Data Analytics of Outcome Based Education

Submitted By

Joshi Neeti Dattuprasad

15MCEI09

DEPARTMENT OF COMPUTER ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2017

Data Analytics of Outcome Based Education

Major Project

Submitted in partial fulfillment of the requirements

for the degree of

Master of Technology in Computer science and Engineering

(Information and Network Security)

Submitted By

Joshi Neeti Dattuprasad

(15MCEI09)

Guided By

Dr. Sanjay Garg

Dr. Priyanka Sharma

DEPARTMENT OF COMPUTER ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2017

Certificate

This is to certify that the major project entitled **"Data Analytics of Outcome Based Educaion "** submitted by **Joshi Neeti Dttuprasad (Roll No: 15MCEI09)**, towards the partial fulfillment of the requirements for the award of degree of Master of Technology in Computer Science and Engineering of Nirma University, Ahmedabad, is the record of work carried out by him under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination. The results embodied in this major project part-I, to the best of my knowledge, haven't been submitted to any other university or institution for award of any degree or diploma.

Dr. Sanjay Garg
Guide & HOD,
CE Department,
Institute of Technology,
Nirma University, Ahmedabad.

Dr. Priyanka Sharma
Guide & Professor,
PG Coordinator M.Tech(CSE),
CE Department,
Institute of Technology,
Nirma University, Ahmedabad.

Pro. Sharada Valivety
Associate Professor,
PG Coordinator M.Tech(INS),
CE Department,
Institute of Technology,
Nirma University, Ahmedabad.

Dr. Alka Mahajan
Director,
Institute of Technology,
Nirma University, Ahmedabad

Statement of Originality

I, **Joshi Neeti Dattuprasad**, Roll. No. **15MCEI09**, give undertaking that the Major Project entitled "**Data Analytics Of Outcome Based Education**" submitted by me, towards the partial fulfillment of the requirements for the degree of Master of Technology in **Computer science and Engineering with specialization in Information and Network Security** of Institute of Technology, Nirma University, Ahmedabad, contains no material that has been awarded for any degree or diploma in any university or school in any territory to the best of my knowledge. It is the original work carried out by me and I give assurance that no attempt of plagiarism has been made.It contains no material that is previously published or written, except where reference has been made. I understand that in the event of any similarity found subsequently with any published work or any dissertation work elsewhere; it will result in severe disciplinary action.

Signature of Student

Date:

Place:

Endorsed by

Dr.Sanjay Garg,

Dr.Priyanka Sharma

Acknowledgements

It gives me immense pleasure in expressing thanks and profound gratitude to **Dr. Priyanka Sharma**, Professor, Computer Engineering Department, Institute of Technology, Nirma University, Ahmedabad for her valuable guidance and continual encouragement throughout this work. The appreciation and continual support she has imparted a great motivation to me in reaching a higher goal.

It gives me an immense pleasure to thank **Dr. Sanjay Garg**, Head of Computer Science and Engineering Department, Institute of Technology, Nirma University, Ahmedabad for his kind support and providing basic infrastructure and healthy research environment.

A special thank you is expressed wholeheartedly to honourable **Dr. Alka Mahajan**, Director, Institute of Technology, Nirma University, Ahmedabad for the unmentionable motivation she has extended throughout course of this work.

I would also thank the Institution, all faculty members of Computer Engineering Department, Nirma University, Ahmedabad for their special attention and suggestions towards the project work.

I am highly indebted to **Mrs. Madhushree B.** (Ph.D scholar) and **Mr. Vaibhav Jain** (JRF) for providing necessary information regarding the project and also for their support in completing the project.

I would like to express my special gratitude to my father **Dr. Dattuprasad Joshi** (Ex HOD of physics department at M S University, Vadodara) for their constant motivation and inspiration , my mother **Mrs. Rajeshree Joshi** whose never-ending encouragement and all my family members (**Er. Jaimin Joshi** and **Dr. Vidhi Joshi**) which help me in completion of this project.

- Joshi Neeti Dattuprasad

15MCEI09

Abstract

Outcome based education (OBE) refers to the analysis of student performance on the basis of program outcomes, course Learning outcomes, assessment matrix and rubrics for each course. Data analysis can assist in terms of predicting and analyzing the performance based on machine learning algorithms. However, these analysis take into consideration, only the academic performance of the students. An adaptive approach that incorporates both academic information and personal characteristics of the student can be used for a more precise prediction. By using the different machine learning algorithms, the accurate prediction and more significant analysis could be procure. This study presents a comparative study of machine learning algorithms for data analysis of outcome based education to obtain accurate algorithm, predictive analysis of grade marks and behavioural analysis of students to find out most impactful behaviour on study. The experimental results contain the predictive analysis, data analysis and comparative analysis of student performance for an intricate analysis of the OBE based implementation.

Abbreviations

OBE	Outcome Based Education.
AEHS	Adaptive Education Hypermedia System.
LMS	Learning Management System.
LO	Learning Outcomes
PO	Program Outcomes

–

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

Outcome Based Education aims that student should be able to learn and understand the specific course work as per the program outcomes(PO) and Learning outcomes(LO). It is the process of Teaching to Learning. Basically the performance of the student is mapped with the different measures of PO and LO by classifying in different rubrics. And at the end of the specific course the student should up the mark as per the defined outcomes. It is largely accepted in different countries. For Example

- Australia (Since 1990)
- Europe (Since December 2012)
- Hong Kong (Since 2005)
- Malaysia (Since 2008)
- South Africa (Since 2005)
- United States (Since 2001)
- Pakistan (Since 2016)
- India (Since 2014)

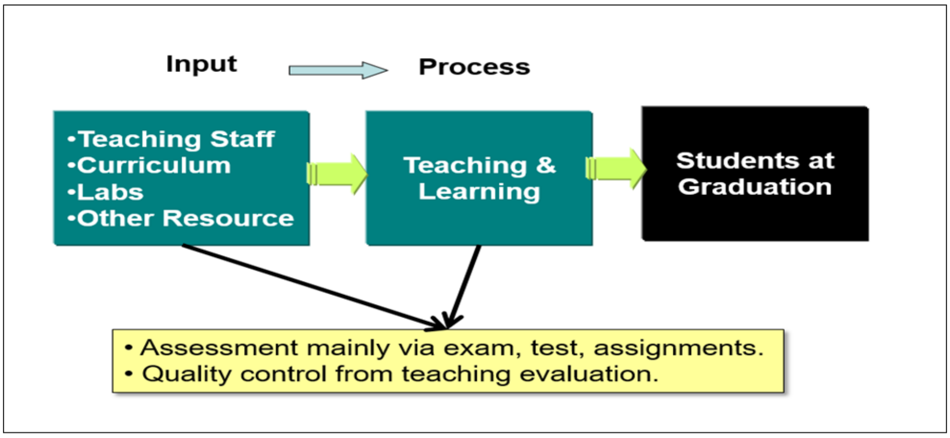

Figure 1.1: Traditional Higher Education System

1.2 Issues of Traditional Higher Education System

1.2.1 Process of Traditional Higher Education System

Traditional Higher education system belongs to classroom teaching system in which teacher teaches the students in the class rooms and labs as per the designed syllabus. Students prepare the subjects from the text book and memorize it. Then Students give the examinations and pass the exams by achieving minimum passing grades and they became a graduates.

The student assessment is perform via exams, test, assignment. There are no interactive assessment which measures the student intelligence towards the subject. Education quality can only be measured and modified by assessment system. In this education system assessment of student is very easy because the course and exam structure is very iterative and repetitive. The exam contains the questions from the question bank which repetitively ask in the exams. Teachers ask direct questions in the exams so students can easily pass the exams. So Traditional Higher Education system is very direct and assessment process is very easy to achievable for the students.

1.2.2 Drawbacks of Higher Education System

- **Teacher-Centric Approach**

  The teachers ask the same repetitive question in the examination so that students can easily pass. So the main focus is to pass maximum number of students in the examinations.

- **Objective is to gain the good marks in the examinations**

  Main motive of exam is to pass the exams with good marks rather than pass the exams with good concepts. For achieving good grades students only prepare the important topics of the subjects. They may remain unaware about rest of the topics.

- **Teaching Method: Lecture Notes and Text Books**

  Students learn the subjects from thee text books and Lecture notes. So students may have the lack of broad knowledge of specific subject. This kind of teaching methods doesn't give proper technical and practical knowledge of subject.

- **No importance for Social Development**

  There is no social development like leadership, group participation, interpersonal relationship etc. Due to lack of social development student gets failure to become a good employee. They may suffocate in corporate environment due to less socialism.

- **Memorisation is the main to gain good marks** Each and every students memorise the topic whether they have understood the concept or not. They prefer only book and notes for exam preparation. Memorisation doesn't gives proper conceptual knowledge of specific topic or subject.

- **Main aim is to pass the exams with good marks**

  There is no importance to acquire proper concepts of subject but all the students study the subject to gain good marks instead of to achieve good knowledge. Due to main focus on marks most of the students fail to gain proper technical knowledge towards subjects.

1.3 Why Outcome Based Education?

1.3.1 Process of Outcome Based Education

Outcome Based Education belongs to e-learning, classes, quizzes, group discussion. Learning of student is more important than teaching. OBE contains so many modules and students should achieve the ability to pass that each and every module.

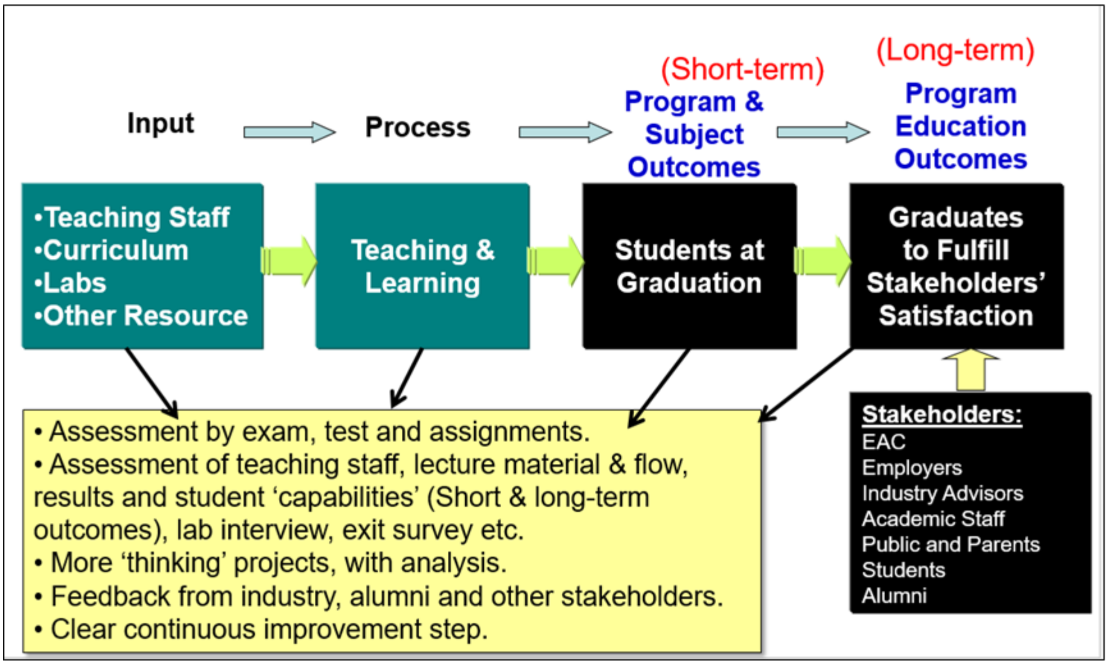

Figure 1.2: Outcome Based Education System

Teacher teaches in the classroom and students learn from the class and also learn from the videos, they solve the quizzes, learn and discuss the topic by group discussion. Teachers give the extra project work for each subject to the student to increase the practical knowledge and concepts.

OBE is not direct learning so the assessment process of OBE is not easily achievable. Teachers are also assessed by superiors like students. Teacher assess the students by test, assignment, project work, presentation work, lab works. Then they map students performance with program outcomes and learning outcome defined by institute. Stakeholders also give the feedback of the students, whether the students are suitable and compatible for corporate world.

Each and every step student is getting assessed and mapped with PO, LO and institutional vision and mission. So if student clears all the levels then only he/she can pass the course. Assessment process of OBE is tedious but it gives qualitative result.

1.3.2 Advantage of Outcome Based Education

- **Student Centric Approach**

  The main goal is continues progress of students. Teachers ask innovative questions in the examinations. There are no direct questions in the exams. So they can check the concepts of the students.

- **Objective is to gain proper conception and retention of the qualitative knowledge**

  Students should have conceptual knowledge and they can represent their knowledge into practical life. Their knowledge should be problem solving. There is no importance to acquire good marks.

- **Teaching method: Hands-on activity, Group Discussion**

  All the student should participate in learning process of subject. Learning process (Teaching Process) contains lectures, online videos, quizzes, group discussion etc. There are no direct questions in the examinations. So memorization would be failure method.

- **Special importance to social development**

  Team work, Interpersonal Relationship, Self Awareness, Leadership are the key parameters of the social development. Due to social development students can easily adopt the corporate environment. So the students can easily satisfy the stakeholders needs.

- **Creativity is a main key role**

  There is no significance to marks. All the student should have creative and innovative idea to solve the technical problems. Teacher assess the students on the basis of their creative ideas and how they have problem solving attitude.

- **Expectation is to gain certain level of knowledge**

  OBE has two parameters i.e. learning outcomes and program outcomes. And both the parameters have rubrics. All the students map with rubrics and classified into the quality indicators. so all the students should be up to the mark in their course. So All the students should reach to the level which is defined by rubrics.

1.4 Summary

This chapter illustrate the introduction of outcome based education, flaw of traditional education system and how outcome based education system overcomes that flaws. In next the chapter we will discuss the brief of outcome based education.

Chapter 2

Outcome Based Education

2.1 Introduction

Outcomes are the parameters which measures the performance of the student. Outcomes are goals which are defined by institute and each students must satisfy those goals. Students assess through and map through all the rubrics with performance of the students and then students would be classified that how many percentage they have achieved their goals, and whether they are successful to getting degree or not. It is an hierarchical structure.

The main aim is to design outcomes to make the students optimistic person, independent learner, concerned human being and problem solving.

There are two parameters in the outcome based education upon which whole the education system established.

- **Program Outcomes [PO]**

  societal outcomes which are defined by institute and they are long term goals to be measured

- **Learning Outcomes [LO]**

  course outcome which is defined by department and they are short term goals to be measured

2.1.1 Program Outcomes [PO]

Program outcome contains long term goals which are measured at end of the course. Knowledge, skills and behaviour are measured as outcomes at the end of the program. It basically measures the performance of the student other than study. Program outcomes may measured by the institute.

The examples of program outcomes as following:

- Student should have potential to apply the knowledge using the principle of engineering and science.
- student should have ability to communicate effectively.
- Student should aware of social, global, cultural and environmental responsibilities
- Student should have problem solving approach
- Student should have good socialism like interpersonal relationship, leadership, team work etc.
- Student should have enthusiasm and always ready to improve their performance

2.1.2 Learning Outcomes [LO]

Learning outcomes contains short term goals which are measured after the examinations. The outcomes which are measured to check the performance of the student of specific course. Learning outcomes are the course outcomes which are measured by the department. Teachers define the student level by mapping the student performance with rubrics, and then teacher decide that how much goal is achieved by student.
The examples of course outcomes is as following: **Ethical Hacking:** (Nirma University: M.Tech subject)[1]

- Understand the core concepts related to malware, hardware and software vulnerabilities and their causes
- Understand ethics behind hacking and vulnerability disclosure
- Appreciate the Cyber Laws and impact of hacking
- Exploit the vulnerabilities related to computer system and networks using state of the art tools and technologies

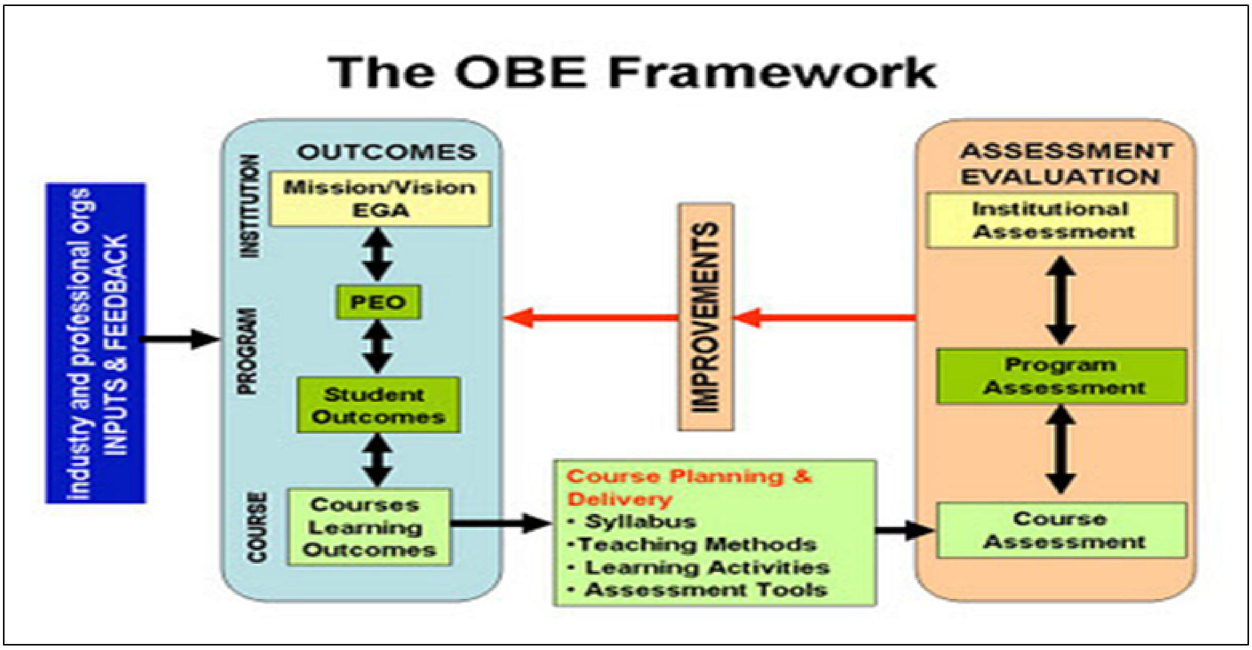

Figure 2.1: OBE Framework

2.2 Framework of OBE

OBE follows the structure in which first student performance would be mapped with the course outcome which are defined by department at the end of the exams and students are classified as per the quality indicators. Course outcomes would be measured at each level of exams.

After the Learning outcomes Program outcomes would be measured at the end of the specific course. Program outcomes contains the rubrics other than the subject.PO measures how the student is problem solving, communication skill, team work etc.

At last the institutional vision and mission mapped with student outcome. whether the student is achieved the goals defined by institute or not. So OBE framework is iterative and hierarchical.

2.2.1 Model Hierarchy

As shown in figure there are four layers of outcomes. First and basic layer is Learning outcome which is measured at each level. Then Program outcome and Program Educational Outcome are the intermediate layers. And Last layer is Vision and Mission of Institute which is measured at end of the program.

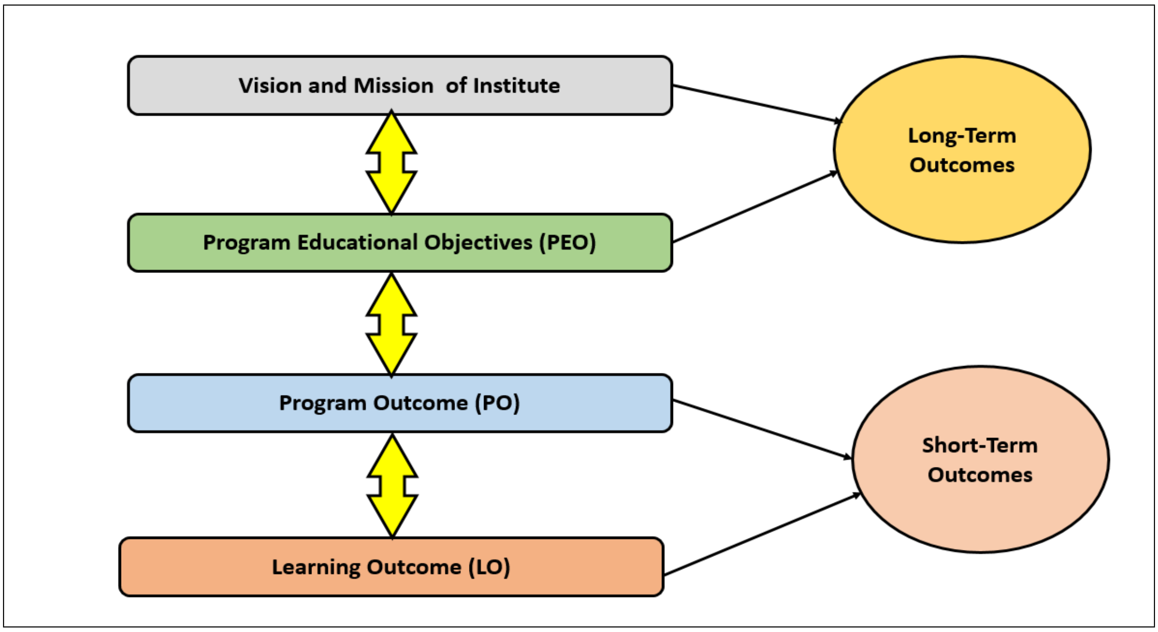

Figure 2.2: Model Hierarchy of OBE

So here Learning Outcomes [LO] and Program Outcomes [PO] are short term outcomes which are measured by department level at short period of time and Program Educational Objectives [PEO] and Vision and Mission of Institute are long term outcomes which are measured by institute at long term period of time.

2.3 Summary

In this chapter we have discussed about the outcomes and it's different parameters. This chapter represents the flow of outcome based education system and it's model hierarchy. In next chapter we will discussed comparative study and on the basis of that we will conclude some challenges of the study.

Chapter 3

Literature Review

3.1 Introduction

Outcome based education used two types of education framework like learning management system and adaptive educational hypermedia system. So, in this chapter the comparative study will be carried out on the basis of following education system.

- **Learning Management System [LMS]:**

  LMS is open source software which provides the learning services to Administrator, Teacher and student.

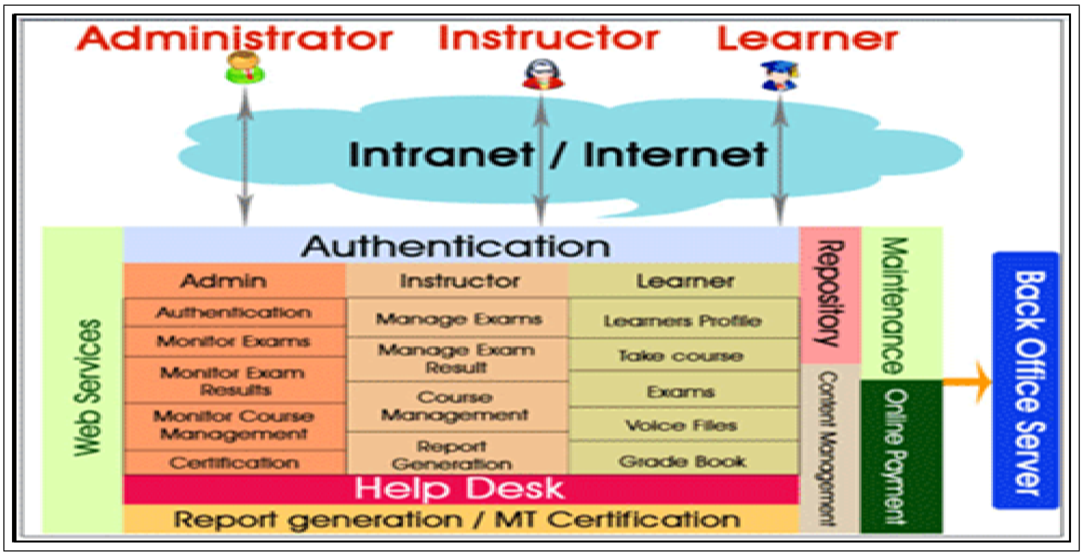

Figure 3.1: LMS Architecture

As shown in figure 3.1 LMS provides different level of accessibility to teachers, students and administrator via nternet.

- **Adaptive Educational Hypermedia System [AEHS]**

  AEHS add on the personal characteristics features in LMS system. All the students may have different hobbies, goals, interests which may lead to different effects on study. So prediction of student result through only academic performance not always gives accurate result. So AEHS gives accurate result by using extra features of students. But it is some times very complex and not also cost effective system.

3.1.1 Comparative Study

Midhun, et al [2] have identified the academically at risk student and predictive analysis of 9th class of CBSE students. They have found out the result using k-means algorithm and multiple regression algorithm. They have defined six clusters to classify the students who contain similar marks. From that one cluster contains the students who have low marks so that cluster would be the result of at-risk student. To find out the Gpa of 4th test they used the training dataset which contains student id, name and marks of four tests. Another dataset is test dataset which contains the student id, name, and the marks of 3 test. So using the multiple regression they have predict the 4rth test marks. The challenge is the accuracy of multiple regression for prediction. By using neural network algorithm, the accuracy can be improved.
Dimokas, et al [3] proposed the statistical, data warehousing and mining methodology to found out the students who fail to complete their study in stipulated time and they compared the male and female who complete their study more faster using Pearson correlation and cross tabulation method. They conclude that 58 percent students complete their study after 4 years (because of strong negative correlation between duration and students) and female completed their study before male. The another algorithm to find the students who would fail to complete their study in stipulated times is using k-means clustering algorithm which would give the clusters similar performance of the students so from that we can defined detainer students and the students who can be detain. Camilo, et al [4] have found out the students who have low academic performance and the subjects in which students frequently fail using decision tree and Naive base classifier algorithm. The result is maths and social science are the subjects in which student frequently fail and students who are younger (23-25 age) have low academic performance. The Apriori algorithm can also be used to find the frequent subjects in which students frequently

failed. Apriori is less complex than decision tree and nave based classifier to find frequent pattern mining.Nguyen, et al [5] suggested the analysis of four algoritms when the data set is inconsistant. They compared four algorithms K-means, SOM, OCFSCM and NPFSCM to find out which algorithm is more consistent for incomplete dataset. They proved that SOM and K-Means is more consistent if the dataset is incomplete more than 50 percent.

Rover, et al [6] proposed survey paper which contains the assessment by ABET at Iowa State University. They have defined different objectives and quality indicators to classify the students. They compare the assessment of the student and teacher and they conclude that teacher have more knowledge than student the result of the survey is hierarchical assessment gives more accurate result. But they didnt classify the objectives into LO and PO. So there is no structural analysis of students. Romero, et al [7] presented the survey paper of in journal of IEEE. They explained the data mining techniques, distant learning, e-learning, text mining, web mining and traditional education system. They also describe the different data mining tools. They did comparative analysis of which data mining task is appropriate to which educational system.

Guleria, etal [8] proposed the feedback system from the student. Here feedback (for different parameters like teaching , intrastate, department, institute) of the student is calculated by using the standard deviation to check the significance of the feedback.Almeida, et al[9](2010) suggested This paper contains the survey of AEHS system and they have conclude that Category Theory is very useful in AEHS system. But the limitation is that, for hierarchical assessment the automated tool is needed to implement category theory. To overcome this challenge computer aided engineering tool is needed for adaptive navigation. Karampiperis et al [10] presented a description the concept competence description ontology and learners competence records using Competence Description Ontology. They conclude that Less success rate in complex competence level and highest success rate in simple competence level.Ioannis, et al [8] compared the Learning management system and Adaptive educational hypermedia system features. They conclude that LMS uses only academic information while AEHS uses academic and personal information too. So AEHS gives more accurate result than LMS. But the limitation is to use the AEHS is, very difficult to implement and it is not coast effective. So to overcome the challenges of LMS and AEHS is to use adaptive LMS system.

3.2 Summary

This chapter demonstrate the two different architecture of outcome based education system like Learning management system and Adaptive educational hypermedia system. On the basis of this study we have done comparative analysis and defined challenges. On the based on this comparative analysis, in next chapter we will be doing experimental analysis on the challenges, which have specified in the comparative study.

Chapter 4

Demo Results

4.1 Introduction

In this chapter, we will discuss some demo experiments. This demo experiments contain some demo dataset samples, by using this sample datasets we will talk about the challenges, which we have found in comparative study.
Following are the objectives which will be discussed in this chapter.

- Predictive Analysis of Student Performance
- Finding Out At-Risk Students
- Data Analysis of students using AEHS(Adaptive Educational Hypermedia System) rather than LMS (Learning Management System).
- Personalizes Comparative Analysis of the student performance
- Mapping of student performance with Program outcomes and Learning Outcomes (i.e. Finding out student performance during whole course).
- Finding out the Accuracy of the result by comparative analysis of two or more than two algorithms.

These objectives are experimented on rapid miner tool. The demo results are achieved by using different machine learning algorithms.

4.2 Find the Frequent Subjects Which Are Cause of Low Academic Performance

Goal: Analysis of the subject results. Due to which two frequent subjects (from three subject) students face failure.

4.2.1 Student Dataset Information

- Attributes: 4 [Name, Maths, Science, English]
- Records: 25
- Missing value: None
- Tool: Rapid Miner
- Algorithm: Apriori

D22 fail

	A	B	C	D	E
1	Name	Maths	Science	English	
2	ANURAG K SHAH	pass	pass	pass	
3	RITVIK O GUNTURU	pass	pass	pass	
4	DEV A GUPTA	fail	pass	fail	
5	AMEYA A KAMAT	pass	pass	pass	
6	MRUTHYUNJAY R KHUMBHMETHU	fail	fail	pass	
7	ANURAG S ARIBANDI	fail	fail	pass	
8	SOUMYA A NIMMU	fail	fail	pass	
9	YASHARYAN S GAIKWAD	pass	pass	fail	
10	MAYANK G SHARMA	pass	pass	pass	
11	MANSI D SINGH	pass	fail	pass	
12	RAMYA S BABU	pass	fail	pass	
13	ADITYA SUDINI	pass	pass	pass	
14	LABDHI V SHAH	fail	fail	pass	
15	SAMPREETI S KALE	fail	fail	pass	
16	DIVYANSH SAKSENA	fail	fail	pass	
17	SHIVANI V ASAWA	fail	fail	pass	
18	RIA R NAIR	fail	pass	pass	
19	NAEEM A SIDDIQUE	pass	pass	pass	
20	AISHWARYA D PATIL	pass	pass	pass	
21	C S KALYANI	fail	fail	pass	
22	HARSHAVARDHAN REDDY	pass	fail	fail	
23	MANASA KOTAMRAJU	fail	fail	pass	
24	NINAD KULKARNI	pass	pass	pass	
25	VISHNU SASIKUMAR	fail	fail	pass	
26	POOJITA REDDY	pass	pass	pass	
27					

Figure 4.1: Student Dataset

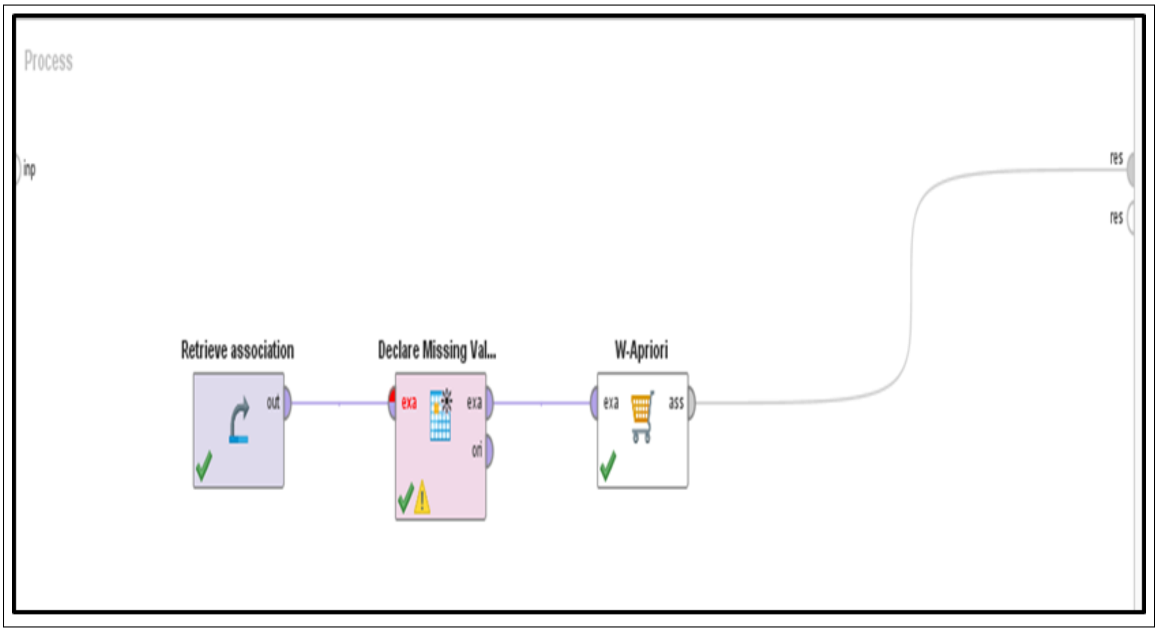

Figure 4.2: Process of Apriori

4.2.2 Process of Apriori Algorithm:

The student dataset is taken as input.If the dataset contains any missing value then it will declare. After this preprocessing phase the Apriori algorithm will apply. Output of the process will display in the result screen.

4.2.3 Result of Apriori Algorithm:

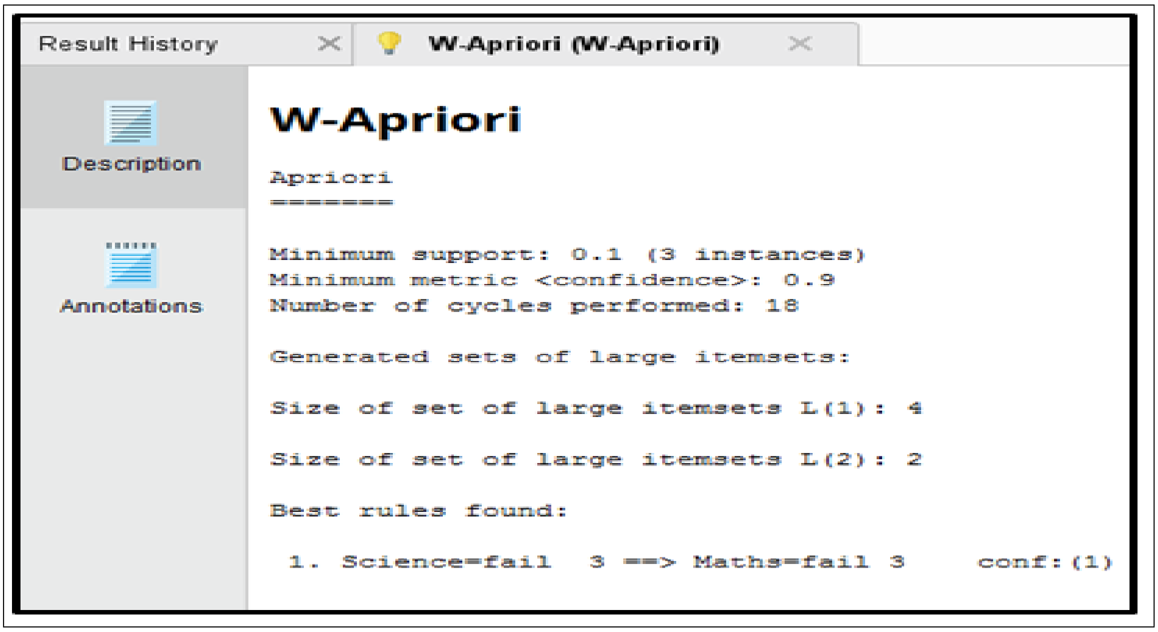

Figure 4.3: Result of Apriori

As shown in figure the result of Apriori algorithm is: **Maths and science**. There are 3 pattern in which students are failed in both maths and science.

4.3 Discriminate the Marks of At Risk Students

Goal: Analysis of at risk student by discriminate in the same cluster.

4.3.1 Student Dataset Information

id	maths	science	english	total out of 300
1001	10	12	10	32
1002	50	55	60	165
1003	20	10	10	40
1004	75	85	80	240
1005	90	95	92	277
1006	60	62	63	185
1007	74	25	85	184
1008	82	85	87	254
1009	90	88	91	269
1010	72	70	85	227
1011	12	100	65	177
1012	13	95	32	140
1013	78	20	91	189
1014	80	14	52	146
1015	100	74	41	215
1016	95	65	68	228
1017	92	35	75	202
1018	42	40	84	166
1019	35	42	29	106
1020	32	55	88	175
1021	78	57	85	220
1022	72	70	64	206
1023	85	87	75	247
1024	71	90	32	193
1025	65	82	74	221

Figure 4.4: Student Dataset

- Attributes: 5 [Id, Maths, Science, English, Total out of 300]
- Records: 25
- Missing value: None
- Tool: Rapid Miner
- Algorithm: K-Means

4.3.2 Process of K-means Algorithm

Process of K-means algorithm is similar as process of apriori. In this process the k-means algorithm would be applied to dataset. As shown in figure 4.5 the missing value function will define the missing value in the datset when the process executes.

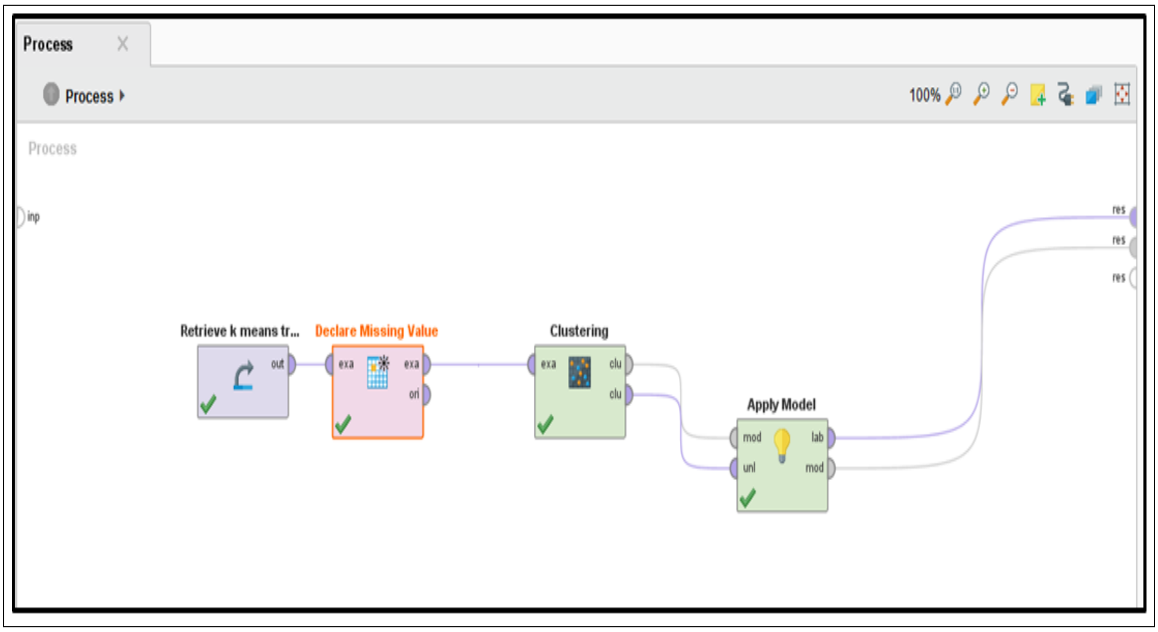

Figure 4.5: Process of K-Means

4.3.3 Result of K-Means Algorithm

ExampleSet (25 examples, 3 special attributes, 3 regular attributes)

Row No.	id	total out of 3...	cluster	maths	science	english
1	1001	32	cluster_0	10	12	10
2	1002	165	cluster_2	50	55	60
3	1003	40	cluster_0	20	10	10
4	1004	240	cluster_3	75	85	80
5	1005	277	cluster_3	90	95	92
6	1006	185	cluster_2	60	62	63
7	1007	184	cluster_4	74	25	85
8	1008	254	cluster_3	82	85	87
9	1009	269	cluster_3	90	88	91
10	1010	227	cluster_3	72	70	85
11	1011	177	cluster_2	12	100	65
12	1012	140	cluster_2	13	95	32
13	1013	189	cluster_4	78	20	91
14	1014	146	cluster_4	80	14	52
15	1015	215	cluster_3	100	74	41
16	1016	228	cluster_3	95	65	68
17	1017	202	cluster_4	92	35	75
18	1018	166	cluster_1	42	40	84
19	1019	106	cluster_0	35	42	29
20	1020	175	cluster_1	32	55	88
21	1021	220	cluster_3	78	57	85
22	1022	206	cluster_3	72	70	64
23	1023	247	cluster_3	85	87	75
24	1024	193	cluster_2	71	90	32
25	1025	221	cluster_3	65	82	74

Figure 4.6: Result of K-Means

Result would be:

- Cluster 0- 3 [id-1001,1003,1019] are At-Risk Students
- Cluster 1- 2 [id-1018,1020]
- Cluster 2- 5 [id-1002,1006,1011,1012,1024]
- Cluster 3- 11 [id-1004,1005,1008,1009,1010,1015,1016,1021,1022,1023,1025]
- Cluster 4- 4 [id-1007,1013,1014,1017]

Attribute	cluster0	cluster1	cluster2	cluster3	cluster4
Maths	21.67	37	41.20	82.18	81
Science	21.33	47.50	80.40	78	23.50
English	16.33	86	50.40	76.55	75.75

Table 4.1: Centroid of clusters

Here we have chosen 5 clusters. Each clusters would have mean values, which are shown in the table. As per the result screen short we can conclude that cluster 0 contains the at risk students.

At risk students are:

- ID-1001, 32 marks
- ID-1003, 40 marks
- ID-1019, 106 marks

4.4 Prediction of the Result and Comparative Analysis of Algorithms

Goal: Predict the SPI result of 4th semester from 1st,2nd and 3rd semester SPI and determine which algorithm is more accurate, Polynomial Regression or Neural Network ?

Here we have train data set which contains SPI of 4 semester. This data set will used to train the model. Test data set contains the SPI of 3 semester. From that the train data the test data wold be predict. 4rth semester SPI is actual value thatwould be predicted in test dataset.

Test Dataset

ID	1st	2nd	3rd
1	7.5	7.9	8.1
2	8.9	7.5	8.1
3	8.5	8.2	7.9
4	7.5	8.2	7.9
5	6.9	8.1	7.5
6	8.2	8.5	7.3
7	8.2	7.3	8.8
8	8.3	7.9	8.1
9	8.2	7.9	8.8
10	7.6	8.1	7.2

Train Dataset

A	B	C	D	E
ID	1st	2nd	3rd	4rth
1	7.5	7.9	8.1	7.9
2	8.9	7.5	8.1	8.2
3	8.5	8.2	7.9	7.9
4	7.5	8.2	7.9	8.4
5	6.9	8.1	7.5	6.8
6	8.2	8.5	7.3	7.4
7	8.2	7.3	8.8	8.7
8	8.3	7.9	8.1	8.2
9	8.2	7.9	8.8	7.8
10	7.6	8.1	7.2	7.9

Figure 4.7: Train and Test data set od student

4.4.1 Polynomial Regression

Process of polynomial regression

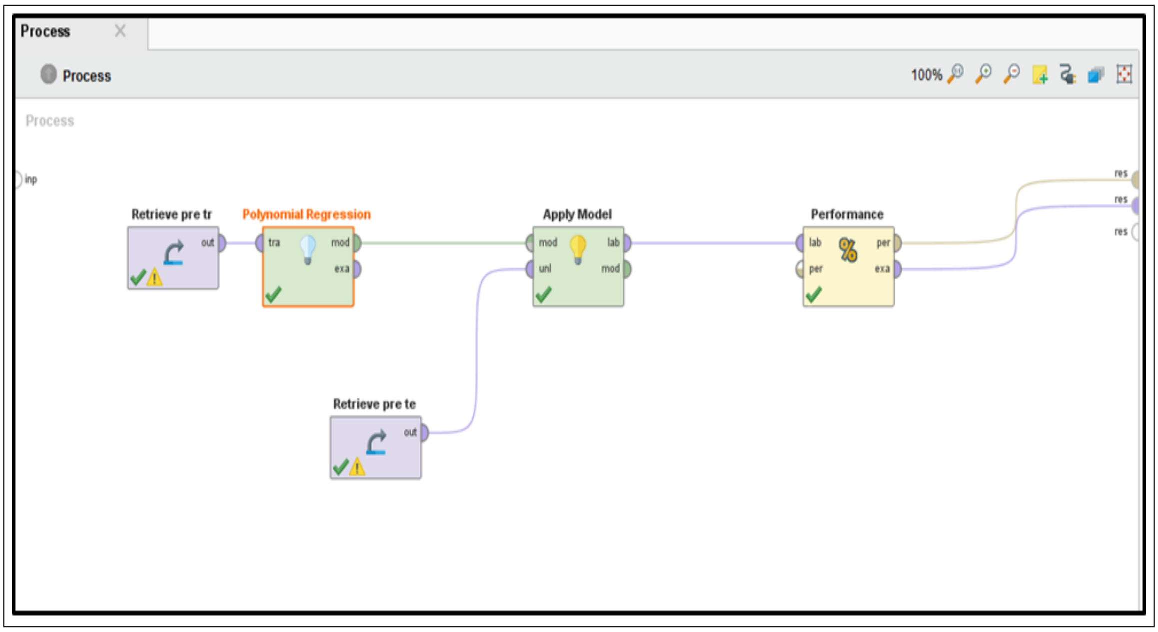

Figure 4.8: Process: Polynomial Regression

Result of Polynomial Regression

ExampleSet (10 examples, 3 special attributes, 2 regular attributes)

Row No.	id	3rd	prediction(3...	1st	2nd
1	1	8.10	6.28	7.50	7.90
2	2	8.10	8.25	8.90	7.50
3	3	7.90	10.45	8.50	8.20
4	4	7.90	7.70	7.50	8.20
5	5	7.50	5.58	6.90	8.10
6	6	7.30	11.04	8.20	8.50
7	7	8.80	5.38	8.20	7.30
8	8	8.10	8.48	8.30	7.90
9	9	8.80	8.21	8.20	7.90
10	10	7.20	7.50	7.60	8.10

Figure 4.9: Result: Prediction of Polynomial Regression

4rth sem(Real Value)	Predicted Value	Accuracy(+/-0.5)
7.9	6.284	0
8.2	8.245	1
7.9	10.447	0
8.4	7.699	0
6.8	5.578	0
7.4	11.038	0
8.7	5.378	0
8.2	8.483	1
7.8	8.208	1
7.9	7.502	1

Table 4.2: Accuracy using Polynomial Regression

Root mean square value is $2.00 + / - 0.00$. As shown in table the accuracy is +/-0.5 SPI. So if the difference beween predicated and real vaule is more than 0.5 then the accuracy would be 0 otherwise it would be 1.

So **Accuracy of polynomial algorithm is 40 percent**

4.4.2 Neural Network:

Process of Neural Network:

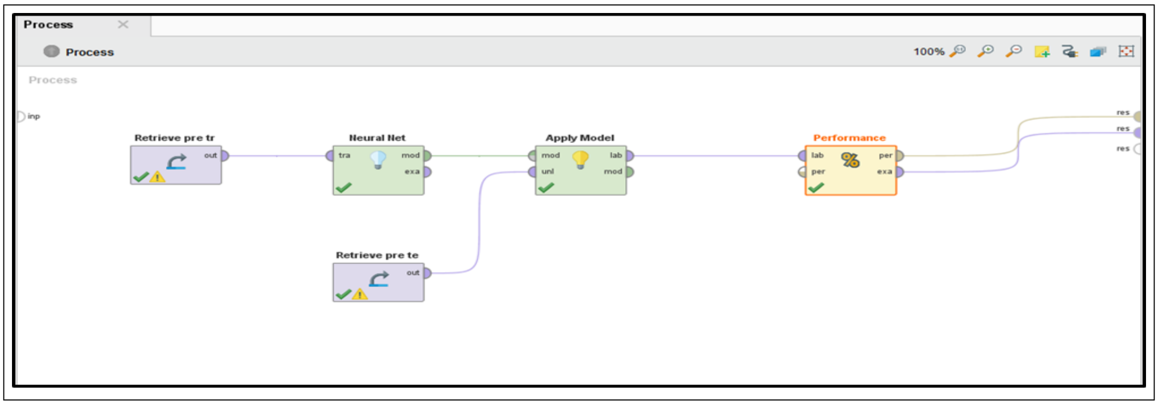

Figure 4.10: Process: Neural Network

Result of Neural Network:

ExampleSet (10 examples, 3 special attributes, 2 regular attributes)

Row No.	id	3rd	prediction(3...	1st	2nd
1	1	8.10	7.89	7.50	7.90
2	2	8.10	8.21	8.90	7.50
3	3	7.90	7.89	8.50	8.20
4	4	7.90	7.30	7.50	8.20
5	5	7.50	7.50	6.90	8.10
6	6	7.30	7.29	8.20	8.50
7	7	8.80	8.74	8.20	7.30
8	8	8.10	8.16	8.30	7.90
9	9	8.80	8.15	8.20	7.90
10	10	7.20	7.43	7.60	8.10

Figure 4.11: Result: Prediction of Neural Network

Root mean square value is $0.39+/-0.00$. As shown in table the accuracy is +/-0.5 SPI. So if the difference beween predicated and real vaule is more than 0.5 then the accuracy would be 0 otherwise it would be 1.

4rth sem(Real Value)	Predicted Value	Accuracy(+/-0.5)
7.9	7.890	1
8.2	8.212	1
7.9	7.890	1
8.4	7.699	0
6.8	7.504	0
7.4	7.288	1
8.7	8.743	1
8.2	8.159	1
7.8	8.146	1
7.9	7.431	1

Table 4.3: Accuracy using Neural Network

So **Accuracy of polynomial algorithm is 80 percent**

4.5 Comparison Between LMS and Adaptive LMS

Goal: Comparison of the student result. [Comparison between knowledge level of each subjects and knowledge level with personal characteristics of each subjects]

4.5.1 Dataset of LMS and Adaptive LMS

❑ **LMS**:

A	B	C	D
Name	subject	level	result
neeti	maths	advance	pass
neeti	science	average	fail
neeti	english	poor	fail
ruchi	maths	average	fail
ruchi	science	advance	pass
ruchi	english	poor	fail
vidhi	maths	poor	fail
vidhi	science	average	fail
vidhi	english	advance	pass

❑ **Adaptive LMS**:

name	subject	level	habits	rubrics	result
neeti	maths	advance	gadgets	high	fail
neeti	maths	advance	emotional	medium	pass
neeti	science	average	gadgets	low	pass
neeti	science	average	emotional	high	fail
neeti	english	poor	gadgets	medium	fail
neeti	english	poor	emotional	high	fail
ruchi	maths	average	gadgets	medium	pass
ruchi	maths	average	emotional	low	pass
ruchi	science	advance	gadgets	high	pass
ruchi	science	advance	emotional	medium	pass
ruchi	english	poor	gadgets	high	fail
ruchi	english	poor	emotional	medium	pass
vidhi	maths	poor	gadgets	low	pass
vidhi	maths	poor	emotional	high	fail
vidhi	science	average	gadgets	medium	fail
vidhi	science	average	emotional	high	fail
vidhi	english	advance	gadgets	medium	pass
vidhi	english	advance	emotional	low	pass

Figure 4.12: Dataset: LMS and Adaptive LMS

Here Adaptive Dataset has two more attributes like habits and rubrics. For example as per the LMS datase if Neeti knows advance level of maths then she will pass the exams bu as par the adaptive LMS dataset if Neeti knows advance level of maths and her use of gadget is maximum then she will fail in exam.

4.5.2 Process of LMS and Adaptive LMS

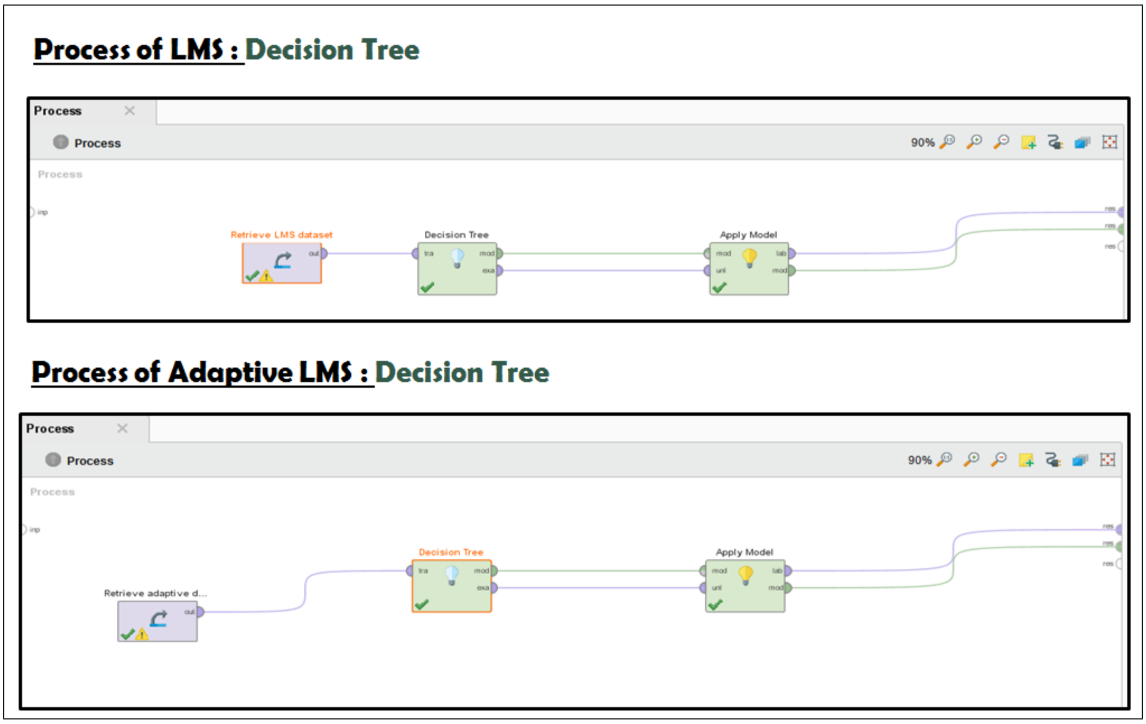

Figure 4.13: Process: LMS and Adaptive LMS

4.5.3 Result Of LMS and Adaptive LMS

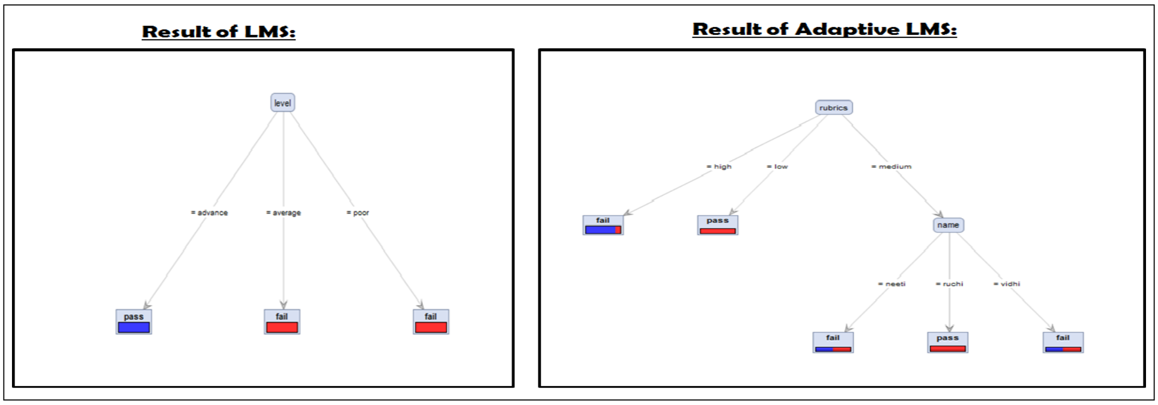

Figure 4.14: Result: LMS and Adaptive LMS

Tree of LMS:

Level = Advance : Pass pass=3, Fail=0

Level = Average : Fail pass=0, Fail=3

Level = poor : Fail pass=0, Fail=3

Tree of Adaptive LMS:

Rubrics = High : Fail Fail=6, Pass=1

Rubrics = Low : Pass Fail=0, Pass=4

Rubrics = Medium

Name= Neeti : Fail Fail=1, Pass=1

Name= Ruchi : Pass Fail=0, Pass=3

Name= Vidhi : Fail Fail=1, Pass=1

4.6 Analysis of Quality of Education

Goal: Analysis of the feedback of students to the Teaching skills of the Teachers.

4.6.1 Dataset of Student Feedback

		Feedback of Teaching skills Parameter (out of 10)			
ID	Knowledge	cooperative	Delivery	Responsiveness	Total(40)
1	9	8	7	9	33
2	7	8	8	9	32
3	7	7	8	8	30
4	9	7	8	9	33
5	9	8	7	7	31
6	8	8	9	9	34
7	7	8	7	8	30
8	8	9	8	5	30
9	7	8	9	9	33
10	8	9	9	8	34

Figure 4.15: Feedback rating

Here there are four teaching parameters, knowledge, Cooperative, delivery, responsiveness. Students have given the feedback out of 10. So last column contains the total sum of the feedback.

4.6.2 Result of Student Feedback

Mean=32

$$\textbf{Standard Deviation}=\sqrt{\frac{\varepsilon[x-\bar{X}]^2}{n-1}}$$

$$=\sqrt{\frac{24}{9}}$$

$$=\sqrt{2.6}$$

$$=1.6$$

Figure 4.16: Result: Student Feedback

Result-As the Standard Deviation value is 1.6 (very small) which shows the Significant feedback From the students.

4.7 Summary

In this chapter we experimented the objectives like predictive analysis of student performance, finding out at-risk students, data analysis of students using AEHS(Adaptive educational hypermedia system) rather than LMS (Learning Management System), personalizes comparative analysis of the student performance, mapping of student performance with program outcomes and learning outcomes, finding out the accuracy of the result by comparative analysis of two or more than two algorithms. In next chapter we will discussed about the proposed work.

Chapter 5

Proposed Work

5.1 Introduction

This chapter, the intended work would be discussed. The proposed work would be briefly describe in this section with the help of architecture. The scope of the study would be determine after the brief study of framework.

5.2 Problem Statement

Predictive and comparative analysis of the students performance based on academic and personal behavioural data using different machine learning algorithms and determine most accurate algorithm.

5.3 Proposed Architecture

The Figure 5.1 illustrates the proposed architecture. The student dataset would be consist of academic information and personal information (characteristics of students). The different machine learning algorithm would be applied for predictive and comparative analysis of students performance. The predictive analysis would be resulted as a prediction of SPI and behavioral analysis which will be (the prediction of most significant personal characteristics which impacts on the marks). The comparative analysis would carry the best accurate machine learning algorithm among all.

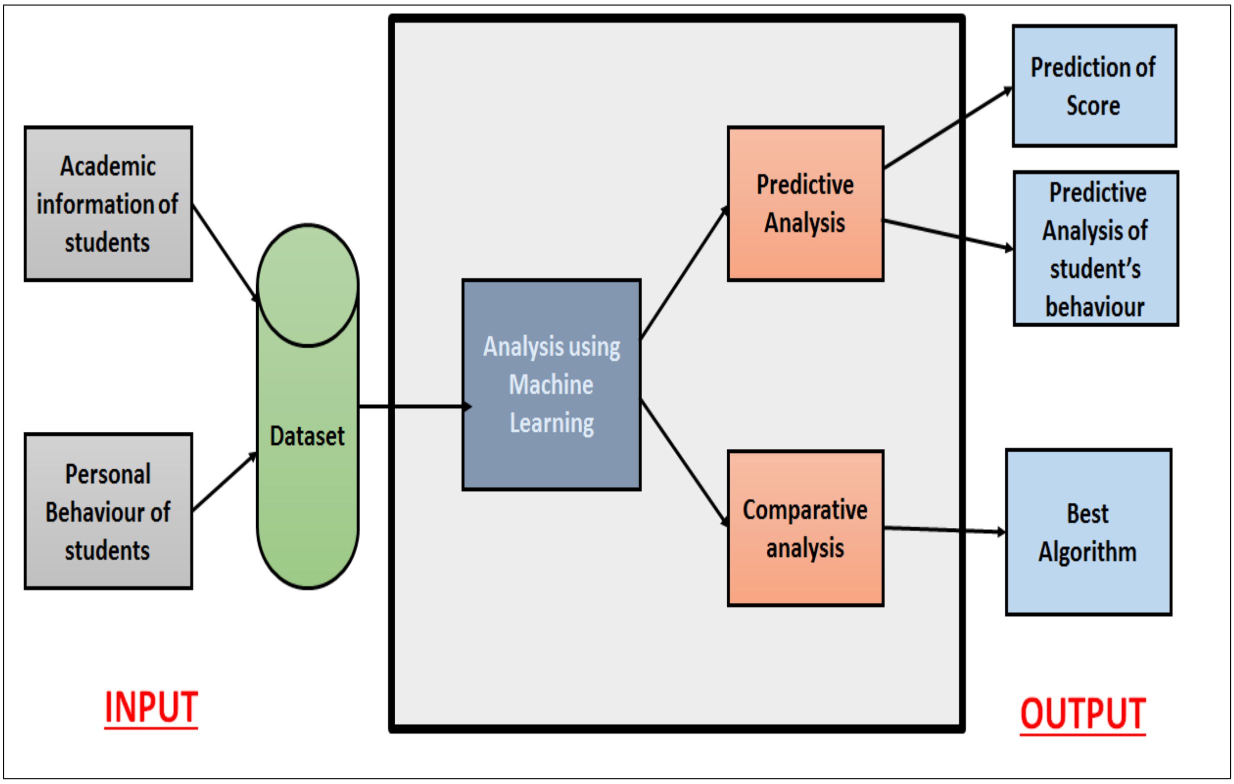

Figure 5.1: Proposed Architecture

5.4 Scope of the Project

- In the beginning of the implementation, the best accurate machine learning algorithm will be carried out on the basis of accuracy.
- After choosing the best accurate machine learning algorithm, the prediction of the score of students will be implemented.
- The most affecting behavioural characteristics of the students (which would be having more impact on the marks) will be determined.

5.5 Summary

This chapter has proposed the framework architecture. This architecture contains the comparative and predictive analysis of the student performance. In next chapter, the implementation will be carried out as per the framework.

Chapter 6

Implementation

This section contain the main student performance analysis by using different machine learning algorithms on the the specific dataset. The dataset information is given below. This student performance analysis wold be carried out the predictive and comparative analysis. The data analysis is process is split into 4 part like data collection and preparation, data selection and transformation, implementation model and result analysis. The result analysis would be followed through predictive and comparative analysis of the student performance.

6.1 Problem Statement

Predictive and comparative analysis of the student performance in maths subject using academic and personal data and finding out the most efficient algorithm.

6.2 Data Collection and Preparation

- **Sources:** Paulo Cortez, University of Minho, Portugal.(Donated on 27 November 2014)
- **Data Collection:** Kalboard 360 (Learning Management System)
- **Attributes:** 33 (Academic, Demographic, Personal)
- **Instances:** 396
- **Missing value:** None
- **Used Algorithm:**

1. Decision Tree
2. Nave Base Classifier
3. Artificial Neural Network

- **Attributes Information:**[11][12]

  (a) **Demographic Data**

   i. school - student's school (binary: 'GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira)
   ii. sex - student's sex (binary: 'F' - female or 'M' - male)
   iii. age - student's age (numeric: from 15 to 22)
   iv. address - student's home address type (binary: 'U' - urban or 'R' - rural)
   v. famsize - family size (binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3)
   vi. Pstatus - parent's cohabitation status (binary: 'T' - living together or 'A' - apart)
   vii. Medu - mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)
   viii. Fedu - father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)
   ix. Mjob - mother's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'athome' or 'other')
   x. Fjob - father's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'athome' or 'other')
   xi. reason - reason to choose this school (nominal: close to 'home', school 'reputation', 'course' preference or 'other')
   xii. guardian - student's guardian (nominal: 'mother', 'father' or 'other')
   xiii. traveltime - home to school travel time (numeric: 1 - ¡15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - ¿1 hour)

xiv. studytime - weekly study time (numeric: 1 - ¡2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - ¿10 hours)

xv. failures - number of past class failures (numeric: n if 1¡=n¡3, else 4)

xvi. schoolsup - extra educational support (binary: yes or no)

xvii. famsup - family educational support (binary: yes or no)

xviii. paid - extra paid classes within the course subject (Maths) (binary: yes or no)

(b) **Personal Data:**

i. activities - extra-curricular activities (binary: yes or no)

ii. nursery - attended nursery school (binary: yes or no)

iii. higher - wants to take higher education (binary: yes or no) internet - Internet access at home (binary: yes or no)

iv. romantic - with a romantic relationship (binary: yes or no)

v. famrel - quality of family relationships (numeric: from 1 - very bad to 5 - excellent)

vi. freetime - free time after school (numeric: from 1 - very low to 5 - very high)

vii. goout - going out with friends (numeric: from 1 - very low to 5 - very high)

viii. Dalc - workday alcohol consumption (numeric: from 1 - very low to 5 - very high)

ix. Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)

x. health - current health status (numeric: from 1 - very bad to 5 - very good)

xi. absences - number of school absences (numeric: from 0 to 93)

(c) **Academic Information:**

i. G1 - first period grade (numeric: from 0 to 20)

ii. G2 - second period grade (numeric: from 0 to 20)

iii. G3 - final grade (numeric: from 0 to 20, output target)

6.3 Data Selection and Transformation

This dataset contains the categorical attributes. The G3 grade marks is considered as the label attributes, which are the final grade marks. Decision tree and naive base classifier algorithm would be used categorical attributes while artificial neural network would be used numerical attributes.

As per the requirement by the algorithm, in the transformation procedure, the "yes" instance would be represented as "1" and for "No" instance, the "0" input would be consider.

6.4 Implementation Model

In this experimental analysis, we would be using three machine learning algorithms i.e. Decision tree, Naive based and Artificial neural network. These three algorithms would be applied on two different datasets. First dataset has three types of attributes like Demographic data, Personal data and Academic data. This dataset would be used for behavioral analysis. Second dataset has two types of attributes like Demographic data and Academic data. This dataset would be performed the non behavioral analysis.

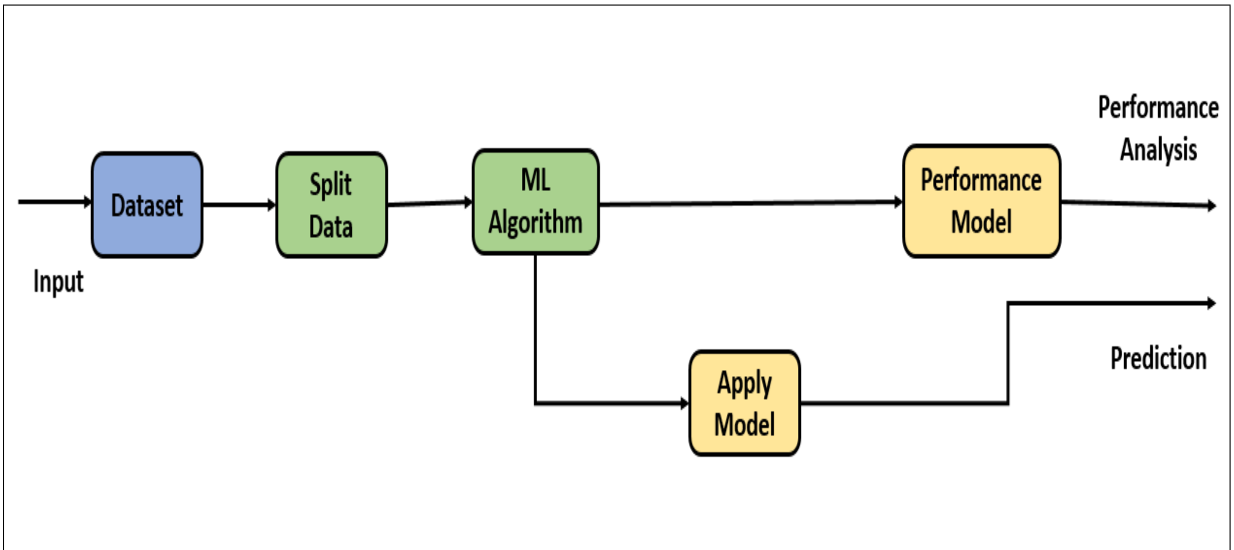

Figure 6.1: Implementation Process

The figure 6.1 represents the implementation process using all three machine learning algorithms using Rapid Miner tool. The behavioral or non behavioral dataset will be given as input. The split data function used to train the dataset for prediction purpose. Here the dataset would be split into 8:2 ratio. So 80 percent of the dataset will be considered as train data and 20 percents dataset will be considered and test data. After the splitting the machine learning algorithm (Decision tree, Naive based, ANN) would be applied. The apply model function will give the prediction as an output. In order to achieve performance analysis like accuracy of the algorithms, performance model would use. From this function, the best accurate algorithm would be able to identify.

6.5 Result Analysis:

This section consists of comparative analysis of the three ML algorithms, which analysis would be useful to choose best accurate algorithm using performance analysis matrix, the prediction of the student marks (i.e. Prediction of G3 marks) would achieve after deciding the best accurate algorithm, identifying the most impactful characteristics of the student which affects the study.

6.5.1 Comparative Analysis

	Decision Tree		Naive Base Classifier		Artificial Neural Network	
	Behavioural Prediction (%)	Nonbehavioral Prediction(%)	Behavioural Prediction(%)	Nonbehavioral Prediction(%)	Behavioural Prediction	Nonbehavioral Prediction
Accuracy	69.87	14.23	85.92	68.45	91.92	80.51
Root mean Squared Error	0.47+/-0.00	0.92+/-0.00	0.35+/-0.00	0.48+/-0.00	0.29+/-0.00	0.46+/-0.00

Figure 6.2: Performance Analysis

The Figure 6.2 describes the performance analysis of different ML algorithms like decision tree, naive based, and ANN uses behavioral data and non behavioral data. Here we have identified the accuracy and root mean squared error to determine the best ML algorithm. Here accuracy refers to the correctly classified instances and root mean squared error represents the deviation between actual value and predicted value.

So as the result we conclude from the figure 6.2, behavioral analysis is more accurate to predict student performance and for that artificial neural network is most accurate algorithm because ANN has 91.92 percent accuracy and root mean squared error is $0.29+/-0.00$ which is lowest among all three algorithms.

Analysis of Decision Tree algorithm using confusion Matrix:

accuracy: 69.87%

	true 6	true 10	true 15	true 11	true 19	true 9	true 12	true 14	true 16	true 5	true 8	true 17	true 18	true 13	true 20	true 7	true 0	true 4	class pre...
pred. 6	9	0	0	0	0	1	0	0	0	1	2	0	0	0	0	0	1	0	64.29%
pred. 10	0	25	0	4	0	1	1	0	0	0	2	0	0	2	0	0	1	1	67.57%
pred. 15	0	0	16	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	88.89%
pred. 11	0	6	0	24	0	0	7	0	0	0	1	0	1	1	0	1	0	0	58.54%
pred. 19	0	0	0	0	2	0	0	0	0	0	0	0	0	0	1	0	0	0	66.67%
pred. 9	0	3	0	0	0	13	0	0	0	0	4	0	0	1	0	1	2	0	54.17%
pred. 12	0	0	0	0	0	1	8	1	0	0	0	0	0	0	0	1	1	0	66.67%
pred. 14	0	0	1	0	0	0	1	15	0	0	0	0	0	3	0	0	0	0	75.00%
pred. 16	0	0	3	0	0	0	0	0	9	0	0	2	1	0	0	0	0	0	60.00%
pred. 5	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	1	0	0	75.00%
pred. 8	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	1	2	0	76.92%
pred. 17	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	50.00%
pred. 18	0	0	0	0	1	0	0	0	0	0	0	0	4	0	0	0	0	0	80.00%
pred. 13	0	0	0	0	0	0	2	0	0	0	0	0	0	12	0	0	0	0	85.71%
pred. 20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred. 7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred. 0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	16	0	94.12%
pred. 4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
class recall	100.00%	73.53%	80.00%	85.71%	66.67%	76.47%	42.11%	93.75%	90.00%	75.00%	52.63%	25.00%	57.14%	63.16%	0.00%	0.00%	69.57%	0.00%	

Figure 6.3: Confusion matrix of Decision tree for behavioral data

The figure 6.3 represents Confusion matrix of Decision tree for behavioral data. This confusion matrix represents the accuracy of behavioral data which represents the recall values for each class. It represents 69.87 percent accuracy.

accuracy: 14.23%

	true 6	true 10	true 15	true 11	true 19	true 9	true 12	true 14	true 16	true 5	true 8	true 17	true 18	true 13	true 20	true 7	true 0	true 4	class pre...
pred. 6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred. 10	9	34	20	28	3	17	19	16	10	4	19	4	7	19	1	5	23	1	14.23%
pred. 15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred. 11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred. 19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred. 9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred. 12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred. 14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred. 16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred. 5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred. 8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred. 17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred. 18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred. 13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred. 20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred. 7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred. 0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred. 4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
class recall	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	

Figure 6.4: Confusion matrix of Decision tree for non behavioral data

The figure 6.4 represents confusion matrix of Decision tree for non behavioral data. It has 14.23 percent accuracy which is very poor.

Analysis of Naive Base algorithm using confusion Matrix:

accuracy: 85.92%

	true 6	true 10	true 15	true 11	true 19	true 9	true 12	true 14	true 16	true 5	true 8	true 17	true 18	true 13	true 20	true 7	true 0	true 4	class pre...
pred. 6	14	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	93.33%
pred. 10	0	43	0	4	0	3	2	0	0	0	3	0	0	0	0	0	1	0	76.79%
pred. 15	0	0	30	0	0	0	0	1	3	0	0	0	0	0	0	0	0	0	88.24%
pred. 11	0	3	0	33	0	1	4	0	0	0	1	0	0	0	0	0	0	0	78.57%
pred. 19	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	100.00%
pred. 9	0	4	0	0	0	21	0	0	0	0	2	0	0	0	0	0	1	0	75.00%
pred. 12	0	0	0	3	0	0	19	1	0	0	0	0	0	1	0	0	0	0	79.17%
pred. 14	0	0	0	0	0	0	0	21	0	0	0	0	0	2	0	0	0	0	91.30%
pred. 16	0	0	0	0	0	0	0	0	11	0	0	0	1	0	0	0	0	0	91.67%
pred. 5	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	100.00%
pred. 8	0	0	0	0	0	0	0	0	0	0	22	0	0	0	0	0	2	0	91.67%
pred. 17	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	100.00%
pred. 18	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	100.00%
pred. 13	0	0	0	1	0	0	3	1	0	0	0	0	0	25	0	0	0	0	83.33%
pred. 20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	100.00%
pred. 7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	100.00%
pred. 0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	30	0	96.77%
pred. 4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	100.00%
class recall	100.00%	86.00%	100.00%	78.57%	100.00%	84.00%	67.86%	87.50%	78.57%	100.00%	75.86%	100.00%	90.91%	89.29%	100.00%	100.00%	88.24%	100.00%	

Figure 6.5: Confusion matrix of Naive Base for behavioral data

The figure 6.5 determine the confusion matrix of Naive Base algorithm for behavioral data. This confusion matrix represents the 85.92 percent accuracy which can be considerable.

accuracy: 68.45%

	true 6	true 10	true 15	true 11	true 19	true 9	true 12	true 14	true 16	true 5	true 8	true 17	true 18	true 13	true 20	true 7	true 0	true 4	class pre...
pred. 6	13	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	81.25%
pred. 10	0	32	0	9	0	8	0	0	0	0	3	0	0	0	0	0	0	0	61.54%
pred. 15	0	0	23	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	79.31%
pred. 11	0	4	0	20	0	0	5	0	0	0	0	0	0	3	0	0	0	0	62.50%
pred. 19	0	0	0	0	5	0	0	0	0	0	0	0	1	0	0	0	0	0	83.33%
pred. 9	0	4	0	1	0	14	0	0	0	0	4	0	0	0	0	0	0	0	60.87%
pred. 12	0	1	0	8	0	0	15	3	0	0	0	0	0	2	0	0	0	0	51.72%
pred. 14	0	0	4	0	0	0	1	19	0	0	0	0	0	5	0	0	0	0	65.52%
pred. 16	0	0	2	0	0	0	0	0	8	0	0	0	1	0	0	0	0	0	72.73%
pred. 5	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	100.00%
pred. 8	0	2	0	0	0	1	0	0	0	0	14	0	0	0	0	0	0	0	82.35%
pred. 17	0	0	1	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	83.33%
pred. 18	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	100.00%
pred. 13	0	0	0	1	0	0	5	2	0	0	0	0	0	18	0	0	0	0	69.23%
pred. 20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	100.00%
pred. 7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	100.00%
pred. 0	1	7	0	3	0	2	2	0	0	0	6	0	0	0	0	1	34	0	60.71%
pred. 4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	100.00%
class recall	92.86%	64.00%	76.67%	47.62%	100.00%	56.00%	53.57%	79.17%	57.14%	83.33%	48.28%	100.00%	81.82%	64.29%	100.00%	87.50%	100.00%	100.00%	

Figure 6.6: Confusion matrix of Naive Base for non behavioral data

The figure 6.6 constitutes the confusion matrix of Naive Base algorithm for non behavioral data. This matrix has 68.45 percent accuracy.

Analysis of Artificial neural network algorithm using confusion Matrix:

accuracy: 91.92%

	true 6	true 10	true 15	true 11	true 19	true 9	true 12	true 14	true 16	true 5	true 8	true 17	true 18	true 13	true 20	true 7	true 0	true 4	class pre...
pred. 6	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	87.50%
pred. 10	0	27	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	96.43%
pred. 15	0	0	16	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	88.89%
pred. 11	0	0	0	23	0	0	0	0	0	0	0	0	0	1	0	0	0	0	95.83%
pred. 19	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	100.00%
pred. 9	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	100.00%
pred. 12	0	0	0	0	0	0	13	0	0	0	0	0	0	1	0	0	0	0	92.86%
pred. 14	0	0	0	0	0	0	1	11	0	0	0	0	0	0	1	0	0	0	84.62%
pred. 16	0	0	1	1	0	0	0	0	8	0	0	0	0	0	0	0	0	0	80.00%
pred. 5	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	100.00%
pred. 8	0	0	0	0	0	0	1	0	0	0	15	0	0	0	0	0	0	0	93.75%
pred. 17	0	0	0	0	0	1	0	0	0	0	0	3	0	0	0	0	0	0	75.00%
pred. 18	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	100.00%
pred. 13	0	0	0	0	0	0	0	1	0	0	0	0	0	14	0	0	0	0	93.33%
pred. 20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred. 7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	100.00%
pred. 0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	19	0	86.36%
pred. 4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
class recall	100.00%	96.43%	94.12%	95.83%	100.00%	78.57%	86.67%	84.62%	100.00%	100.00%	93.75%	100.00%	100.00%	87.50%	0.00%	60.00%	100.00%	0.00%	

Figure 6.7: Confusion matrix of ANN for behavioral data

The figure 6.7 describes the confusion matrix of ANN algorithm for behavioral data. ANN algorithm is having 91.92 percent accuracy which is the most accurate result. So artificial neural network would be our proposed algorithm.

accuracy: 84.34%

	true 6	true 10	true 15	true 11	true 19	true 9	true 12	true 14	true 16	true 5	true 8	true 17	true 18	true 13	true 20	true 7	true 0	true 4	class pre...
pred. 6	7	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	77.78%
pred. 10	0	25	0	3	0	1	1	0	0	0	1	0	0	1	0	0	0	0	78.12%
pred. 15	0	0	15	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	88.24%
pred. 11	0	0	0	21	0	1	1	2	0	0	2	0	0	0	0	0	0	0	77.78%
pred. 19	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	100.00%
pred. 9	0	2	0	0	0	11	1	0	0	0	1	0	0	0	0	0	0	0	73.33%
pred. 12	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	100.00%
pred. 14	0	0	2	0	0	0	2	11	0	0	0	0	0	1	0	0	0	0	68.75%
pred. 16	0	0	0	0	1	0	0	0	7	0	0	0	0	0	0	0	0	0	87.50%
pred. 5	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	100.00%
pred. 8	0	0	0	0	0	0	0	0	0	2	11	0	0	0	0	0	0	0	84.62%
pred. 17	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	100.00%
pred. 18	0	0	0	0	0	0	0	0	0	0	0	0	6	0	1	0	0	0	85.71%
pred. 13	0	0	0	0	0	0	0	0	0	0	0	0	0	14	0	0	0	0	100.00%
pred. 20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred. 7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	100.00%
pred. 0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	19	0	95.00%
pred. 4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
class recall	100.00%	89.29%	88.24%	87.50%	50.00%	78.57%	66.67%	84.62%	87.50%	50.00%	68.75%	66.67%	100.00%	87.50%	0.00%	100.00%	100.00%	0.00%	

Figure 6.8: Confusion matrix of ANN for non behavioral data

The figure 6.8 represents the confusion matrix of ANN algorithm for non behavioral data. ANN algorithm as 84.34 percent accuracy.

6.5.2 Predictive Analysis

NO	G1	G2	G3	Decision Tree		Naïve Based Classifier		Artificial Neural Network	
				Behavioural	Nonbehavioral	Behavioural	Nonbehavioral	Behavioural	Nonbehavioral
1	5	6	6	6	6	6	6	6	6
2	5	5	6	6	6	6	6	6	6
3	7	8	10	10	10	8	10	10	6
4	15	14	15	14	15	15	14	15	15
5	6	10	10	10	10	10	10	10	10
6	15	15	15	15	15	15	15	15	15
7	12	12	11	11	11	12	12	11	11
8	6	5	6	6	6	6	6	6	6
9	16	18	19	19	19	19	19	19	19
10	14	15	15	15	15	15	15	15	15

Figure 6.9: Prediction of Marks

The figure 6.9 represents the prediction of behavioral and non behavioral data (first 10 records) using decision tree, naive based and artificial neural network algorithms.

ExampleSet (395 examples, 20 special attributes, 28 regular attributes) Filter (395 / 395 examples): all

Row No.	G3	prediction(G3)	confidence(6)	confidence(...	confidence(...	confidence(...	confidence(...	confidence(9)	confidence(...	confidence(...	confidence(...	confidence(5)	confidence(8)	confidence(...	confidence(...	confidence(...	con
1	6	6	0.98	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.0
2	6	6	0.94	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0
3	10	10	0.00	0.96	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0
4	15	15	0.00	0.00	0.92	0.00	0.00	0.00	0.00	0.03	0.02	0.00	0.00	0.00	0.00	0.02	0.0
5	10	10	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0
6	15	15	0.00	0.00	0.95	0.00	0.03	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.0
7	11	11	0.00	0.00	0.00	0.92	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0
8	6	6	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0
9	19	19	0.00	0.00	0.03	0.00	0.90	0.00	0.00	0.00	0.02	0.00	0.00	0.01	0.03	0.01	0.0
10	15	15	0.00	0.00	0.97	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.0
11	9	9	0.00	0.02	0.00	0.00	0.00	0.89	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.0
12	12	10	0.00	0.45	0.00	0.38	0.00	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0
13	14	15	0.00	0.00	0.47	0.00	0.00	0.00	0.04	0.00	0.31	0.00	0.00	0.03	0.07	0.03	0.0
14	11	11	0.00	0.11	0.00	0.86	0.00	0.01	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.0
15	16	16	0.00	0.00	0.00	0.02	0.04	0.00	0.00	0.00	0.92	0.00	0.00	0.01	0.00	0.00	0.0
16	14	14	0.00	0.00	0.02	0.00	0.01	0.00	0.00	0.96	0.00	0.00	0.00	0.00	0.00	0.00	0.0
17	14	14	0.00	0.00	0.05	0.03	0.00	0.00	0.00	0.88	0.00	0.00	0.00	0.03	0.01	0.00	0.0
18	10	10	0.00	0.96	0.00	0.01	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0
19	5	5	0.02	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.89	0.06	0.00	0.00	0.00	0.0
20	10	10	0.00	0.97	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.0
21	15	15	0.00	0.00	0.93	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.0
22	15	15	0.00	0.00	0.95	0.00	0.02	0.00	0.01	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.0
23	16	16	0.00	0.00	0.03	0.00	0.00	0.00	0.05	0.00	0.87	0.02	0.00	0.00	0.00	0.01	0.0
24	12	12	0.00	0.04	0.00	0.00	0.00	0.00	0.85	0.00	0.05	0.00	0.02	0.00	0.00	0.02	0.0
25	8	11	0.38	0.00	0.00	0.61	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0
26	8	8	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.0

Figure 6.10: Prediction of behavioral data using ANN algorithm

As per the figure 6.9, the behavioral data using ANN algorithm has more accuracy. So here figure 6.10 represents the prediction marks of behavioral data using ANN.

6.5.3 Predictive Analysis for Behavior of the Students

In this section we will find out the student behavior, which cause the low performance.

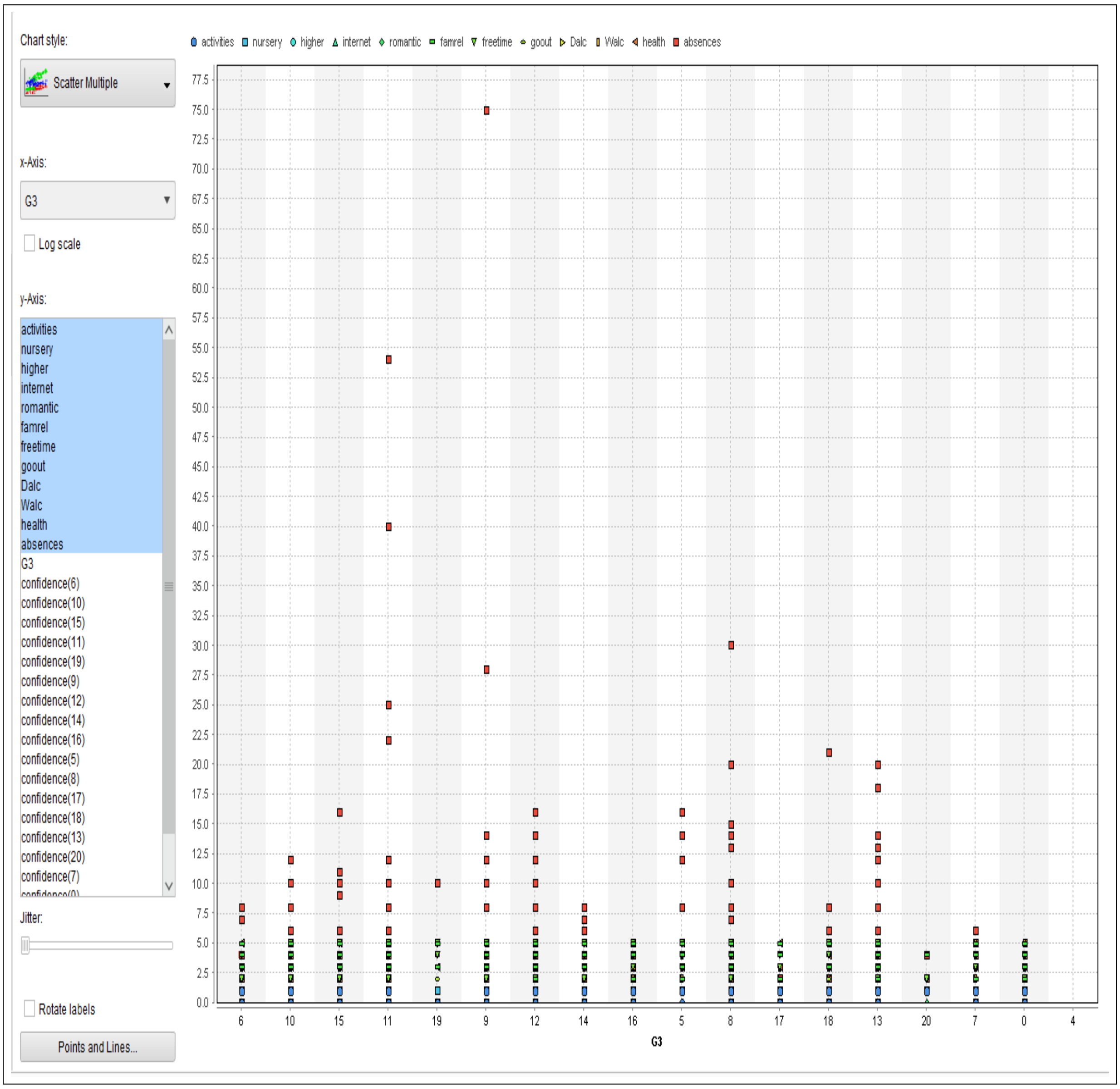

Figure 6.11: Prediction of behavior for G3 exam

In figure 6.11, it represents the behavioral analysis of the students for G3 exam. In this graph X axis contains the values of G3 exam and y axis contains the frequency of the different characteristics of the students. Here, each characteristic is plotted with different colors. For example, absence attributes plotted by red color.

As per diagram, we can conclude that for each grade mark absent attribute has large frequency. So we can say that if the students are remain absent in the classroom than it will affect the study.

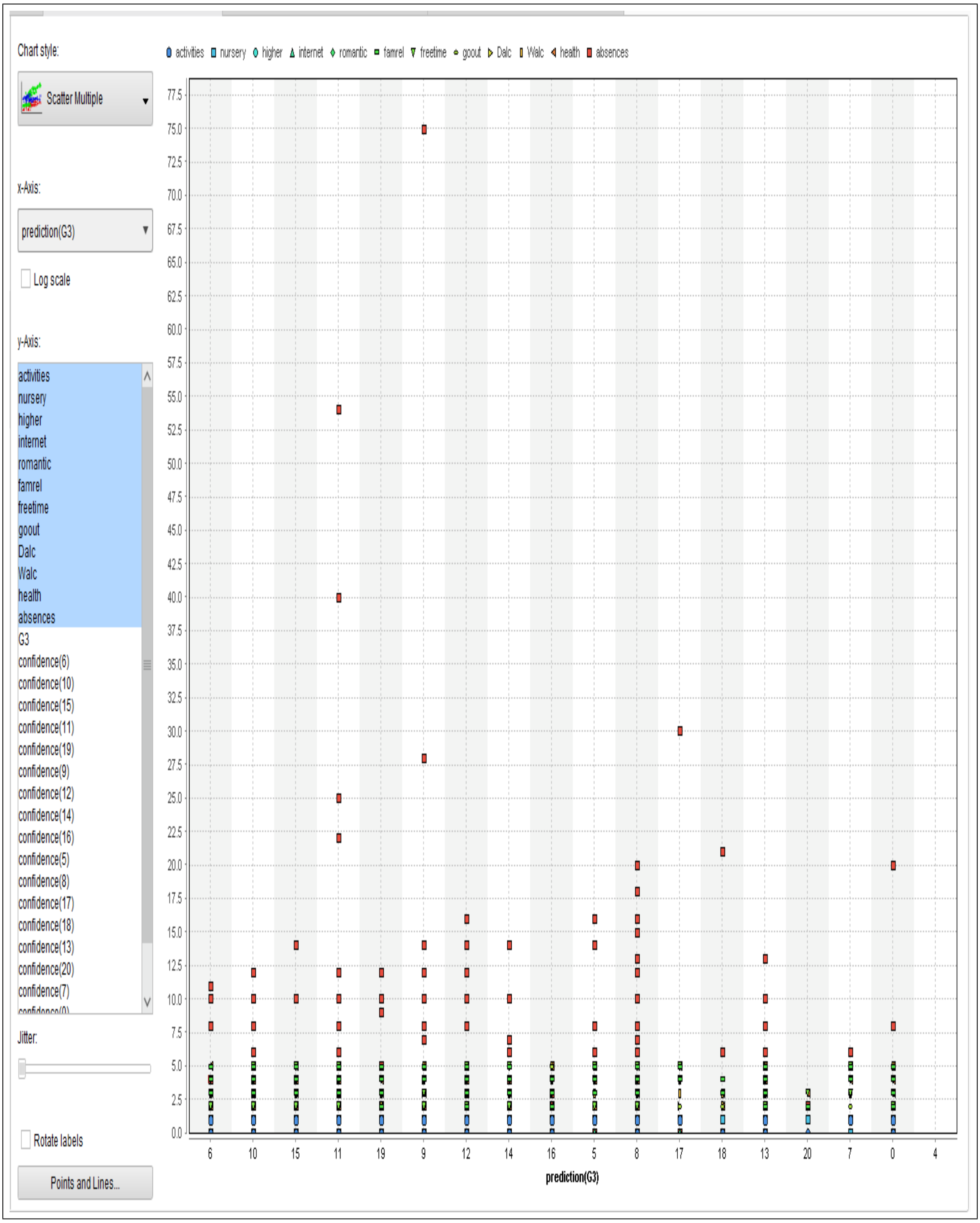

Figure 6.12: Predicted characteristic of predicted G3

The figure 6.12 determines the predicted characteristic of predicted grade of G3. Like grade 3, the large absence of the student will affect in prediction of G3.

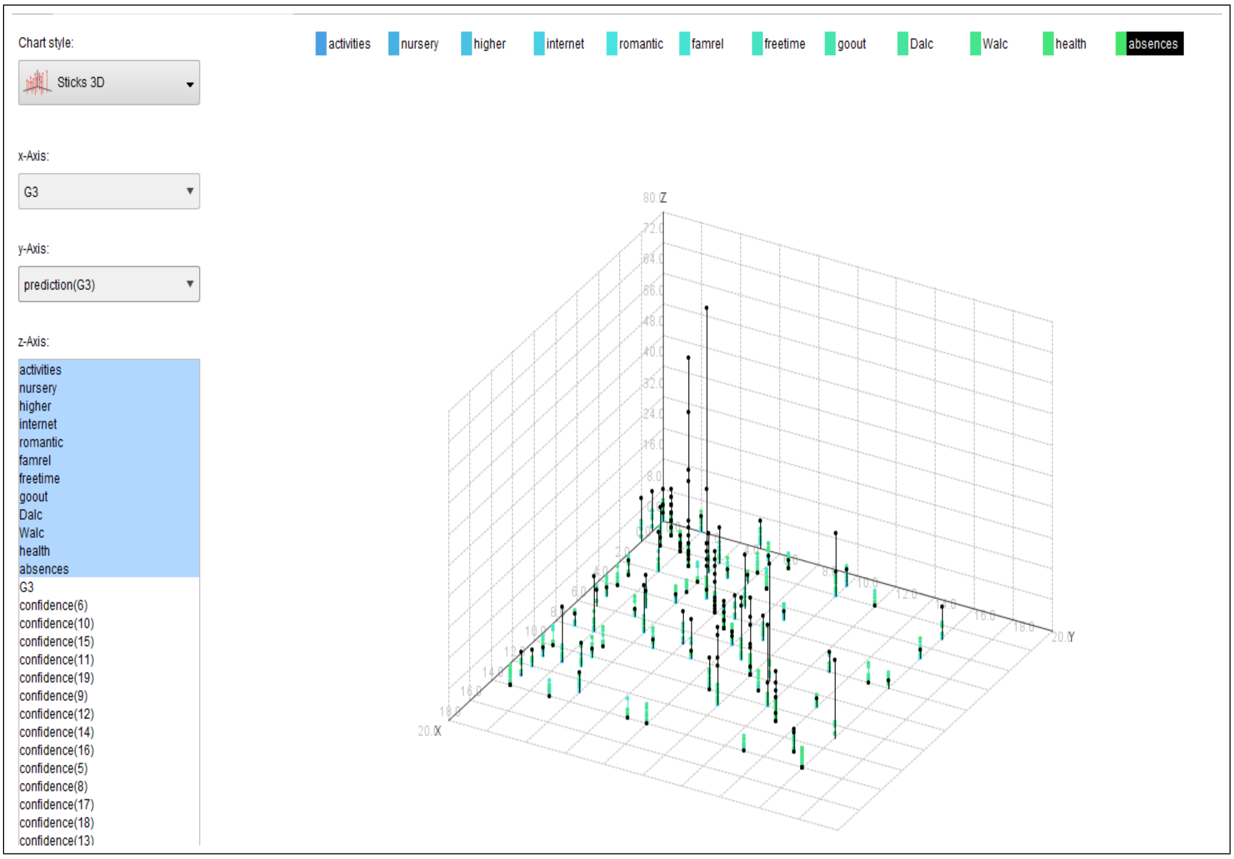

Figure 6.13: Prediction of characteristics using G3 and predicted G3

In figure 6.13 x axis contains the actual value of G3 and Y axis contains the predicted value of G3. Z axis contains the frequency of different attributes. In this graph each attributes are plotted with different colors.

Here we are predicting the characteristic for accurate result. As per the fig 6.13, when the actual value (x axis) and the predicted value will same at that time, frequency of absent attribute would also high. So we can conclude that for accurate prediction results, absent attribute has highest effectiveness.

From all the three results we can conclude that, when the students frequently remain absent in the classroom, they will be having lower performance.

6.6 Summary

This chapter resulted the predictive and comparative analysis of the student performance on the based on student's academic information and personal data. This experiment conclude that ANN is best accurate algorithm and behavioural analysis of the student performance gives precise analysis.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

This study delivers that machine learning approach provides more accuracy in student performance analysis in outcome based education system. By comparative analysis, the most precise machine learning algorithm can be found out which would be having highest accuracy. After determining the accurate algorithm, more precise predicted results can be carried out by using it. In this thesis,it has been proved that student behaviour is also an effective parameter to determine the performance of a student. Hence,overall there are several aspects to be considered while determining the student's performance and among them behaviour is one of the most prominent factor which has been proved in the thesis.

7.2 Future Work

In this thesis,the accuracy has been achieved as and this work can be extended by amalgamating several other algorithms. More effective results can be generated through use of normalized dataset. Different algorithms can be used on various other datsets so as to get more accuracy as we consider different perspective in different dataset.

Bibliography

[1] Nirma, "3cs2202 - ethical hacking - computer engineering - information and network security - institute of technology - nirma university," 2014.

[2] M. M. Mohan, S. K. Augustin, and V. K. Roshni, "A bigdata approach for classification and prediction of student result using mapreduce," pp. 145–150, 2015.

[3] N. Dimokas, N. Mittas, A. Nanopoulos, and L. Angelis, "A prototype system for educational data warehousing and mining," pp. 199–203, 2008.

[4] I. Karagiannis and M. Satratzemi, "Comparing lms and aehs: Challenges for improvement with exploitation of data mining," pp. 65–66, 2014.

[5] C. E. L. Guarín, E. L. Guzmán, and F. A. González, "A model to predict low academic performance at a specific enrollment using data mining," *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje*, vol. 10, no. 3, pp. 119–125, 2015.

[6] V. T. N. Chau, N. H. Phung, and V. T. N. Tran, "A robust and effective algorithmic framework for incomplete educational data clustering," pp. 65–70, 2015.

[7] D. T. Rover, D. Jacobson, A. Kamal, and A. Tyagi, "Implementation and results of a revised abet assessment process," 2013.

[8] P. Guleria, M. Arora, and M. Sood, "Increasing quality of education using educational data mining," pp. 118–122, 2013.

[9] M. A. F. Almeida and F. M. de Azevedo, "A theoretical model of the adaptive navigation support," pp. 195–200, 2010.

[10] K. Phythagoras and D. Samson, "Adaptive learning objects sequencing for competence–based learning," 2006.

[11] “”http://archive.ics.uci.edu/ml/datasets/student+performance”,”

[12] A. Satyanarayana and M. Nuckowski, “Data mining using ensemble classifiers for improved prediction of student academic performance,” 2016.