International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-4, November 2019

Predicting Faculty Performance in Higher Education using Machine Learning

T. Manjunath Kumar, R. Murugeswari

Abstract*: Higher education is witnessing significant change in transferring of knowledge and experience from the faculty to the student community on a large scale. Domains requiring a high degree of skill development such as engineering technology education need a place in the university of deeper faculty communication and information percolation process. Despite various measurement and accounting system are prevailing in the current educational system. Known advanced methods for automated evaluation of faculty performance has emanated successfully yet. Here we propose a Machine Learning technique which can interact both with faculty and students to collect feature-oriented parameters of the academic processes prevailing in engineering domain to qualitative and quantitatively access the faculty performance in the modern ways. By using various machine learning methods such as Support Vector Machine, Logistic Regression, Naïve Bayes, Random Forest and Deep learning are used to build the models. The proposed model could identify the faculty's performance who are likely to improve the course teaching for the students to enhance the teaching – learning process. Model accuracy is used to evaluate the performance of the faculty in the course. To identify the Machine learning algorithm performance F1-Measure and Area under Curve (AUC) value are compared. Evaluation result indicates that Random Forest algorithm is best suitable algorithm for faculty performance prediction for course evaluation through feedback questionaries' given by the students. It has a 73.1% higher accuracy, 78.7% F1- measure and 0.73 value of AUC when compared to other models. This model is used for decision making support in the higher educational university.

Keywords: Machine Learning, Performance Prediction, Feature Selection, Feedback questionnaires, Decision making, Course evaluation.

I. INTRODUCTION

Performance prediction of the faculty plays a major role in providing high quality education to the student. In higher education, the faculty are important resource to ensure the student success. There is huge amount of data generated in the institute which are raw data. These data are collected using the feedback questionnaires. To convert the raw data into useful decision-making data, a machine learning model is needed. The model will provide immediate feedback to the student and faculty about academic performance. Feedback is designed to improve the faculty performance, which will measure the teaching-learning process, teaching-evaluation process and overall performance of the faculty [1]. Feedback is the instructional process in which institute can monitor the faculty performance on what is being learned from the student. Feedback is crucial for improving learning, which provide information to faculty and classify what good performance is.

Nowadays the educational institutes are facing too many problems in providing standard education to the students. So, it undertakes the process of allocating its resources to their right position or field and some of these resources are allocated in a manual way. It adds more effort and consume the HR resources in this process due to huge number of students and faculty in the institutes. The educational institutes consider students and faculties as their main assets and they look forward to improving their key process indicators by effective and efficient use of their assets [2]. But unfortunately, little attention has been paid to rethinking the use of existing institutional resources -such as faculties- the most important and expensive resource in the academic institutes.

One of the primary objectives of the faculty allocation process is to maximize students ' learning capacity by allocating the finest faculty to the correct courses based on the qualifications, skills and abilities of the teacher. The educational institutes should be able to handle resources like faculty, lecture times tables, students, courses and topics in order to accomplish the prior objective. This ability depends – to some extent – on how to efficiently use existing data on these resources. For many exciting and difficult data mining apps, the current information in the scholarly and educational domain provides a fertile ground. These requests can assist administrators and teachers at academic universities / institutes in improving the quality of the teaching process.

Data mining is the implementation of particular algorithms for pattern extraction and the process involving the use of advanced information analysis methods to find earlier unknown, valid features and relationships in huge information set. With the rapid growth of academic information and the much-needed decision-making support, information mining embedded in academic systems will be in great demand over the coming years. The data mining offers a lot of methods and techniques to be benefited from the huge amount of existing data in the academic institutes and schools.

Manuscript published on November 30, 2019.
* Correspondence Author
T. Manjunath Kumar*, Computer Science and Engineering, Kalasalingam Academy of Research and Education, KrishnanKoil, India. Email: manjunathkumar.t@klu.ac.in
Dr.R. Murugeswari, Computer Science and Engineering, Kalasalingam Academy of Research and Education, KrishnanKoil, India. Email: r.murugeswari@klu.ac.in

Using data mining on analyzing academic related data and extract knowledge from it to support the decision making and its application [3]. "It is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better students understanding, and the settings which they learn in".

Machine Learning can play an important role in solving the lecturer recommendation in the academic institutes and universities .A systematic review [4] and Case study[5] of ML used the different methods and techniques such as, classification, association role and prediction depending on the history data available at the Management Information Systems (MIS) as a source, the output of applying data mining on this data can be used as a guide or heuristic base on the academic decision making process.

In this work, to enhance the level of teaching quality in higher education. The performance of teachers is compared with the dataset collected from university in the south tamilnadu of India. It contains some attributes about faculties characteristics made of student course assessment questionnaire and using these we predict the performance of new teacher when he begins to teach a new semester.

The rest of the paper is structured into 5 sections. In section II presents related works. Section III contains methodology. In section IV gives the performance evaluation and implementation result, insights about future work are included. Finally, section V includes the conclusion and future work.

II. RELATED WORKS

In this paper, different related works are studied and investigated. It is divided into two categories. In category A, we will give some related works about using machine learning techniques in human performance prediction in general. In category B, we will present some related works about faculty's performance using machine learning techniques. Finally, we will conclude about this.

A. *Performance prediction using Machine Learning techniques*

In this section, we present some related works that use educational data mining techniques in predicting performance of people in general. We try to focus on performance in the field of education to be closer to our problem.

In [6], Used two specific techniques of data mining known as " Association rule discovery " and " Classification." The attributes Q1 to Q12 are chosen to perform this evaluation. REPTree classification algorithm provide the best result for the selected data with 84.26% and 0.772 are Correctly Classified features and F1 measure respectively. The work is related to our work because the problem is a classification problem. Some differentiations between this study and our work, 1- our goal is predicting the performance of new instructor not to determine exist instructor to teach the course again or not. 2- This study was focused on student side by predicting student performance in a course if he will fail or pass.

In [7], Predicted software project participant output based on previous system or training sets. Techniques for classification such as ID3, CART and C4.5 are used. It was previously assumed that the performance was highest for candidates with a good college average. This paper describes the importance of using data mining in analysis and predicting performance of human resources in organizations, they help software companies to select the right people to perform better for the process.

In [8] gathered data of 300 students from different degree colleges and institutions in Awadh University, Faizabad. It uses Bayesian method to analyze the performance of the student. Our opinion of this work that he uses a good series of procedures to predict the student performance, that we will use in our work, but with differences because we will focus on lecturer performance side, and the amount of data that we will use. Here the researchers used a sample consist of 300 students only, it makes the results less reliable.

we try to found some related work that investigated the performance predicting in general, the term performance predicting is a wide term that use in most things: variant organizations, factories, human, machines, etc. we can benefited from the methodologies that used to predict performance, so we try to explore works that closed to our work field, machine learning.

B. *Faculty's performance prediction using Machine Learning techniques*

This section presents some works that studied the evaluation of performance of university faculty using machine learning techniques.

In [9], Introduced to analyze faculty performance, using machine learning techniques, based on different measures. Naive Bayes, ID3, CART and LAD are the classification algorithm used. Naive Bayes algorithm has 80.35% of highest accuracy and ID3 has the lowest 65.17%. The attributes that use by the research belong to students like student attendance and student result.

In [10], Collect data to identify issues related to the teaching quality of faculty in the Department of Management Information Systems (MIS) of the University of Bogazici in Turkey The study present prediction of lecturer performance based on course characteristics and lecturer characteristics from students view, it use only two attributes about personal characteristics: whether the faculty is part time or not and teaching first time the given course.

In [11], Data and findings were obtained from a survey of 104 teachers at the Sanandaj Daughter Vocational Faculty on teaching activities in the classroom, followed by machine learning algorithms such as Association Law and Decision Trees (j48), evaluating and predicting teacher acceptance for further teaching. It also suggested that the student rating of assessment is a very important variable for many institutions to obtain this data regarding faculty evaluations. By using machine learning and J48 tree as a decision tree, new rules result in educational administrators being able to use these rules for submitting new teachers in future decisions and continuing with selected old faculty.

The study limited on four variables related to faculty, without determining the accuracy of this model, and not exploited the historical data for mining.
In [12], using the current questionnaire to define the variables and predictors of teacher quality among undergraduates at a private university in Malaysia. They use multiple linear regressions and the step-by-step approach to assess the major determinants of overall lecturer output. The final model findings showed that the characteristics of professor and teacher, subject characteristics, and learning tools and facilities represented 61.9 percent of the difference in overall student lecturer quality. The perspective of our research that the features of the professor and teacher were not statistically comparable in terms of gender. This study depends on statistical methods not data mining methods, it deals with samples not with historical data, it can benefit for us that the study results show that overall performance are measured using most relevant feature which act as a predictor called the faculty characteristics.

We conclude by dividing these studies into two categories A and B. Category A contain studies that investigated in the field of performance prediction using machine learning techniques. In Category B, we present studies about prediction of faculty performance using machine learning techniques. We notice that most studies built its prediction on course characteristics, student characteristics, or faculty characteristics poorly. The studies that involved faculty characteristics are involved as secondary predictors. In our work, we will predict the faculty performance depending mainly on faculty characteristics. We get dataset contain 10 attributes about faculty characteristics from a university in the south tamilnadu of India.

III. REVIEW CRITERIA

Machine learning was the technique used in this research. Data set and Pre-processing, feature selection, Modelling using five distinct algorithms, Model Evaluation and performance measure were the four stages involved in this technique (see Fig. 1). Data collected in information exploration and pre-processing were based on all accessible faculty information that would probably be applicable to evaluate the efficiency of their performance. In the modelling phase the prediction model is developed using the machine learning techniques. These models use Cross-validation to learn, examine and evaluate the dataset. the performance measures were done to provide adequate data in the faculty prediction in the teaching course.

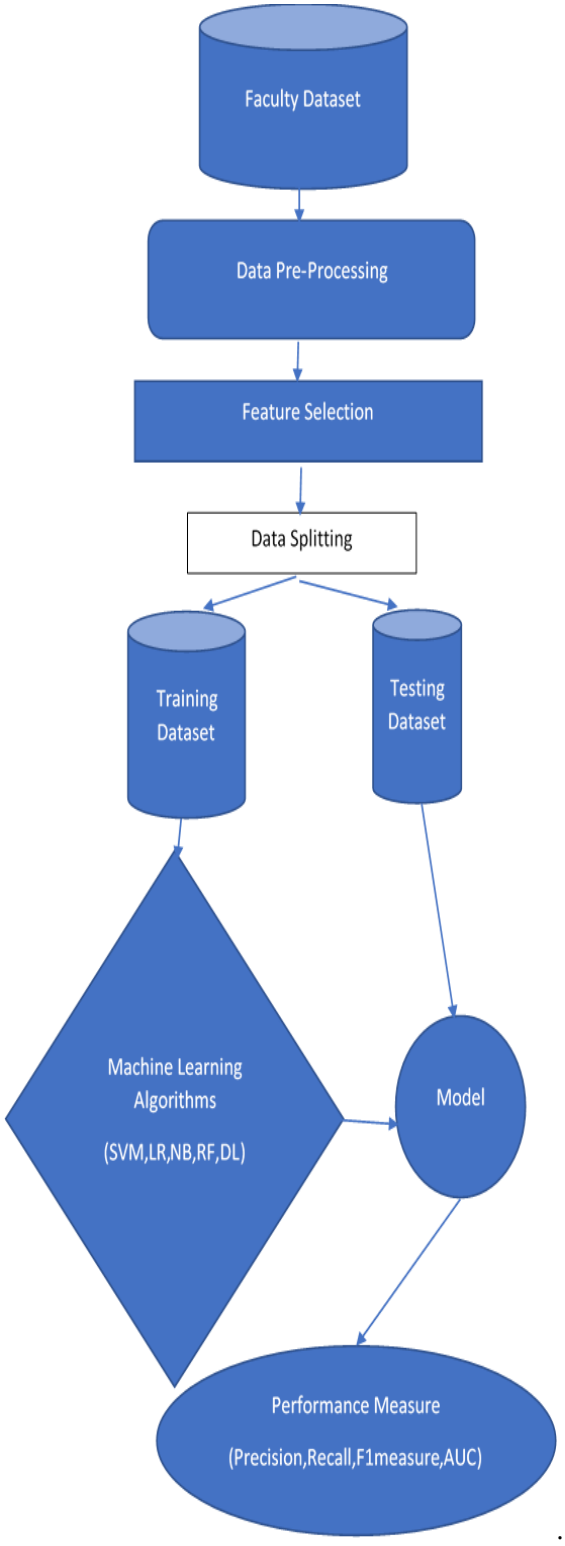

Fig. 1. Overall Framework of the proposed work

A. *Data Pre-processing and Feature Selection*

The dataset is based on the student feedback record of the various department in the university. From 2016 to 2017 batch, there were 892 students who gave their corresponding feedback about the course thought by the faculty have been viewed in this research. The data gathered is categorized according to the features.

Table- I shows features and label features along with their respective possible values. Using Optimize Selection Evolutionary, the genetic algorithm is used to know which attributes are most likely relevant and useful in classification.

Table- I: Feature and Label feature of the dataset

Feature	Description	Values
BCC	teaches basic concept of the course clearly	1.0 – 5.0
FPC	coming to the class with fully prepared	1.0 – 5.0
RWP	Covering the course topics with some real work problems	1.0 – 5.0
WSC	Writing and speaking about the course is clear and audible	1.0 – 5.0
ITC	Assignment/exams increases the ability to think creatively	1.0 – 5.0
ITM	Uses Innovative Teaching Methodology	1.0 – 5.0
CCT	Comes to class on time	1.0 – 5.0
SGB	Deals with Students General Behavior	1.0 – 5.0
AFB	Provide appropriate feedback on how to improve in all aspects	1.0 – 5.0
SQT	Rate your satisfaction about the quality of Teaching the Course	1.0 – 5.0
Result	Satisfactory or not	1/0

B. *Implementation*

We use RapidMiner studio 9.0 to predict the faculty performance. Fig. 2 shows the process of predicting and analyzing the algorithms. First, to load the data set, retrieve operator is used. Then to create the copies of dataset, the multiply operator is used. Model statistical efficiency is predicted by Cross Validation operators. This operator uses optional parameters like number of folds and sampling types as 10-fold validation and Stratified sampling as the automatic sampling. Training and testing methods are handled through cross validation.

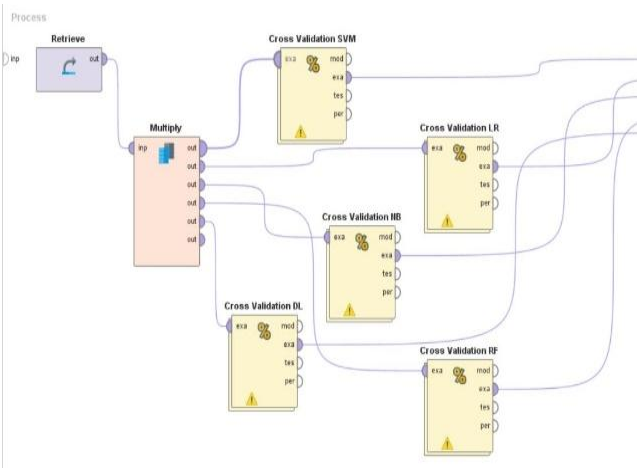

Fig. 2. RapidMiner used for prediction and evaluating various algorithms

Fig. 3 shows that to select the most relevant feature in the dataset, Optimize Selection operator (Evolutionary) is used. Using genetic algorithm, Features are selected. GA is a heuristic search algorithm which reflects the process of natural evolution. It provides the best solution for the searching and optimization problem. Select by Weight operator is used to choose the only features with a weight higher or equal to one in the classification dataset. AdaBoost operator and optimization parameters are used to enhance algorithm efficiency. For optimization and modelling, dataset is replicated.

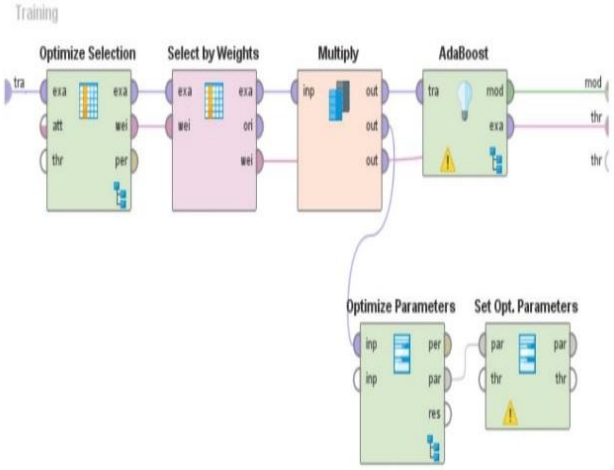

Fig. 3. RapidMiner used for Training Phase

Fig. 4 shows that the Select by Weight Operator uses features used in the training phase also uses same in testing phase. Machine Learning algorithm performance is calculated using the Performance Operator. The performance outcomes of each algorithm are analyzed to find the best machine learning algorithm for predicting the faculty performance in the student course using 10-fold Cross validation.

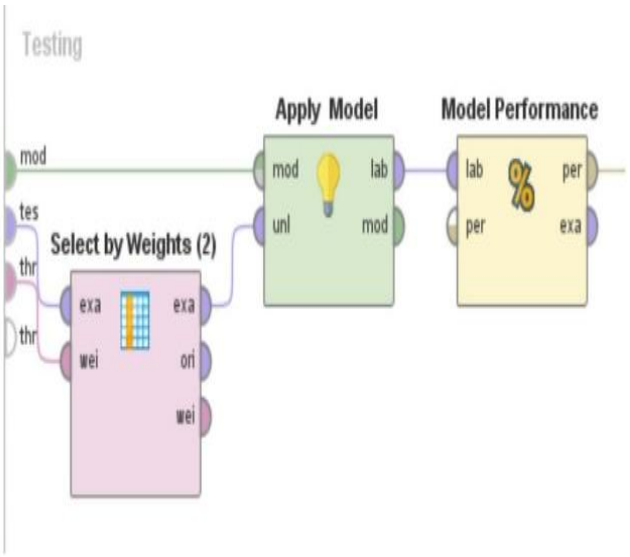

Fig. 4. RapidMiner used for Testing Phase

IV. PERFORMANCE EVALUATION AND RESULT

This section discusses the results of the evaluation. The main objective of the prediction model is to determine which features will act as a predictor. Table- II shows the most relevant features used in the dataset using optimization selection operator. It is interesting to note that the features BCC, FPC, ITC, AFB and SQT are most relevant features which act as a predictor with weight 1.

Retrieval Number: D9750118419/2019©BEIESP
DOI:10.35940/ijrte.D9750.118419
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Table- II: Most Relevant Features in the Dataset

Features	Weight
BCC	1
FPC	1
ITC	1
AFB	1
SQT	1
RWP	0
WSC	0
ITM	0
CCT	0
SGB	0

Performance evaluation is used to measure the prediction model in terms of correctly predicted decision of the models. Class variables can be assumed to be Positive (P) and Negative (N). Actual positive (P) which the classifier properly labels as positive are called true positive (TP). Actual positives wrongly labelled by the classifier as negative are known to be false negatives (FN). Actual negatives (N) which the classifier properly labels as negatives are called true negatives (TN) Actual negatives wrongly labelled by the classifier as positives are known to be false positives (FP). Confusion Matrix uses these terms which shows in Table- III.

Table III: Confusion matrix

	Predicted Positive	Predicted Negative	Total
Actual Positive	TP	FN	P
Actual Negative	FP	TN	N
Total	P'	N'	P + N

Accuracy, Precision, Recall, Specificity and F1-measure is used to Calculate the performance evaluation which are given below

$$Accuracy = \frac{(TP + TN)}{(P + N)}$$

$$Precision = \frac{TP}{(TP + FP)}$$

$$Recall = \frac{TP}{P}$$

$$Specificity = \frac{TN}{N}$$

$$F1 - Measure = 2 * \frac{(Precision * Recall)}{(Precision + Recall)}$$

Table- IV Shows the confusion matrix of the SVM model, which have 448 True positive and 114 True Negative instances Whereas we have 218 False Positive and 112 False Negative instances as a classifier model result.

Table- IV: SVM Confusion Matrix

	Predicted Satisfactory	Predicted Not Satisfactory	Total
Actual Satisfactory	448	112	560
Actual Not Satisfactory	218	114	332
Total	666	226	892

Table- V Shows the confusion matrix of the LR model, which have 472 True positive and 68 True Negative instances Whereas we have 264 False Positive and 88 False Negative instances as a classifier model result.

Table- V: LR Confusion Matrix

	Predicted Satisfactory	Predicted Not Satisfactory	Total
Actual Satisfactory	472	88	560
Actual Not Satisfactory	264	68	332
Total	736	156	892

Table- VI shows the confusion matrix of the NB model, which have 490 True positive and 62 True Negative instances Whereas we have 270 False Positive and 70 False Negative instances as a classifier model result.

Table- VI: NB Confusion Matrix

	Predicted Satisfactory	Predicted Not Satisfactory	Total
Actual Satisfactory	490	70	560
Actual Not Satisfactory	270	62	332
Total	760	132	892

Table- VII shows the confusion matrix of the RF model, which have 452 True positive and 200 True Negative instances Whereas we have 132 False Positive and 108 False Negative instances as a classifier model result.

Table- VII: RF Confusion Matrix

	Predicted Satisfactory	Predicted Not Satisfactory	Total
Actual Satisfactory	452	108	560
Actual Not Satisfactory	132	200	332
Total	584	308	892

Table- VIII shows the confusion matrix of the DL model, which have 428 True positive and 158True Negative instances Whereas we have 174 False Positive and 132 False Negative instances as a classifier model result.

Table- VIII: DL Confusion Matrix

	Predicted Satisfactory	Predicted Not Satisfactory	Total
Actual Satisfactory	428	132	560
Actual Not Satisfactory	174	158	332
Total	602	290	892

Table- IX shows the precision, Recall, and Specificity of the SVM model are 67.27%, 80%, and 34.34% respectively. LR model are 64.13%, 84.29%, and 20.48% respectively. NB model are 64.47%, 87.50%, and 18.67% respectively. RF model are 77.40%, 80.71%, and 60.24% respectively. DL model are 71.10%, 76.43%, and 47.59% respectively.

Table- IX: Precision, Recall, and Specificity of Various model

	Precision	Recall	Specificity
SVM	67.27	80	34.34
LR	64.13	84.29	20.48
NB	64.47	87.50	18.67
RF	77.40	80.71	60.24
DL	71.10	76.43	47.59

Fig. 5 shows the Accuracy, Area Under Curve, and F1-measuire of various model. In these models, Random Forest with 73.10% as highest accuracy and next Deep Learning with 65.67% accuracy then AUC of RF and DL values are 0.730 and 0.701 respectively and F1-measure with 78.73 for RF and 74.24% for NB.

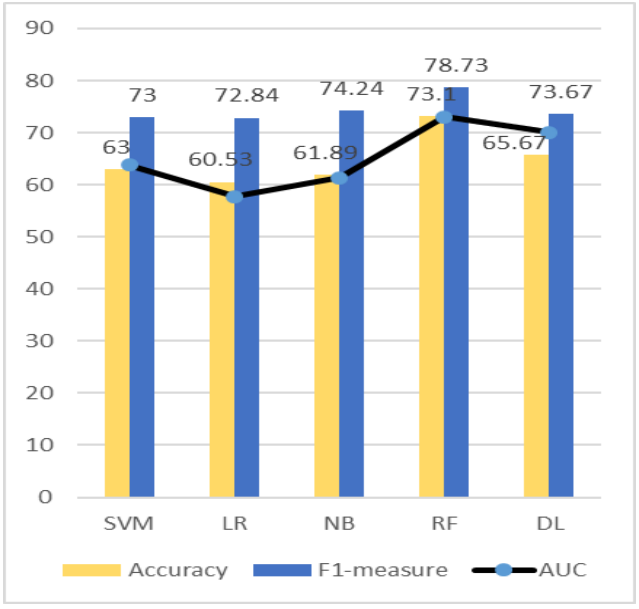

Fig. 5. Overall Performance of the Model

Only Random Forest algorithm is achieved highly predicted classifier among the five-model used, which will be helpful to predict the faculty performance in the Course evaluation for teaching and Learning process.

V. CONCLUSION AND FUTURE WORK

To evaluate the faculty performance in the educational institution/university EDM methods are used with the emergence of machine learning. In this research, by using models, machine learning method is implemented to predict faculty performance in course evaluation through student feedback questionnaires. Only BCC, FPC, ITC, AFB and SQT are identified as the most important predictors in the dataset among the group of features. Evaluation results indicate that faculty should think well about the teaching course which improves the student knowledge as this provides a huge effect in their Teaching-Learning Process.

We have used various ML algorithm in the faculty dataset. Among these algorithm, Random forest has the highest precision and AUC value of 77.4% and 0.73 respectively. So, using this model, the institution/University can be able to identify the faculty performance in teaching the course for students. These faculty will be given huge priority during the selection of course and review meeting which will be able to satisfy the student in teaching the courses and it also be helpful for the institution/university to get the higher ranking, accreditation and improves the students' knowledge, placement as well.

Future work can also be carried out to enhance the model's efficiency, can also work with huge dataset and additional features can be used for further research. Also, the development of Recommendation System to improve the faculty performance should be take place.

ACKNOWLEDGMENT

The first author would like to thank the management of Kalasalingam Academy of Research and Education (KARE) for providing fellowship to carry out the research.

REFERENCES

1. CARLY STEYN, CLINT DAVIES & ADEEL SAMBO, Eliciting student feedback for course development: the application of a qualitative course evaluation tool among business research students, *Assessment & evaluation in higher education*, 2018.
2. NAIK, R. L., SARMA, S., MANJULA, B., & RAMESH, D., Study of Trends in Higher Education, *International Journal of Computer Trends and Technology*, 6-10, 2011.
3. AGASISTI, T., BOWERS, A.J., Data Analytics and Decision-making in Education: Towards the Educational Data Scientist as a Key Actor in Schools and Higher Education Institutions, *Handbook of Contemporary Education Economics*, p.184-210, 2017.
4. ASHISH DUTT, MAIZATUL AKMAR ISMAIL, & TUTUT HERAWAN, A Systematic Review on Educational Data Mining, *IEEE Access*, VOLUME 5, 2017.
5. B.M. MONJURUL ALOM., MATTHEW COURTNEY, Educational Data Mining: A Case Study Perspectives from Primary to University Education in Australia, *International Journal of* Information Technology and Computer Science, 1-9, 2018.
6. HAJIZADEH, N., & AHMADZADEH, M., Analysis of factors that affect students 'academic performance-Data Mining approach, *International Journal of advanced studies in Computer Science and Engineering, VOLUME 3,2014.*
7. GUPTA, S., & SUMA, V., Prediction of Human Performance Capability during Software Development Using Classification. Paper presented at the ICT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India, 2014.
8. BHARDWAJ, B. K., & PAL, S., Data Mining: A prediction for performance improvement using classification, *International Journal ofComputer Science and Information Security*, VOLUME 9, 2011.
9. PAL, A. K., & PAL, S., Evaluation of Teacher's Performance: A Data Mining Approach, *International Journal of Computer Science and Mobile Computing*, VOLUME 2, 2013.
10. MARDIKYAN, S., & BADURE, B., Analysing Teaching Performance of Instructors Using Data Mining Techniques, Paper presented at the Informatics in Education, Vilnius University, Lithuania,2011.
11. AHMADI, F., & ABADI, S., Data Mining in Teacher Evaluation System using WEKA, *International Journal of Computer Applications*, 63(10), 12-18,2013.
12. SOK-FOON, Y., SZE-YIN, J. H., & YIN-FAH, B. C., Student evaluation of lecturer performance among private university students. *Canadian Social Science*, *8*(4), 238-243,2012.
13. MUSTAFA AGAOGLU, Predicting Instructor Performance Using Data Mining Techniques in Higher Education, *IEEE Access*, VOLUME 4, 2016.

AUTHORS PROFILE

T. Manjunath Kumar currently pursuing Ph.D. in the department of Computer Science and Engineering at Kalasalingam Academy of Research and Education, Krishnankoil. He Completed M.Tech Degree in Kalaslingam university and Received B.Tech degree in Park college of engineering and technology, Coimbatore..

Dr. R. Murugeswari completed her Ph.D. (CSE) from Kalasalingam University, Krishnankoil and M.E (CSE) from Arulmigu Kalasalingam College of Engineering, Krishnan oil under Anna University, Chennai. She is currently working as a Associate Professor in the Department of Computer Science and Engineering at Kalasalingam University, Krishnankoil.She has more than 18 years of Teaching Experience. Her research interests are wireless network, Evolutionary Algorithm, Big Data Analytics and Social Network Analysis. She has published more than 15 papers in various International journals and International Conferences.