A Comparative Analysis for GPA Prediction of Undergraduate Students Using Machine and Deep Learning

Ibrahim Alnomay, Abdullah Alfadhly*, and Aali Alqarni

King Abdulaziz City for Science and Technology KACST, Riyadh, Saudi Arabia

Email: ialnomay@kacst.edu.sa (I.A.); fadhly@kacst.edu.sa (A.A.); aaalqarni@kacst.edu.sa (A.A.)

*Corresponding author

Manuscript received July 21, 2023; revised August 8, 2023; accepted August 29, 2023; published February 18, 2024

***Abstract*—Recently the research field of machine learning has experienced a huge rise in popularity and growth. Machine Learning (ML) is a way of improving computational prediction models by allowing the computer to generate its own algorithm to predict outcomes, based on an existing dataset. In this paper, we demonstrate the application of Machine Learning to enhance the educational processes. We implemented regression and supervised learning techniques on data from King Saud University, Riyadh, Saudi Arabia, to construct a predictive model for student performance. This allows for timely interventions in students' academic paths. We utilized extensive and diverse course records, encompassing several academic years and programs, to conduct a comparative analysis of various Machine and Deep Learning methodologies, assessing their efficacy through performance metrics. The developed ML/DL algorithms use Grade Point Averages (GPAs) of courses and semesters as explanatory features to predict the student's final GPA, which is the target value of the models. Based on the results, the linear and bagging regression models have the best Mean Absolute Error (MAE) performance metric. To ensure there will be enough time for academic intervention, data of early courses and semesters are used.**

***Keywords*—deep learning, Grade Point Average (GPA) prediction, machine learning, student performance**

I. Introduction

The application of machine learning (ML) to the prediction of student grades has been studied intensively in recent years. Many researchers [1–6] have proposed machine learning models to predict Grade Point Averages, with their findings showing promising outcomes. To predict student grades in diverse outcomes requires collecting information from a variety of sources. However, a significant challenge persists in predictive modelling for unbalanced datasets. Addressing this challenge has become a focal point, necessitating further exploration and study. This is because unbalanced datasets are skewed towards one class, making it difficult for the model to learn to predict the minority class. Jishan *et al*. [7] attempted to address this issue by employing oversampling techniques. They specifically utilized the Synthetic Minority OverSampling Technique (SMOTE) to handle unbalanced datasets and enhance the experimental outcomes of their study.

In this paper, we investigate the power of machine learning in predicting students' expected Grade Point Average (GPA). In order to perform this task we firstly collected a dataset of students graduated in a range of disciplines from the largest university in Saudi Arabia, King Saud University. The total number of students exceeded 12,499 student, number of graded courses is 766,278 grades. The dataset include both male and female. Finally, practical considerations had to be taken into account during the data processing stage, such as the fact that many courses were transferred from different universities, and courses were changed during the student academic life cycle, and courses were taken more than once, all of that have been taken in consideration in the cleansing and data processing stage.

The contribution of our work is in the ability to deal with large academic dataset that span more than 20 years, serving multiple academic plans. Most of the research that predict the student performance, uses a specific plan to develop their machine learning algorithms. This may not be the optimal solution in situations where big data are available to university administration and they need a way to deal with such dataset. In our work, we devised multiple steps to convert the currently available data to tables with unified features that can be fed to the machine learning algorithms. Moreover, since we did not depend on a specific academic plan, we derived a way from the course code in order to know when these courses are likely taken by students during their academic study. This paper is a part of a larger project to achieve the goal of new generation of Smart universities, by which Artificial Intelligence (AI) will play a greater rule in the education system.

In the following sections of the paper, we highlight works done to utilize AI in the education sector in Section II. In Section III, we described our research methodology, consisting of preprocessing data, features selection, and machine learning development. Our results are discussed in detail in Section IV.

II. Related Work

Jishan *et al*. [7] employed SMOTE approaches to enhance the accuracy of learners' absolute grade prediction. A number of classification methods including Decision Tree (DT), Naive Bayes (NB), and Neural Network (NN), were used to divide learners' grades into five groups: A, B, C, D, and F. They found that using SMOTE, NN and NB exceeded different approaches with the same elevated accuracy of 75%. However, as compared to NN, NB performed better since the best period to apply the prediction instances is immediate.

Nowadays, educational settings can create a large amount and diversity of data, such as those linked to students' school records, evaluation files, curriculum reports, and records of e-mail contacts between students and professors. Educational environments create a lot of data, which is useful for decision-making and improving the learning process. This can be achieved through analysis of student data, and their behaviour, contentment, and performance [8]. Data Mining (DM) approaches may be used to extract information from this data and, as a result, improve the quality of education [1].

doi: 10.18178/ijiet.2024.14.2.2050

Polyzou and Karypis [2] established a strategy for forecasting future course grades acquired from the University of Minnesota. The findings showed that Matrix Factorization (MF) and Linear Regression (LinReg) achieved better than current conventional approaches based on the proposed methodologies. The author also discovered that using a course-specific dataset can help forecast future course grades with greater accuracy.

Another study used MF, Collaborative Filtering (CF), and Restricted Boltzmann Machines (RBM) approaches to predict student grades in various courses using 225 actual data from undergraduate students [3]. They conclude that when compared to MF, utilizing CF did not have great predictive power, specifically when the dataset is sparse. However, their observations showed that the suggested RBM delivers effective instruction and greater prediction accuracy than CF and MF.

Another research [4] created a predictive model for predicting students' exam results in the curriculum at the beginning of the semester. They used Waikato Environment for Knowledge Analysis (WEKA) to compare eleven machine learning algorithms in five categories. They used attribute selection for details preparation to decrease high dimensionality and imbalanced data. The authors used SMOTE to equalize the instance's distribution in three distinct classes.

Al-Barrak *et al*. [5] discovered classification methods to predict the Grade Point Average (GPA) of students based on their grades in preceding studies using the DT. They used 236 students who completed from King Saud University's Computer Science College in 2012. They discovered the DT classification algorithm that may detect early indicators and extract valuable information for students based on their GPA, which can help them to improve their performance.

Abana [6] used several DT algorithms to predict a student's grade performance. Cross-checking was utilized for accessing the prediction model's interpretation. According to the statistics, Random Tree (RT) had a maximum accuracy of 75.188 %, which was superior to other algorithms.

The prediction model's accuracy may be enhanced by increasing the experiment number and characteristics of the dataset.

In Ahmad *et al*.'s [9] research, they established a framework for forecasting student academic achievement. The study employed 399 records of students only from the departmental registry throughout the courses of eight years of admissions, which included student demographics, past academic histories, and family background information. In comparison to DT and NB, the Rule-Based (PART) model was shown to be the most accurate, with 71.3 % accuracy. However, due to the incomplete and incorrect values in the dataset, adopting a limited sample size affected the accuracy of this study.

From 2006 through 2015, Anderson [10] conducted experimental research on 683 students at California State University's Craig School of Business, using a machine learning approach system to predict academic achievement. The best classifier, according to the study, is Support Vector Machine (SVM). It regularly beats a simple average strategy that optimizes each data class with the lowest error rate. The outcome for a large dataset might be different due to huge changes in the structure and format of the historical grade information.

Several works can be found in the literature which identify various educational issues. Depending on the eventual user's perspective, these works have varied goals (students, instructors, administrators, or other stakeholders). Some of these intended works will be presented in the sequel. The scope of this study does not allow for a comprehensive examination of these works; nevertheless, further information may be found in [11, 12].

Iam-On and Boongoen [13] have addressed a different educational issue, concentrating on a critical issue in higher education: student dropout. Early discovery of susceptible students, as described in this paper, can lead to the success of any engagement approach. Academic and administrative help would be offered to at-risk students to boost their chances of finishing the course. Based on students' pre-university traits, admission data, and first academic performance at university, the proposed work seeks to reveal intriguing patterns that might help forecast students' performance and dropout. The authors presented a new development in the social approach to increase the accuracy of traditional classifiers and, as a result, optimize students' attention while also proposing courses depending on their progress. With the advancement and widely usage of on-line education, machine learning is playing a crucial role in improving the quality of teaching. The education sector, in general, could benefit in different aspects [14] among them, which are: 1—students expected performance, 2—students best suitable field of study, 3—student future career, and in many cases 4—the teachers performance. The recent advancement of E-learning and the widely acceptance of Learning Management Systems (LMS) give deep insight into the date collected for the learning activities, providing more detailed information than simply the grades and gender of the students. A deep analysis of the performance of e-learning systems is now with the help of machine learning is more accredited and accessible.

In [15], the authors utilized a plethora of studies in Educational Data Mining (EDM) for predicting a student's academic success at graduation time, investigating which of the individual course grades or grade averages is more relevant for predicting student graduation academic performance. Although both types of data are interchangeably used in the literature, there is no study comparing the performance of EDM models using grade averages vs. individual course grades. It is unknown when and how to use these two college performance representations to attain best predictive power. To elucidate this matter, a comprehensive set of experiments were conducted on the recent student data compiled from the second author's college.

III. Research Methodology

Generally, in machine learning research, two phases are conducted. First, preparing and pre-processing the data. Second, applying machine learning algorithms in order to figure out the one which has the best performance among them.

A. Data Description

The data under investigation are the official records of more than 12,000 students graduated from King Saud

University in two colleges (College of Engineering and College of Computer) and 35 departments as shown in Table 1. For each graduate student the dataset consist of grades of all completed courses based upon an official curriculum. Graduated GPA is also available for each student. Moreover, transferred courses are also highlighted and the student gender is also identified. Table 2 shows all features contained within the dataset.

Table 1. Dataset description

Variable	Value
Total students	12,499
Studied semesters	73
Years	20
No of Departments	35
No of Graded Courses	766,278

Table 2. Dataset features

Feature	Type	Example
Student ID	Numerical	222197475
Gender	Binary	Female, Male
College	Nominal	Engineering
Department	Nominal	Electrical Engineering
Specialization	Nominal	Power
The university transferred from	Nominal	KFUPM
Cumulative GPA	Numerical	3.5

B. *Analysis*

We have analyzed in two categories. First, we used GPAs obtained by students in courses as predictors of the final GPA. We utilized courses of level one and two given the fact that it is these courses are usually taken by students early on in their academic plan. We will show later how we determine the course level. Second, we studied the effect of an accumulated semester's GPA in predicting the final GPA. We started with one semester and added up one more at a time until we reached the fourth semester.

C. *Data Preprocessing and Cleansing*

The original dataset is composed of two large files. The first file has all the metadata of all students, and the second file has all graded courses. We construct one table for each academic department. Each row in that table represent one student, and each column represent a graded course.

D. *Feature Selection*

Due to the many changes in the curriculum in almost all departments, different courses have been taken by the students. This complicates the selection for our ML models. In order to solve this, we selected only the common courses taken by all the students and ignore other courses.

In this study, we have selected the electrical engineering program to be the dataset used in this analysis. This is mainly due to the consistency of the curriculum over the studied period. We found 34 common courses that have been taken by all students over all of the examined semesters. If a student took a course more than once, we computed the grade average. Fig. 1 shows the steps taken during the data preprocessing stage. To be able to predict in an early stage, we further limited the selected features by choosing a lower course level. We developed a criteria based on the course code to determine the course level. We chose the course level based on the most significant digit in the numerical part of the course code as shown in Fig. 2. Specifically, we selected courses with a level 1 or 2 as these courses are typically taken in the early stages of the program and are more likely to be predictive of overall academic performance.

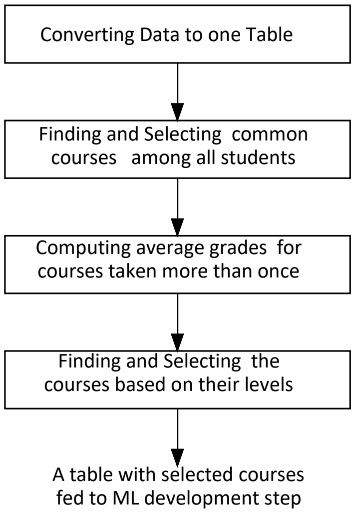

Fig. 1. Data preprocessing steps.

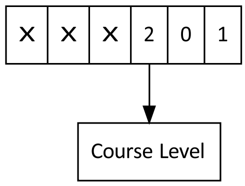

Fig. 2. Criteria of the level of the course

E. *Machine Learning Model Development*

As we identified in the previous section, changes of courses names over the years, adding new courses, and removing others cause some inconsistency in data. Therefore, during the preprocessing step, and before applying machine learning algorithms, we selected departments with the fewest changes so that more courses are used as features in predicting the final GPA.

To predict the final GPA with a high accuracy, we evaluated different machine learning regression algorithms including ensemble techniques and deep learning methods. These algorithms are:

- Linear Regression: The linear regression models a relationship between single or multiple explanatory variables (features) and a target variable. The equation of linear regression with multiple explanatory variables is defined as follows:

$$y = w_0x_0 + w_1x_1 + \ldots + w_mx_m \tag{1}$$

Here w_0 is the y axis intercept with $x_0 = 1$. The objective of the model solution is to find the best line fit that relate dependent variables to the output (target) value.

- Random Forest Regression: The random forest algorithm is a combination of multiple decision trees. Due to the randomness that helps reduce model variance, a random forest typically achieves superior generalization performance than an individual decision tree. Additionally, random forests are less sensitive to anomalies in the dataset and do not require extensive parameter tuning compared to other machine learning models . Typically, the only parameter that requires optimization? in random forests is the number of trees in the ensemble.

- Bagging Regression: The bagging algorithm is an ensemble technique similar to random forest. However, rather than using the same training set to model the individual classifiers in the ensemble, bootstrap samples (random samples with replacement) are drawn from the initial training set, which is why bagging is also known as bootstrap aggregating.
- Adaptive Boosting Regression: In boosting, the ensemble is comprised of extremely simplistic base models, also referred to as weak learners, that have a marginal performance advantage over random guessing. A typical illustration of a poor learner is a decision tree trunk. The central idea of boosting is to focus on training samples that are difficult to predict, i.e., to let weak learners learn from misclassified training samples to enhance the ensemble's performance. In contrast to bagging, the initial formulation of boosting, the algorithm utilizes random subsets of training samples drawn without replacement from the training dataset.
- Gradient Boosting Regression: Gradient Boosting Regression is similar to Adaptive Boosting Regression in terms of using weak learners. However, In Adaptive Boosting Regression, shift is done by up-weighting observations that were mispredicted before where Gradient Boost identifies difficult observations by large residuals computed in the previous iterations.
- Deep Learning: Deep Learning is based on multilayer artificial neural networks which consists of three layers: input layer, hidden layers, and output layer. The number of hidden layers between the input and output layer is configurable and they are created as a hyper-parameter to the problem that is to be solved. In this paper, we constructed an architecture for our deep neural networks which consist of one input layer, multiple hidden layers, and one output layer. The total size of the electrical engineering dataset used in the above algorithms is 231 entries. We allocated 70% for the training dataset, and the remaining for testing.

To evaluate the performance of the algorithms, two metrics are computed:

- Mean Absolute Error (MAE): MAE is defined as follows:

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \quad (2)$$

where y_i is the predicted value and x_i is the actual value. The number of samples in the dataset is n. MAE is commonly used to measure the performance of regression models.

- Accuracy: Accuracy is defined as follows:

$$Accuracy = \frac{True\ Positive + True\ Negative}{Number\ of\ Samples} \quad (3)$$

This metric is often used to measure the performance of supervised models.

IV. Results and Discussions

We conducted our prediction analysis in two categories: the first by using selected courses (features), and the second category by using semester GPAs. During the data exploratory phase, we computed the correlation coefficients between individual courses. Achieved GPAs courses A and B in Fig. 3 show linearity with the final GPA in contrast with GPAs of courses D and C where there is no linearity. This is an interesting finding showing the importance of examining the skill sets of each student before choosing their field of study.

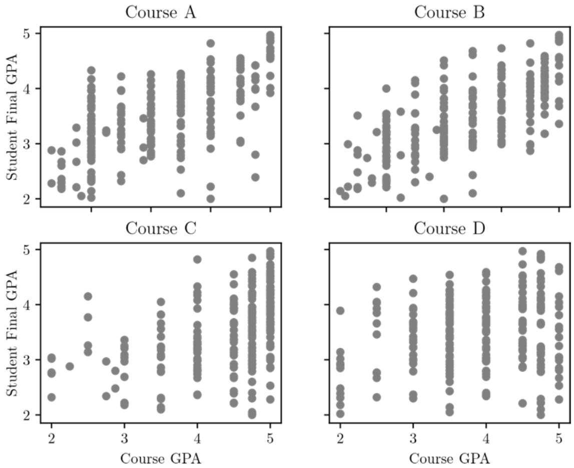

Fig. 3. Achieved GPA versus final GPA of students.

A. Prediction by Selected Courses

Table 3 shows the MAE results of the Linear Regression Model for different selected features (courses) based on the level of the course. As we mentioned earlier, it is better to choose features based on a lower level to predict the final results of the students in their early semesters.

Table 3. Linear regression MAE results based on the selected courses level

Level	No. of Selected Courses(Features)	MAE
First	8	0.37
Second	15	0.21
Third	28	0.12
Fourth	34	0.12

As shown in the Table 3 the MAE decreases if meaningful features are added before applying the machine learning mode. However, after Level 3, the MAE does not decrease since the added features do not add any improvement to the final prediction. Level 2 has a MAE of 0.21 and has a good warning time ahead before later semesters. Therefore, the remaining results in this section are based on this level.

The predicted and actual final GPA values for the first 15 students are shown in Fig. 4. The predicted values are remarkably close to the actual value, as indicated in the graph with a MAE of 0.21 as we discussed later.

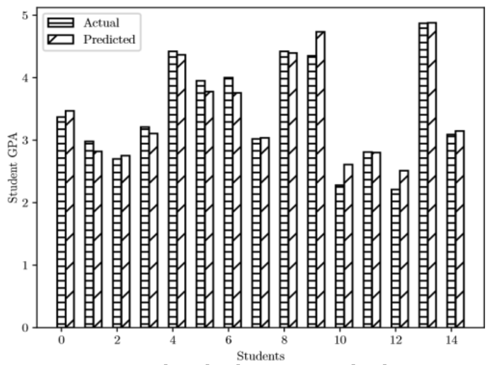

Fig. 4. Predicted values vs Actual values.

To reach a low MAE, we evaluated several regression algorithms. Table 4 shows the evaluated algorithms and MAE values obtained for each algorithm.

Table 4. Evaluated ML regression algorithms

Algorithm	MAE
Linear Regression	0.21
Random Forest Regression	0.27
Bagging Regression	0.21
Gradient Boosting Regression	0.24
Adaptive Boosting Regression	0.22

Per the above evaluation, Linear regression and Bagging regression have the best performance with a MAE value of 0.21. This is due to the high linearity between the target value and the attributes (courses) in the dataset. Tree based regressors perform better than linear regressors if a non-linear relationship exists.

To complete our analysis, we applied Deep Learning technique to our dataset. We created a deep neural network of input layer, two internal layer of size 64, and output layer.

The Deep Learning model has the worst performance analysis with MAE 0.49 as shown in Fig. 5.

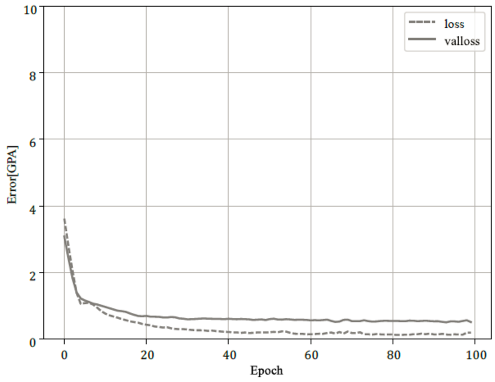

Fig. 5. Loss values of DL algorithm.

Finally, we converted the problem from regression to supervised Machine Learning by predicting if the final student grade fall into specific categories. The categories are Fair, Good, Very Good, Excellent. We used Decision Tree Classifier algorithm, and get an accuracy of 84%. The chosen max depth of the tree is three. If we did not specify a max depth, the algorithm will reach a perfect fit on the training set, but the performance is worse on the testing set, reaching an accuracy of 79%. Our result is comparable with the result in [16] where they used four categories as we did and achieved an accuracy between 48% and 86% for tree-based classifier. However, they used a larger dataset of 530 rows and 64 attributes.

B. Prediction Using Semester GPAs

In this analysis category, we predict the student final GPA by using the GPA of semesters. At first, we predict by only the first semester and see the results, then we add the GPA of the second semester to the selected features, and so on. Table 5 shows the results of the evaluated performance metrics.

Table 5. Prediction performance results using semester GPAs

Semester	MAE
First	0.47
Second	0.41
Third	0.36
Fourth	0.31

As shown in the table, we restrict the selected features until the fourth semester so that the prediction time is suitable to warn students in an early stage of their academic journey. However, the MAE of the first method, using selected courses is better for predicting the final GPAs.

V. Conclusion

In this work, we applied machine learning and deep learning algorithms to a dataset of undergraduate student records collected from king Saud university, Riyadh. The dataset covers twenty years and seventy three semesters, and therefore spans multiple academic plans. We first chose a department with fewer changes to its academic plan and then restricted selected features to the courses that are taken by all students. To reach the best performance, linear regression, ensemble techniques, and deep neural networks were evaluated and compared. The linear regression and bagging algorithms have the best performance with MAE of 0.2. Finally, we tested whether the performance was improved if the dataset was analyzed using supervised learning with definite categories. This second approach did not exceed the linear regression, resulting in an accuracy of 84%. We finally complemented our analysis by predicting by the GPA per semester. We started at the first semester and add the next semesters to the selected features till the fourth semester.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Ibrahim Alnomay conducted the research, did the literature review, and developed the used methodologies. Abdullah Alfadhly and Aali alqarni wrote the analysis code, developed and tested the machine learning and deep learning algorithms. All authors wrote the paper text, reviewed it, and approved the final version.

References

[1] M. Ezz and A. Elshenawy, “Adaptive recommendation system using machine learning algorithms for predicting student’s best academic program,” *Education and Information Technologies*, vol. 25, no. 4, pp. 2733–2746, Jul. 2020.

[2] A. Polyzou and G. Karypis, “Grade prediction with course and student specific models,” *Advances in Knowledge Discovery and Data Mining, ser. Lecture Notes in Computer Science*, Springer International Publishing, 2016, pp. 89–101.

[3] Z. Iqbal, J. Qadir, A. N. Mian, and F. Kamiran, “Machine learning based student grade prediction: A case study,” arXiv:1708.08744 [cs], Aug. 2017, arXiv: 1708.08744.

[4] I. Khan, A. Al Sadiri, A. R. Ahmad, and N. Jabeur, “Tracking student performance in introductory programming by means of machine learning,” in *Proc. 2019 4th MEC International Conference on Big Data and Smart City (ICBDSC)*, Jan. 2019, pp. 1–6.

[5] M. A. Al-Barrak, M. Al-Razgan, and King Saud University, Saudi Arabia, “Predicting students final GPA using decision trees: A case study,” *International Journal of Information and Education Technology*, vol. 6, no. 7, pp. 528–533, 2016.

[6] E. Abana, “A decision tree approach for predicting student grades in research project using weka,” *International Journal of Advanced Computer Science and Applications*, vol. 10, Jul. 2019.

[7] S. T. Jishan, R. I. Rashu, N. Haque, and R. M. Rahman, “Improving accuracy of students’ final grade prediction model using optimal equal width binning and synthetic minority oversampling technique,” *Decision Analytics*, vol. 2, no. 1, p. 1, Mar. 2015.

[8] M. W. Rodrigues, S. Isotani, and L. E. Zárate, “Educational Data Mining: A review of evaluation process in the e-learning,” *Telematics and Informatics*, vol. 35, no. 6, pp. 1701–1717, Sep. 2018.

[9] F. Ahmad, N. H. Ismail, and A. A. Aziz, "The prediction of students' academic performance using classification data mining techniques," *Applied Mathematical Sciences*, vol. 9, pp. 6415–6426, 2015.
[10] T. Anderson and R. Anderson, "Applications of ma-chine learning to student grade prediction in quantitative business courses," *Global Journal of Business Pedagogy*, vol. 1, no. 3, pp. 13–23, Dec. 2017, publisher: Institute for Global Business Research.
[11] B. Bakhshinategh, O. R. Zaiane, S. ElAtia, and D. Ipperciel, "Educational data mining applications and tasks: A survey of the last 10 years," *Education and Information Technologies*, vol. 23, no. 1, pp. 537–553, Jan. 2018.
[12] C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," *WIREs Data Mining and Knowledge Discovery*, vol. 10, no. 3, May 2020.
[13] N. Iam-On and T. Boongoen, "Improved student dropout prediction in Thai University using ensemble of mixed-type data clusterings," *International Journal of Machine Learning and Cybernetics*, vol. 8, no. 2, pp. 497–510, Apr. 2017.
[14] H. Luan and C.-C. Tsai, "A review of using machine learning approaches for precision education," *Educational Technology & Society*, vol. 24, no. 1, pp. 250–266, 2021, International Forum of Educational Technology & Society.
[15] S. Liu, X. Wang, M. Liu, and J. Zhu, "Towards better analysis of machine learning models: A visual analytics perspective," *Visual Informatics*, vol. 1, no. 1, pp. 48–56, Mar. 2017.
[16] N. Alangari and R. Alturki, *Predicting Students Final GPA Using 15 Classification Algorithms*, p. 12.