SEBHA UNIVERSITY CONFERENCE PROCEEDINGS, VOL.03 NO. 3 2024
DOI: 10.51984/SUCP.V3I3.3813

المؤتمر السادس للعلوم الهندسية والتقنية
The Sixth Conference for Engineering Sciences and Technology (CEST-6)
Conference Proceeding homepage: https://cest.org.ly

Arabic Predicting Course Difficulty in Online Education Using Machine Learning

Abeer Farag Al-Nair

Higher Institute of Science and Technology - Tarhuna

Keywords:
Online Education
Course Difficulty Prediction
Machine Learning

A B S T R A C T

The booming proliferation of online educational platforms over the last several years has transformed the landscape of learning resources. On one hand, this transformation has generated many promising opportunities to optimize course delivery and student performance, while, on the other hand, it has posed various dilemmas, one of which is accurately predicting the course level of difficulty. Since the assessment of course difficulty using traditional methods varies greatly and is often determined subjectively, it is difficult for learners to assess whether they have the necessary skills and knowledge to manage the course or not. To resolve this problem, in the present study, we develop a machine learning-based framework to predict course difficulty levels on Coursera Course Dataset. Our contributions include the employment of three strong classifiers, which we compare to one another: GB, RF, and XGBoost. We also conducted a considerable amount of preprocessing, such as missing values, categorical variables encoding, and SMOTE for balancing the dataset. The evaluation results demonstrate the superiority of the XGBoost model with an accuracy of 96.4% and excellent precision, recall, and F1 scores for all classes. The implications of this study include not only its potential for enhancing course recommendation systems and personalizing online education but also for further refinement by introducing more features and real-time predictions.

تنبؤ بصعوبة الدورة التعليمية عبر الإنترنت باستخدام التعلم الآلي

عبير فرج الناير

المعهد العالي للعلوم والتكنولوجيا - ترهونة.

الكلمات المفتاحية
التعليم عبر الإنترنت
التنبؤ بصعوبة الدورات
التعلم الآلي

الملخص

الانتشار الواسع لمنصات التعليم عبر الإنترنت في السنوات الأخيرة قد غيّر بشكل كبير مشهد الموارد التعليمية. من ناحية، أدت هذه التحولات إلى خلق العديد من الفرص الواعدة لتحسين تقديم الدورات وأداء الطلاب، ومن ناحية أخرى، طرحت عدة تحديات، أحدها هو التنبؤ بدقة بمستوى صعوبة الدورات. نظرًا لأن تقييم صعوبة الدورات باستخدام الأساليب التقليدية يختلف بشكل كبير وغالبًا ما يكون قائمًا على التقدير الذاتي، يصبح من الصعب على المتعلمين تقييم ما إذا كانوا يمتلكون المهارات والمعرفة اللازمة لإدارة الدورة بنجاح أم لا. لحل هذه المشكلة، قمنا في هذه الدراسة بتطوير إطار عمل يعتمد على تعلم الآلة للتنبؤ بمستويات صعوبة الدورات باستخدام مجموعة بيانات دورات "كورسيرا". تشمل مساهماتنا استخدام ثلاثة مصنفات قوية قمنا بمقارنتها مع بعضها: Gradient Boosting (GB)، Random Forest (RF)، و XGBoost. كما قمنا بتنفيذ العديد من عمليات المعالجة المسبقة، مثل التعامل مع القيم المفقودة، وترميز المتغيرات الفئوية، واستخدام تقنية SMOTE لموازنة مجموعة البيانات. أظهرت نتائج التقييم تفوق نموذج XGBoost بدقة تصل إلى 96.4% بالإضافة إلى نتائج ممتازة في الدقة والاسترجاع وقيمة F1 لجميع الفئات. تشمل دلالات هذه الدراسة إمكانيات تحسين أنظمة توصية الدورات وتخصيص التعليم عبر الإنترنت، فضلاً عن المزيد من التحسين من خلال إدخال ميزات جديدة وتنبؤات في الوقت الفعلي.

1- Introduction

High-quality educational resources have become increasingly accessible to people all over the word thanks to rapid progress in

*Corresponding author:
E-mail addresses: abeerfarag82@yahoo.com
Article History : Received 14 March 2024 - Received in revised form 29 August 2024 - Accepted 21 October 2024

recent years on digital education platforms [1].With the rapid spread of digital courses both the type and level of content has widened [2].And yet, the increase in course availability also causes students and schools to face the problem of selecting a class with an appropriate degree of difficulty [3].In much cross-disciplinary research, appraising the difficulty of a class is important for several reasons: it will help one understand the situation and what kinds of difficulties you may face, and with predictive analytics you can get a heads-up on sections that need to be reviewed beforehand [4].Such guidance will benefit both students and schools which need to prepare for strenuous academic tasks.

Traditionally, course difficulty has been evaluated subjectively, either by educators assessing what they believe to be difficult for students or by students' self-assessments [3]. These methods, however, can be inconsistent and biased, often resulting in mismatches between course expectations and student readiness. Such discrepancies can lead to higher dropout rates as students become frustrated or lose motivation when courses are either too advanced or too elementary for their skill level. Consequently, there is a pressing need for a data-driven approach to determine course difficulty levels [4].

Machine learning is one of the powerful tools that has proven to be beneficial and efficient in numerous domains, and education is no exception [5]. ML algorithms have an outstanding capability to learn from existing data, detect patterns, and make forecasts. Such levels of automation and improvement are hard to achieve with any existing methods. In terms of a course's difficulty, machine learning models receive data on hundreds or thousands of online courses, including descriptions, ratings, and the number of enrolled. As a result, a predictive model can define the difficulty with a high level of accuracy. More than that, it gives an extrinsic standard with which one can compare courses and decide on a learning path personally [6].

Research problem

Established traditional means to estimate day difficulty subject are infrequently considerable and air of impartial estimation. This disconnect between course design and student readiness can result in negative learning experiences for students, often leading to decreased retention. With the continued and rapid expansion of online education even more so, there is strong imperative to provide better course-difficulty predictions based on data.

The main objective of this study is to take an effort for the development of a new framework using machine learning techniques, which can be helpful in predicting hardness level courses. Using state-of-the-art machine learning methods, our goal is to develop a model that can give good estimates of the difficulty; thereby not only helping students prepare for exam conditions but also providing educators with helpful information. We then compare the performance of different ML models for getting accurate prediction. This will allow us to contribute towards the emerging area of educational data mining, providing more personalized online learning experiences.

2- Literature review

The paper [7] proposes a predictive model for identifying at-risk students in online learning platforms. It utilizes various machine learning and deep learning algorithms to analyze students' study behavior and performance. By considering factors like assessment scores, engagement intensity, and time-dependent variables, the model aims to predict student dropout risks and enable timely intervention by instructors. Experimental results show that the Random Forest algorithm performs best in terms of accuracy, precision, recall, and F-score, offering potential for reducing dropout rates and enhancing student engagement in online courses.

The paper [8] introduces a new model employing machine learning algorithms to predict undergraduate students' final exam grades using midterm exam scores as input. It evaluates the performance of various algorithms including random forests, nearest neighbor, support vector machines, logistic regression, Naive Bayes, and k-nearest neighbor. With data from 1854 students enrolled in a Turkish Language-I course, the proposed model achieves a classification accuracy of 70–75% using only midterm exam grades, department, and faculty data. The study underscores the significance of data-driven approaches in higher education for establishing learning analysis frameworks and aiding decision-making processes, particularly in early identification of students at risk of failure. The paper [9] addresses the challenge of dropout rates in online learning by proposing predictive models aimed at early identification of at-risk students. Recognizing dropout prediction as a sequence labeling or time series problem, the study introduces two models: Logistic Regression with regularization and the Input-Output Hidden Markov Model (IOHMM). Results demonstrate an 84% accuracy in predicting at-risk students compared to baseline machine learning models. These predictive models offer instructors timely intervention opportunities, potentially mitigating dropout rates and enhancing the continuity and growth of online courses.

The paper [10] presents a student academic performance prediction model utilizing supervised machine learning algorithms such as support vector machine and logistic regression. Through various experiments employing different technologies, it compares the results, demonstrating that the sequential minimal optimization algorithm surpasses logistic regression in accuracy. The research aims not only to forecast students' future performance but also to identify impactful features like teacher performance and student motivation, which can aid in categorizing student performance as good or bad and potentially reducing dropout rates in educational institutes.

The study [11] explores the efficacy of video-based learning, particularly in the context of flipped teaching, to enhance student academic performance in higher educational institutions (HEI). Utilizing data from 772 students enrolled in e-commerce and e-commerce technologies modules, the research aims to predict students' overall performance using video learning analytics and data mining techniques. Eight classification algorithms are applied to analyze data from various sources including the student information system, learning management system, and mobile applications. Data preprocessing techniques, such as feature reduction through genetic search and principal component analysis, are employed to refine the analysis. The results indicate that Random Forest achieves an accuracy of 88.3% in accurately predicting successful student outcomes, offering insights into effective teaching methods in HEIs.

The study [12] investigates predicting student dropout at the Karlsruhe Institute of Technology (KIT) using logistic regression and decision trees. Examination data, readily available at all universities, forms the basis for the models, suggesting a practical approach applicable to other institutions. While decision trees yield slightly better results than logistic regressions, both methods achieve high prediction accuracies of up to 95% after three semesters. Importantly, even after just one semester, classification accuracy exceeds 83%, indicating early detection potential for dropout risk.

The paper [13] addresses the societal shifts catalyzed by events like the COVID-19 pandemic, emphasizing the transformation of education through information and communication technologies, particularly learning management systems. It proposes integrating artificial intelligence and data analysis with these systems to enhance learning experiences, reflecting a shift towards robust educational models in the "new normal." The aim is to leverage technologies like virtual assistants to support and guide students in their online learning endeavors. The paper [14] addresses the importance of student success in higher education and the potential of machine learning techniques to predict and support at-risk students. It highlights the challenges educators face in implementing data mining strategies due to technical barriers and aims to provide comprehensive guidance for utilizing these techniques effectively. By synthesizing existing literature and offering a methodical approach, the paper seeks to empower educators to leverage data mining tools to improve student outcomes, ultimately lowering the barrier to entry for these technologies in the classroom.

1- Methodology

A. Dataset

This study used the Coursera Course Dataset which was last updated in August 2020 and obtained from the Kaggle website [23], which was scraped from the official Coursera website as part of a hackathon project for an intelligent course recommendation system. The major aim of the project was to help new learners quickly find courses of interest by allowing them to answer a few simple questions. This dataset served as the foundation for the model, which used regression to predict the difficulty of courses. It contains comprehensive information about 890 different courses available on the Coursera platform, including six critical columns that provide essential features of the courses:

- **course title**: Consists of course titles that give a broader description of what the course might entail.
- **course organization**: Contains the organizations/institutions that offered the course.
- **course Certificate type**: Details the types of certification that one gets after completing the course.
- **course rating**: The rating that each course attained, based on numerous learners who had taken it.
- **course difficulty**: Provides the basis for how the rest of the variables were fetched using the model project.
- **course students enrolled**: Contains the number of students that took the course, indicating its popularity and providing insight into the course's relevance to learners.

The dataset also makes it possible for one to conduct an in- depth analysis and generate predictive models that accurately assess and predict course difficulty. As a result, the dataset promotes personalized learning by recommending courses based on students' abilities and learning ambitions. Furthermore, the code and other sources associated with the data scraping for the dataset generation are available on GitHub for more analysis.

B. Dataset Exploratory Data Analysis

Exploratory Data Analysis (EDA) plays a crucial role in our study, serving as the foundation for understanding the Coursera Course Dataset and uncovering meaningful insights that drive our machine learning models. During the EDA process, we meticulously examined the dataset to identify patterns, trends, and anomalies [15].

The pie chart depicted in Figure 1 illustrates the proportion of various certification options available for courses in our dataset. The chart is divided into three distinct segments, each representing a different type of certificate.

The largest segment, constituting 65.3% of the total, is labeled "COURSE." This indicates that the majority of the courses in our dataset offer a general course completion certificate. This type of certificate is typically awarded upon successful completion of a course and may not necessarily indicate a specialization in a particular subject area.

The second-largest segment, comprising 33.3%, is labeled "SPECIALIZATION." This signifies that a significant portion of the courses provide specialization certificates. Specialization certificates usually require the completion of a series of related courses, reflecting a more in-depth focus and expertise in a specific domain.

The smallest segment, making up 1.3% of the total, is labeled "PROFESSIONAL CERTIFICATE." This suggests that only a small fraction of the courses offer professional certificates, which are often designed to meet the requirements of specific industries or professional standards. These certificates typically have more rigorous criteria and may be recognized by employers as indicative of a certain level of professional competency.

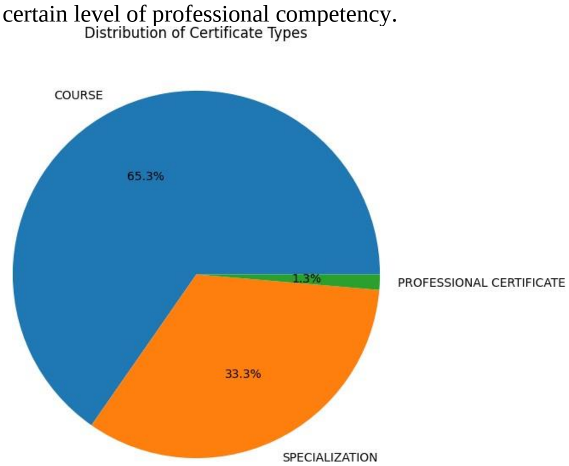

Fig. 1: Distribution of Certificate Types

Figure 2 shows the distribution of course ratings in the Coursera Course Dataset. The plot is a graphical representation, a histogram overlaid with a density plot, indicating the percentage of course ratings on the dataset scale. The plot indicates course ratings spread, illustrating that the largest group of courses has high ratings. The course ratings mainly cluster around 4.5 to 5.0. This situation reflects that most courses on Coursera have been rated highly, with few courses attracting a rating lower than 4.0. The high peak around 4.75 indicates scale by a huge volume of course ratings probably rating very high. The right-skewed distribution along the tail to low rating indicates that still a few courses' ratings are rated low.

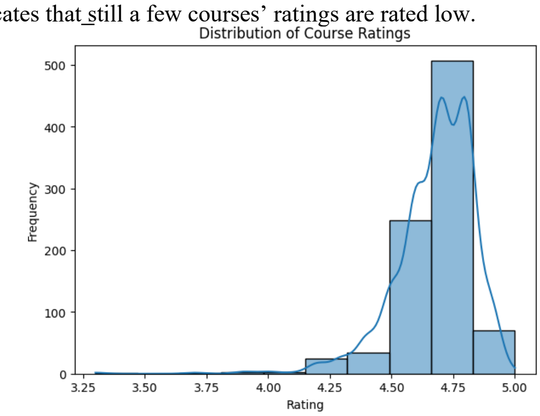

Fig. 2: *Distribution of Course Ratings*

Figure 3 visualizes the distribution of course difficulty levels in the Coursera Course Dataset. This bar chart divides courses into the four levels of difficulty: Beginner, Intermediate, Mixed, and Advanced. As we can see, the largest category is Beginner, in which almost 500 courses can be found. This shows a prominent focus on basic knowledge and skills to cover the demand from students who have no prior knowledge of the subject. The Intermediate and Mixed parts are at about the same level, each accounting for approximately 200 courses. It means that there is enough content for students who want to deepen their fundamental knowledge and for those who want to medium version of different difficulty comprised in one course. Finally, the Advanced level is the smallest category, with very few courses. This means that, although Coursera offers services for those who want in-depth, high-level knowledge, such courses are much rarer than for beginners and intermediate students.

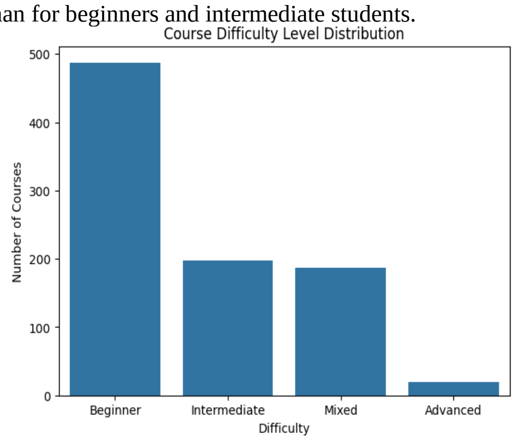

Fig. 3: *Course Difficulty Level Distribution*

The histogram illustrated in Figure 4 shows the distribution of student enrollment across courses covered in the Coursera Course Dataset. The histogram comes with a density plot and indicates that most courses have low enrollment among students, evidenced by a bell-shaped pattern with sharp peaks at the bottom. In other words, there are many courses that have low enrollment number. The frequency of such high enrollment numbers rapidly decreases as the number of enrolled student increase, leading to a right-skewed distribution. More than half of the courses have enrollment not exceeding 500,000 students while very few have enrollment number not exceeding or equal to 1 million. There is a massive long tail towards the right on the graph, implying that very few courses are overly popular relative to the others. The diagram sums up the courses' popularity relative to one another, with only a few being highly popular and another composed of the least popular ones.

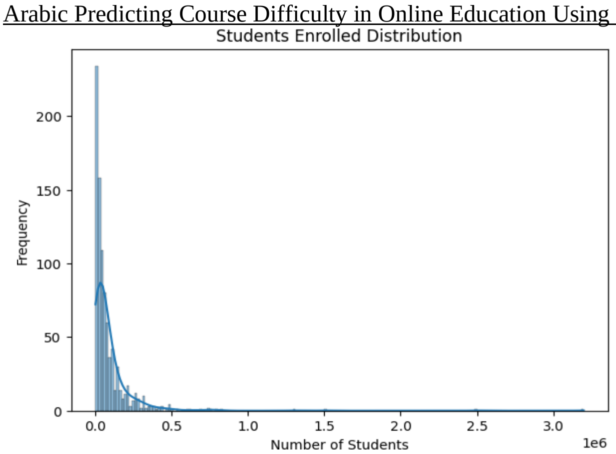

Fig. 4: Course Difficulty Level Distribution

The box plot in Figure 5 below shows the distribution of course ratings based on the course's difficulty level, which include Beginner, Intermediate, Mixed, or Advanced in the Coursera Course Dataset. The courses score high median ratings mostly above 4.5 regardless of the course's difficulty level. As shown, the Beginner's quartiles are narrow, meaning the ratings are consistent with only a few outliers. However, the Intermediate and Mixed difficulty levels have a wider IQR with multiple low outliers with ratings below 4.0, which indicates mixed feelings towards the course. Although the Advanced difficulty level does not have a similar number of ratings, it has the widest range of IQRs and multiple low outliers, indicating advanced level courses have varying levels of learner satisfaction with some falling below the median rating of 4.0. This clearly demonstrates the coarse-related challenges and learner satisfaction.

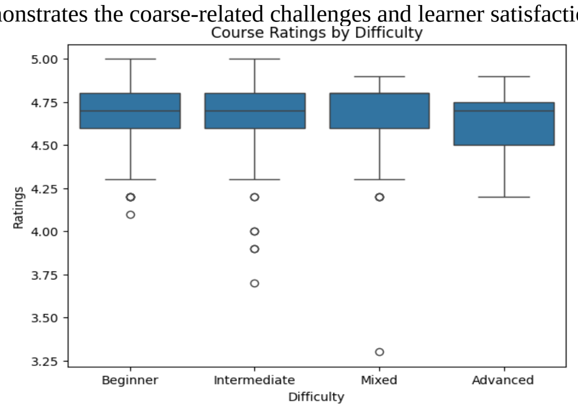

Fig. 5: Course Ratings by Difficulty

Figure 6 presents the relationship between course ratings and student enrollments using the Coursera Course Dataset. It indicates a positive correlation whereby a course with a higher rating attracts more students. Most notably, each rated course between 4.5-5.0 reveals a different number of students, with one course boasting over a million students, which implies that the highly rated courses attract higher enrollments. The lowrated courses are less popular due to their fewer students, with any course below 4.0 indicating a low number of students. Additionally, most courses are rated between 4.0 to around 4.75 are associated with a specific number of enrollment and are well received by students. Few courses fall below 3.75 are less popular with few students enrollments and more received students. In conclusion, the scatter plot indicates the quality of a course determines its student population in the online learning course.

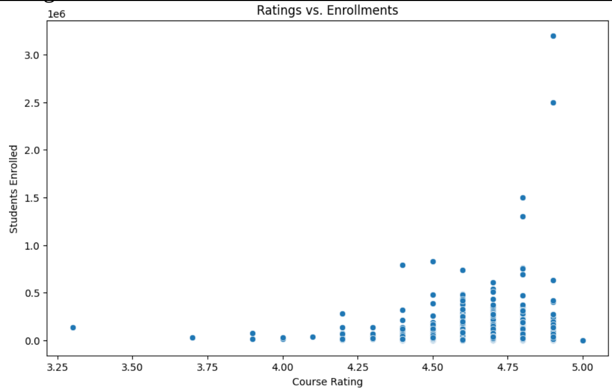

Fig. 6: Ratings vs. Enrollments

Figure 7 presents a heatmap illustrating the course ratings across different difficulty levels (Beginner, Intermediate, Mixed, and Advanced) for various organizing institutions. The color intensity represents the average rating, with darker shades indicating higher ratings and lighter shades indicating lower ratings. The heatmap reveals that many institutions consistently receive high ratings across all difficulty levels, as evidenced by the dominance of dark red shades. For instance, institutions such as the Georgia Institute of Technology, IE Business School, and Emory University exhibit high ratings for both Beginner and Intermediate courses. Conversely, the Intermediate and Mixed difficulty levels show greater variability, with some institutions like the University of Pennsylvania and the University of Michigan maintaining high ratings, while others exhibit lower ratings, shown by lighter shades. Advanced courses, though fewer in number, display the widest range of ratings, with notable high ratings for some institutions like the University of California, Irvine, and lower ratings for others. This variability indicates the diverse challenges associated with higher-level courses.

C. Dataset Pre-processing

The data preprocessing step of our work consists of several crucial operations to make the dataset clean, structured, and ready for training the machine learning model [16]. First, a check for missing values in all columns was performed to assess the gaps in the data that need to be filled. The missing values in the columns 'course title' and 'course difficulty' were filled with a placeholder 'Unknown', the values of the column 'course Certificate type' were loaded with a value 'None'. The missing values of the column 'course rating' were loaded with the mean rating and that of 'course students enrolled' were loaded with '0', so there are no empty cells left.

The next operation concerned the column 'course students enrolled', and integer type columns were transformed. The source data is a string which sometimes contains suffixes like 'k' for thousands or 'm' for millions. The source data was transformed by replacing 'k' with '000', 'm' with '000000' by regular expression, and then the column was turned to float decimals. This operation is required to make correct numerical work with this column later.

Then, the categorical variables were encoded. The text values of the columns 'course organization', 'course title', 'course Certificate type', and 'course difficulty' were transformed into numerical values one-hot encoded with a LabelEncoder. This part allows us to keep the categorical nature of the data but transformed them into the format which can be processed by the models. After that, the target variable 'course difficulty' was separated from the features set.

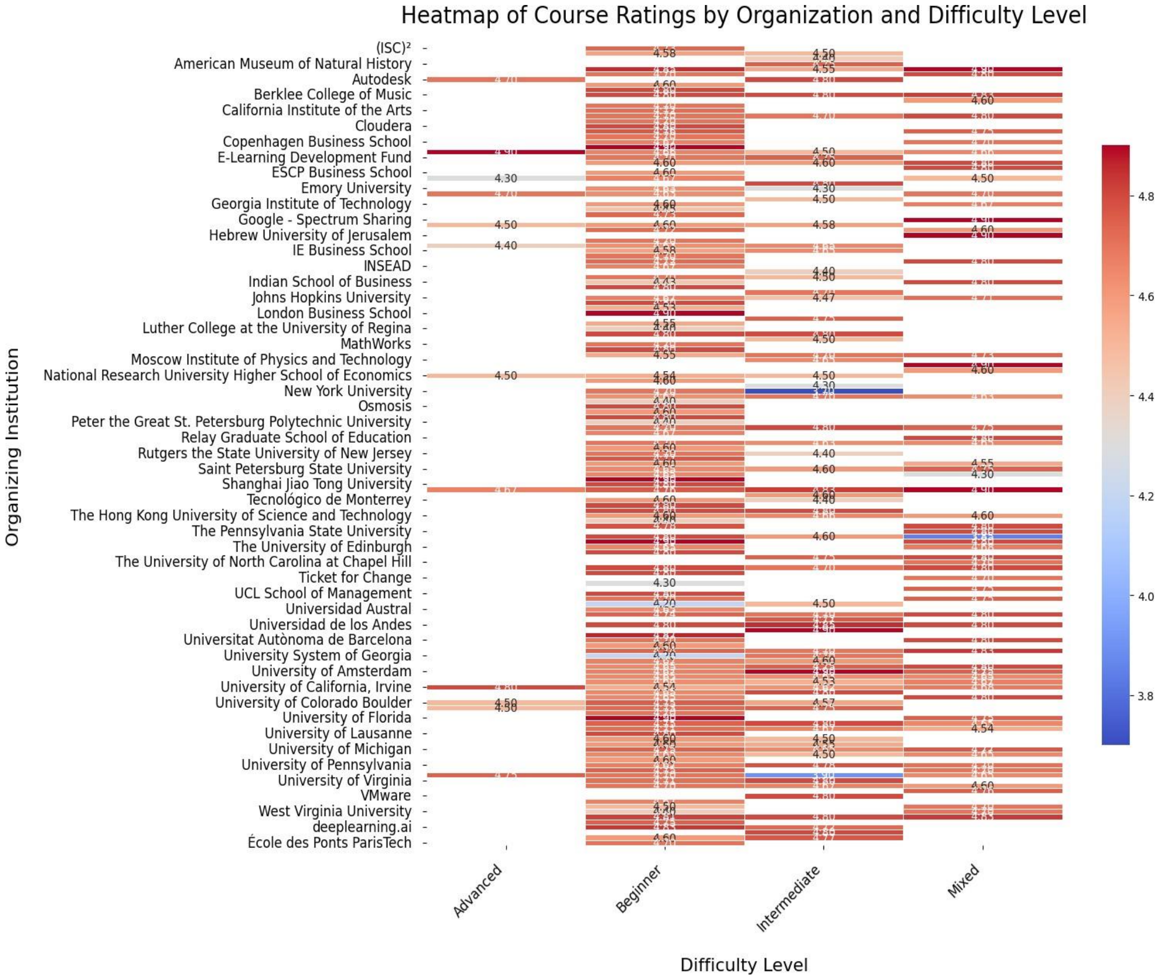

Figure 7.Heatmap of Course Ratings by Organization and Difficulty Level

The scaling data with StandardScaler was applied to standardize the values of all features. The values of all features were transformed into values with mean of 0 and standard deviation of 1, diminishing the effect of columns with the broader extent. SMOTE was applied to balance the dataset due to the possible imbalance between classes, especially in the context of the proportion between levels of course difficulty. SMOTE synthesizes synthetic samples of the unsufficiently represented class and balances the dataset.

Figure 8 above represents a bar chart that illustrates the distribution of course difficulty levels after applying the Synthetic Minority Over-sampling Technique. The figure shows the distribution of course level difficulty after balancing the dataset. In this case, course levels difficulty are denoted 0, 1, 2 and 3. In each bar, the number of courses in level difficulty is almost equal with each difficulty level having 500 courses. As observed, the balancing technique worked well to correct the initial class imbalance, such that there is an even representation of the number of courses of varying difficulty during the training processes of the machine learning model. This is critical in ensuring a strong and unbiased model ideal for predicting course difficulty with neutral coverage of all categories.

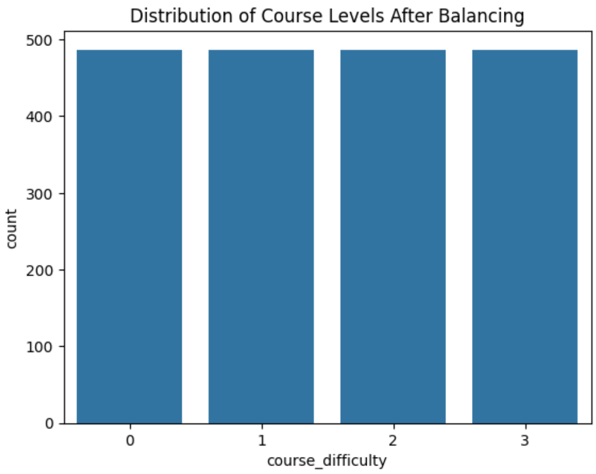

Figure 8. Distribution of Course Levels After Balancing

The splitting procedure was applied to split the dataset into training and testing datasets. Initially, the balanced dataset was split into 90% training set and 10% temporary set, which was split again to produce a small test set and additional validation set. The final test set was balanced. The small random noise factor 0.02 was added to the test features to replicate the real-world conditions better.

D. Modeling

In the modeling step of our study, we utilized three proven and effective machine learning algorithms in the modeling step of our study. These algorithms are Gradient Boosting [17], Random Forest [18], and XGBoost [19].

Gradient Boosting is an ensemble learning approach that fits models in a sequence; each new model corrects errors made by earlier models. It trains a new model based on residual errors made by the available ensemble.

Concretely, Gradient Boosting combines the predictions of weak learners such as decision trees while minimizing a loss function. The continuous model retraining based on the errors of earlier trained ones helps enhance the accuracy of the model and mitigates overfitting or underfitting by focusing on misclassified difficult cases [20].

Random Forest is an ensemble learning algorithm that creates a forest with numerous decision trees while training the model. In a classification context, the Random Forest model will predict the highest mode of the class; while a regression model, the mean of individual trees. Random Forest is an effective model due to its use of multiple decision trees with random samples of the dataset and features. Aspects of Random Forest minimize the risk of underfitting the model and improve model robustness or generalization when making predictions. Each tree vote is considered, with the popular class noted as the final prediction. The algorithm is useful when capturing a broad range of feature interactions while minimizing variance [21].

XGBoost is a specialized Gradient Boosting implementation using an advanced efficient unified framework. By optimizing the gradient boosting algorithm, XGBoost uses a more regularized model formalization to avoid overfitting. It also incorporates system optimizations and algorithmic features to improve model performance. The high-speed algorithm builds an ensemble of trees sequentially. Each builds a model to correct the errors demonstrated by its predecessor, although there are other elements like tree pruning, parallel computing, and missing values. Using these three models, our goal was to benefit from the unique strength of the independent models to make accurate and dependable suggestions about course difficulty [22].

Gradient Boosting and XGBoost work best in enhancing accuracy. We also used Random Forest to improve the robustness and generalization of the model by basing predictions on multiple trees. The integration of multiple approaches offers a comprehensive platform to handle the complications of predicting course difficulty on a scale.

E. Evaluation Metrics

To evaluate the performance of our multiclass classification models, we used the following key metrics: the confusion matrix, accuracy, precision, recall, and F1-score. These metrics provide a well-rounded view of how well the models are performing. They include the following:

The confusion matrix is a table that summarizes the classification model's performance. It includes the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). In the case of multiclass classification, the confusion matrix becomes a matrix with the same number of rows and columns to indicate the classes predicted. The confusion matrix helps us determine where the model is performing well and where it falls short.

- Accuracy is the proportion of correctly predicted data points. and is calculated as:
- $$ACC = \left(\frac{TP+TN}{TP+TN+FP+FN}\right) \quad (1)$$

In the context of multiclass classification, accuracy gives an overall performance measure but can be misleading if the classes are imbalanced.

- Precision is the ratio of true positive predictions to the total predicted positives and is calculated as:

$$PREC = \left(\frac{TP}{TP+FP}\right) \quad (2)$$

Precision indicates the accuracy of the positive predictions made by the model, which is critical when the cost of false positives is high.

- Recall (or Sensitivity) is the ratio of true positive predictions to the total actual positives and is calculated as:

$$REC = \left(\frac{TP}{FN+TP}\right) \quad (3)$$

Recall measures the model's ability to correctly identify all relevant instances, which is important in scenarios where false negatives are particularly costly.

- F1-score is the harmonic mean of precision and recall, providing a single metric that balances both concerns. It iscalculated as:

$$F1 = \left(\frac{2 \times REC \times PREC}{REC+PREC}\right) \quad (4)$$

The F1-score is especially useful when dealing with imbalanced classes, as it considers both false positives and false negatives.

3- Results and discussion

The Gradient Boosting Classifier predicts the course difficulty level with great performance, having an overall accuracy of 0.937. Likewise, it demonstrates high precision, recall, and F1-score, indicating the model's high capacity in solving the multiclass classification problem. In more detail, the classification report presented in Figure 9 reveals that the model performance level is very good for all classes, with precision and recall ranging between 0.90 and 0.98 and F1- scores ranging from 0.90 to 0.98. Class 0 and Class 3 have outstanding performance with records of 0.98 and a record of 0.99 for Class 0 and 3 and 0.98 for Class 3. The recall for Class 2 is lower than the others, recording 0.88, implying that there is a possibility that the model might have some difficulties in detecting those cases. However, the overall macro averages for precision, recall, and F1-score are 0.94, whereas for the weighted averages, they are 0.94, highlighting that the model gives accurate and balanced predictions on which level of difficulty of the course. Hence, it can be concluded that the Gradient Boosting Classifier provides robust predictive power and validity for this problem.

Classification Report:

	precision	recall	f1-score
0	0.98	0.99	0.98
1	0.91	0.90	0.90
2	0.96	0.88	0.92
3	0.90	0.98	0.94
accuracy			0.94
macro avg	0.94	0.94	0.94
weighted avg	0.94	0.94	0.94

Figure 9. Classification report of Gradient Boosting

The confusion matrix built for the Gradient Boosting Classifier outlines the performance of the model in predicting the difficulty level of the course as demonstrated in Figure 10. It can be seen that the model predicts most categories mostly in line with the reality, which is signified by the large number of the diagonal elements. Firstly, it is apparent that the majority of the elements are located on the diagonal, which signifies that the model operates well. Secondly, Class 0, which is the Beginner course, has nearly perfect performance, with 438 instances predicted correctly and minimal predicted other classes. Class 1 Intermediate as well performs well, where 402 of 448 instances are well predicted, although there is a considerable amount of predictions into Class 2 and Class 3. Class 2 Mixed has 392 of 446 instances correctly predicted, with most of the predictions being confused with Class 1.

Lastly, Class 3 Advanced has the highest number of well-predicted values 447 of 454, with virtually no confusion with other classes. The described performance signifies the gradient boosting classifier's efficiency in predicting various course difficulty levels with high precision and recall for extreme classes. In addition, the highly accurate performance with minimal confusion in many instances signified classification possibilities made by the model.

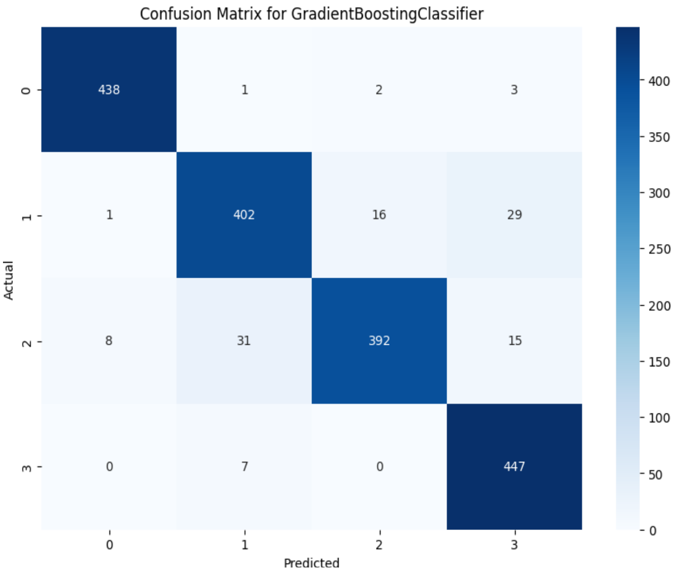

Figure 10. Confusion matrix of Gradient Boosting

The classification report of the Random Forest model also indicates an outstanding performance of outcomes in all classes with an accuracy of 0.98 ((Figure 11)). The presision, recall and F1-scores of the class are higher, which implies that the model shows higher performance to correctly classify instances. Class 0 has 1.00 precision, recall and F1-score implying that the model perfectly predict the outcomes of this class. Class 1 and 2 show relatively higher outcomes of presision, recall and F1-score above 0.94, implying that they are not misclassified. Class 3 has 0.92 precision, 1.00 recall and 0.96 F1-score, which still indicates that this model has a strong accuracy with a lower precision. The macro and weighted average of presision, recall and F1-score is 0.98 implying that the model performs consistently and equally to all the classes. This outcome shows that Random Forest is a robust model that reliably predicts the course's difficulty level accurately.

Classification Report:

	precision	recall	f1-score
0	1.00	1.00	1.00
1	1.00	0.94	0.97
2	1.00	0.97	0.98
3	0.92	1.00	0.96
accuracy			0.98
macro avg	0.98	0.98	0.98
weighted avg	0.98	0.98	0.98

Figure 11. Classification report of Random Forest

The confusion matrix for the Random Forest Classifier indicates an exceptional ability to predict course difficulty levels, and most of its predictions coincide with the actual values as shown in Figure 12 below. Class 0, which is the Beginner class, has 442 correct predictions out of 444, which is to an extent an extremely stifled misclassification rate. Class 1, which is the Intermediate class, has 421 correct predictions out of 448, with a few proper misclassifications, mainly class 3. With respect to Class 2, the Mixed class, the score is impressive at 431 out of 446. Moreover, a few of the correct instances are misclassified, mainly to class 3. Finally, class 3, which is the Advanced class, reports a perfect class classification, with 454 cases out of 454 classified correctly. A high overall accuracy score with not many errors makes the model robust enough to distinguish satisfactorily well between the four course difficulty levels. Random Forest maintains an excellent recall and precision score throughout for all classes, which means that it is a proper model for multiclass classification, as proved by the confusion matrix.

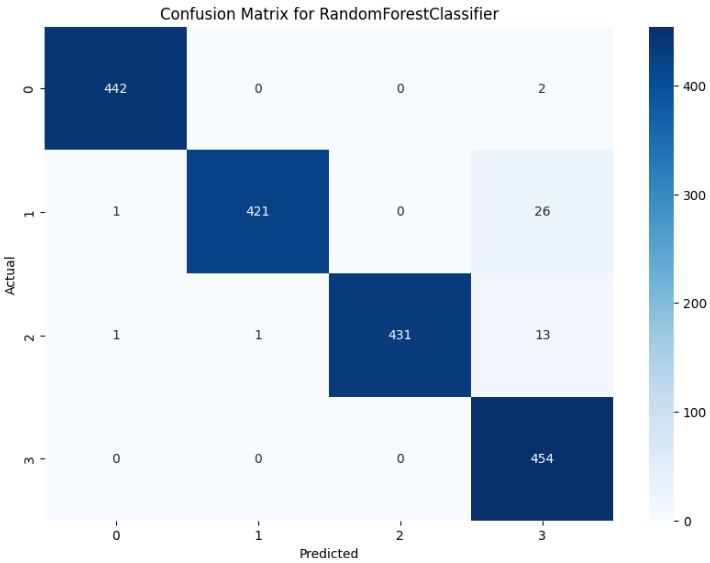

Figure 12. Confusion matrix of Random Forest

The classification report of the XGBoost model presented in Figure 13 vividly proves the high performance of this model in predicting the level of course difficulty, as the overall accuracy achieved equals 0.964. Indeed, the model demonstrates outstanding precision, recall, and f1-scores for all classes, and the macro and weighted average equal 0.96 for each metric. In particular, Class 0, or Beginner, reveals a precision and recall of 0.99, with an f1-score of 0.99, which implies near-perfect classification. Class 1, or Intermediate, also keeps high performance at precision of 0.96, recall of 0.94, and f1-score of 0.95. Meanwhile, Class 2, or Mixed, has a precision of 0.97, recall of 0.93, and f1-score of 0.95; each reflects strong predictive capabilities that contain exceedingly minimal errors. Class 3, or Advanced, likewise performs very well at precision of 0.94, recall of 1.0, and an f1-score of 0.97. Overall, these results testify to the robustness and dependability of the XGBoost model and its capacity to correctly forecast the level of course difficulty, act within the context of multi-class classification, and handle existing subtleties.

```
Classification Report:
              precision    recall  f1-score

           0       0.99      0.99      0.99
           1       0.96      0.94      0.95
           2       0.97      0.93      0.95
           3       0.94      1.00      0.97

    accuracy                           0.96
   macro avg       0.96      0.96      0.96
weighted avg       0.96      0.96      0.96
```

Figure 13. Classification report of XGBoost

The confusion matrix of the XGBoost Classifier depicted in Figure 14 highlights its excellent performance in the prediction of different course difficulties. The classifier has high accuracy within each class, as shown by equal sizes of correctly classified instances that follow the diagonal. For Class 0 which is Beginner, the classifier correctly classified 440 out of the available 444 instances with only 4 misclassification incidences. Class 1, which is Intermediate, also has very high performance as 421 out of the available 448 instances were classified correctly, but some were misclassified as classes 2 and 3. Class 2, which is Mixed, had 414 out of 446 correct predictions with a fair amount of misclassification especially to Class 1. For Class 3, which is Advanced, the performance was near-perfect as 453 out of the available 454 instances. Overall, the XGBoost Classifier accurately differentiates between different levels of difficulty, as shown by Fig. 13, with a high level of precision and recall, and at the same time minimizing misclassification errors. The balance in the performance across all classes presented in the confusion matrix confirms the effectiveness of this classifier in this multi-class classification task.

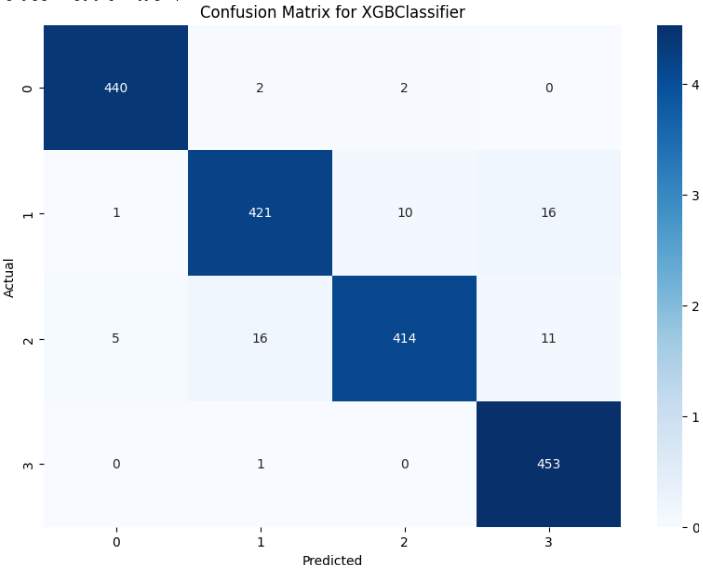

Figure 14. Confusion matrix of XGBoost

Comparison charts presented in Figure 15 explain a thorough performance metric for three classifiers of Gradient Boosting, Random Forest, and XGBoost in four key evaluation measures being Accuracy, Precision, Recall, and F1 Score. Each of the provided models demonstrates a significant degree of efficiency in classifying course difficulty levels. However, specific performance metrics were somewhat different.

Accuracy. XGBoost leads all three classifiers with high accuracy, followed by Random Forest and Gradient Boosting. This score indicates that XGBoost has a relatively correct proportion of predictions.

Precision. Positive predictions score comparison between the classifiers indicated that the three classifiers had a relatively similar performance with a small variance from one another. Recall. XGBoost was somewhat superior to the classifiers, but in general, the score performance was relatively similar to all the classifiers.

F1 Score. XGBoost once again showed a little advantage on performance rather than Gradient Boosting and Random Forest. This score affirms that both XGBoost and Random Forest have a similar balanced score distribution compared to Gradient Boosting. Therefore, the comparative measure indicates that XGBoost is slightly better performing than Gradient Boosting and very closely related to Random Forest. Gradient Boosting classifier is marginally inferior to Random Forest and is thus rated in position against the three classifiers. Summary: The results presented in the comparison analysis show that while all three classifiers of Gradient Boosting, Random Forest, and XGBoost are equally effective in predicting course difficulty as a multiclass classification problem, XGBoost slightly outperformed the former two in each metric. This high performance of XGBoost can be attributed to sophisticated optimization measures employed, such as regularization and handling of missing values, large dataset and high dimensionality, and others. XGBoost has an advanced form of gradient boosting with a regularization term that helps mitigate overfitting and enhance generalization abilities. The algorithm also employed parallel processing and distributed computing, thus rendering XGBoost a competitive algorithm of choice compared to Gradient Boosting and Random Forest. Therefore, for an application that requires a high prediction precision metric, XGBoost is more preferred. Random Forest closely follows XGBoost, while Gradient Boosting comes last.

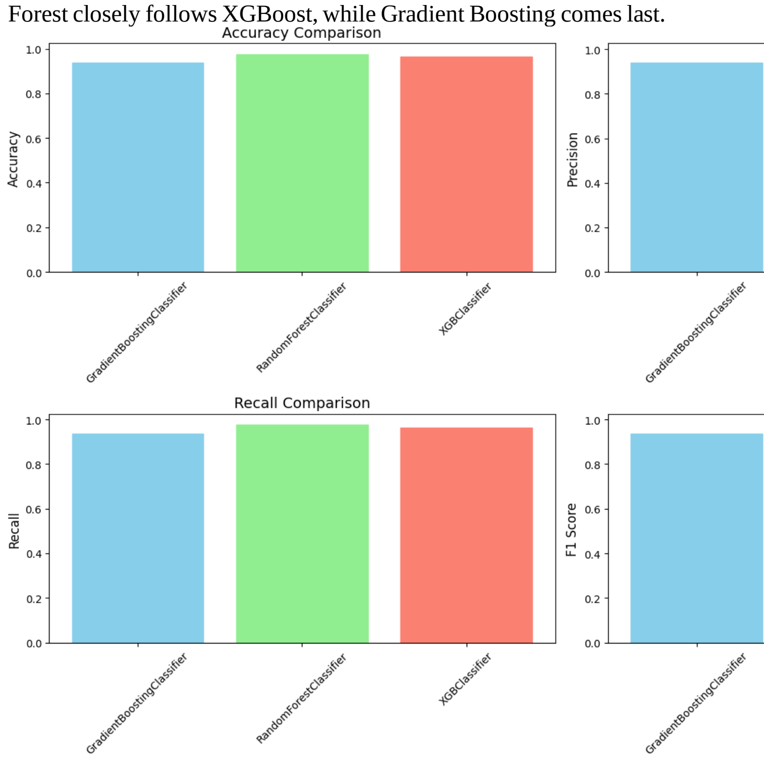

Figure 15. Comparison of the performances of the models

Conclusion

In conclusion, the study exemplifies the usefulness of machine learning models in predicting the level of course difficulty with high accuracy in online education. From the overall assessment, it was identified that XGBoost consistently demonstrated better performance than both Gradient Boosting and Random Forest classifiers based on all the key metrics, including accuracy, precision, recall, and F1 score. The higher efficiency of XGBoost can be linked to its innovative optimization techniques, including regularization, effective treatment of missing values, and parallel processing. These findings highlight the potential of using highly reliable machine learning algorithms to improve the effectiveness and personalization of online learning platforms. In future research, it is anticipated that additional features, such as student engagement and course content analysis, will be integrated to boost the prediction accuracy. In addition, enhanced performance can be achieved by evaluating the functionalities of deep learning models and hybrid approaches. As an extension, the potential of real-time predictions can be explored, making it possible to switch the recommendations on the fly from one course to another based on the dynamics of learner performance data, creating a more responsive and adaptive education system.

References

[1]- H. Ulum, “The effects of online education on academic success: A meta-analysis study,” *Educ. Inf. Technol.*, vol. 27, pp. 429–450, Jan. 2022.

[2]- N. P. Morris, M. Ivancheva, T. Coop, R. Mogliacci, and B. Swinnerton, “Negotiating growth of online education in higher education,” *Int. J. Educ. Technol. High. Educ.*, vol. 17, pp. 1–16, Dec. 2020.

[3]- R. Lourdusamy and P. Magendiran, “A systematic analysis of difficulty level of the question paper using student’s marks: a case study,” *Int. j. inf. tecnol.*, vol. 13, pp. 1127–1143, June 2021.

[4]- P. M, S. G, V. V, and G. J. Chowdary, “Course Difficulty Estimation Based on Mapping of Bloom’s Taxonomy and ABET Criteria,” *arXiv*, Nov. 2021.

[5]- I. T. Sanusi, S. S. Oyelere, H. Vartiainen, J. Suhonen, and M. Tukiainen, “A systematic review of teaching and learning machine learning in K-12 education,” *Educ. Inf. Technol.*, vol. 28, pp. 5967–5997, May 2023.

[6]- M. A. Ayanwale, R. R. Molefi, and S. Oyeniran, “Analyzing the evolution of machine learning integration in educational research: a bibliometric perspective,” *Discov. Educ.*, vol. 3, pp. 47–21, May 2024.

[7]- M. Adnan, A. Habib, J. Ashraf, S. Mussadiq, A. A. Raza, M. Abid, M. Bashir, and S. U. Khan, “Predicting at-Risk Students at Different Percentages of Course Length for Early Intervention Using Machine Learning Models,” *IEEE Access*, vol. 9, pp. 7519–7539, Jan. 2021.

[8]- M. Yağcı, “Educational data mining: prediction of students’ academic performance using machine learning algorithms,” *Smart Learn. Environ.*,vol. 9, pp. 11–19, Mar. 2022.

[9]- A. A. Mubarak, H. Cao, and W. Zhang, “Prediction of students’ early dropout based on their interaction logs in online learning environment,” *Interactive Learning Environments*, July 2022.

[10]-Engr. S. Bhutto, I. F. Siddiqui, Q. A. Arain, and M. Anwar, “Predicting Students’ Academic Performance Through Supervised Machine Learning,” in *2020 International Conference on Information Science and Communication Technology (ICISCT)*, pp. 08–09, IEEE.

[11]-R. Hasan, S. Palaniappan, S. Mahmood, A. Abbas, K. U. Sarker, and M. U. Sattar, “Predicting student performance in higher educational institutions using video learning analytics and data mining techniques,” *Applied Sciences*, vol. 10, no. 11, p. 3894, 2020.

[12]-L. Kemper, G. Vorhoff, and B. U. Wigger, “Predicting student dropout: A machine learning approach,” *European Journal of Higher Education*, Jan. 2020.

[13]-W. Villegas-Ch, M. Román-Cañizares, and X. Palacios-Pacheco, “Improvement of an Online Education Model with the Integration of Machine Learning and Data Analysis in an LMS,” *Appl. Sci.*, vol. 10,p. 5371, Aug. 2020.

[14]-A. Alam and A. Mohanty, “Predicting Students’ Performance Employing Educational Data Mining Techniques, Machine Learning, and Learning Analytics,” in *Communication, Networks and Computing*, pp. 166–177, Cham, Switzerland: Springer, Sept. 2023.

[15]-E. Camizuli and E. J. Carranza, “Exploratory data analysis (eda),” *The encyclopedia of archaeological sciences*, pp. 1–7, 2018.

[16]-S. Garc´ıa, S. Ram´ırez-Gallego, J. Luengo, J. M. Ben´ıtez, and F. Herrera,“Big data preprocessing: methods and prospects,” *Big data analytics*, vol. 1, pp. 1–22, 2016.

[17]-C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, “A comparative analysis of gradient boosting algorithms,” *Artificial Intelligence Review*,vol. 54, pp. 1937–1967, 2021.

[18]-A. Parmar, R. Katariya, and V. Patel, “A review on random forest: An ensemble classifier,” in *International conference on intelligent data communication technologies and internet of things (ICICI) 2018*, pp. 758– 763, Springer, 2019.

[19]-T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, R. Mitchell, I. Cano, T. Zhou, *et al.*, “Xgboost: extreme gradient boosting,” *R package version 0.4-2*, vol. 1, no. 4, pp. 1–4, 2015.

[20]-A. Natekin and A. Knoll, “Gradient boosting machines, a tutorial,” *Frontiers in neurorobotics*, vol. 7, p. 21, 2013.

[21]-P. Probst, M. N. Wright, and A.-L. Boulesteix, “Hyperparameters and tuning strategies for random forest,” *Wiley Interdisciplinary Reviews: data mining and knowledge discovery*, vol. 9, no. 3, p. e1301, 2019.

[22]-T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,”in Proceedings of the 22nd acm sigkdd international conference onknowledge discovery and data mining, pp. 785–794, 2016.

[23]-Siddharth M., "Coursera Course Dataset," *Kaggle*, Updated 4 years ago. [Online]. Available: https://www.kaggle.com/datasets/siddharthm1698/coursera-course-dataset. [Accessed: Aug. 26, 2024].