

Dataset of Students' Performance Using Student Information System, Moodle and the Mobile Application "eDify"

Raza Hasan ^{1,2,*} , Sellappan Palaniappan ¹, Salman Mahmood ¹, Ali Abbas ² and Kamal Uddin Sarker ³ 

¹ Department of Information Technology, School of Science and Engineering, Malaysia University of Science and Technology, Petaling Jaya 47810, Malaysia; sell@must.edu.my (S.P.); salman.mahmood@pg.must.edu.my (S.M.)

² Department of Computing, Middle East College, Knowledge Oasis Muscat, P.B. No. 79, Al Rusayl 124, Oman; aabbas@mec.edu.om

³ School of Informatics and Applied Mathematics, University Malaysia Terengganu, Kuala Terengganu 21030, Malaysia; ku_sarker@yahoo.com

* Correspondence: raza.hasan@pg.must.edu.my; Tel.: +968-9819-9513

Abstract: The data presented in this article comprise an educational dataset collected from the student information system (SIS), the learning management system (LMS) called Moodle, and video interactions from the mobile application called "eDify." The dataset, from the higher educational institution (HEI) in Sultanate of Oman, comprises five modules of data from Spring 2017 to Spring 2021. The dataset consists of 326 student records with 40 features in total, including the students' academic information from SIS (which has 24 features), the students' activities performed on Moodle within and outside the campus (comprising 10 features), and the students' video interactions collected from eDify (consisting of six features). The dataset is useful for researchers who want to explore students' academic performance in online learning environments, and will help them to model their educational datamining models. Moreover, it can serve as an input for predicting students' academic performance within the module for educational datamining and learning analytics. Furthermore, researchers are highly recommended to refer to the original papers for more details.

Dataset: <https://zenodo.org/record/5591907> (accessed on 18 October 2021).

Dataset License: CC-BY 4.0.

Keywords: educational datamining; learning management system; prediction; student academic performance; student information system



Citation: Hasan, R.; Palaniappan, S.; Mahmood, S.; Abbas, A.; Sarker, K.U. Dataset of Students' Performance Using Student Information System, Moodle and the Mobile Application "eDify". *Data* **2021**, *6*, 110. <https://doi.org/10.3390/data6110110>

Academic Editors: Leonardo Grilli, Carla Rampichini, Maria Cecilia Verri and Donatella Merlini

Received: 10 August 2021

Accepted: 19 October 2021

Published: 22 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Summary

Higher educational institutions (HEIs) employ a variety of learning approaches based on information and communications technology (ICT). These approaches involve different learning environments to facilitate the teaching and learning process with ease and dissemination of knowledge to their learners. Moreover, these environments keep track of the users and their interactions within these environments for auditing and recovery purposes. The logs can help stakeholders with valuable learning data, and when analyzed effectively, can help to provide a better learning experience to learners. Reports generating different users/courses can be used to evaluate the efficacy of the courses and the progress of the learners. Insights can help cater different learning styles, which helps to determine the complexity of courses, identifying specific parts of the content that cause problems in understanding the concepts and gaining insights into the future performance of learners.

Many HEIs use machine learning (ML) to discover associated patterns from these learning environments for better decision making and datamining (DM) to improve decision-making models using artificial intelligence (AI). HEIs require educational datamining

(EDM) for a better understanding of learners' behaviors in these learning environments that have the potential to impact educational practices [1].

The provided dataset was from the students of Middle East College (MEC), Muscat, Oman, studying in a computing specialization from the sixth semester and above. Historical data about student academics (extracted from SIS), student logs (extracted from Moodle, where the time spent engaging in activities was considered), and video interactions on blended learning material (extracted from logs of the mobile application eDify). For the students' academic data, the SIS parameters included student demographic data, academic data, degree plan, and academic integrity violations (AIVs). The academic and AIV data were considered. For the students' activities, the Moodle log parameters included logs of course activity, logs of site activity, live logs, site administration settings, and view log capabilities—although, only logs of course activity were considered. Different from the aforementioned, the eDify logs consist of video interactions for each student, indicating attributes such as played, paused, likes, and number of segments replayed within the video—all of which were selected. The Supplementary Material provided within this paper is the raw and filtered data.

To predict student performance based on the datamining approach, many studies have been carried out [2–7]. Nevertheless, these studies have primarily focused on demographic data, and predictions have been carried out based on activities performed in the online environment. However limited research has been conducted based on analyzing the video interactions of learners in a video-assisted course [8–12]. The provided dataset contains SIS, Moodle, and video-assisted course data (eDify), which can help researchers to understand video learning analytics using EDM, thereby enhancing the teaching and learning process.

The dataset provided aimed to predict student performance using EDM. The dataset contained 326 observations, where each observation represents an individual student and has 40 attributes. The application of the dataset can provide the research community to benchmark EDM tasks performed on longitude and latitude datasets. This can help to understand student academic performance (SAP) modeling and prediction using datamining techniques [13,14]. Furthermore, it can be combined with other online environments such as Moodle and online video streaming to understand the behaviors of their learners [15–17].

2. Data Description

The presented dataset was classified into three categories: Student academic information (Sections 2.1 and 2.2), student activity (Sections 2.3 and 2.4), and student video interactions (Section 2.5). First, student academic information was collected from SIS. Second, student activity information was collected from the activities performed on Moodle. Lastly, student video interactions were collected from the mobile application “eDify”. Figure 1 shows the mapping on how the dataset was formed.

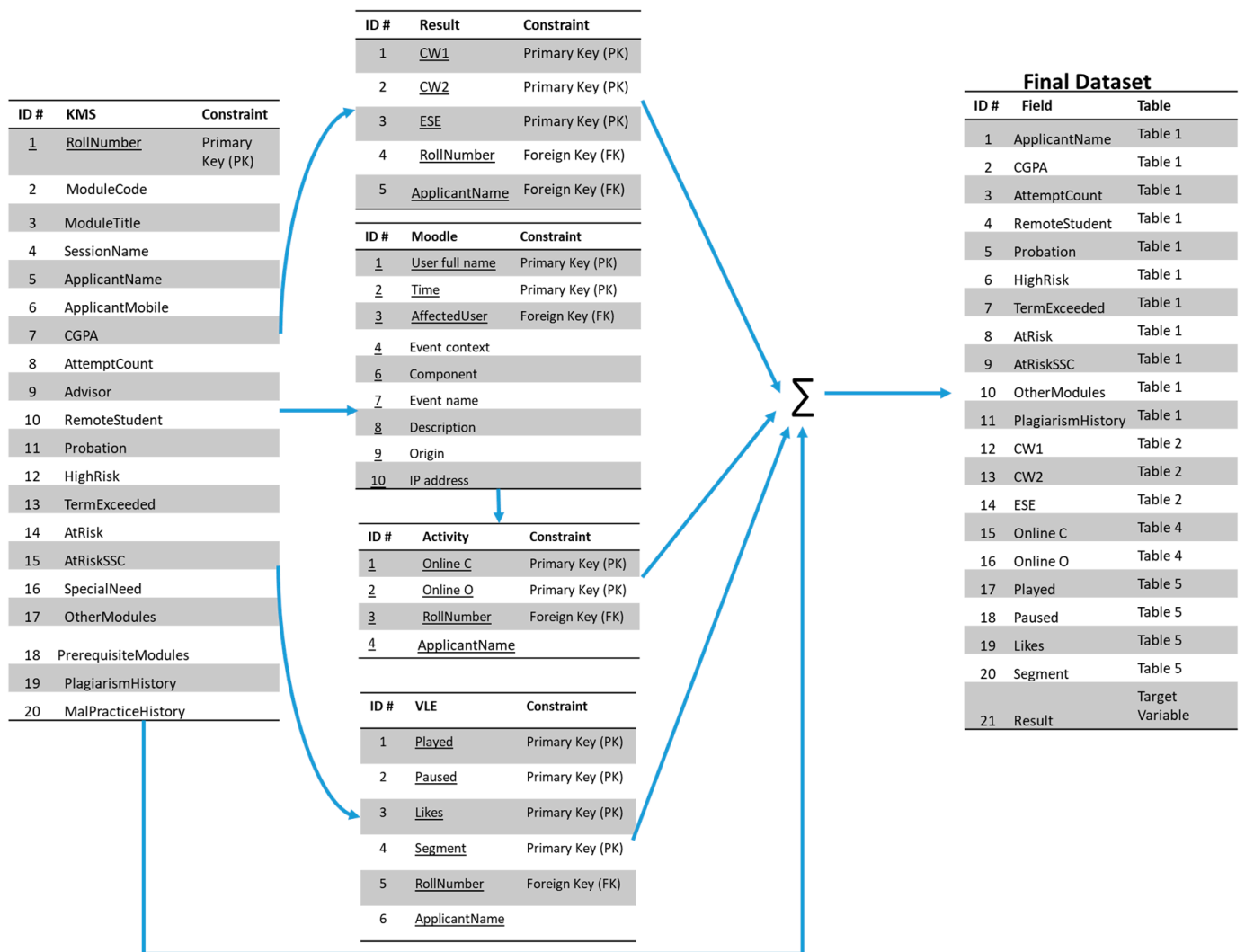


Figure 1. Dataset mapping.

2.1. Student Academic Information

Ten comma-separated value (CSV) files of “KMS Module <Number> <Semester>,” which contain “Know My Student” detail features, were extracted from SIS, with 20 attributes. Table 1 summarizes these attributes, accompanied by a brief description.

Table 1. Attribute and descriptions exported from SIS.

| Attribute | Description |
|--------------------|--|
| ModuleCode | Code of the module in which the student has been registered, with a nominal data type such as “Module 1” |
| ModuleTitle | Title of the module in which the student has been registered, with a nominal data type such as “Course 2” |
| Session | Shows the session in which the student has been registered, with a nominal data type such as “Session-A” |
| RollNumber | Identification number of the student, with a nominal data type such as “21S1234” |
| ApplicantName | Name of the student, with a nominal data type such as “Student 1” |
| ApplicantMobile | Mobile number of the student, with a discrete data type such as “12345678” |
| CGPA | Cumulative grade point average of the student, with a discrete data type such as “4.0” |
| AttemptCount | The number of attempts in the module, with a discrete data type such as “1” |
| RemoteStudent | Either the student is under remote study mode or not, with a nominal data type such as “Yes/No” |
| Probation | Either the student has a backlog of modules to clear, with a nominal data type such as “Yes/No” |
| HighRisk | The high failure rate in a module, with a nominal data type such as “Yes/No” |
| TermExceeded | Progression rate of the student in the degree plan, with a nominal data type such as “Yes/No” |
| AtRisk | Previously failed two or more modules, with a nominal data type such as “Yes/No” |
| AtRiskSSC | Whether the student been registered by the student success center for any educational deficiencies, with a nominal data type such as “Yes/No” |
| SpecialNeeds | Whether the student been registered by the student success center for any special needs, with a nominal data type such as “Yes/No” |
| OtherModules | A student registered in any other modules in the current semester, with a numeric data type such as “1” |
| PrerequisiteModule | Prerequisite module registration, with a nominal data type such as “Yes/No” |
| PlagiarismHistory | Onto which modules the student has been booked for academic integrity violation, including module and academic year, with a nominal data type such as “Module 3” |

2.2. Student Academic Performance

Ten comma-separated value (CSV) files of “Result Module <Number> <Semester>”, containing the overall results in the modules extracted from SIS, and six attributes (including “RollNumber”, “ApplicantName” and “Session”) and three new attributes were extracted. Table 2 summarizes these attributes, accompanied by a brief description.

Table 2. Attributes and descriptions exported from SIS.

| Attribute | Description |
|-----------|--|
| CW1 | Marks obtained by the student in their first coursework, with a discrete data type such as “86.5” |
| CW2 | Marks obtained by the student in their second coursework, with a discrete data type such as “86.5” |
| ESE | Marks obtained in the end semester examination, with a discrete data type such as “86.5” |

2.3. Student Moodle Logs

Ten comma-separated value (CSV) files of “Moodle Module <Number> <Semester>” containing nine attributes were extracted. Table 3 summarizes these attributes, accompanied by a brief description.

Table 3. Attributes and descriptions exported from Moodle.

| Attribute | Description |
|----------------|--|
| Time | Timestamp of each user in the specific module, with a discrete data type such as “30/06/18, 22:22” |
| User full name | Registered name of the user in Moodle when logged in, with a nominal data type such as “Student 1” |
| Affected user | The registered user when they upload any file in Moodle, with a nominal data type such as “Student 3” |
| Event context | The activity accessed in Moodle by the registered user, with a nominal data type such as “File: Week4” |
| Component | The component accessed in Moodle by the registered user, with a nominal data type such as “System/File” |
| Event name | The specific event triggered in Moodle by the registered user, with a nominal data type such as “Course module viewed” |
| Description | Description of the activity performed in Moodle by the registered user, with a nominal data type such as “The user with id ‘4357’ viewed the course with id ‘158’” Origin: Origin of the activity performed in Moodle by the registered user, with a nominal data type such as “web” |
| IP address | Internet protocol address of the registered user accessing the Moodle, with a nominal data type such as “192.168.99.3” |

2.4. Student Online Activity on Moodle

Ten comma-separated value (CSV) files of “Activity Module <Number> <Semester>” containing “RollNumber” and “ApplicantName” and two new attributes were extracted. Table 4 summarizes these attributes, accompanied by a brief description.

Table 4. Attributes and descriptions exported from Moodle.

| Attribute | Description |
|-----------|--|
| Online C | User-performed activities within campus (in minutes), with a discrete data type such as “25” |
| Online O | User-performed activities outside of campus (in minutes), with a discrete data type such as “25” |

2.5. Student Video Interaction

Ten comma-separated value (CSV) files of “VL Module <Number> <Semester>” containing “RollNumber” and “ApplicantName” and four new attributes were extracted. Table 5 summarizes these attributes, accompanied by a brief description.

Table 5. Attributes and descriptions exported from eDify.

| Attribute | Description |
|-----------|--|
| Played | The number of times the video has been played |
| Paused | The number of times the video has been paused |
| Likes | The number of times the student has liked the video |
| Segment | The number of times a student has played a specific portion of the video by using the slider |

2.6. Data Pre-Processing

The pre-processing .csv file contained the consolidated data, and out of that 24 attributes, the following were selected for this study: “ModuleCode”, “ModuleTitle”, “SessionName”, “ApplicantName”, “CGPA”, “AttemptCount”, “RemoteStudent”, “Probation”, “HighRisk”, “TermExceeded”, “AtRisk”, “AtRiskSSC”, “OtherModules”, “PlagiarismHistory”, “CW1”, “CW2”, “ESE”, “Online C”, “Online O”, “Played”, “Paused”, “Likes”, “Segment” and “Result” (mapped with the outcome of the student either having passed or failed the module) based on the grading scheme, as shown in Table 6. Eight attributes were converted from numeric to ordinal values, as shown in Table 7, and three attributes were converted from different numeric to ordinal values, as shown in Table 8. The criteria used to convert the grading scheme and marks to ordinal were in line with the assessment evaluation classification range used at MEC. This conversion was carried out to map the outcome of the target variable “Result”.

Table 6. Grading scheme.

| Grade | Grade Point | Marks |
|-------|-------------|--------|
| A | 4.00 | 91–100 |
| A– | 3.75 | 87–90 |
| B+ | 3.50 | 84–86 |
| B | 3.25 | 80–83 |
| B– | 3.00 | 77–79 |
| C+ | 2.75 | 74–76 |
| C | 2.50 | 70–73 |
| C– | 2.25 | 66–69 |
| D+ | 2.00 | 60–65 |
| D | 1.75 | 50–59 |
| F | 0.00 | <50 |

Table 7. Numeric to ordinal conversion 1.

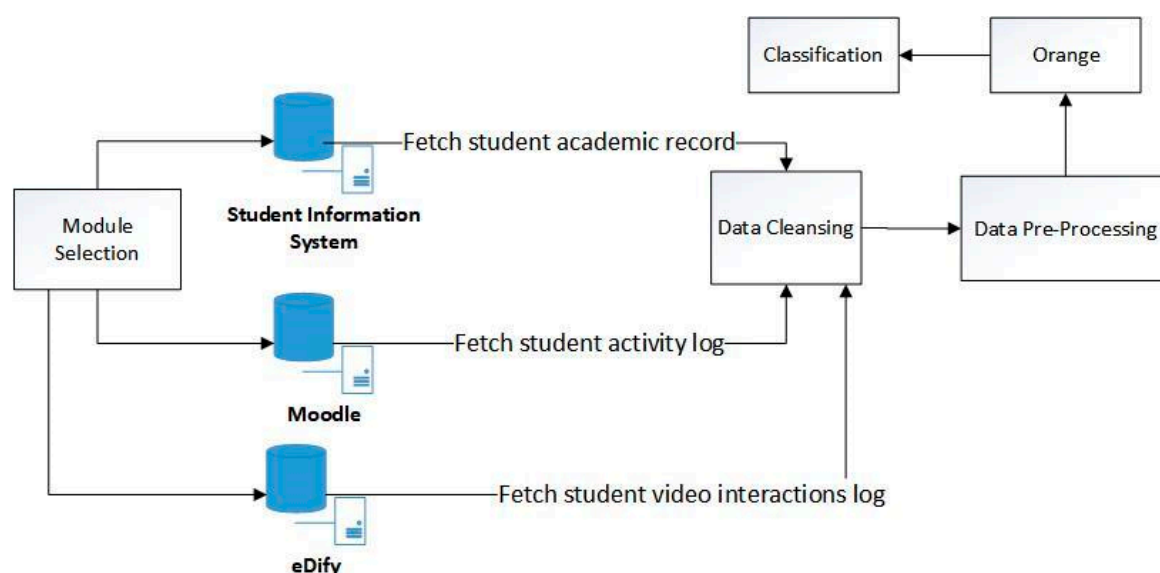
| Nominal Value | CGPA | CW1 | CW2 | ESE | Online Activity |
|---------------|-------|-----|-----|-----|-----------------|
| Excellent | ≥3.75 | ≥87 | ≥87 | ≥87 | ≥300 |
| Very good | ≥3.25 | ≥80 | ≥80 | ≥80 | ≥200 |
| Good | ≥2.75 | ≥74 | ≥74 | ≥74 | ≥100 |
| Fair | ≥2.25 | ≥66 | ≥66 | ≥66 | ≥50 |
| Adequate | ≥1.75 | ≥50 | ≥50 | ≥50 | ≥30 |
| Poor/fail | <1.75 | <50 | <50 | <50 | <30 |

Table 8. Numeric to ordinal conversion 2.

| Nominal Value | Attempt Count | Other Modules | Plagiarism History |
|---------------|---------------|---------------|--------------------|
| High | 3 | 3 | 3 |
| Medium | 2 | 2 | 2 |
| Low | ≤ 1 | ≤ 1 | ≤ 1 |

3. Methods

Before starting the data collection, the first step was to identify the modules. The data were extracted from SIS, Moodle, and eDify. Figure 2 shows the design, materials, and methods used in the process. The raw data were collected from SIS in two phases. First, “Know My Student” details were extracted from the chosen modules. Second, the results of those specific students in the particular modules were extracted. The logfiles from Moodle and eDify were extracted from the selected modules for data cleansing. After the data cleansing, the files were merged, and pre-processing was carried out by merging them into a single consolidated .csv file. Figure 2 shows the procedure on how the data were collected, processed, and made available for the datamining tool (Orange) to predict the students’ academic performance using SIS, Moodle activity data, and video interactions through the mobile application.

**Figure 2.** Data acquisition and processing.

3.1. Module Selection

The sixth semester modules from the computing department at MEC were selected based on the difficulty level (level 2 and 3 modules). Through sampling, it was found that 188 samples were sufficient for this study. Data were collected only from the Spring 2017 to Spring 2021 semesters in the respected modules. The next step was to obtain informed consent from the students who were enrolled on the module; module leaders/instructors obtained this consent, as the study posed no potential risk or discomfort for the students. Ensuring confidentiality and privacy, the identity of the students was coded and mapped accordingly in the data cleansing process. The character marking and generalization method was used to anonymize the data where necessary and applicable.

3.2. Data Cleansing

- For data extracted from SIS, the data were complete and had no missing values. From the first extraction out of 20 attributes, few were not relevant to the study. “Roll-

Number", "ApplicantName" and "ApplicantMobile" were encoded to "xxxxxxxxxx". The "Advisor" attribute were also encoded to "xxxxx". From the second set, only "RollNumber" and "ApplicantName" were also encoded as the first set to make the data anonymous.

- For the data extracted from Moodle, the faculty and moderator logs were filtered out, as they were not required for the study. After the removal of the entries, "User Full Name" and "Affected user" were encoded to "-" instead.
- For the data extracted from eDify, "RollNumber" and "ApplicantName" were encoded to "xxxxxxxxxx".

3.3. Data Pre-Processing

The pre-processing .csv files contained all of the data in a single file that could be pre-processed before being used for classification in any datamining tool. The step carried out here was merging all data into one single data file, where we identified 24 attributes that were useful for this study. The next step was to convert the ordinal values to nominal values, as shown in Tables 7 and 8.

For the data extracted from Moodle, "Affected user", "Event context", "Component", "Event name", "Description" and "Origin" were not relevant to this study, so they were omitted and only the time spent by the individual user in Moodle courses was converted into minutes. IP addresses were used to identify the login timings, either connected from within or outside the campus. 192.168.x.x IP was considered as the within the campus IP address, and the rest were IP addresses from outside the campus. The access time was then converted into minutes to understand the time spent on the activities within or outside the campus.

For data extracted from eDify, all four attributes were taken and no conversion was performed on the data.

3.4. Final Dataset

The final .csv dataset was the complete dataset, with 21 out of 40 attributes that could be used for this study. This dataset can be used with any datamining tool for classifying and predicting student academic performance using EDM.

From SIS, 15 out of the 24 attributes were selected for the final dataset: "Applicant-Name", "CGPA", "AttemptCount", "RemoteStudent", "Probation", "HighRisk", "TermExceeded", "AtRisk", "AtRiskSSC", "OtherModules", "PlagiarismHistory", "CW1", "CW2", "ESE" and "Result (Target Variable)".

From Moodle, two attributes were selected based on the activities performed on Moodle from outside or within the campus: "Online C" and "Online O".

From eDify, four attributes were selected: "Played", "Paused", "Likes" and "Segment".

The final dataset can help researchers to better understand the learning behaviors of the students in the online learning environment setting.

4. Conclusions

This article provides the dataset with multiple learning environments, which will be useful for researchers who want to explore students' academic performance in online learning environments. This will help them to model their educational datamining models. The dataset will be useful for researchers who want to conduct comparative studies on student behaviors and patterns related to online learning environments. It will further help to form an educational datamining model that can be applied to different classification algorithms to predict successful students. Moreover, feature selection techniques can be applied, which can provide a better accuracy rate for predicting students' academic performance.

For future studies, weekly video interaction records can be considered to provide better insights into video learning analytics and student performance. Furthermore, the data can be used with the predictive churn model to act as an early warning system for the dropouts in the course.

5. Patents

Hasan, Raza, Palaniappan, Sellappan, Mahmood, Salman, and Asif Hussain, Shaik. A novel method and system to enhance teaching and learning and the student evaluation process using the “eDify” mobile application. AU Patent Innovation 2021103523, filed 22 June 2021.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/data6110110/s1>, Data S1: csv files.

Author Contributions: Conceptualization and methodology, R.H.; supervision, S.P.; data curation and validation, S.M.; writing—original draft preparation and visualization, A.A.; investigation and writing—review and editing, K.U.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The authors confirm that the data supporting the findings of this study are available within the article and/or its Supplementary Materials.

Acknowledgments: The authors of this data article are extremely thankful to all of the faculty and students who participated in this study.

Conflicts of Interest: The authors declare that they have no known competing financial interest or personal relationship that have, or could be perceived to have, influenced the work reported in this article.

References

- Romero, C.; Ventura, S.; de Bra, P. De knowledge discovery with genetic programming for providing feedback to courseware authors. *User Model. User-Adapt. Interact.* **2004**, *14*, 425–464. [\[CrossRef\]](#)
- Yaacob, W.F.W.; Nasir, S.A.M.; Yaacob, W.F.W.; Sobri, N.M. Supervised data mining approach for predicting student performance. *Indones. J. Electr. Eng. Comput. Sci.* **2019**, *16*, 1584–1592. [\[CrossRef\]](#)
- Shetty, I.D.; Shetty, D.; Roundhal, S. Student performance prediction. *Int. J. Comput. Appl. Technol. Res.* **2019**, *8*, 157–160. [\[CrossRef\]](#)
- Zohair, L.M.A. Prediction of Student’s performance by modelling small dataset size. *Int. J. Educ. Technol. High. Educ.* **2019**, *16*, 27. [\[CrossRef\]](#)
- Daud, A.; Aljohani, N.R.; Abbasi, R.A.; Lytras, M.D.; Abbas, F.; Alowibdi, J.S. Predicting student performance using advanced learning analytics. In Proceedings of the 26th International Conference on World Wide Web Companion, Perth, Australia, 3–7 April 2017; pp. 415–421.
- Tomasevic, N.; Gvozdenovic, N.; Vranes, S. An overview and comparison of supervised data mining techniques for student exam performance prediction. *Comput. Educ.* **2020**, *143*, 103676. [\[CrossRef\]](#)
- Saa, A.A.; Al-Emran, M.; Shaalan, K. Mining student information system records to predict students’ academic performance. In *Advances in Intelligent Systems and Computing*; Springer: Cham, Switzerland, 2020; pp. 229–239. ISBN 9783030141172.
- Giannakos, M.N.; Chorianopoulos, K.; Chrisochoides, N. Making sense of video analytics: Lessons learned from clickstream interactions, attitudes, and learning outcome in a video-assisted course. *Int. Rev. Res. Open Distrib. Learn.* **2015**, *16*, 260–283. [\[CrossRef\]](#)
- Lau, K.H.V.; Farooque, P.; Leydon, G.; Schwartz, M.L.; Sadler, R.M.; Moeller, J.J. Using learning analytics to evaluate a video-based lecture series. *Med. Teach.* **2018**, *40*, 91–98. [\[CrossRef\]](#) [\[PubMed\]](#)
- Mohammed, Q.A.; Naidu, V.R.; Said, M.; Al Harthi, M.S.A.; Babiker, S.; Al Balushi, Q.; Al Rawahi, M.Y.; Al Riyami, N.H.S. Role of online collaborative platform in higher education context. *IJAEDU-Int. E-J. Adv. Educ.* **2020**, *6*, 220–227. [\[CrossRef\]](#)
- Hasnine, M.N.; Akcapinar, G.; Flanagan, B.; Majumdar, R.; Mouri, K.; Ogata, H. Towards final scores prediction over clickstream using machine learning methods. In Proceedings of the ICCE 2018—26th International Conference on Computers in Education, Metro Manila, Philippines, 26–30 November 2018.
- Yang, T.-Y.; Brinton, C.G.; Joe-Wong, C.; Chiang, M. Behavior-based grade prediction for MOOCs via time series neural networks. *IEEE J. Sel. Top. Signal Process.* **2017**, *11*, 716–728. [\[CrossRef\]](#)
- Hasan, R.; Palaniappan, S.; Raziff, A.R.A.; Mahmood, S.; Sarker, K.U. Student academic performance prediction by using decision tree algorithm. In Proceedings of the 2018 4th International Conference on Computer and Information Sciences (ICCOINS), Kuala Lumpur, Malaysia, 13–14 August 2018; pp. 1–5.

14. Hasan, R.; Palaniappan, S.; Mahmood, S.; Sarker, K.U.; Abbas, A. Modelling and predicting student's academic performance using classification data mining techniques. *Int. J. Bus. Inf. Syst.* **2020**, *34*, 403–422. [[CrossRef](#)]
15. Hasan, R.; Palaniappan, S.; Mahmood, S.; Shah, B.; Abbas, A.; Sarker, K.U. Enhancing the teaching and learning process using video streaming servers and forecasting techniques. *Sustainability* **2019**, *11*, 2049. [[CrossRef](#)]
16. Hasan, R.; Palaniappan, S.; Mahmood, S.; Abbas, A.; Sarker, K.U.; Sattar, M.U. Predicting student performance in higher educational institutions using video learning analytics and data mining techniques. *Appl. Sci.* **2020**, *10*, 3894. [[CrossRef](#)]
17. Hasan, R.; Palaniappan, S.; Mahmood, S.; Sarker, K.U.; Sattar, M.U.; Abbas, A.; Naidu, V.R.; Malali Rajegowda, P. eDify: Enhancing teaching and learning process by using video streaming server. *Int. J. Interact. Mob. Technol.* **2021**, *15*, 49–65. [[CrossRef](#)]