IJCSEInternational Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.8, Issue.4, Apr 2020 E-ISSN: 2347-2693

A Comparative Study on Student Academic Performance Prediction Using ID3 and C4.5 Classification Algorithms

Kandepi Suneetha

Dept. of CSE, Gayatri Vidya Parishad College of Engineering for Women, Visakhapatnam,Andhra Pradesh, India

DOI: https://doi.org/10.26438/ijcse/v8i4.106111 | Available online at: www.ijcseonline.org

Received: 5/Apr/2020, Accepted: 23/Apr/2020, Published: 30/Apr/2020

***Abstract*:** The ability to predict a student's performance on a given concept is an important tool for the education institutions, as it allows them to understand the ability of students and derive important methods to enhance their knowledge levels. It is the responsibility of educational institutions to have an approximate prior knowledge of their students to predict their performance in future academics and to train them in various activities. It is used to identify bright students and also provides them an opportunity to pay attention to and improve the slow learners. For predicting the student academic performance a data mining technique under classification is used. I have analyzed the data set containing information about students, such as full name, Roll number, scores in board examinations of classes X and XII, Rank in Eamcet examinations, branch and admission type. ID3 and C4.5 classification algorithms are applied to predict the performance of newly admitted students in their future examinations. In this paper, the performance of ID3 and C4.5 algorithms are compared in terms of parameters like accuracy, error rate and the execution time and the experimental Results shown that C4.5 was found to be best in terms of execution time.

Keywords: *ID3, Classification, Prediction.*

I. INTRODUCTION

Data mining has been attracting a significant amount of research, industry. Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information that can be used to increase revenue, cut costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified.

Classification is a data mining technique that maps data into predefined groups or classes. It is a supervised learning method which requires labelled training data to generate rules for classifying test data into predetermined groups or classes. It is a two phase process. The first phase is the learning phase, where the training data is analysed and classification rules are generated. The next phase is the classification phase, where test data is classified into classes according to the generated rules. Since classification algorithms require the classes to be defined based on data attribute values, we had created an attribute "class" for every student, which can have a value of either "Pass" or "Fail".

A Decision Tree is a classification scheme which generates a tree and a set of rules, representing the model of different classes, from a given dataset. It is a flowchart 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 the classes or class distributions. The topmost node in a tree is the root node. We can easily derive the rules corresponding to the tree by traversing each leaf of the tree starting from the node. It may be noted that many different leaves of the tree may refer to the same class labels, but each leaf corresponds to a different rule.

Over the past decade there has been a rapid growth in higher education system. A lot of new institutions have come up both from public and private sector offering variety of courses for under graduating and post graduating students. The rates of enrollments for higher education has also increased but not as much as the number of higher institutions are increasing. It is a concern for today's education system and this gap has to be identified and properly addressed to the learning community. Hence it has become important to understand the requirement of students and their academic progression. Educational Data Mining helps in a big way to answer the issues of predictions and profiling of not only students but other stake holders of education sectors.

Prediction of student academic performance using ID3 and C4.5 Decision Tree Algorithm is a software application which predicts the student's performance based on their past performance which includes their marks scored in the board
examinations of classes X and XII, Eamcet rank, admission type and branch classifies the student's performance as pass or fail in his/her first semester examinations as precisely as possible and also brings out the accuracy, error rate and execution time taken by the

algorithm for bulk evaluation.

Every year, educational institutes admit students under various courses from different locations, educational background and with varying merit scores in entrance examinations. Moreover, schools and junior colleges may be affiliated to different boards, each board having different subjects in their curricula and also different level of depths in their subjects. Analyzing the past performance of admitted students is the main purpose of the proposed system. This can very well be achieved using the concepts of data mining.

Here, we focus on the ID3 and C4.5 decision tree classification algorithm to classify the academic attributes so that the exact result could be predicted by the system based on the user provided entries, which would provide a better perspective of the probable academic performance of students in the future.

The rest of the paper is organized as follows: Related Work is detailed in Sect. 2. In Sect. 3, Proposed Methodology and Experimental results and discussions are described in Sect. 4. The conclusion is in Sect. 5.

II. RELATED WORK

Kalpesh Adhatrao, Aditya Gaykar, Amira Dhawan, RohitJha and Vipul Honrao [1] predicted the results of students currently in the first year of engineering, based on the results obtained by students currently in the second year of engineering during their first year. Attributes such as merit, gender, percentage, admission type were used to construct the decision tree. The decision tree algorithms that were used are ID3 and C4.5.

Brijesh Kumar Bharadwaj and Saurabh Pal [2] in Data Mining Educational Data to Analyze Students Performance. They use ID3 algorithm for classifying the dataset. This paper used dataset obtained from VBS Paranuchal University, Jaunpur. These analyze the performance of the students based on their Previous Semester Marks, Class Test Grade, Seminar Performance, and Assignment and end semester marks. From this paper, the students with low performance can be easily identified and high concentration can be provided in order to improve the performance level of such students.

Surjeet Kumar Yadav[3] used C4.5, ID3 and CART decision tree algorithms are applied on engineering student's data to predict their performance in the final exam in ―Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification‖. This paper generate the result of the decision tree predicted the number of students and the C4.5 algorithm has given highest level of accuracy 67.77% compared to other classification.

Crist´obal Romero [4] Educational data mining (EDM) is an emerging interdisciplinary research area that deals with the development of methods to explore data originating in an educational context. EDM uses computational approaches to analyze educational data in order to study educational questions. This paper surveys the most relevant studies carried out in this field to date.

III. PROPOSED METHODOLOGY

The proposed system is for managing the student database and maintaining the user friendly interface to predict student performance. The system was build using the programming language. No specific formats required to input the data and datasets are provided in the form of excel sheets. Here we are going to compare the two algorithms in terms of accuracy and execution time.

The entire implementation is divided into five stages. In the first stage, information about students who have been admitted to the first year was collected. This included the details submitted to the college at the time of enrolment. In the second stage, extraneous information was removed from the collected data and the relevant information was fed into a database. The third stage involved applying the ID3 and C4.5 algorithms on the training data to obtain decision trees of both the algorithms. In the next stage, the test data, i.e. information about students currently enrolled in the first year, was applied to the decision trees. The final stage consisted of developing the front end in the form of a web application.

These stages of implementation are depicted in Figure 1.

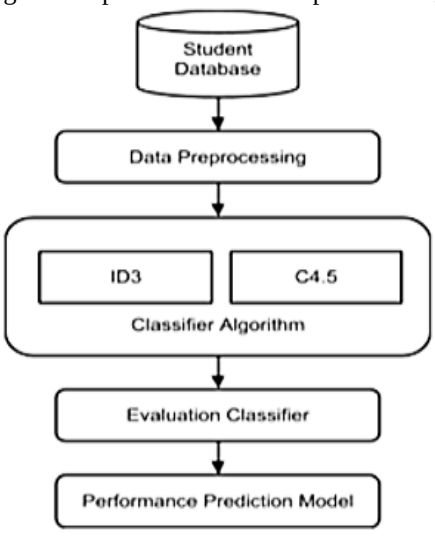

Figure 1: Proposed Flow diagram

A. STUDENT DATABASE

This is an educational data set which is collected from our college. We were provided with a training dataset consisting of information about students admitted to the first year. This data was in the form of a Microsoft Excel 2007 spreadsheet and had details of each student such as full name, Roll number, gender, percentage of marks obtained in board examinations of classes X and XII,

marks obtained in the Eamcet examination, admission type, branch. For ease of performing Data mining operations, the data was filled into a MySQL database by the admin through a front end interface.

B. DATA PREPROCESSING

Once we had details of all the students, we then segmented the training dataset further, considering various feasible splitting attributes, i.e. the attributes which would have a higher impact on the performance of a student.

The attributes that had been retained are those ranks in Eamcet entrance examination, percentage of marks scored in the board examination of class X and XII, branch, roll-number and admission type. Finally, the "class" attribute was added and it held the predicted result, which can be either "Pass" or "Fail".

Since the attributes for marks would have discrete values, to produce better results, specific classes were defined. Thus, if the score of the student was 70 or above, "Grade A" if the percentage was less than 70 and greater than or equal to 60, then it was classified as "Grade B" if the percentage was less than 60, it was classified as "Grade C". The attribute for admission type is labelled "type" and the value held by a student for it can be either "COMMON" if the student was admitted in the college through his/her Eamcet rank or "NRI" if the student was admitted through other means.

	A	B	C	D	E	F	G	
1	RollNo	Name	Branchcode	AdmnType	SeatCat	DOB	ParentName	
2	16U1A0501	ABDUL ZUBAIDHA BEGUM	CSE	COMMON	BC-E	06/04/1996	ABDUL KHANI	D NO-13-1324, RAJEEV NAGAR, ARILOVA, VISAKHAPA
3	16U1A0502	ADDEPALLI KRISHNA MANASA	CSE	COMMON	OC	09/10/1996	ADDEPALLI N SRINIVASARAO	D NO-12-5-32/TT, TOP FLOOR, SRINIVASA RESIDENCY
4	16U1A0503	ADIRAJU SWATHI	CSE	NRI		24/04/1997	ADIRAJU SIVA KUMAR	D NO:3-21, SARADANAGAR 1ST LINE, NARSIPATNAM
5	16U1A0504	ALLURI MANISHA	CSE	NRI		13/03/1997	A SUDARSHANA RAJU	201, 3RD FLOOR, , SRI NIDHI ENCLAVE, VISAKHAPATN
6	16U1A0505	AMBAKKAT PRIYANKA NAIR	CSE	NRI		10/05/1996	KEZHEKEDATH DEVADAS	32-10-49/1,SF-3, SHEELA NAGAR, VISAKHAPATNAM.
7	16U1A0506	ANALA PRIYANKA	CSE	NRI		06/07/1997	ANALA RAJU	4-131, CHINNA VEEDHI OPP RAMMANDHIR, BALAJIPE
8	16U1A0507	ANDIBOYINA ANUSHA	CSE	COMMON	BC-B	19/11/1996	ANDIBOYINA PYDI RAJU	D NO-5-6-52/32/34,RAJEEV_SARMA NAGAR, LAMBAD
9	16U1A0508	ANGURU DIVYA	CSE	NRI		14/07/1997	ANGURU BABU RAO	18-31-52, GREEN GARDENS, MARRIPALEM, VISAKHAP
10	16U1A0509	ARE NAGA SAI AKHILA	CSE	COMMON	OC	10/11/1996	ARE SRINIVASA RAO	D NO-6-65-8, SRANIKA NAGAR, CHINNA GANTYADA,
11	16U1A0510	ATCHI GOWTHAMI	CSE	COMMON	OC	22/01/1997	ATCHI SRINIVASA REDDY	D NO-53-12-29, ATCHI VARI STREET, MADDILAPALEM,
12	16U1A0511	BADIREDDI HEMALEKHA	CSE	COMMON	BC-D	30/11/1996	BADIREDDI KRISHNA	D NO-49-53-9/6, SEBORA APTS,FLAT NO-001, B S LAY
13	16U1A0512	BANDARU DHAMINI	CSE	COMMON		23/10/1996	BANDARU ESWARA RAO	D NO-39-18-63/2, MOTHER ENCLAVE, MADHAVADHA
14	16U1A0513	BANISETTI MRUDULA	CSE	NRI		04/11/1996	BANISETTI SRINIVAS	80-26-5/B A V A ROAD T2-NVR, HEIGHTS KODANDARA
15	16U1A0514	BASWA SREEVANI	CSE	NRI		06/03/1997	B KRISHNA RAO	B.KRISHNA RAO,JANAPRIYA CASTLE, H.NO.1-9-286/2
16	16U1A0515	BEESETTI HEMA SHIRDI SAI KUMARI	CSE	COMMON	BC-D	03/05/1997	B SRINIVASA RAO	D NO-38-31-150, GREN GARDENS, MARRIPALEM, VISA
17	16U1A0516	BELLANA RENUKA PUSHPA	CSE	COMMON	SC	20/03/1997	BELLANA SRINIVASARAO	D NO-40-2-79, SANTHI NAGAR, KAILASAPURAM, VISA
18	16U1A0517	BETHALA SONY	CSE	COMMON	SC	18/07/1997	BETHALA YEDUKONDALU	D NO-3-261/4,KONDALAMMA TEMPLE STREET, PALAN
19	16U1A0518	BHAGYA LAKSHMI PRASANNA REDDY	CSE	COMMON	OC	09/02/1997	V V SATYANARAYANA MURTHY REDDY	D NO-15-6-3, CHENAVARI STREET, MANDAPETA, E.G.I
20	16U1A0519	BHAIKAR MELAVI DEVI	CSE	COMMON	OC	05/11/1996	BHAIKAR RAJA MITRAJI ANAND KUMAR	D NO-7-2-18/A, RANGIREJU STREET, VIZAYANAGARAM
21	16U1A0520	BHEEMINI LAKSHMI DEVI	CSE	COMMON		05/08/1996	BHEEMINI DHARMA SATYA RAO	D NO-44-64-12B TYPE-3,Q NO-20, D L B QUARTERS, KA
22	16U1A0521	BODDANI SREE SRAVYA	CSE	NRI		10/03/1997	BODDANI SATYANARAYANA	FLAT NO:2 PAMPANAS RESIDENCY, SEETHAMMADHA
23	16U1A0522	BODDETI PREETI	CSE	COMMON		12/07/1996	BODDETI APPARAO	D NO-58-1-219/1, SF-3, SRI NILAYAM, OLD KARASA, V
24	16U1A0523	BUDDHARAJU VENKATA SINDHUSHA	CSE	NRI		09/08/1997	BUDDHARAJU VENKATAPATHI RAJU	5-29-5, , BANK COLONY, TUNI, VISAKHAPATNAM.
25	16U1A0524	CHIRA NIKHITHA	CSE	COMMON	SC	25/08/2014	CH MURALI KRISHNA	Q.NO-121-E,SECTOR-III, UKKUNAGARAM, STEEL PLAN

Figure 2: Student Database

	A	B	C	D	E	F
1	Roll-number	Admission-type	branch	Eamcet-Rank	10th Grade	12th-Grade
2	15JG1A0407	COMMON	ECE	52991	A	A
3	15JG1A0458	COMMON	ECE	88617	A	A
4	15JG1A0483	NRI	ECE	80100	A	A
5	15JG1A0501	COMMON	CSE	11526	A	A
6	15JG1A0502	COMMON	CSE	22183	A	A
7	15JG1A0503	NRI	CSE	101110	A	A
8	15JG1A0504	NRI	CSE	64130	A	A
9	15JG1A0505	COMMON	CSE	21844	A	A
10	15JG1A0506	COMMON	CSE	16087	A	A
11	15JG1A0509	COMMON	CSE	12747	A	A
12	15JG1A0511	COMMON	CSE	16195	A	A
13	15JG1A0512	COMMON	CSE	12585	A	A
14	15JG1A0514	COMMON	CSE	12744	A	A
15	15JG1A0517	NRI	CSE	84596	A	A
16	15JG1A0518	COMMON	CSE	21727	A	A
17	15JG1A0519	COMMON	CSE	125401	A	B
18	15JG1A0520	NRI	CSE	29285	A	A
19	15JG1A0522	COMMON	CSE	19666	A	A
20	15JG1A0523	NRI	CSE	31770	A	A
21	15JG1A0524	NRI	CSE	32825	A	A
22	15JG1A0525	COMMON	CSE	25628	A	A
23	15JG1A0526	NRI	CSE	61652	A	A
24	15JG1A0527	COMMON	CSE	25538	A	A
25	15JG1A0529	NRI	CSE	65911	A	A

Figure 3: student database after removing irrelevant data.

C. CLASSIFIER ALGORITHM

Algorithm 1: ID3

In decision tree learning, ID3 (Iterative Dichotomiser 3) is an algorithm invented by Ross Quinlan used to generate a decision tree from the dataset. ID3 is typically used in the machine learning and natural language processing domains. The decision tree technique involves constructing a tree to model the classification process. Once a tree is built, it is applied to each tuple in the database and results in classification for that tuple.

The ID3 algorithm is a classification algorithm based on Information Entropy, its basic idea is that all examples are mapped to different categories according to different values of the condition attribute set; its core is to determine the best classification attribute form condition attribute sets. The algorithm chooses information gain as attribute selection criteria; usually the attribute that has the highest information gain is selected as the splitting attribute of current node, in order to make information entropy that the divided subsets need smallest. According to the different values of the attribute, branches can be established, and the process above is recursively called on each branch to create other nodes and branches until all the samples in a branch belong to the same category. To select the splitting attributes, the concepts of Entropy and Information gain are used.

Entropy: Given probabilities p1, p2, …, ps , where $\sum pi = 1$, Entropy is defined as $H(p1, p2, \ldots, ps) = \sum - (pi \log pi)$. Entropy finds the amount of order in a given database state. A value of H = 0 identifies a perfectly classified set. In other words, the higher the entropy, the higher the potential to improve the classification process.

Information Gain: ID3 chooses the splitting attribute with the highest gain in information, where gain is defined as difference between how much information is needed after the split. This is calculated by determining the differences between the entropies of the original dataset and the weighted sum of the entropies from each of the subdivided datasets. The formula used for this purpose is: $G(D, S) = H(D) - \sum P(Di)H(Di)$:

```
Function ID3 (I, O, T) {
/* I is the set of input attributes
* O is the output attribute
* T is a set of training data
* Function ID3 returns a decision tree
*/
If (T is empty) {
return a single node with the value "Failure";
}
if (all records in T have the same value for O) {
return a single node with that value;
}
If (I is empty) {
return a single node with the value of the most frequent
value of O in T;
/* Note: Some elements in this node will be incorrectly
classified */
```


}
/* Now han4dle the case where we can't return a single node */
Compute the information gain for each attribute in I relative to T;
Let X is the attribute with largest gain(X,T) of the attributes in I;
Let {x_j | j=1, 2…., m} be the values of x;
Let {t_j | j=1, 2….. m} be the subsets of T;
When T is partitioned according the value of x;
return a tree with the root node labelled x and arcs labelled x_1,x_2,….., x_m, where
the arcs go to the trees ID3(I-{x},0,t-1), ID3(I-{x},0,t-2)…., ID3(I-{x},0,T-m);
}

Algorithm 2: C4.5

C4.5 is a well-known algorithm used to generate a decision trees. It is an extension of the ID3 algorithm used to overcome its disadvantages. The decision trees generated by the C4.5 algorithm can be used for classification, and for this reason, C4.5 is also referred to as a statistical classifier.

The C4.5 algorithm made a number of changes to improve ID3 algorithm. Some of these are:

- Handling training data with missing values of attributes.
- Handling differing cost attributes.
- Pruning the decision tree after its creation.
- Handling attributes with discrete and continuous values.

Let the training data be a set S = s1, s2 ... of already classified samples. Each sample Si = xl, x2... is a vector where xl, x2 ... represent attributes or features of the sample. The training data is a vector C = c1, c2..., where c1, c2... represent the class to which each sample belongs to. At each node of the tree, C4.5 chooses one attribute of the data that most effectively splits data set of samples S into subsets that can be one class or the other. It is the normalized information gain (difference in entropy) that results from choosing an attribute for splitting the data. The attribute factor with the highest normalized information gain is considered to make decision. The C4.5 algorithm then continues on the smaller sub-lists having next highest normalized information gain.

1: Tree= {}
2: if D is "pure" OR other stopping criteria met then
3: terminate F
4: end if
5: for all attribute a D do
6: Compute information theoretic criteria if we split on a
7: end for
8: a best=Best attribute according to above computed criteria
9: Tree= Create a decision node that tests a best in the root
10: D v= Induced sub-datasets from D based on a best
11: for all Dv do
12: Tree v=C4.5 (Dv)
13: Attach Tree v to the corresponding branch of tree
14: end for
15: return Tree

D. CLASSIFIER EVALUATION

Singular Evaluation

Once the decision trees were mapped as class methods, we built a web page for admin to feed values for the name, application ID and splitting attributes of a student, as can be seen in Figure 4. These values were then used to predict the result of that student as either "Pass" or "Fail" as seen in figure 5.

Singular Evaluation is beneficial when the results of a small number of students are to be predicted, one at a time. But in case of large testing datasets, it is feasible to upload a data file in a format such as that of a Microsoft Excel spreadsheet, and evaluate each student's record. For this, admin can upload a spreadsheet containing records of students with attributes in a predetermined order.

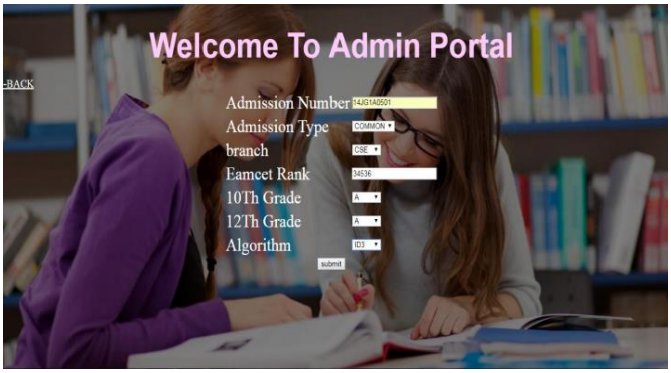

Figure 4: Singular Evaluation

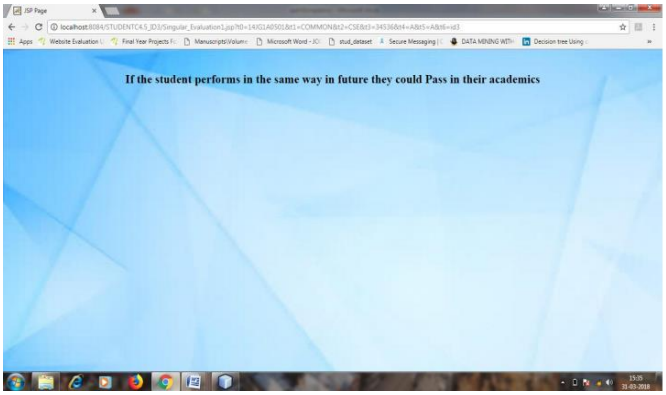

Figure 5: Singular evaluation Result

Bulk Evaluation

Under the Bulk Evaluation tab, an admin can choose an uploaded dataset to evaluate the results, along with the algorithm to be applied over it as seen in figure 6. After submitting the dataset and algorithm, the predicted result of each student is displayed in a table as the value of the attribute "class". A sample result of Bulk Evaluation can be seen in Figure 7.

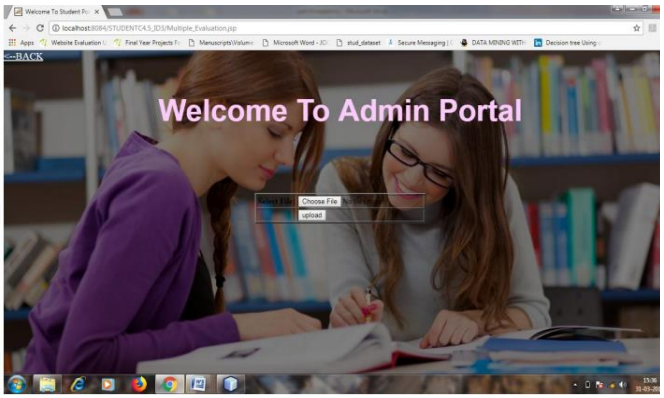

Figure 6: Bulk Evaluation

Figure 7: Bulk Evaluation Result

Verifying Accuracy of Predicted Results

The accuracy of the algorithm results can be tested under the Verify tab as seen in figure 8. An admin has to select the uploaded verification file which already has the actual results and the algorithm that has to be tested for accuracy. After submission the predicted result of evaluation is compared with actual results obtained and the accuracy is calculated. Figure 7 shows that the accuracy achieved is 85.65% for both ID3 and C4.5 algorithms. Figure 7 shows the mismatched tuples, i.e. the tuples which were predicted wrongly by the application for the current test data.

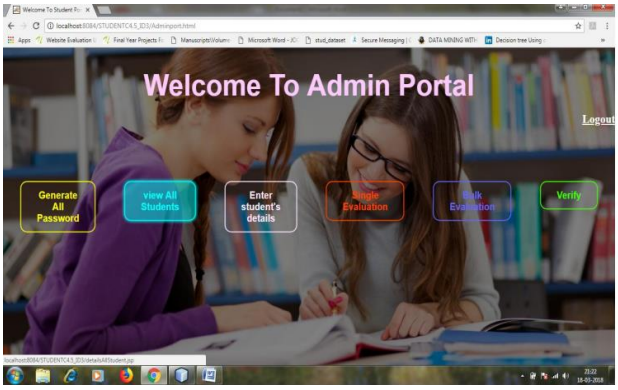

Figure 8: Admin portal

IV. EXPERIMENAL RESULTS AND DISCUSSIONS

Accuracy, error rate and execution time of ID3 and C4.5 are compared and the resultant graph is shown in figure 9.

The results of Bulk evaluation are shown in Table1 and the results of Singular evaluation are shown in Table 2.

.

ID3 and c45 algorithms accuracy are identical for the dataset which was given as input(test set), but there was a huge difference with time taken for execution by the two algorithms. ID3 takes 19937Milliseconds where as c4.5 takes 1933 Milliseconds for Bulk evaluation of 230 students.

Table 1. Results of Bulk Evaluation

Algorithm	ID3	C4.5
Total Students	230	230
Students whose results are correctly predicted	197	197
Accuracy (%)	0.8565	0.8565
Execution Time (in milliseconds)	19937 Milli seconds	1933 Milli seconds
Error rate	0.1434	0.4347

Table 2. Results of Singular Evaluation

Algorithm	ID3	C4.5
Total Students	10	10
Students whose results are correctly predicted	8	8
Accuracy (%)	0.8	0.8
Error rate	0.2	0.2

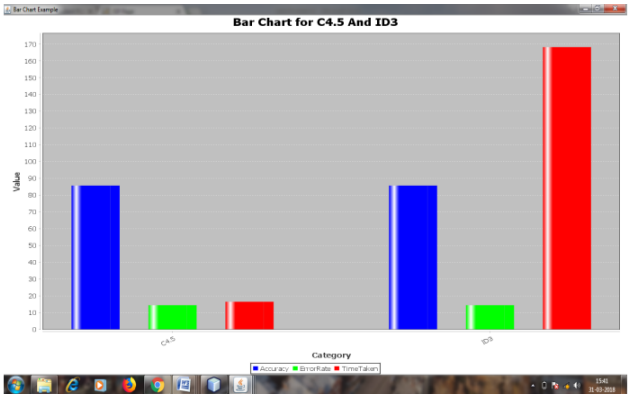

Figure 9: Comparison Graph of ID3 and C4.5

V. CONCLUSION

Early Prediction of student's performance is to enhance the quality of education in various traits. It helps in predicting at risk students during the course time itself and not at the end of the course. In this project the students past performance can be given as input and their future semester performance can be predicted as precisely as possible. So that students graduate on time without reappearing in the semester and are prepared to flourish in actual course and get back to careers on time. In the proposed system we compared ID3 and C4.5 classification algorithms, C4.5 was found to be the best in terms of execution time.

REFERENCES

[1] Kalpesh Adhatrao, Aditya Gaykar, Amiraj Dhawan, Rohit Jha and Vipul Honrao "*Predicting Students' Performance using ID3 and C4.5 classification algorithms*", International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.**3**, No.**5,** pp.**39-52**,September **2013.**

[2] Brijesh Kumar Bharadwaj and Saurabh Pal, *"Mining Educational Data to Analyze Students Performance"*, International Journal of *Advances* in Computer Science and Applications,Vol. **2(6)** , pp. **63- 69**, **2011.**

[3] Surjeet Kumar Yadav, "*Data Mining: A Prediction for Performance improvement of Engineering Students using Classification*‖", World of Computer Science and Information Technology Journal (WCSIT) ISSN: 2221- 0741 Vol**. 2**, No. **2**, **51-56, 2012.**

[4] Crist´obal Romero," *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,** November **2010**.

[5] Mining, H., Wenying, N. and Xu, L., (2009) "*An improved decision tree classification algorithm based on ID3 and the application in score analysis"*, Chinese -Control and Decision Conference (CCDC), pp **1876- 1879.**

[6] Xiaoliang, Z., Jian, W., Hongcan Y., and Shangzhuo, W., (2009) "*Research and Application of the improved Algorithm C4.5 on Decision Tree*", International Conference on Test and Measurement (ICTM), Vol. **2,** pp **184-187.**

[7] M. Mayilvaganan,D .Kalpanadevi," *Comparison of Classification Techniques for predicting the Performance of Students Academic Environment*," in International Conference on Communication and Network Technologies (ICCNT), **2014.**

[8] R.S.J.D Baker and K.Yacef, "*The State of Educational Data Mining in 2009: A Review and Future Visions*" , Journal of Educational Data Mining, **1**, Vol **1**, No **1**, **2009**.

Authors Profile

Mrs.*K.suneetha* pursed Bachelor of Technology from Jawaharlal Nehru Technological University Hyderabad of Telangana, India in 2005 and Master of Technology from Jawaharlal Nehru Technological University Hyderabad in year 2009. She is currently working as Assistant Professor in Department of Computer Science and Engineering, Gayatri vidya Parishad College of engineering for women, Visakhapatnam since 2009. She had 10 years of teaching experience. Her areas of interests are Data Mining, Machine Learning and Deep Learning.

