

Student performance classification and prediction in fully online environment using Decision tree

Musaddiq Al Karim, Mst. Yeasmin Ara, Md. Mahadi Masnad, Mostafa Rasel, Dip Nandi

Abstract— Knowledge Discovery and Data Mining (KDD) is a multidisciplinary field of study that focuses on methodologies for extracting useful knowledge from data. During the latest Covid-19 pandemic, there was a significant uptick in online-based learning (e-learning) operations as every educational institution moved its operations to digital channels. To increase the quality of education in this new normal, it is necessary to determine the key factors in students' performance. The main objective of this study is to exploit the regulating factors of education via digital platforms during the covid-19 pandemic by extracting knowledge and a set of rules by using the Decision Tree (j48) classifier. In this study, we developed a conceptual framework using four datasets, each with a different set of attributes and instances, collected from "X-University" and Microsoft teams. 'Final term' and 'Mid-term' examinations acted as the root node for all four datasets. The findings of this study would benefit higher education institutions by helping instructors and students to recognize the shortcomings and influences controlling students' performance in the online platforms during the covid-19 pandemic, as well as serve as an early warning framework for predicting students' deficiencies and low school performance.

Index Terms— Decision tree, Data mining, Association rules, Student performance analysis, Prediction, E-learning, Covid-19.

Musaddiq Al Karim
American International University-Bangladesh (AIUB)
Email: musaddiqmk19@gmail.com

Mst. Yeasmin Ara
American International University-Bangladesh (AIUB)
Email: yeasminararenu007@gmail.com

Md. Mahadi Masnad
American International University-Bangladesh (AIUB)
Email: atime691@gmail.com

Mostafa Rasel
American International University-Bangladesh (AIUB)
Email: mostafarasel9@gmail.com

Dip Nandi
Professor and Director, Faculty of Science and Technology American International University-Bangladesh (AIUB)
Email: dip.nandi@aiub.edu

I. INTRODUCTION

For the past few decades, there has been an emerging advancement in the educational sectors by adapting different technologies. But since the dreadful Covid-19 situation the education field has faced a substantial change. In order to maintain social distance and operate educational assessments uninterrupted, educational institutes had to quickly adapt to the situation that resulted in an unprecedented push to online learning [26]. All kinds of educational operations have been shifted online [27]. Everything that was being conducted manually is now being conducted on digital platforms.

Now almost every educational institute has digitized its educational information system. A large amount of data of these diversified factors and assessments is being generated by universities through several sources such as educational software, university databases, university portal, offline classroom environments, assignments, and coursework. Student information system of these educational institutes stores these vast amounts of data which possess immersed potential data which can be used in the enhancement of educational systems.

In the offline environment, it was much easier to pay attention to each student and do the needful. Even though there were some challenges at first as every educational operation was moved to online platforms, instructors and students have already become accustomed to the new normal. Online Education has become more important during Covid-19. In a circumstance when the whole world has shut down due to the pandemic, to stop the virus from spreading. As a result, education has changed dramatically, with the distinctive rise of e-learning, whereby teaching is undertaken remotely and on digital platforms [28]. In these circumstances continuing education offline would have been so dangerous. In online education educators and students are able to take and attend classes and conduct exams by staying at home.

Different data mining techniques made this procedure more effective. The purpose of data mining methods is to extract meaningful knowledge from data [29]. There are many kinds of data mining techniques and applications. Major applications include Classification, Numeric Prediction, Association, and Clustering. The application of data mining methods to educational data is referred to as Educational Data Mining (EDM) [30]. Manually analyzing huge amounts of data is time-consuming, can be incorrect, and may skip meaning full

results due to calculation error where data mining techniques made the automated process more efficient and provides more meaningful information. Educational Data Mining techniques are used to analyze and predict student performance. This aids in the improvement of educational quality by allowing educators to monitor students' academic progress.

This study focuses on determining the potential factors and generates a set of rules that imply generating students' performance analysis in online platforms during the covid-19 pandemic. We conducted this research by creating four datasets (with attribute and instance disparities) based on the data of students' performance in several courses on an online platform during the covid-19 pandemic.

This paper is organized accordingly: The second section of the document provides a summary and context report, the third section deals with the data collection, pre-processing, and the analyses of the data, and the fourth section shows the findings finally section 5 draws the conclusion, limitations, and future works.

II. LITERATURE REVIEW

The increase of online classes, degree programs, and educational institutions has been driven and continues to be facilitated by advances in computer-mediated communication technology [25]. Due to the covid-19 pandemic, all the educational institutes have quick-shifted to online platforms. It's a necessity to execute analysis to track which aspects perform a vital role in student performance on online platforms.

The performance of an individual student depends on diversified factors and assessments such as information about the academic progress of students, such as formative and additive assessment data, coursework, instructor observations, demographic information, psychological and socioeconomic data. Students' performance is the margin of academic progress, classifying their current stages. Extracting these data from the student information systems, it is possible to provide meaningful feedback to different stakeholders to analyze and predict students' performance as well as advance the quality of the educational process. The most significant aspect of this study is to effectively transform data into information to analyze and predict the future performance of students by learning from previous data. It is difficult for the faculty members, administration to effectively analyze this huge amount of data manually in the decision-making process. Whereas it was much easier to measure student performance in on Campus environment, measuring student performance in the online environment is equally difficult. This is where Educational Data Mining Techniques come to the rescue. For this study, we have extracted potential attributes and algorithms from previous papers on student performance analysis and prediction. This section- will explore the important factors from the background study in student performance analysis and prediction.

Researchers collected the most responsive and desired attribute values for analyzing student performance. A good number of studies used 'Gender' as a valuable attribute. [1, 2, 4, 6, 12, 13, 15, 17, 20, 21, 23]. As there are some different learning techniques of male and female students. Several

researchers look at assignment submission, quiz scores, and lab work, class test marks, seminal performance as influential factors in their studies [2, 3, 5, 9, 11, 16, 18, 22, 23, and 24]. A number of researchers regarded CGPA as one of the primary attributes from the academic perspective [12, 14, 15, 17, 18, and 21]. Furthermore, mid-exam marks, final-exam marks, final result previous semester grades are all established intellectual factors in predicting a student's future performance, and almost every researcher utilized these variables in their paper. Mostly final marks or result were divided into some predefined nominal values based on their numeric values for classification purposes and acted as classifier attribute [1,2,3,5,6,7,9,10,11,16,18,19,22,23]. As the attendance percentage of a student indicates a student's dedication towards study, few researchers also included this parameter in their work [3, 5, 10, 11, 12,13,16,18, 20, 22, and 23]. Along with attendance, a few researchers considered the number of absences for their study [12, 21]. A number of analysts considered some personal attributes like parents' occupation, parents' qualification, financial status, and living location, information about family members [1, 4, 8, 9, 10, 13, 18, 21, 22, and 23].

One of the most popular predictive modeling techniques is the decision tree, which is used for predicting and categorizing a given data object based on a previously generated model. ID3, Simple CART, C4.5, J48 were the most efficient decision trees used for classification purposes in predicting student's performance [1-11]. A decision tree produces sequences of rules that help for decision-making [10]. Consistently, investigators generated rules from visualized decision trees for a better understanding of the most impactful attribute and also of the final outcome [1-12, 14-21, 24].

Authors in [1] achieved the highest accuracy with 80.15% and 82.58% in 10-fold cross-validation method and then, generated 25 classification rules. Other decision tree classifiers were ranging from 77.86 to 79.58%. Another outline of the research [2] could find high potential attributes from the set of rules after applying ID3 classifier. In another study [5], generated rules showed that students who got poor marks in-session more likely to fail in the final exam. Furthermore, [7] found students who got A+ in that C++ course graduated with an excellent GPA. They applied the J48 decision tree classification algorithm. In [9], the author implemented ID3 decision tree on their dataset and got 92.5% accuracy. They knew from decision tree rules that family problems made students to dropout mostly and in terms of institutional factors, the campus environment caused the most dropouts of students. Moreover, all other studies got some valuable information from decision tree classification rules that assisted in their research work.

TABLE I: Common attributes used in the reviewed papers

| Factor | Attribute | Reference |
|-------------|-----------|---|
| Demographic | Gender | [1] [4] [5] [7] [9] [10] [13] [14] [15] [16] [17][20] |
| | Age | [1] [9] [10] [11] [14] [15] [16] [17] |

| | | |
|----------------------------------|---------------------------------|---------------------------------|
| | Fathers' Occupation | [4] [5] [11][13] [14] [15] [16] |
| | Mothers' occupation | [4] [5] [11][13] [15] [16] |
| | Fathers' Education Background | [4] [5][11][13][15] [16] [20] |
| | Monthly/ annual income | [4] [5] [13] [20] |
| | Mothers' Educational Background | [4] [5] [13] [15] [16] [20] |
| | Number of siblings | [4] [5] [7] [13] [15] [16] [20] |
| Academic | CGPA | [6] [7] [9] [11] [12] |
| | Course grade | [1] [10] [11] |
| | Previous semester grades | [1] [8] [9] [13] [20] |
| | Pre-requisite course grades | [1] [8] [9] [13] |
| | Class timing | [1] |
| | Section size nominal | [1] [11] |
| | Medium of teaching | [4] [5] [13] |
| | Attendance percentage | [1] [4] [8] [11] [13] [19][20] |
| | Number of absences | [1] [15] [16] |
| | Scholarship status | [1] [6] [7] [13] [15] |
| | High school name | [1] [6] [7] [9] [10] [14] [16] |
| | Course Load per semester | [19] |
| | Admission test marks | [1] [7] [9] [10] [17] |
| | Class test | [8] [11] [20] |
| | Assignment | [8] [11] [19][20][21] |
| | Lab evaluation | [8] [11] [19] |
| | Mid exam | [19] |
| Final Exam/End semester | [8] [11][12][19] [20] | |
| Psychological and socio-economic | Extra-curricular activities | [4] [6][12][15] [16] |
| | Health status | [4] [10] [16] |
| | Time spent on social media | [5] [6] [7] |

III. DATA COLLECTION, PREPROCESSING AND ANALYSIS

The previous chapter listed some possible attributes from various articles but not all of these attributes had been available due to the pandemic and the transition to online channels. We extracted 589 instances of seven courses from "X-University" and Microsoft Teams software. However, not all of these courses have the same characteristics. We have therefore developed four separate datasets based on their general characteristics, and our study focuses on them.

A. Dataset 1

This dataset was created using the following courses: the Algorithms, the Artificial Intelligence and the Expert System, the Introduction to Programming (Lab), Introduction to Programming (Theory), Object-Oriented Programming 1(Java), Object-Oriented Programming 2 (C#), and the Web Technologies. This dataset contains 6 attributes of 589 instances. Attributes of our first dataset are described in table 1.

TABLE II: Dataset 1

| Dataset Version 1 | | |
|-------------------|---------|--|
| Attribute name | Type | Summary |
| Gender | Nominal | F = 133, M = 456, {It was generated based on their names.} |
| Attendance (10) | Numeric | Generated via Microsoft teams software, scaled to 10. |
| Absence (10) | Numeric | Calculated by subtracting attendance value from 10. |
| Mid-term (100) | Numeric | Based on mid-term attendance, mid quiz marks, mid assignments, mid-term lab exams and Mid-term exam. |
| Final-term (100) | Numeric | Based on final term attendance, final term quiz marks, final term assignments, final term lab exams and final term exam |
| Class | Nominal | High Performer (Total marks>85) = 203 Medium Performer (70<Total marks<85) = 262 Low Performer (50<Total marks<69) = 93 Failure (0<Total marks<50) = 8 Dropped (Marked as -2) = 23 |

B. Dataset 2

Algorithms, Introduction to Programming (Theory), Object-Oriented Programming 1(Java), Object-Oriented Programming 2 (C#), and Web Technologies are the courses included in this dataset. In this update, we've introduced two new attributes to this dataset. There are now 8 attributes of 330 instances in this edition. We had to drop 259 instances from our previous dataset since the previous version did not include two attributes in this dataset. We included Quiz 1 and Quiz 2 attributes in this data collection, as they influence student success analysis by 40 percent. Attributes of our second dataset are described in table III below.

TABLE III: Dataset 2

| Dataset Version 2 | | |
|-------------------|---------|---|
| Attribute name | Type | Summary |
| Gender | Nominal | F = 72, M = 258 {It was generated based on their names} |
| Attendance (10) | Numeric | Generated via Microsoft teams software, scaled to 10 |
| Absence (10) | Numeric | Calculated by subtracting attendance value from 10. |
| Quiz 1 | Numeric | Midterm quiz average |
| Quiz 2 | Numeric | Final term quiz average |
| Mid-term (100) | Numeric | Based on mid-term attendance, mid quiz marks, mid assignments, mid-term lab exams and Mid-term exam. |
| Final-term (100) | Numeric | Based on final term attendance, final term quiz marks, final term assignments, final term lab exams and final term exam |
| Class | Nominal | High Performer (Total marks>85) = 114 Medium Performer |

| | | |
|--|--|--|
| | | (70<Total marks<85) = 125 Low Performer (50<Total marks<69) = 65 Failure (0<Total marks<50) = 5 Dropped (Marked as -2) = 21 |
|--|--|--|

C. Dataset 3

This dataset was created using the following courses: Algorithms, Introduction to Programming (Theory), Object-Oriented Programming 1 (Java), and Object-Oriented Programming 2 (C#). It has 12 attributes for a total of 280 instances. Attendance, Absence, and Quiz marks are split into two categories: Midterm and Final-Term. The rest of the attributes are the same as in previous datasets. Attributes of our third dataset are described in table IV below.

TABLE IV: Dataset

| Dataset Version 3 | | |
|----------------------------|---------|--|
| Attribute name | Type | Summary |
| Gender | Nominal | F = 61, M = 219, {It was generated based on their names} |
| Mid-term Attendance (10) | Numeric | Generated via Microsoft teams software for midterm only, scaled to 10 |
| Mid-term Absence (10) | Numeric | Calculated by subtracting midterm attendance value from 10. |
| Mid-term Quiz 1 | Numeric | First Quiz of Midterm |
| Mid-term Quiz 2 | Numeric | Second Quiz of Midterm |
| Mid-term (100) | Numeric | Based on mid-term attendance, mid quiz marks, mid assignments, mid-term lab exams and Mid-term exam. |
| Final-term Attendance (10) | Numeric | Generated via Microsoft teams software for final term only, scaled to 10 |
| Final-term Absence (10) | Numeric | Calculated by subtracting final term attendance value from 10. |
| Final-term Quiz 1 | Numeric | First Quiz of Final term |
| Final-term Quiz 2 | Numeric | Second Quiz of Final term |
| Final-term (100) | Numeric | Based on final term attendance, final term quiz marks, final term assignments, final term lab exams and final term exam |
| Class | Nominal | High Performer (Total marks>85) = 101 Medium Performer (70<Total marks<85) = 103 Low Performer (50<Total marks<69) = 54 Failure (0<Total marks<50) = 5 Dropped (Marked as -2) = 17 |

D. Dataset 4

We considered 10 attributes from two courses in this dataset: Algorithms and Web Technologies. There are two new attributes in this dataset: CGPA and Lab results. CGPA is one of the most promising attributes for analyzing and forecasting student results. There are only 91 examples. We were unable to discuss alternative courses due to a lack of evidence. There are 91 instances only. Due to a lack of data, other courses were dropped from this dataset. Attributes of our fourth and final dataset are described in table V below.

TABLE V: Dataset 4

| Dataset Version 4 | | |
|-------------------|---------|--|
| Attribute name | Type | Summary |
| CGPA | Numeric | CGPA until current semester (While extracting data). |
| Gender | Nominal | F = 22, M = 69, {It was generated based on their names} |
| Attendance (10) | Numeric | Generated via Microsoft teams software, scaled to 10. |
| Absence (10) | Numeric | Calculated by subtracting attendance value from 10. |
| Quiz 1 | Numeric | Midterm quiz average |
| Quiz 2 | Numeric | Final term quiz average |
| Lab Performance | Numeric | Based on lab quiz, lab exam and lab report marks. |
| Mid-term (100) | Numeric | Based on mid-term attendance, mid quiz marks, mid assignments, mid-term lab exams and Mid-term exam. |
| Final-term (100) | Numeric | Based on final term attendance, final term quiz marks, final term assignments, final term lab exams and final term exam |
| Class | Nominal | High Performer (Total marks>85) = 19 Medium Performer (70<Total marks<85) = 41 Low Performer (50<Total marks<69) = 21 Failure (0<Total marks<50) = 0 Dropped (Marked as -2) = 10 |

IV. RESULTS AND DISCUSSION

From this study, by achieving a pattern from decision tree (J48) classifier, 'Final-term' and 'Mid-term'; these two attributes acted as the most dominating factor to classify students' performance. The derived decision trees in fig 1-4, show that the 'Final-term' attribute acted as the root node for all datasets in our classification model.

In dataset-1, Attendance played a very impactful role in the performance of students. The male students who got less than 66 marks in the Mid-term but got better marks in the Final-term having an attendance of more than 7.75 become 'Medium performer'. In spite of having more than 78 marks in both mid and final terms, less than 9.75 attendance marks students become 'Medium performer'. In the same case, more than 9.75 attendance students became a 'High performer'.

In dataset-2, Quiz was the most influential attribute. But the students, who got lesser marks on Quiz, ultimately did better at the end of the semester. Most of the students who lag behind in the early days of a semester tried to do better in the final term.

On dataset-3, in the Mid-term, students having more than 77.5 to 84 marks and in the final term having more than 79 to 85 marks and also having a good number in quiz 1 more than 18 became a 'High performer'. Here also Quiz marks influence student's performance besides examination marks.

In our fourth dataset, after getting a decent number in both mid and final terms because of low attendance of less than 6.6, in the end, students got a low grade. Otherwise, students could become medium or high performers.

The knowledge reflected by the decision tree (j48) can be extracted and expressed as IF-THEN law. Extraction of hidden

information from the study reveals that the classification criteria can be met using only three or four attributes. The best rules were chosen, with the 'Class' designated as the students' focus. The rules obtained from the decision tree (J48) classifier are listed below according to datasets.

TABLE VI: Derived set of rules from Dataset 1

| Dataset Version 1: |
|---|
| R1: Final term ≤ -2 , then class = 'Dropped' |
| R2: Final term ≤ 66 and Mid-term > 79 then class = 'Medium Performer' |
| R3: Final term ≤ 66 and Mid-term ≤ 73 then class = 'Low performer' |
| R4: Final term ≤ 66 and Mid-term ≤ 79 , Mid-term > 73 and Final term ≤ 61 then class = 'Low performer' otherwise = 'Medium Performer' |
| R5: Final term > 66 , Mid-term ≤ 66 and Gender = F then class = 'Medium Performer' |
| R6: Final term > 66 , Mid-term ≤ 59 and Gender = M then class = 'Low Performer' |
| R7: Final term > 66 , Mid-term ≤ 66 and Gender = M, Mid-term > 59 and attendance < 7.75 , then class = 'Low performer' |
| R8: Final term > 66 , Mid-term ≤ 66 and Gender = M, Mid-term > 59 and attendance > 7.75 , and attendance < 8.75 then class = 'Medium performer' |
| R9: Final term > 66 , Mid-term ≤ 66 and Gender = M, Mid-term > 59 and attendance > 8.75 , and attendance ≤ 9.75 then class = 'Low performer' otherwise class= 'Medium Performer' |
| R10: Final term > 78 , Mid-term ≤ 81 , Final term ≤ 87 , Mid-term ≤ 78.2 then class = 'Medium Performer' |
| R11: Final term > 78 , Mid-term ≤ 81 , Final term ≤ 87 , Mid-term > 78.2 , Mid-term ≤ 80 , attendance ≤ 9.75 then class = 'Medium performer' otherwise class= 'High Performer' |
| R12: Final term > 78 , Mid-term ≤ 81 , Final term ≤ 87 , Mid-term > 78.2 , Mid-term > 80 then class = 'Medium performer' |
| R13: Final term > 78 , Mid-term ≤ 81 , Final term > 87 , Mid-term ≤ 75 , Final-term ≤ 89 then class = 'Medium performer' otherwise class= 'High Performer' |
| R14: Final term > 78 , Mid-term ≤ 81 , Final term > 87 , Mid-term > 75 the class= 'High Performer' |
| R15: Final term > 78 , Mid-term > 81 , Final term ≤ 83 , Gender = F then class= 'High Performer' |
| R16: Final term > 78 , Mid-term > 81 , Final term ≤ 83 , Gender = M, Mid-term ≤ 86 and attendance ≥ 9.75 then class = 'Medium performer' |
| R17: Final term > 78 , Mid-term > 81 , Final term ≤ 83 , Gender = M, Mid-term ≤ 86 and attendance > 9.75 , Final term ≤ 79 then class = 'Medium performer' |
| R18: Final term > 78 , Mid-term > 81 , Final term ≤ 83 , Gender = M, Mid-term ≤ 86 and attendance > 9.75 , Final term ≤ 79 , Mid-term ≤ 84 then class = 'Medium performer' otherwise class= 'High Performer' |
| R19: Final term > 78 , Mid-term > 81 , Final term > 83 then class= 'High Performer' |
| R20: Final term > 78 , Mid-term > 86 , Final term ≤ 83 , Gender = M then class= 'High Performer' |

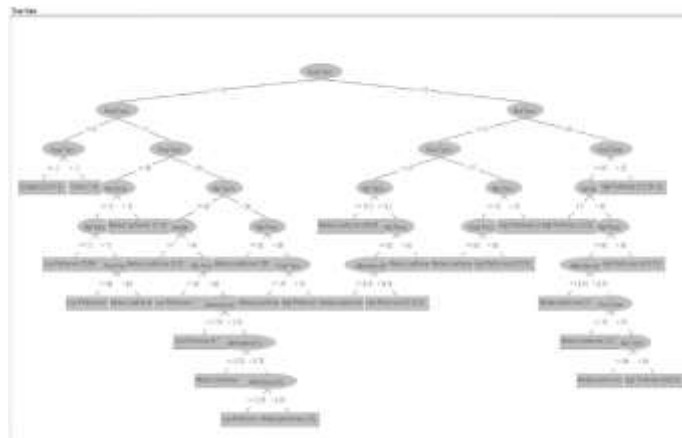


Fig. 1. Decision tree derived from Dataset 1

TABLE VII: Derived set of rules from Dataset 2

| Dataset Version 2: |
|---|
| R1: Final term ≤ -2 , then class = 'Dropped' |
| R2: Final term ≤ 0 , Final term > -2 then class = 'Failure' |
| R3: Final term > 0 , Final term ≤ 66 , Mid-term ≤ 78 , Quiz 1 ≤ 3 then class= 'Medium Performer' otherwise class= 'Low Performer' |
| R4: Final term > 66 , Mid-term ≤ 65 , Quiz 2 ≤ 8.5 , Quiz 1 ≤ 7 the class = 'Medium Performer' otherwise class= 'Low Performer' |
| R5: Final term > 66 , Mid-term > 65 then class= 'Medium Performer' |
| R6: Final term > 79 , Mid-term ≥ 77.5 , Final term ≤ 89 , Quiz 1 ≤ 5 then class = 'High Performer' otherwise class= 'Medium Performer' |
| R7: Final term > 79 , Mid-term ≤ 77.5 , Final term > 89 then class = 'High Performer' |
| R8: Final term > 79 , Mid-term > 84 then class = 'High Performer' |
| R9: Final term > 79 , Mid-term > 77.5 , Mid-term ≤ 84 , Final term ≤ 89 , Quiz 1 ≤ 10 then class = 'High Performer' otherwise class= 'Medium Performer' |
| R10: Mid-term > 77.5 , Mid-term ≤ 84 , Final term > 89 then class = 'High Performer' |

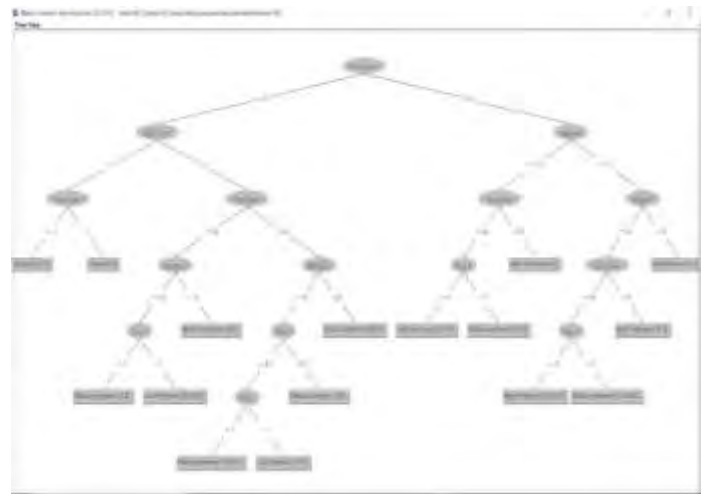


Fig. 2. Decision tree derived from Dataset 2

TABLE VIII: Derived set of rules from Dataset 3

| Dataset Version 3: |
|--|
| R1: Final term ≤ -2 , then class = 'Dropped' |
| R2: Final term ≤ 0 , Final term > -2 then class = 'Failure' |
| R3: Final term ≤ 74 , Mid-term ≤ 67 then class = 'Low Performer' |
| R4: Final term ≤ 79 , Mid-term ≤ 67 , Final term > 74 , Mid-term ≤ 58 then class = 'Low Performer' otherwise class = 'Medium Performer' |
| R5: Mid-term > 67 , Final term ≤ 66 , Mid-term ≤ 78 then class = 'Low Performer' otherwise class = 'Medium Performer' |
| R6: Final term ≤ 79 , Mid-term > 65 , Final term > 66 then class = 'Medium Performer' |
| R7: Final term > 79 , Mid-term ≤ 77.5 , Final term ≤ 89 then class = 'Medium Performer' otherwise class = 'High Performer' |
| R8: Final term > 79 , Mid-term > 77.5 , Mid-term ≤ 84 , Final term ≤ 85 , Final term Quiz 1 ≤ 18 , Final term Quiz 1 ≤ 7 then class = 'High Performer' otherwise class = 'Medium Performer' |
| R9: Final term > 79 , Mid-term > 77.5 , Mid-term ≤ 84 , Final term ≤ 85 , Final term Quiz 1 > 18 then class = 'High Performer' |
| R10: Mid-term > 77.5 , Mid-term ≤ 84 , Final term > 85 then class = 'High Performer' |
| R11: Final term > 79 , Mid-term > 84 then class = 'High Performer' |

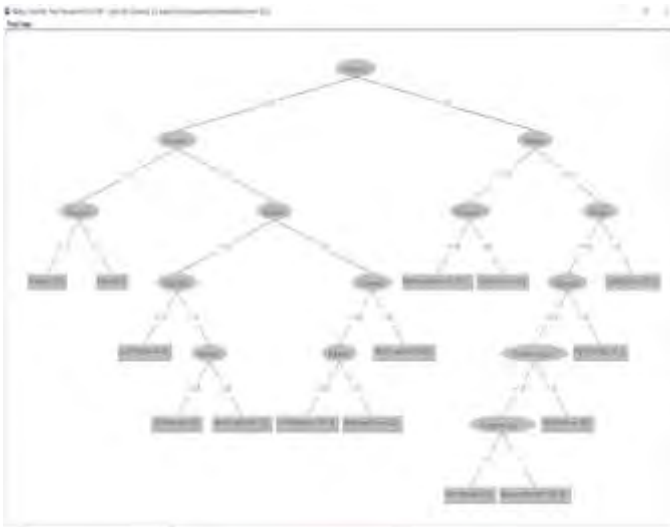


Fig. 3. Decision tree derived from Dataset 2

TABLE IX: Derived set of rules from Dataset 4

| Dataset Version 4: |
|---|
| R1: Final term ≤ -2 , then class = 'Dropped' |
| R2: Final term ≤ 65 , Final term ≤ -2 , Mid-term ≤ 77 then class = 'Low Performer' otherwise class = 'Medium Performer' |
| R3: Final term > 65 , Final term ≤ 80 , Attendance ≤ 6.6 then class = 'Low performer' otherwise class = 'Medium performer' |
| R4: Final term > 80 , Mid-term ≤ 73 then class = 'Medium Performer' otherwise class = 'High Performer' |

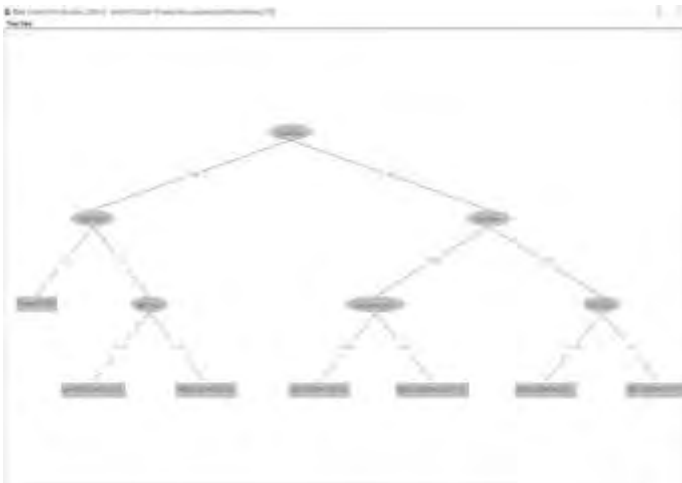


Fig. 4. Decision tree derived from Dataset 4

V. CONCLUSION

The main objective of this study was to generate a set of rules and determining the dominating factors to analyze and predict students' performance on online platforms during the covid-19 pandemic. We extracted students' data from a university server and a digital platform 'Microsoft Teams' and created four different datasets for this experimental approach. Using the 10-folds cross-validation technique on the decision tree (j48) classifier, we generated a set of rules for all four datasets. Knowledge extracted from the decision trees shows that 'Final term' and 'Mid-term' are the most prominent attributes to analyze and predict students' performance where other attributes i.e. Quiz, Gender, Attendance had less impact.

We didn't add the remaining attributes because we didn't find any influence in our research. However, this study had a few limitations. First of all, this study focused on only 7 courses. Secondly, only one semester of data was studied in this research. Thirdly, only one classifier was used to build this model. Future researches are encouraged to include more courses, longer periods of data, and other classifiers to build a model to extract knowledge to predict and analyze students' performance.

REFERENCES

- [1] Khan, A., & Ghosh, S. K. (2018). Data mining based analysis to explore the effect. Springer.
- [2] Angeline, D. M. D. (2013). Association Rule Generation for Student Performance Analysis using Apriori Algorithm. The SIJ Transactions on Computer Science Engineering & Its Applications (CSEA), 01(01), 16–20. <https://doi.org/10.9756/sijcsea/v1i1/01010252>.
- [3] Pandey, M., & Kumar Sharma, V. (2013). A Decision Tree Algorithm Pertaining to the Student Performance Analysis and Prediction. International Journal of Computer Applications, 61(13), 1–5. <https://doi.org/10.5120/9985-4822>.
- [4] Angeline, D. M. D. (2013). Association Rule Generation for Student Performance Analysis using Apriori Algorithm. The SIJ Transactions on Computer Science Engineering & Its Applications (CSEA), 01(01), 16–20. <https://doi.org/10.9756/sijcsea/v1i1/01010252>.
- [5] Raut, A. B., & Nichat, A. A. (2017). Students Performance Prediction Using Decision Tree Technique. International Journal of Computational Intelligence Research, 13(7), 1735–1741. Retrieved from <http://www.ripublication.com>.
- [6] Badr El Din Ahmed, A., & Sayed Elaraby, I. (2014). PER: A prediction for Student's Performance Using Decision Tree ID3 Method. India - World Journal of Computer Application and Technology, 2(2), 43–47. Retrieved from <http://www.hrpub.org>
- [7] Al-Barrak, M. A., & Al-Razgan, M. (2016). Predicting Students Final GPA Using Decision Trees: A Case Study. International Journal of Information and Education Technology, 6(7), 528–533. <https://doi.org/10.7763/ijiet.2016.v6.745>.
- [8] Long, X., & Wu, Y. (2012). Application of decision tree in student achievement evaluation. Proceedings - 2012 International Conference on Computer Science and Electronics Engineering, ICCSEE 2012, 2, 243–247. <https://doi.org/10.1109/ICCSEE.2012.169>.
- [9] Long, X., & Wu, Y. (2012). Application of decision tree in student achievement evaluation. Proceedings - 2012 International Conference on Computer Science and Electronics Engineering, ICCSEE 2012, 2, 243–247. <https://doi.org/10.1109/ICCSEE.2012.169>.
- [10] Lakshmi, T. M., Martin, A., Begum, R. M., & Venkatesan, V. P. (2013). An Analysis on Performance of Decision Tree Algorithms using Student's Qualitative Data. International Journal of Modern Education and Computer Science, 5(5), 18–27. <https://doi.org/10.5815/ijmecs.2013.05.03>.
- [11] Kumar, B., & Pal, S. (2011). Mining Educational Data to Analyze Students Performance. International Journal of Advanced Computer Science and Applications, 2(6). <https://doi.org/10.14569/ijacsa.2011.020609>.
- [12] Saa, A. A., Al-Emran, M., & Shaalan, K. (2020). Mining Student Information System Records to Predict Students' Academic Performance. Advances in Intelligent Systems and Computing (Vol. 921). Springer International Publishing. https://doi.org/10.1007/978-3-030-14118-9_23.
- [13] Ramaswami, M., & Bhaskaran, R. (2010). A CHAID Based Performance Prediction Model in Educational Data Mining, 7(1), 10–18. Retrieved from <http://arxiv.org/abs/1002.1144>.
- [14] Shahiri, A. M., Husain, W., & Rashid, N. A. (2015). A Review on Predicting Student's Performance Using Data Mining Techniques. Procedia Computer Science, 72, 414–422. <https://doi.org/10.1016/j.procs.2015.12.157>.
- [15] Suljic, M. (2012). Data Mining Approach for Predicting Student Performance. Economic Review: Journal of Economics and Business, X (1), 3–12.

- [16] Pisani, S., Fioriti, D., Conte, M. P., Chiarini, F., Seganti, L., & Degener, A. M. (2002). Involvement of herpes simplex virus type 2 in modulation of gene expression of human papillomavirus type 18. *International Journal of Immunopathology and Pharmacology*, 15(1), 59–63. <https://doi.org/10.1177/039463200201500108>.
- [17] Kabakchieva, D. (2013). Predicting student performance by using data mining methods for classification. *Cybernetics and Information Technologies*, 13(1), 61–72. <https://doi.org/10.2478/cait-2013-0006>.
- [18] Mueen, A., Zafar, B., & Manzoor, U. (2016). Modeling and Predicting Students' Academic Performance Using Data Mining Techniques. *International Journal of Modern Education and Computer Science*, 8(11), 36–42. <https://doi.org/10.5815/ijmecs.2016.11.05>.
- [19] Hasan, R., Palaniappan, S., Raziff, A. R. A., Mahmood, S., & Sarker, K. U. (2018). Student Academic Performance Prediction by using Decision Tree Algorithm. 2018 4th International Conference on Computer and Information Sciences: Revolutionising Digital Landscape for Sustainable Smart Society, ICCOINS 2018 - Proceedings, 1–5. <https://doi.org/10.1109/ICCOINS.2018.8510600>.
- [20] Abu, A. (2016). Educational Data Mining & Students' Performance Prediction. *International Journal of Advanced Computer Science and Applications*, 7(5), 212–220. <https://doi.org/10.14569/ijacsa.2016.070531>.
- [21] Salal, Y. K., Abdullaev, S. M., & Kumar, M. (2019). Educational data mining: Student performance prediction in academic. *International Journal of Engineering and Advanced Technology*, 8(4C), 54–59.
- [22] Yassein, N. A., M Helali, R. G., & Mohomad, S. B. (2017). Predicting Student Academic Performance in KSA using Data Mining Techniques. *Journal of Information Technology & Software Engineering*, 07(05). <https://doi.org/10.4172/2165-7866.1000213>.
- [23] Patil, R., & Tamane, S. (2018). A comparative analysis on the evaluation of classification algorithms in the prediction of diabetes. *International Journal of Electrical and Computer Engineering*, 8(5), 3966–3975. <https://doi.org/10.11591/ijece.v8i5.pp3966-3975>.
- [24] Hooshyar, D., Pedaste, M., & Yang, Y. (2020). Mining educational data to predict students' performance through procrastination behavior. *Entropy*, 22(1), 12. <https://doi.org/10.3390/e22010012>.
- [25] Nandi, D., Hamilton, M., Harland, J., & Mahmood, S. (2015). Investigation of Participation and Quality of Online Interaction. *International Journal of Modern Education and Computer Science*, 7(8), 25–37. <https://doi.org/10.5815/ijmecs.2015.08.04>.
- [26] Teräs, M., Suoranta, J., Teräs, H., & Curcher, M. (2020). Post-Covid-19 Education and Education Technology 'Solutionism': a Seller's Market. *Postdigital Science and Education*, 2(3), 863–878. <https://doi.org/10.1007/s42438-020-00164-x>.
- [27] Abdalla, Jihan. "US universities switch to online courses due to coronavirus", *AL JAZEERA*, 10 Mar 2020.
- [28] Li, Cathy ; Lalani, Farah. "The COVID-19 pandemic has changed education forever. This is how", *WORLD ECONOMIC FORUM*, 29 Apr 2020.
- [29] Han, Jiawei, Micheline Kamber, and Jian Pei. *Data Mining: Concepts and Techniques*, Third Edition. 3rd ed. Waltham, Mass.: Morgan Kaufmann Publishers, 2012.
- [30] Baker, R. S., & Yacef, K. (2009). The State of Educational Data Mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining*, 1(1), 3-17. <https://doi.org/10.5281/zenodo.3554657>.



Musaddiq Al Karim has completed his Bachelor of Science in Computer Science and Engineering degree from American International University-Bangladesh. He worked as an Undergraduate Teaching Assistant intern at American International University-Bangladesh and Software Quality Assurance Engineer intern at Olivine Limited. Research enthusiast and highly interested in Artificial Intelligence, Data mining, Big Data, Machine Learning, Blockchain, IoT and Human computer Interaction. Also, he is planning for pursuing MSc degree.



Mst. Yeasmin Ara has completed her Bachelor of Science (BSc) in Computer Science & Engineering (CSE) at American International University - Bangladesh (AIUB). She has completed her internship as a Business Analyst at Akij Group department of Information Technology. And currently she is planning for pursuing MSc degree. She is highly interested in IoT, Data Mining as research area and Software Quality and Assurance.



Md. Mahadi Masnad has received his BSc in Computer Science and engineering from American International University-Bangladesh. He also completed his internship as a Teaching Assistant at American International University-Bangladesh. His areas of interest in research include artificial intelligence, machine learning, cloud computing, human computer interaction and also in different algorithms.



Mostafa Rasel completed his BSc in Computer Science & Engineering (CSE) at American International University - Bangladesh (AIUB). He has held an internship at Banglalink and another at Global Islami Bank Ltd. He is highly interested in IoT and networking as a research area. Currently, he is planning for his MSc degree.



Dr. Dip Nandi has completed his Master Degree on Information Systems from The University of Melbourne, Australia in 2009. Later he finished his Doctor of Philosophy (PhD) in Computer Science from RMIT University Melbourne, Victoria, Australia.

Dr. Dip Nandi is the current Professor and Honorable Director of Faculty of Science & Technology (FST), at American International University- Bangladesh (AIUB). He is a former lecturer at RMIT University Melbourne, Australia from the year 2010 to 2012. Dr. Nandi has a vast range of research activities and contributions in various filed of Computer Science and Multidimensional researches. His profound knowledge over many leading domains and areas include the concept of Algorithmic Design, Software Engineering model & process, Machine learning, Data Warehousing, E Learning are mostly notable. Dr. Dip Nandi also has his Contributions in Alzheimer's disease and Dementia detection using Neural Networks. The education-based researches are also his area of expertise.