

## The Use of Educational Data Mining Technique in Technology Assimilation Evaluation: A Response Assessment



**Marilou B. Carnicer**

College of Technology, Surigao State College of Technology, Surigao City, Philippines

[mbscaggii@gmail.com](mailto:mbscaggii@gmail.com)

### ABSTRACT

This study predicts the accuracy of the students' responses in the actual evaluation conducted in Surigao State College of Technology (SSCT) to be used for decision making and technology assimilation in classroom learning. In this paper, a data mining technique like prediction using the Naïve Bayes algorithm was used to test the accurateness of the responses as perceived by the 53 computer science and information system student-respondents of SSCT. Simulation result showed that the algorithm depicted a high 88.75% accuracy on the used dataset performed in 10-folds cross validation using WEKA software. It is suggested that the output of this study may be used as input and avenue for future researches or as data proliferation that will influence managerial decisions for the betterment of educational practices. Future researchers may utilize other data mining techniques with notable reputations in determining the accuracy of the same dataset. It is hoped that the study will contribute on the two major literatures; in data mining and in higher education mining.

**Key words:** Data mining, Educational data mining, Higher Education, Technology in education, Prediction

### 1. INTRODUCTION

Education is regarded as the medium for developing human knowledge and quality just as improving their potential and competency. Students, being one of the most important element, ought to have the capacity to self-learn and aim for self-advancement. Educational management process must fortify student's natural development with full potential. Students and teachers may learn together from instructional media and technology sources since learning can be dealt everywhere at any time. There is a notion that information technology should be more applied since the use of information technology as the tool for perceiving news and knowledge through self-earning or two-way communication with teachers or colleagues is the current trend [1].

As accentuated by [2], effective teaching in the digital generation involves two basics; embracing the tools that the students are immersed and using these tools to engage students in core curriculum topics. Since the present generation is a digital generation, the so called screen agers, the 21<sup>st</sup> Century teacher should be a "digital native" and not just a digital immigrant. They should be equipped with the knowledge, skills, attitude, and values demanded by the time .

Furthermore, they should have to make the classroom similar to the outside world of digital technology. The teacher injects life into the subject matter by using technology that would bridge the classroom and the high-tech world of students [3].

In the confines of Surigao State College of Technology (SSCT), Surigao, City Campus, a need for the faculty to develop instructional materials is highly encouraged and recommended particularly in the Information Technology (IT) Division. This is in support to the effort of the aforementioned division for curricular program accreditation. During the 1<sup>st</sup> Accrediting Agency of Chartered Colleges and Universities of the Philippines (AACCUP) survey, a strong recommendation to the committees under the curriculum and faculty area is to address the dearth of institutional materials. To be university-ready, instructors are highly encouraged to apply technology in the classroom bringing global scenarios to local settings.

Aligned to the call of the university status is the need of the college to strengthen its function on production. The most feasible contribution that instructors can extend in this area is the development of learning materials like learning guide or manual. This would enable the students to learn independently making them more responsible and more functional in the society as citizens and professionals in the future. It is then a challenge that technology assimilation is done through the learning materials prepared by instructors especially in general sociology. This would also strengthen the skills of the students to relate technology to social science lessons.

Moreover, students enrolled in Information and Communication Technology (ICT) courses are expected to become experts in terms of using computer. Curricular activities along these programs must contribute to the development of the students' skills relative to their chosen course. It is then a challenge to the academic subject instructors to deliver course content with the use of technology in which students are immersed. Teaching general sociology subject must be taught with the use of computer technology so that the development of students in this program will be strengthened. With this, [4] prompted to delve in developing an instructional guide with technology assimilation in the context of constructivism for general sociology to put in place the facilitative role of teachers in the classroom and find out its effectiveness and efficacy to be at par with other skills that advocates pedagogy-technology integration. The study assessed and determined the effectiveness of computer technology using multimedia as an aid to instruction as perceived by computer science and information system students in SSCT.

Despite the advent of technology, there are still researchers who used the old predefined queries and charts in conducting research. They are often unlikely to accomplish a well-established knowledge extraction from databases using the old and traditional methods. The methods employed by other researchers are qualitative and quantitative methods. This is realized with the help of researcher-made questionnaires. The responses on these evaluations are tallied and are kept in databases while interpretation is drawn using statistics and basic mathematical methods. However, the abovementioned methods are outdated and the use of the latter limits the researchers in achieving their quality objectives. Now, with the advent of technology, a new paradigm called data mining has emerged [5].

Data mining (DM) also called as Knowledge Discovery from Databases (KDD) is the process of extracting implicit information or knowledge from huge databases using data mining and machine learning algorithms. Some of the renowned function in data mining used diverse approaches of DM analysis such as decision tree, Bayes classifiers, association rules, clustering, neural networks, genetic algorithms, support vector machines, predicting or forecasting and many more [6]. These approaches encourage data analysis in its full potential with the help of the right algorithms. Among the many data mining approaches, the prediction is considered as the prime method that is commonly used in educational data mining, business, health, and even in almost all sectors of the society [7].

This study used the famous Naïve Bayes algorithm to predict the accuracy and evaluate the responses used in the study of [4]. The result of this study would serve as basis in the development of an improved technology assimilation instructional guide and would be added to the literature of data mining particularly the application of Naïve Bayes algorithm.

## 2. LITERATURE REVIEW

The usage of technology in education became considerably popular in recent years [7]. Introduction of technology to the education environments enables to generate effective learning environments through developing learning materials. The usage of computers in education (internet, simulations, animations, visual and audio presentations, etc.) ushers a new age in transmitting information and presents education tools which potentially can change some traditional and non-effective educational methods. Technology is an important support element in the development of learning processes of the students; computers are tools which provide technologic support to the development of effective learning and teaching processes. The reasons for using education technologies could be supporting the lecturer during the presentation of information, enhancing the success in the lessons, enhancing the permanence of information, improving the interest and motivation of the students for the lessons.

The study of [8] evaluated the use of computers in the academe. Results showed that there is an increase in the use of computers by students and teachers. It was found that students were more enthusiastic and everyone searched for information over the net and they seemed to use home

computers increasingly. It was consistent with the work of [9] who studied on practical courses changing by conducting the research with 1st–3rd year students of University of Missouri. It was found out that the implementation of internet in learning could improve teacher's instruction efficiently

The work of [10] studied the effects of using e-mail by the undergraduate students and examined the changes on the use of electronic services of graduate students. The results showed that educational technology should be integrated in the curriculums for better management in instructions. In addition, there should be practices on e-mail skills between teachers and students. The research reflected that students were able to express their opinions freely that could not be happened through other instructional methods. In general, it is concluded that e-mail should be used by students and teachers.

Premised in the application of data mining in all areas in research, this trend is undoubtedly inevitable. Data mining also termed as Knowledge Discovery in Databases (KDD), is a medium of discovering novel and potentially valuable information from large amounts of data [11]. The field of Educational Data Mining is fresh, new, and developing in the field of education sector which can also be applied in areas such as government, accounts, sports, transportation and a lot more [12]. This new emerging field concerns with developing methods that discover knowledge from data originating from educational environments. Data mining techniques such as decision trees, Naïve Bayes, K- Nearest neighbor, neural networks, K-means clustering and many others are instrumental in extracting data from the datasets [13].

The use of data mining techniques has attracted more and more attention in this big data era [14]. To name some, a study analyzed crimes such as theft, homicide, and various drug offenses along with suspicious activities, noise complaints, and burglar alarm by using qualitative and quantitative approach [15]. Using K-means clustering data mining approach on a crime dataset from the New South Wales region of Australia, crime rates of each type of crimes and cities with high crime rates have been found.

Furthermore, prediction, as one of the data mining techniques, can be made utilizing autoregressive integrated moving average (ARIMA) algorithm that used historical data in predicting cases such as in education, society, climate, health, and others. In other countries, the use of ARIMA algorithm in forecasting incidence of hemorrhagic fever with renal syndrome in China was observed [16]. Meanwhile, the ARIMA model was also used in forecasting dengue hemorrhagic fever cases in Southern Thailand [17]. Further, the potential and effectiveness of using ARIMA modeling in the prediction of travel time to the urban roadway was also proven [18].

Moreover, the data mining techniques were implemented to understand specific trends and pattern of terrorist attacks in India. K-means clustering was used to determine the year wherein the terrorist groups were most active and also which terrorist group has affected the most [19].

On the other note, the literature on representations and distance measures for time-series, clustering, and classification is extensive [20]. Time series analysis method is a kind of data mining method, which is a sequence of data

points, typically consisting of successive measurements made over a time interval. It is a method for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values [21].

### 3. METHODOLOGY

#### 3.1 Datasets

The dataset that were used in this study are the responses of the fifty three 2<sup>nd</sup> year computer science and information system students from SSCT in the academic year of 2014. Table 1 shows the profile of the respondents in the experimental group in terms of their age, gender, and level of computer exposure in games, google, social media, yahoo, and YouTube. The dataset that will be used are the tallied responses of the students from the 40-item test questionnaire deployed. The accuracy of responses as perceived by the students are validated and tested using Naïve Bayes Algorithm. This is to prove the effectiveness and the accurateness of the responses of the student-respondents.

**Table 1:** Profile of the respondents

Profile	f(n=53)	Percent
Age	16-20	29
	21-25	18
	26-30	6
Gender	Male	21
	Female	32
Games	None	4
	1-2 hours	40
	3-4 hours	9
Google	1-2 hours	38
	3-4 hours	15
Social Media	1-2 hours	44
	5 hours or more	9
Yahoo	None	8
	1-2 hours	45
YouTube	None	6
	1-2 hours	35
	3-4 hours	8
	5 hours or more	4

#### 3.2 Naïve Bayes Algorithm

The Naïve Bayes classifier is a probabilistic classifier based on Bayes' theorem. It attributes are fully independent [22] making it reliable and simple to use. Some of the notable advantages of the NB Algorithm includes simplicity to execute and its ability to cope up with noise and irrelevant

data. For the complete formula of Naïve Bayes Algorithm, please refer to the study of [23].

#### 3.3 Prediction Evaluation Tools

Countless forecasting and prediction models found in the literature are evaluated using the various forecast error statistical tools. The following tools listed below will be used:

Root Mean Squared Error (RMSE) (1)

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

Mean Absolute Error (MAE) (1)

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

### 4. RESULT AND DISCUSSION

The use of Waikato Environment for Knowledge Analysis (WEKA) was instrumental to the training of datasets using the predictive model named Naïve Bayes algorithm. To test the predictive capability of the Naïve Bayes algorithm, the dataset was loaded using the 10-folds cross-validation scheme. The prediction model obtained a prediction accuracy of 88% making the responses of the students in the evaluation of the use of technology in instructional guide for general sociology reliable for assimilation and decision making as shown in Table 2.

**Table 2:** Prediction model accuracy evaluation

Model	Accuracy %	Precision	Recall	F-Measure
Naïve Bayes	88.7513 %	0.882	0.845	0.887

**Table 3:** Evaluation tools

AIC	RMSE	MAE
176.53	32.792342	17.6595627

The degree of accuracy of the trained datasets measured by precision, recall, and F-measures are as follows:

Precision - The trained dataset gained an overall score of 88.20% out of 53 respondents. This denotes that out of 53 tested instances, 88.20% of the total responses were retrieved correctly by the algorithm with the use of WEKA. The Precision of the system in testing the accurateness of the responses in the evaluation of the use of technology in sociology courses is good.

Recall - The tested datasets gained an overall score of 84.50% out of 53 instances. This denotes that 84.50% of the total datasets were retrieved by the system. The Recall of the system in retrieving responses is good.

F-Measure - Using the formula for computing F-Measure, the overall computed score for dataset is 88.70%. The F-Measure determines the performance of the system.

## 5. CONCLUSION

Predicting the accuracy of the students' responses in the evaluation conducted by [4] to be used for decision making and assimilation programs is important. Nowadays, it is not enough that a research is conducted especially when research methods used are predefined methods that are outdated. Simulation result shows the effectiveness of data mining technique particularly the use of Naïve Bayes algorithm in knowledge extraction and validation. The use of Naïve Bayes algorithm in predicting the accuracy of the responses helped the researcher in validating the acceptability of the knowledge acquired when the predefined methods were used. It is recommended that the output of this study be used in data proliferation and influence managerial decisions. Future researchers may also utilize other data mining techniques to further test their accuracy and perform another knowledge extraction.

## REFERENCES

- [1] C. Waree, "A Development of Online Lessons of the Sufficiency Economy Philosophy for Sustainability of Learning," *Int. J. Environ. Sci. Educ.*, vol. 9, no. 5, pp. 375–379, 2019.  
<https://doi.org/10.18178/ijet.2019.9.5.1230>
- [2] B. B. Corpuz and I. P. Lucido, "Educational technology 1," in *Lorimar Publishing Inc*, 2012.
- [3] B. B. Corpuz, "Educational technology 1," in *Lorimar Publishing Inc*, 2012.
- [4] M. B. Carnicer, "Technology Assimilation in the Context of Constructivism for General Sociology," 2014.
- [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] 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.
- [7] 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.
- [8] J. M. Casey, "Student Teacher Travel the Information Highway," 2018.
- [9] R. Laroe, "Moving to Virtual Curriculum." 1995.
- [10] J. A. Russett, "Using telecommunications with pre service teachers," *J. Comput. Math. Sci. Teach.*, pp. 65–76, 1992.
- [11] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2012.
- [12] P. Kaur, M. Singh, and G. S. Josan, "Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector," *Procedia Comput. Sci.*, vol. 57, pp. 500–508, 2015.  
<https://doi.org/10.1016/j.procs.2015.07.372>.
- [13] S. K. Yadav and S. Pal, "Data Mining : A Prediction for Performance Improvement of Engineering Students using Classification," *World Comput. Sci. Inf. Technol. J. WCSIT*, vol. 2, no. 2, pp. 51–56, 2012.
- [14] C. Jinyin, L. Xiang, Z. Haibing, and B. Xintong, "A novel cluster center fast determination clustering algorithm," *Appl. Soft Comput. J.*, vol. 57, pp. 539–555, 2017.  
<https://doi.org/10.1016/j.asoc.2017.04.031>
- [15] J. Agarwal, "Crime Analysis using K-Means Clustering," *Int. J. Comput. Appl.*, vol. 83, no. 4, pp. 975–8887, 2013.  
<https://doi.org/10.5120/14433-2579>
- [16] Q. Liu, X. Liu, B. Jiang, and W. Yang, "Forecasting incidence of hemorrhagic fever with renal syndrome in China using ARIMA model," 2011.
- [17] S. Promprou, M. Jaroensutasinee, and K. Jaroensutasinee, "Forecasting dengue haemorrhagic fever cases in Southern Thailand using ARIMA Models," *Dengue Bull.*, vol. 30, pp. 99–106, 2006.
- [18] D. Billings and Y. Jiann-Shiou, "Application of the ARIMA Models to Urban Roadway Travel Time Prediction-A Case Study. Systems, Man and Cybernetics, 2006. SMC'06," *IEEE Int. Conf.*, pp. 2529–2534, 2006.  
<https://doi.org/10.1109/ICSMC.2006.385244>
- [19] P. Gupta, A. S. Sabitha, and T. Choudhury, "Terrorist Attacks Analysis Using Clustering Algorithm," © *Springer Nat. Singapore Pte Ltd.*, pp. 317–328, 2018.  
[https://doi.org/10.1007/978-981-10-5547-8\\_33](https://doi.org/10.1007/978-981-10-5547-8_33)
- [20] E. Keogh and S. Kasetty, "On the need for time series data mining benchmarks," *Proc. eighth ACM SIGKDD Int. Conf. Knowl. Discov. data Min. - KDD '02*, p. 102, 2002.  
<https://doi.org/10.1145/775060.775062>
- [21] W. Hongbin, D. Shuangyu, D. Jianzhuo, R. Ming, and D. Ming, "Study on Condition Pre-warning Method of Power Transformer based on Load Time Series Model \*," pp. 223–227, 2015.  
<https://doi.org/10.1109/STA.2015.7505103>
- [22] D. Soria, J. M. Garibaldi, F. Ambrogi, E. M. Biganzoli, and I. O. Ellis, "A 'non-parametric' version of the naive Bayes classifier," *Knowledge-Based Syst.*, vol. 24, no. 6, pp. 775–784, 2011.  
<https://doi.org/10.1016/j.knosys.2011.02.014>
- [23] W. Chen, X. Yan, Z. Zhao, H. Hong, D. T. Bui, and B. Pradhan, "Spatial prediction of landslide susceptibility using data mining-based kernel logistic regression, naive Bayes and RBFNetwork models for the Long County area (China)," *Bull. Eng. Geol. Environ.*, pp. 1–20, 2018.  
<https://doi.org/10.1007/s10064-018-1256-z>