ISSN 2278-3091
Volume 9, No.3, May - June 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse169932020.pdf
https://doi.org/10.30534/ijatcse/2020/169932020

The Use of Schoology as Learning Management System in the College of Computing Education: A Response Assessment using Data Mining Techniques

Allemar Jhone P. Delima[1], Jan Carlo T. Arroyo[2], Michelle C. Elape[3], Markdy Y. Orong[4]
[1,2]College of Computing Education, University of Mindanao, Davao City, Davao del Sur, Philippines
[1]College of Engineering, Technology and Management, Cebu Technological University-Barili Campus, Cebu, Philippines
[3]Computer Education Department, ACLC College of Butuan, Butuan City, Agusan del Norte, Philippines
[4]College of Computer Studies, Misamis University, Ozamiz City, Philippines
allemardelima@umindanao.edu.ph [1], jancarlo_arroyo@umindanao.edu.ph[2], michellecomillaselape@gmail.com[3], markdy.orong@mu.edu.ph[4]

ABSTRACT

This paper evaluates the accuracy of the students' responses in the actual evaluation conducted in assessing student satisfaction in the use of the Learning Management System (LMS) in the College of Computing Education of the University of Mindanao in Davao City, Philippines. The use of data mining algorithms, namely the C4.5, K-Nearest Neighbor (KNN), and Naïve Bayes algorithms in the prediction of LMS assessment dataset consisting of 257 instances and 24 variables, performed using the 10-folds cross-validation scheme in WEKA software, were undertaken. Simulation results revealed that the optimal model used for prediction is the Naïve Bayes algorithm with 100% prediction accuracy. The high percentage accuracy denotes that the knowledge generated from the actual study is worthy of implementation.

Key words: C4.5 algorithm, data mining, HEI, K-Nearest Neighbor algorithm, Naïve Bayes algorithm, prediction

1. INTRODUCTION

A nation's growth relies on the education of its citizens as education is regarded as the bedrock for the development of every nation. An ever-changing mode of education has been prevalent for over centuries as challenges arise ranging from changes in school curriculum to closing down of schools due to protest by either staff and students, instability of government, or outbreak of pandemic of which, most countries have closed down schools today due to coronavirus disease 2019 (Covid-19) where the Philippines is of no exception [1].

However, the use of online technologies established a potential academic paradigm shift that transitions classroom learning to distance learning platform to pursue education despite the abovementioned challenges [2]. The use of learning management systems enables the anytime anywhere education concept with platforms developed according to the needs of learners. The use of learning management systems bridge learning gaps constrained either by time, distance, or both [3].

The College of Computing Education (CCE) of the University of Mindanao in Davao City, Philippines, uses a blended learning approach in delivering quality education. A study of [4] assesses the use of Schoology as the LMS employed in the college. A 24-item survey questionnaire was deployed where responses obtained were used to generate the result of the study.

However, the use of qualitative and quantitative methods with old predefined queries and charts in conducting research often unlikely to extract well-established knowledge from the dataset. With the advent of technology, a new paradigm called data mining [5] has taken the spotlight in the conduct of researches within the educational context.

Data mining (DM), coined from the term knowledge discovery from databases (KDD), is a technique used in extracting information from databases by using mathematical and machine learning techniques and algorithms [6]–[11]. Some of the notable approaches in data mining include prediction, clustering, association, and classification, among others [12]. Among these, the prediction is identified as the commonly used method in health, business, and educational data mining researches, among others [9].

Therefore, this paper evaluates the responses used in the study of [4] and predicts the accuracy of the dataset obtained from the actual survey they conducted. To realize, the use of C4.5, KNN, and Naïve Bayes algorithms are observed. Further, this study aims to identify the optimal algorithm needed for the prediction of the accuracy of the LMS acceptability assessment dataset. The study is hoped to contribute to the two major literatures; (1) higher education mining, and (2) data mining algorithms.

2. RELATED LITERATURE

The study of [13] used predictive learning analytics to develop a model that measures students' success based on student's behaviors in using a learning management system. In the study, five essential features were classified to examine how these relate to the learning performance of a student. These features include the number of times the learners asked questions, downloaded course materials, viewed announcements, participated in discussions, and their absences. Dataset used was from an online repository that contained 480 learner data spanning two semesters. After data pre-processing, a prediction model was developed using the C4.5 algorithm. A 10-fold cross-validation method was used and attained a prediction accuracy of 68.54%. It has been concluded that downloading course materials is the most important feature that affects students' performance.

A behavior-based students' performance prediction model using Naïve Bayes was presented in the paper of [14]. The dataset used in the study contained a sample of 480 students, wherein sixteen 16 variables were identified and used to develop the prediction model based on the background of the students and their behaviors in using a learning management system. The model was assessed for its accuracy using a 10-fold cross-validation technique. Naïve Bayes returned an accuracy of 67.7082% accuracy with an error rate of 32.2917%.

In the study [15], a prediction model was developed to identify students' personality and dominant preference based on the Myers-Briggs Type Indicator theory. Through this, students are made aware of how their personality affects their study habits, and this also allows educators to further understand their students better. The model was built upon the dataset collected from a learning management system and a social network, wherein ten attributes from 240 instances were classified. The study used ten classification algorithms to wit: RandomTree, OneR, KNN/IBK, J48(C4.5), NaiveBayes, BayesNet, Random forest, Kstar, JRIP, and Decision Table and determined the model with the highest accuracy. Findings show that the OneR algorithm obtained the highest accuracy of 97.40%, followed by random forest with 93.23% accuracy and C4.5 algorithm with 92.19% prediction accuracy.

Another study confirmed whether learning management system activities and prior knowledge through national exam scores are valid predictors for passing programming courses. There were 153 data instances gathered from two semesters within a span of two school years, wherein ten attributes were classified using C4.5, JRip, and PART algorithms. The use of 10-folds cross-validation scheme was instrumental in the classification, with results range at around 83% to 84%, wherein C4.5 performed best. Findings in the study yield that LMS activities were better predictors for success in programming subjects as against prior knowledge [14].

3. METHODOLOGY

3.1 Dataset

The dataset used in this study are the responses of 257 random first year to fourth-year student-respondents that used Schoology under the Bachelor of Science in Entertainment and Multimedia Communication (BSEMC), Bachelor of Science in Information Systems (BSIS), Bachelor of Science in Computer Science (BSCS), and Bachelor of Science in Information Technology (BSIT) programs under the College of Computing Education of the University of Mindanao during the 1st Semester of School Year 2019-2020. The use of 10-folds cross-validation scheme performed in Waikato Environment for Knowledge Analysis (WEKA) was instrumental for the prediction of the dataset with 24 variables having divided into four parts to wit: usefulness, ease of support, learning support, and satisfaction as shown in Table 1 below.

Table 1: Variables used in the study

Category	Description	Variable	Response
Usefulness	Schoology helps me to be more effective.	U1	{1,2,3,4,5,6,7}
	Schoology helps me to be more productive.	U2	{1,2,3,4,5,6,7}
	Schoology gives me more control over the activities in my life.	U3	{1,2,3,4,5,6,7}
	Schoology makes the things I want to accomplish easier to get done.	U4	{1,2,3,4,5,6,7}
	Schoology saves me time when I use it.	U5	{1,2,3,4,5,6,7}
	Schoology meets my needs.	U6	{1,2,3,4,5,6,7}
	Schoology does everything I would expect it to do.	U7	{1,2,3,4,5,6,7}
Ease of Use	Schoology is simple to use.	E1	{1,2,3,4,5,6,7}
	Schoology is user friendly.	E2	{1,2,3,4,5,6,7}
	Schoology requires the fewest steps possible to accomplish what I want to do.	E3	{1,2,3,4,5,6,7}
	Schoology is flexible.	E4	{1,2,3,4,5,6,7}
	Schoology provides simple and clear instructions.	E5	{1,2,3,4,5,6,7}
	Schoology provides recovery of activities quickly and easily.	E6	{1,2,3,4,5,6,7}
	Schoology has no downtime; it functions very well.	E7	{1,2,3,4,5,6,7}
Learning Support	Schoology allows me to understand lessons more.	L1	{1,2,3,4,5,6,7}
	Schoology provides features that facilitate my learning process.	L2	{1,2,3,4,5,6,7}
	Schoology is a good platform to reinforce learning in the classroom.	L3	{1,2,3,4,5,6,7}
	Schoology allows me to appreciate my subject.	L4	{1,2,3,4,5,6,7}
	Schoology helps me manage my academic preparations.	L5	{1,2,3,4,5,6,7}
Satisfaction	I would recommend Schoology to a friend.	S1	{1,2,3,4,5,6,7}
	Schoology is fun to use.	S2	{1,2,3,4,5,6,7}
	I feel I need to have a Schoology.	S3	{1,2,3,4,5,6,7}
	Schoology helps me achieve my learning goals.	S4	{1,2,3,4,5,6,7}
	Schoology is an excellent tool to manage learning.	S5	{1,2,3,4,5,6,7}

3.2 Naïve Bayes Algorithm

The Naïve Bayes is a renowned learning algorithm for machine learning and data mining, known for its efficiency and effectivity [16]–[19]. It is a probabilistic classifier based on Bayes' Theorem with an assumption that all features employ conditional independence given the class variable. Naïve Bayes classifies objects given its attribute values using the Bayes' rule expressed as:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad (1)$$

where, $P(c|x)$ is the posterior probability of class c given predictor x, $P(c)$ is the prior probability of class c, $P(x|c)$ is the likelihood which the probability of predictor x given class c, and $P(x)$ is the prior probability of predictor x.

3.3 K-Nearest Neighbor (KNN) Algorithm

The KNN is a lazy learning algorithm that uses the entire dataset for training and testing. On the other hand, KNN being non-parametric means classification can be done even without prior knowledge about the data distribution. KNN works on feature similarities by predicting the classification of a new data point using a collection of similar classified data points [20]–[22]. In this study, the use of KNN with k value of 3 is observed.

3.4 C4.5 Algorithm

The C4.5 algorithm is a decision tree classifier developed by Ross Quinlan in 1993 [23]. The algorithm one of the widely used classifiers known for its straightforward interpretation [24], [25]. To perform, compute the gain ratio of each attribute first. Attributes whose gain ratio is at maximum will be identified as the root node of the tree. The algorithm uses a pessimistic pruning approach in removing unnecessary branches in the decision tree to increase the classification accuracy [9], [26].

3.5 Prediction Evaluation Tools

In this study, the prediction models are evaluated using the root mean squared error (RMSE) and mean absolute error (MAE) statistical tools with equations below, along with the precision, recall, and f-measure.

$$R.M.S.E. = \sqrt{\sum_{t=T+1}^{T+h} (\hat{y}_t - y_t)^2 / h} \quad (2)$$

$$M.A.E. = \sum_{t=T+1}^{T+h} |\hat{y}_t - y_t| / h \quad (3)$$

4. RESULTS AND DISCUSSION

To predict the accurateness of the responses of the students in the LMS assessment conducted, the dataset was loaded to the Waikato Environment for Knowledge Analysis (WEKA) software and was tested along with the C4.5, KNN, and Naïve Bayes algorithms using the 10-folds cross-validation scheme.

Simulation results in Tables 2-3 revealed that the Naïve Bayes algorithm is the optimal model for prediction as it obtained the highest accuracy of 100% as against the C4.5 and KNN with a 99.61% prediction accuracy for both algorithms. However, both C4.5 and KNN algorithms differ in statistical error values with 0.0624 and 0.0416 error rates, respectively. Further, results show that the KNN algorithm is the second-best method to predict the accuracy of the LMS assessment dataset next to the Naïve Bayes algorithm as evident on the result of the precision, recall, and f-measures, supported by the forecast error evaluation tools used as the model with a lower error rate has the better forecast ability. Furthermore, the accuracy of the dataset is optimal, making the responses of the students in the study of [4] reliable and worthy of implementation. The graphical representation of the results is shown in Figures 1and 2.

Table 2: Prediction model accuracy evaluation

Model	Accuracy %	Precision	Recall	F-Measure
C4.5	99.6109%	0.996	0.996	0.996
KNN	99.6109%	0.996	0.996	0.996
Naïve Bayes	100%	1.000	1.000	1.000

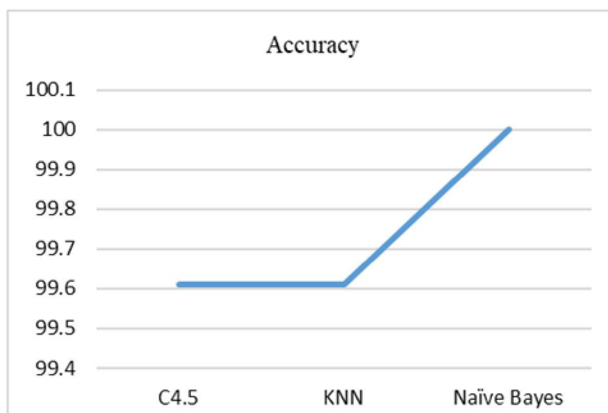

Figure 1: Indexed comparison of prediction accuracies

Table 3: Forecast error evaluation

Model	RMSE	MAE
C4.5	0.0624	0.0039
KNN	0.0416	0.0036
Naïve Bayes	0.0041	0.0004

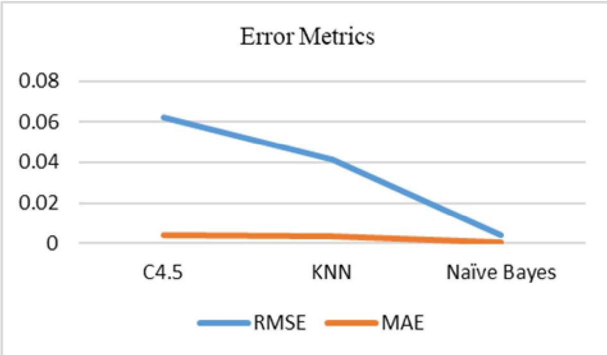

Figure 2: Error rates of the three algorithms

5. CONCLUSION AND RECOMMENDATION

Predicting the accurateness of the students' responses in the study conducted by [4] is important. There is a need to reevaluate the responses of the students to know if the knowledge generated from the actual prior research conducted really needs to be implemented. In this paper, the use of the famous C4.5, KNN, and Naïve Bayes algorithms in the prediction of the accuracy of the LMS assessment dataset are conducted. Simulation results revealed that the data mining algorithms used attained a very satisfactory result with 100% accuracy, validating the acceptability level of the knowledge generated in [4]. It is recommended that the use of the university's Blackboard LMS may be assessed as well.

REFERENCES

[1] C. Owusu-Fordjour, C. K. Koomson, and D. & Hanson, "The impact of COVID-19 on learning-the perspective of the Ghanaian student.," *Eur. J. Educ. Stud.*, vol. 7, no. 3, pp. 88–101, 2020.

[2] R. C. Chick *et al.*, "Using Technology to Maintain the Education of Residents During the COVID-19 Pandemic," *J. Surg. Educ.*, 2020. https://doi.org/10.1016/j.jsurg.2020.03.018

[3] N. D. Vaughan, M. Cleveland-Innes, and D. R. Garrison, *E-Learning in the 21st century: A framework for research and practice, Second edition*. 2011.

[4] J. C. T. Arroyo and C. Villamor, "Use of Schoology as Learning Management System: Evidence of Practice in College of Computing Education," 2020.

[5] K. Rajalakshmi, S. S. Dhenakaran, and N. Roobini, "Comparative Analysis of K-Means Algorithm in Disease Prediction," *Int. J. Sci. Eng. Technol. Res.*, vol. 4, no. 7, pp. 2697–2699, 2015.

[6] A. J. P. Delima, "An Enhanced K-Nearest Neighbor Predictive Model through Metaheuristic Optimization," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 1, pp. 72–80, 2020. https://doi.org/10.14569/IJACSA.2020.0110109

[7] A. J. P. Delima, "Predicting Scholarship Grants Using Data Mining Techniques," *Int. J. Mach. Learn. Comput.*, vol. 9, no. 4, pp. 513–519, 2019. https://doi.org/10.18178/ijmlc.2019.9.4.834

[8] U. O. Cagas, A. J. P. Delima, and T. L. Toledo, "PreFIC : Predictability of Faculty Instructional Performance through Hybrid Prediction Model," *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 7, pp. 22–25, 2019.

[9] A. J. P. Delima, "An Experimental Comparison of Hybrid Modified Genetic Algorithm-based Prediction Models," *Int. J. Recent Technol. Eng.*, vol. 8, no. 1, pp. 1756–1760, 2019.

[10] A. J. P. Delima, A. M. Sison, and R. P. Medina, "Variable Reduction-based Prediction through Modified Genetic Algorithm," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 5, pp. 356–363, 2019.

[11] A. P. Dela Cruz *et al.*, "Higher Education Institution (HEI) Enrollment Forecasting using Data Mining Technique," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 2, pp. 2060–2064, 2020. https://doi.org/10.30534/ijatcse/2020/179922020

[12] E. Susnea, "Using data mining techniques in higher education," in *The 4th International Conference on Virtual Learning ICVL 2009*, 2009, vol. 1, no. 1, pp. 371–375.

[13] K. E. Teng and L. C. Kit, "A Learning Analytics Approach to Model and Predict Learners' Success in Digital Learning," *Australas. Soc. Comput. Learn. Tert. Educ. 2019 Conf. Proc. Publ.*, pp. 476–480, 2019.

[14] J. D. Maulida and Kariyam, "Students Academic Performance Based on Behavior," in *AIP Conference Proceedings*, 2017.

[15] M. S. Halawa, M. E. Shehab, and E. M. R. Hamed, "Predicting student personality based on a data-driven model from student behavior on LMS and social networks," in *2015 Fifth International Conference on Digital Information Processing and Communications (ICDIPC)*, 2015, pp. 294–299.

[16] D. D. Lewis, "Naive (Bayes) at forty: The independence assumption in information retrieval," in *Machine Learning: ECML-98*, 1998, pp. 4–15. https://doi.org/10.1007/BFb0026666

[17] N. Ben Amor, S. Benferhat, and Z. Elouedi, "Naive Bayes vs Decision Trees in Intrusion Detection Systems," in *Proceedings of the 2004 ACM Symposium on Applied Computing*, 2004, pp. 420–424.

[18] A. McCallum and K. Nigam, "A Comparison of Event Models for Naive Bayes Text Classification," *AAAI*, 1998.

[19] H. Zhang, "Exploring Conditions for the Optimality of Naive Bayes," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 19, no. 2, pp. 183–198, 2005. https://doi.org/10.1142/S0218001405003983

[20] Y. Wu, K. Ianakiev, and V. Govindaraju, "Improved k-nearest neighbor classification," *Pattern Recognit.*, vol. 35, no. 10, pp. 2311–2318, 2002.

[21] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN Model-Based Approach in Classification," in *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, 2003, pp. 986–996.

[22] J. M. Keller, M. R. Gray, and J. A. Givens, "A fuzzy K-nearest neighbor algorithm," *IEEE Trans. Syst. Man. Cybern.*, vol. SMC-15, no. 4, pp. 580–585, Jul. 1985.

[23] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993.

[24] S. Ruggieri, "Efficient C4.5," *IEEE Trans. Knowl.*

Data Eng., vol. 14, no. 2, pp. 438–444, 2002. https://doi.org/10.1109/69.991727

[25] H. Jantan, "Human Talent Prediction in HRM using C4 . 5 Classification Algorithm," *Int. J. Comput. Sci. Eng.*, vol. 02, no. 08, pp. 2526–2534, 2010.

[26] R. Benkercha and S. Moulahoum, "Fault detection and diagnosis based on C4.5 decision tree algorithm for grid connected PV system," *Sol. Energy*, vol. 173, pp. 610–634, 2018. https://doi.org/10.1016/j.solener.2018.07.089