

# A STUDY OF STUDENT'S ACADEMIC PERFORMANCE USING DECISION TREE

Sanjeev Gour<sup>1</sup>

<sup>1</sup> Asst. Professor & Head, Dept.of Computer Science, L.B.S. College, Harda, M.P., India.

## ABSTRACT

*The main objective of this paper is an attempt to use data mining methodologies to study and try to analyze and predict the student's performance in the courses in any educational Institution, also using data mining processes, particularly classification, to help in enhancing the educational institutions by evaluating and classifying student data to study the main attributes that may affect the student performance in courses. It present an approach to classifying students in order to predict their final grade based on certain features extract from educational data bases which are collected from an Institutes . It included student's academic as well as personal characteristics attributes which influence the performance of student and also help to predict the result whether he/she can get admitted or not in particular course . For this purpose, ESTARD DATA MINER V3.0 software is used for mining student related academic data. In this study we used supervised real dataset from an academic Institution. This experimental study show how the academic and personal attributes and both together of any student help to make better decision about any educational policies like which students are most likely are eligible to take admission in any course, which one will get what grades in any course on the basis of their Academic and Personal Characteristics collected as a Database.*

**Keywords:-** Educational Data Mining, Classification, Decision Tree/Rules.

## 1 INTRODUCTION

Data mining is a technology used in different disciplines to search for significant relationships among variables in large data sets. Data mining is mainly used in commercial applications. In this study, we concentrated on the application of data mining in an education environment. Data Mining is a powerful tool for academic intervention. Through data mining, a university or college, for example, predict with 85 % accuracy which student will get which grades in any courses. Colleges and many other educational institutions can use classification, for example Decision Rules /Decision Tree to analyze the student data and try to predict the performance of student .Data mining enables educational institutions to use their current major attributes of student to uncover and understand hidden pattern in vast databases. These patterns are then built into data mining models and used to predict student behavior, for example, give an institution the necessary information to take action before student drops out, to estimate accurately how many students will take a particular course or how many students will get passed or failed in any particular course.

## 2 DATA MNING IN EDUCATIONAL SYSTEM:

Data mining is an important data analysis methodology that has been successfully employed in many domains, with numerous applications in educational matters [1],[3]. The Educational Data Mining was defined as “The process of converting raw data from educational systems to useful information that can be used to inform design decisions and answer research questions”[6].

### 3 RELATED WORKS DONE:

Many Research has done about Mining student academic data using Decision trees in which only Academic Attribute data of student was analyzed and evaluated ,article "Mining student Data Using Decision Trees"[2] and "Data Mining Model for Better Higher Educational System".[5] ,In this paper we are trying to analyzed and classified academic attributes as well as Personal characteristics attributes of students and relate both the attributes with common ID field and also predicted the performance of student with 93 %Accuracy using WHAT-IF Analyzer.

### 4 PROPOSED EXPERIMENTAL STUDY:

#### Data mining with Decision Tree Algorithms:

Decision Tree is tree-structure that represents series of rules that leads to sets of decisions. Theses decision generates rules for the classification of data sets. They can manage both continues and categorical variables. Decision trees can perform Classification as well as Regression. The predicted value from a decision tree is not simply a numerical value but can be predicted category such as PASS/FAIL, FIRST/LAST, And TRUE/FALSE etc.

#### Methodology

The process of Data Mining in this application consists of three steps: Data Preparation, building classification model using decision tree and decision rules, prediction using WHAT-IF Analyzer.

### 5 DATABASE PREPARING

#### Data Set

The data gathered from the students of the LBS College of Professional Studies, Harda was used in this study. It included records of 210 students. The academic information of students are collected from Office of the Institute and the personal characteristics information of students are collected using a questionnaire that was passed among undergraduate students of the Science and Arts faculty.

#### Database

The database management system used in the study was the Microsoft EXCEL 2003. This system was used for two reasons; the software used in analysis was compatible and efficient to use with the database management system, and the data to be analyzed was maintained in the database prior to the study.

#### Application Software

The programming environment for the application was Estard Data Miner 3.0. The Estard Data Miner software application was suitable for the development of the application, and compatible with the Excel 2003 in which the data was maintained. The C4.5 and ID3 algorithms used in the application was defined in the Estard software .

Although ESTARD Data Miner is able to work with "uncleaned" data, duplicate records, empty or incomplete records and mixed data types in fields will most likely make the data mining results incomplete, or result in incorrect rules and decision trees. This is why it is recommended to use "clean" data. Besides, in case if you want to use several tables from one database, these tables will need and ID field - a field containing only unique values and identifying connection between records in two or more tables.

The student data that was maintained in Two tables named Academic –Info(table-1) and Personal –Info(Table-2) was joined using the SNo.(Serial Number) field as the key ID-field.

**Academic-Info Table:****Table-1:** The Academic Details Of Student with possible Class-Values.

SNO	Attribute	Attribute-Details	Class –Values
1	Stu-Gender	Student Gender	Male/Female
2	Stu-Age	Student Age	17-22
3	Pre-Marks	Previous Year Marks	0-100
4	HS-Sub	High Secondary Subject	Science/Arts
5	HS-Grade	High School Grade	A,B,C,D
6	Test-Result	Subject Test Result	PASS/FAIL
7	Attendance	Attendance in Percentage	0-100
8	Extra-Curri.	Extra curriculum(NCC,NSS,Sports)	YES/NO
9	Basic-comp Know.	Knowledge of Basic Computer	YES/NO
10	Final-Grade	Final Grade of Student	A,B,C,D

**Personal-Info Table.****Table-2:** The Personal Details Of Student with possible Class-Values.

SNO	Attribute	Attribute –Details	Class-Values
1	Self-Confidence.	Self-Confidence measure in three label	High/Medium/Low
2	Financial -Status	Financial -Status measure in three label	Rich/Average/Poor
3	Discipline-Status	Discipline-Status measure in three label	Best/Good/Worst
4	Reference-Status	Reference-Status	YES/NO
5	Admission -Status	Admission -Status	YES/NO

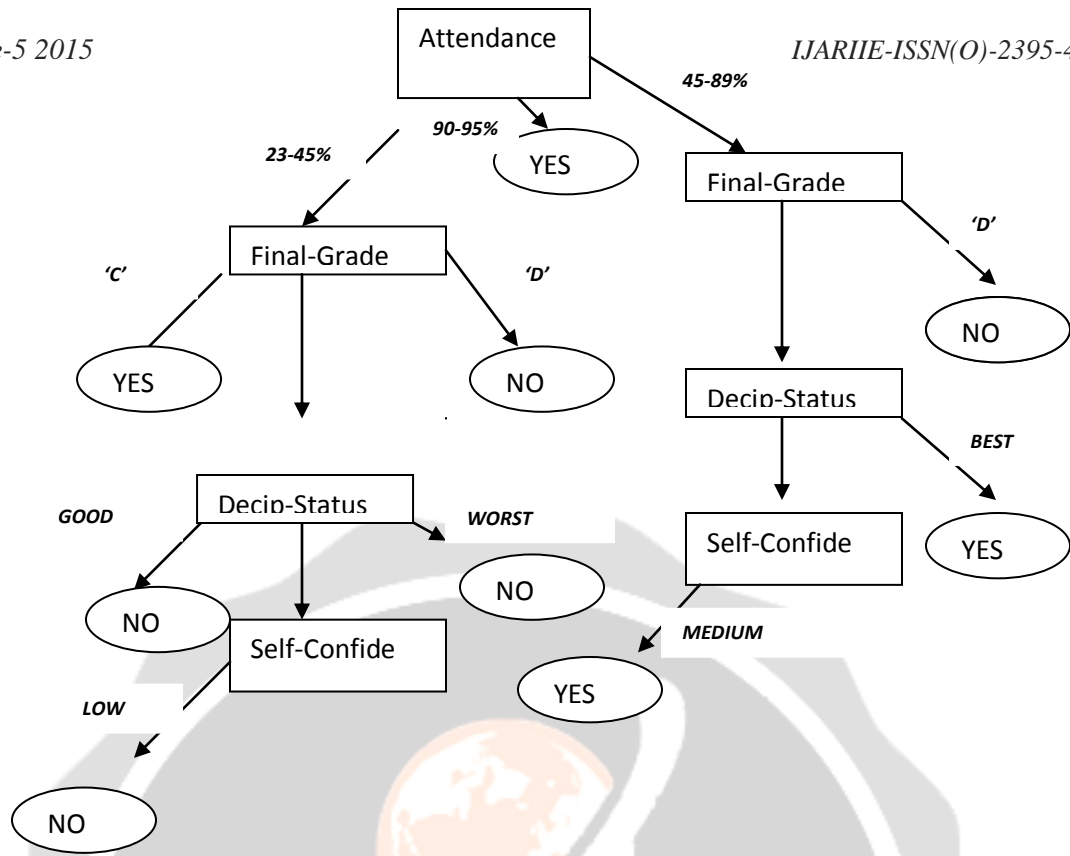


Figure 1: Decision tree-I

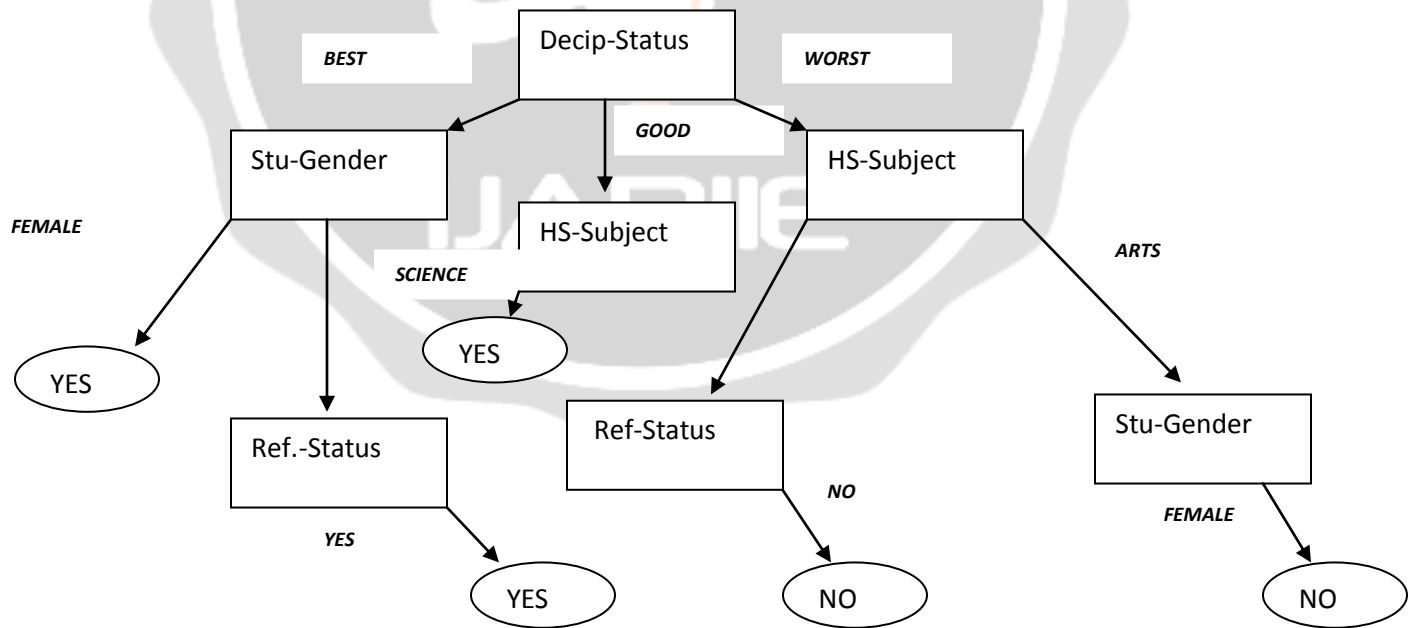


Figure 2: Decision Tree -II

**6 BUILDING DECISION TREE / DECISION RULES.**

For Experimental study here we make some SETS by combining of both academic and Personal Attributes randomly and make a Decision Tree also generate the rules by converting it. (Figure 1 and 2) . Here are some Rules in the form of Tables with rule probability blow:

**Examined Class- ADMISSION-STATUS (YES/NO).**

(For SET1-**Academic-info:** 1)Basic comp. know. 2)Final –grade 3)Test-results.

**Personal –Info :** 1)self-confi 2)reference status 3)Financial staus 4) decip-status.);

(For SET 2- **Academic-info:** 1)Extra-curri. 2)St-Gender3)Attedance 4)Pre-Marks ;

**Personal –Info :** 1)reference status 2) decip-status.);

**TABLE- For Examined Class-Admission Status (Categorical Values-YES/NO)**

**Table 3:** Some IF-THEN Rules Generated using Decision-Rules by ESTARD-DM.

Rule No.	IF-THEN Rules	No. of Cases met	No.of Element in the rule	Rule Probability
S1 #9	IF (Personal-Info.Decip-status='Worst') AND (Academic-Info.Basic-Comp Know='NO' THEN Class-Value Result = " NO";	20	2	81.5789 %
S1 #2	IF (Academic-Info.Basic-Comp Know='NO') AND (Academic-Info.FINAL-GRADE-'D') THEN Class-value= "NO" ;	33	2	99.9999 %
S1 #12	IF (Personal-info.Self-Conf='Medium/High' )AND (Personal-info.Reference-Status='YES') AND (Personal-info.Financial-Status='Average/Rich')AND (Personal-info.Decip-Status='Good') AND (Academic-info.Basic-Comp Know='YES') AND (Academic-info.Final-Grade='B') AND (Academic-Info.Test-Result='PASS') THEN Class Value Result= "YES	18	7	91.6667 %
S2 #10	IF (Personal-info.Reference-Status='YES') AND (Personal-info.Decip-Status='Good') AND (Academic-info.Extra-Curri.='YES') AND (Academic-info.Attedance='45-89') AND (Academic-Info.PRE-MARKS='40-90')AND(Academic-Info.St-Gender='FEMALE') THEN Class Value Result= "YES "	20	6	85.5932%
S2 #4	IF (Personal-info.Decip-Status='Worst') AND (Academic-info.Attedance='23-45') AND (Academic-Info.PRE-MARKS='23-40') THEN Class Value Result= "NO ";	29	3	94.5484%

**Examined Class-FINAL-GRADE (A,B,C,D).**

For this all attributes/fields (Query Parameters) are taken recommended by Data miner.

**TABLE: For Examined Class-Admission Status (Categorical Values: A, B, C, D)**

**Table 4:** Some IF-THEN Rules Generated using Decision-Rules by ESTARD-DM.

Rule No.	IF-THEN Rules	No. of Cases met	No.of Element in the rule	Rule Probability
# 3	IF (Academic-Info.HS-Grade='A') AND (Academic-info.Attendance='88-92') THEN Class-Value Result = 'A'	31	2	99.89 %
# 7	IF (Academic-Info.Pre-Marks=58-67) AND (Academic-info.Attendance='52-71') THEN Class-Value Result = 'C'	11	2	90.000 %
#10	IF (Academic-Info.HS-Grade='D') AND (Academic-info.Attendance='23-45') THEN Class-Value Result = 'D'	18	2	93.5484%

### 6.1 Some of the discovered interesting rules by Decision Tree are:

\* If student's Discipline-Status = 'BEST' and Final-Grade = 'D', Then their Admission Status = 'NO' i.e. He/she will not get admitted. While if a student is female then she will get admission in the course.

\* If a student's Discipline-Status = 'BEST' or 'GOOD', Reference-status='YES', High School subject='Science' and student is female then Admission -Status='YES'. While if a student with Reference-Status='NO', High School Subject = 'Arts' and even student is female then their Admission -Status = 'NO'. (DT-II).

### 7 PREDICTIONS USING WHAT-IF ANALYZER.

In this section we have set new (changed) attribute values of academic and personal data of student from supervised datasets to the values for which we have to see the statistical analysis result. Here using WHAT-IF Analyzer we can predict the Result Value of examined class by setting new attributes values even some result are up to 93 % accurate with known values(Supervised Data).The Report generated by DM for following random selected attributes is here.

**ESTARD Data Miner v 3.0 Report.**

**What-If**

Database: E:\i data\my data\Student-Dataset.xls  
 Table: ' Academic-info\$', ' Personal-info\$'  
 Number of Records in a Table: 210, 210

**Class/Examined Field:**

Selected Class Field: FINAL-Grade  
 Number of Classes: 4

**Table 5 :** Resultant table from What –If analyzer

SN#	Table	Field	Value
0	' Academic-info\$'	Extra Curri	YES
2	' Academic-info\$'	Basic-comp Know#	YES
4	' Academic-info\$'	HS-Grade	B
5	' Academic-info\$'	HS-Sub	SCI
7	' Academic-info\$'	St-Age	18-20
8	' Academic-info\$'	St-Gender	F
9	' Academic-info\$'	Test-Result	PASS
10	' Personal-info\$'	Decip#-Status	Best
12	' Personal-info\$'	Reference-status	NO

**Table 6:** Statistics Analysis Result:

SN.	Final-Grade(Examined-Class)	Probability
1	A	90.0869 %
2	B	9.6302 %
3	C	0.0289 %
4	D	0.0000 %



## 8 CONCLUSIONS:

This research is an initial attempt to use data mining methods to analyze and evaluate student academic and personal data and to enhance the quality of the educational institute. The result obtained from the analysis of student's academic performance has showed that the decision tree data mining software (ESTARD) was able to study the data set used and able to mine the academic and personal information of student from the Database. From these main attributes we classified the student dataset and make the Decision tree (also converted into Decision-Rules) to analyze the educational policy and by changing the attribute values from the supervised dataset we have also get the Statistical Analysis for class value so we predicted the final Grade of student in any course. one of the most excited future work is to collect large dataset from the university student databases and apply the Classification and Regression , Clustering ,Association and Correlation ,and other Data mining techniques to uncover many interesting facts about the student which also can improve the policies ,strategies of any higher educational system.

## 9 REFERENCES :

- [1] Bresfelean ,Vasile Paul ."Data Mining Application in Higher Education and Academic Intelligence Management, MPRA ,paper no.21235,27 June 2008.
- [2] Qasem A.,Al-Radaideh, Emad M. Al-Shawakfa, and Mustafa I. Al-Najjar, "Mining Student Data Using Decision Trees" , ACIT'2006.
- [3] Adesean Barnabas Adeyemo ,Gbemisola Kuye ."Mining Students' Academic performance USING Decision Tree Alogrithms ", Journal Of Information Technology Impact,Vol 6.No-3,pp 161-170,2006.
- [4] Jing Luan ."Data Mining Applications in Higher Education" Executive report ,SPSS.
- [5] K.shyamala and S.P.Rajagopalan .,"Data Mining Model foe a Better Higher Educational System" .Information Technology Journal 5(3):560-564,2006.
- [6] Jiawei Han Micheline Kamber .Book "Data Mining : Concept and Techniques .II-edition,2006.