The Application of Data Mining for Predicting Academic Performance Using K-means Clustering and Naïve Bayes Classification

Zainab Mohammed Ali, Noor Hasan Hassoon, Wasan Saad Ahmed
and Hazim Noman Abed

Abstract--- *Data Mining is a multidisciplinary analyzing process that concentrating to extract and discover useful knowledge from data and information. The field of high education managements is giving a big concern to the knowledge discovery in the academic performance among different courses. Therefore, The education quality is judged by the level of the student's success besides which accreditation that the educational institution can preserve its students. This is for the reason that, there are several factors affecting the academic performance and then the quality of education. The Naïve Bayes classifier is perhaps the most broadly applied probabilistic classifier approaches that can be used for the data exploration. This paper is using the Naïve Bayes classifier for educational data mining process to help in enhancing the quality distinction of the instructive system in higher education. This is by mining student evaluation data related to the instructor's performance to study the main attributes that may affect the educational performance in different courses. Therefore, this paper is using k-means clustering algorithm, which is used to decide the ideal cluster center, so it can be the cluster centroid. Furthermore, the Naïve Bayes algorithm of classification process is applied for the academic evaluation data to generate rules which are studied and evaluated to predict the educational performance. The proposed system helps identifying the dropouts and provides the appropriate advising or counseling for educational management in performing knowledgeable decisions for considering and restructuring the educational curricula. Also, to enhance the academic experience of instructors that would ultimately improve the quality of educational environment of an educational institution.*

Keywords--- *Educational Data Mining, Academic Performance, Educational Management, Naïve Bayes, Classification.*

I. INTRODUCTION

The education quality is judged by the level of the student's success besides which accreditation that the educational institution can preserve its students. Predicting the performance of academic instructor aid recognizing the educational management functioning for providing on time support and attain crucial steps to the academic instructors for improving their overall performance. The capabilities of predicting the overall performance related with the educational institutions is imperative in the educational sector (Gašević et al., 2016). Using data mining methods that explore hidden information group is very important in supporting the decision-making efficiency. It can identify distinctive aspects that can affect the students learning actions and the general performance across the

Zainab Mohammed Ali, Computer Science Department, College of Science/ University of Diyala, Diyala, Iraq.
Noor Hasan Hassoon, Computer Science Department, College of Education of pure Science/ University of Diyala, Diyala, Iraq.
Wasan Saad Ahmed, Computer Science Department, College of Science/ University of Diyala, Diyala, Iraq.
Hazim Noman Abed, Computer Science Department, College of Science/University of Diyala, Diyala, Iraq.
E-mail: hazim_numan@sciences.uodiyala.edu.iq

DOI: 10.37200/IJPR/V24I3/PR200962
Received: 20 Jan 2020 | Revised: 06 Feb 2020 | Accepted: 12 Feb 2020

educational sector in line with the moving to upgrade toward smart technologies and applications (Mohammed et al., 2018). Prediction advancement by using the classification methods of data mining can be employed on the fundamentals of discovered predictive keyword (Malhotra, 2015; Kotsiantis et al., 2006)

Progressive the efficiency of academic instructors is not a straightforward task for the academic field of higher education. The functionality of academic instructors in the educational organizations across bachelor's educational lane is mostly relies on the evaluation scores provided by students in a survey method (Morris, 2016). This kind of critical information can help the lecturers and the educational management to minimize the lack of success ratio to an imperative rate and improve the overall academics functionality. IT general shortage of education methods was totally obvious in strategy, since the industrial sectors are quite often complained the fact that fresh and new computer science graduated students was indeed unprepared to get dealing with the real-world software system projects (Romiszowski, 2016; Pugh and Aspray, 1996; Razaque et al., 2017). Higher education institutions were in fact also believed in this shortage and trying to response with a wide variety of innovative developments to establish training assignments much closely to industrial sectors. It was considered including purposely brought in realistic problems into academic projects, retained large scale projects, carried on assignment which usually distinct sets of students and scholars work on as each semester (Razaque et al., 2017), the needful of university students to perform practical assignment funded by industrial agency (Kezar et al., 2015), incorporate perhaps many educational institutions and professions into industrial project (Ralph and Stubbs, 2014), and some others.

The proposed study is based upon Naive Bayes of data mining approach combined with data clustering which usually empower the analysis to predict the lecturer’s performance, which is provide a very important action to improve the overall performance of bachelor's academic students. Likert-type data was usually used pointer of effectively measured the course assessment which had to be maintained. Subsequently, Likert-type static continued to be largely distributed contributing factor employed by many academics to evaluate the educational progression in academic community (Michaelson and Stacks, 2017; Taylor, 2016). Several performance factors which can include the course objectives and offered knowledge, which usually shown the existing educational performance during the academic course. All these characteristics have recently been concentrated by lecturers in innovating regulations to move forward the education level of students and as well , the overall performance of academics all the way through the strategy of observing performance evolution (Romiszowski, 2016 ;Taylor, 2016). While using the Naïve Bayes along with clustering procedure of data mining approach, it had become possibility to discover the key features of forthcoming prediction. The clustering of data is a course of action for extracting unidentified and hidden pattern coming from huge dataset. The clustering method was in fact used to divide the academic instructors directly onto identical groups in accordance to their performance and capabilities that will help the lecturers and management to greatly improve the level of quality in higher education.

The sections that outlined in this paper are structured as the following: In Section II, we present the method of the conducted research and materials that used for the analysis of the selected dataset. In Section III we present the selected dataset, the experimental analysis along with the achieved results. Finally, the conclusion of this work is delivered in Section IV.

II. Method And Material

The methodology employing Naïve Bayes classifier is implied to evaluate the academic instructor’s general performance. The type of aspects tested for the overall evaluation of the academic instructors was considered the educational value, delivered aims and objectives, and additionally course load activities. The critical information used to be among the educational dataset, which in turn was extractable with the aid of data mining techniques. The classification process was applied to approximate the effectiveness of course knowledge delivery. While certainly, there are various techniques can be implemented for classification, the Naïve Bayes classifier utilized to get help for extracting information to define the overall performance of the academic instructors in each academic course. When it comes to one particular academic system, the entire performance that delivered by the academic instructors had to be determined by way of internal assessment. It is carried out by using the academic actions of each lecturer based on their particular performance. Therefore, in this paper, the approach initiated by the definition of the problem, then identifies the particular dataset, performing the pre-processing procedure, experiments of the proposed methodology, obtain the results, and finally knowledge representation process. Figure1 illustrate the system framework.

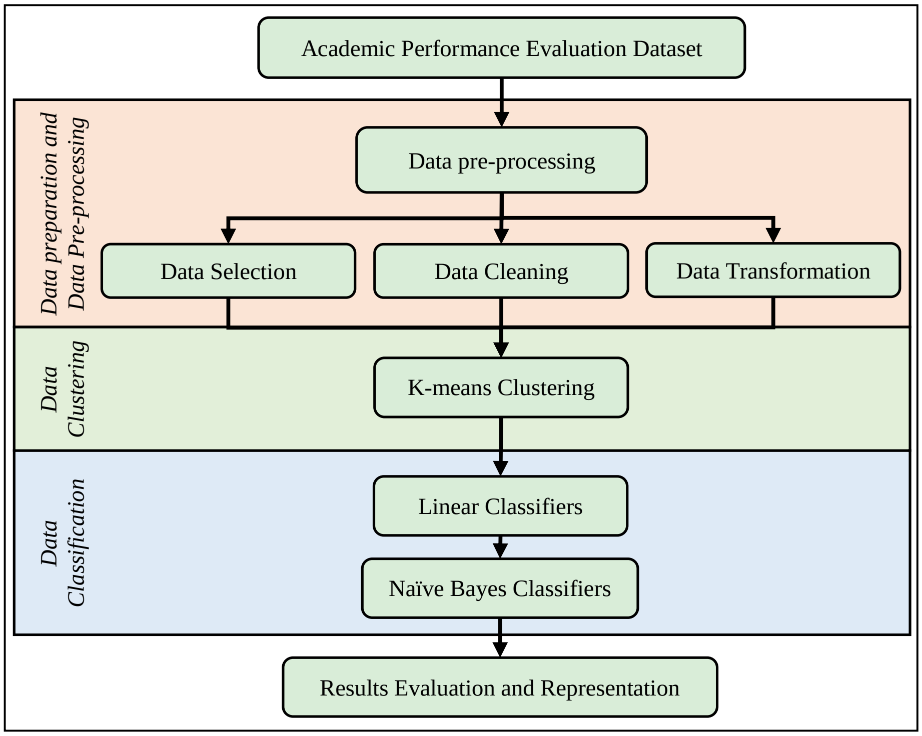


Figure 1: The System Framework

A. 2.1 Data Preparation (The Pre-processing Stage)

The determination of data Pre-processing has been performed to examine and transform the dataset to be able to construct significantly more improved high-quality dataset. When not having data Pre-processing, hidden information may not reached and accessed by utilizing data mining typical models (Witten, 2016). The data Pre-processing stage has been performed to really improve and refine the quality of data by simply removing corrupted data values. Furthermore, in this stage, a data selection and transformation has been performed by determining the

DOI: 10.37200/IJPR/V24I3/PR200962
Received: 20 Jan 2020 | Revised: 06 Feb 2020 | Accepted: 12 Feb 2020

preferred variables, even though particular variables of data have been mined from dataset.

B. 2.2 Data Clustering

The data clustering is an unsupervised approach. Furthermore, the clustering of data is usually done by a statistical methods of data analysis to categorize duplicated information in line with identical cluster group (Hassan and Iskandar, 2017; Talib et al., 2018). It can be employed to handle large datasets for discovering hidden patterns and its matching relationships, so that will help to achieve decisions quickly and proficiently. The clustering analysis divide bulky dataset in line with subsets specified for the instance cluster (Md Shah et al., 2018). Each cluster includes a set of objects that linked to each other and located in an equivalent cluster but also in the same time it disparate to the objects among other clusters (Witten, 2016). In this research, we are using the k-means algorithm as a line of nonhierarchical clustering techniques intended for determining the task of the available data type as well as, the specified operation of analysis.

The K-means algorithm is one of the most widespread clustering techniques for the purpose of clustering data. Furthermore, this algorithm has been used very widely in many different domains such as statistical data analysis, data mining, and several other industry solutions. The K-means clustering algorithm creates clusters by RFM attributes that represents (R: refers to Recency, F: refers to Frequency, M: refers to Monetary). Furthermore, the K-means clustering algorithm was proposed by (Hassan and Iskandar, 2017) aimed to the purpose of assigning every single alternative into a cluster that have the closest centroid i.e. mean. Additionally, the k-means clustering approach have the ability to generate accurately k different clusters of highest available dissimilarities and furthermore the ideal amount of clusters k will lead to extensive decisiveness, which is not really referred to a priori and then needs to be calculated right from the data.

1. The algorithmic procedure for K-Means Clustering will be as the following:
2. Let X = {x1, x2, x3, …, xn} will be the set of data points, also V = {v1, v2, v3, ..., vc} be the set of centres.
3. Selecting the ‘c’ cluster centers in random manner.
4. Calculate the distance among all data points and cluster centers.
5. Assigning the data points for cluster centre which the distance from the cluster centre will be the smallest among entirely cluster centers.
6. Recalculating the new cluster centre using the following formula.

$$v_i = (1/c_i)\sum_{j=1}^{c_i} x_i$$

7. Recalculating the distance among all data points in addition to the new gained cluster centers.
8. In the case of not any data point has been reassigned then stop, otherwise repeat from step c.

The flow chart of the k-means algorithm is shown in Figure 2.

DOI: 10.37200/IJPR/V24I3/PR200962
Received: 20 Jan 2020 | Revised: 06 Feb 2020 | Accepted: 12 Feb 2020

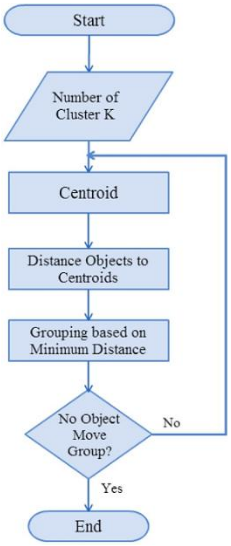

Start
Number of Cluster K
Centroid
Distance Objects to Centroids
Grouping based on Minimum Distance
No Object Move Group?
No
Yes
End

Figure 2: Flow chart of k-means clustering.

2.3 Classification

Classification is a method that mostly utilized through the educational data mining for predicting groups exist in the educational dataset (Frank and Hall, 2011). It used to be employed by academics in educational environment to get enhanced comprehend actions for the academic performance. This is to actually improve the teaching capability, as well as, delivering alternative solutions for the purpose of solving issues related with the management to enrich the quality of education (Romiszowski, 2016 ; Razaque et al., 2017). The student evaluation of the academic instructor's capabilities appears to be predicted by using expending data mining technique that known as classification rules. The Naïve Bayes classification algorithm has been applied to predict the academic instructor's productivity and performance. The Naïve Bayes classifier is a simple probabilistic classifier based on linking Bayes theorem by way of naive impartiality assumptions (Burhanuddin et al., 2018). The Naïve Bayes classifiers is trained highly expeditious in supervised educational area. It is rather simple to understand, unconcerned for not related data features, requesting a training data for diverse parameters estimation, well handling for real and distinct data (Cohen et al., 2014; Shmueli et al., 2017).

III. Experimental Analysis And Results

3.1 Tools Used for Analysis

In this research, we have been used the Weka (Waikato Environment for Knowledge Analysis) software (Hall, 2009), which is a popular suite of machine learning software written in Java, developed at the University of Waikato, New Zealand. Weka is free software available under the GNU General Public License. Accordingly, the enormous data stored in the selected educational dataset will be analysed by Weka software for extracting valuable information to predict the performance of academic instructors.

3.2 Data Preparation and Pre-processing

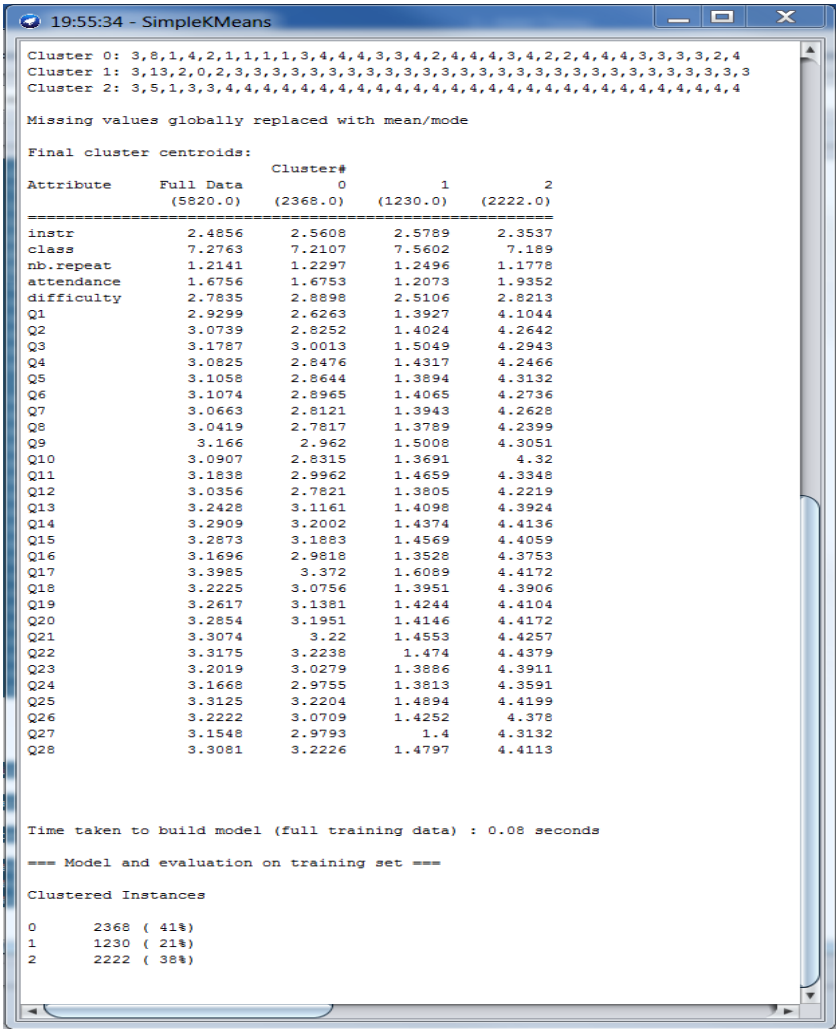
The dataset used in this paper is a student evaluation of the academic instructor's performance and is available at

DOI: 10.37200/IJPR/V24I3/PR200962
Received: 20 Jan 2020 | Revised: 06 Feb 2020 | Accepted: 12 Feb 2020

UCI website (Asuncion and Newman, 2007). The downloaded data came in csv and R format. Thus, in order to use the data set in Weka that specified by attribute-relation file format (.arff), the data set has been pre-processed with ARFF Viewer tool, which is integrated with Weka software. The dataset contains a total 5820 evaluation scores provided by university students for the evaluation of the academic instructor's performance. There is a total of 33 questions out of which 28 are course specific questions and additional 5 attributes were prepared for data mining method. The data relating to course specific questions involves a Likert-type scale, which has the response values of the form {1, 2, 3, 4, 5}. The additional 5 attributes are; instructor, course code values, how many times that the students are taking this course, the course attendance rates, and values for level of the course difficulty as perceived by the students. Twenty-eight course specific questions have been utilized in this paper as outlined in the Table 1.

3.3 Analysis, Results and Discussion

The clustering method applied is the k-means clustering, which is used to adopt the best cluster center to be centroid (Witten, 2016). The selected clustering method has been formed three clusters. The Figure 3 shows the result of the centroid counter, furthermore, it presents the average value of each cluster.



```
Cluster 0: 3,8,1,4,2,1,1,1,1,3,4,4,4,3,3,4,2,4,4,4,3,4,2,2,4,4,4,3,3,3,3,2,4
Cluster 1: 3,13,2,0,2,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3
Cluster 2: 3,5,1,3,3,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4

Missing values globally replaced with mean/mode

Final cluster centroids:
                         Cluster#
Attribute     Full Data         0          1          2
               (5820.0)  (2368.0)   (1230.0)   (2222.0)
=======================================================
instr            2.4856    2.5608     2.5789     2.3537
class            7.2763    7.2107     7.5602      7.189
nb.repeat        1.2141    1.2297     1.2496     1.1778
attendance       1.6756    1.6753     1.2073     1.9352
difficulty       2.7835    2.8898     2.5106     2.8213
Q1               2.9299    2.6263     1.3927     4.1044
Q2               3.0739    2.8252     1.4024     4.2642
Q3               3.1787    3.0013     1.5049     4.2943
Q4               3.0825    2.8476     1.4317     4.2466
Q5               3.1058    2.8644     1.3894     4.3132
Q6               3.1074    2.8965     1.4065     4.2736
Q7               3.0663    2.8121     1.3943     4.2628
Q8               3.0419    2.7817     1.3789     4.2399
Q9                3.166     2.962     1.5008     4.3051
Q10              3.0907    2.8315     1.3691       4.32
Q11              3.1838    2.9962     1.4659     4.3348
Q12              3.0356    2.7821     1.3805     4.2219
Q13              3.2428    3.1161     1.4098     4.3924
Q14              3.2909    3.2002     1.4374     4.4136
Q15              3.2873    3.1883     1.4569     4.4059
Q16              3.1696    2.9818     1.3528     4.3753
Q17              3.3985     3.372     1.6089     4.4172
Q18              3.2225    3.0756     1.3951     4.3906
Q19              3.2617    3.1381     1.4244     4.4104
Q20              3.2854    3.1951     1.4146     4.4172
Q21              3.3074      3.22     1.4553     4.4257
Q22              3.3175    3.2238      1.474     4.4379
Q23              3.2019    3.0279     1.3886     4.3911
Q24              3.1668    2.9755     1.3813     4.3591
Q25              3.3125    3.2204     1.4894     4.4199
Q26              3.2222    3.0709     1.4252      4.378
Q27              3.1548    2.9793        1.4     4.3132
Q28              3.3081    3.2226     1.4797     4.4113




Time taken to build model (full training data) : 0.08 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      2368 ( 41%)
1      1230 ( 21%)
2      2222 ( 38%)
```

Figure 3: Clusters captured by using the k-means Clustering algorithm.

DOI: 10.37200/IJPR/V24I3/PR200962
Received: 20 Jan 2020 | Revised: 06 Feb 2020 | Accepted: 12 Feb 2020

The cluster analysis of the course aims and objectives attribute as example of Q2 shows that centroid of full data is 3.0739 which is an average rating for this question on Likert scale. The centroid of cluster 0 is 2.8252 and 41% of the students fall in this cluster. The centroid of cluster 1 is 1.4024 and 21% of the students fall in this cluster. Furthermore, 38% of students in cluster 2 with centroid 4.2642, indicate an opinion that it is contrary to the cluster 1. This shows that majority of faculty members provided course aims and objectives at the start of the semester. The following Table 1 shows the centroids statistics of each cluster. In addition, all the questions of the dataset that included in the Table 1 are available in (Asuncion and Newman, 2007).

Table 1: Final cluster centroids statistics.

Attribute	Full Data	Cluster 0	Cluster 1	Cluster 2
No. of Instances	(5820.0)	(2368.0)	(1230.0)	(2222.0)
Instructor's identifier	2.4856	2.5608	2.5789	2.3537
Course code	7.2763	7.2107	7.5602	7.189
Number of times the student is taking this course	1.2141	1.2297	1.2496	1.1778
level of attendance	1.6756	1.6753	1.2073	1.9352
Level of course difficulty	2.7835	2.8898	2.5106	2.8213
Q1	2.9299	2.6263	1.3927	4.1044
Q2	3.0739	2.8252	1.4024	4.2642
Q3	3.1787	3.0013	1.5049	4.2943
Q4	3.0825	2.8476	1.4317	4.2466
Q5	3.1058	2.8644	1.3894	4.3132
Q6	3.1074	2.8965	1.4065	4.2736
Q7	3.0663	2.8121	1.3943	4.2628
Q8	3.0419	2.7817	1.3789	4.2399
Q9	3.166	2.962	1.5008	4.3051
Q10	3.0907	2.8315	1.3691	4.32
Q11	3.1838	2.9962	1.4659	4.3348
Q12	3.0356	2.7821	1.3805	4.2219
Q13	3.2428	3.1161	1.4098	4.3924
Q14	3.2909	3.2002	1.4374	4.4136
Q15	3.2873	3.1883	1.4569	4.4059
Q16	3.1696	2.9818	1.3528	4.3753
Q17	3.3985	3.372	1.6089	4.4172
Q18	3.2225	3.0756	1.3951	4.3906
Q19	3.2617	3.1381	1.4244	4.4104
Q20	3.2854	3.1951	1.4146	4.4172
Q21	3.3074	3.22	1.4553	4.4257
Q22	3.3175	3.2238	1.474	4.4379
Q23	3.2019	3.0279	1.3886	4.3911
Q24	3.1668	2.9755	1.3813	4.3591
Q25	3.3125	3.2204	1.4894	4.4199
Q26	3.2222	3.0709	1.4252	4.378
Q27	3.1548	2.9793	1.4	4.3132
Q28	3.3081	3.2226	1.4797	4.4113

Further, the classification method is used to predict effectively the course assessment of each academic instructor. By way of suggesting several data classification approaches, the method of Naïve Bayes as of linear classifier is applied in this research. The performance of the proposed methodology has been evaluated via 10-fold cross validation technique that it is usually used for the reason of ensuring further steady results. The Naïve Bayes classification algorithm was applied using its default parameters. The Naïve Bayes classification algorithm has been analyzed correctly with a very high accuracy result percentage of 98.866 %. When we experimented the dataset with the simple classification method of Naïve Bayes, the accuracy was at 84.24 %. However, after using our proposed methodology, Naïve Bayes showed 14.626% increase to boast the accuracy performance at 98.866 %. Given the nature of Naïve Bayes, this clearly demonstrates that the simple classification method does not serve as supporting information for building a classifier. By studing and observing respectively clusters; the analysis forms a table of defining the characteristics to each cluster and gives an evaluation among all clusters, as presented in Table 2. The Naïve Bayes classifier predicted the sensitivity of detection (True Positive Rate) for each cluster as the Cluster 0 was best cluster. The results obtained are shown in the (Table 2).

Table 2: the presented accuracy by cluster class

Class	TP Rate	FP Rate	Precision	Recall	F-Measure
Cluster 0	0.995	0.015	0.978	0.995	0.986
Cluster 1	0.977	0.000	0.999	0.977	0.988
Cluster 2	0.989	0.003	0.995	0.989	0.992
Weighted Avg.	0.989	0.008	0.989	0.989	0.989

The final evaluation procedure that conducted for the proposed methodology can be delivered by comparing and matching the results by using the F-measure values which can be identified as the harmonic mean of Precision and Recall. On the other hand, the precision can be recognized as the rate of correctly classified clusters among all classifier results. Furthermore, the recall is correspondingly well-defined as a measure of the rate of correctly classified clusters consistent with the other clusters to be classified correctly. Therefore, we have been selected the F-measure for the reason that both the precision and also the recall ratios are considered within the F-measure. The above (Table 2) shows the performance of F-measure in different classified clusters based on the proposed methodology. Accordingly, the performances of the F-measure in terms of classifying Cluster 2 is better than the other clusters. Thus, the performance of clustering by way of the Naïve Bayes-classifier is delivering a data mining task that realizing the strong points of student evaluation towards the academic instructor's performance to have its place in a single cluster to be more associated to each other rather than to a student evaluation that can be relevant to wide-ranging clusters.

IV. CONCLUSION

It was presented a classification framework for student evaluation by using K-means clustering algorithm and Naïve Bayesian classifier to predict the academic instructor's performance. This is to enhance the academic experience of instructors that would ultimately improve the quality of educational environment of an educational institution. All these and alike hidden patterns could serve as an important feedback for instructors, curriculum planners, academic managers, and other stakeholders in making informed decisions for evaluating and restructuring curricula. Besides, teaching and assessment methodologies with a view to improve the students' performance in their respective programs.

REFERENCES

[1] Gašević, D., Dawson, S., Rogers, T., &Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. The Internet and Higher Education, 28, 68-84.

[2] Mohammed, A. A. J., Burhanuddin, M. A., Basiron, H., & Tunggal, D. (2018). Key enablers of IoT strategies in the context of smart city innovation. J. Adv. Res. Dyn. Control Syst, 10(4)

[3] Malhotra, R. (2015). A systematic review of machine learning techniques for software fault prediction. Applied Soft Computing, 27, 504-518.

[4] Kotsiantis, S. B., Zaharakis, I. D., &Pintelas, P. E. (2006). Machine learning: a review of classification and combining techniques. Artificial Intelligence Review, 26(3), 159-190.

[5] Morris, A. M. (2016). The Implementation of an Accountability and Assessment System: A Case Study of Organizational Change in Higher Education.

[6] Romiszowski, A. J. (2016). Designing instructional systems: Decision making in course planning and curriculum design. Routledge.

[7] Pugh, E., &Aspray, W. (1996). A history of the information machine. IEEE Annals of the History of Computing, 18(2), 70-76.

[8] Razaque, F., Soomro, N., Shaikh, S. A., Soomro, S., Samo, J. A., Kumar, N., &Dharejo, H. (2017). Using naïve bayes algorithm to students' bachelor academic performances analysis. In 2017 4th IEEE International Conference on Engineering Technologies and Applied Sciences (ICETAS) (pp. 1-5). IEEE.

[9] Kezar, A., Chambers, A. C., &Burkhardt, J. C. (Eds.). (2015). Higher education for the public good: Emerging voices from a national movement. John Wiley & Sons.

[10] Ralph, M., & Stubbs, W. (2014). Integrating environmental sustainability into universities. Higher Education, 67(1), 71-90.

[11] Michaelson, D., & Stacks, D. W. (2017). A Professional and Practitioner's Guide to Public Relations Research, Measurement, and Evaluation. Business Expert Press.

[12] Taylor, B. K. (2016). Pre-service Teachers' Knowledge of Reading and Assessment for Providing Differentiated Instruction to Struggling Readers and How This Knowledge Relates to Their Perceptions for the Use of Retention (Doctoral dissertation).

[13] Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann.

[14] Hassan, A. A. H., &Iskandar, M. F. (2017). Clustering Methods for Cluster-based Routing Protocols in Wireless Sensor Networks: Comparative Study. Int. J. Appl. Eng. Res, 12(21), 11350-11360.

[15] Talib, M. S., Hassan, A., Hussin, B., Abas, Z. A., Talib, Z. S., &Rasoul, Z. S. (2018). A novel stable clustering approach based on Gaussian distribution and relative velocity in VANETs. IJACSA) Int. J. Adv. Comput. Sci. Appl, 9(4), 216-220

[16] Md Shah, W., Othman, M. F. I., Hassan, H., Abdul, A., Talib, M. S., & Mohammed, A. A. J. (2018). K nearest neighbor joins and mapreduce process enforcement for the cluster of data sets in bigdata. Journal Of Adv Research In Dynamical & Control Systems, 10, 690-696.

[17] Frank, E., & Hall, M. A. (2011). Data mining: practical machine learning tools and techniques. Morgan Kaufmann.

[18] Burhanuddin, M. A., Ismail, R., Izzaimah, N., Mohammed, A. A. J., &Zainol, N. (2018). Analysis of Mobile Service Providers Performance Using Naive Bayes Data Mining Technique. International Journal of Electrical and Computer Engineering, 8(6), 5153.

[19] Cohen, P., West, S. G., & Aiken, L. S. (2014). Applied multiple regression/correlation analysis for the behavioral sciences. Psychology Press.

[20] Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R., &LichtendahlJr, K. C. (2017). Data mining for business analytics: concepts, techniques, and applications in R. John Wiley & Sons.

[21] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. ACM SIGKDD explorations newsletter, 11(1), 10-18.

[22] Asuncion, A., & Newman, D. (2007). UCI machine learning repository.

DOI: 10.37200/IJPR/V24I3/PR200962
Received: 20 Jan 2020 | Revised: 06 Feb 2020 | Accepted: 12 Feb 2020