INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688

STUDENT PERFORMANCE PREDICTION USING MACHINE LEARNING ALGORITHMS

[1]GUJJULA VEERANJANEYA REDDY, [2]Dr. YND ARAVIND

[1]PG Scholar, Dept. of MCA, Newton's Institute of Engineering, Guntur, (A.P)

[2]Professor, Dept. of CSE, Newton's Institute of Engineering, Guntur, (A.P)

Abstract: *An enormous measure of computerized information is being produced over a wide assortment in the field of data mining strategies. The creation of student achievement prediction models to predict student performance in academic institutions is a key area of the development of Education Data Mining. A prediction system has been proposed by using their 10th, 12th and previous semester marks. We used Radial Basis Function network, Multilayer Perceptron, C4.5 and Random Forest Algorithms for classification. Initially classification accuracy was computed individually by the classification algorithms. The Radial Basis Function network, Multilayer Perceptron, C4.5 and Random Forest Algorithm's individual classification gave the accuracy of 72.9167%, 75.4167%, 75% and 73.125% vice versa. To increase more accuracy of classification algorithm the Radial Basis Function network is combined with multilayer perceptron. This hybrid algorithm provides 75.625% of classification accuracy. Then we combined C4.5 algorithm with random forest algorithm which gives 76.4583% classification accuracy.*

Keywords: *Hybrid classification algorithm, Student performance prediction, Random Forest.*

I. INTRODUCTION

Due to the huge amount of data in educational databases, predicting the performance of students has become more difficult. The shortage of an established framework for evaluating and tracking the success of students also isn't currently being considered. There are two primary reasons why such kind of occurring. First, the research on existing methods of prediction is still insufficient to determine the most appropriate methods for predicting student performance in institutions. Second, is

ISSN: 2057-5688

the absence of inquiry of the specific courses. The real goal is to have an overview of the systems of artificial intelligence that were used to predict academic learning. This research also focuses on how to classify the most relevant attributes in student data by using prediction algorithm. Using educational machine learning methods, we could potentially improve the performance and progress of students more efficiently in an efficient manner. Students, educator and academic institutions could benefit and also have an impact [1].

The rapid digitization of educational institutions is very significant in helping educators and education personnel to collect data. The abundance of data on educational institutions is also moving very dynamically accompanied by changes in learning patterns and governance of educational institutions from previously offline-based to online-based. Similar to other industries affected by the pattern of 4.0 Industry, explicit programming which is currently commonly used as a tool to provide solutions to problems is no longer sufficient in some cases. In this case, machine learning fills the gap and provides a solution for processing these very large and dynamic data. Prediction of student's performance can be used as a basis for early intervention on the potential failure of students to achieve learning objectives; and at the same time able to make changes to learning strategies in order to facilitate student diversity. This is also supported by the availability of student data that can be processed to make predictions such as behavioural data, type and frequency of activities carried out (both in online and offline learning settings), age, height and weight, previous academic achievement; as well as latent data such as personality and motivation as well as external data such as parenting patterns, parental support makes the data contained in educational institutions abundant, multi-data and qualified to be analysed[2].

ISSN: 2057-5688

Student's performance in the educational process can literally be defined as something that is obtained from changes in the behaviour of students based on their experiences, besides that learning outcomes are also a realization of the potential or capacity possessed by students. These learning outcomes from students can be seen from their behaviour, both behaviour in the form of understanding knowledge, thinking skills, or motor skills; an outcome of the process of changing student behaviour after attending lessons. The concrete form of student's performance can be seen from their understanding of the knowledge being studied, their expertise in processing information and making decisions based on certain thoughts or motor skills [3]. Based on those understandings, student's performance can be observed and measured in the realm of students' knowledge, attitudes and skills after following a series of lessons. Student's performance depends on the teaching and learning process they go through, so that learning outcomes can be used as considerations in improving the quality of the learning process.

II. LITERATURE SURVEY

Supervised learning is divided into two types, namely classification and regression. The purpose of classification is to predict the class in the labelled data, which has provided a choice from a list of possibilities. Meanwhile, the main purpose of regression is to predict the description of the regression relationship from the data, namely a floating-point number. The second type is unsupervised learning. In unsupervised learning, labelling is not done on all input data. Algorithms in supervised learning will read the data as it is without any relation to the label given so that the program will draw information directly from the existing data. In general, there are two things that are generally done in unsupervised learning, namely by transforming datasets or by clustering. Dataset transformation is carried out with an algorithm that allows the program to

ISSN: 2057-5688

convert existing datasets into datasets that are easier to understand. While in clustering, the algorithm will separate the data into several clusters based on the similarity of data types. One example of the application of the clustering algorithm is the compilation of photos uploaded to social media into albums based on the faces of those who appear in the photos. Solving problems using machine learning effectively requires an understanding of the problem to be solved in its entirety. Understanding the problem will affect the decision on what data is needed and what algorithm should be used to solve the problem. Algorithms that will be used in machine learning (both categorized in supervised learning and unsupervised learning) need to be adapted to the existing data and the analysis objectives to be achieved.

Vladislav Miskovic [4] evaluated predictive accuracy of pharmaceutics symptomatic, e-commerce, retailing, and economic analysing problems using direct, indirect and hybrid machine learning models. He has used C4.5, C45Rules, KNN, Random Forest algorithms for hybrid classification.

Thaddeus Matundura Ogwoka et al [5] construct a model to predict the students' scholastic presentation using a hybrid k-means and decision tree algorithms. This algorithm improves the accuracy. That helps to the institution simply implemented of further improvement to do forecasting of students' execution.

Akanksha Ahlawat, Bharti Suri [6] has been applied hybrid algorithm, which uses the concept of clustering and pattern evaluation and representation of result which is in form of decision trees. Their result shows enhanced accuracy which is tested on real life datasets. Hamza Turabieh [7] used hybrid feature selection algorithm for predicting the student performance. He used KNN, CNN, NB and C4.5 algorithms for better prediction. Their research data set consist of student marks, school information and demographics.

ISSN: 2057-5688

Bindhia et.al [8] found novel foretelling technique for estimating the student's execution in academic who have been evolved based on both classification and clustering techniques. Their result shows that the hybrid algorithm of integrating clustering and classification gives better results for prediction.

Abhishek Lal and Kumar [9] made hybrid classification algorithm by amalgamation of Decision Tree and Naïve Bayes algorithms classifies the health data set. While they used individual classifier, the identity of the data set is could not achieve the preferred level. After making hybrid classifier the classification accuracy was increased.

Ankita Dewan and Meghna Sharma [10] applied hybrid classification algorithm to predict heart disease of the patients. Prediction of the patient disease they used best data mining algorithms such as Back propagation, naïve bayes, Multilayer perceptron, Support vector machine and C4.5. Their result shows that back propagation algorithm works better than other algorithms

An automated evaluation system has been proposed to evaluate student performance and to analyse the student achievement. Here the author uses tree algorithm for predicting student performance accurately. In the proposed system Education Data Mining (EDM) is used for the classification. Clustering data mining technique is used for analysing the large set of student database.

This technique will speed up the searching process and the also yield the classification result more accurately. A new Learning model has been proposed by using the student information from the college registration. The final dataset is provided as input so ML algorithms which can apply and predict student's academic performance. They selected 13 algorithms from 5 categories of ML they are Naïve Bayes, SVM, MLP, IBK, Rules and tree. A comparative study on supervised learning for student prediction has been proposed. The

ISSN: 2057-5688

author handles with 14 feature set for classification. The tools used for classification are: KNN, Decision tree, Navie Bayes Psychometric analysis of the student behaviour has been proposed by using intellectual parameters of the student which affect their study. Various mining techniques are used to determine the educational data covering psychological factors. The accuracy rate of the previous study is 89% but by using the proposed system the accuracy rate has increased to 90%. Here the author used Radical Basis Function Kernel to produce higher accuracy.

III. IMPLEMENTATION

In this research work we have collected data from UCI repository. Totally 480 data samples were applied to predict the student performance. Weka Machine Learning tool was used to predict the student performance.

PREPROCESSING

Pre-processing is the process of getting suitable attribute for classification. In this step the edudata.arff file has been given as an input to the weka tool. Originally the data set holds 17 features. These features are submitted to data pre-processing step to make enhanced classification. CfsSubsetEval attribute selection method was applied to yield the best attribute for Classification step. Figure 1 shows all the attributes that are used for pre-processing.

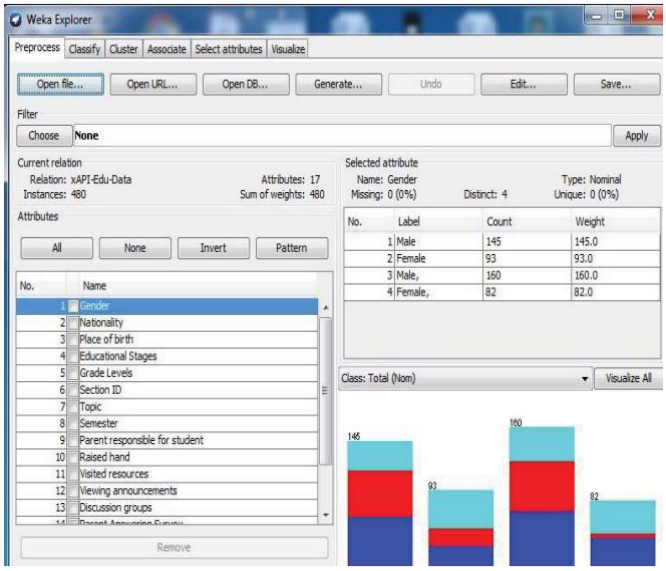


Fig.1 Data Preprocessing

CfsSubsetEval

It calculates the implication of a subtype of attributes. These features are taking into separate prognostic capability of every feature along with the degree of repetition between them.

INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE
ISSN: 2057-5688

Using CfsSubsetEval feature selection method 6 attributes were selected for next step. The figure 2 shows the top 6 attributes selected by using CfsSubsetEval attribute selection method.

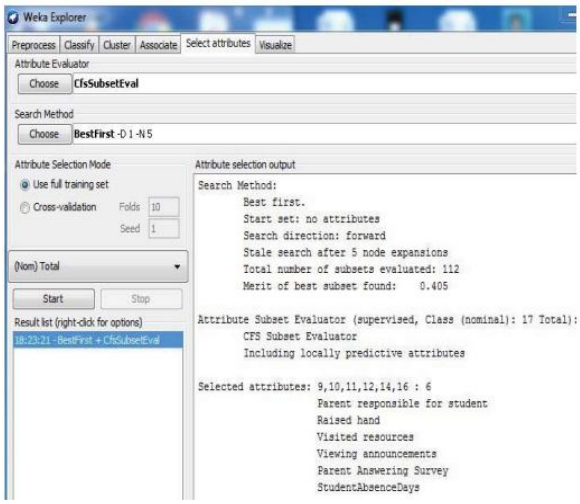


Fig.2 Attribute Selection

Attributes Description

Parents Responsible for Students- This attribute is considered for whether the parents are supported for their children's Education. This factor is an important factor to improve the student performance.

Raised Hand- While teachers taking class the students can ask doubt. When they raised the hand the teachers can understand they are active in the class.

Visited Resources- This attribute shows the interest of the students in learning. It calculates how many times the students are visited the resources.

CLASSIFICATION

In this step two classification algorithms are individually classified. To classify the data 10-fold cross-validation technique was applied. First, we used Radial Basis Function Network and then used Multilayer perceptron, Decision tree, Random forest classifiers.

Radial Basis Function Networks:

A Radial Basis Function (RBF) is a class of functions whose values increases or decreases with the distance from a central point. RBF network uses the Gaussian activation Function. It contains three layers.

INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE
ISSN: 2057-5688

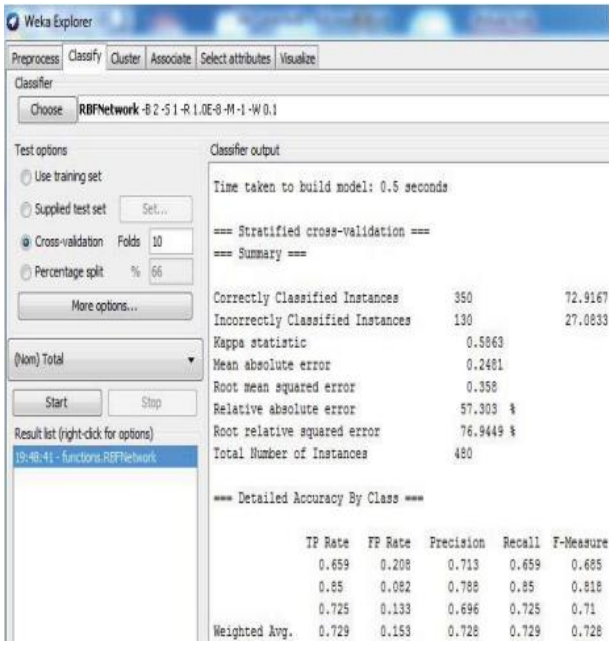


Fig.3 Classification of RBF Network

Multilayer Perceptron: Multilayer Perceptron contains multiple layers. MLP is a network of perceptrons. The neurons are placed in layers with outputs always following toward the output layer [1]. The multilayer perceptron be trained through the practice of back propagation Network. It is trained in a test condition under person expound better output cost resultant function which is enhanced in the training. A multilayer perceptron can map out and recognize adequate difficulty to discover an estimated mathematical representation for the given circumstances [2]. The figure 4 shows the result of multilayer perceptron which classifies 362 examples accurately and 118 examples are falsely classified. The classification precision of multilayer perceptron is 75.4167%

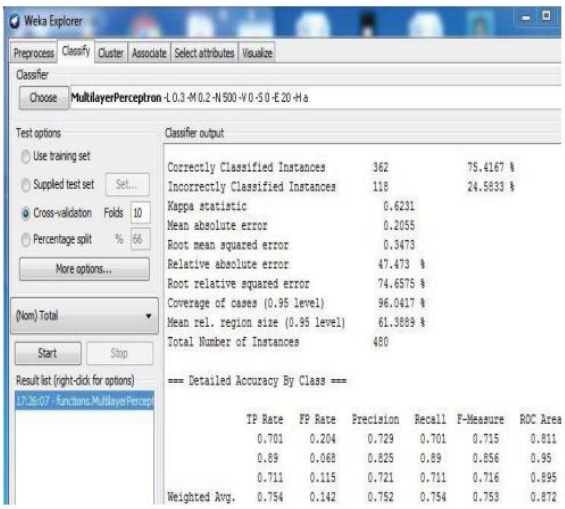


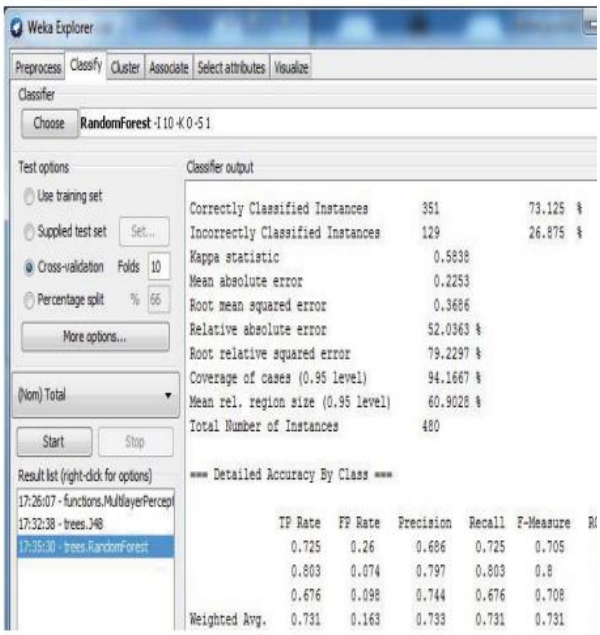
Figure 4. Classification of Multilayer Perceptron

Random Forest:

Random forest is a class of ensemble learning algorithm which is exclusively intended for tree classifier. It is one of the supervised learning algorithms. The benefit of Random Forest tree is that can be applied reciprocally classifications in addition regression approach. It is established

INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE
ISSN: 2057-5688

to get the better of the problem of overfitting of data in decision trees [2].



Decision Tree

Decision Tree classifier is the regression model which is represented in the form of tree structure. The purpose of Decision Tree classifier is to breakdown the dataset into smaller subset. The tree consists of decision nodes and leaf nodes.[19] In our proposed architecture the attribute which delivers maximum information will act as a decision node. The node which is present as the top most of the decision node acts as a predictor which is called as root node.

A. K-Nearest Neighbour K-Nearest Neighbour is one of the basic and essential classification algorithms in machine learning. It is non-parametric and makes any underlying assumptions about the distribution of data. The steps involved the KNN is listed below:

- File the training data in a sample points array.
- The Euclidean distance measures.
- Make the least distance range available

The proposed system handles with the student individual mark that include 10th, 12th mark and there semester mark. The prediction of the system has the following task: Case 1: The student who have secured below 50 percentage in their 10th and 12th.

Case 2: The student who have failed in internal. Case 3: The student having less attendance percentage / irregular. On the above three cases if any one achieved the student may not complete his/her degree successfully.

ISSN: 2057-5688

By this scenario the system has been implemented.

IV. RESULTS AND DISCUSSION

In this proposed work we have choose 4 classification algorithms for prediction. First, we classified the algorithm individually. Then we combined two algorithms for better prediction. In this section we compare the accuracy and correctly classified instances of 4 algorithms.

Table.1 Accuracy between various algorithms

Algorithm	**Correctly classified instances**	**Accuracy**
RBF	350	72.9167%
MLP	362	75.4167%
DT	360	75%
RF	351	73.125

Table 1 shows the result of the performance of 4 algorithms while making classification individually. Multilayer perceptron algorithm gives more accuracy than other 3 algorithms. It gives 75.4167 % of accuracy and 362 correctly classified instances. An RBF, J48 and Random Forest algorithms correctly classified instance are 350, 360, and 351 and gives the accuracy of 72.9167%, 75% and 73.125%

V. CONCLUSION

In this study we have used Radial Basis Function network, Multilayer perceptron, and Decision tree and Random Forest classification algorithms to forecast the students' academic performance. These four algorithms are individually classified and classification accuracy was computed. Then RBF and MLP algorithms are combined together and accuracy was computed. Then we made another hybrid classification algorithm DT and Random Forest. This algorithm gives better accuracy of 76.4583 compared to RBF and MLP hybrid classification. So, we concluded that DT and Random Forest hybrid classification algorithm works better than RBF and MLP hybrid classification algorithm.

REFERENCES

ISSN: 2057-5688

[1] N.V.Krishna Rao, Dr.N.Mangathayaru, Dr.M.Sreenivasa Rao," Evolution and Prediction of Radical MultiDimensional E-Learning System with Cluster based Data Mining Techniques", International Conference on Trends in Electronics and Informatics,2017, PP.701-707.

[2] Pushpa S.K, Manjunath T.N, "Class result prediction using machine learning", InternationalConference On Smart Technologyfor Smart Nation,2017,pp1208-1212. Micheal Bowles, Machine Learning in Python: Essential Techniques for Predictive Analysis. John Wiley & Sons, Inc. 2015.

[3] Trishul Chilimbi, Yutaka Suzue, Johnson Apacible, and Karthik Kalyanaraman. Project Adam: Building an efficient and scalable deep learning training system. In 11th USENIX Symposium on Operating Systems Design and Implementation (OSDI 14), pages 571–582, 2014. www.usenix.org/system/files/conference/osdi14/osdi14paper-chilimbi.pdf.

[4] Jack Clark. Google turning its lucrative web search over to AI machines, 2015. www.bloomberg.com/news/articles/2015-1026/googleturning- its-lucrative-web-search-over- to-aimachines.

[5] J. Xu, K. H. Moon, and M. Van Der Schaar, "A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs," IEEE J. Sel. Top. Signal Process., vol. 11, no. 5, pp. 742–753, 2017.

[6] Prasadu Peddi (2018), Data sharing Privacy in Mobile cloud using AES, ISSN 2319-1953, volume 7, issue 4.

[7] A. M. Shahiri, W. Husain, and N. A. Rashid, "A Review on Predicting Student's Performance Using Data Mining Techniques," in Procedia Computer Science, 2015.

[8] Y. Meier, J. Xu, O. Atan, and M. Van Der Schaar, "Predicting grades," IEEE Trans. Signal Process. vol. 64, no. 4, pp. 959–972, 2016

[9] Prasadu Peddi (2017) "Design of Simulators for Job Group Resource Allocation Scheduling In Grid and

ISSN: 2057-5688

Cloud Computing Environments", ISSN: 2319- 8753 volume 6 issue 8 pp: 17805-17811.

[10] Uday Chandrakant Patkar, Sushas Haribabu Patil and Prasad Peddi, "Translation of English to Ahirani Language", *International Research Journal of Engineering and Technology(IRJET)*, vol. 07, no. 06, June 2020.