International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075, Volume-8 Issue-7 May, 2019

PreFIC: Predictability of Faculty Instructional Performance through Hybrid Prediction Model

Unife O. Cagas, Allemar Jhone P. Delima, Teresita L. Toledo

Abstract*: The Higher Education Institutions have amplified the practice of incorporating datamining in extracting information from the data in relation to educational context. As one of the regarded quest of HEIs, predicting faculty instructional performance has made easy; and the accuracy of the result has become more reliable through the application of data mining algorithms and techniques. This study proposed a hybrid model in predicting the instructional performance of faculty in the four State Universities and Colleges (SUC) in Caraga Region, Philippines by integrating k-means segmentation on the C4.5 algorithm prior to prediction. A total of 597 records of student-respondents was used for simulation using the 10-folds cross validation scheme. Simulation result showed that with integration of k-means algorithm, the identified prediction accuracy of 86.09% using C4.5 algorithm alone has increased to 87.93%. Future researchers may utilize other hybrid algorithms in the quest on improving the literature of educational data mining.

***Index Terms*: accuracy enhancement, hybrid model, instructional performance, prediction**

I. INTRODUCTION

Today, there are some identified challenges that each of the higher education institutions (HEIs) are currently facing. This includes the proliferation of data and how those data will be of use to increase the excellence of academic curriculums and services as well as how those data will influence managerial decisions. There are existing procedures, such as qualitative and quantitative methods that the HEIs used in data extraction but it limits them in achieving their quality objectives. The methods used by HEIs are mostly grounded on predefined queries and charts. Therefore, these approaches inhibit the full potential of data analysis and knowledge discovery due to its inability to reveal new and complete valuable hidden information [1].

Recently, one of the quests the HEIs delved in is the evaluation of faculty instructional performance in the classroom. The most commonly used tool to evaluate such performance is by getting the responses of the students with regards to the course and its instructor in the form of surveys where deployment of questionnaire is observed. Since the introduction of student evaluation in higher education systems in the 1920's, an unending argument on the consistency and validation of the assessments has become prevalent up to date [2]. Students who do not have adequate knowledge and deep understanding to assess the program and the faculty; and student's evaluation being influenced by the reputation of the program or handling faculty; as well as grades received by student from such faculty are some of the concerns about student evaluations [3]. The result of these evaluations are tallied and are stored in databases where interpretation is drawn from the field of statistics and basic mathematical methods.

Now, the analysis and processing of huge data has become a stronghold technique that is being applied in extracting useful information and discovering patterns from databases [4]. The literature suggest that concealed data from huge databases is best investigated through the integration of data mining (DM) analysis. DM, as coined with Knowledge Discovery in Databases (KDD), discovers patterns of information and hidden knowledge within large databases where it opt to make predictions for certain outcomes or behaviors [5]. Some of the generally accepted functions in data mining includes association, classification, clustering, estimation and prediction [6].

The use of data mining techniques in higher education has been amplified lately. With this, a new area termed as educational data mining (EDM) has emerged [7]. This field concerns on the development of approaches that explores information from scholastic settings with the drive of providing valuable info and facts from educational context. In educational data mining, further understandings can be obtained from scholastic entities such as better allocating resources [8], objective managerial decision support [9], predicting student performance [10] [11], e-learning systems [12], pedagogical support [13], and more.

With the aforementioned quest where the HEIs delved in, the authors aimed to show the impending benefits of DM in predicting effectiveness towards the faculty instructional performance as observed by students from the identified State Universities and Colleges (SUCs) in Caraga Region, Philippines to be added to the literature of knowledge for hybrid algorithms and in EDM. In this study, the K-means segmentation was performed prior to the prediction using the infamous C4.5 algorithm. The goal of this paper is to increase the prediction accuracy when C4.5 algorithm is used by incorporating K-Means algorithm. To realize, the datasets were loaded to WEKA software application for prediction after the K-Means segmentation.

Revised Manuscript Received on May 10, 2019

Unife O. Cagas, College of Engineering and Information Technology, Surigao State College of Technology, Surigao City, Philippines.

Allemar Jhone P. Delima, College of Engineering and Information Technology, Surigao State College of Technology, Surigao City, Philippines.

Teresita L. Toledo, College of Engineering and Information Technology, Surigao State College of Technology, Surigao City, Philippines.

*Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication*

II. LITERATURE REVIEW

A. Hybrid Prediction Models

With the advent of hybridization of various algorithms, prediction models have become more effective and efficient in performing the job. According to [14], in analyzing datasets, data mining algorithms are required coined with new parallelized implementations such as hybridization of methods to produce much better results.

According to the literature, most of the prediction models clustered the training data before prediction stage takes place to elevate the prediction accuracy. Further, it is evident in the literature that the K-means algorithm is the frequently used clustering algorithm that is combined in various prediction models in order to produce admirable prediction rate. After clustering the data, a prediction model will be used to obtain the desired predictive values through graph or a decision trees. Literature revealed that there are existing studies that prefer to use decision trees for prediction [15].

Meanwhile, various researchers employed different machine learning and data mining algorithms for prediction. Algorithms such as C4.5 [16], artificial neural network (ANN) [17][18], recurrent neural network, LPBoost [19], random forest [20], naive bayes [21], bayes net [22], K-nearest neighbor [23], linear regression [24], ARIMA algorithm, and others are some of the commonly used for such purpose.

III. METHODOLOGY

This study used the 597 records of student-respondents in the evaluation of the faculty instructional performance from the four State Universities and Colleges (SUC) in Caraga Region, Philippines. There are 30 variables in the dataset that will be clustered prior to prediction in a 10-folds cross validation scheme. Each items are represented by its mean value obtained from the male and female respondents. The cluster analysis was done using the software called KNIME analytics. Meanwhile, the software application called WEKA was instrumental for the prediction. Table 1 depicts the variables used in this paper.

Table 1. Variables used

Category	Variable	Mean Value	
		Male Respondent	Female Respondent
Methodology	M1	3.27333	3.13230
	M2	3.40667	3.32296
	M3	3.40667	3.39300
	M4	3.38667	3.31518
	M5	3.28000	3.22179
Classroom Management	C1	2.9867	3.0117
	C2	2.62	2.5798
	C3	3.0267	2.9222
	C4	3.4133	3.1712
	C5	3.2667	3.2763
Student Discipline	SD1	3.23333	3.24903
	SD2	3.14	3.1284
	SD3	3.40667	3.48249
	SD4	3.46667	3.4786
	SD5	3.29333	3.32685
Assessment of Learning	A1	3.05444	3.16342
	A2	3.13333	3.08949
	A3	3.29333	3.25
	A4	3.16667	3.05837
	A5	3.08667	3.0856
Student-teacher relationship	ST1	3.43333	3.36187
	ST2	3.26	3.24514
	ST3	3.4	3.23346
	ST4	3.28	3.2607
	ST5	3.20667	3.15953
Peer relationship	P1	3.14	3.2179
	P2	3.1267	3.0778
	P3	3.1	3.1245
	P4	3.2067	3.3191
	P5	3.22	3.2257

A. Clustering

According to [25], data clustering is an unsupervised classification method that create groups of objects, or clusters, where object in the same cluster has similar traits, otherwise, distinct. Being said, the cluster analysis is the first step towards knowledge discovery. The K-Means is one of the commonly used algorithm for cluster analysis [26]. The clustering algorithm works as follows:

1. Randomly initialize centroid distance.
2. For each data point X_i, where i = 1, …n, the D distance is calculated from cluster centroid Z_i, where i = 1, …n and every data point is allocated to the group with the least distance. To calculate the distance, the formula is observed:

$$D\,(X_l, Z_j) = \sqrt{\textstyle\sum_{i=1}^{k} (X_{li},\ Z_{ji})^2}, \qquad 8$$

where X_l is the lth vector while Z_j is the centroid (C) distance of cluster j.

3. Calculate the C distance using:

$$Z_j = \frac{1}{n_j} \sum_{\forall X_l \in Z_j} X_l, \qquad 9$$

B. C4.5 Algorithm

According to [27], the C4.5 algorithm is a descendant to ID3 model which was created by J. Ross Quinlan. The C4.5 is the most famous decision tree algorithm in machine learning. To perform, compute the gain ratio of each attribute first. Attributes whose gain ratio is at maximum will be identified as the tree's root node. The algorithm uses a pessimistic pruning approach in removing redundant branches of the tree in increasing accuracy of the classification method.

IV. RESULT AND DISCUSSION

A. Variable Segmentation using K-Means Algorithm

The variables found in the dataset were clustered into two. This is the first step of the process before prediction. The cluster analysis was done using KNIME analytics. Variables within each clusters have high resemblance with each other that is why they are grouped in the same cluster. The groupings of the variables is shown in Table 2.

Table 2. Cluster result using KNIME

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Variable	Mean Value		Cluster
	Male Respondent	Female Respondent	
M1	3.27333	3.13230	cluster_1
M2	3.40667	3.32296	cluster_1
M3	3.40667	3.39300	cluster_1
M4	3.38667	3.31518	cluster_1
M5	3.28000	3.22179	cluster_1
C1	2.9867	3.0117	cluster_2
C2	2.62	2.5798	cluster_2
C3	3.0267	2.9222	cluster_2
C4	3.4133	3.1712	cluster_1
C5	3.2667	3.2763	cluster_1
SD1	3.23333	3.24903	cluster_1
SD2	3.14	3.1284	cluster_2
SD3	3.40667	3.48249	cluster_1
SD4	3.46667	3.4786	cluster_1
SD5	3.29333	3.32685	cluster_1
A1	3.05444	3.16342	cluster_2
A2	3.13333	3.08949	cluster_2
A3	3.29333	3.25	cluster_1
A4	3.16667	3.05837	cluster_2
A5	3.08667	3.0856	cluster_2
ST1	3.43333	3.36187	cluster_1
ST2	3.26	3.24514	cluster_1
ST3	3.4	3.23346	cluster_1
ST4	3.28	3.2607	cluster_1
ST5	3.20667	3.15953	cluster_2
P1	3.14	3.2179	cluster_2
P2	3.1267	3.0778	cluster_2
P3	3.1	3.1245	cluster_2
P4	3.2067	3.3191	cluster_1
P5	3.22	3.2257	cluster_1

Result showed that from 30 variables, 18 of them belongs to cluster 1 while the remaining 12 variables falls under cluster 2. The items that belongs to cluster 1 are the most observed instructional performance by the faculty as perceived by the student-respondents upon comparing the simulation result to the actual result of the conducted survey. The indexed result is shown in Table 3 and Fig.1.

Table 3. Indexed cluster analysis

Clusters	List of Variables	Total Number of Variables
1	M1, M2, M3, M4, M5, C4, C5, SD1, SD3, SD4, SD5, A3, ST1, ST2, ST3, ST4, P4, P5	18
2	C1, C2, C3, SD2, A1, A2, A4, A5, ST5, P1, P2, P3	12

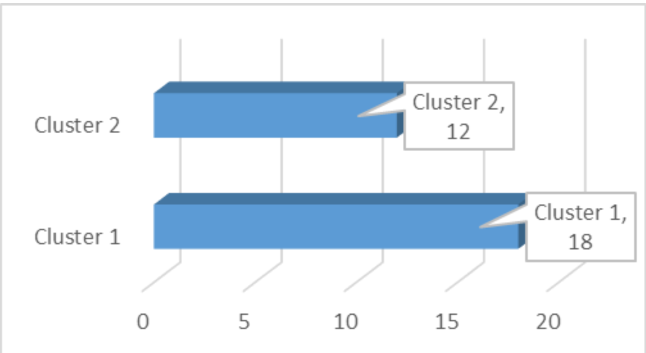

Fig. 1. Graphical representation of cluster count

B. Prediction Model Accuracy Enhancement

A 10-folds cross validation scheme was used in predicting the faculty instructional performance through WEKA. First, the dataset were tested using the complete number of variables to identify if there are changes in the accuracy of prediction. Next, a prediction using the data containing the variables from cluster 1 with 597 instances was done followed by the prediction using variables from cluster 2.

Table 4. Indexed prediction result using C4.5 algorithm with K-means segmentation

Model	Accuracy %	Precision	Recall	F-Measure
C4.5	86.0972%	0.860	0.861	0.860
C1+C4.5	87.6047%	0.875	0.876	0.876
C2+C4.5	87.9397%	0.878	0.879	0.878

Table 4 shows the simulation result for prediction with the integration of k-means segmentation. It is obvious in the simulation that when k-means clustering algorithm is integrated, the accuracy of the C4.5 algorithm has managed to increase. This is due to the reduced number of variables prior to prediction. The prediction model with C4.5 algorithm alone has achieved an 86.09% accuracy. With the reduced number of variables such that in cluster 1, the prediction has reached to 87.60% while 87.93% for cluster 2.

V. CONCLUSION

A prediction model with k-means segmentation, like the works found in the literature, has proven to be effective in increasing the accuracy rate of such models. An increase in the prediction was depicted making the study a success. It has again been proved that maximized accuracy through the reduced number of attributes and better understandability and interpretability of the results are among the many benefits perceived in data reduction. Future researchers may utilize other techniques in reducing the variables prior to prediction using other decision tree algorithms for EDM purposes.

REFERENCES

1. A. M. Abaidullah, N. Ahmed, and E. Ali, “Identifying Hidden Patterns in Students’ Feedback through Cluster Analysis,” Int. J. Comput. Theory Eng., vol. 7, no. 1, pp. 16–20, 2015.
2. S. Calkins and M. Micari, “Less-Than-Perfect Judges: Evaluating Student Evaluations,” NEA-Thought & Action, vol. 73, no. 3, pp. 7–33, 2010.
3. L. Coburn, “Student Evaluation of Teacher Performance,” ERIC Clearinghouse on Tests Measurement and Evaluation Princeton NJ, 1984. [Online]. Available: http://ericae.net/edo/ed289887.htm.
4. C. W. Tsai, C. F. Lai, H. C. Chao, and A. V. Vasilakos, “Big data analytics: a survey,” J. Big Data, vol. 2, no. 1, pp. 1–32, 2015.
5. A. Savaliya, A. Bhatia, and J. Bhatia, “Application of Data Mining Techniques in IoT: A Short Review,” Int. J. Sci. Res. Sci. Eng. Technol., vol. 4, no. 2, pp. 218–223, 2018.
6. M. Brilliant, DwiHandoko, and Sriyanto, “Implementation of Data Mining Using Association Rules for Transactional Data Analysis,” 3rd Int. Conf. Inf. Technol. Bus., pp. 177–180, 2017.
7. M. Zaffar, M. Ahmed, K. S. Savita, and S. Sajjad, “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.
8. L. Jing, “Data Mining and Knowledge Management in Higher Education,” in Annual Forum for the Association for Institutional Research, 2002, pp. 1–18.
9. E. Sugiyarti, K. A. Jasmi, B. Basiron, M. Huda, K. Shankar, and A. Maseleno, “Decision support system of scholarship grantee selection using data mining,” Int. J. Pure Appl. Math., vol. 119, no. 15, pp. 2239–2249, 2018.
10. S. B. Sarkar and B. C. Mohan, First International Conference

on Artificial Intelligence and Cognitive Computing. Springer Singapore, 2018.
11. P. N. P. Maya, J. R. Aguilar, R. A. Zamora, and J. M. Barron, "Predictive Model Design applying Data Mining to identify causes of Dropout in University Students," Strateg. Technol. Soc., vol. 7, no. 2, pp. 11–39, 2018.
12. H. M. Truong, "Integrating learning styles and adaptive e-learning system: Current developments, problems and opportunities," Comput. Human Behav., vol. 55, no. 2, pp. 1185–1193, 2016.
13. J.-L. Hung and S. M. Crooks, "Examining Online Learning Patterns with Data Mining Techniques in Peer-Moderated and Teacher-Moderated Courses," J. Educ. Comput. Res., vol. 40, no. 2, pp. 183–210, 2009.
14. J. C. Jackson, V. Vijayakumar, M. A. Quadir, and C. Bharathi, "Survey on programming models and environments for cluster, cloud, and grid computing that defends big data," Procedia Comput. Sci., vol. 50, pp. 517–523, 2015.
15. M. Y. Orong, A. M. Sison, and R. P. Medina, "A Hybrid Prediction Model Integrating a Modified Genetic Algorithm to K-means Segmentation and C4.5," in TENCON 2018 - 2018 IEEE Region 10 Conference, Jeju, Korea (South), 2018, pp. 1853–1858.
16. 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, no. April, pp. 610–634, 2018.
17. S. Gurgen, B. Unver, and I. Altın, "Prediction of cyclic variability in a diesel engine fueled with n-butanol and diesel fuel blends using artificial neural network," Renew. Energy, no. 117, pp. 538–544, 2018.
18. A. S. Park et al., "Predicting PM 10 Concentration in Seoul Metropolitan Subway Stations Using Artificial Neural Network (ANN)," J. Hazard. Mater., no. 341, pp. 75–82, 2018.
19. K. M. Asim, F. Martínez-Álvarez, A. Basit, and T. Iqbal, "Earthquake magnitude prediction in Hindukush region using machine learning techniques," Nat. Hazards, vol. 85, no. 1, pp. 471–486, 2017.
20. P. Silva, L. T. Cao, and W. B. Hayes, "SpArcFiRe : Enhancing Spiral Galaxy Recognition Using Arm Analysis and Random Forests," Galaxies, vol. 6, no. 3, pp. 1–19, 2018.
21. X. Feng, S. Li, C. Yuan, P. Zeng, and Y. Sun, "Prediction of Slope Stability using Naive Bayes Classifier," J. Civ. Eng., vol. 22, pp. 941–950, 2018.
22. A. Rovinelli, M. D. Sangid, H. Proudhon, Y. Guilhem, R. A. Lebensohn, and W. Ludwig, "Predicting the 3D Fatigue Crack Growth Rate of Small Cracks Using Multimodal Data via Bayesian Networks: in-situ Experiments and Crystal Plasticity Simulations," J. Mech. Phys. Solids, 2018.
23. M. Huang, R. Lin, S. Huang, and T. Xing, "A novel approach for precipitation forecast via improved K-nearest neighbor algorithm," Adv. Eng. Informatics, vol. 33, pp. 89–95, 2017.
24. P. M. Buscema, G. Massini, and G. Maurelli, "Artificial adaptive systems to predict the magnitude of earthquakes," Boll. di Geofis. Teor. ed Appl., vol. 56, no. 2, pp. 227–256, 2015.
25. A. Bansal, M. Sharma, and S. Goel, "Improved K-mean Clustering Algorithm for Prediction Analysis using Classification Technique in Data Mining," Int. J. Comput. Appl., vol. 157, no. 6, pp. 35–40, 2017.
26. S. Kant, T. Mahara, V. Kumar Jain, D. Kumar Jain, and A. K. Sangaiah, "LeaderRank based k-means clustering initialization method for collaborative filtering," Comput. Electr. Eng., vol. 69, pp. 598–609, 2018.
27. 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.

AUTHORS PROFILE

Unife O. Cagas is an Associate Professor under the College of Engineering and Information Technology at Surigao State College of Technology, Surigao City, Philippines. She received her Doctor in Technology Education degree at Mindanao University of Science and Technology at Cagayan de Oro City in 2015 while she obtain her master in information technology degree at St. Paul University Surigao in the year 2000. Her research interest is in educational data mining.

Allemar Jhone P. Delima was born in Surigao City, Philippines. He is a faculty under the College of Engineering and Information Technology at Surigao State College of Technology. He received his B.S. degree from Surigao State College of Technology (SSCT), Surigao City, in 2014 and his Masters degree from the same institution in 2016, both in Information Technology. He is currently pursuing his Doctor in Information Technology degree in the Technological Institute of the Philippines – Quezon City. His research interest is in Data Mining, Data Analytics and Machine Learning.

Teresita L. Toledo is a faculty under the College of Engineering and Information Technology at Surigao State College of Technology, Surigao City, Philippines. She is pursuing her doctoral degree in information technology at Cebu Institute of Technology University at Cebu City.