technologies*, *25*(4), 2635–2664. https://doi.org/10.1007/s10639-019-10063-9

Krawczyk, B. (2016). Learning from imbalanced data: Open challenges and future directions. *Progress in Artificial Intelligence*, *5*(4), 221-232. https://doi.org/10.1007/s13748-016-0094-0

LaetitiaVanCauwenberge (2016, February 20). *11 important model evaluation techniques everyone should know*. Data Science Central. https://www.datasciencecentral.com/7-important-model-evaluation-error-metrics-everyone-should-know/

Le, B., & Nguyen, H. (2015). Twitter sentiment analysis using machine learning techniques. In: H. Le Thi, N. Nguyen, & T. Do (Eds.), *Advanced computational methods for knowledge engineering* (pp. 279-289). Springer. https://doi.org/10.1007/978-3-319-17996-4_25

Lemay, D. J., & Doleck, T. (2020, October). Grade prediction of weekly assignments in MOOCS: Mining video-viewing behavior. *Education and Information Technologies*, *25*(2), 1333–1342. https://doi.org/10.1007/s10639-019-10022-4

Li, B., Hou, Y., & Che, W. (2022). Data augmentation approaches in natural language processing: A survey. *AI Open*, Online First. https://doi.org/10.1016/j.aiopen.2022.03.001

Mahboob, K., Ali, S. A., & Laila, U. E. (2020). Investigating learning outcomes in engineering education with data mining. *Computer Applications in Engineering Education*, *28*(6), 1652-1670. https://doi.org/10.1002/cae.22345

Marivate, V., & Sefara, T. (2020, August). Improving short text classification through global augmentation methods. *Proceedings of the International Cross-Domain Conference for Machine Learning and Knowledge Extraction* (pp. 385-399). Dublin, Ireland: Springer. https://doi.org/10.1007/978-3-030-57321-8_21

Namoun, A., & Alshanqiti, A. (2021). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Applied Sciences*, *11*(1), 237. https://doi.org/10.3390/app11010237

Nisonger, T. E. (2008). The "80/20 rule" and core journals. *The Serials Librarian*, *55*(1-2), 62-84. https://doi.org/10.1080/03615260801970774

Queen Mary University of London (2014). *Good practice guide on writing aims and learning outcomes.* London: The Learning Institute. https://www.yashada.org/yash/ttt_new/static_pgs/TC/Aims-and-Outcomes-Guide.pdf

Rao, N. J. (2020). Outcome-based education: An outline. *Higher Education for the Future*, *7*(1), 5-21. https://doi.org/10.1177/2347631119886418

Romero, C., Romero, J. R., & Ventura, S. (2014). A survey on pre-processing educational data. In A. Peña-Ayala (Ed.), *Educational data mining* (pp. 29–64). Springer. https://doi.org/10.1007/978-3-319-02738-8_2

Romero, C., & Ventura, S. (2013). Data mining in education. In *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *3*(1), 12-27. https://doi.org/10.1002/widm.1075

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(3), e1355. https://doi.org/10.1002/widm.1355

Slater, S., Joksimović, S., Kovanovic, V., Baker, R. S., & Gasevic, D. (2017). Tools for educational data mining: A review. *Journal of Educational and Behavioral Statistics*, *42*(1), 85-106. https://doi.org/10.3102/1076998616666808

Solak, S., Ucar, M. H., & Albadwieh, M. (2020). Computer based evaluation to assess students' learning for the multiple-choice question-based exams: CBE-MCQs software tool. *Computer Applications in Engineering Education*, *28*(6), 1406-1420. https://doi.org/10.1002/cae.22309

Spady, W. G. (1994). *Outcome-based education: Critical issues and answers.* Arlington, VA: American Association of School Administrators. https://files.eric.ed.gov/fulltext/ED380910.pdf

Suthaharan, S. (2015). *Machine learning models and algorithms for big data classification: Thinking with examples for effective learning*. In R. Sharda, & S. Voß (Eds.), Integrated Series in Information Systems, Vol 36. Springer.

Umasankar, N. (2021, August 25). *NLPAUG – A Python library to augment your text data*. Analytics Vidhya Blog. https://www.analyticsvidhya.com/blog/2021/08/nlpaug-a-python-library-to-augment-your-text-data/

Yağcı, M. (2022, March). Educational data mining: Prediction of studentsacademic performance using machine learning algorithms. *Smart Learning Environments*, *9*(1), 1-19. https://doi.org/10.1186/s40561-022-00192-z

Yogeswari, S., & Rajermani, T. (2022). A review on the attainment of outcome-based education using educational data mining. *Journal of Data Science*, *2022*(03), 1-11. http://eprints.intimal.edu.my/1586/

AUTHORS

Yessine Hadj Kacem is a senior researcher with Computer and Embedded System (CES) research team at the National Engineering School of Sfax, Tunisia. He obtained his Ph.D. in Computing Systems from the National Engineering School of Sfax in 2012. Dr Yessine is currently an associate professor in Gafsa university, Tunisia. He was the head of computer sciences department in the college of computer Science Abha, King Khalid University, Saudi Arabia from January 2018 till June 2021. He contributed there in NCAAA and ABET accreditation as well as the development of new study plans for bachelor and master programs. Dr. Yessine is specialized in software engineering.

Safa Alshehri obtained in 2016 her Bachelor degree in Information Systems from King Khalid University KKU, the Master Degree in Information System in 2021 from the same university. Since 2017, she has been Quality Matters Coordinator in KKU, eLearning Deanship. She is Specialized in quality management and responsible of educating instructors while using e-learning systems and quality matters rubric.

Talal Qaid was born in Taiz, Yemen, in 1968. He received the Ph.D. degree in computer Science from Cairo University, Egypt, in 2009. From 2009 to 2014, he was an Assistant Professor with Computer Science Department, Faculty of Computer Science and Engineering, Hodeidah University, Hodeidah, Yemen. Since 2014, he has been an Assistant Professor with Computer Science Department, Faculty of Computer Science, King Khalid University, Abha, Saudi Arabia. His research interests include recommender systems, data mining, and artificial intelligence.