Abstract—Education is one of the sectors that established for the reason of fruitful production of the student and society in the world. While doing their education student faces different factor that affects the academic performance. The performance of the student is one of the crucial aspects of every educational institute. Knowing performance of the student has significant benefit for the student as well as the Educational institute. Educational data mining is the use of data mining tools related to education. Data. Different data mining tools used to predict the performance of the student using the available data set based on the dominant attribute. With the help of this educational data mining techniques, an academic performance of the student and different factor affect that lead the student to the failure and success identified. The result of the performance of algorithm is vary based on the different condition. Data set size, missing value and attribute selection determines which data mining tools perform best from all the other.

I. INTRODUCTION

In the present information world, education plays a major role in the progress of a nation’s economy and literate society development. Education also increases decision-making process and create the competitive generation. Different data mining tools ar used to predict the student performance. The performance of the student varies throughout a year and the number of a student failure increase due to the decrement in the student performance with many factors. The most known cause of the failure of the student is lack of deep knowledge about the course, difficult to adapt the new environment, social media usage, romantic relationship and teacher approach to the student and course. The prediction of the student performance has the advantage of indicating earlier the factor and provide prevention solution. Data mining is used to generate the meaningful information from massive data set using some patterns. Data mining used in many Applications such as educational data mining, web mining and Text mining [1]. Data mining is one of the paces in KDD Process. Knowledge discovery (KDD) focus at the uncover of foremost information from immense collections of data [2].

The utilization of Data Mining in the Educational perceptive is mentioned as Educational Data Mining (EDM) and explained By the International Educational Data Mining Society in as “new coming discipline, concerned with improving Methods for investigating the unique types of data that come from Educational settings. Data mining techniques applied to foretell the educational achievement of the learners based on their socio-economic condition and previous Academic potential [3]. Using data mining techniques assistant for anticipate the student accomplishment, indicate resource and aids to make a decision. Forecasting the students’ achievement is a vital Part of a higher education, as the whole growth of the Education system is directly proportional to improve the success rate of the learner, increase students’ outcome and decrease the dropout rate. Therefore there are many Situations where the performance of the students’ needs to be predicted, for example, to differentiate allowed students for join in placement activities and to find the weak students so that different necessary action took to change. A decision tree is a flow –charts like tree structure which is made of nodes and arcs. Each internal nodes represented by rectangles and the leaf node represented By Oval. Decision tree mostly used for the use of decision making. It always begins from the node to take action, and from this point, the user spilled each node continuously [4]. Classification techniques is one of the most used applications of data mining. The major function of classification is allocating class label to a set of possible class values to an unseen instance constitute a set of variables.
Classification is done by using classifier. In the recommended paper classification techniques applied to foresee the student achievement. Using data mining techniques aids for estimating the student accomplishment, indicate resource and helps to make a decision. Foretelling the educational accomplishment is indispensable Part of a higher education, as the whole growth of the Education system is directly proportional to improve the success rate of the students increase students’ outcome and decrease the dropout rate. Therefore there are many circumstances where the achievement of the students’ needs to be predicted, for example, to differentiate allowed students for join in placement activities and to find the weak students so that different necessary action can be taken for their improvement [5].

II. LITERATURE SURVEY

Tripti Mishra et al. [6] An early prediction of students’ failure may be the management provide timely manner counseling as well coaching to increase the success rate and student retention, which uses the j48 and random forest to predict the third-semester student performance.

V.shanmugarajeshwar et al. [7] they proposed C5.0 classification algorithm and it has high accuracy. The attribute used are students’ name, roll number, previous semester marks, attendance, lab works which nourish to predict academic performance and feature selection used.

Camilo e Jopiz et al. [3].Educational data mining techniques endorse to model academic loss using two classification algorithm such as Bayes and decision tree classifier. Academic and nonacademic data set are utilized. When using academic data, the accuracy is improved .Bayes classifier increased the performance while using academic data in first enrollments and decrease in the second enrollment.

Kamal bunkar et al. [4]. Proposed the use of classification algorithm which identifies the chief attribute that affects the higher education student. Decision tree rule formed which predict the grade of the student. The decision tree algorithm is ID3, C4.5 and CART employed. The performance of the decision tree classifier evaluated by using a confusion tree. The classifier differentiate student who is at risk, this assist to strengthen the institution education.

Mustafa Agaoglu. [5] Various classification algorithm applied to predict instructor performance. The classification algorithm employed are decision tree, support vector machine, artificial neuron network and discriminant analysis. Among the given Classification the C5.0 has higher and best accuracy. The questionaries’ has given to the student then the performance of the instructor measured. The perception of the student the performance result might be satisfactory and not satisfactory.

Evandro B.costa,baldino Fonseca.[8]suggested that they exercise data mining technique such as support vector machine,j48, Neural network and naive Bayes to early predict student failure in the introductory programming courses .support vector machine perform higher accuracy than the other. Concepción Burgos a.et al [9] This research employ classification techniques which supports to predict the dropout out of the student. Logistic regression is one of the classification techniques that assist to predict discrete variable.The paper decrease the dropout rate in the course and increase the accuracy when using student grade. Anjana Pradeep et al. [10]. The research suggested that the use of educational data mining to predicts the factor that affects the academic performance. The paper identifies the weak performance of the student. In this paper, the induction rule and decision tree used to classify the failure and success of the student. It identifies the students who are at high risk of failure.

Ali Daud et al. [11] The paper describe the use of classification algorithm .The paper predicts for the scholarship students from different university whether they complete their degree or not .Learning analytics used.Different classification algorithm used support vector machine, CART, Bayes network and naive Bayes, Prediction based on generative and descriptive classification model. Support vector machine found that it most effective classifier.

Ishwank Singh et al. [12] the proposed describe the use of clustering algorithm to predict the student performance using the k-means. In this paper, the parameter is the marks of the student, Projects internship and skills set. The performance of the student calculated and different results observed.

Mirmal Pandey, S. Taruna et al [14]. Accurate predictions of students’ academic performance at early stages of the degree programme helps in identification of the weak students .Enable management to take the corrective actions to prevent them from failure .The integrated classifier consists of three complementary algorithms, namely Decision Tree, K-Nearest Neighbour, and Aggregating One-Dependence Estimators (AODE). The proposed method is also compared with KSTAR, OneR, ZeroR, Naive Bayes, and NB tree classifiers as well as with the individual classifiers and voting methodology are used. Integrated model is constructed using AODE, IBK, and j48 with the help of voting.

Rashmi Bansal et al. [15] This paper uses educational data mining to detect and make a prediction of the student behavior and conduct. The paper mostly focuses on different problem of the student which group them according to the level of the difficulty of the situation. Finally, it becomes easy for the teacher to identify the weak areas of the student.

Xiaojian long et al [16] In this paper decision tree algorithm used to build a model for the student achievement .C4.5 is decision tree algorithm uses to build the model .C4.5 provide advantage for the improvement of teaching and learning quality of student and teacher. At the same pruning algorithms are used to reduce the complexity of the tree finally. C4.5 builds tree which gives higher accuracy when analyzing the data.

Munaisyab Abdullah et al [17] The study describe the application of data mining to foretell the student enrollment in the higher education .The researchFind the best method for making a decision for the management.

Brijesh Kumar Baradwaj et al [18] classification and decision tree algorithm are used to predict the performance of the student. The classification performed by using the previous
database. Various information like attendance, class test, assignment results are used to foretell the performance of student achievement.

Lotfi Najidi et al [19] Predictive modeling system random forest and CART are operated to predict the student retention and graduation. In this system the student at risk of dropout identified. Ensemble method bagging utilized in the system to increase the performance of the algorithm. The vital attribute in the paper are GPA and secondary school. The accuracy of the algorithm is 88%. The machine learning tools this model uses is R.

Anne-Sophie Hoffait et al [20] In the decision support system data mining techniques random forest, artificial neuron network and logistic regression applied to detect the student performance early during registration time. The ensembling method bagging combine two the decision tree to the increment of the efficacy of the algorithm.

Sattar Ameri et al [21] This research done in the higher education student. Data mining and statistical method applied to predict the dropout and distinguish the student who is at endanger. The two framework Cox proportional hazards (cox) and time -dependent cox indicate the dropout of the student. The attribute in this paper are family background, high school information, financial, enrollment data and GPA.

Asep Nurhuda et al [22] Artificial neuron network applied to indicate number of student graduate in each year of the school. A pattern is generated using artificial neuron network to forecast number of student graduate. Input layer, hidden layer, and output layer and activation function applied to forecast the precision.

Natthekankan lam –On and Tossaapon Boongoen [23] Descriptive model generated using clustering algorithm. A number of student information who perform well and who perform less clustered. This clustering aids to detect dropout ration of the student. R-KM clustering approach groups student.

Shaymaa E.Sorour et al [24] The study suggests evaluating the performance using the comment provided from student in the class in each lesson of the subject. The design method use Latent semantic analysis and data mining techniques. The techniques are support vector machine and artificial neuron network. The machine learning tools aids to forecast the final grade. The accuracy of support vector machine is 50.7 % and also ANN is 48.7%. SVM is better than ANN in this paper.

Dursun Delen [25] In this research student retention is key issues addressed for the benefit of the institute and student. Machine learning tools applied to build analytical model which forecast the cause and effect of the attrition of the student. Various data mining techniques used such as decision tree, neural network, support vector and logistic regression. Among the given techniques SVM has higher accuracy when compared with other techniques. The comparison of data mining algorithm takes place in these research. The result found as SVM, ANN, decision tree and logistic regression, respectively. Ensemble method used to increase accuracy of algorithm by combing two or more classification techniques together. Another important techniques is sensitive analysis which distinguish the input and output character by observing the cause and effect relationship.

III. METHODOLOGY USED IN DATA MINING TECHNIQUES

The prediction of the student performance is mostly implemented using different data mining techniques namely ID3, random forest, artificial neural network and logistic regression.

A. ID (Iterative Dichotomiser 3)

It is the algorithm created by Ross Quinlan. ID3 is the precursor to the C4.5. J48 is the implementation of the C4.5. A decision tree can be formed using ID3 using the dataset. ID3 most of the time helps to develop the least decision tree. C5.0 classifier is the highest in performance and accuracy followed by SVM [5].

B. C4.5

It comes after ID3 made by Quinlan Ross. C4.5 forms decision tree and aids to handles categorical and continuous attribute. C4.5 use for gain ratio to make decision tree. Pruning helps to eradicate unwanted branches in the tree. This is the most important to increase the accuracy and the performance of the algorithm [15].

C. Artificial neural network (Multilevel perception)

ANN, another technique used in the different paper which improves the student and instructors performance. Similar to a human brain, Artificial neuron network are predominantly shown as a system of interconnected neurons that interchange information between each other. Multilevel perception performs best among all classifier and more efficient when there are large data set. It is found in the weka tools to make analysis by using the name sequential minimal optimization. Accuracy values, which evaluating the effectiveness of the models, are all at least approximately 90% [5].

D. Random forest

Random forest constructs a tree that examine k randomly chose attributes at each node without pruning. Random tree more power than j48 for predicting the student achievement [6]. Random forest is often over fit. Its accuracy is not better than other methods for dataset with small and not as easy to visually interpret.

E. Logistic regression

It is a method best for a type of problem (the variable to predict is discrete and the predicting variables are discrete) Where it can take only two values, “0” and “1”. Represent outcomes such as pass/fail and win/lose. Its values bound between 0 and 1. The function which represent logistic regression represented is 0<=f (n) <=1. Numeric vales are used in the logistic regression due to this it increases the accuracy of the student grade [9].
IV. CONCLUSION

Most of the paper discussed above related to student performance and academic dropout. The academic failure and success based on historical data, educational data and nonacademic data. Those papers used to predict the student performance. The need for analyzing student performance is to help the student give attention if they are at risk point or not. Another advantage is benefit of the institution to take different action to right track the student. Data mining techniques utilized especially, classification algorithm to analyze the student data and form model. Classification model contribute for prediction. The accuracy of the decision algorithm compared and the one which has high selected to predict the performance. Prediction of student performance in different paper differ. The difference is duration takes place for some forecast the last semester, some in the middle semester and other early prediction of the student performance. The academic achievement in all paper based on the result of the student. The student GPA is the main attribute that determines the performance of the student. The limitation found in various student performance papers are the difficult to detect early the factors that lead to failure.

References


