iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 2 **Issue:** XI **Month of publication:** November 2014

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Predicting the Performance of Students in Higher Education Using Data Mining Classification Algorithms - A Case Study

Jai Ruby[1], Dr. K. David[2]
[1] Research Scholar, Research & Development Centre, Bharathiar University, Tamilnadu, India
[2]Associate Professor, Dept. of Computer Science and Engg, Rover Engg. College,Perambalur, Tamilnadu, India

Abstract: ***In recent years, Educational Data Mining has put on a mammoth recognition within the research realm and it has become a vital need for the academic institutions to improve the quality of education. Higher education does categorize the students by their academic performance. In higher education institutions a substantial amount of knowledge is hidden and need to be extracted using Knowledge Discovery process. Data mining helps to extract the knowledge from available dataset and should be created as knowledge intelligence for the benefit of the institution. Many factors influence the academic performance of the students. The factors that describe student performance can be used for predicting students performance by using a number of well - known data mining classification algorithms such as ID3, REPTree, Simplecart, J48, NB Tree, BFTree, Decision Table, MLP and Bayesnet. The study model is mainly focused on analyzing the prediction of the academic performance of the students by all the classification algorithms. The algorithms are analyzed based on the precision of predicting the result.***
Key Words: Educational Data Mining, Academic Performance, Higher Education, Prediction, Classification

I. INTRODUCTION

Educational data has become a vital resource in this modern era, contributing much to the welfare of the society. Educational institutions are becoming more competitive because of the number of institutions growing rapidly. To stay afloat, these institutions are focusing more on improving various aspects and one important factor among them is quality learning. For providing quality education and to face new challenges, the institutions need to know about their potentials which are explicitly seen and which are hidden. The truths behind today's educational institutions are a substantial amount of knowledge is hidden. To be competitive, the institutions should identify their own potentials hidden and implement a technique to bring it out. In recent years, Educational Data Mining has put on a mammoth recognition within the research realm as it has become a vital need for the academic institutions to improve the quality of education. The higher education institutions has potential knowledge such as academic performance of students, administrative accounts, potential knowledge of the faculty, Non academic scores of students, demographic details of the students and many other information in a hidden form. For the higher education institutions to enhance their quality it is a must for them to extract a substantial amount of hidden knowledge. The technique behind the extraction of the hidden knowledge is Knowledge Discovery process. In higher education institutions a substantial amount of hidden knowledge need to be extracted using Knowledge Discovery process. Data mining helps to extract the knowledge from available dataset and should be created as knowledge intelligence for the benefit of the institution. Higher education does categorize the students by their academic performance. Many factors influence the academic performance of the student. The study model (Jai Ruby and K. David, 2014) is mainly focused on exploring various indicators that have an effect on the academic performance of the students. The extracted information that describes student performance can be stored as intelligent knowledge for decision making to improve the quality of education in institutions. The knowledge stored is used for predicting the students' performance in advance.

Data mining, also called Knowledge Discovery in Databases (KDD), is the field of discovering and extracting hidden and potentially useful information from large amounts of data. Data mining is applied in various fields like medical, marketing, databases, machine learning, artificial intelligence, and customer relations etc., Recently Data mining is widely used on educational dataset. Educational Data mining (EDM) has become a very useful research area (Baker R & Yacef K, 2009). Educational Data Mining refers to techniques, tools, and research designed for automatically extracting meaning from large repositories of data generated by or related to people's learning activities in educational settings. Key uses of EDM include learning and predicting student performance in order to recommend improvements to current educational practice. EDM can be considered as one of the learning sciences, as well as an area of data mining (Monika Goyal & Rajan Vohra,2012). The

researcher applied the educational data mining concerns with developing methods for discovering knowledge from data that come from educational domain and used to analyze learning behavior. Students data was collected, preprocessed and data mining techniques were applied to discover association, classification, clustering and outlier detection rules and in all tasks knowledge was extracted that described students behaviour (Alaa El-Halees,2009). Some of the benefits of data mining in an education sector are identifying students' preferences towards course choices, their selection of specialization and predicting students' knowledge, grades, and final results (Mohd Maqsood Ali,2013). Sir Francis Bacon (1597) commented, "Knowledge is power" and in today's context it may be rephrased as "Knowledge sharing is power". The extracted information from the data can be transferred as knowledge and can be stored in decision making for the betterment of the institution. Institutions of Higher Learning (IHL) are similar to knowledge businesses, in that both are involved in knowledge creation, dissemination, and learning (Rowley, 2000). However, people in business world are concerned with the profit they could gain by exploiting knowledge through the implementation of KMS whereas IHL consider that KMS could improve the quality of service deliveries and sustained competitive advantages in the academic world (Lubega, J. T., Omona, W., & Weide, T. V. D., 2011).

This paper makes a novel attempt to look into the higher educational domain of data mining to analyze the students' performance. Section 2 gives the overview of data mining techniques available to extract the hidden information. Section 3 provides the general account of the data under study and details the classification techniques. Section 4 predicts results by various classification algorithms and the comparative analysis. Conclusion and a discussion on future work are in the final section.

A. Related Work

(Zlatko J. Kovacic & John Steven Green, 2010) have predicted student's academic performance using various attributes like gender, parent education, economic background etc. (M.N. Quadri & N.V. Kalyankar, 2010) explained that the previous academic result plays a key task to predict the students who are a threat to be unsuccessful in the exam. (Bhardwaj & Pal, 2011) performed a study on the student performance among 300 students. By means of Bayesian classification method on 17 attributes, it is noted that the influencing factors like grade in senior secondary exam, medium of teaching, living location, mother's qualification, other habit, income and family status plays a vital role in the student academic performance. (Pandey & Pal, 2011) performed a study on the student performance on 600 students. The attributes like category, language and background qualification were used to predict student performance. (Hijazid & Naqvi, 2006) performed a study on the student performance by selecting a sample of 300 students (225 males, 75 females). The study reveals that "Student's attendance, hours spent in study, family income and mother's education are significantly related with student performance" using linear regression. If data mining techniques such as clustering, decision tree and association are applied to higher education processes, it would help to improve students' performance.

(Al-Radaideh Q., Al-Shawakfa E. & AI-Najjar M,2006) used data mining classification techniques to enhance the quality of the higher educational system by evaluating students' data that may affect the students' performance in courses. They used three different classification methods ID3, C4.5 and the NaiveBayes. The results indicated that the decision tree model had better prediction accuracy than the other models. (Z. J. Kovacic,2010) presented a study on educational data mining to identify up to what extent the enrolment data can be used to predict student's success. The algorithms CHAID and CART were used. (K.Shanmuga Priya & A.V.Senthil Kumar, 2013) applied a Classification Technique in Data Mining to enhance the student's performance by extracting the discovery of knowledge from the end semester mark.

Mohammed M. Abu Tair & Alaa M. El-Halees (2012) applied the data mining for discovering knowledge from data that come from educational environment. Student's data has been collected from the college of Science and Technology for a period of 15 years [1993-2007]. The collected data was preprocessed and data mining techniques are applied to improve graduate students' performance, and overcome the problem of low grades of graduate students. (Muslihah W., Yuhanim Y., Norshahriah W., Mohd Rizal M., Nor Fatimah A., & Hoo Y. S., 2009) have compared Artificial Neural Network and the combination of clustering and decision tree classification techniques for predicting and classifying student's academic performance. Students' data were collected from the data of the National Defense University of Malaysia (NDUM). (H. W. Ian & F. Eibe,2005) gave a case study that used educational data mining to identify behavior of failing students to warn students at risk before final exam. (S. B. Kotsiantis, C. J. Pierrakeas & P. E. Pintelas,2003) have compared six classification methods (Naive Bayes, Decision Tree, Feed-forward Neural Network, Support Vector Machine, 3-nearest Neighbor and Logistic Regression) to predict drop-outs in a course. The data set contained records of 350 students. Their best classifiers, Naive Bayes and Neural Network, were able to achieve a performance of 80%.

(Nguyen N., Paul J. & Peter H., 2007) compared the accuracy of decision tree and Bayesian network algorithms for predicting the academic performance of students of Under Graduate and Post Graduate students. The decision tree classifier provided better accuracy than the Bayesian network classifier. (Ramaswami M., & Bhaskaran R,2010) have constructed a predictive model using 772 students' records with 7-class response variables by using highly influencing predictive variables obtained

through feature selection. The accuracy of the present model was compared with other models and it has been found to be satisfactory. (Brijesh Kumar Baradwaj, Saurabh Pal,2011) conducted a study on a data set of size 50 MCA students for mining educational data to analyze students' performance. Decision tree method was used for classification and to predict the performance of the students. Different measures that are not taken into consideration were economic background; technology exposure etc. (Han & M. Kamber, 2001) depicted the data mining process and the methods to analyze data from different perspective and the steps to mine knowledge. (Monika Goyal & Rajan Vohra, 2012) discussed on various applications of data mining in order to improve the quality of higher education. (Anupama Kumar & Vijayalakshmi, 2011) have used classification algorithms like C4.5 and Random Tree to mine the records and predict the outcome of the students in the V semester. The algorithms are analyzed using parameters like the number of instances predicted correctly, algorithm accuracy, and precision value of the algorithm. (Y. Bengio, J. M. Buhmann, M. Embrechts, and J. M. Zurada, 2000) discussed that neural networks are suitable in data-rich environments and are typically used for extracting embedded knowledge in the form of rules, quantitative evaluation of these rules, clustering, self-organization, classification and regression. Neural networks have an advantage, over other types of machine learning algorithms, for scaling.

(J. A. Moriana, F. Alos, R. Alcala, M. J. Pino, J. Herruzo, and R. Ruiz, 2006) studied the possible influence of extra- curricular activities like study-related (tutoring or private classes, computers) and/or sports-related (indoor and outdoor games) on the academic performance of the secondary school students in Spain. A total number of 222 students from 12 different schools were the samples and they were categorized into two groups as a function of student activities (both sports and academic) outside the school day. Analysis of variance (ANOVA) was used to verify the effect of extracurricular activities on the academic performance and it was observed that group involved in activities outside the school yielded better academic performance.

II. DATA MINING AND CLASSIFICATION TECHNIQUES

Data mining also termed as Knowledge Discovery in Databases (KDD) refers to extracting or "mining" knowledge from large amount of data. (Han & M. Kamber,2001) Knowledge Discovery process involve various steps like Data cleaning, transformation, data mining, pattern evaluation in extracting knowledge from data. Knowledge Discovery is involved in a multitude of tasks such as association, clustering, classification, prediction, etc. Classification and prediction are functions which are used to create models that are constructed by analyzing data and then used for assessing other data. Clustering is a way of identifying similar classes of objects. Association is mainly used to relate frequent item set among large data sets. Data mining for small data sets has a real potential to become a serious part of higher education teachers' Knowledge Management Systems (Srecko Natek, Moti Zwilling, 2013).

This study is carried out using a small dataset with a number of attributes to predict the performance of the students using classification algorithms. Two steps are involved in classification. Building a model and using the model for classification. In the first step, a model that describes a predetermined set of classes or concepts is made by examining a set of training dataset. The learning is known as supervised learning as the class labels of all the records of the dataset are known. The models are usually in the form of classification rules or decision tree. In the second step, the model is put to test using a different data set that is used to estimate the predictive accuracy of the model. Various methods like holdout, random sub sampling, k-fold cross validation, stratified cross validation, bootstrapping are used to estimate the accuracy of the model. If the accuracy of the model is considered acceptable, the model can be used to classify the dataset for which the class label is not known in advance (Han & M. Kamber,2001).

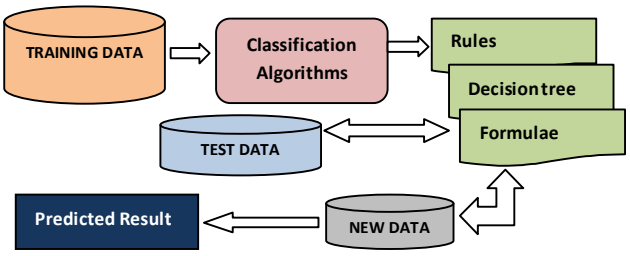


Fig. 1 – Learning, Classification and Prediction

Classification techniques can be applied on the educational data for predicting the students' behavior, performance in examination etc.

Basic techniques for classification are decision tree induction, Bayesian classification and neural networks. Other approaches like genetic algorithms, rough sets, fuzzy logic, case based reasoning can also be used for classification. Decision Tree classifier is a powerful and popular classification and prediction technique (Chaudhuri, 1998). Some of the decision tree classifiers are J48, NBTree, ID3, CART, REPTree, Simplecart, BFTree and others. A Decision Tree is a flow-chart like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test and leaf nodes represent cases or class distributions (Han & M. Kamber,2001). These decisions generate rules for the classification of a dataset. Each rule can be created for each path from the root to a leaf. The leaf node holds the class prediction.

ID3 - Iterative Dichotomiser 3 is a decision tree Induction algorithm. The ID3 algorithm was originally developed by J. Ross Quinlan at the University of Sydney. The ID3 algorithm induces decision trees from data. It is a supervised learning algorithm that is trained by examples for different classes. After being trained, the algorithm should be able to predict the class of a new item. To determine which attributes are the most important, ID3 uses the statistical property of entropy. Entropy measures the amount of information in an attribute. This is how the decision tree, which will be used in testing future cases, is built.

J48 - is an open source Java implementation of the C4.5 algorithm in the Weka data mining tool. This algorithm was developed by Ross Quinlan. C4.5 is a program that creates a decision tree based on a set of labeled input data by recursive partitioning of data. The decision is grown using Depth-first strategy. The decision trees generated by C4.5 can be used for classification. The algorithm considers all the possible tests that can split the data set and selects a test that gives the best information gain. In every internal node the condition of some attribute is being examined, and every branch of the tree represents an outcome of the study. The branching of the tree ends with leaves that define a class to which examples belong.

NBTree - The Naive Bayesian tree learner, NBTree (Kohavi 1996), combined Naive Bayesian classification and decision tree learning. Bayesian classifiers are statistical classifier. The Naive Bayes algorithm is a simple probabilistic classifier that calculates a set of probabilities by counting the frequency and combinations of values in a given data set. In an NBTree, a local naive Bayes is deployed on each leaf of a traditional decision tree, and an instance is classified using the local naive Bayes on the leaf into which it falls. After a tree is grown, a naive Bayes is constructed for each leaf using the data associated with that leaf. An NBTree classifies an example by sorting it to a leaf and applying the naive Bayes in that leaf to assign a class label to it.

MLP - Multilayer Perceptron algorithm is one of the most widely used and common neural networks. Multilayer Perceptron is a feed forward artificial neural network model trained with the standard back propagation algorithm that maps sets of input data onto a collection of acceptable output. An MLP consists of multiple layers of nodes in a directed graph, with every layer totally connected to the consequent one. These are supervised networks so they require a desired response to be trained. They learn how to transform input data into a desired response, so they are widely used for pattern classification and prediction.

SimpleCart - It is a non-parametric decision tree learning technique that produces classification trees, depending on whether the dependent variable is categorical or numeric. CART (Classification and Regression Trees) is very similar to C4.5, but it differs in that it supports numerical target variables (regression) and does not compute rule sets. CART constructs binary trees using the feature and threshold that yield the largest information gain at each node. Trees are grown to a maximal size without the use of a stopping rule and then pruned back to the root via cost-complexity pruning. The next split to be pruned is the one contributing least to the overall performance of the tree on training data. Tree performance is always measured on independent test data and tree selection proceeds only after test-data-based evaluation.

REPTree - *It* is a fast decision tree learner. It builds a decision/regression tree using information gain/variance and prunes it using reduced-error pruning (with back-fitting). The algorithm only sorts values for numeric attributes once. Missing values are dealt with by splitting the corresponding instances into pieces.

Decision Table - Decision Tables are classification models elicited by machine learning algorithms and are used for creating predictions. A decision table consists of a hierarchical table within which entry in a higher level table gets broken down by the values of a pair of additional attributes to make another table.

III. METHODOLOGY

The dataset used for this study for performance analysis was taken from PG Computer Application course offered by an Arts and Science College between 2007 and 2012. The data of 165 students were collected. Student personal and academic details along with their attendance were collected from the student information system. The collected information was integrated into a distinct table. Student dataset contains various attributes like Theory Scores, Laboratory scores, Medium of study, UG course, Family Income, Parental Education, First Generation Learner, Stay, Extracurricular activities etc. Among the different attributes initially present using feature selection techniques like chi square, info gain, gain ratio, correlation and regression it is found that the high impact attributes that contribute for the performance of the students are Theory, Medium of Study, Previous Course

studied, UG Percentage, Stay, Extra Curricular Activities and Family Income (Jai Ruby and K. David, 2014). The influencing attributes are selected and are used to classify and predict the student performance using weka data mining tool.

Weka (Waikato Environment for Knowledge Analysis) is a popular suite of machine learning software written in Java, developed at the University of Waikato, New Zealand. Weka is a free software available under the GNU General Public License. The Weka workbench contains a collection of visualization tools and algorithms for data analysis and predictive modelling, together with graphical user interfaces for easy access to this functionality. Weka tool contains many packages which include Filters, Classifiers, Clusters, Associations, and Attribute Selection. The Visualization tool in weka allows datasets and the predictions of Classifiers in a pictorial form. Weka is a collection of machine learning algorithms for solving real-world data mining problems. It is written in Java and runs on almost any platform. The algorithms can either be applied directly to a dataset or called from your own Java code.

In Weka datasets should be formatted to the ARFF format. The initial dataset of 165 records was split up into two sets. Two thirds of the data are allocated to the training set and the remaining one third is allocated to the test set. The training set help in building the model and it is used for classification. For estimating classifier accuracy k-fold cross-validation is used. Training and testing is performed k-times. The accuracy estimate is the overall number of correct classifications from the k iterations divided by the total number of samples in the initial data (Han & M. Kamber,2001). The classify panel in the weka tool facilitates to apply classification algorithms and to estimate the accuracy of the predictive model. Seven different classifiers ID3, J48, NBTree, RepTree, Multi Layer Perceptron (MLP), SimpleCart and Decision table were used in the study. These classifiers are some of the common classifiers used in most practical classification problems. This study analyse the accuracy of the algorithms in predicting the performance of the student.

IV. RESULTS AND DISCUSSION

Rule, Function and Tree based classification algorithms were used to analyse the student data. ID3, J48, NBTree, REPTree, SimpleCart were tree based, MLP was a Neural Network function and Decision Table a rule based algorithm. The student data set of 165 records was split into two sets consisting of two-third as training set and one-third as testing set. The training set is used to build a model and the test set is used to estimate the accuracy of the classifier. The Fig. 2 shows the classification of the training data set using MLP algorithm and Fig. 3 shows the classification of test data using MLP algorithm via weka tool. If the accuracy of the model is acceptable then it is used for the prediction of data for which the class label is unknown. The sample set was divided into 5 sets of distinct two-third records and 5 sets of distinct one-third records. They are the data set used in Run1 through Run5 respectively. In each run, three new set of data whose class labels are unknown was given for prediction. Since we use three different sets of new data for prediction, the average of the three results was considered for each run. Since we have 5 training data set and 5 test data set we get 5 results for each algorithm. The experiment shows the accuracy of the different Classification algorithms. Fig. 4 shows the prediction of new test data whose label is unknown using MLP algorithm via weka.

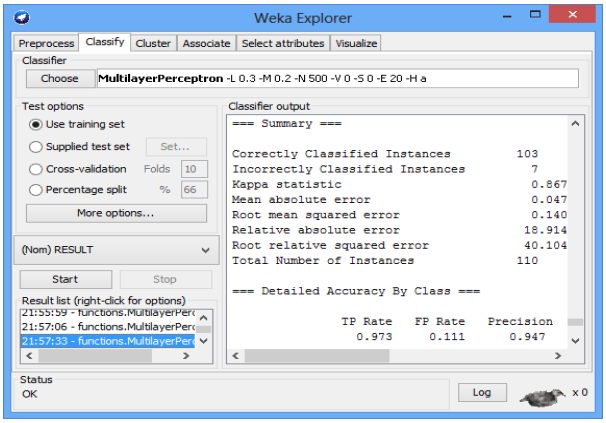


Fig 2 : Classification of Training data set using MLP

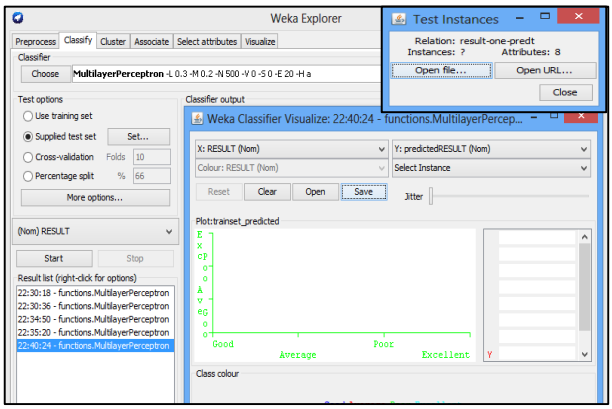


Fig. 3 : Classification of Test data set using MLP

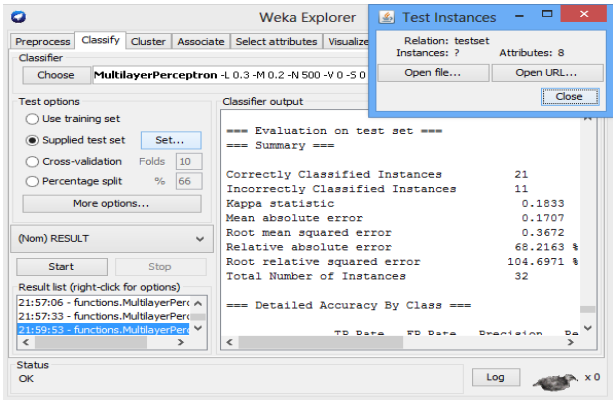


Fig. 4 : Prediction of New test data set using MLP

Algorithm	Run - 1	Run - 2	Run - 3	Run - 4	Run – 5
ID3	82.1	65.15	61.2	75.9	80.3
J48	69.1	59	69.3	69.1	77.4
REP Tree	69.8	60.1	68.3	79.5	76.75
Simple Cart	69.8	57.6	68.3	79.5	76.8
NB Tree	75.7	62.1	59.9	79.6	78.3
MLP	78.1	58.3	71.2	93.9	72.2
Decision Table	69.8	58.1	69.3	69.2	74.1

Table 1 : Prediction Accuracy of the Classification Algorithms in percentage

Table-1 shows prediction accuracy of the Classification Algorithms in percentage for a different combination of training data set, testing data set and new datasets for which prediction of class label was carried out. The average of the 5 runs of the different algorithms was calculated and tabulated in Table-2.

ID3	J48	REP Tree	Simple Cart	NB Tree	MLP	Decision Table
73	68.8	70.9	69.5	71.1	74.8	68.1

Table 2 : Average Accuracy of 5 runs of various Classification Algorithms in percentage

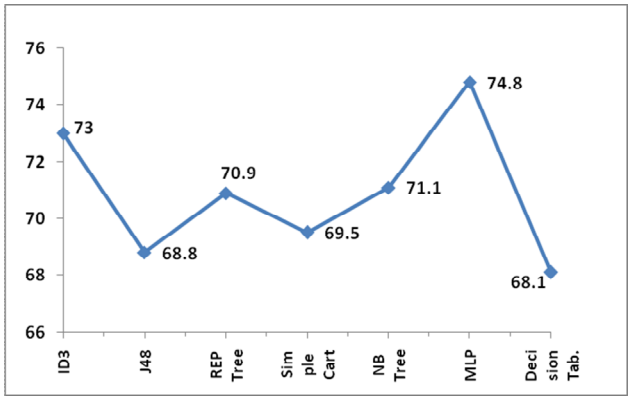


Fig. 5: Prediction accuracy of various Data mining Algorithms

Fig.5 shows the prediction accuracy of various data mining algorithms. The results show that MLP, a neural network based classification show best result of 74.8% accurate prediction which is followed by ID3 showing an accuracy of 73%. NBTree and REPTree classification tree based algorthims show an accuracy of around 71% and other Classifiication algorithms of J48, SimpleCart and Decision Table show a prediction accuracy of 68.8%, 69.5% and 68.1% respectively.

V. CONCLUSION

This paper deals with the prediction of student academic performance using various classification algorithms. This study paper help the institution to know the academic status of the students in advance and can concentrate on weak students to improve their academic results. All the classification algorithms MLP, ID3, J48, REPTree, NBTree, SimpleCart and Decision Table considered in the study shows prediction accuracy above 68% for the student dataset. The study was carried out using only a small dataset and it can be extended to a large dataset. Also, the educational data set can be tested using a combination of two or more classification algorithms for better prediction which would be the future work.

REFERENCES

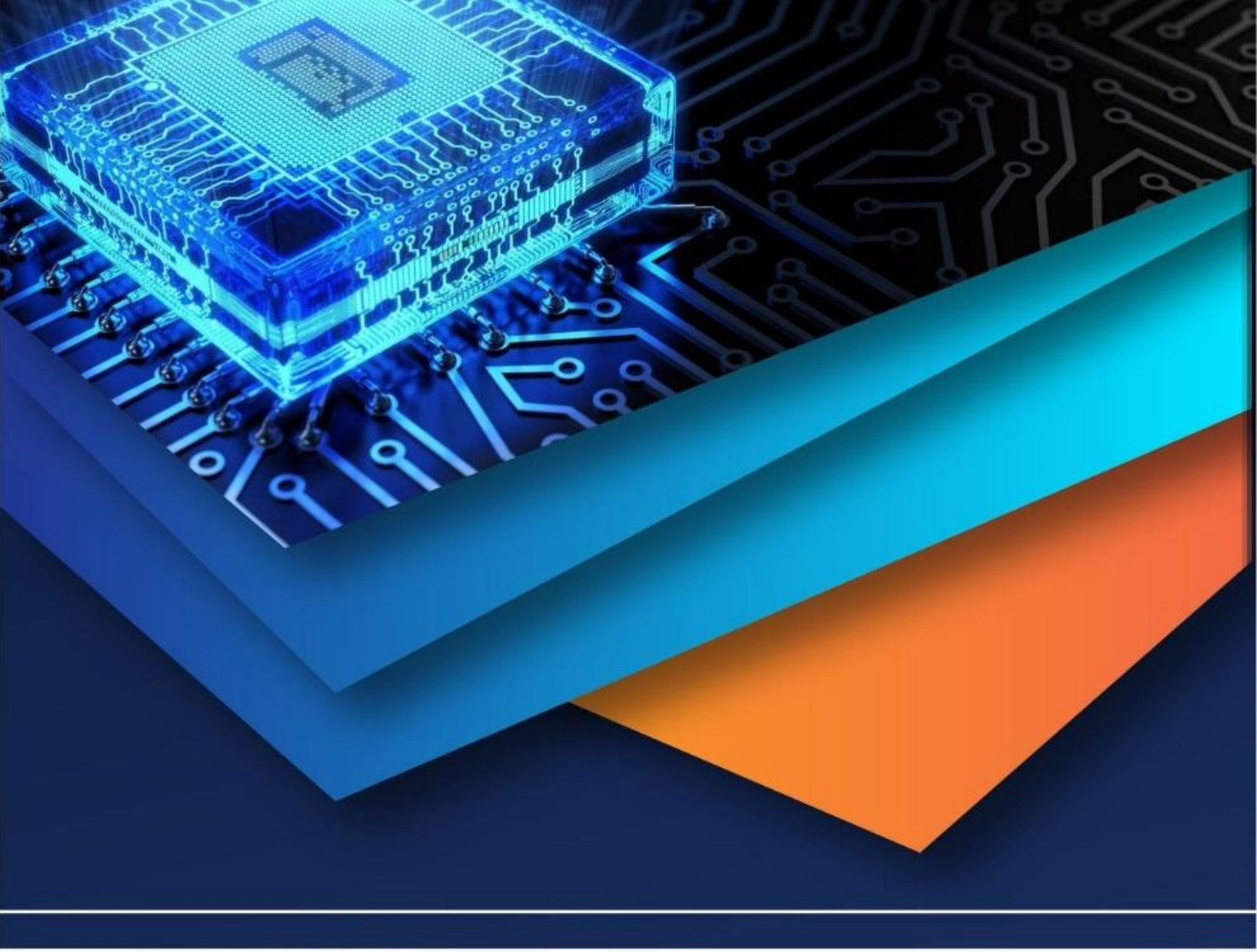
[1] Alaa M El-Hales, "Mining Educational Data to Analyze Learning Behaviour A case study", 2009
[2] Al-Radaideh Q., Al-Shawakfa E., & AI-Najjar M., Mining Student Data using Decision Trees, In Proceedings of the International Arab Conference on Information Technology (ACIT'2006), Yarmouk University, Jordan, 2006.
[3] Anupama Kumar. S & Vijayalakshmi M.N., "Envision of student"s concert using supervised learning techniques", IJCES International Journal of Computer Engineering Science , Volume1 Issue2, November 2011 ISSN : 2250:3439
[4] Baker R.S.J.D., & Yacef K, 2009, The state of educational data mining in 2009: A review and future vision, Journal of Educational Data Mining, I, 3-17.
[5] Bengio Y., Buhmann J. M., Embrechts M., & Zurada J. M., "Introduction to the special issue on neural networks for data mining and knowledge discovery," IEEE Trans. Neural Networks, vol. 11, pp. 545-549, 2000.
[6] Bharadwaj. B.K & S. Pal. "Data Mining: A prediction for performance improvement using classification", International Journal of Computer Science and Information Security (IJCSIS), Vol. 9, No. 4, pp. 136-140, 2011.
[7] Brijesh Kumar Baradwaj & Saurabh Pal, IJACSA, Vol.2, No.6,2011,"Mining Educational Data to Analyze Students' Performance"
[8] Cesar V., Javier B., liela S., & Alvaro O., Recommendation in Higher Education Using Data Mining Techniques, In Proceedings of the Educational Data Mining Conference, 2009.

[9] Cristobal Romero & Sebastian Ventura, "Educational Data Mining: A Review of the State of the Art," IEEE Transactions on Systems, Man, and Cybernetics—Part c: Applications and Reviews, vol. 40, no. 6, 2010, pp. 601-618.
[10] Han. J & Kamber. M (2001). "Data mining concepts and techniques", San Francisco, USA, Morgan Kaufmann.
[11] Hijazi S. T., &. Naqvi R. S. M. M, "Factors affecting student's performance: A Case of Private Colleges", Bangladesh e-Journal of Sociology, Vol. 3, No. 1, 2006.
[12] http://en.wikipedia.org/wiki/Weka_(machine_learning)
[13] http://www.cs.waikato.ac.nz/ml/weka/
[14] Ian H. W. & Eibe F., "Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations," California: Morgan Kaufmann, 2005
[15] Jai Ruby & K. David, "A study model on the impact of various indicators in the performance of students in higher education", IJRET International Journal of Research in Engineering and Technology, Vol. 3, Issue 5, May-2014, pp.750-755.
[16] Kotsiantis, S. B., Pierrakeas C. J., & Pintelas P. E., "Preventing Student Dropout in Distance Learning Using Machine Learning Techniques," In proceedings of 7th International Conference on Knowledge-Based Intelligent Information and Engineering Systems (KES 2003), pp. 267-274, 2003. ISBN 3-540-40803-7.
[17] Kovacic Z. J., "Early prediction of student success: Mining student enrollment data", Proceedings of Informing Science & IT Education Conference 2010.
[18] Lubega, J. T., Omona, W., & Weide, T. V. D., "Knowledge management technologies and higher education processes_: approach to integration for performance improvement", International Journal of Computing and ICT Research, 2011, vol. 5(Special Issue), pp. 55–68.
[19] Mohammed M. Abu Tair & Alaa M. El-Halees, Mining Educational Data to Improve Students' Performance: A Case Study, 2012
[20] Mohd Maqsood Ali, International Journal of Computer Science and Mobile Computing Vol.2 Issue. 4, April- 2013, pg. 374-383
[21] Monika Goyal & Rajan Vohra, "Applications of Data Mining in Higher Education" IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 1, March 2012, pp.130-120.
[22] Moriana J. A., Alos F., Alcala R., Pino M. J., Herruzo J., & Ruiz R., "Extra Curricular Activities and Academic Performance in Secondary Students", Electronic Journal of Research in Educational Psychology,Vol. 4, No. 1, 2006, pp. 35-46
[23] Muslihah W., Yuhanim Y., Norshahriah W., Mohd Rizal M., Nor Fatimah A., & Hoo Y. S., Predicting NDUM Student's Academic Performance Using Data Mining Techniques, In Proceedings of the Second International Conference on Computer and Electrical Engineering, IEEE computer society, 2009.
[24] Nguyen N., Paul J., & Peter H., A Comparative Analysis of Techniques for Predicting Academic Performance. In Proceedings of the 37th ASEE/IEEE Frontiers in Education Conference. pp. 7-12, 2007.
[25] Pandey U.K & S.Pal, Data Mining A prediction of performer or underperformer Using classification IJCSIT International Journal of Computer Science and Information Technology, Vol. 2(2),pp.686-690,ISSN:0975-9646,2011
[26] Quadril.M.N & Dr. Kalyanka N.V. Drop Out Feature of Student Data for Academic Performance Using Decision Tree, Global Journal of Computer Science and Technology Vol. 10 Issue 2 (Ver 1.0), April 2010.
[27] Ramaswami M., & Bhaskaran R., CHAID Based Performance Prediction Model in Educational Data Mining, IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 1, No. 1, 2010.
[28] Rowley, J., "Is higher education ready for knowledge management?",International Journal of Educational Management, 2000, vol. 14(7), pp. 325–333.
[29] Shanmuga Priya. K, & Senthil Kumar A.V., Improving the Student's Performance Using Educational Data Mining, 2013.
[30] Srecko Natek & Moti Zwilling, 2013, "Data mining for small student data set – knowledge management system for higher education teachers"
[31] Zlatko J. Kovacic & John Steven Green, Predictive working tool for early identification of 'at risk' students, New Zealand, 2010.

BIOGRAPHIES

Jai Ruby is a Research Scholar in Research & Development Centre, Bharathiar University, Tamilnadu, India. She has 14 years' experience in teaching field and research. Her current areas of research are Data Mining and Mobile Communications.

Dr. K. David is working as an Associate Professor and Head, Dept. of Computer Science and Engg, OAS Institute of Tech. & Mgt., Tamilnadu, India. He has over 15 years of teaching experience and about 4.5 years of Industry experience. He has published scores of papers in peer reviewed journals of national and international repute and is currently guiding 6 Ph.D scholars. His research interests include, UML, OOAD, Knowledge Management, Web Services and Software Engineering.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY