A Comparative Study of Classification Algorithms of Moodle Course Logfile using Weka Tool

Iman Al-Kindi* and Zuhoor Al-Khanjari
Sultan Qaboos University, Muscat, SULTANATE OF OMAN

Abstract

Learning Management Systems (LMSs) have been widely used in the deployment of e-learning in higher education institutions. One of the most famous LMS used is Moodle. In Moodle environment, classification has been used for several reasons, including finding students who share similar traits and forecasting student performance. Therefore, this study looks at two classification algorithms that were used on a dataset gathered from a Moodle LMS course logfile. The goal is to conduct a thorough theoretical and experimental examination of classification data mining techniques, as well as a comparison study, to determine which methodology is the best for identifying student performance with the support of their engagement, behavior, and personality during different activities of the course. The algorithms under investigation are Naive Bayes (NB) and Random Forest (RF). The performance of the classification of the two algorithms is compared using the tool Weka "Waikato Environment for Knowledge Analysis" as an open-source software package that includes data preparation, algorithm implementation and visualization tools. According to the study of the comparison results, the classification algorithms with the best accuracy is the Random Forest, with 97.36 % correctly predicted instances. In a Moodle environment, the classification techniques might be used to predict students' performance.

Key Words: Moodle, logfile, classification algorithms, student performance, student engagement.

1 Introduction

Recent developments in education technology are all driving global education reform [3]. The construction of a smart learning environment is also the foundation for altering education and learning techniques. Higher education institutions require personalized and smart learning environments with learning materials to meet the needs of their students, who have a wide range of demands [2, 5]. Learning Management Systems (LMSs) are also becoming increasingly common in universities, schools, and businesses, with individual professors using them to supplement traditional face-to-face sessions with online technology [27]. Modular Object-Oriented Dynamic Learning Environment "Moodle" is an open-source and one of the most popular Learning Management Systems (LMSs) [16]. There is a lot of interest these days in evaluating and mining Moodle interaction data for forecasting students' final grades in blended learning [29]. Moodle log data, particularly that relating to students' interactions with educational materials, may be quite interesting and useful for developing student behavior models [28]. Educational Data Mining (EDM) is a valuable technique for analyzing this data [326]. It is the process of discovering information from LMS datasets [26]. There are several data mining tools available. DBMiner and SPSS are instances of commercial mining tools, whereas Weka and Keel are examples of public domain mining tools [37].

This study expands previously published papers entitled: "Exploring Factors and Indicators for Measuring Students' Performance in Moodle Learning Environment." And "Tracking Student Performance Tool for Predicting Students EBPP in Online Courses," using the same dataset that has been used in the previously published papers in the International Journal of Emerging Technologies in Learning. The first paper used manual analysis to get the results. While the second paper used the tracking tool to get the results. Hence, this paper used different methods to analyze Moodle log file of the same dataset.

The authors distinguish some of the classification algorithms that will be chosen for analyzing a real dataset in order to have a better idea of how students will perform, engage, behave, and treat when dealing with Moodle courses. This is based on the Engagement, Behavior, Personality "EBP" predictive model that has been proposed by the authors previously. The prediction model is built on the course log files, which primarily indicate student's engagement, behavior and personality in the course. The instructor can create patterns of students from those log files to aid in the preparation of customized learning courses tailored to the needs of his/her students and therefore are meant to aid in the monitoring of the student's performance [1, 8].

As a result, in this paper, the authors examine and compare two different classification algorithms. Those algorithms are Naive Bayes (NB) and Random Forest (RF). These algorithms are compared in terms of their time taken to build the model, accuracy, correctly classified and incorrectly classified instances. The key target of this study is the detailed performance analysis of the two classification algorithms selected in the Weka tool (version 3.8.5) and to do a comparison among them, alongside, to assist the instructors in how to

* Email: ir.alkindi@gmail.com.

develop better learning strategies for their students based on their requirements.

The paper is organized as follows: “Background” section discusses the main concepts covered in this study. The section “Review of Relevant Literature” goes over the relevant literature for this study. The data and experiments are described in the section “Method.” The section "Results" contains and examines the steps to apply the selected algorithms in this study. The “Comparison” section highlights the main comparison results between NB and RF algorithms. Also, the section “Discussion” summarizes the main findings. Finally, section “Conclusions” concludes this study.

2 Background

2.1 Moodle Logfile

At Sultan Qaboos University (SQU), the teaching style is largely focused on a blended learning strategy that incorporates both online and traditional learning. Meanwhile, whenever academic year data is available, the student's engagement, behavior, personality, and performance are critical. Students’ successes in completing course assignments, as well as class participation and engagement, are important variables in improving the learning environment [8]. Moodle’s environment not only allows students convenient access to educational resources but also allows higher education institutions to collect massive amounts of data on student activity. This data is useful for assessing student behavior and determining whether there are any trends that contribute to enhanced learning outcomes [4]. Moodle logs maintain tracking of what materials students have accessed, updated, produced, and removed, as well as every click student and teacher make while navigating the site [19]. Table 1 shows an example of logfile in Moodle which consists of nine dimensions; “Time, User Full Name, Affected User, Event Context, Element, Event Name, Description, Origin and IP address”.

2.2 Student Engagement

Engaging in an activity is described as active, deliberate, and prolonged action and it is a hallmark of students’ genuine engagement with academic activities [38]. Moreover, Student engagement with activities of the course may be used to diagnose other crucial motivational processes that are not always obvious [21]. Experts are aware of the value of engagement in learning. Furthermore, empirical research has frequently shown the correlation between engagement and students’ performance [20].

2.3 Student Behavior

Studies of real online student behavior are necessary to discover the behaviors that contribute to online persistence and accomplishment for students and educators [32]. More and more, “student behaviors are significantly associated with many desirable academic and personal development outcomes of college,” as stated by the National Survey of Student Engagement. Standard college experiences for instance involve in collaborative learning activities and the number of hours spent on homework per week are among these behaviors [9]. Further, Zaiane stressed the need for distinguishing between various students' online activities [40]. Cantabella and colleagues advocated using big data technology to try to understand student behavior and draw conclusions about how to improve students’ performance by enhancing their learning process [12].

2.4 Student Personality

“A consistent predictor of student satisfaction, academic motivation, and academic performance” is personality [10]. Personality refers to individual differences in thought, mood, and behavior patterns [13]. McGeown and colleagues identified personality as “A set of fundamental qualities that influence

Table 1: An example of log file of the course in Moodle

Time	Full Name	Affected User	Event Context	Element	Event Name	Description	Origin
20/05/1916:00	SH	-	Forum: Alerts and Circulars	Forum	The course was reviewed	The user with id '13966' viewed the 'forum' activity with course module id '7453'.	web
12/05/1913:50	S H	-	Course: Internet search strategies	System	The course was reviewed	The user with id '13966' viewed the section number '-1' of the course with id '7453'.	web
4/03/1913:33	TA	-	Course: Internet search strategies	System	The course was reviewed	The user with id '26716' viewed the course with id '7453'.	web
19/02/1900:20	T L	-	Course: Internet search strategies	System	The course was reviewed	The user with id '26716' viewed the section number '-1' of the course with id '7453'.	web
……….	……….	……….	……….	……….	……….	……….	……….

how a person generally acts, thinks, and feels" [31].

2.5 Student Performance

Student performance means that the consequences of the teaching and learning process in terms of knowledge and skills obtained by students from schools and colleges are assessed by exams scores [30]. The quality of interpersonal contact in a course, according to Jaggars, has a favorable and meaningful relationship with student performance [22].

2.6 Classification Algorithms

Because data is not always in the best shape for analysis, new techniques for data analysis are needed to turn it into knowledge and information [33]. The authors are going to use classification algorithms, because it has been widely utilized in educational data mining and, also with the aim of predicting the student's performance along with their engagement, behavior and personality. One of the most common research challenges by machine learning researchers is classification. Predicting the value of the class attribute is based on the values of other attributes [36]. When the expected variable is binary or categorical, classification is employed [25]. The following machine learning algorithms were used to classify the students into three categories: Random Forest classifier (RF), Naive Bayes (NB), Logistic Regression (LR) and k-Nearest Neighbors (kNN) [34]. The authors focus in this paper on NB and RF algorithms.

2.7 Weka

In this paper the authors would like to use Weka (Waikato Environment for Knowledge Analysis), which is a set of machine learning algorithms for data mining. It is one of the most frequently used methods for finding knowledge from databases of data [35]. It consists of data preparation, classification, clustering, association rules and visualization tools [39] and built at the University of Waikato, New Zealand in 1997. It is applied in a wide range of applications, including educational, scientific and research purposes.

3 Literature Review

Methods, techniques, and the process of finding information through using Weka from logfile in Moodle LMS are discussed in this literature review as shown in Table 2.

Considering all existing work in Table 2 and based on authors' knowledge, the study described in this paper is the first to analyze student performance, engagement, behavior, and personality simultaneously using data from Moodle course logfiles with support of Weka classification algorithms.

Table 2: Literature review of using Weka to analyze Moodle log file

Author	Objective of the Study	Method	Dataset	Results	Future work/Limitations
[24]	Three classification data mining approaches for the detection of information presentation dimension (visual/verbal) learning style were compared using the Felder-Silverman Learning Style Model and the behavior of students in the Moodle course.	Using Weka an open-source software provides tools for implementation of several algorithms such as J48, Naive Bayes and PART.	Questionnaire and log data of Moodle	The best accuracy was achieved by the Naive Bayes algorithm, which achieved 71.18 % accuracy.	The authors want to incorporate Weka into Moodle so that they may update students' learning styles and adjust Moodle material using the data mining approach discovered in this study on Moodle log data.
[17]	This study used historical data from Moodle logs to preprocess and create machine learning models using Weka to track student performance and reduce the failure rate. Predictor qualities relating to student study behavior, such as Course Viewing Time, Quiz Taken, and so on, were used in this study.	Using Weka of J48, Random Forest, JRip, and OneR algorithms	Moodle Log of five courses	Students' performance was found to be significantly associated with predictive variables such as Activities Completed, Course Views, and Assignment Passed.	Other data mining tools and platforms are recommended to be used for comparative examination of the predictive analytics framework.
[23]	The purpose was to learn more about instructors' behavior and to create clusters based on the activities they did on the platform. The objective of this study is to boost the teaching process by devising particular approaches that will help students achieve greater success.	Weka and Hadoop	Moodle Log	Extracted knowledge from the activities of teachers	New found knowledge was used to enhance the teaching process and develop new instructional approaches in the future.

[18]	Educational Data Mining was applied to Moodle logs in order to see if the level of participation indicated in the amount of time spent using LMS services improved students' academic achievement.	Using machine learning algorithms of clustering and classification from WEKA system tools	Moodle Log	Findings showed that there is a significant relationship between the usage of Moodle resources and students' academic achievement. The findings are beneficial for strategic academic planning at the institution using LMS data.	Suggested that this study could be done with a variety of courses from other disciplines to see if academic disciplines had an impact on students' performance
[14]	Evaluating student prediction performance in Moodle and MOOCs based on their involvement with eLearning activities	Decision Tree, Artificial Neural Network, Support Vector Machine and KNN algorithms using Weka	Moodle log	The rate of interaction with the E-learning environment has a major influence on their performance, according to the study of log files, as students with the highest interactivity on the Moodle tend to do better than those with a low interactivity rate. Also, students spend more time on E-learning Moodle than MOOCs.	Not mention

4 Method

4.1 Materials and Dataset

The authors collected data from Moodle LMS log files for course "Search Strategies on the Internet" of 38 students (Table 1), offered by Department of Information Studies, College of Arts and Social Sciences, Sultan Qaboos University. Students' ability to discover information using multiple techniques such as search engines; library catalogs and online databases was improved via this course. The data taken contained 241896 logs.

4.2 Data Preprocessing

Firstly, the authors downloaded and extracted the Moodle logfile of the course. Secondly, feature selection and data filtering are all chosen during the preprocessing process. This study was carried out utilizing feature selection from the eight attributes listed in Table 3, the authors used Microsoft Excel for performing the preprocessing process. Then the feature selection was made into eight attributes that can be found in Tables 3 and 4, respectively. The next step is to run an algorithm test. The findings are documented and presented in tables and graphs in the next section's results.

The engagement, behavior, personality, and performance traits in numbers for each student are transformed into three categories by dividing the number using the percentile approach (High, Average and Low). It operates by dividing the data into irregular intervals, each pointing to a different category [5], as seen below in Table 4.

Table 5 shows the final feature selection. This file as CVS format will be imported later in Weka tool to compare the NB and RF.

Table 3: Attributes generated from data summarization of Moodle logfile

Attributes	Technical Definition	Type of Data
Engagement	A factor reflects the level of students' interactions with the activities of course in Moodle as getting exams, submit an assignment, etc.	Represent the data in numbers out of 100%
Engagement Category		Shows data in three categories (Low, Average and High)
Behavior	A factor gives the percentage of components that the student interacted with.	Represent the data in numbers out of 100%
Behavior Category		Shows data in three categories (Low, Average and High)
Personality	A factor represents the count of the accessed elements by the student.	Represent the data in numbers out of 100%
Personality Category		Shows data in three categories (Low, Average and High)
Performance	A factor indicates all the marks of students during the course.	Represent the data in numbers out of 100%
Performance Category		Shows data in three categories (Low, Average and High)

Table 4: Range with category division

Range	Category
0.00 - 35	Low
35.1 - 75	Average
75.1- 100.0	High

5 Results

Using the explorer application from Weka interface of Weka, explorer application is one of five available applications that has been used in this study. The authors applied Naive Bayes Algorithm and Random Forest Algorithm as displayed in Figures1 and 2 below.

Table 5: Nine attributes of the study after pre-processing [5]

Student ID	Engagement	Engagement Category	Behavior	Behavior Category	Personality	Personality Category	Performance	Performance Category
ST1	43.27	Average	64.7	Average	61.52	Average	69.5	Average
ST2	89.35	High	82.4	High	84.59	High	85.25	High
ST3	29.13	Low	70.6	Average	84.59	High	52.75	Average
ST4	61.24	Average	70.6	Average	76.9	High	86	High
ST5	62.44	Average	76.5	High	84.59	High	86.25	High
ST6	32.28	Low	70.6	Average	76.9	High	61	Average
ST7	67.46	Average	76.5	High	84.59	High	65.25	Average
ST8	66.78	Average	58.8	Average	69.21	Average	96.75	High
ST9	66.27	Average	64.7	Average	69.21	Average	88.75	High
ST10	68.82	Average	58.8	Average	69.21	Average	75.75	High
ST11	61.50	Average	64.7	Average	69.21	Average	76.25	High
ST12	53.32	Average	82.4	High	84.59	High	74.5	Average
ST13	52.39	Average	64.7	Average	69.21	Average	72.75	Average
ST14	100.00	High	76.5	High	76.9	High	83.5	High
ST15	58.52	Average	82.4	High	92.28	High	85	High
ST16	38.84	Average	64.7	Average	61.52	Average	56.75	Average
ST17	58.26	Average	88.2	High	99.97	High	86	High
ST18	26.75	Low	64.7	Average	61.52	Average	27.75	Low
ST19	79.81	High	82.4	High	76.9	High	85	High
ST20	92.50	High	70.6	Average	76.9	High	89.25	High
ST21	92.08	High	76.5	High	84.59	High	81	High
ST22	88.93	High	82.4	High	84.59	High	83	High
ST23	37.39	Average	58.8	Average	61.52	Average	68.25	Average
ST24	71.64	Average	76.5	High	84.59	High	87	High
ST25	45.74	Average	76.5	High	84.59	High	74.5	Average
ST26	46.17	Average	64.7	Average	69.21	Average	67.25	Average
ST27	58.94	Average	82.4	High	92.28	High	51.75	Average
ST28	55.11	Average	88.2	High	92.28	High	76.75	High
ST29	71.04	Average	70.6	Average	76.9	High	77.25	High
ST30	83.30	High	82.4	High	92.28	High	96.25	High
ST31	70.78	Average	76.5	High	84.59	High	88	High
ST32	53.92	Average	70.6	Average	76.9	High	75	Average
ST33	90.03	High	82.4	High	92.28	High	98.75	High
ST34	65.50	Average	82.4	High	99.97	High	97	High
ST35	48.98	Average	76.5	High	92.28	High	73.5	Average

ST36	55.20	Average	76.5	High	92.28	High	89.5	High
ST37	58.26	Average	70.6	Average	76.9	High	89.25	High
ST38	43.10	Average	70.6	Average	84.59	High	58	Average

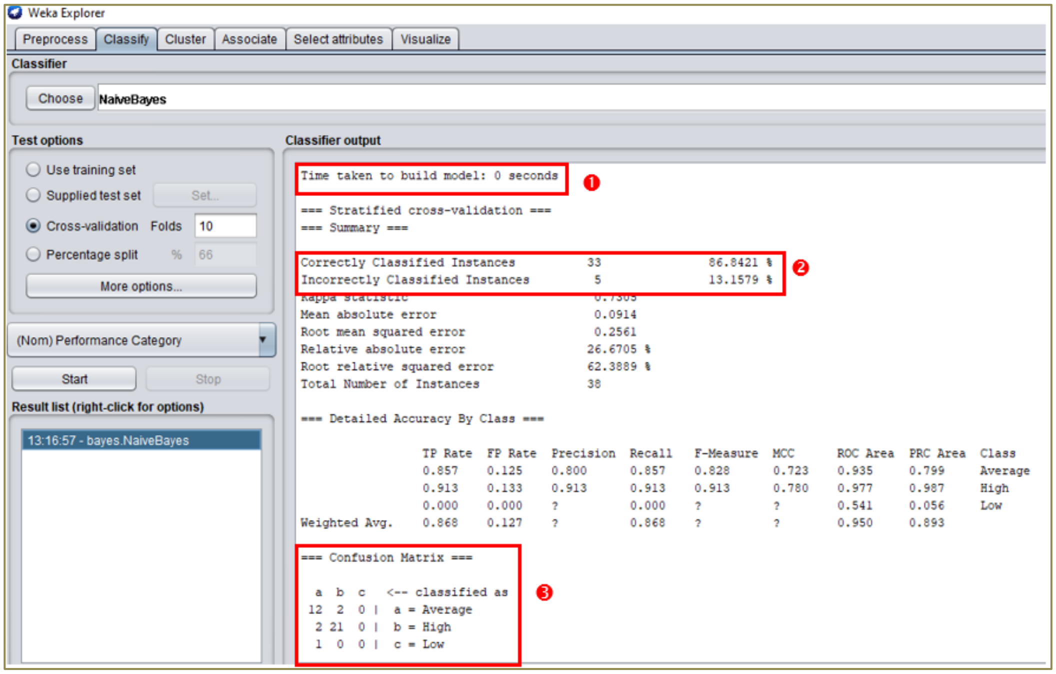

Figure 1: The results of conducting Naïve Bayes algorithm on the dataset

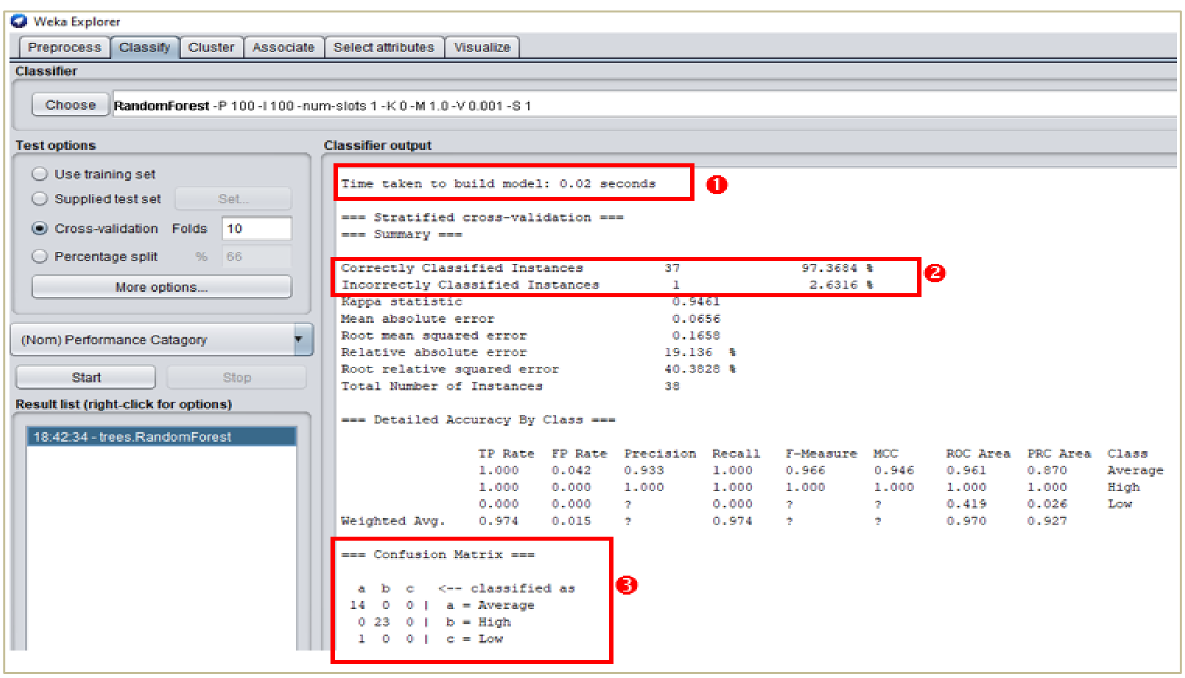

Figure 2: The results of conducting random forest algorithm on dataset

In general, accuracy indicates how often the variable is correct [11]. In Weka, each class has an alphabetical letter as a,b,c. In our experiment, Weka gives “a to Average,” “b to High,” and “c to Low.”

6 Comparison

The previous section discussed using the Weka tool to examine each of the two classification techniques on the “Search Strategies on the Internet” course dataset, which consists of eight attributes and contains data of 38 students. The dataset was classified using two classification algorithms included in the Weka tool: Naive Bayes (NB) and Random Forest (RF). Finally, the findings are as follows:

- **In terms of time taken to build the model**: NB shows 0 second. While RF shows 0.02 second.
- **In terms of accuracy**: The accuracy of NB algorithm is 86.84% compared to the accuracy that we got when we applied RF which is 97.36.%.
- **In terms of confusion matrix**: Using the Confusion Matrix, Naive Bayes classified 12 students correctly as Average and 21 as High which are 33. While 2 students have an average performance and are classified incorrectly as High. In addition, 2 students have a high performance and are classified incorrectly as Average. Lastly, the only student with low performance is classified incorrectly as Average. On the other side, Using the Confusion Matrix, Random Forest classified 14 students correctly as Average and 23 as High which are 37. Lastly, only 1 student also with low performance is classified incorrectly as Average.
- **In terms of correctly classified and incorrectly classified instances**: NB shows 33 correctly classified students from 38 while incorrectly classified 5 students. While RF shows 37 correctly classified students from 38 although incorrectly classified 1 student.

Overall, as referred to Table 6, RF performs the best compared to NB.

7 Discussion

As illustrated in Figure 3, the following features of a comparative analysis of two classification algorithms available in the Weka tools across the selected dataset as a case study in this paper have been examined: Time taken to build the model, accuracy, correctly classified and incorrectly classified instances. As a result of this investigation, the following conclusion may be drawn. We can analyze that the RF algorithm achieves accuracy greater than the NB algorithm. On the other hand, the NB is faster than RF.

Overall, the RF is better than NB because the accuracy is more important than time, especially the difference between the time of two algorithms is very small. In addition, the time of RF is due to the process of building many decision trees to select the best one of them. Based on the training instances and testing, each of them uses the same method which is 10-fold. This method trains all the instances and then tests the algorithm with the same data. This makes the performance more accurate and eliminates any overfitting of results. In addition, RF algorithm was able to correctly classify instances that reach 37 instances out of 38 instances which is the number of all students included in the dataset. In contrast, NB algorithm achieves 33 instances classified correctly. From the result, we see that time to build the RF model is more than using NB and correctly classifying instances are more and prediction accuracy is also greater in RF than the other. Hence it is concluded that RF performed better. To summarize, data mining algorithms are extremely beneficial for examining logfile data within the Moodle LMS to assess students' progress during any course.

8 Conclusions

There is a crucial need to monitor student engagement, behavior and personality in online courses and knowing how to respond to it, which is shown to improve student performance [1]. In such online courses, it is critical for instructors to be able to grasp the needs of each student [5]. Researchers can use good analytic techniques to intelligently examine students’ logfiles in educational systems [15]. Nevertheless, knowledge of the dataset and the type of analysis needed, all work to gather in selecting the best algorithm [33]. With the addition of new knowledge in E-learning, online courses, and the educational environment, such an analysis can allow more academics to dig deeper into the Moodle log file using any approach [6]. It allows for the generation of new data on a student's activity based on their digital profile [7]. So, using the log files of students in a Moodle course, the authors examined two classification data mining methods for forecasting student performance in this paper. The data used were gathered from Moodle log file of 38 students. The algorithms that were used in this study were: NB and RF. The Random Forest algorithm, which obtained 97.36%, had the best accuracy.

Table 6: Comparison of NB and RF algorithms results using the Weka tool

Algorithms	Instances	Attributes	Time Taken to build Model	Accuracy	Correctly Classified Instances	Incorrectly Classified Instances
Naive Bayes (NB)	38	8	0	86.84%	33	5
Random Forest (RF)	38	8	0.02	97.36%.	37	1

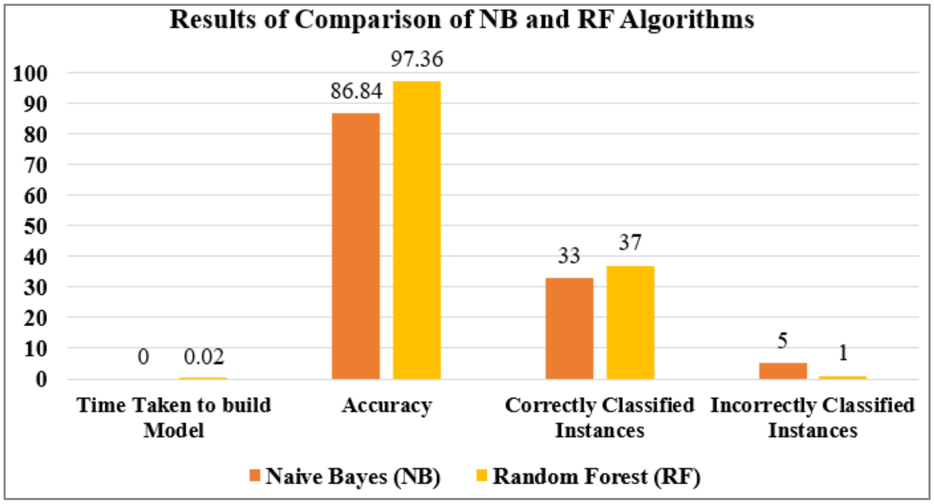

Figure 3: Results of comparison of NB and RF algorithms

Acknowledgement

The authors wish to thank Sultan Qaboos University, College of Science, and the Department of Computer Science. This work is under Prof. Zuhoor Al-Khanjari's supervision supported as a part of a scholarship of Doctoral Program from the Sultan Qaboos University.

References

[1] Z. Al-Khanjari and I. Al-Kindi, "Proposing a Systematic Framework for SQU-Smart Learning Management System (SQU-SLMS)", *International Journal of Computing and Digital Systems*. 10:1-13, https://doi.org /10.12785/ijcds/100169, 2021.

[2] I. Al-Kindi and Z. Al-Khanjari, "The Smart Learning Management System (SLMS)," *4th Free & Open-Source Software Conference* (FOSSC'2019-OMAN), ISSN: 1813-419X, https://fossc.om/2019/wp-content/uploads/ 2019/03/Proceedings.pdf#page=32, 2019.

[3] I. Al-Kindi and Z. Al-Khanjari, "A Novel Architecture of SQU SMART LMS: The New Horizon for SMART City in Oman," *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, IEEE, pp. 751-756, https://doi.org/10.1109/ICSSIT48917.2020. 9214141, 2020.

[4] I. Al-Kindi and Z. Al-Khanjari, "Proposing the EBP Smart Predictive Model Towards Smart Learning Environment," *Journal of Talent Development and Excellence*, 12(2s):2422-2438, https://iratde.com/index.php/jtde/ article/view/959, 2020.

[5] I Al-Kindi and Z. Al-Khanjari, "Exploring Factors and Indicators for Measuring Students' Performance in Moodle Learning Environment," *International Journal of Emerging Technologies in Learning*, 16(12):169-184, https://doi.org/10.3991/ijet.v16i12.22049, 2021.

[6] I. Al-Kindi and Z. Al-Khanjari, "NVivo to Analyze the Definition of Student EBP Qualitatively," *International Conference on IOT Based Control Networks and Intelligent Systems* (ICICNIS 2020), Available at SSRN: https://ssrn.com/abstract=3768217, http://dx.doi.org/ 10.2139/ssrn.3768217, 2021.

[7] I. Al-Kindi and Z. Al-Khanjari, "Tracking Student Performance Tool for Predicting Students EBPP in Online Courses," *International Journal of Emerging Technologies in Learning*. 16(23):140-156, https://doi.org/10.3991/ijet.v16i23.25503, 2021.

[8] I. Al-Kindi, Z. Al-Khanjari, and J. Al-Salmi, "Managing the Triangular Bond of the EBP for SQU Students Through the Proposed Test Model," *International Journal of Engineering and Advanced Technology (IJEAT)*, ISSN: 2249-8958, 10(1):391-402, October 2020, https://doi.org/10.35940/ijeat. A1914.1010120, 2020.

[9] R. D. Axelson and A. Flick, "Defining Student Engagement, Change: The Magazine of Higher Learning," 43(1):8-43, https://doi.org/10.1080/ 00091383. 2011.533096, 2010.

[10] K. K. Bhagat, L. Y. Wu, and C. Y. Chang, "The impact of personality on students' perceptions towards online learning," *Australasian Journal of Educational Technology*, 35(4), https://doi.org/10.14742/ajet 4162, 2019.

[11] Jason Brownlee, "A Gentle Introduction to K-Fold Cross-Validation," Retrieved from: https://machine learningmastery.com/k-fold-cross-validation/. Retrieved on: 18/8/2021, 2020.

[12] M. Cantabella, R. Martínez-España, B. Ayuso, J. A. Yáñez, and A. Muñoz, "Analysis of Student Behavior in Learning Management Systems Through a Big Data Framework," *Future Generation Computer Systems*. 90:262-272, https://doi.org/10.1016/j.future.2018.08.003, 2019.

[13] S. Caswell, J. P. Ambegaonkar, and A. M. Caswell, "Examination of Personality Traits in Athletic Training

Students," *International Journal of Athletic Therapy and Training*, 15(6):37-40, https://doi.org/10.1123/att.15.6.37, 2010.

[14] R. Chweya, S. M. Shamsuddin, S. S. M Ajibade, and S. Moveh, "A Literature Review of Student Performance Prediction in E-Learning Environment," *Journal of Science, Engineering, Technology and Management*, ISSN: 2795-3955, 2(1):1441-1453, Retrieved from https://www.jsetm.org/index.php/jsetm/article/ view/23, 2020.

[15] Data School, "Simple Guide to Confusion Matrix Terminology," Retrieved from: https://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/, Retrieved on: 18/8/2021, 2014.

[16] M. Dougiamas and P. Taylor, "Moodle: Using Learning Communities to Create an Open-source Course Management System," *Proceedings of ED-MEDIA 2003--World Conference on Educational Multimedia, Hypermedia & Telecommunications*, pp. 171-178, Honolulu, Hawaii, USA: Association for the Advancement of Computing in Education (AACE). https://www.learntechlib.org/primary/p/13739/, 2003.

[17] E. D. Evangelista, "Development of Machine Learning Models using Study Behavior Predictors of Students' Academic Performance through Moodle," *International Journal of Innovative Technology and Exploring Engineering* (IJITEE), 8:22-27, https://www.ijitee.org/wpcontent/uploads/papers/v8i6s3/F10050486S319.pdf, 2019.

[18] U. O. Ezekiel and M. Mogorosi, "Educational Data Mining for Monitoring and Improving Academic Performance at University Levels," (IJACSA) *International Journal of Advanced Computer Science and Applications*. 11(11):570-581, https://doi.org/10.14569/IJACSA.2020.0111171, 2020.

[19] I. M. Félix, A. P. Ambrósio, P. S. Neves, J. Siqueira, and J. D. Brancher, "Moodle Predicta: A Data Mining Tool for Student Follow Up," *Proceedings of the 9th International Conference on Computer Supported Education (CSEDU 2017),* 1:339-346, https://doi.org/10.5220/0006318403390346, 2017.

[20] J. D. Finn and K. S. Zimmer, "Student Engagement: What is it? Why Does it Matter?" *Handbook of Research on Student Engagement,* Springer, Boston, MA, pp. 97-131, https://doi.org/10.1007/978-1-4614-2018-7_5, 2012.

[21] C. Furrer, G. Kelly, and E. Skinner, "Can Teachers use Children's Emotions in the Classroom to Diagnose and Treat Underlying Motivational Problems," *Poster Presented at the Biennial Meetings of the Society for Research in Child Development*, Tampa, FL, https://doi.org/10.1037/a0012840, 2003.

[22] S. Jaggars and D. Xu, "Predicting Online Student Outcomes from a Measure of Course Quality," *Journal of Research in Personality*, 35:78-90, https://doi.org/10.7916/D8N29TZH, 2013.

[23] L. Janevski, A. Velinov, and Z. Zdravev, "Analyzing Teachers Behavior Using Moodle Data and Big Data Tools," *Balkan Journal of Applied Mathematics and Informatics*, 2(1): 39-44. https://eprints.ugd.edu.mk/id/eprint/22199, 2019.

[24] A. Kika, L. Leka, S. Maxhelaku, and A. Ktona, "Using Data Mining Techniques on Moodle Data for Classification of Student? S Learning Styles," *Proceedings of International Academic Conferences* (No. 9211567), International Institute of Social and Economic Sciences, https://doi.org/10.20472/IAC.2019.047.010, 2019.

[25] R. G. Leppan, *A Model for Using Learners' Online Behavior to Inform Differentiated Instructional Design in MOODLE*, PhD Thesis, Faculty of Engineering, the Built Environment, and Information Technology, Nelson Mandela University, https://core.ac.uk/download/pdf/270043831.pdf, 2019.

[26] M. P. P. Liyanage, L. G. KS, and M. Hirakawa, "Detecting Learning Styles in Learning Management Systems using Data Mining," *Journal of Information Processing*, 24(4):740-749, https://doi.org/10.2197/ipsjjip.24.740, (2016).

[27] J. López-Zambrano, J. A. Lara, and C. Romero, "Towards Portability of Models for Predicting Students' Final Performance in University Courses Starting from Moodle Logs," *Applied Sciences*, 10(1):354, https://doi.org/10.3390/app10010354, 2020.

[28] J. López-Zambrano, J. A. Lara, and C. Romero, "Improving the Portability of Predicting Students' Performance Models by using Ontologies," *Journal of Computing in Higher Education*, pp. 1-19, https://doi.org/10.1007/s12528-021-09273-3, 2021.

[29] J. M. Luna, C. Castro, and C. Romero, "MDM Tool: A Data Mining Framework Integrated into Moodle," *Computer Applications in Engineering Education*, 25(1):90-102, https://doi.org/10.1002/cae.21782, 2017.

[30] J. H. Maganga, "Factors Affecting Students' Academic Performance: A Case of Public Secondary Schools in Ilala District," *Dar es Salaam (Doctoral Dissertation, The Open University of Tanzania),* http://repository.out.ac.tz/1732/1/JAMILLAH__MAGANGA-Dissertation_14-10-2016-Final.pdf, 2016.

[31] S. P. McGeown, D. Putwain, E. G. Simpson, E. Boffey, J. Markham, and A. Vince, "Predictors of Adolescents' Academic Motivation: Personality, Self-efficacy and Adolescents' Characteristics," *Learning and Individual Differences*. 32:278-286, https://doi.org/10.1016/j.lindif.2014.03.022, 2014.

[32] L. V. Morris, C. Finnegan, and S. S. Wu, "Tracking student behavior, persistence, and achievement in online courses*". The Internet and Higher Education*, 8(3):221-231, https://doi.org/10.1016/j.iheduc.2005.06.009, 2005.

[33] C. Pradana, S. S. Kusumawardani, and A. E. Permanasari, "Comparison Clustering Performance Based on Moodle Log Mining," *IOP Conference Series:*

Materials Science and Engineering, 722(1):012012), IOP Publishing. https://doi.org/ 10.1088/1757-899X/722/1/012012, 2020.

[34] G. S. Ramaswami, T. Susnjak, and A. Mathrani, "Capitalizing on Learning Analytics Dashboard for Maximizing Student Outcomes," *2019 IEEE Asia-Pacific Conference on Computer Science and Data Engineering* (CSDE), pp. 1-6, IEEE. https://doi.org/10.1109/CSDE48274.2019.9162357, 2019.

[35] B. Rawat and S. K. Dwivedi, "An Architecture for Recommendation of Courses in E-learning System" *Int. J. Inf. Technol. Comput. Sci*. (IJITCS), 9(4):39-47, https://doi.org/10.5815/ijitcs.2017.04.06, 2017.

[36] C. Romero, S. Ventura, P. G. Espejo, and C. Hervás, "Data Mining Algorithms to Classify Students," *Educational Data Mining 2008*, https://www.researchgate.net/profile/Michel-Desmarais-2/publication/221570516_Adaptive_Test_Design_with_a_Naive_Bayes_Framework/links/09e41512380f4e4bc1000000/Adaptive-Test-Design-with-a-Naive-Bayes-Framework.pdf#page=8, (2008).

[37] C. Romero, S. Ventura, and E. García, "Data Mining in Course Management Systems: Moodle Case Study and Tutorial," *Computers & Education*, 51(1):368-384, https://doi.org/10.1016/j.compedu.2007.05.016, 2007.

[38] E. A. Skinner and J. R. Pitzer, "Developmental Dynamics of Student Engagement, Coping, and Everyday Resilience," *Handbook of Research on Student Engagement*, Springer, Boston, MA, pp. 21-44, https://doi.org/10.1007/978-1-4614-2018-7_2, 2012.

[39] "Weka 3: Machine Learning Software in Java," Retrieved from: https://www.cs.waikato.ac.nz/ml/weka/, Retrieved on: 22/8/2021

[40] O. Zaiane, "Web Usage Mining for a Better Web-based Learning Environment," Available online at https://era.library.ualberta.ca/items/0a182195-ce39-4b5d-a1c1-291ed91a0f36/view/336c2a34-b149-4d9d-95a1-ad78c08ee35c/TR01-05.pdf, 2001.

Iman Al-Kindi is a PhD graduate student from the Department of Computer Science, College of Science at Sultan Qaboos University, Sultanate of Oman. She received her BSc in software Engineering from Higher College of Technology, Sultanate of Oman, and MSc in Computer Science from Sultan Qaboos University, Sultanate of Oman. She has worked as a visiting lecturer for more than one year at Sultan Qaboos University, Sultanate of Oman.

Zuhoor Al-Khanjari is a Professor in Software Engineering. She worked as the HOD of the Department of Computer Science, College of Science at Sultan Qaboos University, Sultanate of Oman. She received her BSc in mathematics and computing from Sultan Qaboos University, Sultanate of Oman, MSc and PhD in computer science (software engineering) from the University of Liverpool, UK. Her research interests include software engineering, software testing techniques, database management, e-learning, m-learning and mobile computing. Currently, she is the coordinator of the software engineering group in the Department of Computer Science, Sultan Qaboos University, Sultanate of Oman. Also, she is coordinating e-learning facilities in the same department. She is a member of the editorial board of the International Arab Journal of Information Technology (IAJIT) and a member of the executive committee of the International Arab Conference on Information Technology (ACIT).