I. J. Mathematical Sciences and Computing, **2022, 2, 11-27**
Published Online June 2022 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijmsc.2022.02.02

A Comparative Analysis among Online and On-Campus Students Using Decision Tree

Rifat-Ibn-Alam
American International University- Bangladesh
Email: rifatibnalam50@gmail.com

Md. Golam Ahsan Akib
American International University- Bangladesh
Email: ahsanakib47@gmail.com

Nyme Ahmed
American International University- Bangladesh
Email: nymeahmedhimu@gmail.com

Syed Nafiul Shefat
American International University- Bangladesh
Email: syednafiul0621@gmail.com

Dip Nandi
Department of Computer Science, American International University- Bangladesh
Email: dip.nandi@aiub.edu

Received: 07 October 2021; Revised: 22 November 2021; Accepted: 10 December 2021; Published: 08 June 2022

Abstract: COVID-19 hit the world unexpectedly, forcing humans to isolate themselves. It has placed the lives of people in jeopardy with its fury. The global pandemic had a detrimental effect on the worlds' education spheres. It has imposed a global lockdown, with a negative impact on the students' lives. Continuing regular classes on-campus was out of the question. At that moment, online learning came to us as a savior. The quality of online education was yet to be tested on a large scale compared to regular schooling. Educational data mining is a modern arena that holds promise for those who work in education. Data mining strategies are developed to uncover latent information and identify valuable trends that can increase students' performance and, in turn, contribute to the improvement of the educational system in the long run. This research mainly aims to identify a comparative analysis of the students' academic performance between online and on-campus environments and distinguish the significant characteristics that influence their academic endeavors. The impact of the factors on the students' performance is visualized with the help of the Decision Tree Classification Model. This paper will assist in giving a good overview that influences the distinguished factors on students' academic performance. Moreover, educators will also be benefited from this paper while making any important decision regarding the educational activity.

Index Terms: Educational data mining, Data Analytics, Decision Tree, E-learning.

1. Introduction

COVID-19 has wreaked havoc on virtually every aspect of human activity. Not only has the outbreak increased the risk of infection-related death, but it has also had a significant impact on the global education system. While universities worldwide have adapted to academic changes brought about by the pandemic, some universities appear to be falling behind [20]. The universities had to conduct all the academic activities online such as taking the classes and lab remotely, taking the assessment, grading the students, and so on. It was entirely a new experience for both the students and faculty members. Switching to the online platform was mandatory. If they did not switch to the online platform, several problems would arise. One of the most significant setbacks is that the continuity of education would be interrupted, which will undoubtedly lead to a session jam; in general, the nation's overall growth would be at a standstill.

Performance analysis is necessary to evaluate the students' academic progress in an online and on-campus environment. It helps to navigate the process for the advancement of the quality of education. Though there are many differences between online and on-campus learning systems, a comparison is essential to determine the factors behind students' performance. This paper gives an analysis of students' evaluations in the online and on-campus semesters. It also distinguishes between online and on-campus students' performance in some particular core courses and the reasons behind their achieved marks. Various attributes will be considered while executing this comparison, giving a clear overview of the sectors that affect students' academic performance. Four distinct programming courses of the computing students were chosen for the study, allowing for coverage of students from freshman year to the final year of the undergraduate program. This paper is a part of a more extensive research, where in-depth analysis is also done on students' online academic performance. This research can be used as a light bearer if the educational environment needs to be shifted from online to on-campus or vice-versa. It can be used as a guide while making important decisions regarding academic matters, especially for computing students.

The paper is divided into five sections, which are arranged in the following order: The introduction is in section 1, followed by a quick overview of the overall scenario of the education system in section 2, an analysis of the collected data, comparison, and the findings of this paper in section 3, section 4 contains the discussion about the outcomes found in section 3, and lastly, section 5 consists of the conclusion.

2. Related Works

The entire planet is confronted with the greatest challenge of the century. Coronavirus has put the entire world at a standstill. It has affected education, economics, and some other sectors deeply. Especially in the education sector, it has some severe impacts. Due to the global pandemic, all the educational institutions were suddenly shut down. Since then, everything has become online, from classes to offices, grocery shopping everything. This form of life is called the "new normal". Schools and universities are conducting classes, exams, presentations, viva, etc., all online. At first, teachers and students became befuddled and uncertain about coping with the sudden situation that caused the closure of instructional activities. Still, everyone knew that the lockdown had taught them a lot about dealing with the emergence of such pandemics. It has paved the way for educational institutions to develop and embrace virtual learning [19]. The educational environment often uses data mining techniques to explore and evaluate student performance. Educational data mining is used to analyze educational or academic behaviors so that necessary measures can be taken to better the quality of education. Students' performance depends on several aspects such as personal, economic, social, and environmental behaviors [1]. [2] Many educational institutions use the experiment's outcome to generate a pattern of the students' behaviors and develop a new strategy to overcome the drawbacks of academic performance and improve the quality of education. [3] Classification is the way of developing a model that illustrates and categorizes data classes that are derived based on analyzing the training data (data object for the declared class labels). [4] There are various data mining techniques for classifying the students according to their overall performance, such as Naive Bayes, K-Nearest Neighbors, Decision Tree, Apriori algorithm, and many more. Clustering is similar to grouping objects in a particular cluster and divergent to other clusters' objects. [5] Various kinds of clustering algorithms are available such as K-means clustering, hierarchical clustering, mean shift clustering, density-based clustering, and so on. [6] used several classification algorithms such as C5.0, J48, CART, KNN, SVM, Naïve Bayes, and Random Forest on three different datasets from school, college, and virtually conducted classes. According to their findings, C5.0 and Random Forest performed best for all three datasets. [7] Conducted research on two online introductory computing courses (Introduction to Programming and Introduction to IT) from Open Universities Australia to evaluate the students' performance for a certain period. Their analysis was based on several attributes that may affect the overall students' academic performance. A comparative study was conducted on students' performance in face-to-face and virtual classes. It was found that with proper planning and execution, it is possible to ensure the quality of education in online classes [8]. Virtual Learning Environments (VLEs) play a significant role in the educational sector during the global epidemic. It has made it more convenient for students and course instructors to attend classes from home; in other words, it has removed the barrier of location. Students can join the classes from any part of the world [9]. Several factors were used to analyze the most significant factors in the students' academic performance in online classes. Here many aspects of student grades in a MOOC were examined. They found that the best variables were those related to exercises of the MOOC while analyzing the performance of the students [10]. [11] used four classification algorithms, J48, Random Forest, PART, and Bayes Network Classifiers, to classify the students from three different colleges in India. According to them, Random Forest was more accurate (99%) than the other techniques. Twenty-two factors were taken into account while conducting the research, such as students' gender, grades, study hours, class attendance, admission category, and so on. [12] Applied Naïve Bayes to classify and K-means Clustering to evaluate students' academic performance, and 98.866% of accuracy was obtained to forecast the students' academic performance. Here the dataset was taken from the UCI website. Three different decision tree algorithms (ID3, C4.5, CART) were used by [13] for classification. Compared to C4.5 and ID3, the CART algorithm gained the highest classification accuracy. Moreover, the research implies that qualitative factors have a significant influence on the students' academic performance. Several attributes were considered, such as Students' Parent' Qualifications, Living Location, Economic State, Friends & Relative Support, Resource Accessibility, Attendance, and Academic results. [14] Conducted research using three different algorithms (K-Nearest Neighbour, SVM, and Decision Tree) to evaluate students' performance, and while doing so, the authors achieved maximum accuracy with Decision Tree (95.65%). Various attributes were considered, such as Grade Point Average, Students' hometown, Major, Parents Job, and many more. [15] applied three different Educational Data Mining classification algorithms: Naïve Bayes, Neural Network, and Decision Tree for modeling and predicting students' performance, and better accuracy was observed using Naïve Bayes (85.7%). Multiple attributes were considered, such as GPA, Test Average, Assignment Submission, Participation Rate, Attendance, Lab Test, and Final Grades. Four different Educational Data Mining classification techniques such as Artificial Neural Network, Decision Tree, K-Nearest Neighbors, and Naïve Bayes were applied [16]. Among them, the Artificial Neural Network achieved the most accuracy (93.70%). However, when the number of classes increased, the decision tree was performing better than the rest. About 39 attributes were used, such as students' SGPA, gender, curriculum, assessment objective, course curriculum, admission policy, modern devices usage, internet speed, indoor and outdoor, medical facilities, and many more. All the data were collected from individual semesters of students' undergraduate programs. These data were taken by a survey that was a part of the Institutional Quality Assurance Program. This survey was organized by the University Grant Commission, Bangladesh. [17] Used four different Data Mining classification techniques: Neural Networks, Decision Tree, Support Vector Ma, chine (SVM), and Naive Bayes. Two different and independent data were compared in this paper: distance education and on-campus from Brazilian Public University on introductory programming courses. Among these four classifiers, the Decision Tree performed most accurately on both data sources. For classification, several attributes were used, such as age, gender, semester, the performance of the students' weekly activities, and so on. [18] Applied decision tree and clustering algorithms such as K-Means clustering. Dataset was taken from various courses of undergraduate programs from a particular university of Pakistan. Three types of decision trees were applied in between the range of 60.58-69.23% accuracy.

According to [13, 14], [16, 17, 18], we found out that the Decision Tree gives the most accurate results. However, for a limited number of classes, Decision Trees are simple to use. They are the most straightforward to describe and comprehend. The majority of people are familiar with hierarchical trees, and a simple diagram will aid in communicating the findings. Decision Trees are very easy to interpret. People can get a gist of the tree just by giving it a look. A Decision Tree is probably the best place to start if to classify any dataset. It'll give a good overview, and it'll make the classification understandable. Decision Trees have simple features to define the most critical dimensions, manage missing values, and deal with outliers. Therefore, previous researchers have used this classification technique to overcome several classifying problems using Decision Trees. It immensely helps the educators to make crucial decisions while formulating new ideas and their implementation. That is why we are going with the Decision Tree to conduct this research.

3. Data Collection and Analysis

This section consists of the collection of the datasets with a brief analysis. Here, the datasets are obtained from the X University's students' academic reports of 4 different courses in two different semesters. One of them was held on-campus, and the other one was conducted virtually. Several decision trees are also generated by using WEKA to identify the prominent factors that have significant impacts on students' academic performance. Then, a distinction is made among these two semesters based on the factors impacting students' academic performance. Weka's result summary for all courses is also compared. Table 1 demonstrates the sample size of the four courses in both online and on-campus environments of the computing students conducted on the four consecutive years of the undergraduate program.

Table 1. Sample of Population

Year	Course name	Fall 2020-21 (Online)	Fall 2019-20 (On-campus)
1st	Introduction to Programming (IP)	40	93
2nd	Object Oriented Programming- 1 (OOP1)	78	25
3rd	Object Oriented Programming- 2 (OOP2)	127	79
4th	Web Technologies (WT)	46	98

The courses' results, including the entire breakdown of the Mid Term, Final Term, and Overall Grade, are available here. The grades are given based on the marks obtained by the students. In the online semester, 50% of the Overall Grade comes from the mid-term, and the other 50% comes from the final term. On the other hand, in the on-campus semester, the Overall Grade consists of 40% of the mid-term and 60% of the final term. Here, the grades start from 'F' and go all the way up to 'A+.' The grading policy works in the following way-

Table 2. Range of the possible grades

Marks	90-100	85-89	80-84	75-79	70-74	65-69	60-64	50-59	0-49
Grades	A+	A	B+	B	C+	C	D+	D	F

A. *Introduction to Programming (IP) Fall 2020-21 (Online)*

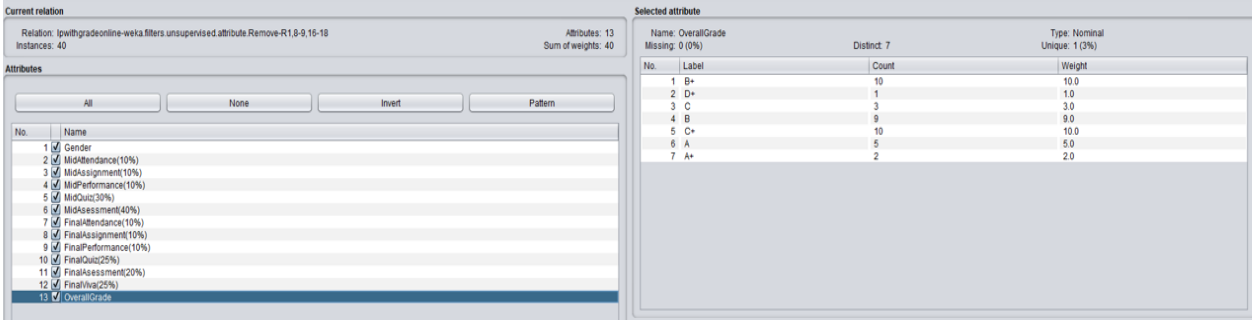
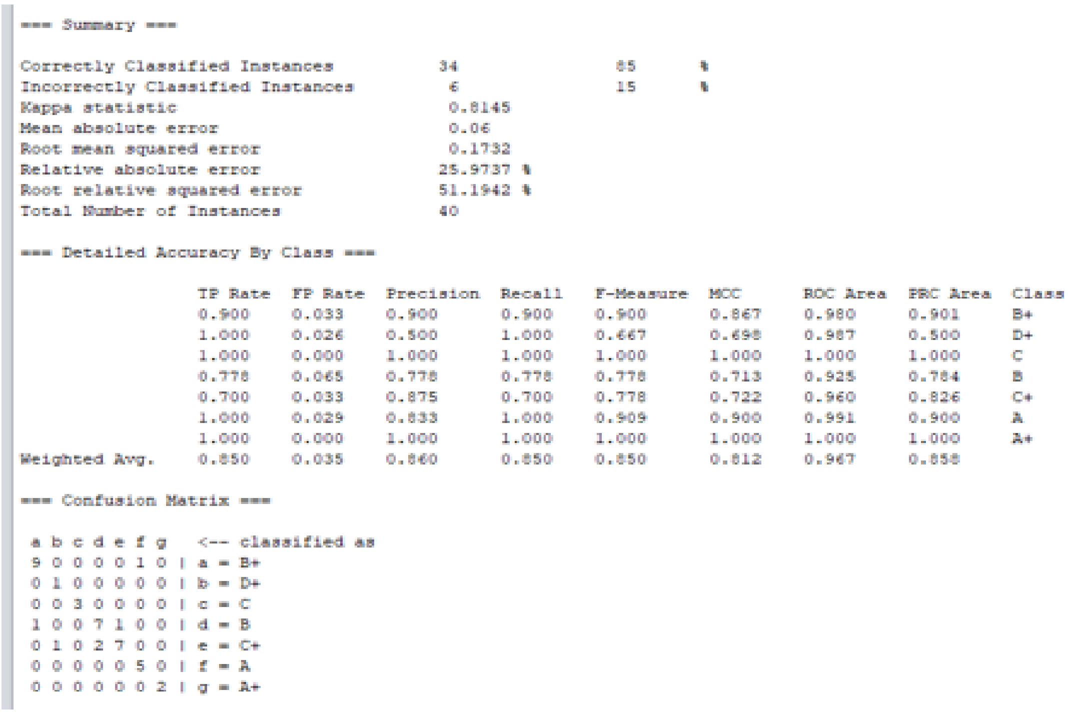


Fig.1. Attribute names and the quantity of IP Overall Grades, Fall 2020-21



```
=== Summary ===

Correctly Classified Instances          34               85      %
Incorrectly Classified Instances         6               15      %
Kappa statistic                          0.8145
Mean absolute error                      0.06
Root mean squared error                  0.1732
Relative absolute error                 25.9737 %
Root relative squared error             51.1942 %
Total Number of Instances               40

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.900    0.033    0.900      0.900    0.900      0.867    0.980     0.901     B+
                 1.000    0.026    0.500      1.000    0.667      0.698    0.987     0.500     D+
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     C
                 0.778    0.065    0.778      0.778    0.778      0.713    0.925     0.784     B
                 0.700    0.033    0.875      0.700    0.778      0.722    0.960     0.826     C+
                 1.000    0.029    0.833      1.000    0.909      0.900    0.991     0.900     A
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     A+
Weighted Avg.    0.850    0.035    0.860      0.850    0.850      0.812    0.967     0.858

=== Confusion Matrix ===

 a b c d e f g   <-- classified as
 9 0 0 0 0 1 0 | a = B+
 0 1 0 0 0 0 0 | b = D+
 0 0 3 0 0 0 0 | c = C
 1 0 0 7 1 0 0 | d = B
 0 1 0 2 7 0 0 | e = C+
 0 0 0 0 0 5 0 | f = A
 0 0 0 0 0 0 2 | g = A+
```

Fig.2. Weka summary report for the IP Overall Grades, Fall 2020-21

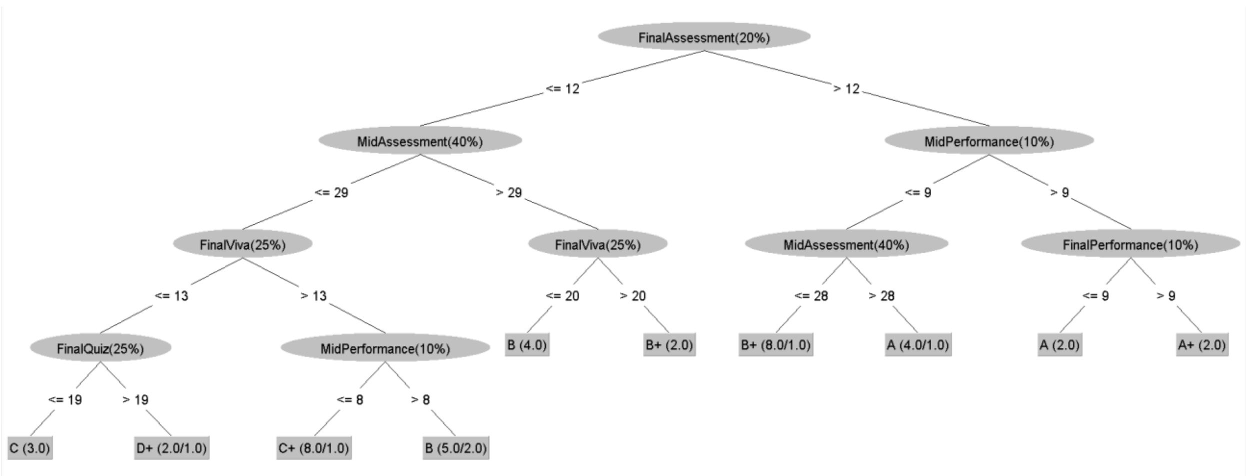


Fig.3. Decision Tree for IP Overall Grades, Fall 2020-21

According to the Decision Tree from fig 3, 'Final Assessment (20%)' got the highest priority for determining students' overall grades. The students' who got 'A+' obtained more than 12 in 'Final Assessment (20%)', greater than 9 in 'Mid Performance (10%)' and 'Final Performance (10%)'. However, the students' who got 'D+' achieved less or equal to

12 in 'Final Assessment (20%), less or equal to 29 in 'Mid Assessment (40%)', less or equal to 13 in 'Final Viva (25%)' and greater than 19 in 'Final Quiz (25%)'. A similar process will be followed for explaining other grades as well from the Decision Tree. This description pattern is applied for illustrating the rest of the Decision Trees.

B. Introduction to Programming (IP) Fall 2019-20 (On-campus)

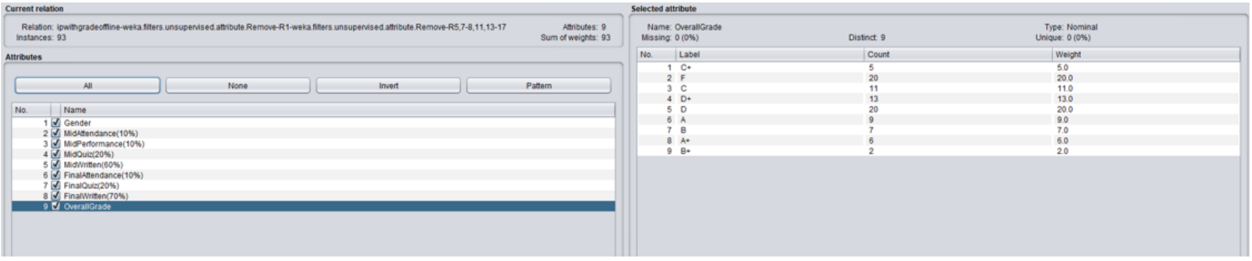


Fig.4. Attribute names and the quantity of IP Overall Grades, Fall 2019-20

```
=== Summary ===

Correctly Classified Instances          86               92.4731 %
Incorrectly Classified Instances         7                7.5269 %
Kappa statistic                          0.9116
Mean absolute error                      0.0244
Root mean squared error                  0.1104
Relative absolute error                 12.8431 %
Root relative squared error             35.9006 %
Total Number of Instances               93

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.600    0.011    0.750      0.600    0.667      0.654    0.972     0.601     C+
                 1.000    0.014    0.952      1.000    0.976      0.969    0.994     0.957     F
                 0.818    0.012    0.900      0.818    0.857      0.840    0.991     0.920     C
                 1.000    0.038    0.813      1.000    0.897      0.884    0.991     0.922     D+
                 0.900    0.000    1.000      0.900    0.947      0.936    0.993     0.973     D
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     A
                 0.857    0.012    0.857      0.857    0.857      0.846    0.993     0.846     B
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     A+
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     B+
Weighted Avg.    0.925    0.011    0.928      0.925    0.923      0.914    0.993     0.932

=== Confusion Matrix ===

  a  b  c  d  e  f  g  h  i   <-- classified as
  3  0  0  1  0  0  1  0  0 |  a = C+
  0 20  0  0  0  0  0  0  0 |  b = F
  1  0  9  1  0  0  0  0  0 |  c = C
  0  0  0 13  0  0  0  0  0 |  d = D+
  0  1  0  1 18  0  0  0  0 |  e = D
  0  0  0  0  0  9  0  0  0 |  f = A
  0  0  1  0  0  0  6  0  0 |  g = B
  0  0  0  0  0  0  0  6  0 |  h = A+
  0  0  0  0  0  0  0  0  2 |  i = B+
```

Fig.5. Weka summary report for the IP Overall Grades, Fall 2019-20

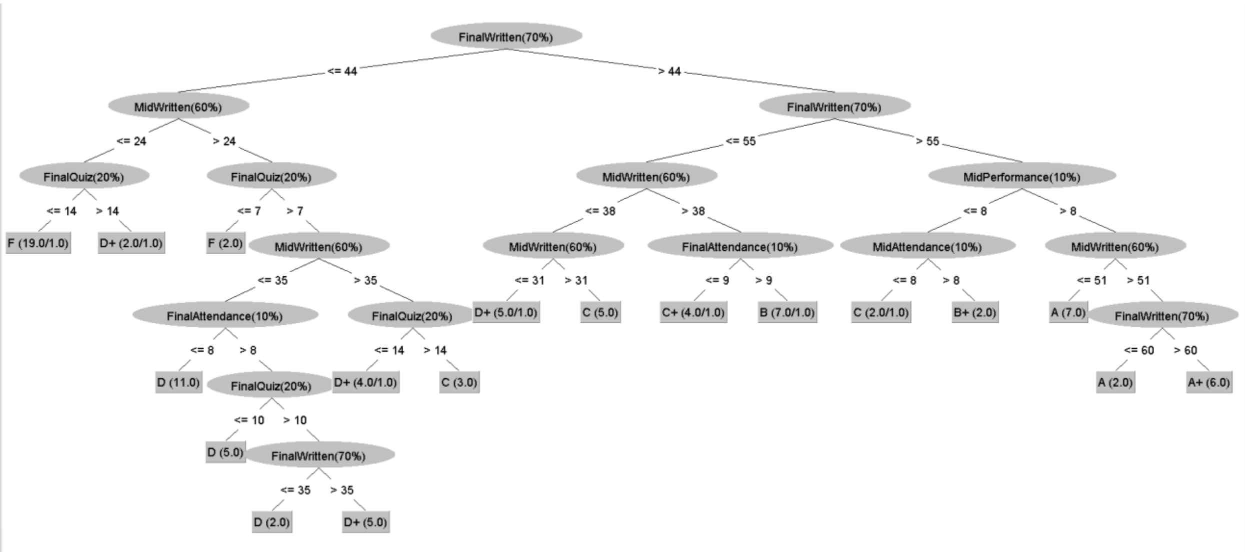


Fig.6. Decision Tree for IP Overall Grades, Fall 2019-20

C. *Object Oriented Programming 1 (OOP1) Fall 2020-21 (Online)*

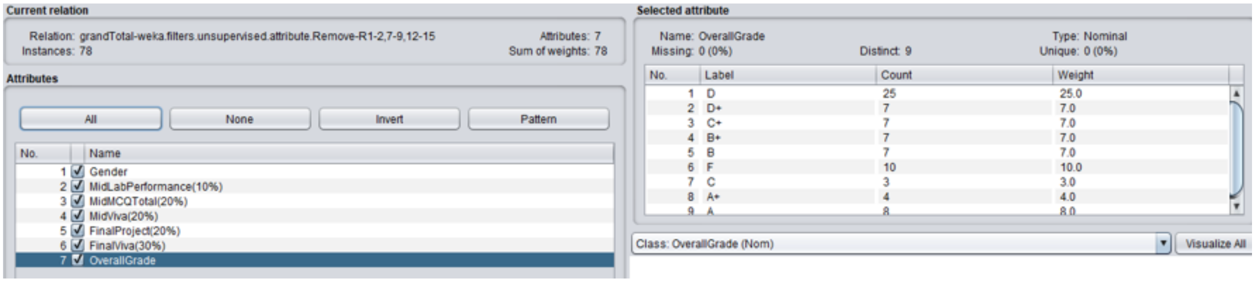


Fig.7. Attribute names and the quantity of OOP1 Overall Grades, Fall 2020-21

```
=== Summary ===

Correctly Classified Instances          69               88.4615 %
Incorrectly Classified Instances         9               11.5385 %
Kappa statistic                          0.86
Mean absolute error                      0.0385
Root mean squared error                  0.1388
Relative absolute error                 20.6379 %
Root relative squared error             45.5674 %
Total Number of Instances               78

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.960    0.094    0.828      0.960    0.889      0.836    0.963     0.870     D
                 1.000    0.014    0.875      1.000    0.933      0.929    0.993     0.875     D+
                 0.857    0.000    1.000      0.857    0.923      0.919    0.983     0.899     C+
                 0.714    0.000    1.000      0.714    0.833      0.833    0.992     0.896     B+
                 1.000    0.028    0.778      1.000    0.875      0.869    0.992     0.873     B
                 0.600    0.000    1.000      0.600    0.750      0.753    0.957     0.817     F
                 0.667    0.013    0.667      0.667    0.667      0.653    0.989     0.644     C
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     A+
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     A
Weighted Avg.    0.885    0.035    0.901      0.885    0.880      0.862    0.979     0.880

=== Confusion Matrix ===

  a  b  c  d  e  f  g  h  i   <-- classified as
 24  1  0  0  0  0  0  0  0 |  a = D
  0  7  0  0  0  0  0  0  0 |  b = D+
  1  0  6  0  0  0  0  0  0 |  c = C+
  0  0  0  5  2  0  0  0  0 |  d = B+
  0  0  0  0  7  0  0  0  0 |  e = B
  3  0  0  0  0  6  1  0  0 |  f = F
  1  0  0  0  0  0  2  0  0 |  g = C
  0  0  0  0  0  0  0  4  0 |  h = A+
  0  0  0  0  0  0  0  0  8 |  i = A
```

Fig.8. Weka summary report for the OOP1 Overall Grades, Fall 2020-21

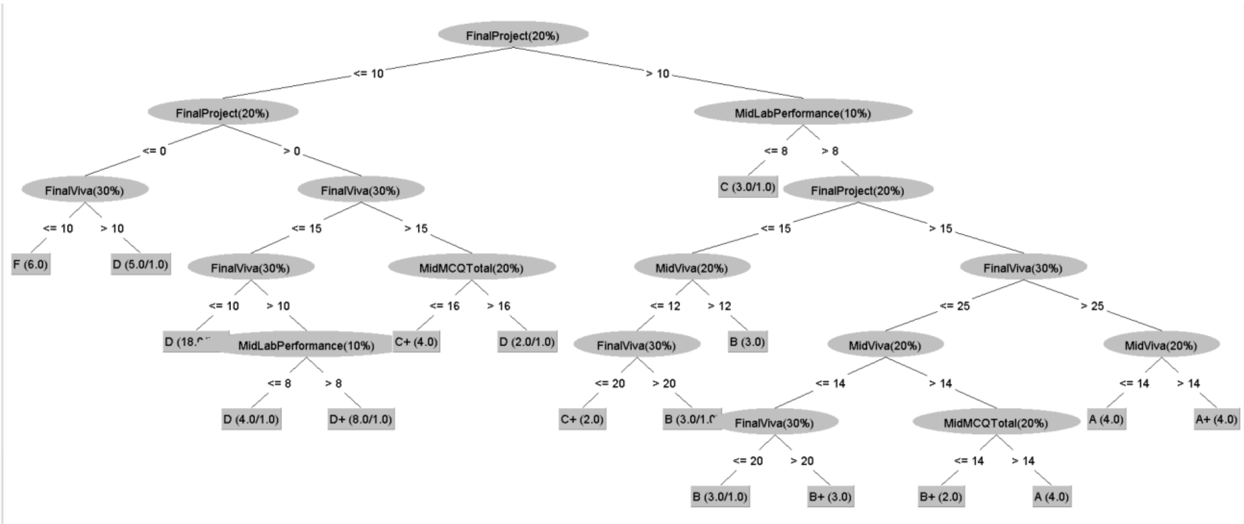


Fig.9. Decision Tree for OOP1 Overall Grades, Fall 2020-21

D. *Object Oriented Programming 1 (OOP1) Fall 2019-20 (On-campus)*

Unfortunately, the final-term data of this course on-campus could not be retrieved due to the campus lockdown since the data were in hard copy and kept on campus. So only mid-term data is used for further works.

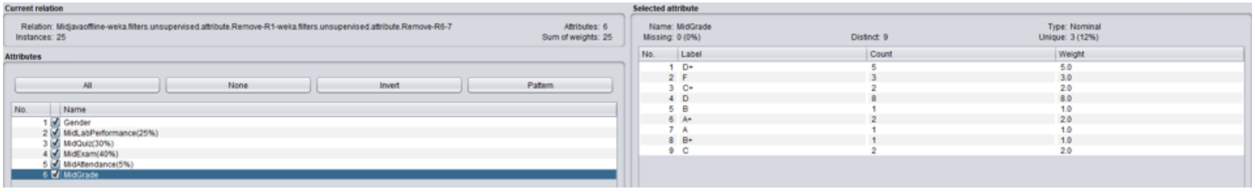


Fig.10. Attribute names and the quantity of OOP1 mid-term Grades, Fall 2019-20

```
=== Summary ===

Correctly Classified Instances          21               84      %
Incorrectly Classified Instances         4               16      %
Kappa statistic                          0.8016
Mean absolute error                      0.0514
Root mean squared error                  0.1602
Relative absolute error                 27.5905 %
Root relative squared error             52.9572 %
Total Number of Instances               25

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     D+
                 0.667    0.000    1.000      0.667    0.800      0.799    0.939     0.758     F
                 1.000    0.043    0.667      1.000    0.800      0.799    0.978     0.667     C+
                 1.000    0.059    0.889      1.000    0.941      0.915    0.971     0.889     D
                 0.000    0.000    ?          0.000    ?          ?        0.958     0.333     B
                 1.000    0.043    0.667      1.000    0.800      0.799    0.978     0.667     A+
                 0.000    0.000    ?          0.000    ?          ?        0.958     0.333     A
                 0.000    0.000    ?          0.000    ?          ?        0.958     0.333     B+
                 1.000    0.043    0.667      1.000    0.800      0.799    0.978     0.667     C
Weighted Avg.    0.840    0.029    ?          0.840    ?          ?        0.973     0.775

=== Confusion Matrix ===

 a b c d e f g h i   <-- classified as
 5 0 0 0 0 0 0 0 0 | a = D+
 0 2 0 1 0 0 0 0 0 | b = F
 0 0 2 0 0 0 0 0 0 | c = C+
 0 0 0 8 0 0 0 0 0 | d = D
 0 0 1 0 0 0 0 0 0 | e = B
 0 0 0 0 0 2 0 0 0 | f = A+
 0 0 0 0 0 1 0 0 0 | g = A
 0 0 0 0 0 0 0 0 1 | h = B+
 0 0 0 0 0 0 0 0 2 | i = C
```

Fig.11. Weka summary report for the OOP1 mid-term Grades, Fall 2019-20

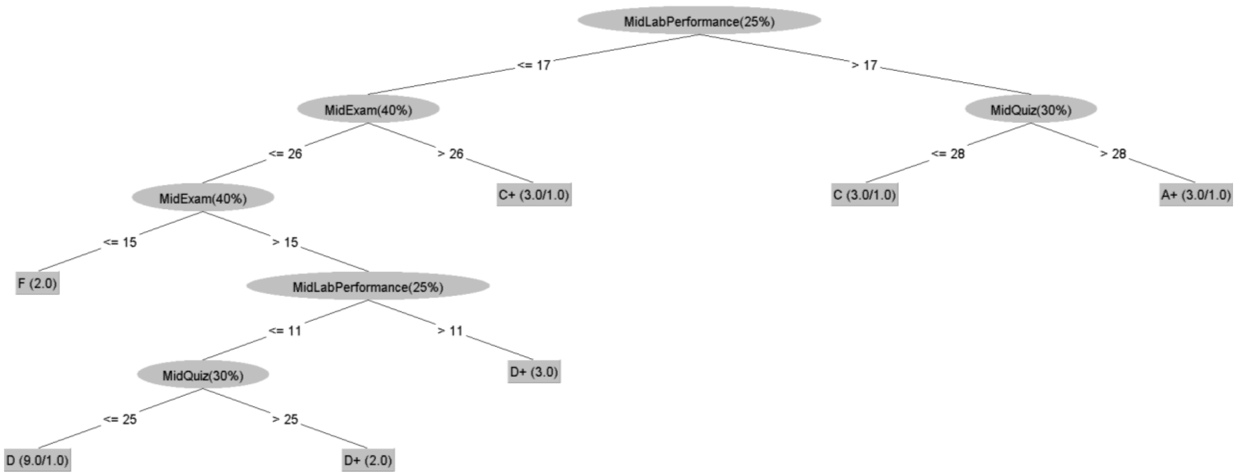


Fig.12. Decision Tree for OOP1 mid-term Grades, Fall 2019-20

E. Object Oriented Programming 2 (OOP2) Fall 2020-21 (Online)

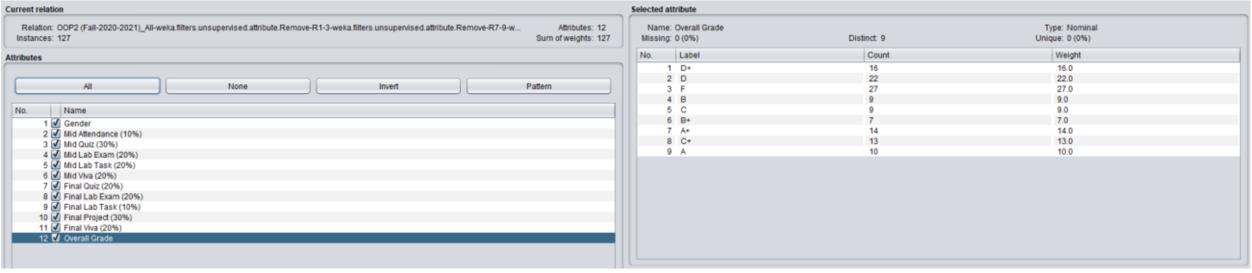


Fig.13. Attribute names and the quantity of OOP2 Overall Grades, Fall 2020-21

```
=== Summary ===

Correctly Classified Instances         109               85.8268 %
Incorrectly Classified Instances        18               14.1732 %
Kappa statistic                          0.8367
Mean absolute error                      0.0418
Root mean squared error                  0.1446
Relative absolute error                 21.6618 %
Root relative squared error             46.5785 %
Total Number of Instances              127

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.813    0.063    0.650      0.813    0.722      0.683    0.970     0.756     D+
                 0.864    0.019    0.905      0.864    0.884      0.860    0.988     0.904     D
                 0.963    0.000    1.000      0.963    0.981      0.976    0.997     0.987     F
                 0.889    0.034    0.667      0.889    0.762      0.750    0.984     0.759     B
                 0.667    0.025    0.667      0.667    0.667      0.641    0.972     0.606     C
                 0.714    0.008    0.833      0.714    0.769      0.759    0.992     0.795     B+
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     A+
                 0.692    0.009    0.900      0.692    0.783      0.769    0.985     0.863     C+
                 0.900    0.000    1.000      0.900    0.947      0.945    0.998     0.967     A
Weighted Avg.    0.858    0.017    0.873      0.858    0.861      0.846    0.988     0.877

=== Confusion Matrix ===

  a  b  c  d  e  f  g  h  i   <-- classified as
 13  1  0  0  1  0  0  1  0 |  a = D+
  3 19  0  0  0  0  0  0  0 |  b = D
  0  1 26  0  0  0  0  0  0 |  c = F
  0  0  0  8  1  0  0  0  0 |  d = B
  3  0  0  0  6  0  0  0  0 |  e = C
  0  0  0  2  0  5  0  0  0 |  f = B+
  0  0  0  0  0  0 14  0  0 |  g = A+
  1  0  0  2  1  0  0  9  0 |  h = C+
  0  0  0  0  0  1  0  0  9 |  i = A
```

Fig.14. Weka summary report for the OOP2 Overall Grades, Fall 2020-21

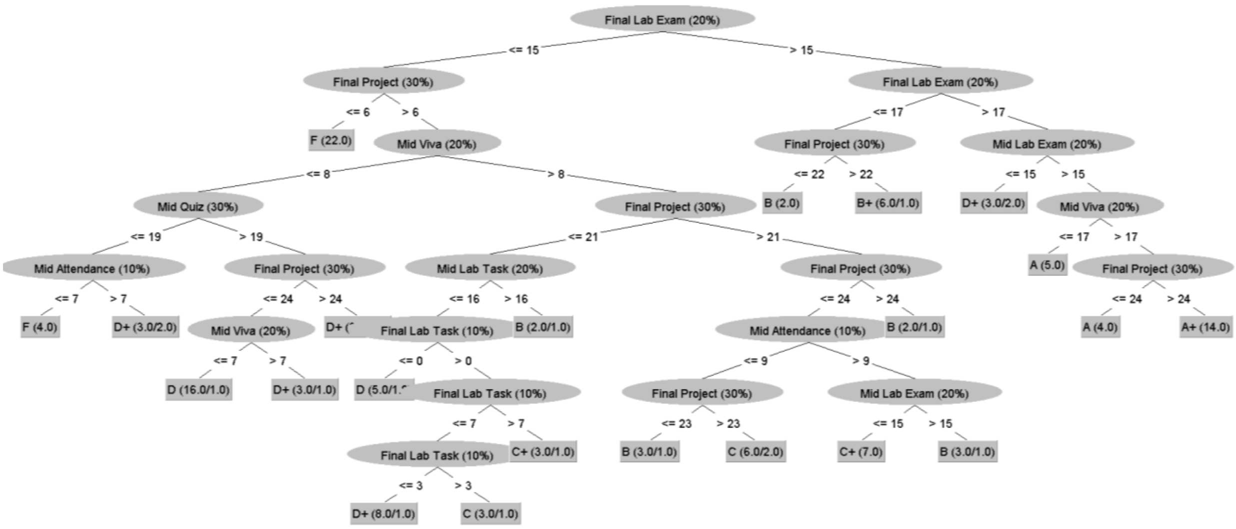


Fig.15. Decision Tree for OOP2 Overall Grades, Fall 2020-21

F. *Object Oriented Programming 2 (OOP2) Fall 2019-20 (On-campus)*

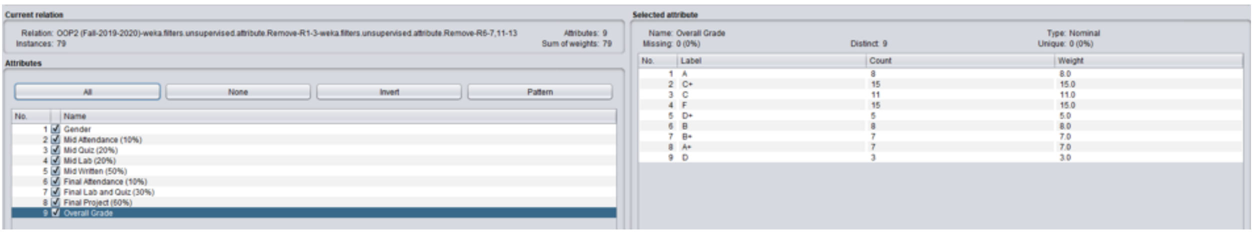


Fig.16. Attribute names and the quantity of OOP2 Overall Grades, Fall 2019-20

```
=== Summary ===

Correctly Classified Instances          63               79.7468 %
Incorrectly Classified Instances        16               20.2532 %
Kappa statistic                          0.7673
Mean absolute error                      0.0672
Root mean squared error                  0.1833
Relative absolute error                 34.8062 %
Root relative squared error             59.0656 %
Total Number of Instances               79

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.750    0.028    0.750      0.750    0.750      0.722    0.960     0.663     A
                 0.733    0.047    0.786      0.733    0.759      0.705    0.936     0.712     C+
                 0.636    0.044    0.700      0.636    0.667      0.617    0.938     0.700     C
                 0.933    0.000    1.000      0.933    0.966      0.959    0.998     0.989     F
                 1.000    0.027    0.714      1.000    0.833      0.834    0.991     0.783     D+
                 1.000    0.042    0.727      1.000    0.842      0.835    0.984     0.795     B
                 0.571    0.014    0.800      0.571    0.667      0.651    0.921     0.636     B+
                 0.857    0.014    0.857      0.857    0.857      0.843    0.990     0.835     A+
                 0.667    0.013    0.667      0.667    0.667      0.654    0.989     0.644     D
Weighted Avg.    0.797    0.027    0.803      0.797    0.795      0.769    0.964     0.772

=== Confusion Matrix ===

  a  b  c  d  e  f  g  h  i   <-- classified as
  6  0  0  0  0  1  0  1  0 |  a = A
  1 11  2  0  0  1  0  0  0 |  b = C+
  0  2  7  0  1  0  0  0  1 |  c = C
  0  0  1 14  0  0  0  0  0 |  d = F
  0  0  0  0  5  0  0  0  0 |  e = D+
  0  0  0  0  0  8  0  0  0 |  f = B
  1  1  0  0  0  1  4  0  0 |  g = B+
  0  0  0  0  0  0  1  6  0 |  h = A+
  0  0  0  0  1  0  0  0  2 |  i = D
```

Fig.17. Weka summary report for the OOP2 Overall Grades, Fall 2019-20

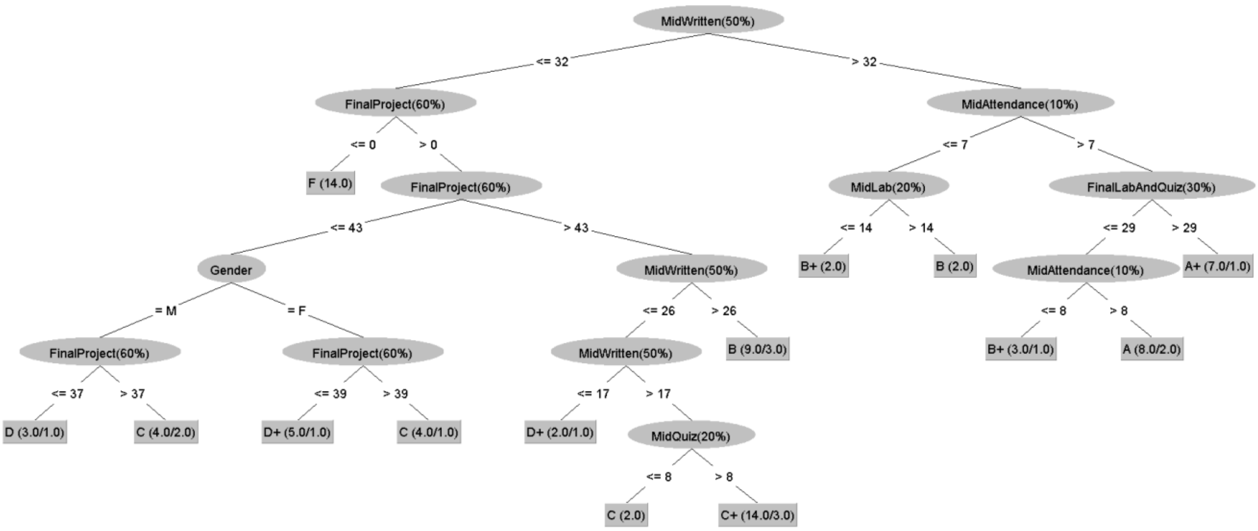
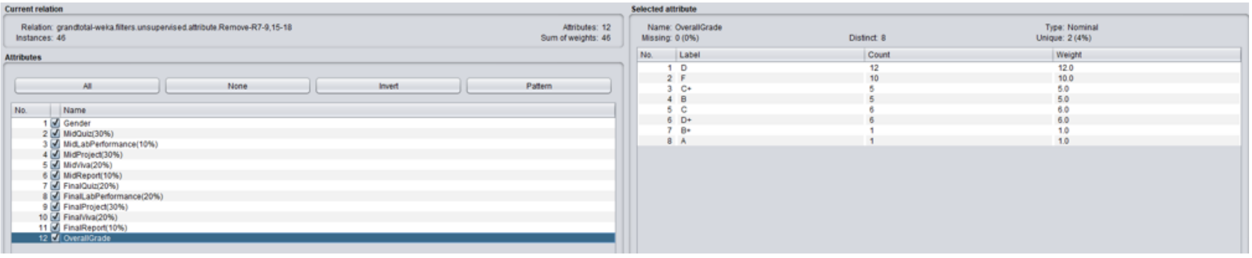


Fig.18. Decision Tree for OOP2 Overall Grades, Fall 2019-20

G. *Web Technologies (WT) Fall 2020-21 (Online)*



Current relation

Relation: grandtotal-weka.filters.unsupervised.attribute.Remove-R7-9,15-18 Attributes: 12
Instances: 46 Sum of weights: 46

Attributes

All | None | Invert | Pattern

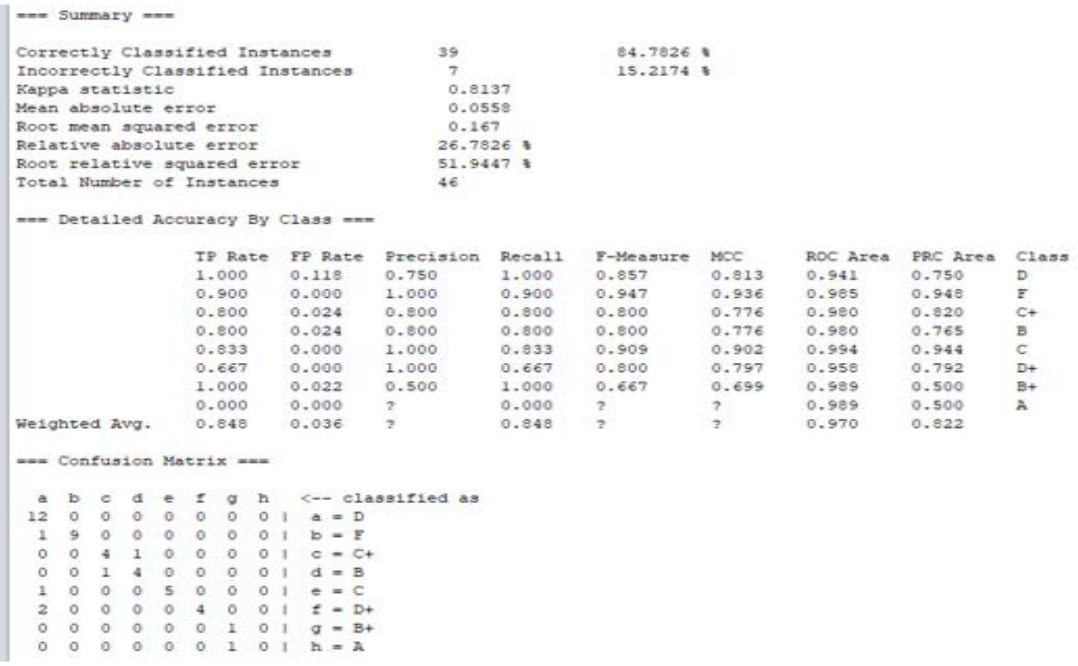
No.	Name
1	Gender
2	MidQuiz(30%)
3	MidLabPerformance(10%)
4	MidProject(30%)
5	MidViva(20%)
6	MidReport(10%)
7	FinalQuiz(20%)
8	FinalLabPerformance(20%)
9	FinalProject(30%)
10	FinalViva(20%)
11	FinalReport(10%)
12	OverallGrade

Selected attribute

Name: OverallGrade Type: Nominal
Missing: 0 (0%) Distinct: 8 Unique: 2 (4%)

No.	Label	Count	Weight
1	D	12	12.0
2	F	10	10.0
3	C+	5	5.0
4	B	5	5.0
5	C	6	6.0
6	D+	6	6.0
7	B+	1	1.0
8	A	1	1.0

Fig.19. Attribute names and the quantity of WT Overall Grades, Fall 2020-21



```
=== Summary ===

Correctly Classified Instances          39               84.7826 %
Incorrectly Classified Instances         7               15.2174 %
Kappa statistic                          0.8137
Mean absolute error                      0.0558
Root mean squared error                  0.167
Relative absolute error                 26.7826 %
Root relative squared error             51.9447 %
Total Number of Instances               46

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.118    0.750      1.000    0.857      0.813    0.941     0.750     D
                 0.900    0.000    1.000      0.900    0.947      0.936    0.985     0.948     F
                 0.800    0.024    0.800      0.800    0.800      0.776    0.980     0.820     C+
                 0.800    0.024    0.800      0.800    0.800      0.776    0.980     0.765     B
                 0.833    0.000    1.000      0.833    0.909      0.902    0.994     0.944     C
                 0.667    0.000    1.000      0.667    0.800      0.797    0.958     0.792     D+
                 1.000    0.022    0.500      1.000    0.667      0.699    0.989     0.500     B+
                 0.000    0.000    ?          0.000    ?          ?        0.989     0.500     A
Weighted Avg.    0.848    0.036    ?          0.848    ?          ?        0.970     0.822

=== Confusion Matrix ===

  a  b  c  d  e  f  g  h   <-- classified as
 12  0  0  0  0  0  0  0 |  a = D
  1  9  0  0  0  0  0  0 |  b = F
  0  0  4  1  0  0  0  0 |  c = C+
  0  0  1  4  0  0  0  0 |  d = B
  1  0  0  0  5  0  0  0 |  e = C
  2  0  0  0  0  4  0  0 |  f = D+
  0  0  0  0  0  0  1  0 |  g = B+
  0  0  0  0  0  0  1  0 |  h = A
```

Fig.20. Weka summary report for the WT Overall Grades, Fall 2020-21

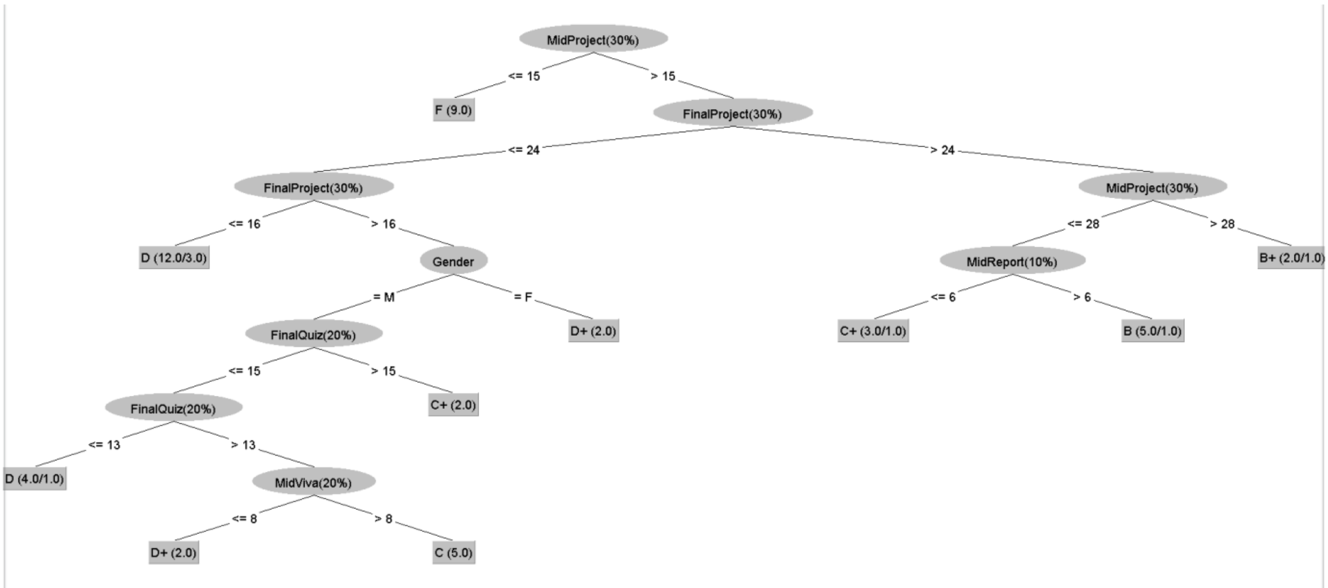


Fig.21. Decision Tree for WT Overall Grades, Fall 2020-21

H. *Web Technologies (WT) Fall 2019-20 (On-campus)*

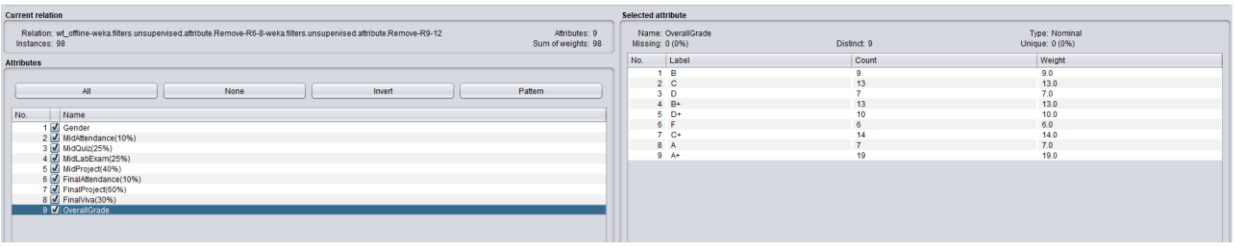


Fig.22. Attribute names and the quantity of WT Overall Grades, Fall 2019-20

```
=== Summary ===

Correctly Classified Instances          89               90.8163 %
Incorrectly Classified Instances         9                9.1837 %
Kappa statistic                          0.8949
Mean absolute error                      0.0281
Root mean squared error                  0.1186
Relative absolute error                 14.4619 %
Root relative squared error             38.0537 %
Total Number of Instances               98

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.889    0.022    0.800      0.889    0.842      0.827    0.986     0.881     B
                 0.692    0.000    1.000      0.692    0.818      0.813    0.988     0.918     C
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     D
                 0.923    0.024    0.857      0.923    0.889      0.872    0.986     0.850     B+
                 1.000    0.034    0.769      1.000    0.870      0.862    0.992     0.903     D+
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     F
                 0.929    0.012    0.929      0.929    0.929      0.917    0.998     0.981     C+
                 0.714    0.000    1.000      0.714    0.833      0.836    0.985     0.867     A
                 1.000    0.013    0.950      1.000    0.974      0.968    0.999     0.995     A+
Weighted Avg.    0.908    0.013    0.919      0.908    0.906      0.898    0.993     0.935

=== Confusion Matrix ===

  a  b  c  d  e  f  g  h  i   <-- classified as
  8  0  0  1  0  0  0  0  0 |  a = B
  1  9  0  0  2  0  1  0  0 |  b = C
  0  0  7  0  0  0  0  0  0 |  c = D
  1  0  0 12  0  0  0  0  0 |  d = B+
  0  0  0  0 10  0  0  0  0 |  e = D+
  0  0  0  0  0  6  0  0  0 |  f = F
  0  0  0  0  1  0 13  0  0 |  g = C+
  0  0  0  1  0  0  0  5  1 |  h = A
  0  0  0  0  0  0  0  0 19 |  i = A+
```

Fig.23. Weka summary report for the WT Overall Grades, Fall 2019-20

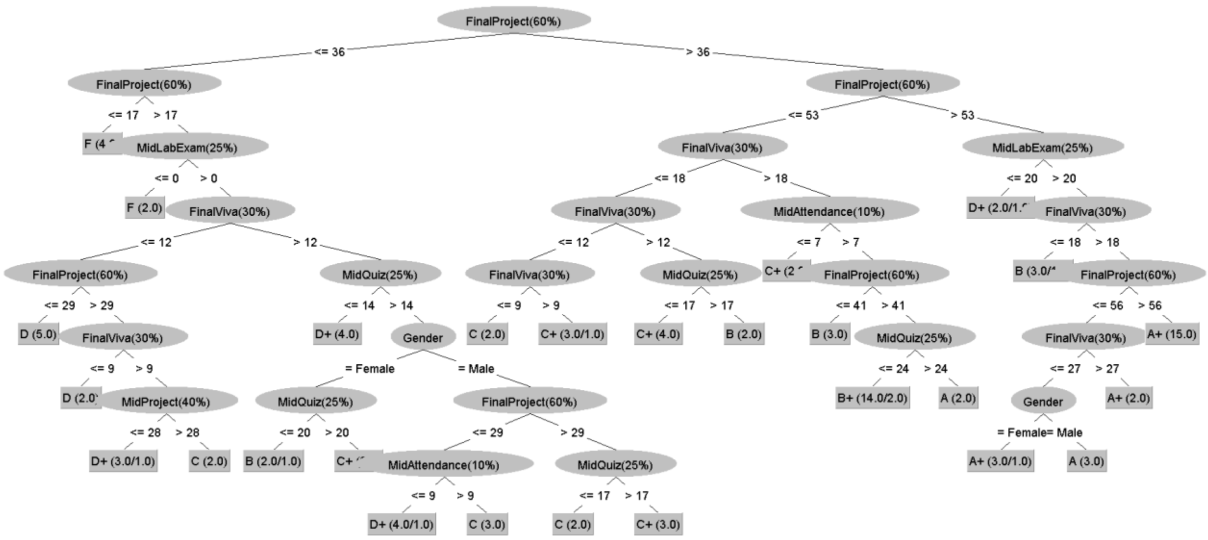


Fig.24. Decision Tree for WT Overall Grades, Fall 2019-20

I. Comparison between Fall 2020-21 (Online) and Fall 2019-20 (On-campus)

However, the course activities, grading criteria, and marks distribution in the online semester are not identical to those in the on-campus semester. Because it is not possible to keep the assessment measures the same for both online and on-campus semesters as some of the marking criteria had to be changed in the online semester. Hence, a slight modification in the grading system was essential. In this situation, the comparison is much more difficult. Therefore, combining several activities into a single action for both online and on-campus semesters was necessary. Besides, for a balanced comparison, the marks for all activities are needed to be similar. As a result, for the sake of a fair comparison, all the marks of the activities are converted to hundred percent.

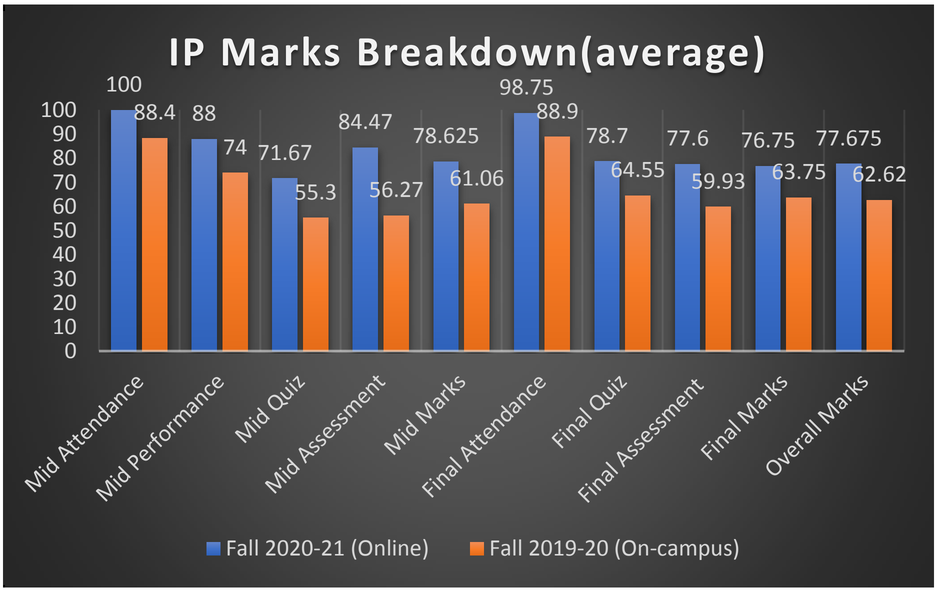


Fig.25. Comparison between attributes of IP in online and on campus.

Fig 25 illustrates the comparison between the Fall 2020-2021 and Fall 2019-2020 semester of Introduction to Programming (IP). Here, all the course activities in the online semester are better than the on-campus semester.

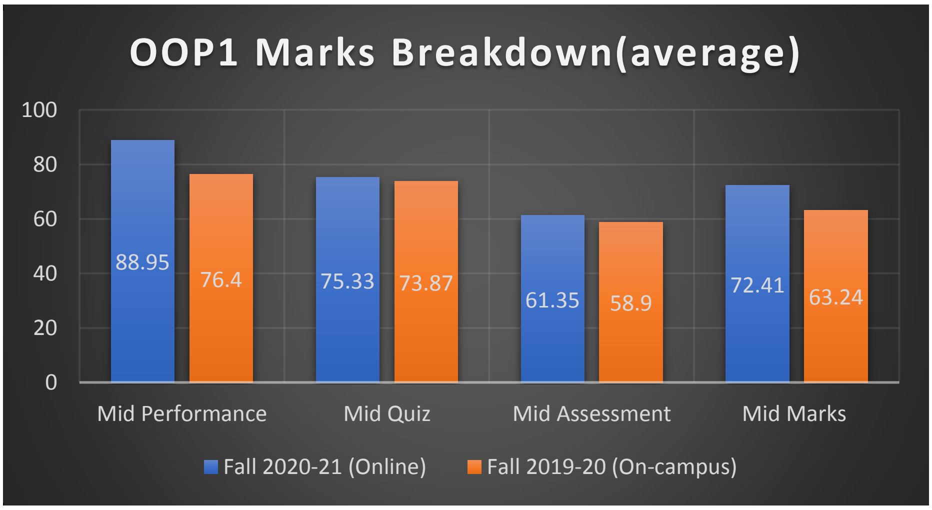


Fig.26. Comparison between attributes of OOP1 (Mid) in online and on campus.

Fig 26 depicts the comparison of mid-term between the Fall 2020-2021 and Fall 2019-2020 semester of Object-Oriented Programming 1 (OOP1). The final term data of Fall 2019-20 (on-campus) is missing, so the comparison is also made on mid-term only. In this course, online semesters' course activities for mid-term are also better than the on-campus semester.

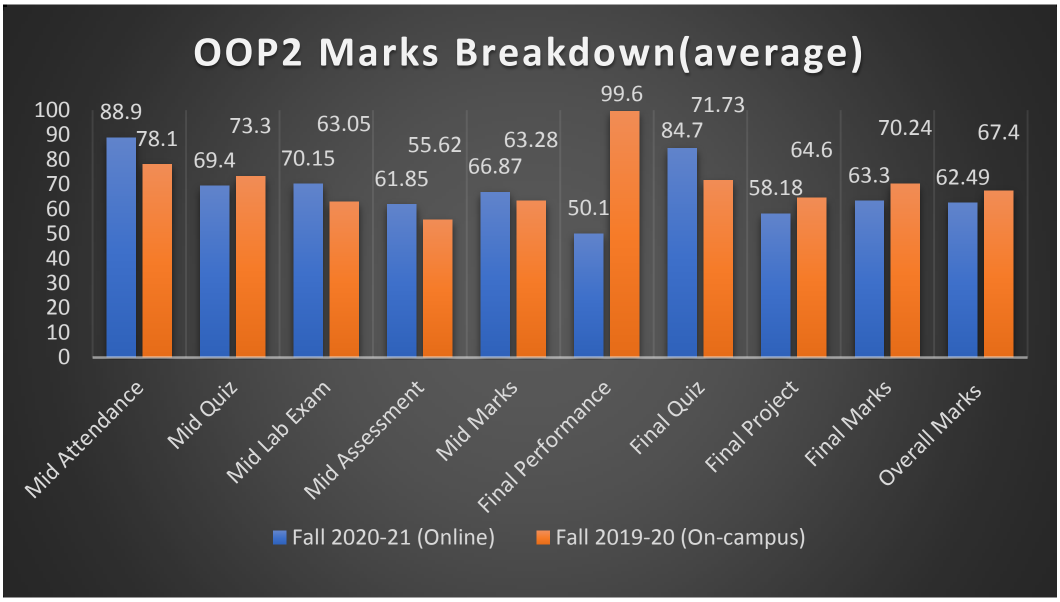


Fig.27. Comparison between attributes of OOP2 in online and on-campus

Fig 27 illustrates the comparison between the Fall 2020-2021 and Fall 2019-2020 semester of Object-Oriented Programming 2 (OOP2). It gives a balanced outcome for this course between online and on-campus. For mid-term activities, students' of the online semester performed better than students' on-campus semester, whereas vice versa happened in final term course activities.

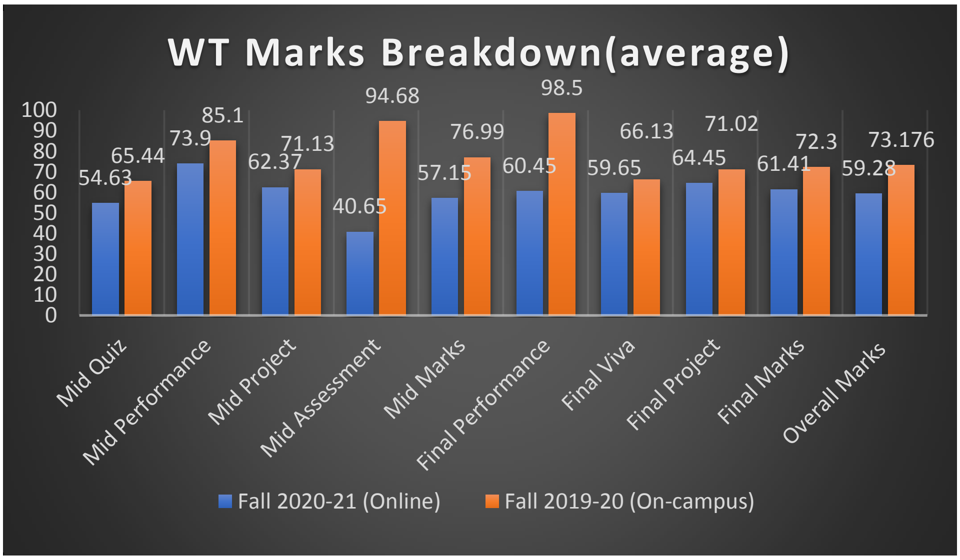


Fig.28. Comparison between attributes of WT in online and on-campus

Fig 28 demonstrates the comparison between the Fall 2020-2021 and Fall 2019-2020 semester of Web Technologies. Interestingly, Fall 2019-20 (on-campus) students achieved better outcomes in course activities than the Fall 2020-21 (online) semester students.

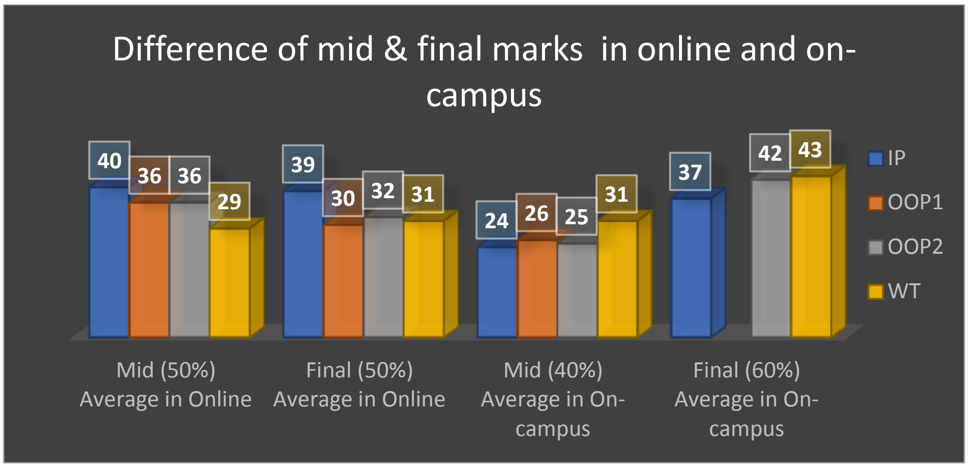


Fig.29. Difference between online and on-campus average marks in four courses.

Fig 29 depicts the average marks obtained by the students in four different courses, and it also contains the breakdown of the mid-term and final term of both online and on campus. Online semester courses had an equal percentage of marks distribution in mid and final, whereas on-campus semester had 40% in mid and 60% in final.

Table 3. Weka summary report of Decision Trees' for all datasets

	Total number of instances	Accuracy	Mean absolute error	Root mean squared error	Relative absolute error	Root relative squared error
IP Fall 2020-21(Online)	40	85%	0.06	0.1732	25.97%	51.19%
IP Fall 2019-20(On-campus)	93	92.47%	0.0244	0.1104	12.84%	35.90%
OOP1 Fall 2020-21(Online)	78	88.46%	0.0385	0.1388	20.64%	45.57%
OOP1 Fall 2019-20(On-campus)	25	84%	0.0514	0.1602	27.59%	52.96%
OOP2 Fall 2020-21(Online)	127	85.83%	0.0418	0.1446	21.66%	46.58%
OOP2 Fall 2019-20(On-campus)	79	79.75%	0.0672	0.1833	34.81%	59.07%
WT Fall 2020-21(Online)	46	84.78%	0.0558	0.167	26.78%	51.94%
WT Fall 2019-20(On-campus)	98	90.82%	0.0281	0.1186	14.46%	38.05%

The data in table 3 are collected from fig 2, 5, 8, 11, 14, 17, 20, and 23. All Decision Trees have reasonable accuracy. Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are two of the most common metrics used to measure accuracy for continuous variables. MAE measures the average magnitude of the errors in a set of predictions without considering their direction. Here, the MAE values of every Decision Tree are less than 0.1, which is excellent. RMSE is a quadratic scoring rule that also measures the average magnitude of the error. According to a rule of thumb, an RMSE of 0.2 to 0.5 indicates that the model can make sound predictions for the data. In table 3, the RMSE value of every Decision Tree is in the range of good predictions, though we will not predict anything in this research. Relative Absolute Error (RAE) is a way to measure the performance of a predictive model. A good forecasting model will produce an RAE ratio close to zero, whereas a poor model will produce a ratio greater than one. In our findings, all models will be suitable for forecasting as their RAE values are close to zero. The root relative squared error is relative to what it would have been if a simple predictor had been used. More specifically, this simple predictor is just the average of the actual values. It gives the idea of the scale of error compared to how variable the actual values are.

4. Discussion

Nowadays, due to the global pandemic, educational institutions are bound to shift all their activities online. The feasibility of online classes is now being questioned, and this research aims to figure out the answer to this question. This study analyzes the data of four programming courses for four consecutive years of the Computer Science department. The results from experimenting with these courses indicate several reasons that influence the students' performance online and on-campus. Several Decision Trees (fig 3, 6, 9, 12, 15, 18, 21, and 24) are generated based on these data, which helped understand the importance of a particular factor on students' overall performance. Besides, the accuracy of these Decision Tree classification models is also very precise (table 3).

A brief comparison is made based on the students' performance between online and on-campus semesters. There is a variation in students' attendance between online and on-campus semesters (fig 25 and 27). Students usually have to overcome several obstacles to join their classes in the on-campus semester, whereas they do not have to face these troubles in online courses. On another note, students can join the classes from their homes even if they are mildly sick, which they could not do if they had to go to the university to attend their classes. In the online semester, much time is saved, as the students do not physically join the classes. Therefore, for all the courses, in the online semester, the attendance marks are significantly higher than those in the on-campus semester (fig 25 and 27).

In the Introduction to Programming (IP), students are taught basic programming techniques and Object-Oriented Programming 1 (OOP1), the students are obliged to submit a simple project. In these courses mentioned above online, we found that students performed better than they performed on-campus (fig 25-26). Thus, the students of Introduction to Programming (IP) and Object-Oriented Programming 1 (OOP1) performed better in the online semester than the students who did the courses on-campus (fig 25-26). On the contrary, Object Oriented Programming 2 (OOP2) and Web Technologies (WT) emphasize real-life projects that require in-depth knowledge and analytical abilities. During the mid-term of Object-Oriented Programming 2 (OOP2), students learn about various techniques about the language and implement those on their regular lab tasks; in the final term, they have to use all the lessons they learned and implement on a comprehensive project. Hence, for Object-Oriented Programming 2 (OOP2), in the online semester, the students performed well in the mid-term (fig 27). However, surprisingly, they could not keep it up in the final, as the project carries a substantial mark (fig 27). On the other hand, for Web Technologies (WT), the case is slightly different. In this course, they have to submit two individual projects in both the mid and final terms. The average marks of the students are comparatively better in the on-campus semester (fig 28) for Web Technologies (WT).

To summarize, first-year students did very well in Introduction to Programming (IP) compared to other courses in the online semester (fig 29). However, in the on-campus semester, students of the Web Technologies (WT) course performed better than others (fig 29). The increased face-to-face contact with the course teacher throughout the on-campus semester enables students to completely grasp all of the concepts regarding WT, allowing them to present outstanding projects.

5. Conclusion

COVID-19 has a significant impact on education communities all around the globe. It forced the conventional form of education to close its doors, and it would have stayed closed if the advent of online education had not occurred. This research aims to identify the variables that influence students' success in the online and on-campus environment in four programming courses, which covers the computing students from four consecutive years of undergraduate. Furthermore, a distinction of student results was made between two semesters, one of which was performed online and the other on-campus. Unfortunately, for a particular course, the data of the final term of the semester conducted on-campus could not be retrieved due to the lockdown as the data were in hard copy stored in the university premises. That is why comparing and classifying students' performance for that course is made based on the mid-term only.

In this research, the WEKA tool is used to classify students' performance in online and on-campus classes using the Decision Tree classification algorithm. All the Decision Tree classification models have outstanding accuracy. However, they have very decent MAE, RMSE, and RAE values, which can make sound predictions for the data. Later, a comparison was made between the four stated courses. This study enabled us to identify several obstacles that students encountered in four specific courses. According to the decision trees, the freshman students in the Introduction to Programming (IP) tend to perform better in that course when they have a good score in the 'Final Assessment' in online and on-campus semesters. In the Object Oriented Programming 1 (OOP1) course, the 'Final Project' has the highest impact on students' performance in the online semester. However, in the on-campus semester, the 'Mid Lab Performance' seems to be the most crucial factor, but that is because the data of the final term could not be gathered and the analysis was conducted solely based on the mid-term. Then in the Object Oriented Programming 2 (OOP2), the 'Final Lab Exam' is the most crucial factor in the semester conducted virtually. On the contrary, the 'Mid Written' has the most significant influence on students' academic performance in the on-campus semester. At last, the final year students who performed well in the 'Mid Project' in the Web Technologies (WT) course did well in the online semester. On the flip slide, in the on-campus semester, the overall performance mainly depended on the submission of an adequate 'Project' in the final term. In short, fresher students admitted to programming courses performed better than senior students enrolled in Object Oriented Programming 1 (OOP1) or Object-Oriented Programming 2 (OOP2) or Web Technologies (WT) in the online semester. But vice versa happened in the on-campus semester. Senior students enrolled in Web Technology (WT) performed better than the rest of the courses in the on-campus semester. As a result, it allows us to see what kinds of factors affect students' academic outcomes. Our finding also indicates that the overall performance of students in the online semester was adequate.

As discussed before, data of the final term in the on-campus semester is missing for a particular course. Furthermore, data on students who dropped the courses in the online and on-campus semesters were not available. Hence, there was no way to distinguish dropped students from online and on-campus semesters. However, these shortcomings can be improved in future implementations, and predictions can be made as our models' accuracy, MAE, RMSE, and RAE values are very good for forecasting. This study will be valuable for the future if any unforeseen event occurs to change the classes from on-campus to online, or vice-versa. As mentioned earlier, this paper is part of a large study. In the other part, we will undertake an in-depth analysis of students' online performance.

References

[1] Syed Tahir Hijaz and SMM Raza Naqvi. Factors affecting students' performance: A case of private colleges in bangladesh. Journal of sociology, 3(1):44–45, 2006.

[2] Brijesh Kumar Bhardwaj and Saurabh Pal. Data mining: A prediction for performance improvement using classification. arXiv preprint arXiv:1201.3418, 2012.

[3] Pedro Strecht, LuÃs Cruz, Carlos Soares, JoÃ Mendes-Moreira, et al. A comparative study of classification and regression algorithms for modelling students' academic performance. International Educational Data Mining Society, 2015.

[4] Jiawei Han, Jian Pei, and Micheline Kamber. Data mining: concepts and techniques. Elsevier, 2011.

[5] Max Bramer. Principles of data mining, volume 180. Springer, 2007.

[6] Sathe, M.T. and Adamuthe, A.C., 2021. Comparative Study of Supervised Algorithms for Prediction of Students' Performance. International Journal of Modern Education & Computer Science, 13(1).

[7] Nandi, D., Hamilton, M., Harland, J. and Mahmood, S., 2015. Investigation of participation and quality of online interaction. International Journal of Modern Education and Computer Science, 7(8), p.25.

[8] Ali, M. T., Rahman, M. A. and Lamagna, C. Z. (2021) "Comparative Analysis of Program Outcomes Achievement between Face-to-Face and Virtual Classes during COVID-19 Pandemic", AIUB Journal of Science and Engineering (AJSE), 20(1), pp. 1 - 7. Available at: http://ajse.aiub.edu/index.php/ajse/article/view/150 (Accessed: 8 July 2021).

[9] Rivas, A., Gonzalez-Briones, A., Hernandez, G., Prieto, J. and Chamoso, P., 2021. Artificial neural network analysis of the academic performance of students in virtual learning environments. Neurocomputing, 423, pp.713-720.

[10] Moreno-Marcos, P.M., Pong, T.C., Munoz-Merino, P.J. and Kloos, C.D., 2020. Analysis of the factors influencing learners' performance prediction with learning analytics. IEEE Access, 8, pp.5264-5282.

[11] Sadiq Hussain, Neama Abdulaziz Dahan, Fadl Mutaher Ba-Alwib, and Najoua Ribata. Educational data mining and analysis of students' academic performance using weka. Indonesian Journal of Electrical Engineering and Computer Science, 9(2):447–459, 2018.

[12] Zainab Mohammed Ali, Noor Hasan Hassoon, Wasan Saad Ahmed, and Hazim Noman Abed. The application of data mining for predicting academic performance using k-means clustering and na ïve bayes classification. International Journal of Psychosocial Rehabilitation, 24(03), 2020.

[13] T Miranda Lakshmi, A Martin, R Mumtaj Begum, and V Prasanna Venkatesan. An analysis on performance of decision tree algorithms using student's qualitative data. International journal of modern education and computer science, 5(5):18, 2013.

[14] Slamet Wiyono and Taufiq Abidin. Comparative study of machine learning knn, svm, and decision tree algorithm to predict student's performance. International Journal of ResearchGranthaalayah, 7(1):190–196, 2019.

[15] Ahmed Mueen, Bassam Zafar, and Umar Manzoor. Modeling and predicting students' academic performance using data mining techniques. International journal of modern education & computer science, 8(11), 2016.

[16] Dipta Das, Asif Khan Shakir, Md Shah Golam Rabbani, Mostafijur Rahman, Syamimi Mardiah Shaharum, Sabira Khatun, Norasyikin Binti Fadilah, Khandker M Qaiduzzaman, Md Shariful Islam, and Md Shohel Arman. A comparative analysis of four classification algorithms for university students performance detection. In InECCE2019, pages 415–424. Springer, 2020.

[17] Evandro B Costa, Baldoino Fonseca, Marcelo Almeida Santana, Fabr ísia Ferreira de Ara újo, and Joilson Rego. Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. Computers in Human Behavior, 73:247–256, 2017.

[18] Raheela Asif, Agathe Merceron, Syed Abbas Ali, and Najmi Ghani Haider. Analyzing undergraduate students' performance using educational data mining. Computers & Education, 113:177–194, 2017.

[19] Jena, P.K., 2020. Impact of pandemic COVID-19 on education in India. International Journal of Current Research (IJCR), 12.

[20] Sakpere, A.B., Oluwadebi, A.G., Ajilore, O.H. and Malaka, L.E., 2021. The Impact of COVID-19 on the Academic Performance of Students: A Psychosocial Study Using Association and Regression Model.

Authors' Profiles

Rifat-Ibn-Alam has completed his Bachelor's in Computer Science & Engineering from American International University- Bangladesh (AIUB) in 2021. He was awarded the Summa Cum Laude award and Dean's List Honors multiple times for academic excellence during his undergraduate. His research interests include Data Mining, Data Analytics, and E-learning, and so on.

Md. Golam Ahsan Akib is pursuing his Bachelor's in Computer Science & Engineering from American International University- Bangladesh (AIUB) in 2021. He was awarded Dean's List Honors for academic excellence. His research interests include Data Analytics, Data Mining, and Machine Learning.

Nyme Ahmed has earned a Bachelor of Science (BSc) in Computer Science and Engineering (CSE) from the American International University-Bangladesh (AIUB) in 2021. During his BSc, he was awarded the Summa Cum Laude and Dean's List Honors for academic excellence several times. He is interested in research areas including Data Mining, Machine Learning, Data Analytics, Graph Theory, and a wide variety of Algorithms and Data Structures.

Syed Nafiul Shefat has received his Bachelor's in Computer Science & Engineering from American International University- Bangladesh (AIUB) in 2021. He was awarded the Summa Cum Laude and Dean's List Honors twice for academic excellence in his undergraduate period. His research interests are Data Mining, Data Analytics, Software Engineering, and Machine Learning.

Dr. Dip Nandi is currently working as a Professor and Director of Faculty of Science and Technology in American International University- Bangladesh. His research interest includes Data Mining, Software Engineering, Management Information Systems, E-learning etc. He can be contacted at dip.nandi@aiub.edu.

How to cite this paper: Rifat-Ibn-Alam, Md. Golam Ahsan Akib, Nyme Ahmed, Syed Nafiul Shefat, Dip Nandi," A Comparative Analysis among Online and On-Campus Students Using Decision Tree ", International Journal of Mathematical Sciences and Computing(IJMSC), Vol.8, No.2, pp. 11-27, 2022. DOI: 10.5815/ijmsc.2022.02.02