Predicting Learners' Performance in Virtual Learning Environment (VLE) based on Demographic, Behavioral and Engagement Antecedents

Ahmed Al-Azawei (✉), Miami Abdul Aziz Al-Masoudy
University of Babylon, Babel, Iraq
`ahmedhabeeb@itnet.uobabylon.edu.iq`

Abstract—This study aims at predicting students' performance in a Virtual Learning Environment (VLE) based on four time periods of the examined online course in order to provide an early prediction model. The research is based on data from one of the scientific courses at the Open University (OU) in Britain. The investigated data consists of 1938 students in which the influence of demographic and behavioral variables was explored first. Then, three features were generated to improve the prediction accuracy as well as examining the effect of learners' engagement on their academic performance. Thus, the prediction accuracy of the original features and their integration with the generated attributes was compared. The findings suggest that some of the demographic variables and all online behavioral features had a significant impact on students' performance. However, the accuracy was highly improved after using the generated features. It was found that the level of the financial and service instability, assessment grades, the total number of clicks, the interaction with different course activities, and students' engagement were significant predictors of academic achievement.

1 Introduction

In contemporary education, universities aim at improving the quality of teaching and learning as well as enhancing students' performance [1]. Currently, different educational modes are available such as Face to Face (F2F) learning, e-learning, blended learning, and online learning. The latter, however, becomes very popular in contemporary education [2]. It can be offered in many forms such as massive open online courses (MOOCs), virtual learning environments (VLE), and learning management systems (LMSs) [3]. On the other hand, earlier literature shows that a high number of students either drop out or fail to achieve good scores in online learning environments [4]. Moreover, the number of dropouts from online learning courses is higher than that in traditional learning [5]. This is more evident in developing nations because leaners still face many barriers in adopting this learning form such as the lack of stu-

dents' motivation and the direct interaction between teachers and students as well as the absence of a learning atmosphere [6][7]. Hence, identifying the actual level of students or guess their possible achievement may be a difficult process. Moreover, teachers may face an issue in providing appropriate advice for students or changing the method of presenting learning content to meet learners' preferences.

Based on the above-discussed disadvantages, it is essential to investigate behavioral and demographic features in understanding learners' online achievement. Educational data mining (EDM) approaches can be used in analyzing factors that may affect learners' performance [4]. EDM is defined as the "process used to extract useful information and patterns from a huge educational database" [8]. The main functions of EDM are discovering and extracting different patterns in order to use them in predicting students' performance [18]. However, there is still a lack of research investigating factors that may influence students' achievement in online courses [8]. Thus, predicting students' performance is considered as an important topic in EDM.

Accordingly, this research aims at covering three key objectives. First, it investigates the effect of demographic and online behavioral factors on learners' achievement in VLE. Furthermore, the study attempts to select the best features that may affect students' performance. Finally, it determines the relationship between students' engagement and achievement. The research outcomes can extend previous work and help overcome obstacles that students may face in VLE.

The rest of this paper is structured as follows. Section two presents the theoretical background and reviews earlier research on predicting students' achievement. The methodology of the study is shown in section three. In section four, the experimental results are presented and discussed. The last section concludes the key points highlighted in this research and suggests possible future directions.

2 Theoretical Background and Previous Work

2.1 Theoretical background

This study adopts the Classification via Regression method. This method classifies a particular data using a regression approach by binarizing the class and building a regression model for each class value [9][10]. The M5P regression algorithm is used here. M5P is a supervised algorithm that combines a conventional decision tree with the possibility of linear regression functions at the leaves [17]. It can be used to predict a numeric target (class) attribute.

2.2 Previous work

Generally, students' academic achievement can be predicted using demographic features, behavioral variables, previous scores achieved during a course, and/or integrating all of them.

Hussain et al. [3] aimed at predicting low-engagement students and identifying the relationship between students' engagement and their course assessment scores. The

selected variables in predicting learners' engagement were the highest level of education, final results, assessment scores, and the number of clicks on VLE. Six classification algorithms were used in which the J48 technique achieved the best accuracy (88.52%). The results also showed that students' clicks on the homepage, forum, course content, and subpages were the best predictors of students' engagement. Moreover, it was found that students' activities in VLE positively affected their engagement and scores.

Daud et al. [11] built a model to predict whether students would complete their degrees successfully or not. Demographic features were used in analyzing the proposed model. Many algorithms were implemented in which Support Vector Machine (SVM) outperformed others with an F1 score of 0.867. Furthermore, the research findings revealed that natural gas expenditure, electricity expenditure, self-employed, and location were the most influential factors in predicting learners' performance.

Eduardo et al. [12] investigated variables that may affect learners' performance based on demographic and previous score features. Demographic factors were collected prior to the beginning of the course, whereas 'absence', ' grades' and 'school subjects' were recorded during the course. The study compared the predictive capability at two different times in the course. A classification model was built based on the Gradient Boost Machine (GBM) for each dataset. The findings of the first classification model (CM-I) suggested that the most important variables were 'neighborhood', 'school', 'city', and 'age'. Such results demonstrated that students' demographic features directly influence the teaching-learning process. On the other hand, the second classification model (CM-II) showed that 'grade', 'absence' and 'school subject' variables were significant determinants of students' final results.

Umer et al. [13] conducted a comparative analysis of four predictive models namely, Random Forest (RF), Naive Bayes (NB), K-Nearest Neighbor (KNN), and Linear Discriminant Analysis (LDA) in order to predict students' final outcomes. The study aimed at identifying students who were at risk of failing. The prediction was performed weekly based on VLE engagement data and assignment scores. The outcomes showed that assignment scores were the most discriminative variable where the prediction accuracy reached to 70% after week-1. Results also showed that Random Forest outperformed other classifiers in all weeks.

Sukhbaatar et al. [14] proposed an early prediction scheme to identify students at risk of failing in a blended learning course. The Neural Networks technique was performed on a set of prediction variables extracted from the online learning activities, in which data of five years were used in validating the proposed model. The integrated variables were online quiz scores, mid-term scores, and final grade information. The study correctly predicted 25% of the failing students after the first quiz. The prediction accuracy gradually increased week by week, reaching 53% after the 8th quiz and 65% after the mid-term exam. The experimental results presented the possibility of developing an early warning system using online learners' activities.

Jiang et al. [15] used a combination of students' Week 1 assignment performance and social interaction within MOOC to predict the final performance. The role of external incentives in predicting final MOOC performance was also investigated. Two logistic regression models were built. The predicted variable for the first model was

the type of certificate learners obtained such as distinction or normal. The predicted variable for the second model was whether students obtained a normal certificate or did not complete the MOOC course. The predictors were: the average quiz score learners obtained in the first week, the number of peer assessments students completed in Week 1, the learners' social network degree in Week 1, and whether or not a learner is an incoming undeclared major student. The outcomes found that assignment performance in Week 1 was a strong predictor of students' performance at the end of the course. The degree of social integration in the learning community in Week 1 was also positively correlated with the achievement of distinction certificates. Furthermore, students with external incentives were more likely to complete the course compared to those in general, even in comparison with students who had similar backgrounds. The first model achieved 92.6% of accuracy, whereas the accuracy of the second was 79.6%.

In this present study, behavioral, demographic, and performance features are used in predicting students' academic achievement in VLE. The prediction is based on multi-time periods. It was after the second (53 days of the course), fifth (after 165 days of the course), sixth (after 207 days of the course) assessments, and a day before the final exam. This research also calculates the level of students' engagement based on a developed formula. Moreover, it generates new features from the available dataset. Finally, the most influential factors that may affect students' academic levels are also highlighted. Thus, this research establishes a new direction in predicting online students' performance in comparison to previous literature.

3   Research Methodology

Figure 1 shows the proposed research model and the main steps followed in order to achieve the key aims of this study.

3.1   Data collection

The Open University Learning Analytics Dataset (OULAD) is used in this research. This is one of the largest universities in Europe. About 200,000 students were enrolled in different courses at the Open University (OU) [3]. There are many modules in this dataset which include Science, Technology, Engineering, and Mathematics (STEM), and three Social Science modules. The dataset includes information about 38,239 students [16]. Table 1 summarizes the main courses and the number of students in each module.
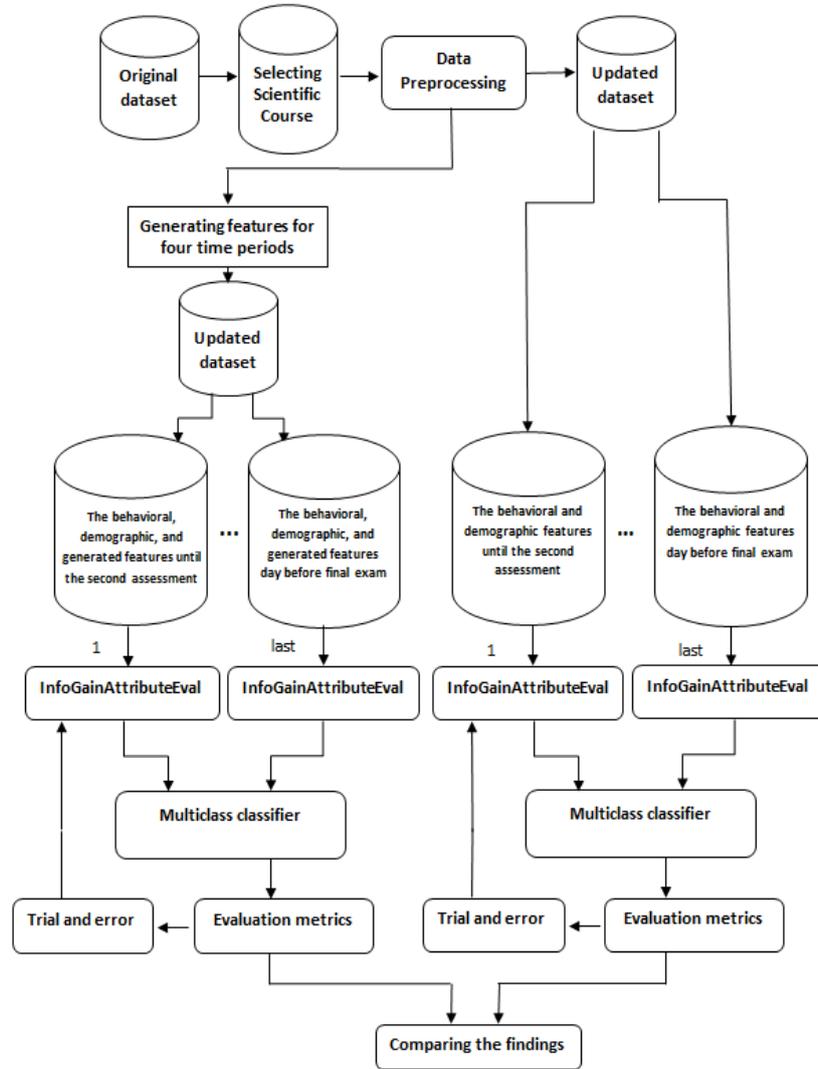
Fig. 1. The proposed research model

Table 1. Module summary and domain information [16]

Module	Domain	Presentations	Students
AAA	Social Sciences	2	748
BBB	Social Sciences	4	7,909
CCC	STEM	2	4,434
DDD	STEM	4	6,272
EEE	STEM	3	2,934
FFF	STEM	4	7,762
GGG	Social Sciences	3	2,534

VLE stores course lectures, materials, and assessment information. Students interacted with the VLE in order to watch lectures, complete assignments, read materials, and communicate with each other in which their interaction with the VLE was recorded and stored in log files [3]. Students' information was stored in seven tables namely, info, student assessment, assessments, student VLE, courses, student registration, and VLE. These tables include the following information [16]:

- The student info table contains the students' demographic information and the results of each course.
- The course table contains information about the courses in which students are enrolled.
- The registration table contains student record timestamps and course enrollment dates.
- The assessment information is recorded in the assessment table.
- The student assessment table contains the assessment results of different students.
- The interaction information of different students regarding different materials and activities is stored in the student-VLE table. The VLE interaction data consists of the number of clicks students made while studying the course material.
- Each course activity is identified with a label (activity type) such as resource, content, forum, homepage, subpage, wiki, URL, collaborate, glossary, externalquize. These types are stored in the VLE table.

The number of times each student clicks on each of the activities is recorded daily in a time-stamped log file that indicates the time students spent on each activity. The forum variable refers to the discussion forum where students can discuss problems with each other. The forum is also a space where students can submit questions to better understand the subject [3]. Resources consist of lecture notes, books, lecture slides, and other course materials in HTML and PDF formats [3]. The content variable contains study materials in HTML format related to a particular course. The subpage variable reveals students' navigation path through the VLE structure [3]. The homepage variable reflects the first screen of every course; these screens are visited by a student before accessing other course material. The glossary includes details about the OU and higher education acronyms. Figure 2 represents the relationship between the dataset tables.

In the present study, data of students (N=1938) who enrolled in a Science module in October 2013 were used. Three different data types are included in the dataset:

Demographic: This represents the basic information about the students. Here, we used gender (the student's gender), region (the geographic region, where the student lived while taking the module-presentation), highest education (highest student education level on entry to the module presentation), imd band (specifies the index of multiple deprivation band of the place where the student lived during the module-presentation), age band (band of the student's age), number of previous attempts (the number times the student has attempted this module), and disability (indicates whether the student has declared a disability) attributes.

Learning behavior: This includes students' interaction with different id_site in the course [3].

Performance: Reflects students' results and achievements during their studies at the OU.

This dataset was subjected to different preprocessing steps due to many issues associated with it. Fig. 2 presents the main tables of this dataset.
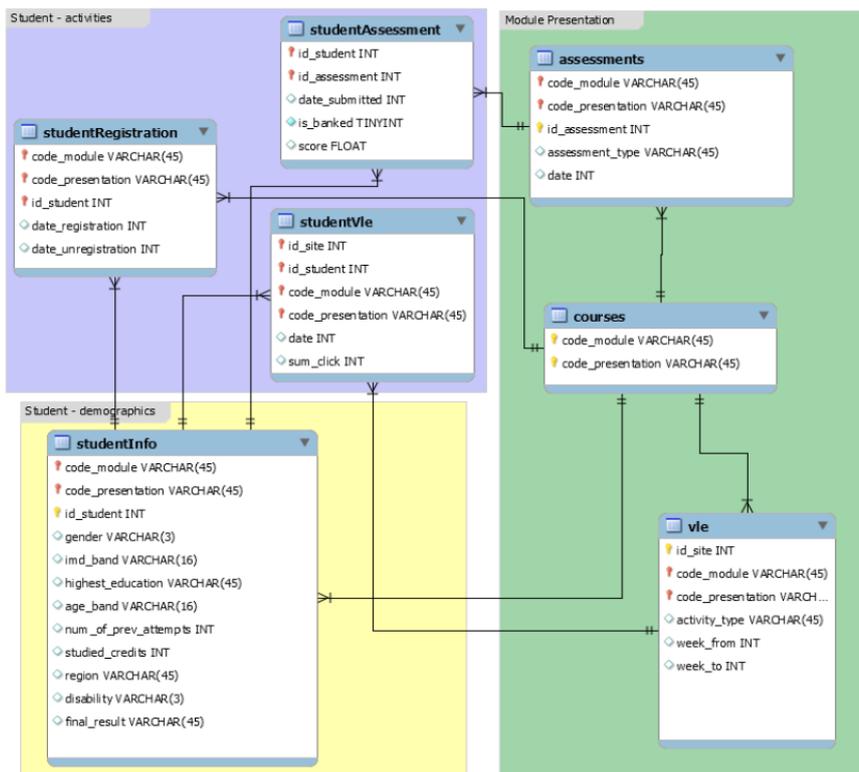

Fig. 2. Main tables in the selected dataset [16]

3.2 Data preprocessing

Preprocessing includes several purposes such as removing noise, handling missing values, and inconsistent data [19]. Preprocessing consists of a number of different strategies and techniques. In general, all of these strategies fall into two categories: selecting data objects and attributes for the analysis or creating/changing the attributes [20]. In this research, many steps were followed in order to prepare the research data for the prediction model.

Data integration: In this study, data were arranged first in order to integrate all attributes in one table. The attributes are classified into three types: demographics, performance, and behavioral features. Demographic attributes were taken directly from the student information table. This includes gender, region, highest education, deprivation band (imd band), age band, disability, and the number of previous attempts.

The behavioral attributes are extracted from the 'student vle' table. This was carried out by adding students' interaction with each site to the total of their interaction with this type to which this site belongs. The type of site is defined through the use of the 'vle' table which contains all id sites and the type of each one. Ten types of id sites were found namely, resource, content, forum, homepage, subpage, wiki, URL, collaborate, glossary, and externalquize. Therefore, ten new attributes were obtained. Students' interaction with these types of activities was calculated at four different intervals. The study calculates the interaction until the second, fifth, and sixth assessments and then a day before the final exam. Initially, each id site that students interacted with was converted into a number that symbolizes its type. These types were coded by numbers from one to ten. Algorithm 1 shows the main steps performed in this process. Algorithm 2 explains the computation of learners' interaction at the four-time periods.

Algorithm 1. Transformation of id sites into numbers

```
Algorithm Transform_To_Type

 Input: id_siteA [1*n], id_site [1*m] and Type [1*m] are
three arrays, where id_siteA represents the interaction
sites by students, id_site contains all interaction
course sites by students and Type is the types of id
sites in the id_site array.
 Output: activity_type [1*n] is an array representing
the type of id sites in the id_siteA array in the order
of their appearance.
 for i=1 to n
  for j=1 to m
  if ( id_siteA[i] = id_site[j] ) then
  if type[j] = resource then
  activity_type[i]=1
  else
  if type[j] = content
  activity_type [i]=2
```

```
  else
  .
  .
  else
  If type[j] = glossary
  activity_type [i]=10
  end if
  end if
 end for
 end for
```

Algorithm 2. Counting the activities of all students in four time periods

```
Algorithm Count_Activities

 Input: id_assess [1xn], all_date [1xn] are two arrays,
where id_assess represents
 the assessment IDs in which students' interactions are
calculated until they are made, all_date contains the
official date for each assessment.

 stu_assess [1xm], stu_assess_id [1xm], assess_date
[1xm] are three arrays, represent the student IDs, the
assessment IDs that students conducted and the dates of
their conducting to that assessment IDs respectively.

 id_stu [1xs], act [1xs], activity_date [1xs], etera-
tion_of_act [1xs] represent the student IDs, the numbers
which represent the type of id sites that students inter-
acted with, the date of conducting these interactions,
and the number of clicks to each id site in each entry
respectively.

 id_stuA [1xh] is the student IDs of the course which
are wanted to calculate their interactions.

 Output: result [hx40] represents the number of stu-
dents' interactions with each activity type for four time
periods. Initially, this is a zero matrix.
 for t=1 to n
 for i= 1 to h
  for k=1 to m
  if (stu_assess [k] = id_stuA [i] ) AND (id_assess [t]
= stu_assess_id [k] ) then
  date=assess_date[k];
```

```
  else
  if ( k = m) AND (stu_assess [k]!=id_stuA[i] OR
id_assess[t] !=assess_id[k] )
  date=all_date[t];
  end if
  end for k
  for j= 1 to s
  if (id_stu[j] = id_stuA[i]) AND (activity_date[j] <=
date )
  if t=1
  result [i,act[j]]= result[i,act] +eteration_of_act[j]
  else
  if t=2
  result [i,act[j]+10]= result
[i,act[j]+10]+eteration_of_act[j]
  else
  if t=3
  result [i,act[j]+20]= re-
sult[i,act[j]+20]+eteration_of_act[j]
  else
  if t=4
  result [i,act[j]+30]= re-
sult[i,act[j]+30]+eteration_of_act[j]
  end if
  end if
  end for j
 end for i
 end for t
```

Generating new features. New features were generated in this study in order to enhance the accuracy of the prediction process. These features are:

- Total number of activities: This attribute was calculated from the students' behavioral attributes with all id sites until the prediction day (days 53, 165, 207).
- Average: This attribute is generated based on the grades of students' assessments until the prediction day. Each assessment takes weight at this average according to its weight relative to the rest of the course assessments. Here, the weights of first, second, third, fourth, fifth and sixth assessments were 10, 12.5, 17.5, 20, 20, and 20 respectively.
- Engagement: This attribute reflects the level of students' motivation until the prediction day. It is expected that this feature would have a significant effect on students' performance. Students were divided into very low, low, and high engagement levels. Therefore, three values (0, 0.25, 1) were used in referring to this attribute respectively. It was calculated based on the following developed formula:

$$\text{Engagement} = \begin{cases} \textbf{0, (average =-1) V (average =0 \& total number of activities=0)} \\ \\ \textbf{0.25, (current\_last\_assessment <50) ^ (average < = 65) ^ (total} \\ \qquad\quad \textbf{number of activities < threshold)} \\ \\ \textbf{1, Otherwise} \end{cases}$$

Where current_last_assessment represents the last assessment made by students within the time period for which the level of students' participation was calculated. This is either the second, fifth, or sixth assessments. The threshold is the average of the interaction of all students.

Missing values: Preprocessing missing values is an important stage in data analysis. In the used dataset, there are some missing values either in the assessment scores or in the deprivation band (imd band) features. Zero was placed instead of missing assessments based on the Open University's assertion of its negligence of all assessment values that students did not perform. Moreover, the most frequent value in the deprivation band (imd band) attribute was considered as a substitute for its missing values.

Normalization: It was performed for all values of numeric features that would be the input for the machine learning algorithms. This includes many features namely, number of previous attempts, dataplus, forum, glossary, collaborate, content, resources, subpage, homepage, URL, the total number of activities, average, and engagement. This step was conducted according to Equation 1 to ensure that the values of all features remain within one range.

$$\text{Xnormalized} = (X - \text{Xminimum}) / (\text{Xmaximum} - \text{Xminimum}) \qquad (1)$$

where X is a numeric feature, Xminimum and Xmaximum represent the minimum and maximum value in a numeric feature X.

3.3 Building and testing the predictive model

Waikato Environment for Knowledge Analysis (WEKA) version 3.8 was used in this research. The most suitable prediction techniques were identified by using demographic and behavioral features. It was found that the highest accuracy for predicting students' final results was obtained by the ClassificationViaRegression method. Furthermore, this research used 3-folds cross-validation technique to train and test the data. Cross-validation is primarily utilized to assess model performance. This technique divides the dataset into three subsets of equal size. Two subsets are used for training, while another is used for testing. This process is iterated three times, where the final result is estimated as the average error rate on the tested examples.

3.4 Evaluation measures

This study adopts one of the most common metrics in evaluating the quality of rating which is known as the accuracy (ACC) method. It is calculated based on Equation

2 (the number of all correct predictions is divided by the total number of the dataset). The best accuracy is 1, whereas the worst is 0.

$$ACC = (TP+TN)/(TP+FP+TN+FN) \qquad (2)$$

4   Results and Discussion

This study aims at predicting students' performance in an online learning setting, particularly VLE. The key objective of this research is to minimize learners' failures and/or dropouts in online learning courses by using an early prediction and continuous alert on the expected final level of students. The prediction model was based on four time periods (after the second, fifth, sixth assessments, and immediately before the final exam). It was performed in such different periods in order to provide an indicator of students' final results if they will remain at the same academic level as they were at the prediction time. Demographic, behavioral, and assessment score features were used in the classification process. Table 2 shows the model accuracy before using the generated features and after the second, fifth, and sixth assessments as well as a day before the final exam.

Table 2. The prediction accuracy of final results by using demographic and behavioral attributes

Date of prediction	Correctly Classified Instances	Accuracy
After 2nd assessment	1227	63.3127 %
After 5th assessment	1340	69.1434 %
After 6th assessment	1562	80.5986 %
A day before the final exam	1652	85.2425 %

Based on the above-presented results, it is clear that despite the importance of demographic and behavioral attributes; they have not sufficiently predicted learners' achievement, particularly at an early period. This may indicate that there is a substantial need to create new attributes in order to enhance the overall accuracy. As such, the generated attributes were used alongside the original features. Table 3 presents the model accuracy after using the generated features in the four periods. It is clear that the accuracy was dramatically increased after using the generated attributes.

Table 3. The prediction accuracy of final results with demographic, behavioral, and the generated attributes

Date of prediction	Correctly Classified Instances	Accuracy
After 2nd assessment	1365	70.4334 %
After 5th assessment	1619	83.5397 %
After 6th assessment	1710	88.2353 %
A day before the final exam	1771	91.3829 %

This research adopts the 'InfoGainAttributeEval' method to select the best features. InfoGainAttributeEval evaluates attributes by measuring their information gain with

respect to the class [10]. Demographic and behavioral characteristics were used for the period from the beginning of the course to the 260th day which was the last day of the course. These attributes are ordered by InfoGainAttributeEval in the following order: homepage, subpage, resource, content, forum, URL, externalquize, wiki, collaborate, glossary, imd band, highest education, region, number of previous attempts, disability, gender, and age band. According to the trial and error method, the first 16 attributes were entered into the Classification via Regression technique and then the first 15, 14, 13 and 12 until we reached the required accuracy. The obtained accuracy is 85.2425% where 1652 instances were correctly classified using the first 12 attributes (homepage, subpage, resource, content, forum, URL, external quize, wiki, collaborate, glossary, imd band, and highest education).

After this stage, the generated features were used alongside behavioral and demographic factors. By applying the "InfoGainAttributeEval" technique, the order of the factors was as follows: average, engagement, total num of activities, homepage, subpage, resource, content, forum, URL, external quize, wiki, collaborate, glossary, imd band, region, highest education, num of previous attempts, disability, gender, and age band. In the trial and error method, the first 19 factors were introduced to the Classification Via Regression technique; then the first 18, 17, 16, 15 and 14 factors were entered until obtaining the best accuracy. The accuracy is 91.4345% in which 1772 instances were correctly classified using the following attributes: average, engagement, total num of activities, homepage, subpage, resource, content, forum, URL, external quize, wiki, collaborate, glossary, and imd band.

Overall, the research outcomes suggest that both demographic and behavioral features were important in predicting students' performance. The study confirmed that the most important demographic factors are the level of the students' educational attainment before enrolling in the course and the level of the financial and service stability based on students' area. This is consistent with other studies [11, 12]. It is worth mentioning that students' participation in different VLE activities had a great effect on their performance. Hence, educational institutions are invited to increase their online learning activities to ensure achieving high online learning outcomes. This should also encourage further integration of educational technologies to enhance the learning process and meet learners' individual needs [22].

After adding the generated factors to the original demographic and behavioral features, it was noted that they had an important role in decision-making as the prediction accuracy was significantly increased. The level of students' participation in the VLE activities, the total number of clicks on different activities, and the average of assessments were significant antecedents of academic achievement. The assessments' presence provides a great incentive for students and thus increases the proportion of their participation in the course activities. This supports the findings of another study [23].

Integrating the proposed factors, however, led to reducing the influence of demographic features except for the deprivation band (imd band). On the other hand, behavioral factors still had a high effect on learners' performance to be consistent with the outcomes presented in [13]. Students' engagement was also a significant predictor of academic achievement. This should encourage educational institutions to adopt

different means in order to motivate students towards learning content. This outcome supports previous literature on the positive role of engagement in different teaching and learning activities [24].

5   Conclusion

Online learning has become a common method in contemporary education. However, predicting students' performance in this learning mode is a complicated process as it is based on several different variables. This paper predicted students' final scores over four different periods in a VLE course. Demographic, scores, and behavioral features were used first. Then, three new features were generated to support the prediction process. These features were the total number of clicks on different activities, average, and engagement.

By integrating the generated features, the highest prediction accuracy obtained in the first assessment was 53 days after the beginning of the course with an accuracy of 70.4334%. The highest accuracy of the last prediction achieved a day before the final exam which was 91.3829%. The study showed that the most important demographic variable is the deprivation band (imd band), whereas all VLE behavioral variables and assessment scores were significant predictors.

In the future, it is possible to identify other factors that may affect students' achievement before assessing their performance. This can assist identifying the most important online learning activities that should be considered further in such settings. Moreover, the accuracy of the educational data mining (EDM) technique adopted in this research can be compared with other techniques to identify the most accurate one in predicting students' academic achievement.

6   References

[1] M. Zaffar, K. S. Savita, M. A. Hashmani, and S. S. H. Rizvi, "A Study of Feature Selection Algorithms for Predicting Students Academic Performance," Int. J. Adv. Comput. Sci. Appl., vol. 9, no. 5, pp. 541–549, 2018.

[2] A. Al-Azawei and H. M. Habeeb, ―Which factors affect learner satisfaction in educational hypermedia systems? A case study of the Moodle system,‖ J. Eng. Appl. Sci., vol. 12, no. Specialissue10, pp. 8864–8874, 2017..

[3] M. Hussain, W. Zhu, W. Zhang, and S. M. R. Abidi, "Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores," Hindawi, vol. 2018, pp. 1–22, 2018. https://doi.org/10.1155/2018/6347186

[4] A. M. Mogus, I. Djurdjevic, and N. Suvak, "The impact of student activity in a virtual learning environment on their final mark," Active Learning in Higher Education, vol. 13, no. 3, pp. 177–189, 2012. https://doi.org/10.1177/1469787412452985

[5] P. S. Muljana and T. Luo, "Factors contributing to student retention in online learning and recommended strategies for improvement: A systematic literature review," J. Inf. Technol. Educ. Res., vol. 18, pp. 19–57, 2019.

[6] A. Al-Azawei, "Modelling E-Learning Adoption: The Influence of Learning Style and Universal Learning Theories", Doctoral Dissertation, University of Reading, UK, 2017.

[7] A. Al-Azawei, P. Parslow, and K. Lundqvist, "The effect of Universal Design for Learning (UDL) application on e-learning acceptance: A structural equation model," Int. Rev. Res. Open Distance Learn., vol. 18, no. 6, pp. 54–87, 2017. https://doi.org/10.19173/irrodl.v18i6.2880

[8] A. M. Shahiri, W. Husain, and N. Abdul Rashid, "A Review on Predicting Student's Performance using Data Mining Techniques," Procedia - Procedia Comput. Sci., vol. 72, pp. 414–422, 2015. https://doi.org/10.1016/j.procs.2015.12.157

[9] Y. Kumar, "AI based Hybrid Ensemble Technique for Network Security," no. Icaet, pp. 1–10, 2016.

[10] E. Frank, M. A. Hall, and I. H. Witten, ―The WEKA workbench,‖ Data Min., pp. 553–571, 2017.

[11] A. Daud, N. R. Aljohani, R. A. Abbasi, M. D. Lytras, F. A. Abbas, and J. S. Alowi-bdi, ―Predicting Student Performance using Advanced Learning Analytics,‖ Int. World Wide Web Conf. Comm., pp. 415–421, 2017. https://doi.org/10.1145/3041021.3054164

[12] F. Eduardo, M. Holanda, M. Victorino, V. Borges, R. Carvalho, and G. Van Erven, ―Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil,‖ J. Bus. Res., vol. 94, no. August 2017, pp. 335–343, 2019. https://doi.org/10.1016/j.jbusres.2018.02.012

[13] R. Umer, T. Susnjak, A. Mathrani, and S. Suriadi, ―A learning analytics approach: Using online weekly student engagement data to make predictions on student performance,‖ 2018 Int. Conf. Comput. Electron. Electr. Eng. ICE Cube 2018, pp. 1–5, 2019. https://doi.org/10.1109/icecube.2018.8610959

[14] O. Sukhbaatar, T. Usagawa, and L. Choimaa, ―An Artificial Neural Network Based Early

[15] Prediction of Failure-Prone Students in Blended Learning Course,‖ Int. J. Emerg. Technol. Learn., vol. 14, no. 19, p. 77, 2019. https://doi.org/10.3991/ijet.v14i19.10366

[16] S. Jiang, A. E. Williams, K. Schenke, M. Warschauer, and D. O. Dowd, ―Predicting MOOC Performance with Week 1 Behavior,‖ Proc. 7th Int. Conf. Educ. Data Min., pp. 273–275, 2014.

[17] Kuzilek J., Hlosta M., Zdrahal Z. Open University Learning Analytics dataset Sci. Data 4:170171 doi: 10.1038/sdata.2017.171 (2017). https://doi.org/10.1038/sdata.2017.171

[18] S. Rana and R. Garg, ―Prediction of Students Performance of an Institute Using Classification ViaClustering and ClassificationViaRegression Shiwani,‖ Proc. Int. Conf. Commun. Networks, Adv. Intell. Syst. Comput., p. 333,343, 2017.

[19] B. K. Baradwaj and S. Pal, ―Mining Educational Data to Predict Students'Academic Performance,‖ Int. J. Adv. Comput. Sci. Appl., vol. 2, no. 6, pp. 63–69, 2015.

[20] A. Jović, K. Brkić, and N. Bogunović, "A review of feature selection methods with applications," 2015 38th Int. Conv. Inf. Commun. Technol. Electron. Microelectron. MIPRO 2015 - Proc., pp. 1200–1205, 2015. https://doi.org/10.1109/mipro.2015.7160458

[21] P. Tang, M. Steinbach, and V. Kumar, "Introduction to data mining", Pearson Education, Inc.,2006.

[22] A. M. Shahiri, W. Husain, and N. Abdul Rashid, ―A Review on Predicting Student ' s Performance using Data Mining Techniques,‖ Procedia - Procedia Comput. Sci., vol. 72, pp. 414–422, 2015. https://doi.org/10.1016/j.procs.2015.12.157

[23] A. Al-Azawei, "What drives successful social media in education and e-learning? A comparative study on Facebook and Moodle," J. Inf. Technol. Educ. Res., vol. 18, pp. 253–274, 2019. https://doi.org/10.28945/4360

[24] N. Holmes, "Engaging with assessment: increasing student engagement through continuous assessment," Active Learning in Higher Education, vol. 19, no. 1, pp. 23–34, 2018. https://doi.org/10.1177/1469787417723230

[25] A. Al-Azawei, W. R. Baiee, and M. A. Mohammed, "Learners' Experience Towards E-Assessment Tools: A Comparative Study on Virtual Reality and Moodle Quiz," Int. J. Emerg. Technol. Learn., vol. 14, no. 05, p. 34, 2019. https://doi.org/10.3991/ijet.v14i05.9998

7   Authors

Ahmed Al-Azawei is a senior lecturer in the Software Department / College of Information Technology, at the University of Babylon, Iraq. He undertook his PhD in Computer Science at the University of Reading, UK. His area of research focuses on learner differences, the application of e-learning, inclusive learning theories, technology adoption, Web technologies, and adaptive educational hypermedia systems.

Miami Abdul Aziz Al-Masoudy undertook her BSc in Computer Science at the University of Babylon, Iraq. She is currently an MSc student in the Software Department at the University of Babylon.