

## A COMPARISON OF STUDENTS PERFORMANCE USING DECISION TREES TECHNIQUES

<sup>1</sup>B.Sundarraaj, <sup>2</sup>M.Sriram

<sup>1,2</sup>Asst. Professor Dept.of CSE,BIST,BIHER,Bharath University, Chennai.

<sup>1</sup>sundarraaj.cse@bharathuniv.ac.in, <sup>2</sup>sriramm.cse@bharathuniv.ac.in

**Abstract:** Data mining techniques can be used in different fields to extract knowledge from large data. Educational data mining is an emerging discipline concerned with developing methods for exploring the unique types of data that comes from educational setting. In research of data mining, machine learning and statistics to information generated from educational setting (eg.universities and intelligent tutorial systems). Educational field identify the techniques of the weak students. The educational institutes are always giving quality education. By using prediction method a model can be developed which can be used to predict student performance. Prediction can be done by using students' academic background and family background. Different Data Mining Techniques like Classification, Clustering, Associations education data mining with decision tree. Mining, Regression etc. can be used for this purpose of prediction of student academic performance in education data mining using with decision tree algorithm. The prediction will help them to improve their performance.

**Keywords :** Educational data mining, Classification, keywords, Decision Tree.

### 1. Introduction

Data Mining is the process of discovering new knowledge from large data set. Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that comes from educational settings and using those methods to better understand students and the settings which they learn. There are different types of educational environments[1-3].off line education (traditional class room): it is used to deliver knowledge and skills are based on face to face contact. In this we have to consider students behaviour performance, curriculum etc. that was gathered in classroom environment.

1. E-Learning and learning management system (LMS)- E-learning providing online instructions. LMS also provides communication, collaboration, administration, and reporting tools. Web mining techniques have been applied to

student's data stored by these systems in log files and databases[4-6].

2. Intelligent tutoring system (ITS) AND Educational hypermedia system (AEHS)It adapts teaching to the needs of each particular student. DM has been applied to data pick up these systems, such as log files, user's models[7].

The performance of third year students depends on factors such as their family background like father's education,fathers' occupation, mother education, mothers' occupation. Predicting students' academic performance, those students are on risk can be find out some remedial action can be taken to improve their performance. Due to this results of the college can improve. The educational data mining can be used to get the feedback for the teachers so the teacher can improve the level of teaching[8].

There are many data mining techniques can be used for this like classification, regression,association rule mining[9].

### 2. Literature Survey

Data mining is the process of analysing data from different prospective and summarizing it into useful information.

1. According to CRISTO OBEL ROMERO & SEBASTIAN VENTURA EDM can be used for predicting students'performance, to get feedback to the teachers, to improve teaching learning method. Different data mining techniques can be used for this like clustering,classification,association rule mining[10].

1. Students performance education in Online Education System vs Traditional Educational System(Udeni Jayasinghe,Anuja Dharmaratne,Ajantha Atukorale)Now a day's most of the educational institutes practice online teaching mechanism rather than using the traditional teacher centred teaching mechanism to enhance the learning ability of the student's centred learning mechanism[11].

2. According to ABEER BADR EL DIN AHMED, IBRAHIM SAYED ELARABY, A prediction for students performance using classification method , currently the amount huge of data stored in educational

database contain the useful information for predict of students performance.

3. CHARANJIT BAMBRAH, MINAKSHI BHANDARI, NIRALI MANIAR, PROF VANDANA MUNDE suggested Data mining is used to extract meaningful information and to develop relationship among variables stored in large Data set.

4. YADAV SURJEET KUMAR AND PAL SAURABH suggested C4.5, ID3, and CART decision tree algorithms can be used to predict the performance of the first year engineering students. It was three classifications as class predictions students were classification pass fail and promoted. This model was good to identifying the students that are most likely to fail.

### 3. Proposal

➤ **DECISION TREE:-**A decision tree is a graph that uses a branching method to illustrate every possible outcome of a decision[11].

➤ Decision tree can be drawn by hand or created with a graphics program or specialized software. Decision tree software is used in data mining to simplify complex is used in Data mining to simplify complex is used in data mining to simplify complex strategic challenges and evaluates the cost effectiveness of research and business decision[12-15].

➤ **TYPES OF DECISION TREE:**

1. ID3(Iterative dicotomiser3)
2. CART (Classification and regression tree)
3. C4.5

➤ 1.ID3(Iterative dichotomiser3)

The ID3 decision tree algorithm is developed by ROSS QUINLAN. The algorithm generates an unpruned full decision tree from a dataset.

➤ 2. CART(Classification and regression tree)

Classification and regression tree is another decision tree algorithm which uses minimal cost complexity pruning.

The main elements of CART are:

I.Rules for splitting data at a node based on the value of one variable.

II.Stopping rules for deciding when a branch is terminal and can be split no more.

III.A prediction for the target variable in each terminal node.

➤ 3.C4.5(J48 algorithm)

➤ The C4.5 decision tree algorithm developed by Ross Quinlan which was the successor of the ID3 algorithm.The C4.5 algorithm are pruning in the generation of a decision tree, where a node could be removed from the tree if it adds little to no value to the final predictive model. C4.5 has the concept of gain ratio as an attribute selection measure to create a decision tree. At first, the gain ratio of every attribute is calculated. The root node is the attribute which has

maximum gain ratio c4.5 uses pessimistic pruning to get rid of unessential branches within the decision tree to enhance the accuracy of classification[16-18].

### 4. Experimental –Setup

1.HARDWARE - WEKA TOOL

2.SOFTWARE – Here we can check the performance of students using decision tree methods.

INTERNAL MARK 1 PASS MARK =25

IF (B2>=25,"PASS,"FAIL")

INTERNAL MARK 2

IF (B2>=25,"PASS,"FAIL")

ASSIGNMENT PASS MARK 10

IF(B2>=10,"PASS,"FAIL")

FOR PERCENTAGE %

B2/210\*100

MODEL

IF (B2>=50,"PASS,FAIL")

SEMESTER

IF(B2>=50,"PASS,"FAIL")

IF

(C2="FAIL","U",IF(B2>90,"A"),IF,(B2>80,"B",IF(B2 >70,"C",IF(B2>60,"D"))))

### 5. Implementation

i). Start weka tool.

ii).Open the data/iris arff dataset. click the “open file”.....button to open a data set and double click on the “data” directory.

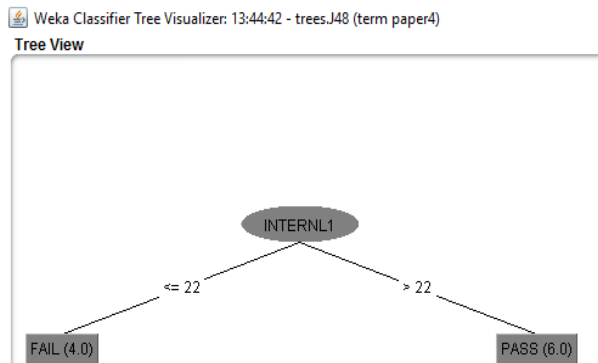
A	B	C	D	E	F	G	H
S NAME	INTERNL	INTRNL2	ASSINM	ATTEN%	MODEL	SEM EXAM	RESULT
PAYAL	35	41	10	79	77	88	PASS
MONI	39	32	10	75	76	90	PASS
SAMPA	38	21	10	65	56	88	PASS
PIYALI	22	32	8	80	78	88	PASS
NITISH	45	39	8	30	88	78	PASS
SUBHAM	41	29	7	55	76	87	PASS
KAJAL	22	20	10	66	79	87	PASS
PUJA	21	21	9	88	54	76	PASS
NAIYNA	26	26	9	80	65	76	PASS
RAKHI	21	37	10	80	77	99	PASS

Figure 1.1. Data before pre-processing

### FOLD CROSS VALIDATION

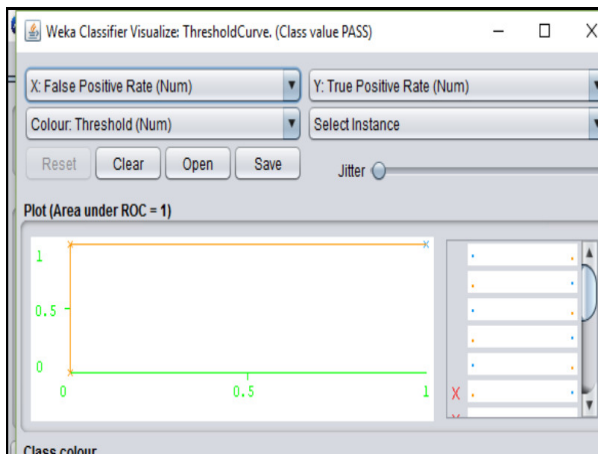
When the data is less then fold cross validation can be used. In this the original sample is randomly partitioned into k attribute[19-20].

**INTERPRETATION**

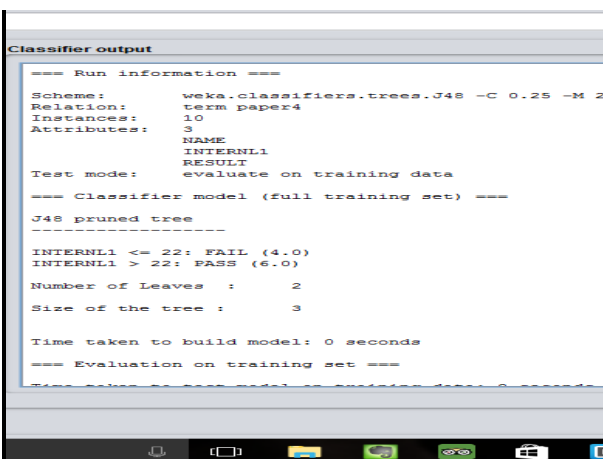


**Figure 1.2.** Prediction for exam

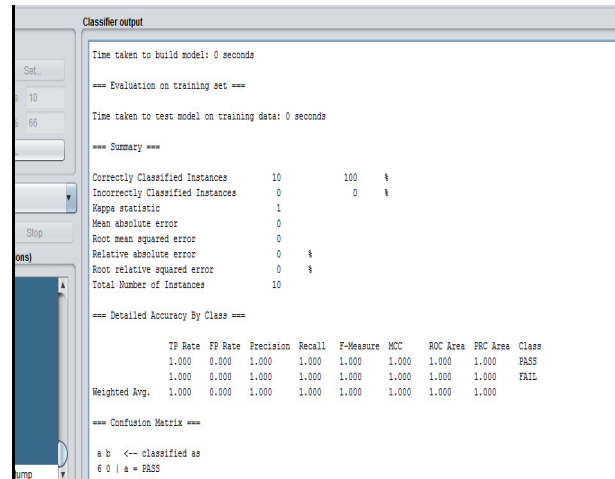
2)These results can show here is for first semester. Taking into consideration the internal marks of the students:



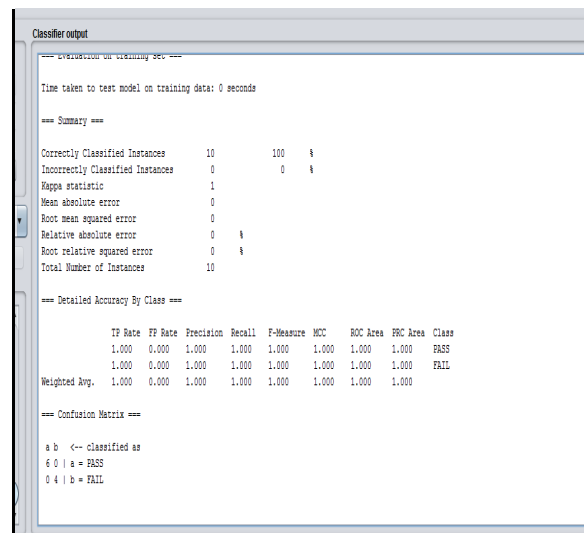
**Figure 1.3.** Ratio of Exam



**Figure 1.4.** Output 1



**Figure 1.5.** Output 2



**Figure 1.6.** Output 3

**6. Conclusion**

The main objective of Educational Data Mining (EDM) is to improve teaching learning process. Predicting students' performance is one of the major applications of EDM. So using decision tree students' performance can be predicted. The students, whose performance is poor, can be warned. The management can take is easy to handle their performance for maximum attention, and taking extra advices etc. Due to such type of measures students' performance may improve .There are number of failure that can be reduced. Ultimately college results also get improved.

## References

- [1] SEBASTIAN VENTURA(2010), "Educational Data Mining" A review of the state of the art ,IEEE transactions on systems ,man, and cybernetics- partc: applications and reviews,vol40 no6, November 2010.
- [2] KUMAR S.ANUPAMA AAD DR.VIJAYALAKSHMI M.N(2011) Efficiency of decision trees in predicting students academic performance computer science and information technology.
- [3] PANDEY UK AND PAL(2011)," Performance prediction of engineering students, using decision trees", International journal of computer applications(0975-8887).
- [4] CHARANJIT BAMBRAH MINAKSHI BHANDARI, NIRALI MANIAR,PROF VANDANA MUNDE "mining association rules in students assessment data" International journal of advanced research in computer and communication engineering vol.3, issue 3,march 2014.
- [5] S.MENAKSHI "Association rule generation for student performance analysis using apriori algorithm" the SIJ transactions on computer science engineering and its applications (CSEA).
- [6] MRINAL PANDEY AND VIVEK KUMAR SHARMA (2013), "A decision tree algorithm pertaining to the students' performance analysis and prediction, "International journal of computer applications (0975-8887).
- [7] BARADWAJ BRIJESH KUMAR AND PAL SAURABH (2011).Mining educational data to analyse student performance.
- [8] YADAV SURJEET KUMAR,PAL SAURABH 'DATA MINING. A prediction for performance improvement of engineering
- [9] Udayakumar R., Kaliyamurthie K.P., Khanaa, Thooyamani K.P., Data mining a boon: Predictive system for university topper women in academia, World Applied Sciences Journal, v-29, i-14, pp-86-90, 2014.
- [10] Kaliyamurthie K.P., Parameswari D., Udayakumar R., QOS aware privacy preserving location monitoring in wireless sensor network, Indian Journal of Science and Technology, v-6, i-SUPPL5, pp-4648-4652, 2013.
- [11] Brintha Rajakumari S., Nalini C., An efficient cost model for data storage with horizontal layout in the cloud, Indian Journal of Science and Technology, v-7, i-, pp-45-46, 2014.
- [12] Brintha Rajakumari S., Nalini C., An efficient data mining dataset preparation using aggregation in relational database, Indian Journal of Science and Technology, v-7, i-, pp-44-46, 2014.
- [13] Khanna V., Mohanta K., Saravanan T., Recovery of link quality degradation in wireless mesh networks, Indian Journal of Science and Technology, v-6, i-SUPPL.6, pp-4837-4843, 2013.
- [14] Khanaa V., Thooyamani K.P., Udayakumar R., A secure and efficient authentication system for distributed wireless sensor network, World Applied Sciences Journal, v-29, i-14, pp-304-308, 2014.
- [15] Udayakumar R., Khanaa V., Saravanan T., Saritha G., Retinal image analysis using curvelet transform and multistructure elements morphology by reconstruction, Middle - East Journal of Scientific Research, v-16, i-12, pp-1781-1785, 2013.
- [16] Khanaa V., Mohanta K., Saravanan. T., Performance analysis of FTTH using GEAPON in direct and external modulation, Indian Journal of Science and Technology, v-6, i-SUPPL.6, pp-4848-4852, 2013.
- [17] Kaliyamurthie K.P., Udayakumar R., Parameswari D., Mugunthan S.N., Highly secured online voting system over network, Indian Journal of Science and Technology, v-6, i-SUPPL.6, pp-4831-4836, 2013.
- [18] Thooyamani K.P., Khanaa V., Udayakumar R., Efficiently measuring denial of service attacks using appropriate metrics, Middle - East Journal of Scientific Research, v-20, i-12, pp-2464-2470, 2014.
- [19] R.Kalaiprasath, R.Elankavi, Dr.R.Udayakumar, Cloud Information Accountability (Cia) Framework Ensuring Accountability Of Data In Cloud And Security In End To End Process In Cloud Terminology, International Journal Of Civil Engineering And Technology (Ijciet)Volume 8, Issue 4, Pp. 376–385, April 2017.
- [20] R.Elankavi, R.Kalaiprasath, Dr.R.Udayakumar, A fast clustering algorithm for high-dimensional data, International Journal Of Civil Engineering And Technology (Ijciet), Volume 8, Issue 5, Pp. 1220–1227, May 2017.
- [21] R. Kalaiprasath, R. Elankavi and Dr. R. Udayakumar. Cloud. Security and Compliance - A Semantic Approach in End to End Security, International Journal Of Mechanical Engineering And Technology (Ijmet), Volume 8, Issue 5, pp-987-994, May 2017.
- [22] Thooyamani K.P., Khanaa V., Udayakumar R., Virtual instrumentation based process of agriculture by automation, Middle - East Journal of Scientific Research, v-20, i-12, pp-2604-2612, 2014.
- [23] Udayakumar R., Thooyamani K.P., Khanaa, Random projection based data perturbation using

geometric transformation, World Applied Sciences Journal, v-29, i-14, pp-19-24, 2014.

[24] Udayakumar R., Thooyamani K.P., Khanaa, Deploying site-to-site VPN connectivity: MPLS Vs IPSec, World Applied Sciences Journal, v-29, i-14, pp-6-10, 2014.

