

Enhancing Educational Data Mining based ICT Competency among e-Learning Tutors using Statistical Classifier

Lalbihari Barik^{1*}, Ahmad AbdulQadir AlRababah², Yasser Difulah Al-Otaibi³

Department of Information Systems, Faculty of Computing and Information Technology in Rabigh^{1,3}
King Abdul Aziz University, Kingdom of Saudi Arabia^{1,3}

Department of Computer Science, Faculty of Computing and Information Technology in Rabigh²
King Abdul Aziz University, Kingdom of Saudi Arabia²

Abstract—The implementation of computer-supported collaborative learning has come to play a pivotal role in e-learning platforms. Educational Data Mining (EDM) is a promising area for the exclusive skill development of e-learning tutors, the major concern being investigations over large datasets. The tutors possessing efficient and sufficient soft skills can teach students within less time and with greater productivity. EDM is a regularly used research area that handles the development of methods to explore new ideas in the educational field. Computer-supported collaborative learning in e-learning and competencies on a real-time perspective among teachers are calculated using statistical classifiers. This paper aims to identify a feasible perspective on EDM based ICT competency over e-learning tutors using statistical classifiers. A set of tutors from diverse e-learning centers of various universities is selected for the evaluation purpose. The teachers from the department of mathematics in the universities are selected to attend a professional Qualified Teacher Status numeracy skills test and tutors' online test. The results of online tests are collected and correlated with the Naive Bayes Classifiers algorithms. Naive Bayes Classifiers are used in this paper to find the classification performance results among teachers. Naive Bayes based classification is beneficial for skill identification and improvement among the teachers. Significantly, the data mining classifiers performed well with the large dataset.

Keywords—Data mining; e-learning tutors; Naive Bayes Classifiers algorithms; ICT; QTS numeracy

I. INTRODUCTION

E-learning is a new universal teaching and learning model through the help of electronic resources. A basic teaching model only depends on the help of classrooms and blackboard. The classrooms and blackboard teaching reaches a certain amount of distance only. Other than the class, attendees cannot learn the taught lessons [1]. Computer and internet-based e-learning are introduced to avoid such discrepancies. E-learning is capable of conveying the skill and knowledge to a huge amount of learners at different places. Initially, e-learning is not accepted because they thought that the computer could not lead the human's intelligence in learning [2]. Compare with humans, much information are readily accessible by a single click. E-learning without any intelligent system is not at all possible. Computers are the shapeless information carriers

system, so provide any coherent knowledge; some intelligent methods are needed [3].

Data Mining is the best coherent knowledge-based intelligent method. Data Mining is the best field of modern technology in e-learning. Data Mining is the best field of modern technology in e-learning to avoid the inherent difficulty of teaching. Data Mining is a set of data testing methods with the collected data understanding, pre-processing, and reproduction procedure to valuation and implementation [4]. Data mining is an exclusive technique that gives individual concentration and adapts any new modern technique with Information Communication Technology (ICT) as well as the database based data mining technique usually deals with large, heterogeneous, and complex databases. Accordingly, the e-learning database is also well fitted with this description [5].

In large database cases, data mining is the finest procedure, which extorts and identifies the required information and consequent knowledge [6]. These extort and identifications are performed with statistical, mathematical, artificial intelligence, and machine learning techniques [7]. Data Mining is capable of extorting knowledge of the e-learning systems during the testing of information with large datasets. In this work, the most important aim is to discover the ICT skills among the teachers of e-learning centre tutors [8].

Educational Data Mining (EDM) is the best field for the ICT skill development over the e-learning tutors. Computer-supported e-learning and ICT skill among the teachers in real-time are calculated using statistical classifiers. A set of e-learning tutors from various areas is selected for validation purposes. A professional Qualified Teacher Status (QTS) numeracy skills test and tutors' online test based results of mathematical department teachers are selected for validation.

These results are collected and correlated with the Naive Bayes Classifiers algorithms. Naive Bayes Classifiers are used in this paper to discover the classification performance results among the teachers. Naive Bayes based classification is most helpful for skill identification and development amongst the teachers.

*Corresponding Author

II. LITERATURE SURVEY

The purpose of the data mining method in the e-Learning model is to support the teachers to increase the e-Learning situation. Data mining is the best method to find out valuable information through plenty of techniques such as prediction, classification, rule-based mining, and clustering [9]. DM is the best procedure to analyze the data in a sequence to find out hidden data that are discovering from the previously hidden patterns [10]. DM is a procedure that utilizes the statistical, mathematical methods to extort and recognize the valuable information is as of large databases [11]. Data Mining is used to extract the knowledge of the e-learning method throughout the investigation of the information obtainable in the form of user-generated data. In this work, the main aim is to find out the ICT skill among the e-learning teachers to discover the teachers learning behavior patterns [12].

The data mining technique is fruitfully integrated into the e-learning environment. The advantage of the data mining technique, as well as perception over the e-Learning system, helps to supports the teachers in developing their teaching skills. Data mining is the development of analyzing the input data sequentially to find out the helpful information based on the earlier anonymous data set [13]. Data mining in e-learning intends to afford an advanced picture of the present application of the data mining technique. To propose a practical arrangement of the data mining in e-learning, the collected dataset alone is not possible for the grade or any other criterions predictions. Based on the needed output criteria, the input data are processed with any soft computing technique [14]. The basic existing used techniques for the data processing are the Neural Networks, Genetic Algorithms, Clustering and Visualization Methods, Fuzzy Logic, Intelligent agent-based techniques. In data mining classification methods, the composed data are processed with any one of the following methods, such as clustering-based techniques, classification based techniques, and prediction based techniques. [15].

In the e-learning tutoring process, information and communication technology skills (ICT) are essential. It develops the educational institutions' standards, sustainability-related educational resources, and the selection priority amongst the students in selecting the particular institution [16, 29]. In the beginning, the ICT competency is improved by the Learning Resource Management courses [17]. After that, the teaching system improved to encompass ICT skills with a real-time environment. Nowadays, the world is attractive towards the digitalized, so the teaching and learning methods also enhanced gradually based on digital techniques. In e-learning systems, the ICT skillful teaching methods improve the students learning capacity [18]. ICT in e-learning has many advantages such as easy understanding, easy access to the study materials, time-saving, and book free study [19].

ICT proficiency is necessary to originate their thought of information over some electronic gadgets. Proficiency gives you an idea about the information to locate, evaluate, and use the needed information effectively to teach their subjects [19]. ICT skills are taken into account for the skillful usage of hardware and software, computer system association configuration [20]. Furthermore, the ICT is included with the

logical capability with the necessary concept and ability about hardware and software application for efficient utilization of information communication technology [21]. To identify the skills of the tutors are necessary to identify their ranks or ranges. The classification is very much essential to predict the student's ranks. The best classifier is the Neural Network (NN) based statistical classifiers [22].

NN is used to solve real-time statistical problems. NN resolves considerable complicated areas such as business, real estate, education, medicine, and pattern recognition [23]. The reason is that the statistical classifier makes the illustration of the nonlinear function map among the input to output. The statistical classifier has trained through powerful as well as computational efficient technique call backpropagation. The statistical classifier has trained through powerful as well as computational efficient technique call backpropagation [24].

In this work, a decision supported system for ICT skill-based classification on Sequential Minimal Optimization (SMO) in Support Vector Machine (SVM). The SVM based SMO algorithm is designed for testing the ICT skills of the e-tutors in the training process of this model [26, 28]. SVM based SMO algorithm is used for its high accuracy and high speed. The reason is the ease of use and better scaling with the training set with a large dataset. SMO is for evaluating the multipart correlation among input and output of the model [27]. The evaluation results are compared with the SVM based SMO algorithm results. NN results are demonstrated, and intelligent assessments are made by various researchers based on predicting the performance of various departments.

EDM in teaching and learning programs NN has been used to compute the students' performance assessments and monitor them using the Classifiers algorithms [24, 25]. Naive Bayes Classifiers algorithms can calculate the performances of students in the time of admissions to various universities in recent times. The results specify that the Classifier algorithm is potentially enhancing the efficiency of the prediction accuracy of teachers' ICT skills. In this work, a multilayer perceptual recurrent neural network is used with backpropagation learning is utilized. Naive Bayes Classifiers algorithms are used in this study to calculate the ITC among the teachers of the distance education center. The computer science department teachers are recommended to obtain the information and communication technology competency for their education.

III. METHODS AND MATERIALS

EDM is a talented field of skill development of e-learning tutors, the major concern being investigations over large datasets. The tutors possessing efficient and sufficient soft skills can teach students within less time and with greater productivity. EDM is a regularly used research area that handles the development of methods to explore new ideas in the educational field. Computer-supported collaborative learning in e-learning and competencies on a real-time perspective among teachers are calculated using statistical classifiers. This paper aims to identify a feasible perspective on EDM based ICT competency over e-learning tutors using statistical classifiers. A set of tutors from diverse e-learning centers of various universities is selected for the evaluation purpose. The teachers from the department of mathematics in

these universities are selected to attend a professional QTS numeracy skills test and tutors' online test. The results of online tests are collected and correlated with the Naive Bayes Classifiers algorithms. Naive Bayes Classifiers are used in this paper to find the classification performance results among teachers. Naive Bayes based classification is beneficial for skill identification and improvement among the teachers. Significantly, the data mining classifiers performed well with the large dataset. The e-learning data mining process consists of different steps in the general data mining process is shown in Fig. 1.

A. Educational Data Mining

EDM is a promising area for discovering the needed predictions from the exclusive type of data sets. EDM is very much helpful for the educational skill-based classifications and the settings of the results based on the classifications. An input of EDM in this work is mining the ICT skills of the e-learning centre tutors. The input data usages of EDM take account of calculating the teacher's performances, as well as lagging skills, which are advised to improve based on the present learning system. In this work, the data mining uses for the investigation and idea of data provides the teaching advice for supports and improves the teacher's skills. This skill improvement brings so many eager learning students to a particular centre.

The first step in EDM is the collection of data. The LMS system is used for collecting, and the interaction information is stored in the database. In this work, we are going to use the e-learning centre teacher's data set. The professional QTS numeracy skills test and tutors' online test information are stored in the database of the DM LMS. In the pre-processing step, the data are arranged in a suitable arrangement to be mined. The database administrator pre-processing tool is used to pre-process the e-learning centre teacher's data set. Then the data are stacked up with an appropriate format.

B. Processes of Data Mining in e-Learning

In the e-learning domain, the learners, as well as the teachers, are the soul of the system. The data mining in e-learning is a comfortable area for the special skill development of e-learning center tutors; the major concern is being investigated over large datasets. In e-learning tutors, efficient and sufficient soft skills can teach students within less time and with greater productivity. EDM is a frequently used research field that handles the development of methods to explore new ideas in the teaching-learning area. Computer-supported collaborative learning in e-learning and ICT competencies on a real-time perspective among teachers are calculated using statistical classifiers. This paper aims to recognize a practical perspective on EDM based ICT competency over e-learning tutors using the statistical classifier. E-learning center teachers of various universities are selected for the evaluation purpose. The e-learning center teachers of mathematics in various universities are selected to attend a professional QTS numeracy skills test and tutors' online test.

Data mining has the site records information such as gathering the teacher's profile, real-time accessed information, academic performances of teachers, and the estimation result. Data mining with any online test based result prediction is the main pathway to skill identification and user behaviors. The online test based data mining does much help in the e-learning field to provide real-time instructions for e-learners. Then construct the construct based on the teacher's field of interest. Then build and right to use these fields of interested ICT skills to improve and can implement for the teaching. The results are updated to the data mining server based on the earlier practiced information. Finally, the identified group of teachers of the comparable fields of interest and sending personalized information to the individuals. The proposed e-learning data mining procedure contains three steps are shown in Fig. 2.

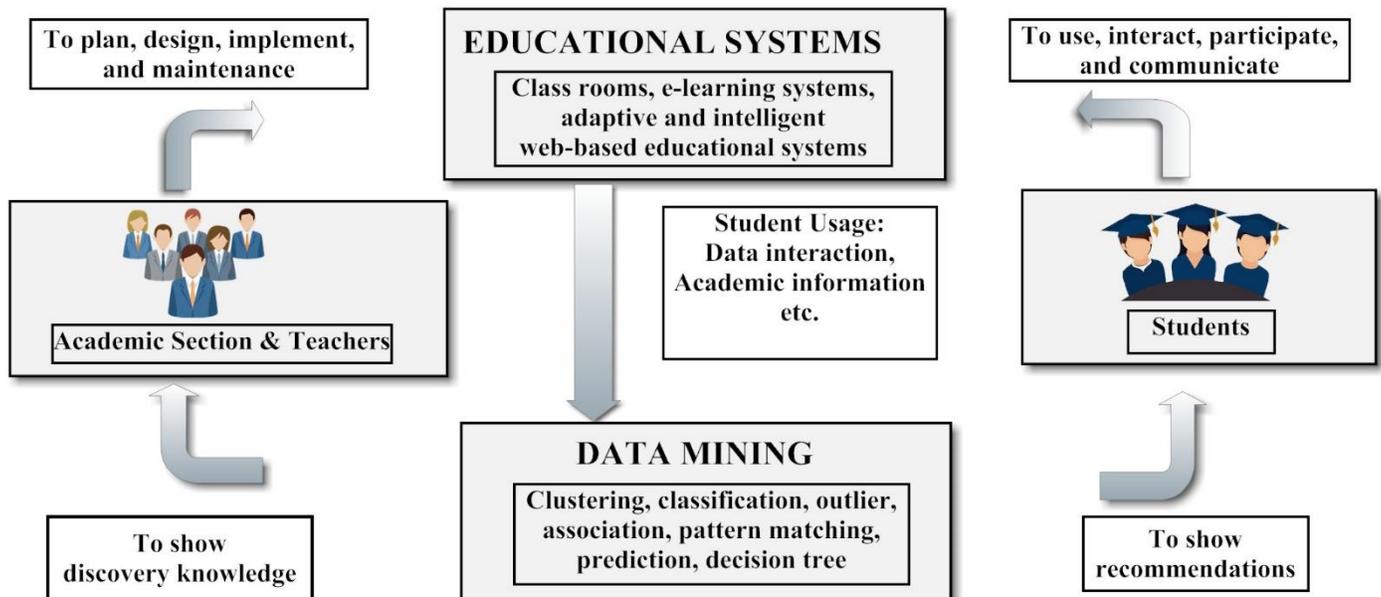


Fig. 1. Block Diagram of Data Mining in the e-Learning System.

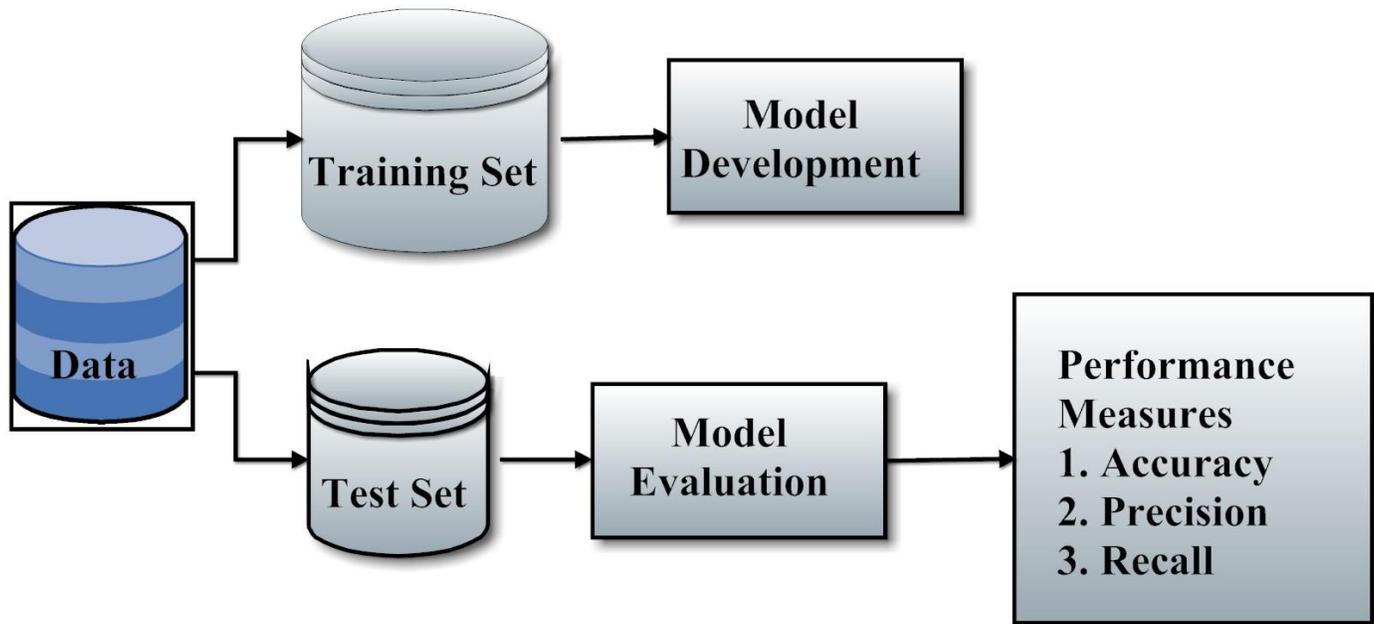


Fig. 2. Data Mining Systems Component.

In this work, data mining with the e-learning development system is iteration based ICT skill of teacher’s prediction. Then the minced information has to form some loop of information for classification along with enhanced their teaching skills. Not simply turn the data into relevant knowledge also pass through a filter the mined information for finding the ICT skills among the teachers. The main purpose of data mining in this work is to find the teachers' exact computer skills and their performance together with pedagogical aspects.

C. Naive Bayes Classifier

The Naive Bayes classification algorithm used to construct and implement the data to discover and summarize the information needed for prediction. The data mining algorithm is useful to generate and implement the method that finds out the information and pattern for the ICT skill test. The data mining based classifier is used to find the ICT competency over e-learning centre teachers in this work. The results of the classifiers are used to find the teachers’ skills with ICT and skill improvement. The professional QTS numeracy skills test and tutors’ online test results are used for deciding the teachers’ ICT skill percentage and ICT learning ability. The discovered information are beneficial for the teachers to improve their e-learning system and process. Fig. 3 shows the class-based Naive Bayes Classifier.

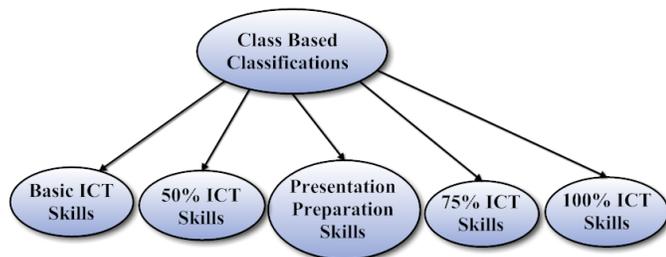


Fig. 3. Naive Bayes Classifier.

The Naive Bayes Classifier belongs to the family of probability classifier with the use of the Bayesian theorem. The reason why it is called ‘Naive’ because it requires a rigid independence assumption between input variables. Therefore, it is more proper to call Simple Bayes or Independence Bayes. Naive Bayes Classifier is the best method to solve the ICT skill categorization problems. In this work, the preprocessed data are characterized and featured as classes. Based on the results of the skill-based tests, the classifications are performed. The whole data are divided into subclasses such as Basic ICT skills as (0), 50% ICT skills as (1), Presentation Preparation skills (2), 75% ICT skills (3), 100% ICT skills (4). The problem of classifying the class-based probability cases belonging to one category or the other. The goal of the Naive Bayes Classifier is to calculate the conditional probability of the teachers’ skill test results, i.e., $P(O_s | Y_1, Y_2, \dots, Y_N)$. For every value of $P(O_s)$, the possible outcomes of classes are O_s . Where ‘s’ is the possible outcomes. Let $Y = (Y_1, Y_2, \dots, Y_n)$. Using Bayesian theorem, we can get:

$$P(O_s|X) = P(O_s)P(O_s|X)|p(O) \alpha P(O_s)P(O_s|X) = P(O_s | Y_1, Y_2, \dots, Y_N) \quad (1)$$

Then the joint probability values can be written as from can be written in equation (2).

$$P(O_s, Y_1, Y_2, \dots, Y_N) \quad (2)$$

$$P(Y_1, Y_2, \dots, Y_N, O_s) = P(Y_1|Y_2, \dots, Y_N, O_s) \cdot P(Y_1, Y_2, \dots, Y_N, O_s) \quad (3)$$

$$= P(Y_1|Y_2, \dots, Y_N, O_s) \cdot P(Y_2|Y_3, \dots, Y_N, O_s) \cdot P(Y_3, Y_4, \dots, Y_N, O_s) \quad (4)$$

$$= P(Y_1|Y_2, \dots, Y_N, O_s) \cdot P(Y_2|Y_3, \dots, Y_N, O_s) \cdot P(Y_3|Y_N, O_s) \cdot O_s \quad (5)$$

Assume that all features Y are mutually independent, we can get:

$$P(Y_1, Y_2, \dots, Y_N, O_s) = P(Y_1|O_s) \quad (6)$$

Therefore, the formula can be written as:

$$P(O_s|X_1, X_2, \dots, X_N) \propto P(O_s, Y_1, Y_2, \dots, Y_N) \quad (7)$$

$$P(Y_1, Y_2, \dots, Y_N, O_s) = P(Y_1|O_s) \cdot P(Y_2|O_s) \dots P(Y_N|O_s) \cdot P(O_s) \quad (8)$$

$$P(Y_1, Y_2, \dots, Y_N, O_s) = P(O_s) \prod_{i=1}^N P(Y_i|O_s) \quad (9)$$

This is the final Naïve Bayes Classifier formula for classification. The class-based predictions are used to calculate parameters and predict the Naive Bayes Classifier. LMS is used to estimate parameters prior to probability and conditional probability.

$$P(O_s) = P(Z = O_s) = \frac{\sum_{x=1}^M Q(Z_i=O_s)}{M} \quad (10)$$

The prior probability equals the number of certain cases of y occurs divided by the total number of records.

$$P(X_1 = B_k|z = O_s) = \frac{\sum_{x=1}^M Q(Y_{1i}=B_k, Z_i=O_s)}{\sum_{i=1}^M (Z_i=O_s)} \quad (11)$$

The conditional probability of $P(Y_1=B_1|y=C_1)$ equals the number of cases when Y1 equals to a1 and y equals to C1 divided by the number of cases when y equals to C1. Naive Bayes Classifier uses the following formula to make a prediction:

$$Z = \text{arg max } P(Z=O_s) \prod P(Y|Z = O_s) \quad (12)$$

IV. RESULTS AND DISCUSSION

ICT learning in e-learning and competencies on real-time perspectives among teachers are calculated using the Naive Bayes statistical classifiers. This paper aims to recognize a practicable point of view on EDM based ICT competency over e-learning tutors using statistical classifiers. The tutors from different e-learning centers of various universities are selected for the evaluation purpose. The teachers from the department of mathematics in the universities are chosen to grace with your presence a professional QTS numeracy skills test and tutors' online test. Then the online test are collected and correlated with the Naive Bayes Classifiers algorithms. Naive Bayes Classifiers are used in this work to calculate the classification performance consequences among teachers. Naive Bayes based classification is beneficial for skill classification and development among the teachers. Extensively, the Naive Bayes statistical classifiers performed well with the large dataset.

Table I contains the basic technical aspects particulars of the e-learning center teachers. The teachers of various universities are collected through the internet. Different ICT competencies are investigated and their outcomes are tabulated. In these ICT skill tests, the major three criterions selected are the online test based results, educational qualification, and the collected feedback from the students. From the above table, the ICT skills of the teachers are not only dependent on their educational qualifications. The ICT skills solely depend on the teachers' ICT skills. So the e-learning teachers must have

proper ICT skills for e-learning tutoring. Student's feedback on the subject of teacher's presentation percentage are tabulated on the average results. The data of teachers are required to be examined separately to find their ICT skill-based online test. The manual evaluation of the teacher's ICT skill test consumes more time.

The tests are validated by the online test to avoid these kinds of discrepancies. Naive Bayes Classifiers algorithm is necessary to analyze the teacher's data and evaluate their performance.

In the e-learning system, the logical assessments amongst the teachers are performed by the class-based statistical classifier. Then the performances of the teacher's performances over ICT skills are done with the professional QTS numeracy skills test and tutors' online test. The presentation skills and the range of the ICT skills of teachers are classified with the class labels. The ICT skill test data are saved in the DM server. Then the ICT skill test data are arranged based on the performance criteria. Finally, the classification is executed with the five different classes based on the ICT skills over mathematical tests. After the class-based label formations, the results are classified by their performances. The performances of the teachers are selected as excellent, good, medium, normal, and bad performances. Naive Bayes class is a straightforward and high-speed classification algorithm, so it is fit for large datasets. Naive Bayes classification is working with the principle of Bayes theorem of probability for prediction of different classes. Fig. 4 shows the proposed Naive Bayes Classifier.

TABLE. I. BASIC TECHNICAL CHARACTERISTIC PARTICULARS

Basic Technical Aspects	Online Test Results (%)	Required Educational Qualification	Performance Feedback
1.a. Basic ICT skills	83%	UG degree, PG degree, Trainees	75%
1.b. Notes downloading ability	85%	PG degree	71%
2.a. Typing skill ability	82%	PG degree	69%
2.b. Media communication ability	84%	Trainees	74%
3.a. Powerpoint preparation ability	87%	UG degree	82%
3.b. Teleconferencing ability	71%	UG degree	80%
4.a. Online test conduction ability	64%	UG degree	68%
4.b. Database management ability	54%	UG degree, PG degree, Trainees	61%
5.a. Computer-Aided animation presentation ability	44%	UG degree, PG degree, Trainees	78%
5.b. Computer-Aided equation making ability	42%	UG degree, PG degree, Trainees	72%

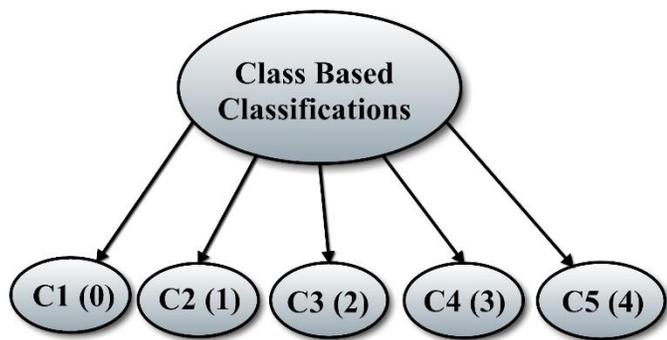


Fig. 4. Proposed Naive Bayes Classifier.

In classification, the initial step is finding the classes which are needed to classify and label them based on performances. Features are characterized with the five classes such as basic ICT skill as C1(0), 50% ICT skills as C2(1), Presentation Preparation skills C3(2), 75% ICT skills C4(3), 100% ICT skills C5(4). This distinctiveness is called features. It is a very useful method to classify teachers. This statistical classification has two input data sets; one is from the online test, and the one is from the professional QTS numeracy skills test. In the testing and training phases, classifiers are used for the evaluation performance. The datasets are divided into test and training set. The validation parameters are validated based on the diverse parameter such as accuracy, precision, and sensitivity.

In the classification part, the ICT skills are classified with the five classifiers that are checked with the depending or not depending based features. If the depending or not depending features are mutually dependent means, it is considered as an independent feature category. This assumption is very important because it reduces computation discrepancy and reduces time consumption. The mutually dependent supposition is named as the class conditional independence. The following equation is the final classifier class-based independence calculation of posterior probability.

$$P(C|F) = \frac{P(F|C)P(C)}{P(F)} \quad (13)$$

In this work, the classes are based on the ICT skills of the teacher. The probability conditions are calculated based on whether the teachers' skills are in C1(0) or elsewhere. The uses of probability conditions for checking whether the skills are depending only on the 1(a, b). If the case is only on 1(a, b), the

class is C1(0). The second case is considered on both skills 1(a, b), 2(a, b) skills are present, means the class is C2(0). The probability P (C|F) is used for the probability finding of the three cases, such as 50% skills, 75% skills, and 100% skill predictions. Which the input skill depends on the second class C2(0) means P (C|F) wants to find the probability of whether the teacher has both the 1(a, b) and 2(a, b) skills. This type of probability that is used to find the priority of the input skill, which depends on the most probable class (F). 'P(C)' is the probability of the skill which is not needed for our prediction or out of our five classes C1(0), C2(1), C3(2)... After the class-based probability predictions, the posterior probability output P (C|F) of our five different labeled classes are predicted. The posterior probability is used for final assumptions of the true class finding of the input skill test data of 'F' given that the class 'C'. The Naive prior probability is used for the assigned five class labels findings without any statistical error. Furthermore, discover the possibility of the class with every characteristic belonging without any error. Then substitute all the predicted class-based probability value in Bayes Formula and calculate posterior probability.

From Table II classes with high probability from the input, value belongs to the high probability classes. In prior and posterior probability calculation, an equation is designed to simplify the probability calculation methods.

Table II is intended to evaluate the posterior probability value. The Posterior probability table encloses the happening of class labels for all characteristics. In the testing and training phases, classifiers are used for the evaluation performance. The datasets are divided into test and training set. Fig. 5 shows that the evaluation parameters are validated based on the diverse parameter such as accuracy, precision, and sensitivity.

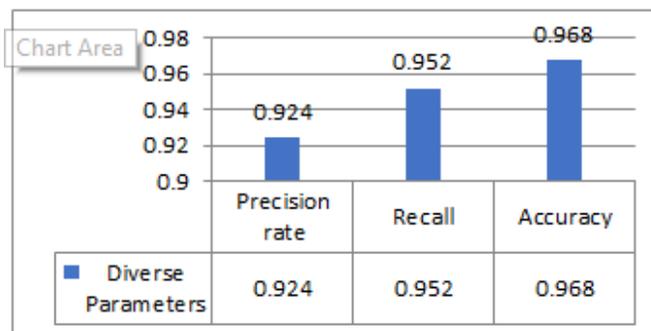


Fig. 5. Evaluation Results Graph.

TABLE. II. POSTERIOR PROBABILITY

Number of class-based/cases	Basic ICT skills as C1(0)	50% ICT skills as C2(1)	Presentation preparation skills C3(2)	75% ICT skills C4(3)	100% ICT skills C5(4)	Posterior probability $P(C F) = \frac{P(F C)P(C)}{P(F)}$
0.2 (only 1class dependency)	Yes	No	No	No	No	0.00275
0.4 (only 2class dependency)	Yes	Yes	No	No	No	0.0042
0.6 (only 3class dependency)	Yes	Yes	Yes	No	No	0.05920
0.8 (only 4class dependency)	Yes	Yes	Yes	Yes	No	0.7934
1 (all 5classes are dependent)	Yes	Yes	Yes	Yes	Yes	0.910

Precision rate is calculated by the ratio between the correctly labeled positive probable classes among the labeled five classes C1 (0), C2 (1), C3 (2), C4 (3), C5 (4) with the total number of positive classes. Accuracy is the value calculated by the ratio between the fraction of predicted classes among the labeled five classes C1 (0), C2 (1), C3 (2), C4 (3), C5 (4) with the total number of predicted classes. Classification accuracy is the performance evaluator which is used to find the overall performances. Accuracy is the ratio between the total number of accurate class C1 (0), C2 (1), C3 (2), C4 (3), C5 (4) prediction with the total number of overall class predictions done. In the machine learning environment, accuracy, precision, and sensitivity values are essential because these values only showcase our performances over other methods. In this work, the attained accuracy, precision, and sensitivity values are 0.924, 0.954, and 0.968, respectively.

Finally, the results are compared with the SMO of the SVM method. A comparatively naive classifier provides the best performances over SVM because the SVM is a decision-based system only. Initially, the data are tested with the SVM based SMO algorithm for finding the accuracy, precision, and sensitivity. The increased results of the Naïve over SVM are accuracy is over 0.114, precision is increased around 0.09, and sensitivity is over 0.131 values. It shows that our model attained a very standalone performance over SVM. SVM based SMO algorithm is generally used for its high accuracy and high-speed performance. From Fig. 6, we can say that the Naive is performed well over the SVM method. The reason that is Naive can ease of use and better scaling with the training set with large datasets. The best part is the Naive algorithm is executed with less time on large datasets as well as the high accuracy and high-speed performances.

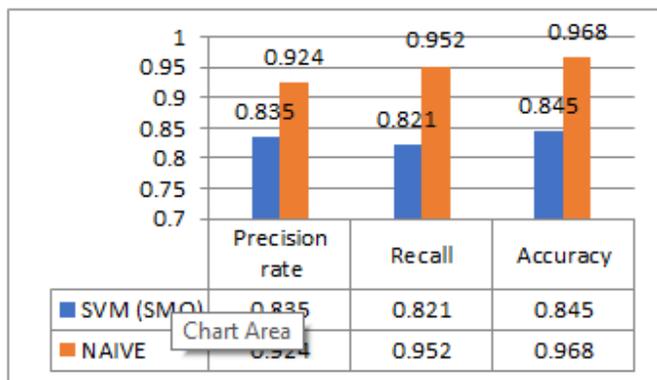


Fig. 6. Comparison with the SMO of the SVM Method.

V. CONCLUSION

E-learning tutors must have ICT skills over other teachers because they are teaching with the help of computers. The Naive Bayes algorithm is used for the skill prediction because of its simplicity and accuracy over prediction. The most important apprehension is the efficiency over large datasets. Also, this method is well suited to multiple class prediction models. Comparatively, the Naive classifiers perform well over with the logistic regression-based models. The teachers' ICT skills are necessary to teach students within less time and with greater productivity. Computer-supported learning in ICT competencies on a real-time perspective among teachers are

calculated with the class-based classifier for efficiency improvement. In this work, a feasible perspective of the posterior probability-based statistical classifier is used for error-free prediction. The teachers from the department of mathematics are not fluent in computer skills. To improve them by knowing them with their result range is very easy to give them training regarding soft skills. Naive Bayes based classification is an appropriate skill identification and improvement among the teacher with large datasets.

Further, the dependent and independent features are taken into account. In practice, individual dependent and independent model alone model statistical predictions are not possible because the prediction depends only on both the features only. The necessities of predictors are must be independent features. According to the real-time case, the predictors are only dependent only on the performance of the classifier. Additionally, classification accuracy can also improve through the use of noise elimination techniques to get rid of an outlier in data sets.

ACKNOWLEDGMENT

The author thanks King Abdulaziz University for this work.

REFERENCES

- [1] Anduela Lile, "Analyzing E-Learning Systems Using Educational Data Mining Techniques," *Mediterr. J. Soc. Sci.*, vol. 2, no. 3, pp. 403-419, 2011. doi: <http://dx.doi.org/10.5901/mjss.2011.v2n3p403>.
- [2] F. Castro, A. Vellido, À. Nebot, and F. Mugica, "Applying Data Mining Techniques to e-Learning Problems," *Studies in Computational Intelligence (SCI)* vol. 62, no. 221, pp. 183-221, 2007.
- [3] Romero, Cristobal, and Sebastian Ventura, eds. "Data mining in e-learning," WIT Press, Vol. 4, 2006.
- [4] L. Jiang, H. Zhang, and Z. Cai, "A Novel Bayes Model : Hidden Naive Bayes," *IEEE Transaction on Knowledge and Data Engineering*, vol. 21, no. 10, pp. 1361-1371, 2009. doi:<http://dx.doi.org/10.1109/TKDE.2008.234>.
- [5] S. Taheri and M. Mammadov, "Learning the naive bayes classifier with optimization models," *Int. J. Appl. Math. Comput. Sci.*, vol. 23, no. 4, pp. 787-795, 2013. doi:<http://dx.doi.org/10.2478/amcs-2013-0059>.
- [6] S. Lawe and R. Wang, "Optimization of Traffic Signals Using Deep," *AI 2016 Adv. Artif. Intell. AI 2016. Lect. Notes Comput. Sci.*, vol. 9992, pp. 403-415, 2016. doi:http://dx.doi.org/10.1007/978-3-319-50127-7_27.
- [7] W. Zhang and F. Gao, "An improvement to naive bayes for text classification," *Procedia Eng.*, vol. 15, pp. 2160-2164, 2011. doi:<http://dx.doi.org/doi:10.1016/j.proeng.2011.08.404>.
- [8] S. Banga, S. Mongia, S. Dhotre, and I. Introduction, "Regression And Augmentation Analytics on Earth ' s Surface Temperature," vol. 5, no. 3, pp. 17-19, 2017.
- [9] X. Wu et al., " Top 10 algorithms in data mining," Springer-Verlag London, vol. 14, no. 1. 2008. doi:<http://dx.doi.org/doi:10.1007/s10115-007-0114-2>.
- [10] A. Choi, N. Tavabi, and A. Darwiche, "Structured features in naive bayes classification," *30th AAAI Conf. Artif. Intell. AAAI*, 2016, pp. 3233-3240, 2016.
- [11] Zhang H., "The Optimality of Naive Bayes," 2004. American Association for Artificial Intelligence (www.aaai.org). 2004.
- [12] T. Calders and S. Verwer, "Three naive Bayes approaches for discrimination-free classification," *Data Min. Knowl. Discov.*, vol. 21, no. 2, pp. 277-292, 2010. doi:<http://dx.doi.org/doi:10.1007/s10618-010-0190-x>.
- [13] Laney D., "3D data management: Controlling data volume, velocity and variety," *META group research note*. 2001 Feb 6; 6(70):1, 2001.

- [14] R. Y. M. Li and H. C. Y. Li, "Have housing prices gone with the smelly wind? Big data analysis on landfill in Hong Kong," *Sustain.*, vol. 10, no. 2, pp. 1–19, 2018. doi:<http://dx.doi.org/doi:10.3390/su10020341>.
- [15] Boyd, Danah and Crawford, Kate, "Six Provocations for Big Data," *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*, September 2011. Available at SSRN: <https://ssrn.com/abstract=1926431> or <http://dx.doi.org/10.2139/ssrn.1926431>.
- [16] K. Swan and L. F. Shih, "on the Nature and Development of Social Presence in Online Course Discussions," *Online Learn.*, vol. 9, no. 3, pp. 115–136, 2019.
- [17] Segaran, Toby, and Jeff Hammerbacher, "Beautiful data: the stories behind elegant data solutions," O'Reilly Media, Inc., p. 257, 2009.
- [18] S. O. Material, S. Web, H. Press, N. York, and A. Nw, "The World ' s Technological Capacity," vol. 60, no. 2011, pp. 60–66, 2014. doi: <http://dx.doi.org/10.1126/science.1200970>. PMID 21310967
- [19] P. Larrañaga, H. Karshenas, C. Bielza, and R. Santana, "A review on evolutionary algorithms in Bayesian network learning and inference tasks," *Inf. Sci. (NY)*, vol. 233, pp. 109–125, 2013. doi:<http://dx.doi.org/10.1016/j.ins.2012.12.051>.
- [20] M. E. Maron, "Automatic Indexing: An Experimental Inquiry," *J. ACM*, vol. 8, no. 3, pp. 404–417, 1961. doi: <http://dx.doi.org/10.1145/321075.321084>.
- [21] Rennie, J.; Shih, L.; Teevan, J.; Karger, D., "Tackling the poor assumptions of Naive Bayes classifiers," *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)*, Washington DC, 2003.
- [22] Rish, Irina. "An empirical study of the naive Bayes classifier." *IJCAI 2001 workshop on empirical methods in artificial intelligence*. Vol. 3. No. 22. 2001.
- [23] Russell SJ, Stuart J. Norvig. *Artificial Intelligence: A Modern Approach*. 2003:111-4.
- [24] R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms," *ACM Int. Conf. Proceeding Ser.*, vol. 148, pp. 161–168, 2006.
- [25] Keerthi S. S., Shevade S. K., Bhattacharyya C., and Murthy K. R. K., "Improvements to Platt's SMO Algorithm for SVM Classifier Design," *Neural Computation*, 13: 637-649, 2001.
- [26] S. K. Shevade, S. S. Keerthi, C. Bhattacharyya, and K. R. K. Murthy, "Improvements to the SMO algorithm for SVM regression," *IEEE Trans. Neural Networks*, vol. 11, no. 5, pp. 1188–1193, 2000.
- [27] N. Matic, I. Guyon, L. Bottou, J. Denker, and V. Vapnik, "Computer aided cleaning of large databases for character recognition," *Proc. - Int. Conf. Pattern Recognit.*, vol. 2, pp. 330–333, 1992.
- [28] Ali Z, Shahzad SK, and Shahzad W. Performance analysis of support vector machine based classifiers. *International Journal of Advanced and Applied Sciences*, 5(9): 33-38, 2018. doi:<http://dx.doi.org/doi:0.21833/ijaas.2018.09.007>.
- [29] C. Jensen, M. Kotaish, A. Chopra, K. A. Jacob, T. I. Widekar, and R. Alam, "Piloting a Methodology for Sustainability Education: Project Examples and Exploratory Action Research Highlights," *Emerg. Sci. J.*, vol. 3, no. 5, pp. 312–326, 2019. doi:<http://dx.doi.org/10.28991/esj-2019-01194>.