

Article

Predicting Student Performance in Higher Educational Institutions Using Video Learning Analytics and Data Mining Techniques

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

- ¹ Department of Information Technology, School of Science and Engineering, Malaysia University of Science and Technology, Petaling Jaya 47810, Selangor, Malaysia; sell@must.edu.my (S.P.); salman.mahmood@pg.must.edu.my (S.M.); mian.usman@phd.must.edu.my (M.U.S.)
- ² Department of Computing, Middle East College, Knowledge Oasis Muscat, P.B. No. 79, Al Rusayl 124, Oman; aabbas@mec.edu.om
- ³ Faculty of Ocean Engineering Technology and Informatics (FTKKI), University Malaysia Terengganu, Kuala Terengganu 21030, Terengganu, Malaysia; ku_sarker@yahoo.com
- * Correspondence: raza.hasan@pg.must.edu.my; Tel.: +968-98199513

Received: 4 May 2020; Accepted: 3 June 2020; Published: 4 June 2020



Abstract: Technology and innovation empower higher educational institutions (HEI) to use different types of learning systems—video learning is one such system. Analyzing the footprints left behind from these online interactions is useful for understanding the effectiveness of this kind of learning. Video-based learning with flipped teaching can help improve student's academic performance. This study was carried out with 772 examples of students registered in e-commerce and e-commerce technologies modules at an HEI. The study aimed to predict student's overall performance at the end of the semester using video learning analytics and data mining techniques. Data from the student information system, learning management system and mobile applications were analyzed using eight different classification algorithms. Furthermore, data transformation and preprocessing techniques were carried out to reduce the features. Moreover, genetic search and principle component analysis were carried out to further reduce the features. Additionally, the CN2 Rule Inducer and multivariate projection can be used to assist faculty in interpreting the rules to gain insights into student interactions. The results showed that Random Forest accurately predicted successful students at the end of the class with an accuracy of 88.3% with an equal width and information gain ratio.

Keywords: classification algorithms; data preprocessing; data mining; data transformation; student academic performance; video learning analytics

1. Introduction

Digitalization has infiltrated into every aspect of life. Emerging new technologies have an impact on our lives and change the way we do our daily work, raising our performance to a new height. The technology used in education has allowed educators to implement new theories to enhance the teaching and learning process. This shift from “traditional learning” has led to a model in which learning can take place outside the classroom, facilitating different learner attributes such as visual, verbal, aural and solitary learning within the “blended learning” approach [1]. Educators use innovative technologies to cater to different learners with blended learning, and learners can use these technologies to improve their cognitive abilities to excel in the courses being taught [2]. Educators use virtual classrooms, webinars, links, simulations or any other online mechanism to deliver the information [3].

In blended learning environments (BLE), the approach to education is a flipped classroom, where content is provided through the Internet in advance before the commencement of the class [4]. The flipped classroom enables students to come to the class prepared, and the educator uses activities in the class that can clear the doubts of the students, as well as helping them in their assessments and the concepts learned from the content provided to them in the form of discussions or activities. The content that is usually provided by the educator for the flipped classroom includes reading along with questions and answers, created video lectures, demonstration videos, an online class discussion room, lecture slides, tutorials and reading from textbooks or reference books [5].

The adoption of flipped classrooms by higher education institutions (HEI) has rapidly increased in recent years [6]. The crucial factor in flipped classrooms is the electronic support that the HEI uses to disseminate knowledge among the learners. Encouraging learners to learn at their own pace, including through video lectures, frees up classroom time for more applied learning or active learning. Educators take support from various educational technologies to replicate the virtual classroom.

At Middle East College (MEC), different systems are in place to store student information. The student information system (SIS) stores the related academic data; the learning management system (LMS)—i.e., Moodle—is used as an e-learning tool to disseminate knowledge at the same time as online behavior is stored in the logs for each student; and the video streaming server (VSS) is used alongside Moodle to share video lectures to the students [7]. As mobile culture is on the rise, an in-house built mobile application—eDify—has been developed to share video lectures to students since spring 2019, enhancing the teaching and learning process. With the use of eDify, the collection of students' video interactions was made easy, providing valuable information about the learners. The data which these systems hold about the learners can be useful for enhancing the teaching and learning process. Learning analytics (LA) help HEIs to gain useful insights about their learners interacting with the system [8–11]. When videos are used to provide useful information about the learners, it is called video learning analytics (VLA) [10,12–14].

The study aimed to use systems such as SIS, LMS and eDify to explore students' overall performance at the end of the semester. Education data mining is used to elucidate this issue as this uses predictive analysis on the data [8,15,16]. The majority of the research suggests that the classification model can be used to predict success in HEIs using data related to student profiles and data logs from Moodle [17–19]. Similarly, research has been carried out to predict student performance using videos, but little work has been done on using both systems to predict the students' academic performance. The study attempts to predict student performance in HEIs using video learning analytics and data mining techniques to help faculty, learners and the management to make better decisions.

The contribution of this paper is threefold: (1) when the classification model is formed using interaction data from different systems used in the learning setting, we determine which algorithm and preprocessing techniques are best for predicting successful students; (2) we determine the impacts that feature selection techniques have on classification performance; and (3) we determine which features have greater importance in the prediction of student performance.

This paper is organized as follows: Section 2 presents a literature review and an overview of related works in this field. Section 3 presents the methodology used in the study. Section 4 presents the results. Section 5 presents the discussion on the study. Finally, Section 6 draws a conclusion and proposes future research.

2. Literature Review and Related Works

2.1. Learning Analytics

LA is defined as the usage of data, statistical analysis, and explanatory and predictive models to gain insight and act on complex issues. LA involves the data analysis of the learners and their activities to enhance the student learning experience [20]. Implementing LA allows higher education institutions to understand their learners and the barriers to their learning, thus ensuring institutional success and

retaining a larger and more diverse student population, which is important for operational facilities, fundraising and admissions. Student success is a key factor in improving academic institutions' resource management.

LA is the measurement, collection, analysis and reporting of data about learners. Understanding context is important for the purposes of optimization and learning and for the environments in which learning occurs [21], as shown in Figure 1.



Figure 1. Learning analytics cycle.

Several studies have been conducted on LA to understand the students' learning behaviors and to optimize the learning process. Thus, HEIs will be able to identify the learners' behavior and patterns to detect low achievers or students who are at-risk. Early warning systems are investigated in order to achieve these objectives. Predictive modelling is usually used to predict the learners' end-of-term academic performance, with the data from different online learning systems being evaluated [22]. Purdue university uses course signals which allow faculty to provide real-time feedback to the students. Grades, demographic data, learning management system (LMS) interaction data, grade history and students' effort are measured with the help of LA. A personalized email is used to communicate with students about their current status with the help of traffic lights. The system helps to retain information, and performance outcomes are thus evaluated [23]. Predictive analysis using PredictED emails students about their behavior to predict their end-of-semester grades with the help of LMS access logs [24]. An LA study showed that students' online activities correlate with their performance in the course, and a prediction can be made regarding the possible outcome of their performance at the end of the course [25]. The study suggests that LA plays a vital role in affecting students' perspectives and the way they learn in the different online settings [26]. LA not only provides the educator with insights on the outcomes of the course, but also provides an opportunity for self-evaluation for the students [9,27].

2.2. Video Learning Analytics

VLA enables us to understand and improve the effectiveness of video-based learning (VBL) as a tool and its related practices [28]. Flipped teaching is an important component of VBL, where an educator uploads a video lecture. The evaluation of this type of education can be enhanced through LA and can predict the students' overall performance in the course. With the rapidly changing environments in teaching and learning, different technologies in VLA can help stakeholders, educators and learners to understand the data generated by these videos. Different research has been carried out to understand this relationship. Regarding the use of YouTube to upload lectures, quantitative data were analyzed by using the trends of the video interactions [13]. The researcher used student interactions in the analyzed video to understand the students' behavior and tried to predict the outcomes of their final scores with the help of machine learning. These interactions within the video are called the "Clickstream" [28,29]. Researchers use different data mining algorithms to predict student performance or for grade prediction [12].

2.3. Educational Data Mining

Data mining (DM) refers to the discovery of associative patterns from large datasets. It is a powerful tool in artificial intelligence (AI) and facilitates the categorization of data into different dimensions, identifying relationships and categorizing the information. This allows stakeholders to use the information extracted from DM and can help improve decision-making. Educational data

mining (EDM) is a growing discipline which is used to discover meaningful and useful knowledge from data extracted from educational settings. Applying DM techniques, EDM provides researchers with a better understanding of student behavior and the settings in which learning happens, as shown in Figure 2 [18,30].

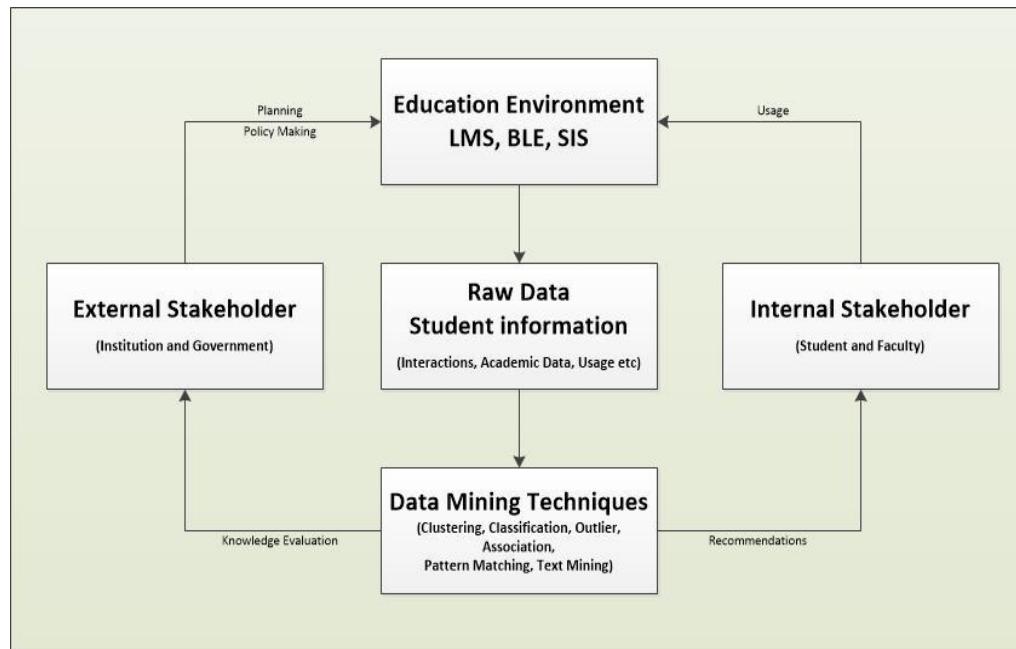


Figure 2. The cycle of applying data mining in education [31].

The EDM process converts raw data from different systems used in HEIs into useful information that has a potential impact on educational practice and research [32]. It is also referred to as Knowledge Discovery in Databases (KDD) due to the hierarchical nature of the data used [18,30]. The methods used for KDD or EDM are as follows.

2.3.1. Prediction

This is a widely used technique in EDM. Here, historical data are used in terms of student grades, demographic data, etc., to predict the future outcomes of the students' results. Many studies have been carried out in which classification is most the commonly used method to make predictions. In order to achieve their research targets, researchers have used the Decision Tree, Random Forest, Naïve Bayes, Support Vector Machines, Linear Regression or Logistic Regression models, and K means approaches [33–38]. The studies mainly suggest the prediction of students' academic performance either before the classes, at the middle of the session or at the end of the term.

2.3.2. Clustering

This is a process of finding and grouping a set of objects, called a cluster, in the same group based on similar traits. This technique is used in EDM, where student participation in online forums, discussion groups or chats is studied. Studies show that classification should be used after clustering. Researchers have used similar algorithms as those for prediction to determine students' academic success [39–42].

2.3.3. Relationship

This method is used to look for similar patterns from multiple tables as compared to other methods. The methods commonly used to investigate relationships are association, correlation, and sequential

patterns. The study suggests that, in EDM, an association rule should be used to predict students' end-of-semester exam results and performance using heuristic algorithms [43,44].

2.3.4. Distillation

This recognizes and classifies features of the data, depicted in the form of a visualization for human inference [18,36,45].

2.4. State-Of-The-Art

EDM aims to predict student performance either using demographic data or socio-economic data or online activity on the learning management system at different educational levels. A problem arises when the data are imbalanced or where researchers have used other techniques alongside a simple classification method such as data balancing, cost-sensitive learning and genetic programming [8,46]. Early dropout studies have been conducted by researchers to determine the factors that can influence student retention. Here, classical classification methods were also used to predict early dropout. Feature selection plays a vital role in the accuracy of the classification algorithms. The genetic algorithm was used and obtained 10 attributes out of 27, and the efficiency was improved [47,48]. For our study into handling imbalanced data, we use the K-fold cross-validation technique; to reduce features, we use ranking and scoring, the genetic search algorithm, principle component analysis and the multiview learning approach using multivariate projection along with the CN2 Rule induction for easier interpretation, as the multiview learning approach uses sparse data where a high percentage of the variable's cells do not contain actual data.

HEIs use different learning systems and the amount of data accumulated from these systems is enormous. The LA provides the analysis of data from these systems and finds meaningful patterns which can be helpful for educators and learners. From the literature review and related works, it is evident that, when analyzing student academic performance prediction, LA is a widely used method in EDM. The classification technique is also used to classify the parameters used for the prediction and success of the model. Due to innovative teaching and learning pedagogies and the implementation of flipped classrooms, the use of VBL is on the increase, where students can study prior to the class at their own leisure and come prepared to the class. Little research has been carried out on VLA and on determining students' attitude and behavior towards VBL. Educators apply different settings within or outside the classroom to create an effective learning experience for learners. For effective knowledge transfer, an educator needs a predictive model to understand the future outcomes for each student. This will help the educator to identify the methods which should be applied, to identify poorly performing students and provide better support to the students to obtain good grades at an early stage.

3. Methodology

The study is exploratory in nature. The quantitative prediction method is used for the study.

3.1. Educational Datamining Model

According to the literature review in the previous section, some of the important activities are recognized in educational datamining, as shown in Figure 3. The classification method is used to predict student academic performance as a widely used method in prediction, as shown in the literature review. For this study, we also use the classification method, and the data are gathered from multiple systems which are already running in the HEI: student academic information is gathered from the student information system (SIS), student online activity from LMS Moodle and student video interactions data from eDify (mobile application). The algorithms used for this study are based on the frequently used algorithms from the existing literature: Classification Tree, Random Forest, k-Nearest Neighbors (kNN), Support Vector Machine (SVM), Logistic Regression, Naïve Bayes, Neural Network and CN2 Rule Induction.

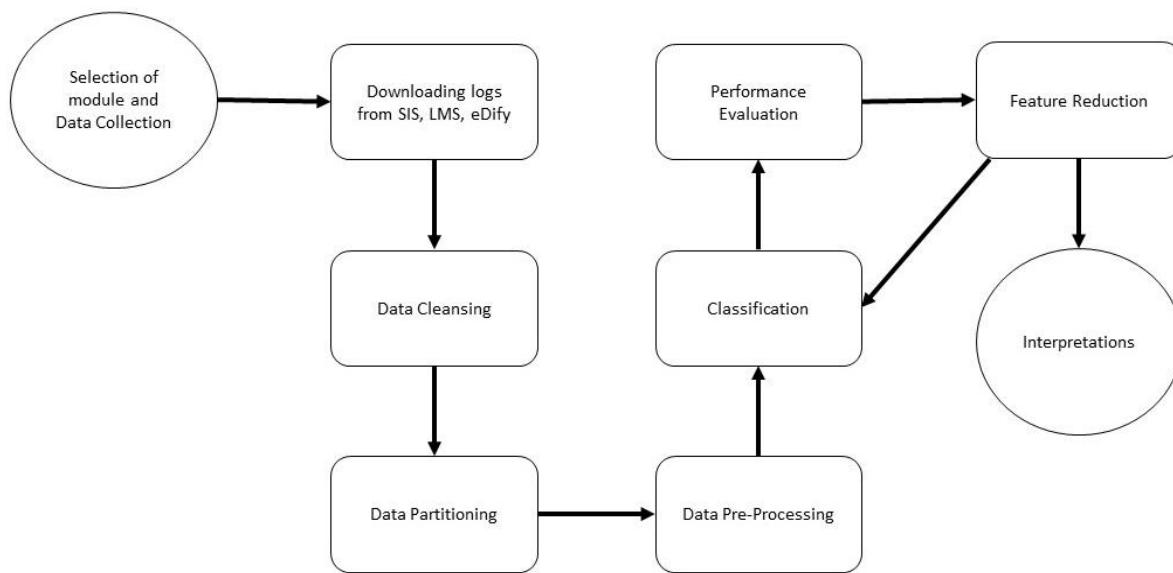


Figure 3. The cycle of applying educational data mining in research.

3.2. Module Selection

The study was conducted at a private HEI in Oman with a dataset consisting of 772 instances of students registered in the sixth semester. The modules chosen were e-commerce (COMP 1008) and e-commerce technologies (COMP 0382). The reason for choosing these modules was that the modules are offered in different specializations, sharing the same course content and yielding the maximum number of students necessary for making the dataset.

3.3. Data Collection

Before starting data collection for any research, ethical approval is necessary. For this study, informed consent was obtained from the applicants, explaining the purpose of study and the description. There were “no potential risks” or discomfort while using eDify, but if the applicant, for any reason, felt discomfort or risk, they were able to withdraw from the study at any time. To ensure confidentiality and privacy, all the information from the study was coded to protect applicant names. No names or other identifying information were used when discussing or reporting data. Once the data was fully analyzed, it was destroyed. This research was voluntary and the applicant retained the right to withdraw from participation at any time; this was also communicated to the students when the semester started.

The data were collected from the Spring 2019 and Fall 2019 semesters after the implementation of eDify (mobile application) supporting VBL and by capturing students’ video interactions. Eight lecture videos were used in the module, but for study purposes, the data from all lecture videos were used.

3.4. Data Cleansing

In this activity, unnecessary data were cleansed and data were separated from information which was not relevant for the analysis. Historical data for each student registered in the two modules were considered for the study. In total, 19 features were selected for study, from which 12 features and one meta attribute was used from SIS; Moodle yielded two features related to the online activity within and outside the campus; and four features were selected from eDify.

3.5. Data Partitioning

After the data cleansing activity, data partitioning was performed. Relevant data were extracted and combined for further analysis, as shown in Table 1.

Table 1. List of extracted features and description. GPA: grade point average.

No	Category	Feature	Values	Description
1	Student Academic Information	Applicant	Text + Numeric	Student ID for identification and mapping
2		Cumulative Grade Point Average (CGPA)	* 0.00–4.00	Student overall GPA
3		Attempt Count	Low (1), Medium (2) and High (>2)	Status of the student attempts in the module
4		High Risk	Yes/No	A student having a high failure rate in the same module
5		Term Exceed at Risk	Yes/No	Shows the students' progression in the degree plan
6		At Risk	Yes/No	A student failed two or more modules previously
7		Student Success Centre (SSC)	Yes/No	Student referred to SSC for any assistance
8		Other Modules	Low (1), Medium (2) and High (>2)	Student register for other modules
9		Plagiarism Count	* Numeric Value	Students have been accused of any plagiarism
10		Coursework 1 (CW1)	* 1–100	Marks obtained in CW1
11		Coursework 2 (CW2)	* 1–100	Marks obtained in CW2
12		End Semester Examination (ESE)	* 1–100	Marks obtained in ESE
13	Students Activity	Activity on Campus	* Time spent in minutes	Student Moodle activity on campus
14		Activity Off Campus	* Time spent in minutes	Student Moodle activity off campus
15	Student Video Interactions	Played	* No. of times video played	eDify access and played video
16		Paused	* No. of times video paused	eDify access and paused the video
17		Likes/Dislikes	Yes/No	eDify access and either liked or disliked video
18		Segment	* No. of segments rewind	eDify access and most-watched segment mapped with ESE
19	Result (Target Variable)		* Numeric value	Overall marks obtained in the module

3.6. Data Pre-Processing

The main reason for testing several algorithms on the dataset was that their performance varies for the selected features. The study suggested that algorithms behave differently; depending on the dataset, the efficiency and performance may also vary. With this approach, it is easier to identify any one algorithm which suits the dataset with better accuracy and performance. For this purpose, a similar approach was used for this study. For the study, the Orange data mining tool was used in accordance with the process shown in Figure 4 [49].

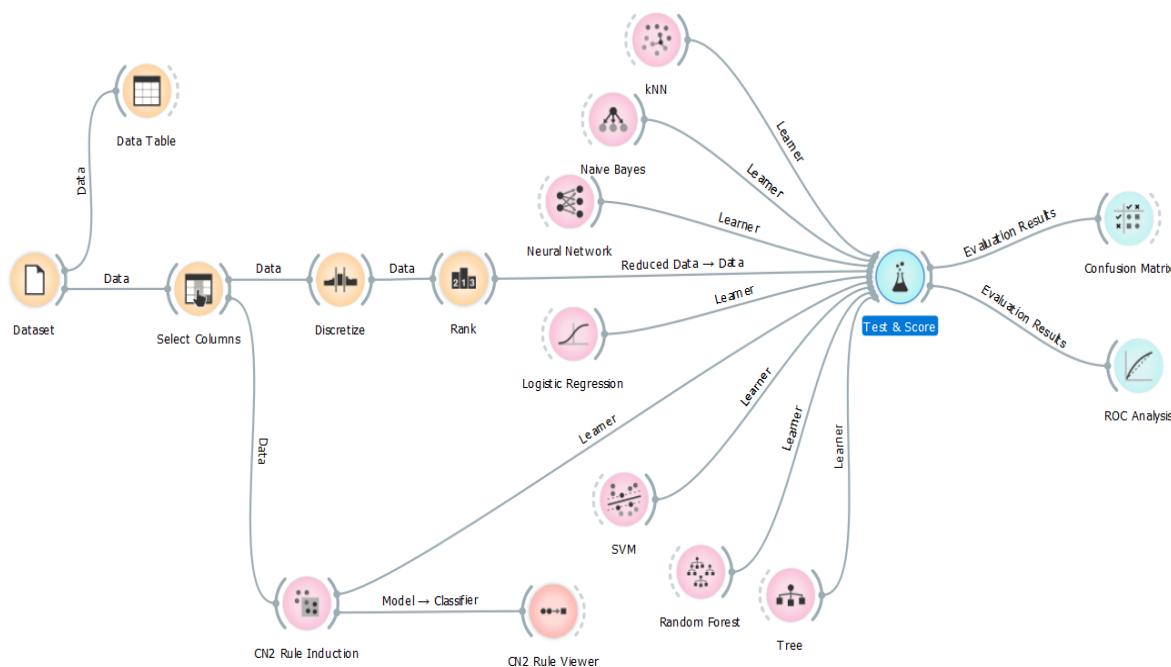


Figure 4. Data mining process.

Pre-processing methods were used to transform raw data collected from these systems into an understandable form. Table 1 shows the features selected from SIS, where CGPA is converted into nominal order as excellent, very good, good, fair, adequate, or poor/fail; plagiarism count into low, medium, or high; coursework (CW)1 into pass or fail; CW2 into pass or fail; and end-of-semester evaluation (ESE) into pass or fail. Online activity was captured from Moodle in terms of minutes and is converted into nominal order as follows: activity on campus—low, medium, or high; and activity off campus—low, medium, or high. From eDify, four features—played, paused, likes/dislikes and segment—were selected, and discretization was applied on the continuous data to transform them into categorical data.

The dataset was fed into the analyzer, and the select column widget was used to select the features from the available variables. Student ID is the meta attribute, and the result was selected as the target variable. The discretize widget was used to transform the variable into categorical data in order to be used in the study. The rank widget scoring technique was used for feature selection, which could be used further for prediction. Information gain, gain ratio, and the Gini decrease weight were compared for the performance of the prediction model. Furthermore, the investigation of the feature selection proceeded using the genetic algorithm, and principle component analysis (PCA) was used to further reduce the features. Multivariate projection was used as an optimization method that finds a linear projection and associated scatterplot that best separates instances of separate classes, uncovering feature interactions and providing information about intra-class similarities.

3.7. Performance Evaluation

The performance of the classification algorithms was determined in the study; performance was based on the four standard evaluation metrics for accuracy, sensitivity, specificity and f-measure. A 10-fold cross-validation for comparison with the baseline method was used, splitting the data into 10 folds and using nine folds for training and one fold for testing. Confusion metrics were used for the analysis of supervised learning, where each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class to avoid mislabeling as shown in Table 2.

Table 2. Sample confusion matrix.

		Predictions	
		Failed	Passed
Actual	Failed	a	b
	Passed	c	d

The entries in the confusion matrix have the following meaning in the context of a data mining problem: a is the correct negative prediction, also called true negative (TN), classified as failed by the model; b is the incorrect positive prediction, also called false positive (FP), classified as passed by the model; c is the incorrect negative prediction, also called false negative (FN), classified as failed by the model; and d is the correct positive prediction, also called true positive (TP), classified as passed by the model.

The performance metrics according to this confusion matrix are calculated as follows.

3.7.1. Accuracy

The accuracy (AC) is the proportion of the total number of predictions that were correct. It is determined using the following equation:

$$AC = (d + a)/(d + a + b + c), \quad (1)$$

3.7.2. Sensitivity

The recall or TP rate is the proportion of positive cases that were correctly identified, as calculated using the following equation:

$$\text{Sensitivity} = d/(d + c), \quad (2)$$

3.7.3. Specificity

The TN rate is the proportion of negatives cases that were correctly classified as negative, as calculated using the following equation:

$$\text{Specificity} = a/(a + b), \quad (3)$$

3.7.4. F-Measure

The confusion matrix belongs to a binary classification, returning a value of either “passed” or “failed”. The sensitivity and specificity measures may lead to biased comments in the evaluation of the model, as calculated using the following equation:

$$\text{F-measure} = 2d/(2d + 2b + c), \quad (4)$$

4. Results

A supervised data classification technique was used to determine the best prediction model that fit the requirements for giving an optimal result. For analysis, the same set of classification algorithms, performance metrics and the 10-fold cross-validation method were used.

4.1. No Feature Selection and Transformation

The CN2 Rule Inducer and Random Forest algorithm exhibited a performance rate of 85.1%, as shown in Table 3. Further investigating the performance metrics of the CN2 Rule Inducer showed good sensitivity, levelling-out the lower specificity and giving a good F-measure score. When no data transformation and no feature selection were undertaken, the CN2 rule inducer showed the highest classification accuracy.

Table 3. Data transformation (none), feature selection (none).

Algorithm	Accuracy	Sensitivity	Specificity	F-Measure
CN2 Rule Inducer	0.851	0.960	0.870	0.913
Random Forest	0.851	0.952	0.875	0.912
Log. Reg.	0.847	0.992	0.846	0.913
Tree	0.843	0.939	0.876	0.907
kNN	0.841	0.939	0.874	0.905
Naïve Bayes	0.821	0.957	0.844	0.897
Neural Network	0.811	1.00	0.811	0.896
SVM	0.785	0.927	0.829	0.875

4.2. No Feature Selection and Equal Frequency Transformation

When data transformation was applied with equal frequency, the Random Forest algorithm predicted with an accuracy of 85%, as shown in Table 4. The predicted scores were similar to the previous results when no data transformation and no feature selection were applied, as shown in Table 3.

Table 4. Data transformation (equal frequency), feature selection (none).

Algorithm	Accuracy	Sensitivity	Specificity	F-Measure
Random Forest	0.850	0.911	0.873	0.911
Log. Reg.	0.846	0.992	0.845	0.913
kNN	0.837	0.939	0.870	0.903
Tree	0.837	0.936	0.872	0.903
CN2 Rule Inducer	0.837	0.938	0.871	0.903
SVM	0.820	0.960	0.841	0.896
Neural Network	0.811	1.00	0.811	0.896
Naïve Bayes	0.804	0.931	0.844	0.885

4.3. No Feature Selection and Equal Width Transformation

When data transformation was applied with equal width, the Random Forest predicted with an accuracy of 85.5%, as shown in Table 5. From the analysis, it was found that the equal width data transformation technique can be further investigated with feature selection, as the result showed slight improvements.

Table 5. Data transformation (equal width), feature selection (none).

Algorithm	Accuracy	Sensitivity	Specificity	F-Measure
Random Forest	0.855	0.973	0.865	0.916
Log. Reg.	0.846	0.990	0.846	0.912
kNN	0.842	0.971	0.854	0.909
Tree	0.842	0.955	0.864	0.907
CN2 Rule Inducer	0.842	0.957	0.863	0.908
SVM	0.820	0.958	0.842	0.896
Naïve Bayes	0.811	0.935	0.848	0.889
Neural Network	0.811	1.00	0.811	0.896

4.4. Information Gain Feature Selection and Equal Width Transformation

The tree-based algorithm showed improvements in accuracy when a data transformation of equal width was applied with the ranking technique, utilizing the scoring methods provided by the Orange data mining tool. The accuracy of Random Forest increased from 85.5% to 87.6%, the CN2 Rule Inducer increased from 85.1% to 87.3%, and Classification Tree increased from 84.3% to 87.2% when the information gain feature was selected. The nine selected features were found in all the three scoring methods—CW1, ESE, CW2, likes, paused, played, segment, Moodle on campus and Moodle off campus. SVM performed worse compared to others, but the accuracy of SVM was slightly improved from 82% to 82.5%, as shown in Table 6.

Table 6. Data transformation (equal width), feature selection (information gain).

Algorithm	Accuracy	Sensitivity	Specificity	F-Measure
Random Forest	0.876	0.981	0.883	0.930
CN2 Rule Inducer	0.873	0.978	0.883	0.928
Tree	0.872	0.978	0.881	0.927
Log. Reg.	0.870	0.992	0.871	0.928
kNN	0.864	0.961	0.886	0.922
Naïve Bayes	0.835	0.935	0.876	0.905
Neural Network	0.835	1.00	0.835	0.910
SVM	0.825	0.949	0.857	0.901

4.5. Information Gain Ratio Feature Selection and Equal Width Transformation

Random Forest's accuracy was improved by 0.7% with better sensitivity and F-measure when the equal width data transformation technique and information gain ratio technique were applied. The CN2 Rule Inducer improved by 0.1% with less sensitivity and better specificity. Interestingly, Classification Tree, Logistic Regression, Naïve Bayes, Neural Network and SVM showed no sign of improvement and remained unchanged. kNN's performance was decreased by 3% from the previous example, as shown in Table 7.

Table 7. Data transformation (equal width), feature selection (information gain ratio).

Algorithm	Accuracy	Sensitivity	Specificity	F-Measure
Random Forest	0.883	0.975	0.895	0.933
CN2 Rule Inducer	0.874	0.964	0.894	0.928
Tree	0.872	0.978	0.881	0.927
Log. Reg.	0.870	0.992	0.871	0.928
Naïve Bayes	0.835	0.935	0.876	0.905
Neural Network	0.835	1.00	0.835	0.910
kNN	0.834	0.922	0.884	0.903
SVM	0.825	0.949	0.857	0.901

4.6. Gini Decrease Feature Selection and Equal Width Transformation

Data transformation with equal width and the feature selection of Gini decrease reduced the overall accuracy of all the selected algorithms in the study, except the Neural Network, as shown in Table 8. Although kNN performed well compared to the other classification algorithms, its accuracy was less than the information gain, as shown in Table 6. Thus, the Gini decrease method was omitted as the accuracy was not enhanced, as shown in Table 8.

Table 8. Data transformation (equal width), feature selection (Gini decrease).

Algorithm	Accuracy	Sensitivity	Specificity	F-Measure
kNN	0.838	0.989	0.844	0.911
Neural Network	0.835	1.00	0.835	0.910
Log. Reg.	0.832	0.995	0.835	0.908
Tree	0.829	0.980	0.842	0.905
CN2 Rule Inducer	0.829	0.980	0.842	0.905
Random Forest	0.828	0.980	0.840	0.905
Naïve Bayes	0.807	0.941	0.845	0.891
SVM	0.773	0.913	0.832	0.871

4.7. Feature Selection Using Genetic Algorithm and Classification

Due to the limitation of the Orange tool to run the genetic algorithm for feature selection, Weka (Waikato Environment for Knowledge Analysis) was used on the same dataset for feature selection using genetic search. Before applying classification algorithms in Orange, relevant features were selected by using feature selection method. Feature selection was performed by using the genetic algorithm, and applicant, at-risk, CW1, CW2, ESE and played were selected. Tree predicted with an accuracy of 87.4%, as shown in Table 9. Feature selection using the genetic search reduced the features to six as compared to the information gain ratio feature selection technique and equal width transformation, but the accuracy was not improved, as shown in Table 7.

Table 9. Feature selection using genetic algorithm and classification.

Algorithm	Accuracy	Sensitivity	Specificity	F-Measure
Tree	0.874	0.997	0.748	0.93
Log. Reg.	0.871	0.992	0.74	0.928
kNN	0.867	0.983	0.717	0.926
SVM	0.867	0.989	0.748	0.926
Naïve Bayes	0.847	0.952	0.685	0.912
Random Forest	0.835	1	1	0.91
CN2 Rule Inducer	0.834	0.922	0.884	0.903
Neural Network	0.828	0.98	0.84	0.905

4.8. Principal Component Analysis

A principle component analysis (PCA) was undertaken to reduce the number of variables from 19 to eight components with a variance of 95.6%, as shown in Figure 5. Table 10 shows the PCA component variances.

PC1 shows that a video being played and paused and the segment correspond to the behavior of a video being watched. PC2 shows that a video being played and the segment correspond to the behavior of a video being watched and rewound. PC3 shows that CGPA indicates high academic achievers. PC4 shows that at risk, ESE and likes correspond to the behavior of a student being at risk; those who scored fewer marks in the ESE will like video content. PC5 shows that plagiarized, likes and Moodle activity outside campus correspond to the behavior of a student who is weak in academic writing but likes the video and watches the video outside the campus. PC6 shows that attempt count,

at risk, plagiarized, CW1, CW2 and Moodle activity in campus correspond to the behavior of a weak student that spends time on Moodle inside the campus to get better grades in the ESE. PC7 shows that not at risk, plagiarized, CW1, ESE and Moodle activity outside the campus correspond to the behavior of weak students who spend time on Moodle activities outside the campus. PC8 shows that not at risk, plagiarized, CW1, CW2, likes, Moodle activity inside the campus, and less Moodle activity outside the campus correspond to the behavior of weak students, with the attributes in common that they spend time on Moodle both inside and outside the campus.

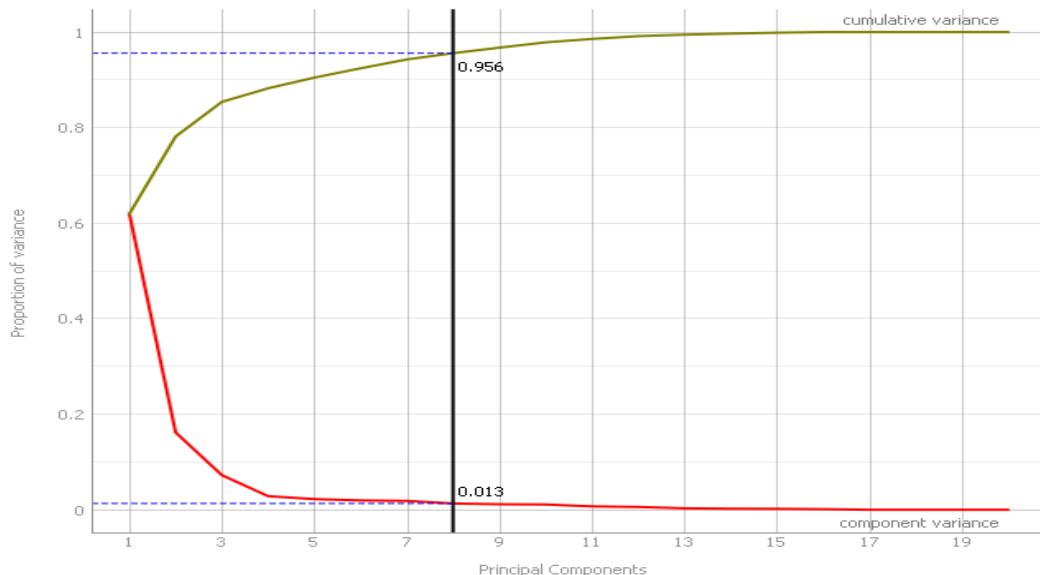


Figure 5. Principal component analysis (PCA) selection.

Table 10. PCA variances.

Component	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
CGPA	0.013	-0.003	0.998	-0.028	0.011	0.017	-0.024	-0.014
Attempt Count	0.002	-0.004	0.007	0.031	0.005	0.145	0.067	-0.082
High Risk = 1	-0.007	0.005	-0.014	-0.117	0.017	-0.151	0.025	0.044
High Risk = 2	0.007	-0.005	0.014	0.117	-0.017	0.151	-0.025	-0.044
Term Exceeded = 1	-0.006	-0.002	-0.015	-0.005	0.005	-0.012	-0.019	0.050
Term Exceeded = 2	0.006	0.002	0.015	0.005	-0.005	0.012	0.019	-0.050
At Risk = 1	-0.020	0.010	0.016	-0.204	0.057	-0.534	0.154	0.261
At Risk = 2	0.020	-0.010	-0.016	0.204	-0.057	0.534	-0.154	-0.261
At Risk SSC = 1	0.000	0.004	0.004	-0.027	-0.003	-0.049	0.016	0.002
At Risk SSC = 2	0.000	-0.004	-0.004	0.027	0.003	0.049	-0.016	-0.002
Other Module Count	-0.008	0.013	-0.011	0.075	0.071	-0.284	-0.265	0.043
Plagiarism	0.016	-0.005	0.003	0.027	0.242	0.287	0.172	0.346
CW1 = 1	-0.011	-0.005	-0.012	-0.065	0.038	-0.132	-0.113	-0.446
CW1 = 2	0.011	0.005	0.012	0.065	-0.038	0.132	0.113	0.446
CW2 = 1	-0.004	-0.003	-0.007	-0.084	0.066	-0.109	-0.075	-0.311
CW2 = 2	0.004	0.003	0.007	0.084	-0.066	0.109	0.075	0.311
ESE = 1	-0.004	0.059	-0.025	-0.638	-0.002	0.232	-0.130	0.052
ESE = 2	0.004	-0.059	0.025	0.638	0.002	-0.232	0.130	-0.052
Played	0.398	0.582	-0.002	0.039	-0.022	-0.020	0.006	-0.010
Paused	0.825	-0.560	-0.014	-0.060	0.001	-0.024	-0.012	0.002
Likes	0.008	0.013	-0.012	0.134	0.522	-0.019	-0.748	0.235
Segment	0.398	0.582	-0.002	0.039	-0.022	-0.020	0.006	-0.010
Moodle Online In campus = 1	-0.009	0.008	0.011	-0.094	0.012	-0.104	-0.047	-0.130
Moodle Online In campus = 2	0.009	-0.008	-0.011	0.094	-0.012	0.104	0.047	0.130
Moodle Outside campus = 1	-0.012	-0.017	0.004	0.033	-0.568	-0.046	-0.325	0.144
Moodle Outside campus = 2	0.012	0.017	-0.004	-0.033	0.568	0.046	0.325	-0.144

4.9. Multivariate Projection

Multivariate visualization is one of the tools used in datamining and provides the starting point in an explorative study. With an increasing number of features, it requires some automatic means of finding good projections that would optimize criteria of quality and expose any inherit structure in the data. FreeViz is used for multivariate projection, which optimizes the linear projection and displays the projected data in a scatterplot. The target projection is found through a gradient optimization approach [50]. Figure 6 depicts the multivariant projection on the dataset; based on the target variable, the projection is made, and two clusters are easily visible. First, the cluster “pass” identified students that passed the module, with relationships in CW1, CW2, ESE, played, paused, likes, segment and Moodle activity either on campus or outside, all of which play a significant role in their overall pass performance; the second cluster, “fail”, identified students that failed the module, with relationships in at risk, SSC, high risk, term exceed, at risk, attempt count, CGPA and Moodle activity within the campus.

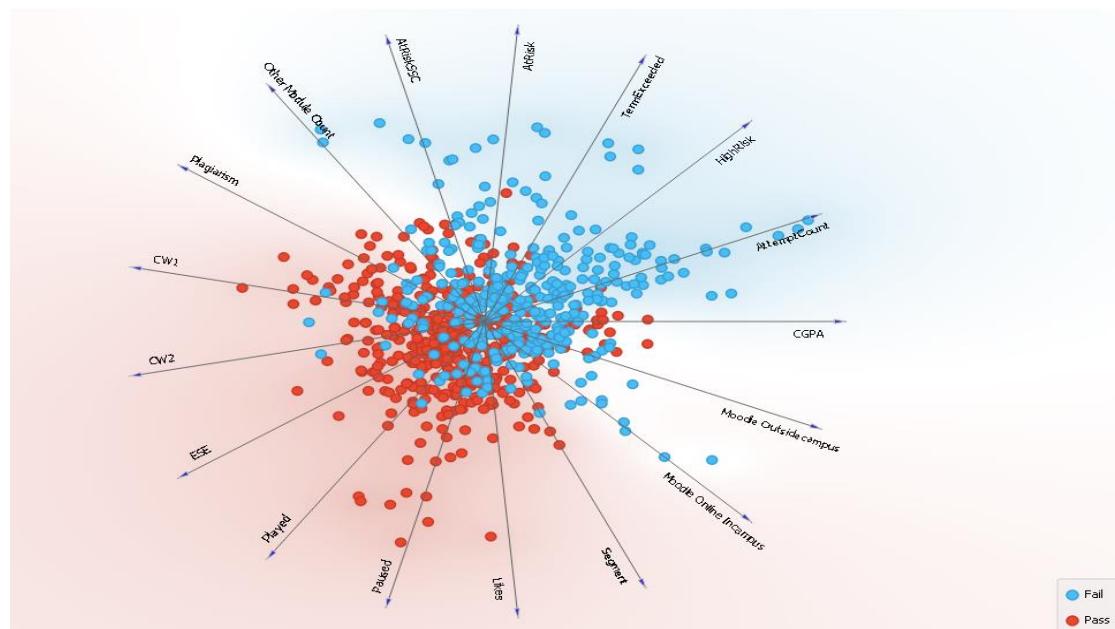


Figure 6. Multivariate projection.

4.10. CN2 Rule Viewer

CN2 Rule Inducer has a closer accuracy to Random Forest; Table 11 depicts the rules set by CN2 Rule Inducer. Out of 45 rules, only 23 rules were selected based on either Moodle activity or video interactions or both. The reason for selecting CN2 Rule Inducer was because it is easier to interpret for non-expert users of data mining. For this case, it would be easier for a faculty member to determine the probabilities of student interactions within the learning setting systems. The interpretation of the rules set by CN2 Rule inducer is that there is a high probability—92%—that a student will pass the module if the student is available on Moodle when active on campus and he/she played the video at least three times or more. A student is likely to pass the module with a probability of 75% if there is Moodle activity in and outside the campus. A student has a high chance of passing the end-of-semester examination if they engaged in the activities outside the campus on Moodle, played the video more than once and paused the video at least three times or more, with a probability of 94%. A student spending time on Moodle activities on campus has a fair chance of passing the module, with a probability of 83%. The student is likely to pass the module if they use Moodle outside the campus and like a video which they paused either twice or more. A video being paused more than or equal to 10 times and accessing Moodle outside the campus indicates a probability of 67% that the

student will not be successful at the end semester examination. A student playing a video equal to or more than five times and exhibiting activity Moodle outside the campus has a probability of 90% of passing the module. If the student pauses a video more than or equal to four times and has low Moodle activity, they are likely to pass the module with a probability of 83%. If a student likes the video content equal to or more than twice with little Moodle activity outside the campus, they have a chance of 88% of passing the module. A student has a probability of 87% of passing the module if the student has Moodle activity outside the campus and played the video at least twice. Students' CW1 marks are important, as if the student failed the CW1 and played the video equal to or less than twice along with rewinding to capture the concepts, they have a chance of 75% of passing the module. A student playing the video content many times, such as more than or equal to five times, and pausing the video many times has a 91% chance of passing the module. A student who engages in rewinding and listening to the content many times has high chances, at 93%, of passing the module. For video content, a student who pauses the videos many times and moves slowly to gather the concepts has an 80% probability of passing the module, but if the frequency is more than or equal to 13 times, the student has an 83% chance of failing the end-of-semester examination. If a student failed in CW1 and exhibits low activity on Moodle on campus, there are high chances of failure in the module, at 90%. If a student plays the video content at least twice, with a low activity on Moodle outside the campus, and pauses the video much more than seven times, they have high chances of passing the module, at 90%. If a student passed in CW1, plays the video at least twice, exhibits a high activity on Moodle outside campus and likes the video more than once, they have high chances of passing the module at 80%. If a student pauses the video less than four times, exhibits a high activity on Moodle outside campus and plays the video more than twice, they have high chances of passing the module, at 90%. Playing video content more than or equal to five times gives a high probability of 83% that the student will fail in the module. Similarly, if the video is paused more than or equal to seven times, the probability goes to 62%. Playing a video up to four times results in a probability of 71% of a student passing the module. It is evident from the analysis and visualization projection (Figure 5) that videos being played, paused, segments and likes along with either in-campus or off-campus use can enhance learning as compared to the normal setting in which this opportunity was not available.

Table 11. Rules set by CN2.

No.	IF Conditions	THEN Class	Probability
1	ESE≠Fail AND Activity on campus=LOW AND Played≥3	Result = Passed	92%
2	ESE≠Fail AND Activity on campus=LOW AND Activity out campus=LOW	Result = Passed	75%
3	ESE≠Fail AND Activity out campus=LOW AND Paused≥3 AND Played≤2	Result = Passed	94%
4	ESE≠Fail AND Activity on campus=LOW	Result = Passed	83%
5	ESE≠Fail AND Activity out campus=LOW AND Likes≥1 AND Paused≥2	Result = Passed	94%
6	ESE≠Fail AND Activity out campus=LOW AND Paused≥10	Result = Failed	67%
7	ESE≠Fail AND Played≥5 AND Activity out campus=LOW	Result = Passed	90%
8	ESE≠Fail AND Activity out campus=LOW AND Paused≥4	Result = Passed	83%
9	ESE≠Fail AND Activity out campus=LOW AND Likes≥2	Result = Passed	88%
10	ESE≠Fail AND Activity out campus=LOW AND Played≥2	Result = Passed	87%
11	ESE≠Fail AND Played≤2 AND CW1≠Fail AND Segment≥1	Result = Passed	75%
12	ESE≠Fail AND Played≥5 AND Paused≤7	Result = Passed	91%
13	ESE≠Fail AND CW1=Fail AND Segment≥2	Result = Passed	83%
14	ESE≠Fail AND Paused≥10	Result = Passed	80%
15	Paused≥13	Result = Failed	83%
16	CW1 = Fail AND Activity on campus≠LOW	Result = Failed	90%
17	Played≤2 AND Activity out campus≠LOW AND Paused≥7	Result = Passed	90%
18	Played≤2 AND CW1≠Fail AND Activity out campus≠LOW AND Paused≥6	Result = Failed	83%
19	CW1≠Fail AND Played≤2 AND Activity out campus≠LOW AND Likes≥1	Result = Passed	80%
20	Paused≤4 AND Activity out campus≠LOW AND Played≥2	Result = Passed	90%
21	Played≥5	Result = Failed	83%
22	Played≥3	Result = Passed	71%
23	Paused≥7	Result = Failed	62%

5. Discussion

The study was conducted to create a model that can predict students' end-of-semester academic performance in the modules e-commerce (COMP 1008) and e-commerce technologies (COMP 0382). End-of-semester results from these two modules were taken into consideration as the performance indicators and were considered as target features. The data of students from these modules was collected using three different systems such as SIS (student academics), LMS (online activities) and eDify (video interactions). Firstly, the performance of the most widely used classification algorithms was compared using the complete dataset. Secondly, data transformation and feature selection techniques were applied to the dataset to determine the impact on the performance on the classification algorithms. Lastly, we reduced the features and determined the appropriate features that can be used in order to predict the students' end-of-semester performance. Four performance metrics—accuracy, sensitivity, specificity and F-Measure—were evaluated in order to compare the performance of the classification models.

The effect of data transformation on classification performance was tested by converting the features into a categorical form using the techniques of equal width and equal frequency. Results indicated that models with categorical data performed better than those with continuous data. Observation showed that equal width data transformation performance results were better compared with equal frequency data transformation.

The impact of feature election on classification performance was tested. For this purpose, a ranking and scoring technique was used with three feature selection methods: information gain, information gain ratio and Gini decrease. Nine features were determined from ranking and scoring techniques—CW1, ESE, CW2, likes, paused, played, segment, Moodle on campus and Moodle off campus—and were tested regarding the performance of the classification models.

Moreover, feature selection using the genetic algorithm was used to further reduce the features. The features were reduced from nine to six: applicant, at risk, CW1, CW2, ESE and played. Here, a meta attribute was selected and the Moodle activity feature was not selected. The performance of the classification model accuracy was less than the data transformation method and the information gain ratio selection technique. Then, a PCA analysis was performed to reduce the features with a variance of 95.6%; PCA derived eight component solutions. The feature selection method enables the prediction model to be interpreted more easily. However, PCA resulted in one fewer component, and it is difficult to interpret for a normal faculty member from a different specialization field.

Furthermore, multivariate projection using a gradient optimization approach and CN2 Rule viewer was used to interpret the relationship within the dataset and the features. Relationship, features that were extracted with help of data transformation method and information gain ratio selection technique showed a resemblance which enhances the understanding for the novice users..

From the results of this study, we can conclude that Random Forest provides a high classification accuracy rate in conditions where equal width data transformation method and information gain ratio selection techniques were used (Table 7). The Random Forest outperformed the other selected algorithms in predicting successful students. The confusion matrix for 10-fold cross-validation derived via Random Forest is provided in Table 12. The classification model correctly predicted 629 students out of 645 students, with an accuracy of 88.3%.

Table 12. Sample confusion matrix.

		Predictions			
Actual		Failed	Passed	Σ	
		Failed	Passed	Σ	
	Failed	53	74	127	
	Passed	16	629	645	
	Σ	69	703	772	

6. Conclusions and Future Works

In this study, a supervised data classification model is proposed with the aim of predicting student academic performance at the end of the semester. Two modules were selected for the study based on the similarity of the course content. The dataset consisted of student academic data gathered from SIS (student academics) and performance in the modules using two different learning environments; i.e., Moodle (LMS) and eDify (mobile application). Activities performed by students in Moodle on campus and outside campus were used. Video interactions (clickstreams) of students in eDify were used for prediction. In total, 18 features and one meta attribute were used to form the dataset, 12 features were extracted from SIS, 2 from Moodle, four from eDify and one result was used as the target feature for prediction. The dataset consisted of 772 samples from one academic year. The complete dataset was tested with eight classification algorithms derived from the literature review and related works.

The Tree-based classification model—specifically, Random Forest—outperformed the other techniques with an accuracy of 88.3%. This accuracy was achieved using the equal width data transformation method and information gain ratio selection technique. To identify which features have greater importance in the prediction of student performance, features were reduced using the genetic algorithm and PCA. The result was inconclusive as the reduced features have low significance in predicting student performance. Multivariate analysis was conducted to inspect the correlations, and nine variables were selected using scoring and ranking techniques to successfully predict the students' academic performance. Thus, the results obtained with reduced features were better than those using all the features. The CN2 Rule Inducer algorithm was the second-best performing algorithm, showing 87.4% accuracy. The reason for using CN2 Rule Viewer is that it provides rules induction with probability, which is easier to interpret for non-expert users such as faculty who can therefore easily relate to the situation, as shown in Table 12.

For future work, a dashboard with data representations from these virtual learning environments would help in projecting the students' performance and interactions. Predicting the students' performance and outcomes on a weekly basis could help faculty to identify poor-performing students. This can act as an early alert system for faculty to intervene with any problems faced by the students within the module. Students could also self-assess their own performance within the module with the help of a dashboard.

Author Contributions: R.H. contributed to the investigation and project administration. S.P. contributed to the supervision- S.M. contributed to the visualization. M.U.S. contributed to the resources and writing—review and editing. A.A. and K.U.S. contributed to the collected data and conducted the pre-processing of the input data. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors would like to thank Middle East College, Oman for the experimental data. The authors are thankful to the Head of Computing Department, Mounir Dhibi (MEC), for his support and encouragement to carry out this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Castro, R. Blended learning in higher education: Trends and capabilities. *Educ. Inf. Technol.* **2019**, *24*, 2523–2546. [[CrossRef](#)]
2. Ali, M.F.; Joyes, G.; Ellison, L. Using blended learning to enhance students' cognitive presence. In Proceedings of the 2013 International Conference on Informatics and Creative Multimedia, ICICM 2013, Kuala Lumpur, 4–6 September 2013.
3. Cleveland-Innes, M.; Wilton, D. *Guide to Blended Learning*; Commonwealth of Learning: Burnbay, BC, Canada, 2018, ISBN 9781894975940.
4. Medina, L.C. Blended learning: Deficits and prospects in higher education. *Australas. J. Educ. Technol.* **2018**, *34*. [[CrossRef](#)]

5. Schmidt, S.M.P.; Ralph, D.L. The flipped classroom: A twist on teaching. *Contemp. Issues Educ. Res.* **2016**, *9*. [[CrossRef](#)]
6. Graham, C.R.; Woodfield, W.; Harrison, J.B. A framework for institutional adoption and implementation of blended learning in higher education. *Internet High. Educ.* **2013**, *18*, 4–14. [[CrossRef](#)]
7. 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](#)]
8. Romero, C.; Ventura, S. Educational data mining and learning analytics: An updated survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2020**, *10*, e1355. [[CrossRef](#)]
9. Naidu, V.R.; Singh, B.; Hasan, R.; Al Hadrami, G. Learning analytics for smart classrooms in higher education. *IJAEDU- Int. E-J. Adv. Educ.* **2017**, 440–446. [[CrossRef](#)]
10. Lester, J.; Klein, C.; Rangwala, H.; Johri, A. Learning analytics in higher education. *ASHE High. Educ. Rep.* **2017**, *43*, 9–135. [[CrossRef](#)]
11. Atif, A.; Richards, D.; Bilgin, A.; Marrone, M. Learning analytics in higher education: A summary of tools and approaches. In Proceedings of the 30th Annual conference on Australian Society for Computers in Learning in Tertiary Education, ASCILITE 2013, Sydney, Australia, 1–4 December 2013.
12. 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](#)]
13. 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](#)]
14. Hasan, R.; Palaniappan, S.; Mahmood, S.; Naidu, V.R.; Agarwal, A.; Singh, B.; Sarker, K.U.; Abbas, A.; Sattar, M.U. A review: Emerging trends of big data in higher educational institutions. In *Micro-Electronics and Telecommunication Engineering*; Lecture Notes in Networks and Systems; Springer: Singapore, 2020; Volume 106, pp. 289–297. [[CrossRef](#)]
15. Chaudhury, P.; Tripathy, H.K. An empirical study on attribute selection of student performance prediction model. *Int. J. Learn. Technol.* **2017**, *12*, 241–252. [[CrossRef](#)]
16. Romero, C.; Ventura, S. Data mining in education. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2013**, *3*, 12–27. [[CrossRef](#)]
17. Romero, C.; López, M.I.; Luna, J.M.; Ventura, S. Predicting students' final performance from participation in on-line discussion forums. *Comput. Educ.* **2013**, *68*, 458–472. [[CrossRef](#)]
18. 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: Revolutionising Digital Landscape for Sustainable Smart Society, ICCOINS 2018—Proceedings, IEEE, Kuala Lumpur, Malaysia, 13–14 August 2018; pp. 1–5.
19. Shana, J.; Venkatachalam, T. Identifying key performance indicators and predicting the result from student data. *Int. J. Comput. Appl.* **2011**, *25*. [[CrossRef](#)]
20. Kash, B.R.P.; Thappa, D.M.H.; Kavitha, V. Big data in educational data mining and learning analytics. *Int. J. Innov. Res. Comput. Commun. Eng.* **2015**, *2*, 7515–7520. [[CrossRef](#)]
21. Siemens, G.; Gasevic, D. Guest editorial - learning and knowledge analytics. *Educ. Technol. Soc.* **2012**, *15*, 1–2.
22. Akçapınar, G.; Altun, A.; Aşkar, P. Using learning analytics to develop early-warning system for at-risk students. *Int. J. Educ. Technol. High. Educ.* **2019**, *16*, 1–20. [[CrossRef](#)]
23. Arnold, K.E.; Pistilli, M.D. Course signals at Purdue: Using learning analytics to increase student success. In Proceedings of the ACM International Conference Proceeding Series, Vancouver, BC, Canada, 29 April–2 May 012.
24. Corrigan, O.; Smeaton, A.F.; Glynn, M.; Smyth, S. Using educational analytics to improve test performance. In Proceedings of the Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Toledo, Spain, 15–18 September 2015.
25. Saqr, M.; Fors, U.; Tedre, M. How learning analytics can early predict under-achieving students in a blended medical education course. *Med. Teach.* **2017**, *15*, 1–11. [[CrossRef](#)]
26. Chatti, M.A.; Schroeder, U.; Jarke, M. LaaN: Convergence of knowledge management and technology-enhanced learning. *IEEE Trans. Learn. Technol.* **2012**, *5*, 177–189. [[CrossRef](#)]

27. Hadhrami, G.A. AI learning analytics dashboard to improve students' performance and success. *IOSR J. Res. Method Educ.* **2017**, *7*, 39–45. [[CrossRef](#)]
28. 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 Distance Learn.* **2015**, *16*. [[CrossRef](#)]
29. 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, Workshop Proceedings, Manila, Philippines, 26–30 November 2018.
30. Hasan, R.; Palaniappan, S.; Mahmood, S.; Abbas, A.; Sarker, K.U. Modelling and predicting student's academic performance using classification data mining techniques. *Int. J. Bus. Inf. Syst.* **2020**, *1*, 1. [[CrossRef](#)]
31. Romero, C.; Ventura, S. Educational data mining: A survey from 1995 to 2005. *Expert Syst. Appl.* **2007**, *33*, 135–146. [[CrossRef](#)]
32. Romero, C.; Ventura, S.; De Bra, P. Knowledge discovery with genetic programming for providing feedback to courseware authors. *User Model. User-Adapt. Interact.* **2004**, *14*, 425–464. [[CrossRef](#)]
33. 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](#)]
34. Shetty, I.D.; Shetty, D.; Roundhal, S. Student performance prediction. *Int. J. Comput. Appl. Technol. Res.* **2019**, *8*, 157–160. [[CrossRef](#)]
35. Abu Zohair, L.M. Prediction of Student's performance by modelling small dataset size. *Int. J. Educ. Technol. High. Educ.* **2019**, *16*. [[CrossRef](#)]
36. 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—WWW '17 Companion, Perth, Australia, 7 April 2017; ACM Press: New York, NY, USA, 2017; pp. 415–421.
37. 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](#)]
38. 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: Berlin, Germany, 2020; pp. 229–239. ISBN 9783030141172.
39. López, M.I.; Luna, J.M.; Romero, C.; Ventura, S. Classification via clustering for predicting final marks based on student participation in forums. In Proceedings of the 5th International Conference on Educational Data Mining, EDM 2012, Chania, Greece, 19–21 June 2012.
40. Yassein, N.A.; M Helali, R.G.; Mohomad, S.B. Predicting student academic performance in KSA using data mining techniques. *J. Inf. Technol. Softw. Eng.* **2017**, *7*, 1–15. [[CrossRef](#)]
41. Govindasamy, K.; Velmurugan, T. Analysis of student academic performance using clustering techniques. *Int. J. Pure Appl. Math.* **2018**, *119*, 309–322.
42. Veeramuthu, P.; Periyasamy, R.; Sugasini, V.; Patti, P. Analysis of student result using clustering techniques. *Int. J. Comput. Sci. Inf. Technol.* **2014**, *5*, 5092–5094.
43. Saneifar, R.; Saniee Abadeh, M. Association Rule Discovery for Student Performance Prediction Using Metaheuristic Algorithms. *Comput. Sci. Inf. Technol. (CS & IT)* **2015**, *5*, 115–123.
44. Chandrakar, O.; Saini, J.R. Predicting examination results using association rule mining. *Int. J. Comput. Appl.* **2015**, *116*, 7–10. [[CrossRef](#)]
45. Ferguson, R.; Macfadyen, L.P.; Clow, D.; Tynan, B.; Alexander, S.; Dawson, S. Setting learning analytics in context: Overcoming the barriers to large-scale adoption. *J. Learn. Anal.* **2014**, *1*, 120–144. [[CrossRef](#)]
46. Márquez-Vera, C.; Cano, A.; Romero, C.; Ventura, S. Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data. *Appl. Intell.* **2013**, *38*, 315–330. [[CrossRef](#)]
47. Márquez-Vera, C.; Cano, A.; Romero, C.; Noaman, A.Y.M.; Mousa Fardoun, H.; Ventura, S. Early dropout prediction using data mining: A case study with high school students. *Expert Syst.* **2016**, *33*, 107–124. [[CrossRef](#)]
48. Cano, A.; Leonard, J.D. Interpretable multiview early warning system adapted to underrepresented student populations. *IEEE Trans. Learn. Technol.* **2019**, *12*, 198–211. [[CrossRef](#)]

49. Demšar, J.; Curk, T.; Erjavec, A.; Gorup, Č.; Hočević, T.; Milutinović, M.; Možina, M.; Polajnar, M.; Toplak, M.; Starič, A.; et al. Orange: Data mining toolbox in python. *J. Mach. Learn. Res.* **2013**, *14*, 2349–2353.
50. Demšar, J.; Leban, G.; Zupan, B. FreeViz-An intelligent multivariate visualization approach to explorative analysis of biomedical data. *J. Biomed. Inform.* **2007**, *40*, 661–671. [CrossRef]



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).