Mining Continuous Assessments Marks to Predict Final Results

Indika Peiris

129160

Dissertation submitted to the Faculty of Information Technology, University of Moratuwa, Sri Lanka for the partial fulfillment of the requirements of the Degree of Master of Science in Information Technology.

March 2016

DECLERATION

I hereby declare that the project work entitled "Mining Continuous Assessment Marks to Predict Final Results" submitted to the University of Moratuwa, Faculty of Information Technology, is a record of an original work done by me under the guidance of Mr. S. C. Premaratne, Senior Lecturer of Department of Information Technology, University of Moratuwa and this project work is submitted in the partial fulfillment of the requirements for the award of the degree of Master of Information Technology. The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or Diploma.

Date: March 15, 2016

Indika Pu.

G.V.A.D.I.P. Peiris

MSCIT/11/062

Supervised by

Mr. Saminda Premaratne

UOM Verified Signature

Signature

...

Date

ACKNOWLEDGEMENT

I would like mostly thank to Mr. S. C. Premaratne for his boundless and instructive help for making this research project success. Without his supervision, guidance and expert knowledge in data mining, this research would have not been possible.

I would like to thank Mr. Fordhous for guiding me to decide the scope of the project and collect necessary data.

Further I would like to thank Ms. Thilini for assisting me by providing necessary students data which was very important for this research project.

Next I would like to thank the staff members of Faculty of Information Technology for motivating and assisting me to do this course of study in numerous ways.

I take this opportunity to extend my gratitude to thank many of the colleagues and friends who have provided help, advice and companionship during my M.Sc. research study.

Finally I thank my parents for their encouragement and patience during research period.

ABSTRACT

Educational Data Mining is used to study the data available in the Universities, Higher Educational Institutes and other educational fields and extract the knowledge from it. As a result of reducing the cost of processing data and storing data, data storage became more easy and cheaper. Education Institutions are facing important and fast growth of the volume of educational data.

Data mining also called as Knowledge Discovery in Database (KDD) and search for inter relationships and patterns that can find, but already hidden among the vast volume of educational data.

Classification methods like decision trees, rule mining, Bayesian network etc can be applied on the educational data for predicting the students performance in examinations. This prediction will help the lecturers, teachers, tutors and students themselves to identify students' performance in the end semester examination. It will help the intelligent students to motivate more to maintain higher standard of marks and motivate weak students score better marks.

The J48 decision tree algorithm is applied on students' internal assessment marks to predict the grade they would gain at the end semester examination. In order do more accurate prediction some personal attributes like gender, their academic district, Advanced level Stream had been considered. With this research, students' who are likely to get higher grade or lower grade will be predicted more accurately. Predicted results can be distributed among teachers and tutors and necessary steps can be taken to improve the performance of the students who will be predicted to get lower grade or fail.

Table of Contents

Table of Figures

Table of Tables

Chapter 1

Introduction

1.1 Introduction

Educational data mining is developing methods to discover knowledge from data originating from educational environments. Main objective of higher education institutions is to provide quality education to its students. One way to achieve higher level of quality in higher education system is by discovering knowledge for predictions regarding students' performance in a particular course. The knowledge is hidden among the educational data set and it is extractable through data mining techniques. Current university education has shown a drift from traditional lecture-hall to technology driven model that merge lecture-hall teaching with web based learning management systems such as e-learning systems.

There are increasing research interests in using data mining in education. This new emerging field, called as Educational Data Mining, concerns with developing methods that discover knowledge from data originating from educational environments. Educational data mining uses many techniques such as Decision Trees, Neural Networks, Nearest Neighbor and many more. Using these techniques, many kind of knowledge can be discovered such as association rules, classifications and clustering.

Data mining provides many tasks that could be used to study the student performance. Data mining techniques such as classification, clustering and association rule mining can be used to provide guidance to students and teacher in activities such as predicting student's performance and failure rate, discovering interesting patterns among student attributes.

1.2 Background and Motivation

During last ten years education institutions have integrated Information Technology with advancements for their educational programs to improve the level of teaching and learning capacities of the students. To improve the learning capabilities of the students, the teachers and tutors should be capable to monitor the overall performance in each student separately and dynamically adjust their teaching methodologies on poor performance students and it will assist the knowledge producers change the knowledge flow and to take immediate decision to improve learning capacities.

Education data mining is a rising research discipline which is concerned with developing various methodologies to extract knowledge from educational data sources to better understand students and the way they learn. The methodologies which are used in educational data mining are differ from traditional data mining which is mainly based on exploiting the multiple levels of meaningful hierarchy in educational data.

1.3 Problem Definition

Educational data mining is developing methods to discover knowledge from data originating from educational environments. In this project, continuous assessments marks of different subjects along with some parameters had been evaluated to check whether there is any impact of those parameters for end semester results.

1.4 Objectives of the Project

The main objective of this project is to use data mining methodologies to study the impact of course assignments/continuous assessments on student performance at the end semester examination and overall final result. While continuing with the project work some attributes like gender, Z-score, Advanced Level Stream and AL District also considered to see whether there is any impact from these attributes to final results to make more accurate prediction.

Such prediction will help students to provide with early warning about their final examination results and will identify those students which need special attention to reduce fail rate and take appropriate actions for the end semester examinations and next semester examinations as well. A timely and appropriate warning to students at a risk could help preventing failing in the final exam.

Therefore Teachers can take informed decisions using such patterns and use them in improving their curriculum and strategy of teaching a class. The assertion that assignment marks have a direct correlation with final exam and overall marks can be a motivating factor for them to perform well in the assignments.

This will be used by the teachers, instructors and tutors also to identify and provide necessary guidance to the students who need more attention and also as an assistance to improve their capabilities on teaching.

Therefore data mining has a lot of potential for education and can bring lots of benefits for students and academics.

1.5 Structure of the Thesis

The introductory chapter, chapter 1provides brief idea of the project and also it provides background information of the study. This chapter wishes to present the objectives and significance of the research. chapter 2 presents different data mining techniques and tools available in the industry. It is on critical review of application of data mining in the area of Higher Education. Chapter 3 represents Technology adapted towards predicting results. Chapter 4 represents overall approach on developing model for predicting results, application of model and validation. Chapter 5 discuss the design of new system by introducing modifications to extracted rules with the intension of increase the accuracy. Chapter 6 discuss the implementation of new system along with hardware, software related implementation of the design. Chapter 7 concluded the thesis with the note on further work.

1.6 Summary

This chapter gave an overall picture of the system.

Chapter 2

Literature Review

2.1 Introduction

Chapter 1 represented the overall picture of the system. This chapter presents different data mining techniques and tools available in the industry. Later part of this chapter is on critical review of application of data mining in the area of Higher Education.

2.2 Data Mining Concept

Generally, 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, cuts 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. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. [6]

2.3 Educational Data Mining

Applying data mining (DM) in education is an emerging interdisciplinary research field also known as educational data mining (EDM). It is concerned with developing methods for exploring the unique types of data that come from educational environments. Its goal is to better understand how students learn and identify the settings in which they learn to improve educational outcomes and to gain insights into and explain educational phenomena.[7]

2.4 Data Mining Techniques

Decision trees have proved to be valuable tools for the description, classification and generalization of data. Work on constructing decision trees from data exists in multiple disciplines such as machine leaning , statistics, , decision theory, pattern recognition, artificial neural network and signal processing[8].

There are four methods for improving the intelligibility of decision tree and thereby making them more knowledge like. Three of the methods involve pruning the decision tree. It is replacing one or more sub trees with its leaves, while the remaining method reformulates the decision tree as a set of production rules [9]

In data mining, association rule learning is a popular and well researched method for discovering interesting relations between variables in large databases. It describes analyzing and presenting important rules discovered in databases using different measures of interestingness. [10]

2.5 Data Mining Tools

It is said that data is money in today's world. Along with the transition to an app-based world, it comes the exponential growth of data. However, most of the data is unstructured and hence it takes a process and method to extract useful information from the data and transform it into understandable and usable form. This is where data mining comes into picture. Plenty of tools are available for data mining tasks using artificial intelligence, machine learning and other techniques to extract data.

Rapid Miner is a famous data mining tool. It is written in the Java Programming language. This tool offers advanced analytics through template-based frameworks. A bonus: Users hardly have to write any code. Offered as a service, rather than a piece of local software, this tool holds top position on the list of data mining tools.

In addition to data mining, Rapid Miner also provides functionality like data preprocessing and visualization, predictive analytics and statistical modeling, evaluation, and deployment. What makes it even more powerful is that it provides learning schemes, models and algorithms from WEKA and R scripts.[11]

Weka is another data mining tool. The original non-Java version of WEKA primarily was developed for analyzing data from the agricultural domain. With the Java-based version, the tool is very sophisticated and used in many different applications including visualization and algorithms for data analysis and predictive modeling. Its free under the GNU General Public License, which is a big plus compared to Rapid Miner, because users can customize it however they please.[11]

R is a language and environment for statistical computing and graphics. R provides a wide variety of statistical (linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering) and graphical techniques, and is highly extensible. The S language is often the vehicle of choice for research in statistical methodology, and R provides an Open Source route to participation in that activity.

One of R's strengths is the ease with which well-designed publication-quality plots can be produced, including mathematical symbols and formulae where needed. Great care has been taken over the defaults for the minor design choices in graphics, but the user retains full control. [12]

2.6 Related Work

B. Minaei-Bidgoli, D.A. Kashy, G. Kortmeyer and W.F.Punch [13] have introduced a classification approach that can be used to predict the final result for course in a web based education system.

M.N. Quadri and Dr. N.V. Kalyankar [14] have predicted students' academic performance using CGPA grade system where the data set comprised of the students gender, his parents education details, his financial background. In [15] author has explored the various variables to predict the students who are at risk to fail the exam.

S. Anupama Kumar and Dr. Vijayalakshmi M.N has done a comparison of J48 and ID3 algorithms to find the efficiency of decision tree in predicting student's academic performance.[16]. With this research they have identified that J48 algorithm is more accurate than ID3 algorithm.

G. Naga Raja Prasad and Dr. A. Vinaya Babu have applied C4.5 decision tree algorithm on students first year and second year marks obtained for MCA to predict their performance in final year examination. The outcome of the decision tree has predicted the number of students who are likely to 'pass' or 'fail'.[17]

2.7 Summary

This chapter contains brief description of available data mining techniques and data mining tools with respect to the current research area. Further it discusses several attributes and different data mining techniques used by some researchers to predict student performance in higher education institutions.

Chapter 3

Analysis and Technology Adapted towards Prediction

3.1 Introduction

Chapter 2 represented literature review for the system. This chapter presents selection of attributes for prediction purpose, find the inter relationships and data analysis processes. Later part of this chapter describes the selection of method to validate the results and find the accuracy of the results.

3.2 Selection of inputs for the system

In order to do predictions it is required to identify the attribute(s) or factor(s) which will be influence the final results. Therefore the following attributes were considered to see the possibility of predictions.

1. Z Score

   From 0.21 to 2.01

2. District

   Kandy, Hambantota, Colombo, Polonnaruwa, Galle, Matara, Kaluthara, Gampaha,Jaffna, Anuradhapura, Nuwara Eliya, Mulathivu, Ampara, Kilinochchi, Rathnapura, Batticaloa, Trincomalee, Kurunegala, Mathale ,Badulla,Moneragala, Puttalam, Mannar, Kegale, Vavuniya

3. Gender

   M - Male F - Female

4. A/L Stream

   Mathematics, Bio-Science, ICT/Maths, Commerce, Arts

5. Continuous Assessment Marks (Grades)

   Range from 40% - 100%

   Marks>=69.5 A

Marks>=54.5 B

Marks>=39.5 C

Marks>=34.4 D

Marks <34.4 F

6. Final Exam Paper Marks (?)

   Range from 0% - 100%

   Marks>=69.5 A

   Marks>=54.5 B

   Marks>=39.5 C

   Marks>=34.4 D

   Marks <34.4 F

   (not considered for predictions due to confidentiality issues)

7. Grade for Final Marks

   Range from 0% - 100%

   Marks>=69.5 A

   Marks>=54.5 B

   Marks>=39.5 C

   Marks>=34.4 D

   Marks <34.4 F

3.3 Process

After the basic attributes or the factors are identified, they will be used to create a model to evaluate the students' end semester results for a particular subject and their performance.

3.4 Data Mining Techniques used in this Project

There are several Data Mining Techniques, such as Association, Clustering, Classification, Regression and Deviation. In this project results prediction has been done using Classification Data Mining Technique with Decision Tree and Rule Based Methods. The classification rule generation process is based on the decision tree as a classification method where the generated rules are studied and evaluated.

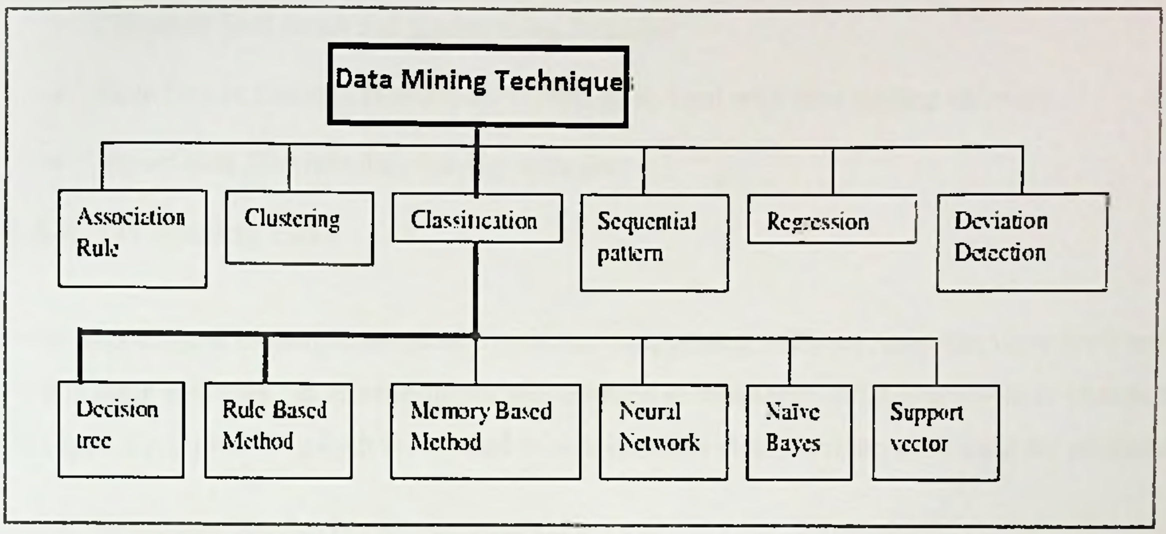


Figure3. 1 Data mining Techniques and Methods

3.5 Expected Output

Predicted Final Results as: A,B,C,D or F

3.6 Data Analysis Process

Data Analysis Process consist of 5 steps as follows:

1. Data understanding and pre-preparation
2. Visualizing data
3. Creating Model for Data Analysis
4. Applying the Model for Data Analysis
5. Validating the Model

3.6.1 Data understanding and Pre-Process

- Data (students' personal records) is stored in an Excel sheet at the pre processing step.
- It is required to delete incomplete records to maintain complete set of records.
- It is required to delete unwanted columns like students' address, contact no, NIC as they do not used for prediction purposes .
- Enter continuous assessment marks and final exam Grade to excel sheets.

- Calculate final result and grades using formulas.
- Save files as Excel files and CSV format to be used with data mining software.
- Import data files into data mining software.

3.6.2 Visualizing Data

After importing or loading data sheets to relevant data mining software, user can view such records. At this stage also user can remove unwanted columns in data sheet and it is possible to change the data type if required. Finally It is required to nominate the class attribute to be used for prediction.

3.6.3 Creating Model for Data Analysis

Predictive analytics detects patterns in large amounts of data with the help of machine learning methods. The result is a model which can be used to predict the outcome for new situations.

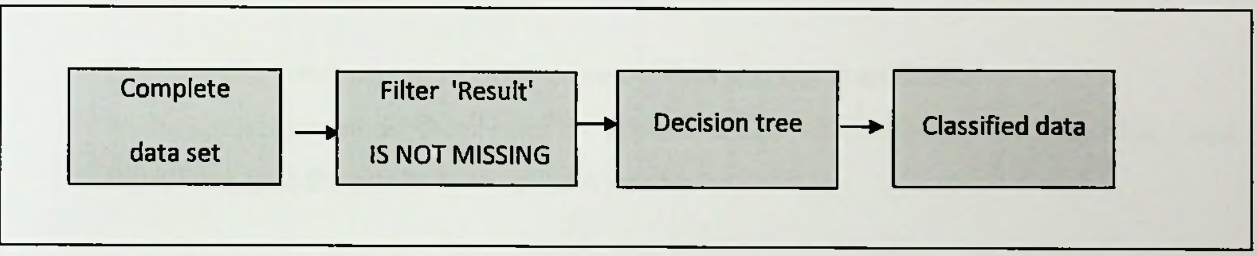


Figure3. 2 Creation of Model for prediction

Preparation of model involves identifying set of rules that can be used for data prediction. Expected outcome is set of rules that can be used for data classification

3.6.4 Apply the Model for Data Analysis

- Take full data set.
- Make another copy of full data set.
- Take one 'Filter' to filter out records where 'result' field 'Is Not Missing'.
- Take another 'Filter' to filer out the records where 'result' field 'Is Missing'.
- Send records through 'Decision Tree' and then 'Apply Model' which are sent out from the first filter.
- Sent the records through 'Apply Model' which are sent out from second filter.
- Get the predicted results.

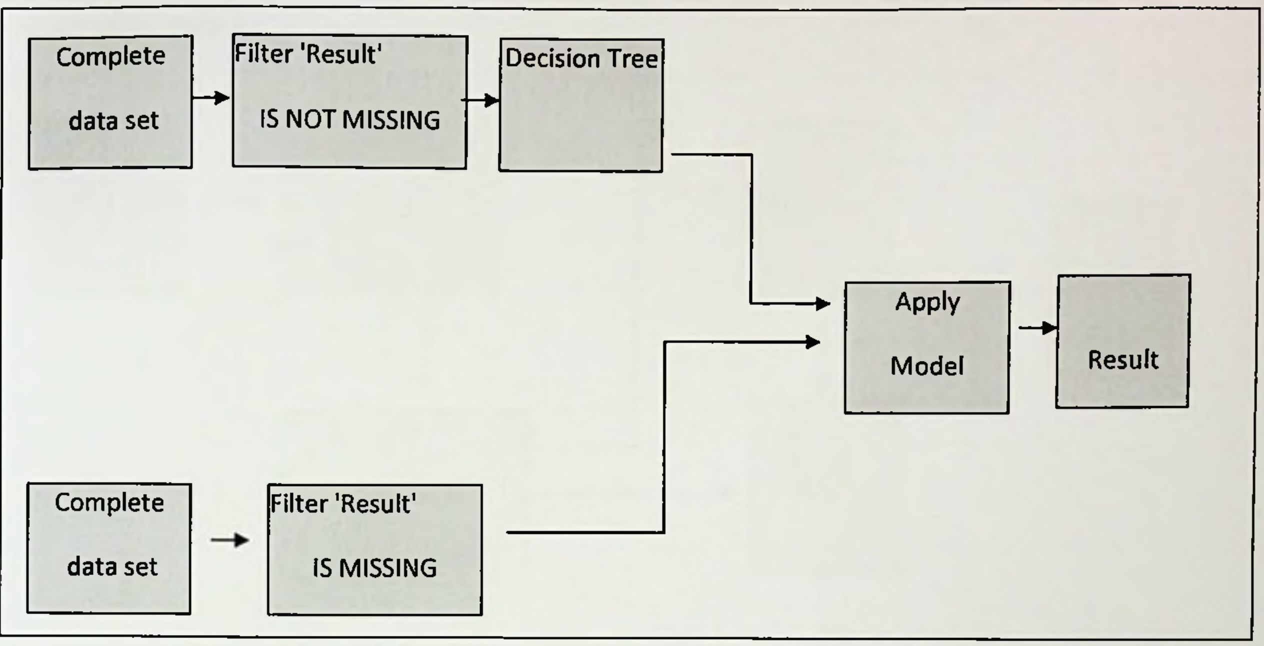


Figure3. 3 Application of Model

- Decision tree creates set of rules based on the data sent out from filter 1.
- At the application model those rules are applied to data which are sent out from filter 1 and filter 2 and give the results base on the created rules.

3.6.5 Validate the Model

- After application of the model, it is possible to view the predicted output (results)
- Then it is required to apply 'Validation' operator
- During validation process complete data set is used as 'training data'.
- Then apply 'apply the model' and 'performance' operator on 'test data'.

Validation Process follows 4 steps

Validation: step 1

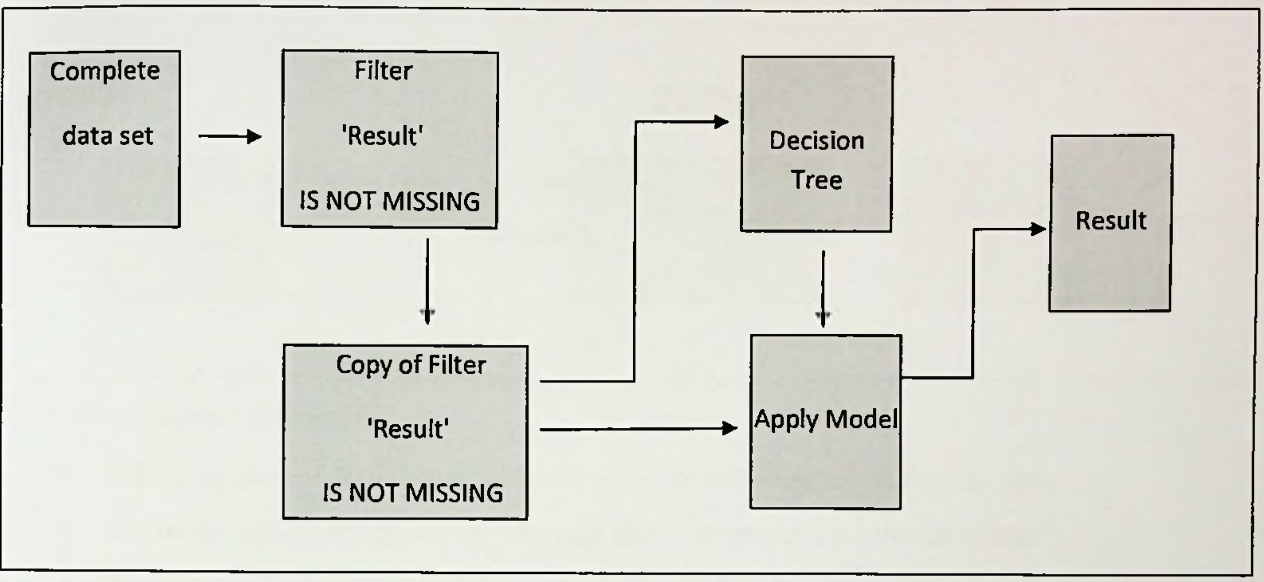


Figure3. 4 Validation of Model step1

- Apply the model and get the predicted results for whole set of data where results field is 'incomplete' or 'missing' data and results filed is 'complete' or 'not missing'.

Validation: step 2

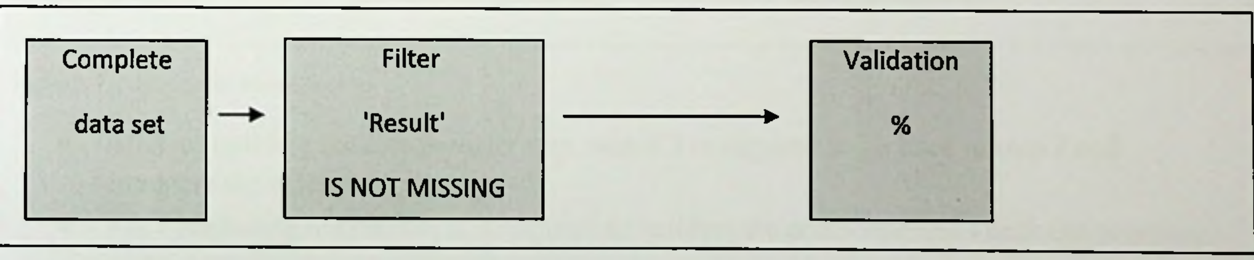


Figure3. 5 Validation of Model Step 2

- In this step 'Validation' operator applies for complete set of data.

Validation: step 3

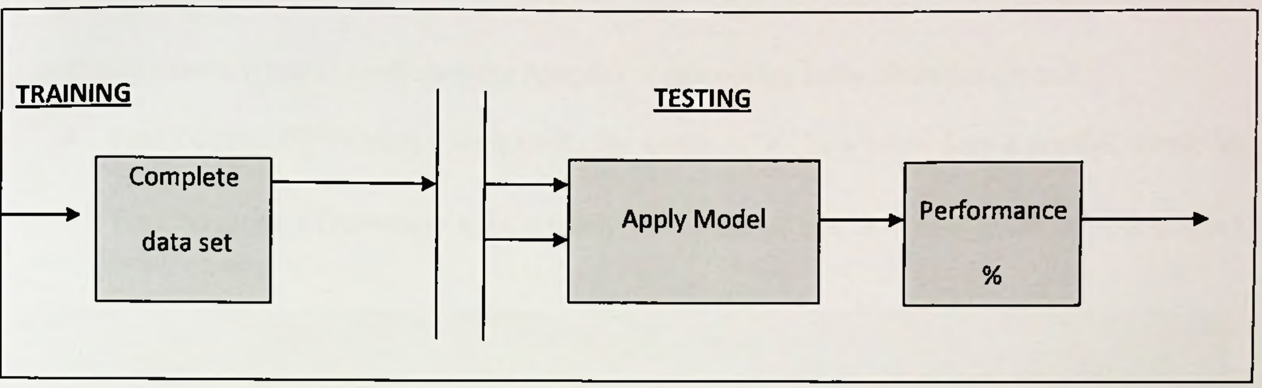


Figure3. 6 Validation of Model step 3

- This is an intermediate step of step two which consist training and testing data.
- **Training**- complete data set sent through decision tree and get the set of rule.
- **Testing**- apply those rules on data and get the predicted results.

Validating step 4

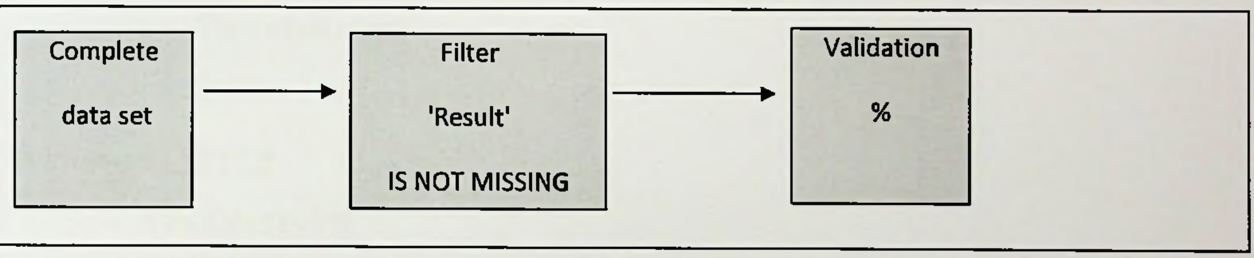


Figure3. 7 Validation of Model step 4

- After completing the intermediate step, step 3 it is required to go back to step 2 and complete the validation process.
- After predicting the results, it is required to validate the prediction and check the accuracy.

3.7 Calculating the Accuracy

Confusion Matrix is used to calculate the accuracy. (How often is the classifier correct?)

- True Positive (TP)=when it is actually the Grade is 'x' ,how often does it predict Grade as 'x'
- True Negative (TN)=when it is actually the Grade is not 'x' ,how often does it predict Grade as not 'x'

		Predicted Class	
Actual Class		Predicted = No	Predicted = Yes
	Actual = No	TN	FP
	Actual = Yes	FN	TP

Table 3. 1 Confusion Matrix

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total of all cases}}$$

$$\textbf{Accuracy} = \frac{\mathbf{TP+TN}}{\mathbf{TP+TN+FP+FN}}$$

3.8 Summary

This Chapter covers the selection of inputs for the system, selection of data mining techniques, preparation of model to analyze and predict data. The later part of this chapter described the method used to validate the results obtained through analysis.

Chapter 4

Development Approach

4.1 Introduction

Chapter 3 described the selection of inputs for the system, preparation of model to predict data and validate the results. This chapter describes the usage of decision tree, comparison of results obtained and application of the prepared model with real data in order to determine the attributes to be used for results prediction.

4.2 Application of Model and Validation

Initial model was created using data mining software and validation is done to verify the accuracy of predictions. Different selected individual attributes and then combination of selected attributes are tested using the model for results prediction. Then the result generated by the model is tested to verify the accuracy of prediction in order to determine which attribute(s) can be used successfully to predict the results.

4.2.1 Prediction of Results Based on One Attribute – Subject3

Classifier: Trees J48

Test mode: 10-fold cross-validation

Software: Weka and Rapid Miner

- **Z-score →Result**

J48 pruned tree

Z-Score <= 1.3702: B (67.0/33.0)

Z-Score > 1.3702: A (193.0/106.0)

Summary

Correctly Classified Instances	109	41.9231 %
Incorrectly Classified Instances	151	58.0769 %

- **District → Result**

J48 pruned tree

District = Ratnapura: B (14.0/7.0)

District = Kurunegala: B (20.0/11.0)

District = Colombo: A (51.0/20.0)

District = Kalutara: B (12.0/7.0)

District = Kandy: A (15.0/5.0)

District = Hambantota: B (7.0/4.0)

District = Ampara: B (7.0/3.0)

District = Galle: A (17.0/6.0)

District = Kegalle: B (13.0/5.0)

District = Badulla: A (9.0/5.0)

District = Jaffna: A (10.0/4.0)

District = Anuradhapura: B (7.0/3.0)

District = Matara: A (16.0/8.0)

District = Mannar: B (3.0/1.0)

District = Puttalam: B (9.0/5.0)

District = Batticaloa: B (6.0/1.0)

District = Nuwara Eliya: B (6.0/2.0)

District = Gampaha: A (19.0/11.0)

District = Vavuniya: C (4.0/2.0)

District = Moneragala: B (4.0/2.0)

District = Trincomalee: B (6.0/2.0)

District = Matale: B (3.0/2.0)

District = Polonnaruwa: B (2.0/1.0)

Summary

Correctly Classified Instances	125	48.0769 %
Incorrectly Classified Instances	135	51.9231 %

- **Gender → Results**

J48 pruned tree

Gender = M: A (123.0/67.0)

Gender = F: B (137.0/81.0)

Summary

Correctly Classified Instances	108	41.5385 %
Incorrectly Classified Instances	152	58.4615 %

- **Continuous Assessments Marks → Results**

J48 pruned tree

CA1 = B: A (123.0/64.0)

CA1 = C: B (85.0/39.0)

CA1 = A: A (52.0/12.0)

Summary

Correctly Classified Instances	145	55.7692 %
Incorrectly Classified Instances	115	44.2308 %

- **Advanced Level Stream → Results**

J48 pruned tree

AL Stream = Biology: B (96.0/46.0)

AL Stream = Maths: A (119.0/52.0)

AL Stream = ICT/Maths: A (3.0)

AL Stream = Arts: C (33.0/19.0)

AL Stream = Commerce: C (9.0/3.0)

Summary

Correctly Classified Instances	137	52.6923 %
Incorrectly Classified Instances	123	47.3077 %

4.2.2 Summary of Prediction using one Attribute -Subject 3

Attribute: Result	Correctly Classified Ratio (%)
Z-score: Result	41%
District: Result	48%
Gender: Result	41%
Continuous Assessments Marks: Result	55%
Advanced Level Stream: Result	52%

Table 4. 1 Summary of Prediction- subject 3

Since CA → Marks score the highest correctly classified ratio, CA is used for further classifications.

4.2.3 Prediction of Results Based on Two Attributes - Subject 3

- **Gender, CA → RESULTS**

Summary

Correctly Classified Instances 145 55.7692 %

Incorrectly Classified Instances 115 44.2308 %

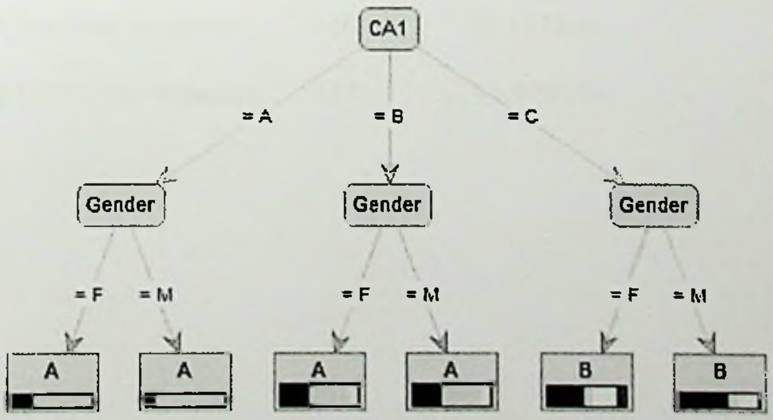


Figure 4. 1 Decision Tree-CA:Gender

```
CA1 = A
|   Gender = F: A {B=7, A=18, F=0, C=0, D=0}
|   Gender = M: A {B=4, A=22, F=0, C=1, D=0}
CA1 = B
```

```
|   Gender = F: A {B=24, A=29, F=0, C=7, D=1}
|   Gender = M: A {B=24, A=30, F=0, C=7, D=1}
CA1 = C
|   Gender = F: B {B=25, A=3, F=1, C=16, D=6}
|   Gender = M: B {B=21, A=4, F=0, C=8, D=1}
```

- **District, CA → RESULTS**

Summary

Correctly Classified Instances 145 55.7692 %

Incorrectly Classified Instances 115 44.2308 %

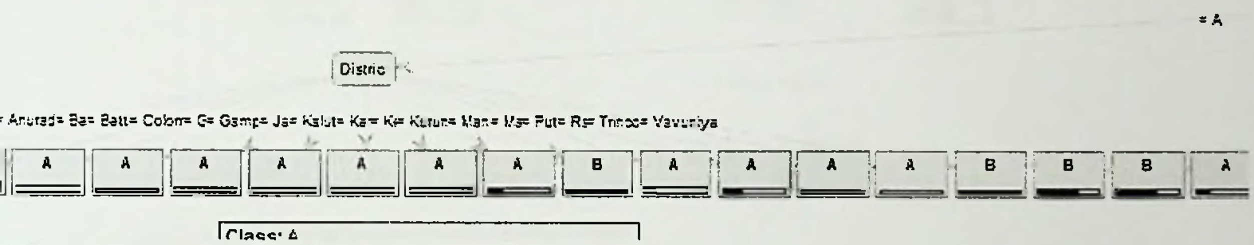


Figure 4. 2 Decision Tree CA=A: District

- **Z-Score, CA → RESULTS**

Summary

Correctly Classified Instances 146 56.1538 %

Incorrectly Classified Instances 114 43.8462 %

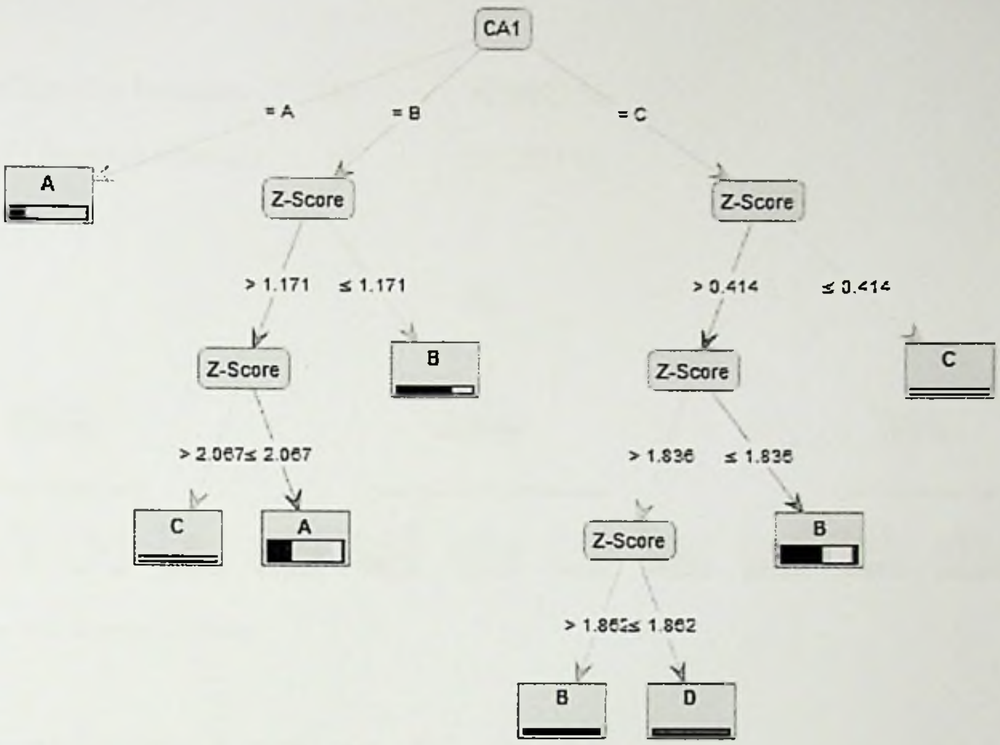


Figure 4. 3 Decision Tree Z score: Results

```
CA1 = A: A {B=11, A=40, F=0, C=1, D=0}
CA1 = B
|   Z-Score > 1.171
|   |   Z-Score > 2.067: C {B=0, A=0, F=0, C=1, D=0}
|   |   Z-Score ≤ 2.067: A {B=36, A=57, F=0, C=11, D=2}
|   Z-Score ≤ 1.171: B {B=12, A=2, F=0, C=2, D=0}
CA1 = C
|   Z-Score > 0.414
|   |   Z-Score > 1.836
|   |   |   Z-Score > 1.862: B {B=1, A=0, F=0, C=0, D=0}
|   |   |   Z-Score ≤ 1.862: D {B=0, A=0, F=0, C=0, D=2}
|   |   Z-Score ≤ 1.836: B {B=45, A=7, F=1, C=22, D=5}
|   Z-Score ≤ 0.414: C {B=0, A=0, F=0, C=2, D=0}
```

- **Advanced Level Stream, CA → RESULTS**

Correctly Classified Instances	163	62.6923 %
Incorrectly Classified Instances	97	37.3077 %

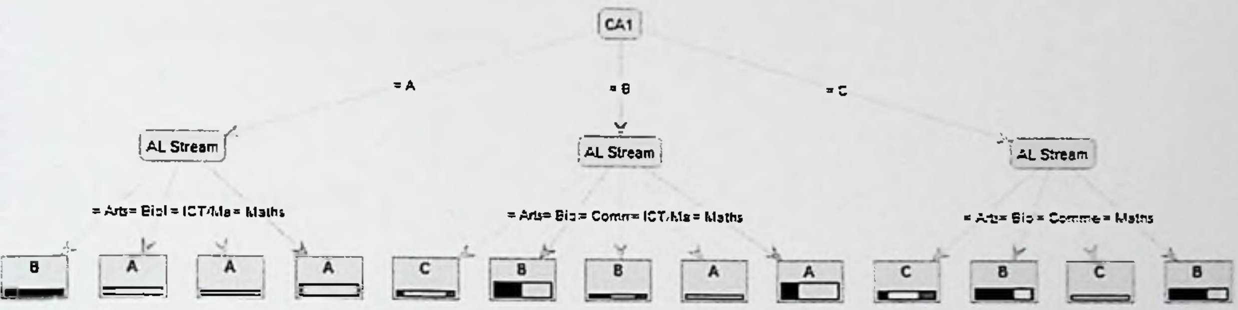


Figure 4. 4 Decision Tree CA: AL Stream

CA1 = A

```
|    AL Stream = Arts: B {B=6, A=0, F=0, C=0, D=0}
|    AL Stream = Biology: A {B=3, A=10, F=0, C=1, D=0}
|    AL Stream = ICT/Maths: A {B=0, A=2, F=0, C=0, D=0}
|    AL Stream = Maths: A {B=2, A=28, F=0, C=0, D=0}
CA1 = B
|    AL Stream = Arts: C {B=1, A=1, F=0, C=4, D=1}
|    AL Stream = Biology: B {B=26, A=23, F=0, C=3, D=0}
|    AL Stream = Commerce: B {B=2, A=0, F=0, C=2, D=1}
|    AL Stream = ICT/Maths: A {B=0, A=1, F=0, C=0, D=0}
|    AL Stream = Maths: A {B=19, A=34, F=0, C=5, D=0}
CA1 = C
|    AL Stream = Arts: C {B=4, A=0, F=0, C=10, D=6}
|    AL Stream = Biology: B {B=21, A=2, F=0, C=6, D=1}
|    AL Stream = Commerce: C {B=0, A=0, F=0, C=4, D=0}
|    AL Stream = Maths: B {B=21, A=5, F=1, C=4, D=0}
```

4.2.4. Summary of Prediction Using Two Attributes-Subject3

Attribute: Result	Correctly Classified Ratio (%)
Gender, CA: Result	55%
District, CA: Result	55%
Z-Score, CA: Result	56%
Advanced Level Stream, CA: Result	62%

Table 4. 2 summary of prediction with two attributes

4.2.5 Prediction with three attributes - Subject3

- **District, AL,CA → RESULTS** has **56%** correctly classified instance rate.

Since Advanced Level Stream, CA score the highest correctly classified ratio, AL Stream and CA is used for further classifications and Predictions.

4.2.6 Application of Model - Subject3

CA, AL, → Results

Total number of records 265

Missing records 50

Row No	RESULTS1	prediction(RESULTS1)	confidence(B)	confidence(A)	confidence(F)	confidence(C)	confidence(D)	Reg No.	AL Stream	CA1
1	B	B	0.478	0.457	0	0.065	0		Biology	B
2	B	B	0.741	0.111	0.037	0.111	0		Maths	C
3	A	A	0.325	0.587	0	0.087	0		Maths	B
4	F	B	0.741	0.111	0.037	0.111	0		Maths	C
5	A	A	0.111	0.869	0	0	0		Maths	A
6	C	A	0.231	0.692	0	0.077	0		Biology	A
7	A	B	0.478	0.457	0	0.065	0		Biology	B
8	C	A	0.326	0.587	0	0.087	0		Maths	B
9	B	B	0.741	0.111	0.037	0.111	0		Maths	C
10	B	A	0.326	0.587	0	0.087	0		Maths	B
11	C	A	0.326	0.587	0	0.087	0		Maths	B
12	A	A	0.111	0.869	0	0	0		Maths	A
13	B	B	0.741	0.111	0.037	0.111	0		Maths	C
14	B	B	0.741	0.111	0.037	0.111	0		Maths	C
15	A	A	0.325	0.587	0	0.087	0		Maths	B
16	A	A	0.111	0.869	0	0	0		Maths	A
17	A	A	0.326	0.587	0	0.087	0		Maths	B
18	A	B	0.478	0.457	0	0.065	0		Biology	B
19	C	B	0.682	0.045	0	0.227	0.045		Biology	C
20	A	A	0.326	0.587	0	0.087	0		Maths	B

Contd..

Row No	RESULTS1	prediction(RESULTS1)	confidence(B)	confidence(A)	confidence(F)	confidence(C)	confidence(D)	Reg No	AL Stream	CA1
144	A	A	0.111	0.889	0	0	0		Maths	A
145	B	B	0.741	0.111	0.037	0.111	0		Maths	C
146	B	B	0.478	0.457	0	0.055	0		Biology	B
147	C	B	0.478	0.457	0	0.055	0		Biology	B
148	A	B	0.741	0.111	0.037	0.111	0		Maths	C
149	B	B	0.741	0.111	0.037	0.111	0		Maths	C
150	?	B	0.741	0.111	0.037	0.111	0		Maths	C
151	?	A	0.111	0.889	0	0	0		Maths	A
152	?	A	0.326	0.587	0	0.087	0		Maths	B
153	?	A	0.326	0.587	0	0.087	0		Maths	B
154	?	B	0.682	0.045	0	0.227	0.045		Biology	C
155	?	A	0.111	0.889	0	0	0		Maths	A
156	?	B	0.682	0.045	0	0.227	0.045		Biology	C
157	?	A	0.326	0.587	0	0.087	0		Maths	B
158	?	B	0.682	0.045	0	0.227	0.045		Biology	C
159	?	B	0.478	0.457	0	0.055	0		Biology	B
160	?	B	0.741	0.111	0.037	0.111	0		Maths	C
161	?	A	0.111	0.889	0	0	0		Maths	A
162	?	B	0.478	0.457	0	0.055	0		Biology	B
163	?	A	0.111	0.889	0	0	0		Maths	A

Table 4. 3 Application of Model - subject 3

4.2.7 Validation of Model – Subject3

CA, AL, → Results

Row No	RESULTS1	prediction(RESULTS1)	confidence(B)	confidence(A)	confidence(F)	confidence(C)	confidence(D)	Reg No	AL Stream	CA1
244	C	B	0.400	0	0	0.400	0.200		Commerce	B
245	B	B	0.500	0.442	0	0.058	0		Biology	B
246	D	C	0.143	0.143	0	0.571	0.143		Arts	B
247	D	B	0.400	0	0	0.400	0.200		Commerce	B
248	C	B	0.500	0.442	0	0.058	0		Biology	B
249	D	C	0.200	0	0	0.500	0.300		Arts	C
250	A	B	0.500	0.442	0	0.058	0		Biology	B
251	D	C	0.200	0	0	0.500	0.300		Arts	C
252	B	A	0.328	0.586	0	0.086	0		Maths	B
253	C	C	0.200	0	0	0.500	0.300		Arts	C
254	B	B	1	0	0	0	0		Arts	A
255	B	B	0.700	0.067	0	0.200	0.033		Biology	C
256	D	C	0.200	0	0	0.500	0.300		Arts	C
257	D	C	0.200	0	0	0.500	0.300		Arts	C
258	A	A	0.328	0.586	0	0.086	0		Maths	B
259	B	C	0.200	0	0	0.500	0.300		Arts	C
260	C	C	0.200	0	0	0.500	0.300		Arts	C
261	B	B	0.700	0.067	0	0.200	0.033		Biology	C
262	C	C	0.200	0	0	0.500	0.300		Arts	C
263	C	C	0.200	0	0	0.500	0.300		Arts	C

Table 4. 4 Validation of Predicted Results - subject 3

4.2.8 Calculating the Accuracy - Subject3

Confusion Matrix is used to calculate the accuracy of the predictions

accuracy: 62.52% +/- 6.68% (mikro: 62.50%)

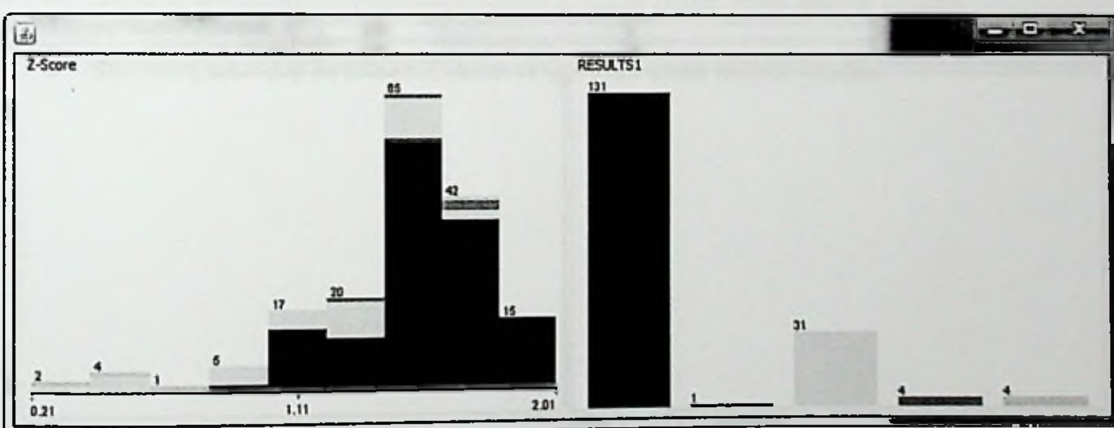
	true B	true A	true F	true C	true D	class precision
pred. B	51	10	1	10	1	69.86%
pred. A	49	96	0	15	1	59.63%
pred. F	0	0	4	0	0	100.00%
pred. C	5	0	0	14	7	53.85%
pred. D	0	0	0	0	0	0.00%
class recall	48.57%	90.57%	80.00%	35.90%	0.00%	

Table 4. 5 Accuracy Calculation with Confusion Matrix - subject 3

4.2.9. Prediction of Results based on One Attribute - Subject 2

Subject 2 (B12)

Z-Score→Results



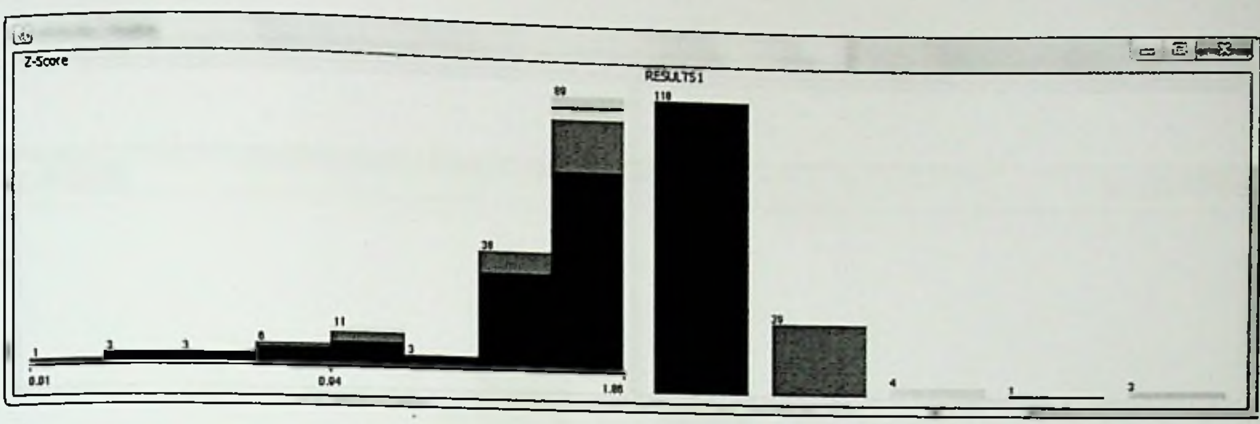
No.	Label	Count
1	A	131
2	D	1
3	B	31
4	F	4
5	C	4

Figure 4. 5 Z-score → results (B12) subject2

- No 'A' grades for student who are in lower range of Z score
- Most students in higher Z score range gets 'A' grade
- Most students who are in lower Z score range get 'C' and 'D' grades
- Students who have fail this subject are in middle range of Z score

Subject 2 (B13)

Z-Score→Results



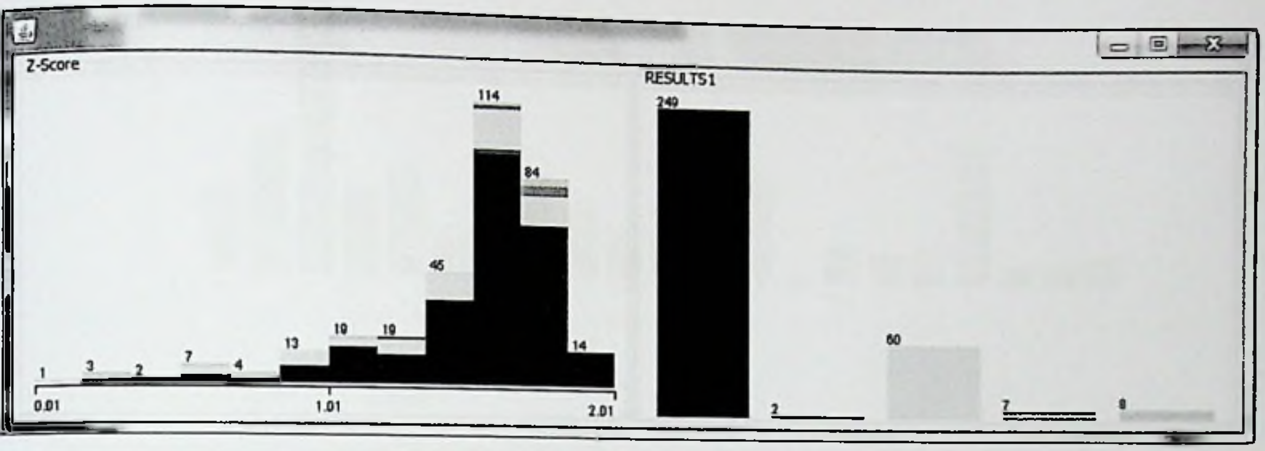
No.	Label	Count
1	A	118
2	B	29
3	C	4
4	D	1
5	F	3

Figure 4. 6 Z-score → results (B13) Subject 2

- Most of the students who have higher Z score range score higher grades.
- Students who are in lower Z score range also score higher grades.

Both Batches (*larger data set*)

Z-Score→Results



No.	Label	Count
1	A	243
2	B	60
3	C	8
4	D	2
5	F	4

Figure 4. 7 Z-score → results (B12+B13) Subject 2

When considering two batches 76.3% students get 'A' grade for this subject.

Heights Z-Score scores Higher Results.

Subject 2 (B12)

District →Results

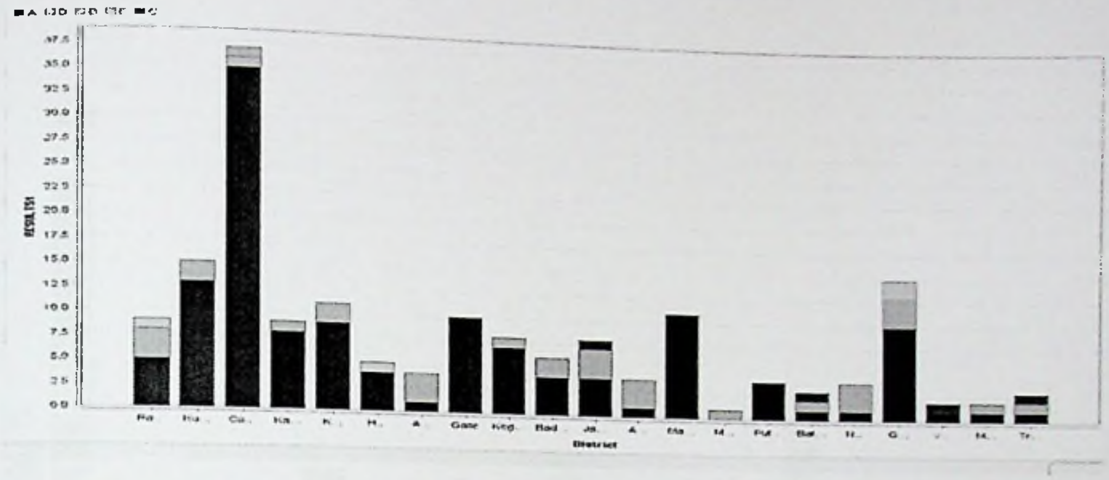


Figure 4. 8 District → results (B12) Subject 2

- All students from Galle, Mathara, Puththalam reach to A pass
- students from Mannar do not reach to A pass

Subject 2 (B13)

District →Results

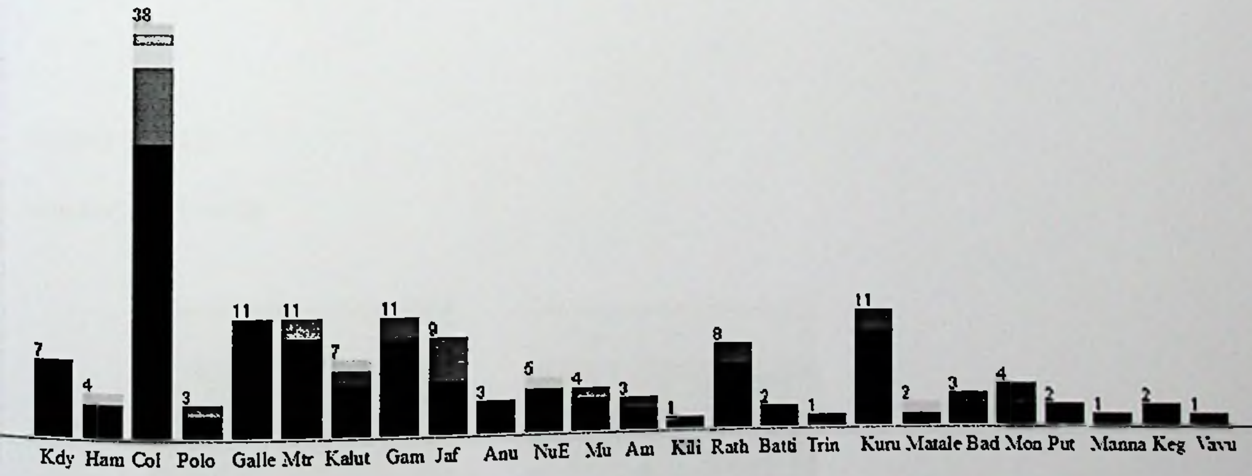


Figure 4. 9 District → results (B13) - Subject 2

- Kegalle and Vavniya do not reach A
- Kandy, Galle, Anuradhapura, Batticaloa, Trincomalee, Puttalam, Mannar all have reached to A pass

Both Batches (larger data set)

District →Results

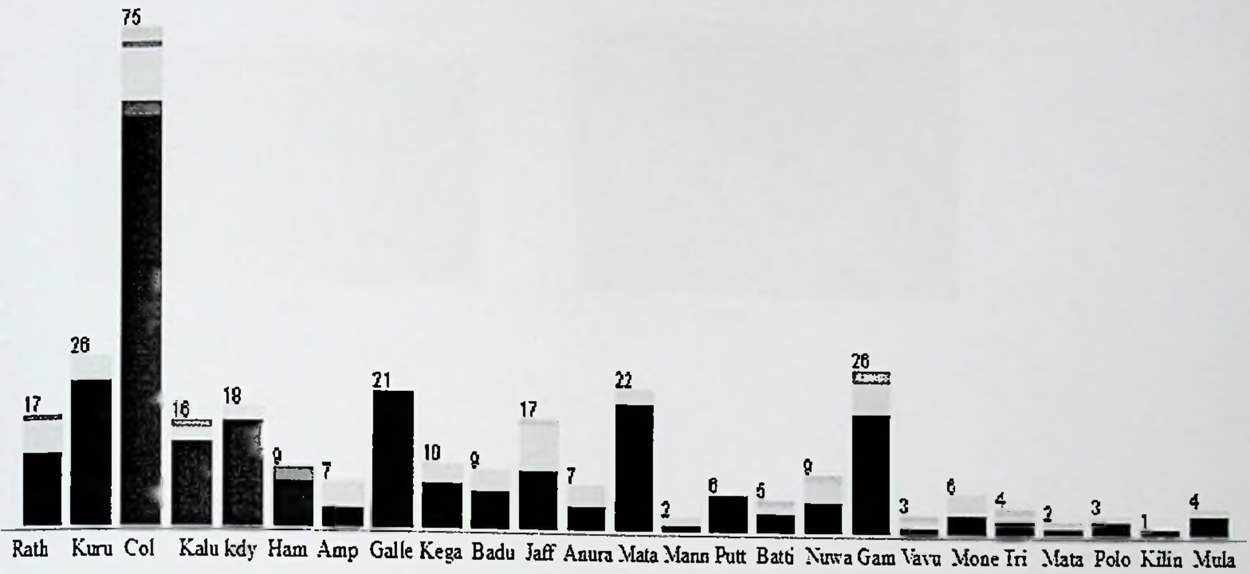


Figure 4. 10 District → results (B12+B13) - Subject 2

- All students from Galle, Puttalam, Kilinochchi reach to 'A' pass.

Subject 2 (B12)

Gender → Results

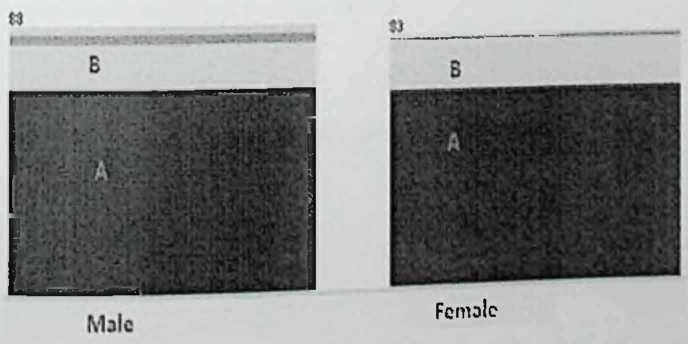


Figure 4. 11 Gender → results (B12) Subject 2

Subject 2 (B13)

Gender → Results

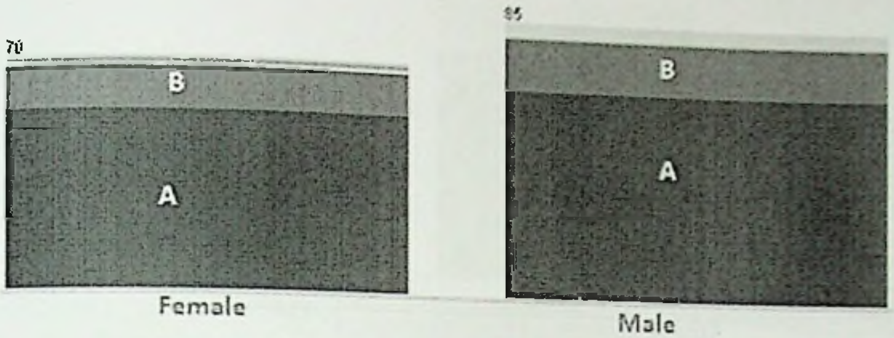


Figure 4. 12 Gender-> results B12 subject 2

Subject 2 (B12+B13)

Gender → Results

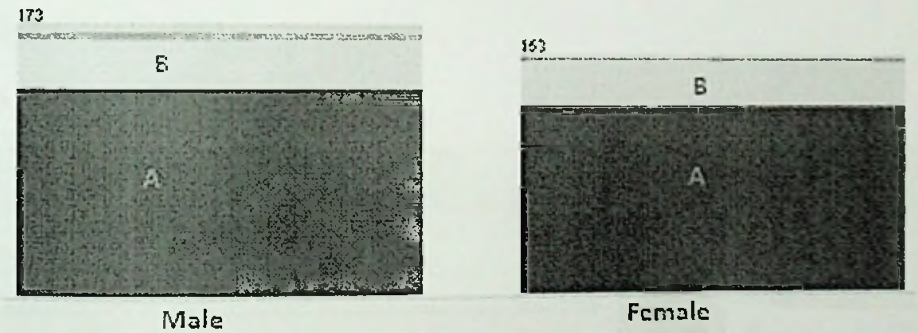


Figure 4. 13 Gender-> Results (both Batches) subject 2

- No significant influence form gender to results.

Subject 2 (B12+B13)

Both Batches

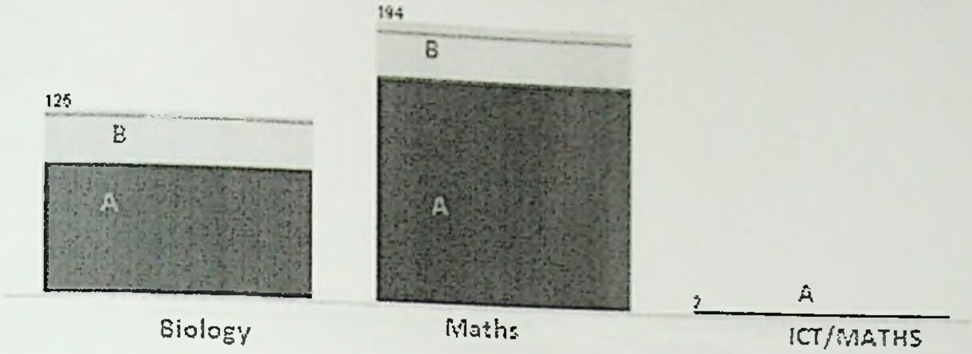


Figure 4. 14AL stream → results (B12+B13) Subject 2

- Most student from Biology stream score 'A' grade
- Most students from Maths stream score 'A' grade
- All students from ICT/Maths score 'A' grade

Subject 2 (B12+B13)

CA → Results

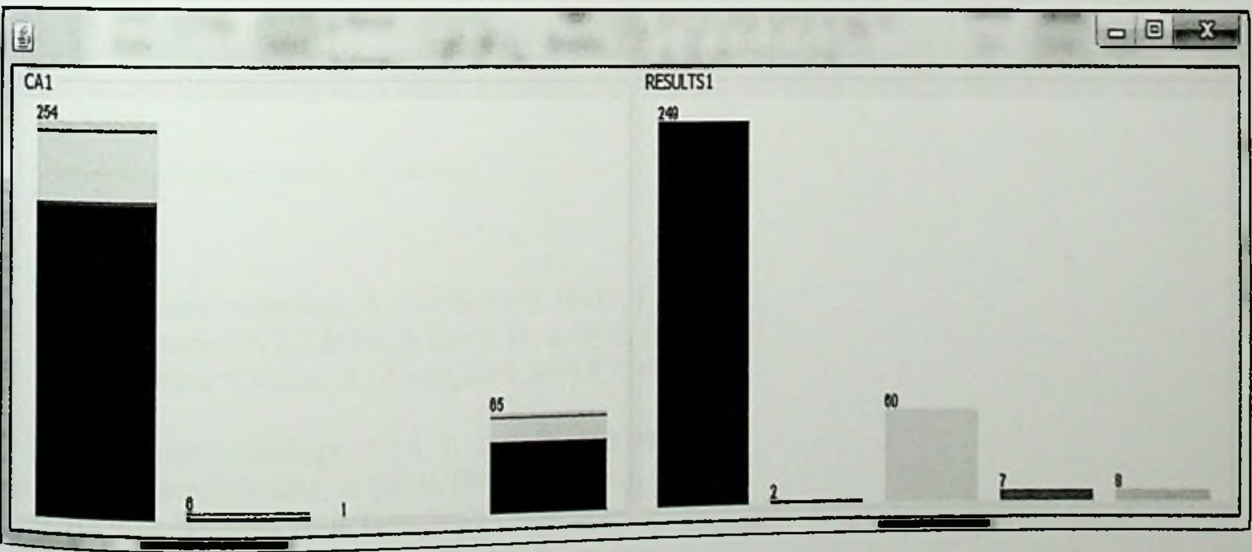


Figure 4. 15 CA → results (B12+B13) Subject 2

- 254 students have scored A grade for CA. There are 249 students score A grade at the final exam.
- 65 students have scored B grade for CA. There are 60 students score B grade at the final exam.

- In this subject, it has higher possibility of getting CA grade for the final grade.

4.2.10 Prediction of Results based on Two Attributes - IN 1310

Subject -2 B12

Decision Tree

AL, CA → Results

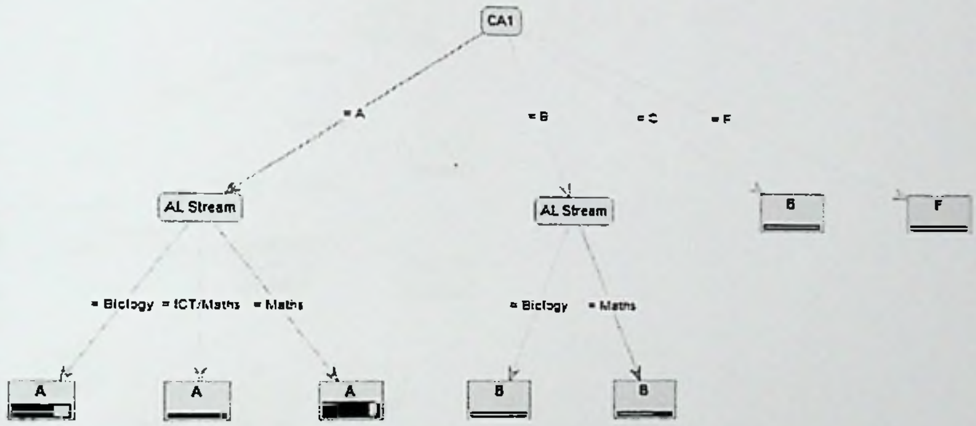


Figure 4. 16 Decision Tree CA, AL → Results

CA1 = A
| AL Stream = Biology: A {A=49, D=0, B=17, F=0, C=0}
| AL Stream = ICT/Maths: A {A=2, D=0, B=0, F=0, C=0}
| AL Stream = Maths: A {A=80, D=1, B=10, F=0, C=3}
CA1 = B
| AL Stream = Biology: B {A=0, D=0, B=1, F=0, C=0}
| AL Stream = Maths: B {A=0, D=0, B=2, F=0, C=1}
CA1 = C: B {A=0, D=0, B=1, F=0, C=0}
CA1 = F: F {A=0, D=0, B=0, F=4, C=0}

District, CA → Result

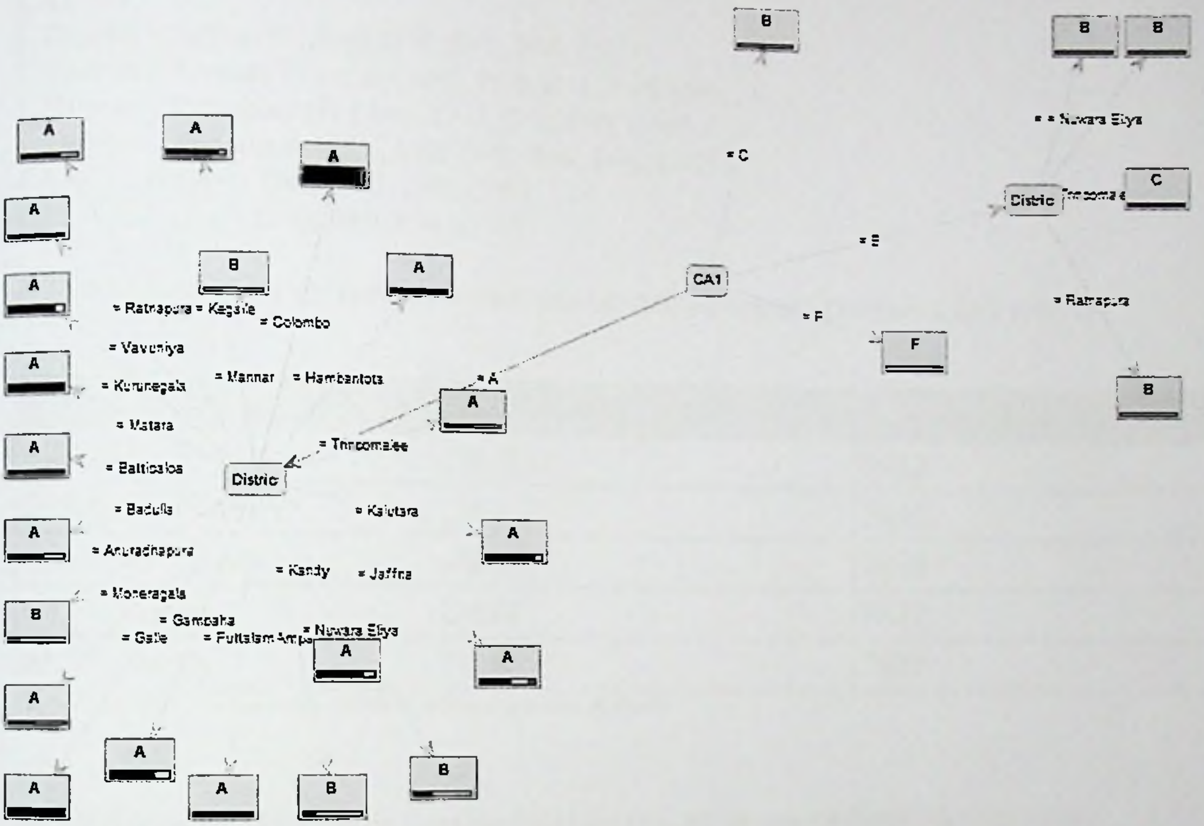


Figure 4. 17 Decision Tree CA, District → Results

CA1 = A
| District = Ampara: B {A=1, D=0, B=3, F=0, C=0}
| District = Anuradhapura: B {A=1, D=0, B=3, F=0, C=0}
| District = Badulla: A {A=4, D=0, B=2, F=0, C=0}
| District = Batticaloa: A {A=1, D=0, B=0, F=0, C=1}
| District = Colombo: A {A=35, D=1, B=1, F=0, C=0}
| District = Galle: A {A=10, D=0, B=0, F=0, C=0}
| District = Gampaha: A {A=10, D=0, B=3, F=0, C=0}
| District = Hambantota: A {A=4, D=0, B=0, F=0, C=0}
| District = Jaffna: A {A=4, D=0, B=2, F=0, C=1}
| District = Kalutara: A {A=8, D=0, B=1, F=0, C=0}
| District = Kandy: A {A=9, D=0, B=2, F=0, C=0}
| District = Kegalle: A {A=7, D=0, B=1, F=0, C=0}
| District = Kurunegala: A {A=13, D=0, B=2, F=0, C=0}
| District = Mannar: B {A=0, D=0, B=1, F=0, C=0}
| District = Matara: A {A=11, D=0, B=0, F=0, C=0}

| District = Moneragala: A {A=1, D=0, B=1, F=0, C=0}
| District = Nuwara Eliya: B {A=1, D=0, B=2, F=0, C=0}
| District = Puttalam: A {A=4, D=0, B=0, F=0, C=0}
| District = Ratnapura: A {A=5, D=0, B=2, F=0, C=0}
| District = Trincomalee: A {A=1, D=0, B=1, F=0, C=0}
| District = Vavuniya: A {A=1, D=0, B=0, F=0, C=1}
CA1 = B
| District = Jaffna: B {A=0, D=0, B=1, F=0, C=0}
| District = Nuwara Eliya: B {A=0, D=0, B=1, F=0, C=0}
| District = Ratnapura: B {A=0, D=0, B=1, F=0, C=0}
| District = Trincomalee: C {A=0, D=0, B=0, F=0, C=1}
CA1 = C: B {A=0, D=0, B=1, F=0, C=0}
CA1 = F: F {A=0, D=0, B=0, F=4, C=0}

4.2.11 Summary of Correctly classified rate for single attribute (Subject 2 B12 & B13)

Attribute	B-12 -correctly classified rate	B-13-correctly classified rate
Z-Score → Results	79.53	76.12
District → Results	76.6	76.12
Gender → Results	78.36	76.12
CA → Results	78.94	76.12
AL → Results	76.6	76.12

Table 4 1 summary of correctly classified instance rate-one attribute

4.2.12 Summary of Correctly classified rate for two attributes (Subject 2 B12 & B13)

Attribute	B-12 -correctly classified rate	B-13-correctly classified rate
Gender,CA → Results	80.71	76.13
District,CA → Results	76.6	76.13
Z-Score, CA → Results	80.11	7.13
AL,CA → Results	80.71	76.13

Table 4 2 summary of correctly classified instance rate-two attribute

4.2.13 Summary of Correctly classified rate for two attributes (Subject 2 B12 + B13)

Attribute	Both Batches -correctly classified rate
Gender,CA → Results	76.38
District,CA → Results	76.38
Z-Score, CA → Results	76.38
AL,CA → Results	76.38

Table 4 3 summary of correctly classified instance rate two attributes - both batches

Since CA & Advanced Level and Gender & CA (B12) score the heights correctly classified rate (80.71), CA , AL and Gender use for further classifications for subject 2 (B12+13)

4.2.14 Application of Model Subject 2(B12 + B13)

CA, AL, → Results

Total records 200

Missing records 50

Row No	RESULTS1	prediction(RESULTS1)	confidence(A)	confidence(D)	confidence(B)	confidence(F)	confidence(C)	Reg No	AL Stream	CA1
146	A	A	0.733	0	0.267	0	0		Biology	A
147	B	B	0	0	1	0	0		Biology	B
148	A	A	0.835	0.013	0.114	0	0.038		Maths	A
149	A	A	0.835	0.013	0.114	0	0.038		Maths	A
150	?	A	0.835	0.013	0.114	0	0.038		Maths	A
151	?	A	0.835	0.013	0.114	0	0.038		Maths	A
152	?	A	0.835	0.013	0.114	0	0.038		Maths	A
153	?	A	0.835	0.013	0.114	0	0.038		Maths	A
154	?	A	0.733	0	0.267	0	0		Biology	A
155	?	A	0.835	0.013	0.114	0	0.038		Maths	A
156	?	A	0.733	0	0.267	0	0		Biology	A
157	?	A	0.835	0.013	0.114	0	0.038		Maths	A
158	?	A	0.733	0	0.267	0	0		Biology	A
159	?	A	0.733	0	0.267	0	0		Biology	A
160	?	A	0.835	0.013	0.114	0	0.038		Maths	A
161	?	A	0.835	0.013	0.114	0	0.038		Maths	A
162	?	A	0.733	0	0.267	0	0		Biology	A
163	?	A	0.835	0.013	0.114	0	0.038		Maths	A
164	?	A	0.835	0.013	0.114	0	0.038		Maths	A
165	?	F	0	0	0	1	0		Maths	F

Table 4 4 Application of Model subject2

4.2.15 Validation of Model Subject 2(B12 + B13)

Row No	RES	prediction(RESULTS1)	confidence(A)	confidence(D)	confidence(B)	confidence(F)	confidence(C)	Reg No	AL Stream	CA1
1	A	A	0.742	0	0.253	0	0		Biology	A
2	A	A	0.851	0.011	0.105	0	0.032		Maths	A
3	A	A	0.851	0.011	0.105	0	0.032		Maths	A
4	D	A	0.851	0.011	0.105	0	0.032		Maths	A
5	A	A	0.851	0.011	0.105	0	0.032		Maths	A
6	B	A	0.742	0	0.258	0	0		Biology	A
7	A	A	0.742	0	0.255	0	0		Biology	A
8	B	A	0.851	0.011	0.105	0	0.032		Maths	A
9	A	A	0.851	0.011	0.105	0	0.032		Maths	A
10	A	A	0.851	0.011	0.105	0	0.032		Maths	A
11	B	A	0.851	0.011	0.105	0	0.032		Maths	A
12	A	A	0.851	0.011	0.105	0	0.032		Maths	A
13	A	A	0.851	0.011	0.105	0	0.032		Maths	A
14	A	A	0.851	0.011	0.105	0	0.032		Maths	A
15	A	A	0.851	0.011	0.105	0	0.032		Maths	A
16	A	A	0.851	0.011	0.105	0	0.032		Maths	A
17	A	A	0.851	0.011	0.105	0	0.032		Maths	A
18	B	A	0.742	0	0.258	0	0		Biology	A
19	B	A	0.742	0	0.258	0	0		Biology	A
20	B	A	0.851	0.011	0.105	0	0.032		Maths	A

Figure 4. 18validation of model subject 2

4.2.16 Calculating Accuracy subject 2(B12 + B13)

accuracy: 77.22% +/- 2.86% (mikro: 77.19%)

	true A	true D	true B	true F	true C	class precision
pred. A	131	1	31	3	4	77.06%
pred. D	0	0	0	0	0	0.00%
pred. B	0	0	0	0	0	0.00%
pred. F	0	0	0	1	0	100.00%
pred. C	0	0	0	0	0	0.00%
class recall	100.00%	0.00%	0.00%	25.00%	0.00%	

Figure 4. 19 Confusion Matrix subject 2

4.2.17 Prediction of Results based on one attribute - subject 1

Predictions on one Attribute

Z-Score → Results

Correctly Classified Instances	66	44.2953 %
Incorrectly Classified Instances	83	55.7047 %

District → Results

Correctly Classified Instances	66	44.2953 %
Incorrectly Classified Instances	83	55.7047 %

Gender → Results

Correctly Classified Instances	66	44.2953 %
Incorrectly Classified Instances	83	55.7047 %

AI → Results

Correctly Classified Instances	58	38.9262 %
Incorrectly Classified Instances	91	61.0738 %

CA → Results

Correctly Classified Instances	93	62.4161 %
Incorrectly Classified Instances	56	37.5839 %

Attribute	Correctly Classified Instance (%)
Z-Score → Results	44.29%
District → Results	44.29%
Gender → Results	44.29%
AI → Results	38.29%
CA → Results	62.41%

Table 4. 6 summary of correctly classified instance rate -one attribute

Predictions on two Attributes

Gender,CA → Results

Correctly Classified Instances	92	61.745 %
Incorrectly Classified Instances	57	38.255 %

District,CA → Results

Correctly Classified Instances	93	62.4161 %
Incorrectly Classified Instances	56	37.5839 %

Zscore,CA → Results

Correctly Classified Instances	93	62.4161 %
Incorrectly Classified Instances	56	37.5839 %

AL,CA → Results

Correctly Classified Instances	99	66.443 %
Incorrectly Classified Instances	50	33.557 %

Attribute	Correctly Classified Instance (%)
Gender,CA → Results	61.74%
District,CA→ Results	62.41%
Zscore,CA → Results	62.41%
Al,CA → Results	66.44%

Table 4. 7 summary of correctly classified instance rate -two attribute

4.2.18 Prediction of Results based on two attributes - subject 1

Gender,AL,CA → Results
Correctly Classified Instances 98 65.7718 %
Incorrectly Classified Instances 51 34.2282 %

Zscore,AL,CA
Correctly Classified Instances 99 66.443 %
Incorrectly Classified Instances 50 33.557 %

Gender,Zscore,District,AL,CA → Results
Correctly Classified Instances 93 62.4161 %
Incorrectly Classified Instances 56 37.5839 %

4.2.19 Prediction of Results based on three attributes - subject 1

Attribute	Correctly Classified Instance (%)
Gender,AL,CA → Results	65.77%
ZScore,AL,CA→ Results	66.44%
Gender,Zscore,District,AL,CA → Results	62.41%

Table 4. 8summary of correctly classified instance rate -two attribute subject1

When it consider the ratios of correctly classified instances, AL, CA → Results and AL,CA,Zscore → Results scores the highest. Therefore for AL, CA → Results will be used for further predictions.

4.3 Gender → Results Analysis

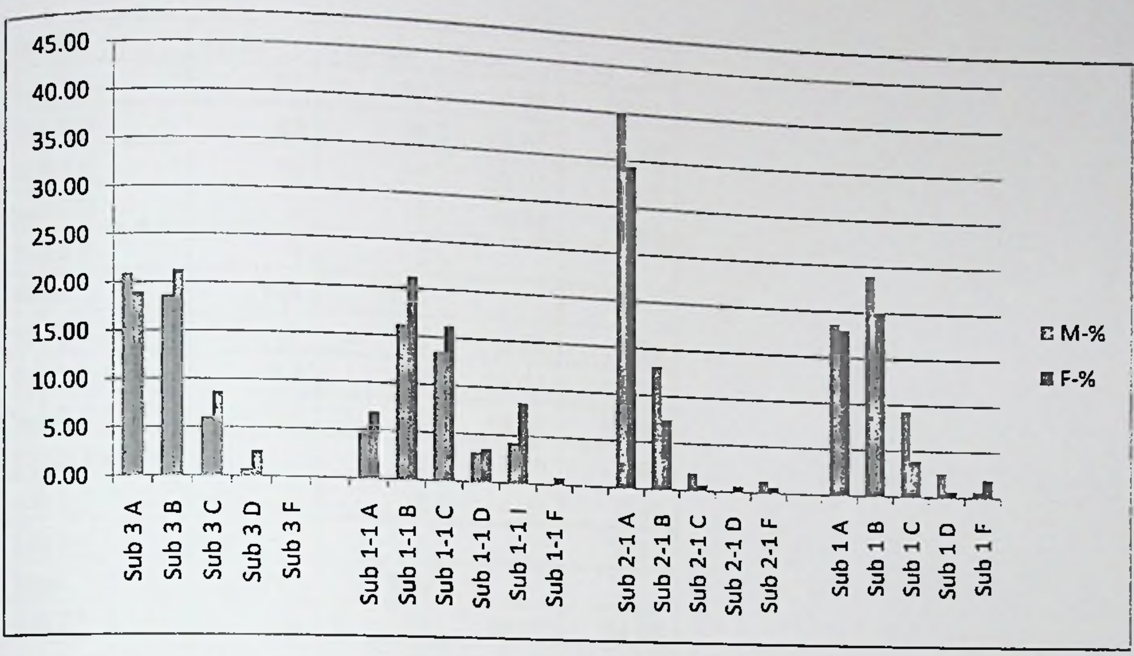


Figure 4. 20 Subject, Gender, Grade Distribution

In subject 1 more male student score A grades. Rest of the subject 1 and Subject 1-1 more female students score each grades. All the other subjects male students score higher grades than female student. In subject 1, rate of female students getting lower grades is higher than male students.

Gender - Grade Percentages for 5 Subjects

Grade	Male	Female	M-%	F-%
Sub 1 A	55	50	20.91	19.01
Sub 1 B	49	56	18.63	21.29
Sub 1 C	16	23	6.08	8.75
Sub 1 D	2	7	0.76	2.66
Sub 1 F	3	2		
Total	**125**	**138**		
Sub 1-1 A	12	18	4.63	6.95
Sub 1-1 B	42	55	16.22	21.24
Sub 1-1 C	35	42	13.51	16.22
Sub 1-1 D	8	9	3.09	3.47
Sub 1-1 I	11	22	4.25	8.49
Sub 1-1 F	3	2		0.77
Total	**111**	**148**		
Sub 2-1 A	64	55	40.00	34.38
Sub 2-1 B	21	12	13.13	7.50
Sub 2-1 C	3	1	1.88	0.63
Sub 2-1 D	0	1	0.00	0.63
Sub 2-1 F	2	1	1.25	0.63
Total	**90**	**70**		
Sub 3 A	28	27	18.67	18.00
Sub 3 B	36	30	24.00	20.00
Sub 3 C	14	6	9.33	4.00
Sub 3 D	4	1	2.67	0.67
Sub 3 F	1	3	0.67	2.00
Total	**83**	**67**		

Table 4. 9 subject, gender, grade distribution

4.4 Summary

This chapter covered the prediction of results based on one attribute, with two attributes and three attributes. The summaries of the correctly classified instant ratios were considered to select the attributes for predictions. Finally attributes which showed the heights correctly classified instant rate were selected for results prediction.

Chapter 5

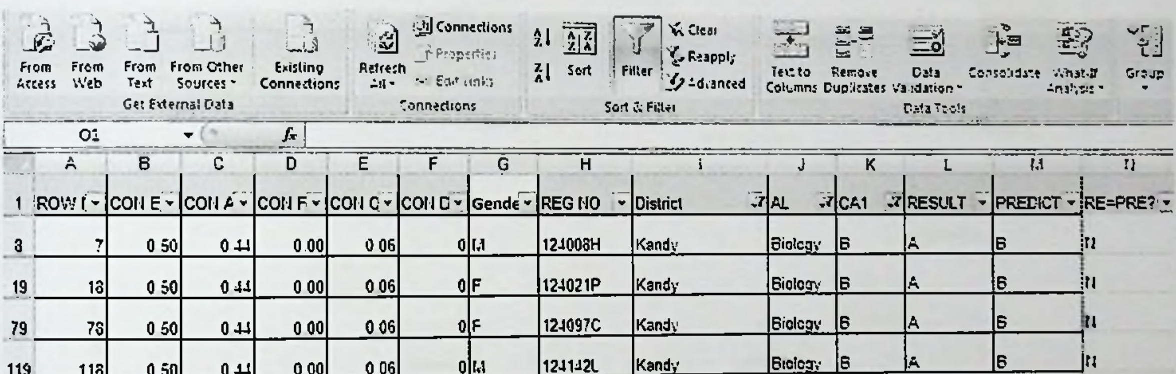
Design

5.1 Introduction

Chapter 4 presented the approach to develop and apply model for results prediction. This chapter elaborates the method of increasing the accuracy of prediction by introducing changes to the rules extracted from decision tree.

5.2 Introducing Changes to Rules

1. MODIFICATION TO RULES



	ROW I	CON E	CON A	CON F	CON C	CON D	Gende	REG NO	District	AL	CA1	RESULT	PREDICT	RE=PRE?
8	7	0.50	0.44	0.00	0.06	0	M	124008H	Kandy	Biology	B	A	B	N
19	18	0.50	0.44	0.00	0.06	0	F	124021P	Kandy	Biology	B	A	B	N
79	78	0.50	0.44	0.00	0.06	0	F	124097C	Kandy	Biology	B	A	B	N
119	118	0.50	0.44	0.00	0.06	0	M	124142L	Kandy	Biology	B	A	B	N

Table 5. 1 results prediction when CA=B and AL=Biology

IF CA=B ^AL=BILOGY^DISTRICT=KANDY

R →A

4 RECORDS CAN BE PREDICT CORRECTLY

0 INCORRECT RECORDS

- According to the decision tree, the predicted result is B
- This prediction is done based on the CA marks and AL stream.
- If more students':
- CA=B and AL=Biology and their final result =B, then the decision tree gives the predicted result as B.
- **IF CA='B' AND AL=BIOLOGY THEN RESULT = 'B'**
- But when it is considered with AL District: If CA=B and District=Kandy, then the decision tree predict results as 'A'

- **IF CA=B AND DISTRICT=KANDY THEN RESULT='A'**
- There fore we can over write the 1st rule with the second rule to get the correct prediction as follows

  IF CA='B' AND AL=BIOLOGY AND DISTRICT=KANDY THEN RESULT = 'A'

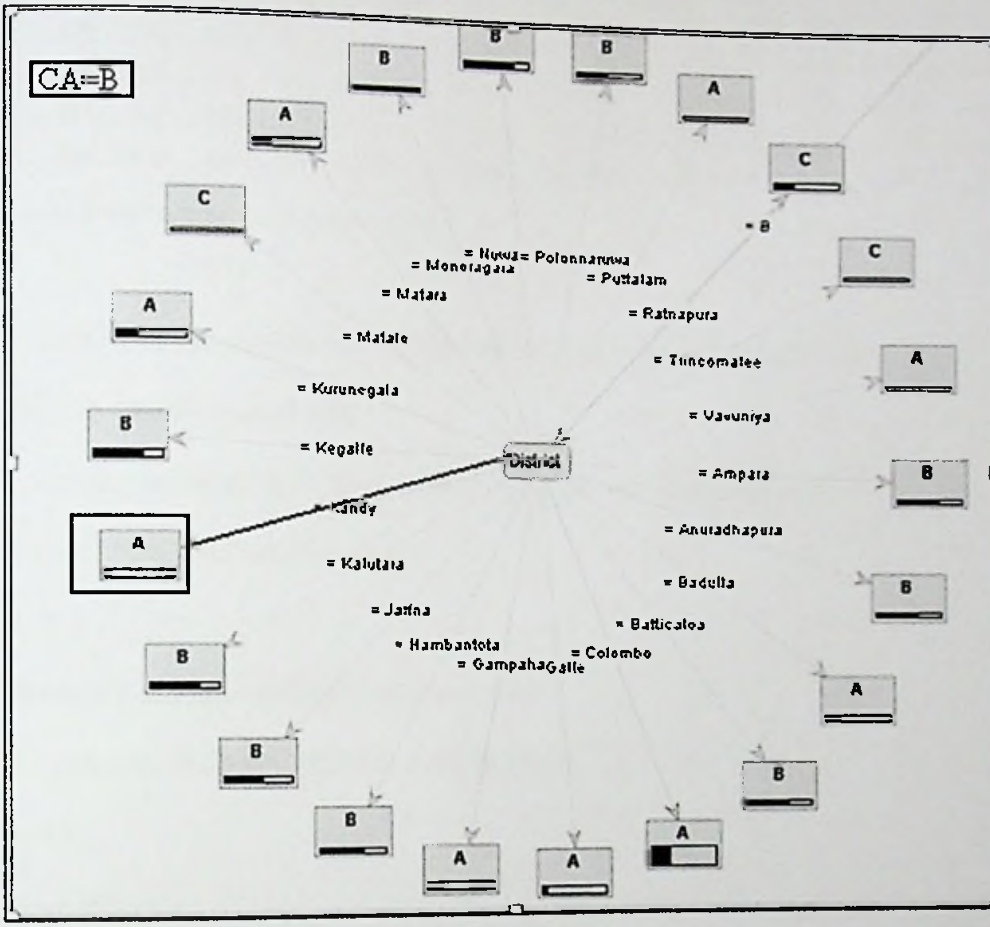


Figure 5. 1 result prediction when CA=B and District=Kandy

- From this way it is possible to get 4 correct predictions instead of 4 incorrect prediction.

	ROW1	COLE	COLA	COLF	COLC	COLC	Gende	REG NO	District	AL	CA1	RESULT	PREDICT	RE=PRE?	
8	7	0.50	0.44	0.00	0.05	0	M	[illegible]	Kandy	Biology	B	A	B	[illegible]	Y
19	18	0.50	0.44	0.00	0.05	0	F	[illegible]	Kandy	Biology	B	A	B	[illegible]	Y
79	78	0.50	0.44	0.00	0.06	0	F	[illegible]	Kandy	Biology	B	A	B	[illegible]	Y
119	118	0.50	0.44	0.00	0.06	0	M	[illegible]	Kandy	Biology	B	A	B	[illegible]	Y

Table 5. 2 result prediction when CA=B and AL=Biology and District=Kandy R=A

2. MODIFICATION TO RULES

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ROW I	CON E	CON A	CON F	CON C	CON D	Gende	REG NO	District	AL	CA1	RESULT	PREDICT	RE=PRE?
30	29	0.50	0.44	0.00	0.06	0	F	124032B	Kurunegala	Biology	B	A	B	no
77	76	0.50	0.44	0.00	0.06	0	M	124095T	Kurunegala	Biology	B	A	B	no
165	164	0.50	0.44	0.00	0.06	0	F	124196E	Kurunegala	Biology	B	A	B	no
244	243	0.50	0.44	0.00	0.06	0	M	125060P	Kurunegala	Biology	B	B	B	yes

Table 5. 3 results prediction when CA=B and AL=Biology

According to the decision tree of CA and AL stream result prediction is:

IF CA=B ^AL=BILOGY → R=B

According to the decision tree of CA and District result prediction is:

IF CA=B ^DISTRICT=KURUNEGALA

R-→ A

Therefore we can modify the rule as follows:

IF CA=B ^AL=BILOGY^DISTRICT=KURUNEGALA

R →A

From this rule

3 RECORDS CAN BE PREDICT CORRECTLY

1 INCORRECT RECORD

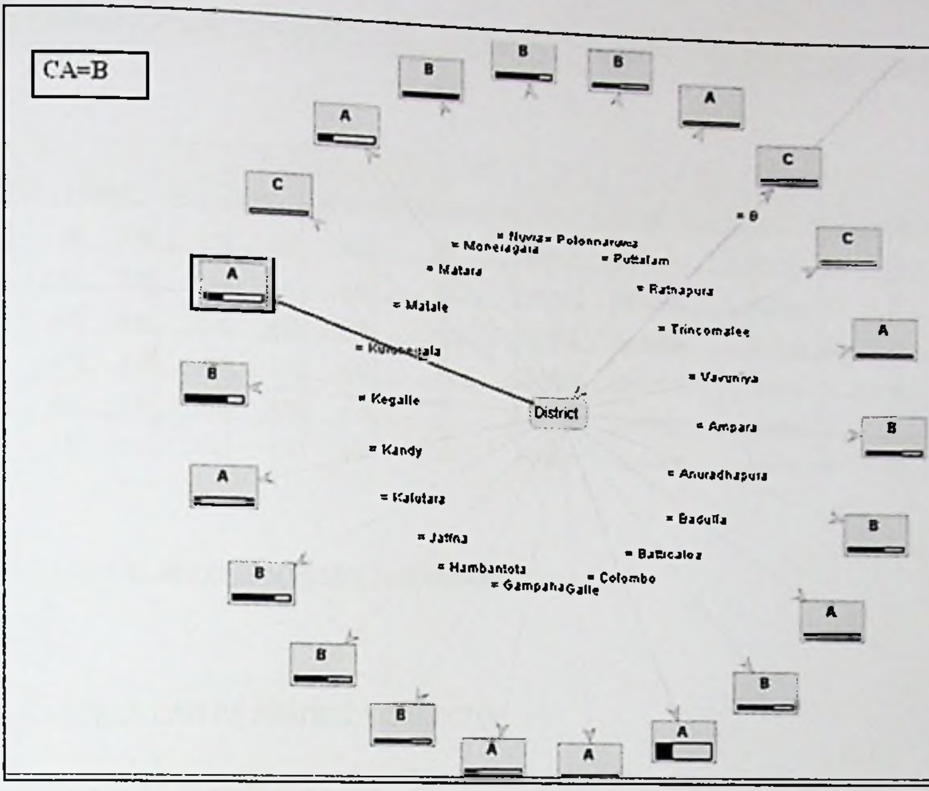


Figure 5. 2result prediction when CA=B and District=Kurunegala

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
1	ROW I	CON B	CON A	CON F	CON C	CON D	Gender	REG NO	District	AL	CA1	RESULT	PREDICT	RES PRED	
30	29	0.50	0.44	0.00	0.06	0	F	124032B	Kurunegala	Biology	B	A	B	N	Y
77	76	0.50	0.44	0.00	0.06	0	M	124095T	Kurunegala	Biology	B	A	B	N	Y
165	164	0.50	0.44	0.00	0.06	0	F	124156E	Kurunegala	Biology	B	A	B	N	Y
244	243	0.50	0.44	0.00	0.06	0	M	125063P	Kurunegala	Biology	B	B	B	Yes	N

Table 5. 4 result prediction when CA=B and AL=Biology and District=Kurunegala R=A

3 RECORDS PREDICT CORRECTLY

1 RECORD PREDICT INCORRECTLY

3. MODIFICATION TO RULES

A	B	C	D	E	F	G	H	I	J	K	L	M	N
ROW I	CON E	CON A	CON F	CON C	CON D	Gender	REG NO	District	AL	CA1	RESULT	PREDICT	RE=PRE?
52	0.50	0.44	0.00	0.06	0	F	124053U	Gampaha	Biology	B	B	B	yes
90	0.50	0.44	0.00	0.06	0	F	124112V	Gampaha	Biology	B	A	B	N
106	0.50	0.44	0.00	0.06	0	F	124130A	Gampaha	Biology	B	A	B	N
107	0.50	0.44	0.00	0.06	0	F	124131D	Gampaha	Biology	B	A	B	N
219	0.50	0.44	0.00	0.06	0	F	125035U	Gampaha	Biology	B	A	B	N
224	0.50	0.44	0.00	0.06	0	F	125040F	Gampaha	Biology	B	B	B	yes

IF CA=B ^AL=BILOGY^DISTRICT=GAMPAHA

R →A

4 RECORDS CAN BE PREDICT CORRECTLY

2 INCORRECT RECORDS

In this manner rules extracted from decision tree have been modified to increase the correctness and the accuracy of prediction. All such rules are included in Appendix A.

5.2 Summary

This section explained the introduction of changes and modifications for the rules extracted from different decision trees with the intension of increase the correct predictions and increase the accuracy of prediction.

Chapter 6

Implementation

6.1 Introduction

Chapter 5 presented the method of introducing changes to the rules extracted from different decision trees in order to increase the amount of correct predictions and increase the accuracy of predictions. This chapter describes the implementation of predictions using the modified rules with the aid of newly designed software.

6.2 Proposed System

6.2.1 Inputs for the proposed system

Excel Sheet with grades for continuous assessments.

6.2.2 Process

Prediction of final grade for a particular subject based on the predefined set of rules.

6.2.3 Output of the proposed system

Excel sheet with predicted final grade.

6.3 Prerequisites for the proposed system

Software Requirement:
It is required to installed jdk 1.6 (java development kit) or higher version for programming purposes
Advanced Installer to create exe. File
IDE Net Beans 6.1 or above
Microsoft Office package -MS Excel to upload and download input and output files

6.4 Proposed System Implementation

Executable file of this program can be saved to the computer and it creates a shortcut on the desktop. User can run the program by double clicking the shortcut icon on the desktop. Then the user can see the predicator interface which consist of following commands.

- Import
- Predict
- Export
- Clear

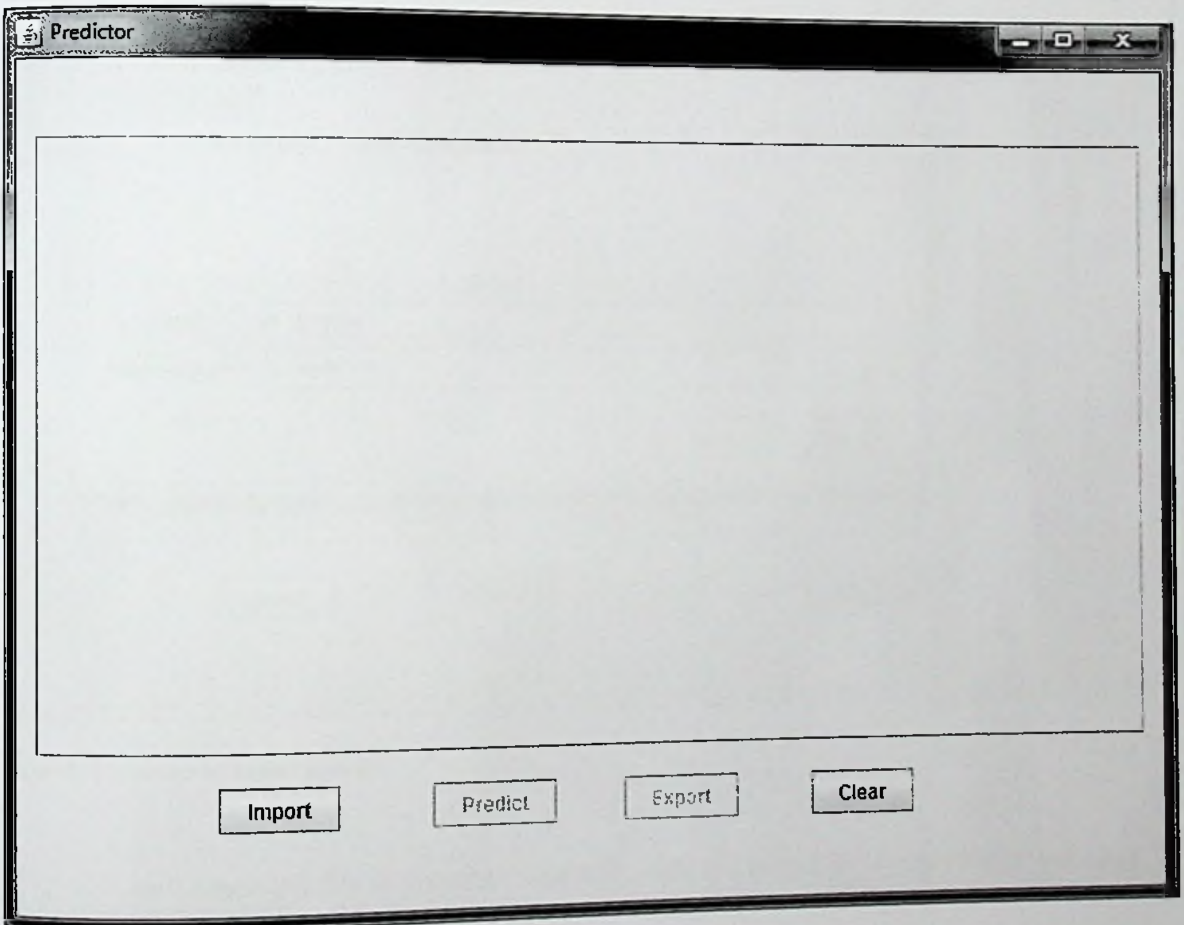


Figure 6. 1 Predicator Interface

- Import - user can import excel file which contain data use to predicted results.
- Predict - predict the result based on the records using predefined set of rules.
- Export- provide facilities with the user to save the predicted records as an excel file in local machine.
- Clear- clear data loaded onto the predicator window.

6.4 Results Prediction Process

Step 1

- Import the Excel file which contains records along with CA Grade of a particular subject.

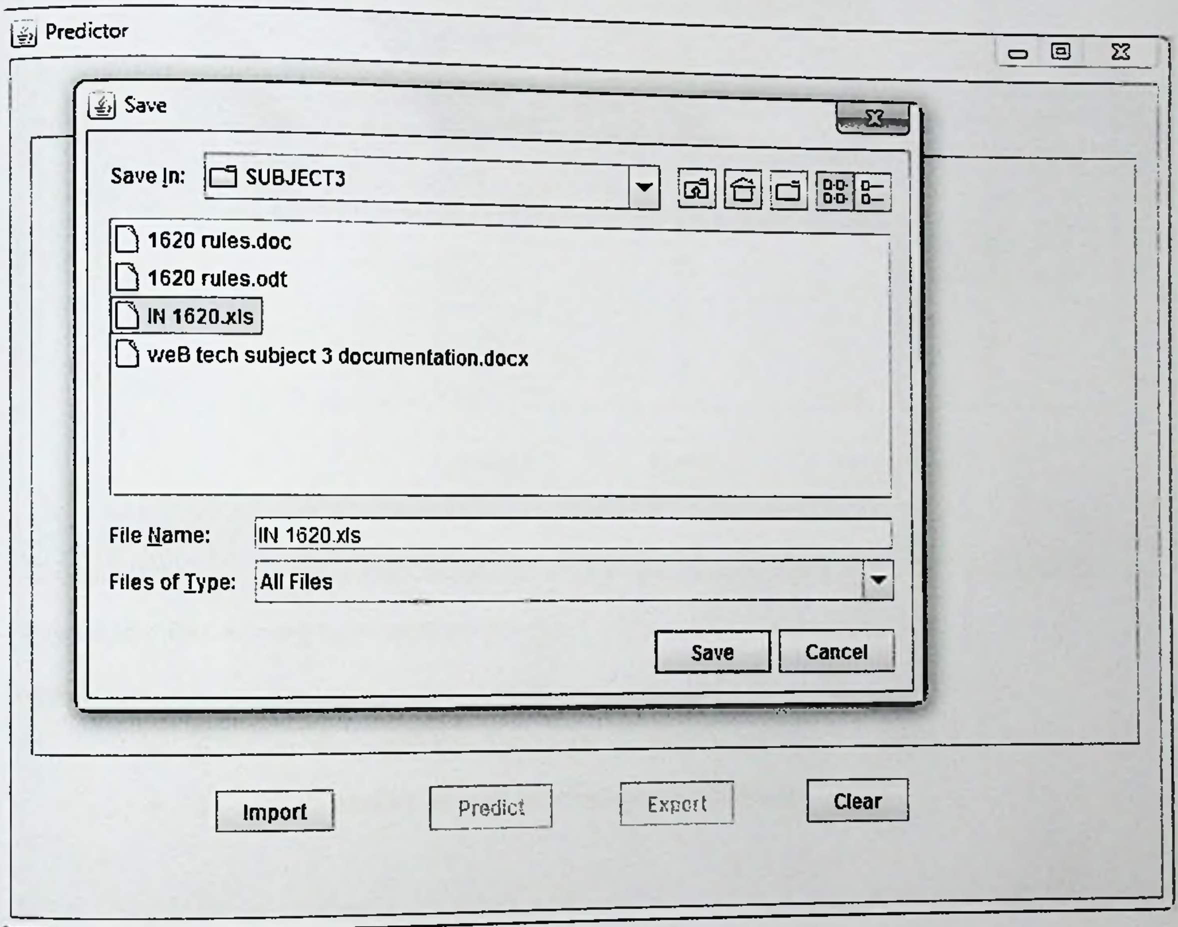
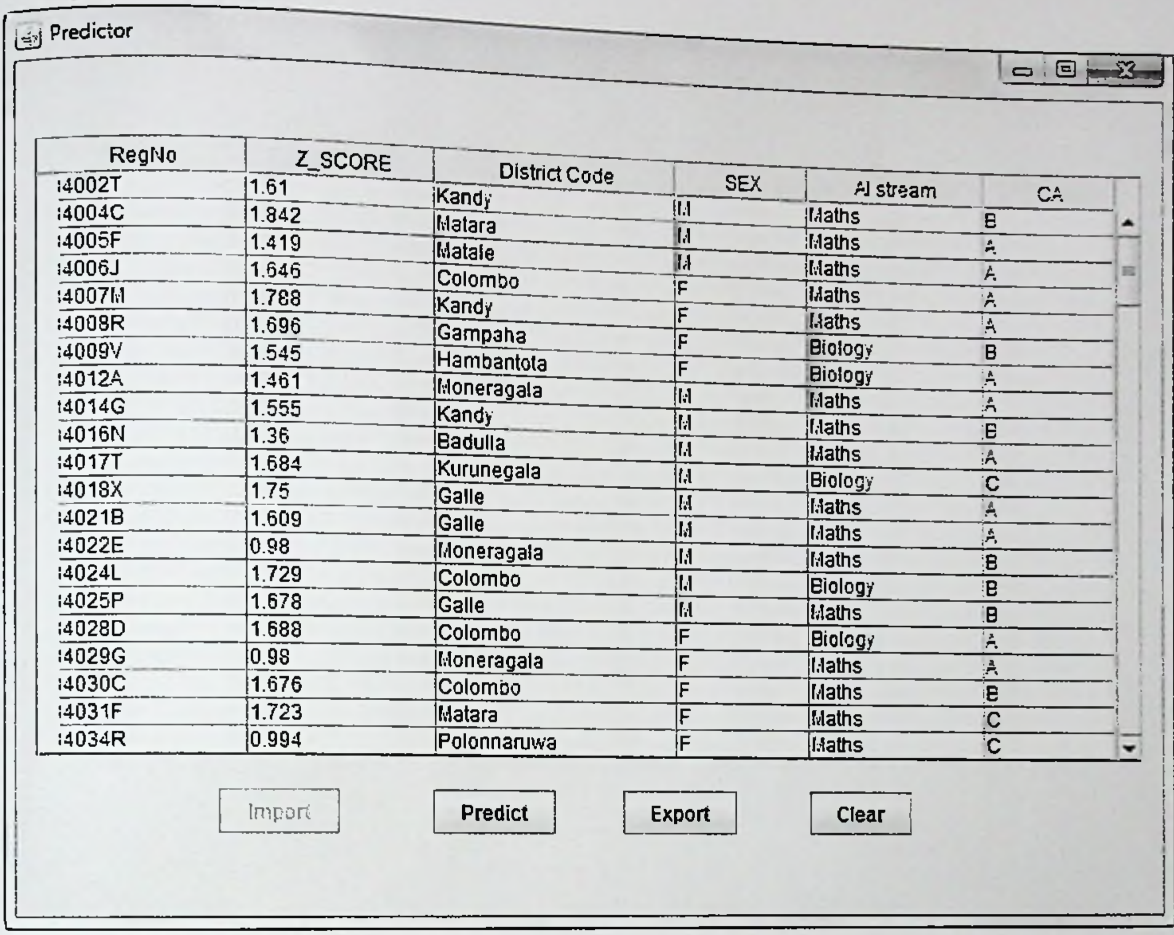


Figure 6. 2 Interface to Import data file

- Once the file is imported user will view the following window with imported records.

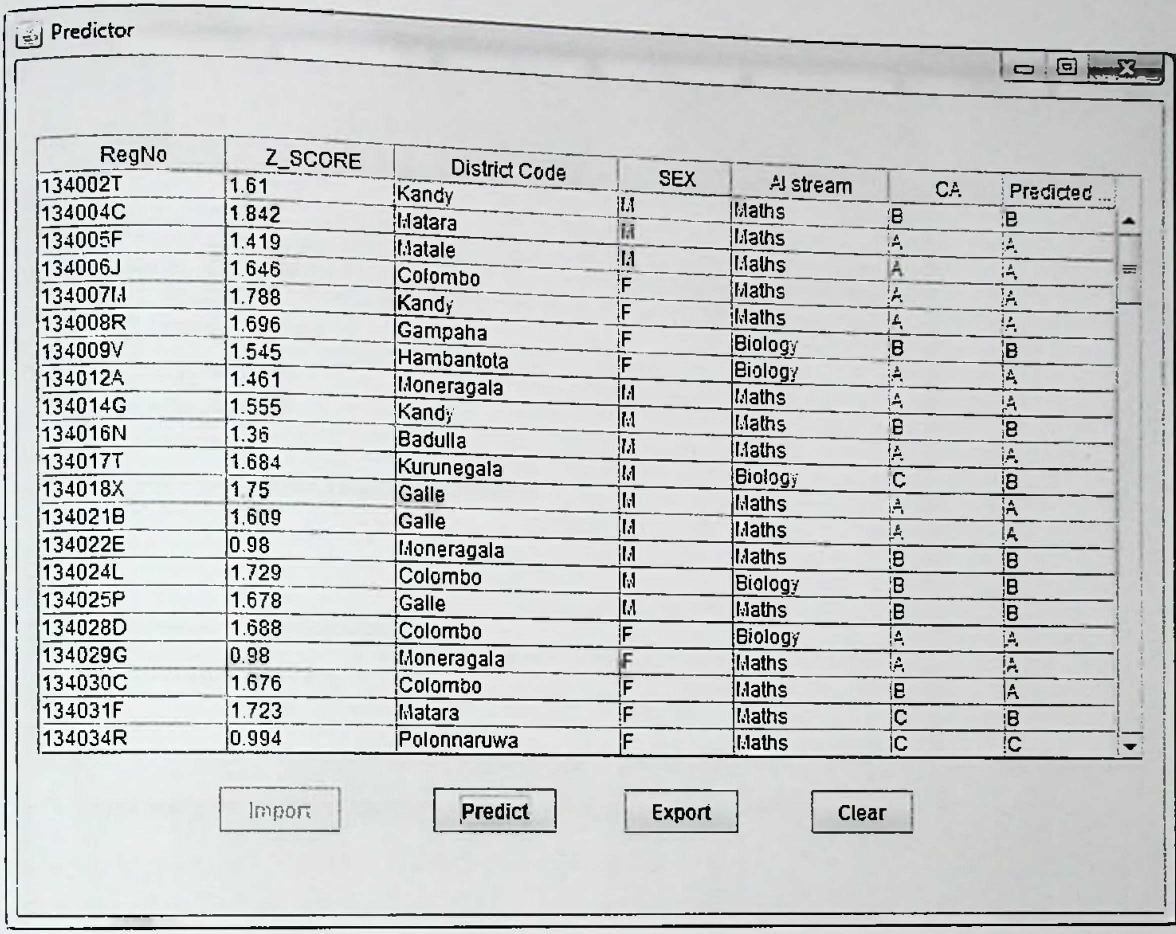


RegNo	Z_SCORE	District Code	SEX	Al stream	CA
4002T	1.61	Kandy	M	Maths	B
4004C	1.842	Matara	M	Maths	A
4005F	1.419	Matale	M	Maths	A
4006J	1.646	Colombo	F	Maths	A
4007M	1.788	Kandy	F	Maths	A
4008R	1.696	Gampaha	F	Biology	B
4009V	1.545	Hambantota	F	Biology	A
4012A	1.461	Moneragala	M	Maths	A
4014G	1.555	Kandy	M	Maths	B
4016N	1.36	Badulla	M	Maths	A
4017T	1.684	Kurunegala	M	Biology	C
4018X	1.75	Galle	M	Maths	A
4021B	1.609	Galle	M	Maths	A
4022E	0.98	Moneragala	M	Maths	B
4024L	1.729	Colombo	M	Biology	B
4025P	1.678	Galle	M	Maths	B
4028D	1.688	Colombo	F	Biology	A
4029G	0.98	Moneragala	F	Maths	A
4030C	1.676	Colombo	F	Maths	B
4031F	1.723	Matara	F	Maths	C
4034R	0.994	Polonnaruwa	F	Maths	C

Figure 6. 3 After data importing to the predicator interface

Step 2

- Click on the predict button to get the predictions for the result
- The result prediction will be displayed as follows:



RegNo	Z_SCORE	District Code	SEX	Al stream	CA	Predicted ...
134002T	1.61	Kandy	M	Maths	B	B
134004C	1.842	Matara	M	Maths	A	A
134005F	1.419	Matale	M	Maths	A	A
134006J	1.646	Colombo	F	Maths	A	A
134007M	1.788	Kandy	F	Maths	A	A
134008R	1.696	Gampaha	F	Biology	B	B
134009V	1.545	Hambantota	F	Biology	A	A
134012A	1.461	Moneragala	M	Maths	A	A
134014G	1.555	Kandy	M	Maths	B	B
134016N	1.36	Badulla	M	Maths	A	A
134017T	1.684	Kurunegala	M	Biology	C	B
134018X	1.75	Galle	M	Maths	A	A
134021B	1.609	Galle	M	Maths	A	A
134022E	0.98	Moneragala	M	Maths	B	B
134024L	1.729	Colombo	M	Biology	B	B
134025P	1.678	Galle	M	Maths	B	B
134028D	1.688	Colombo	F	Biology	A	A
134029G	0.98	Moneragala	F	Maths	A	A
134030C	1.676	Colombo	F	Maths	B	A
134031F	1.723	Matara	F	Maths	C	B
134034R	0.994	Polonnaruwa	F	Maths	C	C

Figure 6. 4 Predicted Results

Step 3

User has provided with the saving facility of these records as an Excel file in the local machine. User can select the export command and save the prediction results. Then user can make printouts if required.

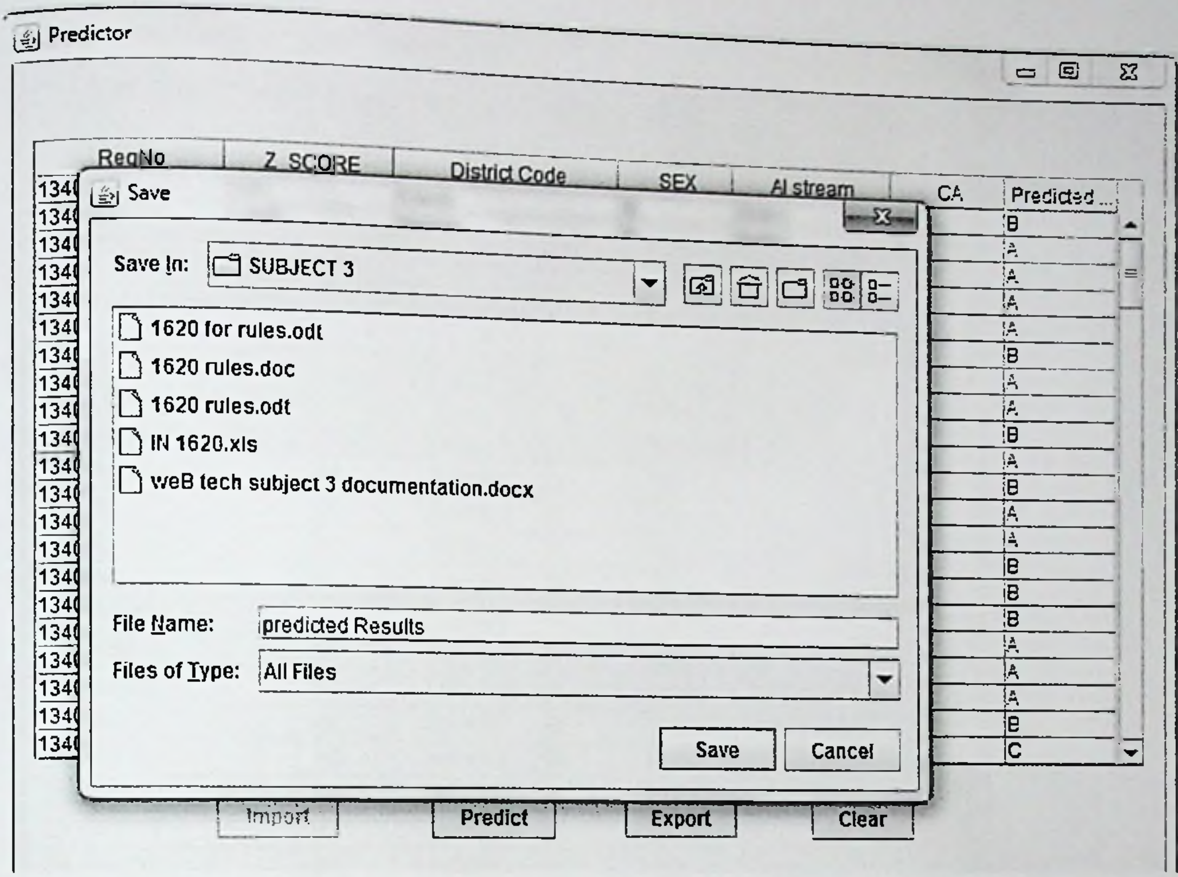


Figure 6. 5 Save predicted results to the local machine

- Open the Excel sheet saved in Local Machine

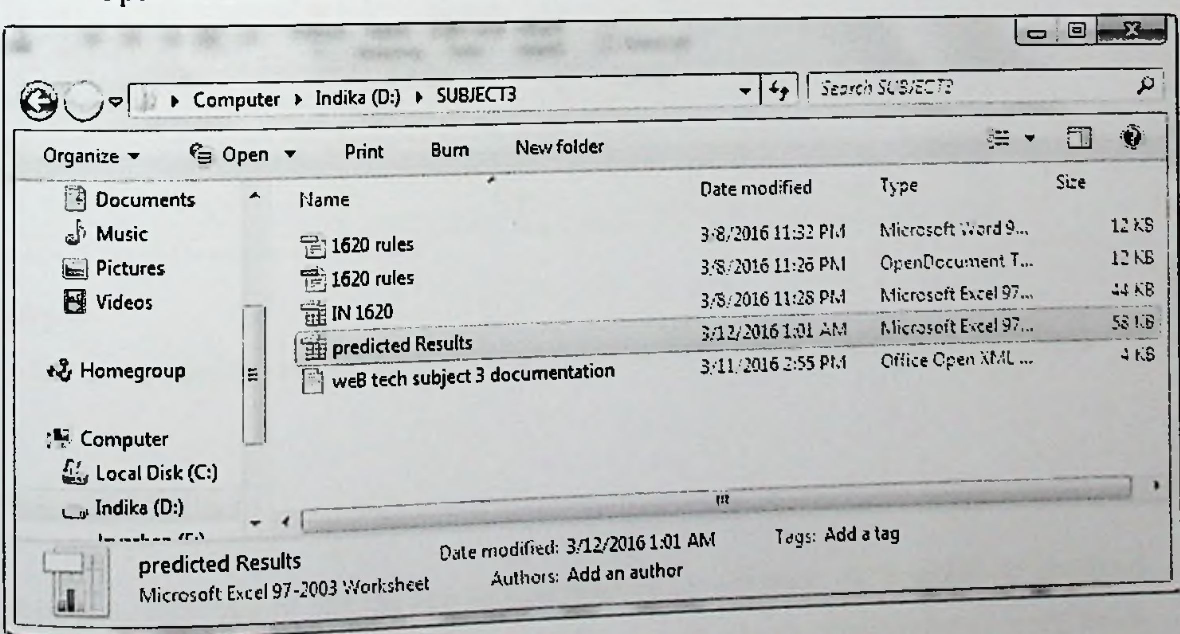
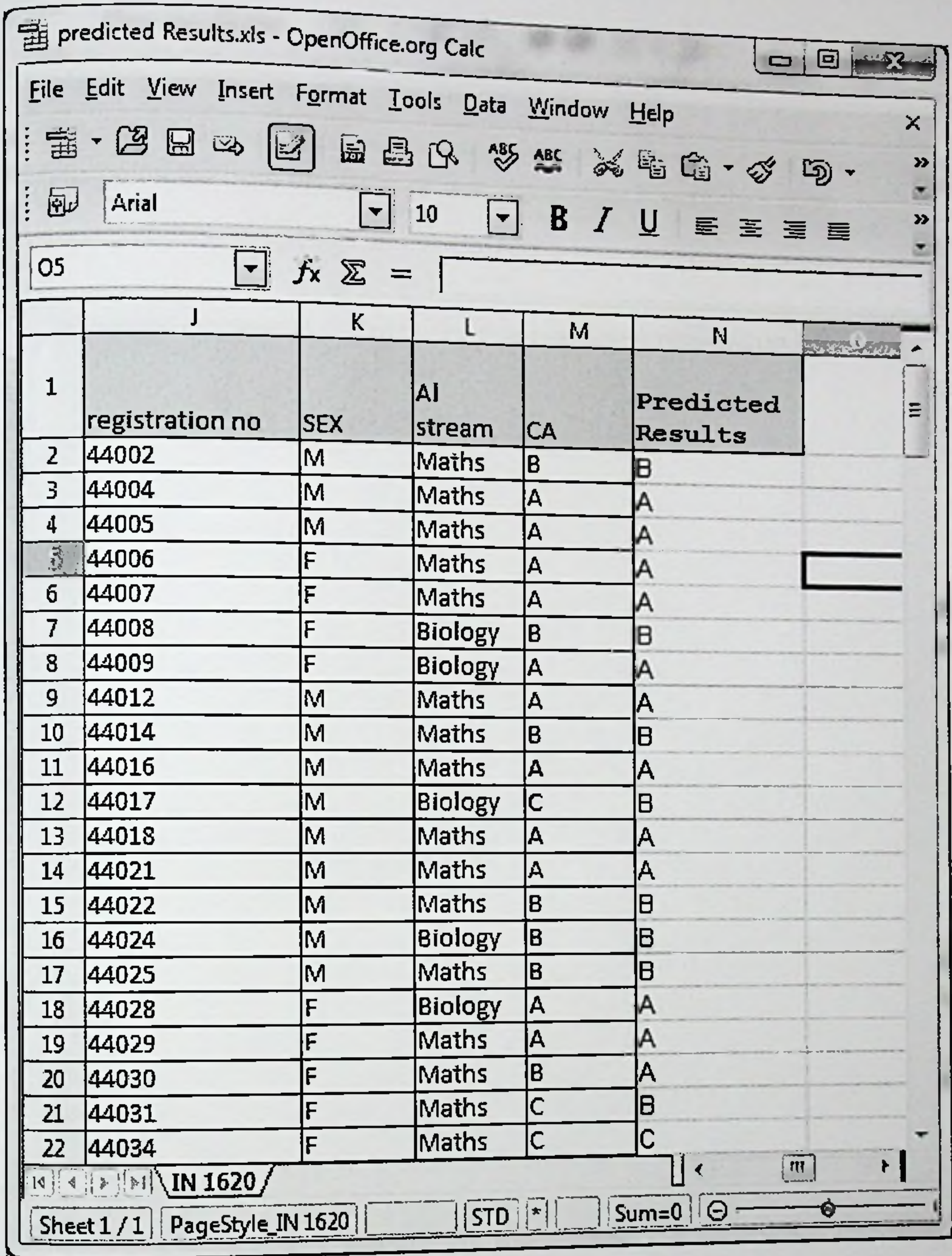


Figure 6. 6 open the excel file saved in the local machine



	J	K	L	M	N
1	registration no	SEX	Al stream	CA	Predicted Results
2	44002	M	Maths	B	B
3	44004	M	Maths	A	A
4	44005	M	Maths	A	A
5	44006	F	Maths	A	A
6	44007	F	Maths	A	A
7	44008	F	Biology	B	B
8	44009	F	Biology	A	A
9	44012	M	Maths	A	A
10	44014	M	Maths	B	B
11	44016	M	Maths	A	A
12	44017	M	Biology	C	B
13	44018	M	Maths	A	A
14	44021	M	Maths	A	A
15	44022	M	Maths	B	B
16	44024	M	Biology	B	B
17	44025	M	Maths	B	B
18	44028	F	Biology	A	A
19	44029	F	Maths	A	A
20	44030	F	Maths	B	A
21	44031	F	Maths	C	B
22	44034	F	Maths	C	C

Figure 6. 7 Excel file with Predicted Results

step 4

After completing the prediction process user can clear the screen with the clear command.

6.5 Testing

It is very important to test the results generated by the software. Each and every predicted result is tested with the predefined rules to check the correctness of the predicted result. Black Box testing method was used to test the results.

6.6 Summery

This chapter described the development and implementation of software to predict results. Results predicted from the software was tested using black box testing.

Chapter 7

Conclusion and Further Work

7.1 Introduction

Chapter 6 described the implementation of proposed software to predicts students' final result with respect to basic attributes identified. This Chapter discusses a summary of the study along with its concept and findings. Then it presents the recommendations deriving from the study.

7.2 Conclusion

This thesis discusses the application of data mining technique to students' data of Faculty of Information Technology to predict the final grade for a particular subject. This experiment provides a model to find out which students gain higher grades, average grades and marginal performances in advance. This prediction will help students to provide with early warning about their final examination result/grade for a selected subject. This will help to identify the students who need special attention to maintain their grades and reduce the fail rate. A timely appropriate warning to students at a risk could help prevent failing the exam or reducing a student getting marginal performance.

The developed methodology consist of following steps:

- Use decision tree techniques to build an understandable model including the inputs and class variables.
- Identification of classification rules among several factors and class variables.
- Introducing modification to classification rules to increase the accuracy of prediction and performance
- Develop a computer software with the identified classification rules to predict students final grade for a selected subject and its confidence.
- Using this new system it is possible to predict more than 75% accurate predictions.

7.3 Further work

This research can be extend to predict final result of subjects using some more attributes like his parents level of education, their family income, family background, the number of attempts of his Advanced Level etc.

This research can be extend to other post graduate programs to check their success or dropper base on their basic degree, age and some other personal attributes.

Further this kind of research can be extend to evaluate the performance of online or distance mode students performance in the final examinations and their successfulness of the degree program.

It is possible to use this type of prediction system for school examinations like GCE/OL and GCE/AL to predict students final results for particular subjects.

It is possible to conduct research using data mining techniques to predict the final result of the degree program using GPA statistics.

References

[1] R.Kohavi and T.Provost. “Confusion Matrix”. Internet :http://www2.cs.uregina.ca ~dbd/cs831/notes/confusion_matrix/confusion_matrix.html, 1998 [Jun. 22,2008]

[2] P.Smith. “Principals of data mining”. Internet:http://media.wiley.com/ product_data/excerpt/24/04712285/0471228524.pdf,2001 [Jan. 25, 2008].

[3] A. Costea, T. Eklund and J. Karlsson. “A Framework for Predictive Data Mining in the Telecommunications Sector”, in Intl Conf. WWW/Internet,2002,pp. 38-46.

[4] G. Holmes, R. Kirkby, and B. Pfahringer. “Mining data streams using option trees (revised edition, 2004).” Technical Report, The University of Waikato, Department of Computer Science, Hamilton, New Zealand, 2004.

[5] Two Crows Corporation. “Introduction to Data Mining and Knowledge Discovery,” in Data Mining Concept, 3rd ed, Vol. 1, Two Crows Corporation, Ed. USA: Potomac, 2005, pp. 1-2

[6] http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining

[7] http://onlinelibrary.wiley.com/doi/10.1002/widm.1075/abstract

[8] K. Sreeram and A. Murthy, ”Automatic construction of decision tree from data”, Siemens corporate research, United Kingdom,2004

[9] S. R. Safavian and D. Landgrebe,”A survey of decision tree classifier methodology”, M.A. thesis, Purdue University, United States, 2005

[10] G. P. Shapiro, “Discovery, analysis, and presentation of strong rules,” in G. P. Shapiro & W. J. Frawley, eds, ‘Knowledge Discovery in Databases’, AAAI/MIT Press, Cambridge, MA, 1991

[11] https://thenewstack.io/six-of-the-best-open-source-data-mining-tools/

[12] https://www.r-project.org/about.html

[13] B. Minaei-Bidgoli, D.A. Kashy, G. Kortmeyer and W.F.Punch "Predicting student performance: an application of data mining methods with an educational web based system" in Frontires in Education,2003 FIE 2003 33rd Annual 2003, vol. 1,p.T2A-13

[14] M.N. Quadri and Dr. N.V. Kalyankar-drop our feature of student data for academic performance using decision tree, Global Journal of Computer Science and Technology Vol.10 Issue 2(Ver 1.0), April 2010.

[15] Zlatko J, Kovacic, John Steven Green, Predictive working tool for early identification of 'at risk students', new Zealand.

[16] Anupama Kumar S. and Dr. Vijayalakshmi M.N. "Efficiency of Decision Tree in Predicting Student's Academic Performance" Banglore India 2011

[17] G. Naga Raja Prasad and Dr. A. Vinaya Bahu "Mining Previous Marks Data to Predict Students Performance in their Final Year Examination" International Journal of Engineering Research and Technology Vol.2, February 2013.

Appendix A

Set of Rules for Subject -3[c]

Accuracy=62 to 78%

1. AL=ARTS^CA=A
   PR=B^C(A)=0.00^C(B)=100.00^C(F)=0.00^C(C)=0.00^C(D)=0.00
2. AL=BIOLOGY^CA=A
   PR=A^C(A)=71.43^C(B)=21.43^C(F)=0.00^C(C)=7.14^C(D)=0.00
3. AL=BIOLOGY^CA=A^DISTRICT=ANURADHAPURA
   PR=B^C(A)=71.43^C(B)=21.43^C(F)=0.00^C(C)=7.14^C(D)=0.00
4. AL=BIOLOGY^CA=A^DISTRICT=KURUNEGALA^GENDER=F
   PR=B^C(A)=71.43^C(B)=21.43^C(F)=0.00^C(C)=7.14^C(D)=0.00
5. AL=ICT/MATHS^CA=A
   PR=B^C(A)=100.00^C(B)=00.00^C(F)=0.00^C(C)=0.00^C(D)=0.00
6. AL=/MATHS^CA=A
   PR=A^C(A)=93.33.00^C(B)=6.67^C(F)=0.00^C(C)=0.00^C(D)=0.00
7. AL=MATHS^CA=A^DISTRICT=BATTICOLOA
   PR=B^C(A)=93.33.00^C(B)=6.67^C(F)=0.00^C(C)=0.00^C(D)=0.00
8. AL=ARTS^CA=B
   PR=C^C(A)=14.29^C(B)=14.29^C(F)=0.00^C(C)=57.14^C(D)=14.29
9. AL=BIOLOGY^CA=B
   PR=B^C(A)=44.23^C(B)=50.00^C(F)=0.00^C(C)=5.77^C(D)=0.00
10. AL=BIOLOGY^CA=B^DISTRICT=KANDY
    PR=A^C(A)=44.23^C(B)=50.00^C(F)=0.00^C(C)=5.77^C(D)=0.00
11. AL=BIOLOGY^CA=B^DISTRICT=KURUNEGALA
    PR=A^C(A)=44.23^C(B)=50.00^C(F)=0.00^C(C)=5.77^C(D)=0.00
12. AL=BIOLOGY^CA=B^DISTRICT=GALLE
    PR=A^C(A)=44.23^C(B)=50.00^C(F)=0.00^C(C)=5.77^C(D)=0.00
13. AL=BIOLOGY^CA=B^DISTRICT=GAMPAHA
    PR=A^C(A)=44.23^C(B)=50.00^C(F)=0.00^C(C)=5.77^C(D)=0.00
14. AL=BIOLOGY^CA=B^DISTRICT=TRINCOMALEE
    PR=C^C(A)=44.23^C(B)=50.00^C(F)=0.00^C(C)=5.77^C(D)=0.00
15. AL=BIOLOGY^CA=B^DISTRICT=MATARA PR=A

^C(A)=44.23^C(B)=50.00^C(F)=0.00^C(C)=5.77^C(D)=0.00

16. AL=BIOLOGY^CA=B^DISTRICT=MATARA^GENDER=F
PR=A^C(A)=44.23^C(B)=50.00^C(F)=0.00^C(C)=5.77^C(D)=0.00
17. AL=BIOLOGY^CA=B^DISTRICT=PUTTALAM
PR=A^C(A)=44.23^C(B)=50.00^C(F)=0.00^C(C)=5.77^C(D)=0.00
18. AL=COMMERCE^CA=B
PR=B^C(A)=0.00^C(B)=40.00^C(F)=0.00^C(C)=40.00^C(D)=20.00
19. AL=COMMERCE^CA=B^DISTRICT=MATHALE
PR=C^C(A)=0.00^C(B)=40.00^C(F)=0.00^C(C)=40.00^C(D)=20.00
20. AL=ICT/MATHS^CA=B
PR=A^C(A)=100.00^C(B)=00.00^C(F)=0.00^C(C)=0.00^C(D)=0.00
21. AL=MATHS^CA=B
PR=A^C(A)=58.62^C(B)=32.76^C(F)=0.00^C(C)=8.62^C(D)=0.00
22. AL=MATHS^CA=B^DISTRICT=KEGALLE
PR=B^C(A)=58.62^C(B)=32.76^C(F)=0.00^C(C)=8.62^C(D)=0.00
23. AL=MATHS^CA=B^DISTRICT=NUWARAELIYA
PR=B^C(A)=58.62^C(B)=32.76^C(F)=0.00^C(C)=8.62^C(D)=0.00
24. AL=MATHS^CA=B^DISTRICT=KALUTHARA
PR=B^C(A)=58.62^C(B)=32.76^C(F)=0.00^C(C)=8.62^C(D)=0.00
25. AL=MATHS^CA=B^DISTRICT=BATTICOLOA
PR=B^C(A)=58.62^C(B)=32.76^C(F)=0.00^C(C)=8.62^C(D)=0.00
26. AL=MATHS^CA=B^DISTRICT=MONERAGALA
PR=B^C(A)=58.62^C(B)=32.76^C(F)=0.00^C(C)=8.62^C(D)=0.00
27. AL=MATHS^CA=B^DISTRICT=JAFFNA^GENDER=F
PR=B^C(A)=58.62^C(B)=32.76^C(F)=0.00^C(C)=8.62^C(D)=0.00
28. AL=ARTS^CA=C
PR=C^C(A)=0.00^C(B)=20.00^C(F)=0.00^C(C)=50.00^C(D)=30.00
29. AL=ARTS^CA=C^DISTRICT=KURUNEGALA
PR=B^C(A)=0.00^C(B)=20.00^C(F)=0.00^C(C)=50.00^C(D)=30.00
30. AL=ARTS^CA=C^DISTRICT=COLOMBO PR=D
^C(A)=0.00^C(B)=20.00^C(F)=0.00^C(C)=50.00^C(D)=30.00
31. AL=ATRS^CA=C^DISTRICT=TRINCOMALEE
PR=B^C(A)=0.00^C(B)=20.00^C(F)=0.00^C(C)=50.00^C(D)=30.00

32. AL=ARTS^CA=C^DISTRICT=HAMBANTHOTA
PR=D^C(A)=0.00^C(B)=20.00^C(F)=0.00^C(C)=50.00^C(D)=30.00

33. AL=BIOLOGY^CA=C
PR=B^C(A)=6.67^C(B)=70.00^C(F)=0.00^C(C)=20.00^C(D)=3.33

34. AL=BIOLOGY^CA=C^DISTRICT=BADULLA
PR=C^C(A)=6.67^C(B)=70.00^C(F)=0.00^C(C)=20.00^C(D)=3.33

35. AL=COMMERCE^CA=C
PR=C^C(A)=0.00^C(B)=00.00^C(F)=0.00^C(C)=0.00^C(D)=100.00

36. AL=MATHS^CA=C
PR=B^C(A)=16.13^C(B)=67.74^C(F)=3.23^C(C)=12.90^C(D)=0.00

37. AL=MATHS^CA=C^DISTRICT=VAVVUNIYA
PR=C^C(A)=16.13^C(B)=67.74^C(F)=3.23^C(C)=12.90^C(D)=0.00

38. AL=MATHS^CA=C^DISTRICT=RATHNAPURA^GENDER=M
PR=A^C(A)=16.13^C(B)=67.74^C(F)=3.23^C(C)=12.90^C(D)=0.00

Set of Rules for subject 2

Accuracy=75.36 to 80.32%

39. CA=A^ AL =Biology^ District=Badulla^ Gender =F → Results=B
^C(A)=0.72^C(B)=0.27^C(C)=0.00^C(D)=0.00^CON(F)=0.01

40. CA=A ^ AL=Biology ^ District=Anuradhapura ^ Gender=F →
Results=B^C(A)=0.72^C(B)=0.27^C(C)=0.00^C(D)=0.00^CON(F)=0.01

41. CA=A AL= Biology ^ District= Moneragala → Results=
B^C(A)=0.72^C(B)=0.27^C(C)=0.00^C(D)=0.00^CON(F)=0.01

42. CA=A AL=Biology → Results= A
^C(A)=0.72^C(B)=0.27^C(C)=0.00^C(D)=0.00^CON(F)=0.01

43. CA=A ^ AL=ICT/Maths ^ Results =
A^C(A)=1.00^C(B)=0.00^C(C)=0.00^C(D)=0.00^CON(F)=0.00

44. CA=A ^ AL= Maths ^ District= Ampara → Results=B ^
C(A)=0.83^C(B)=0.12^C(C)=0.03^C(D)=0.01^CON(F)=0.01

45. CA=A ^ AL=Maths ^ District=Anuradhapura →

Results=B^C(A)=0.83^C(B)=0.12^C(C)=0.03^C(D)=0.01^CON(F)=0.01

46. CA=A ^ AL=Maths ^ District=Mannar →
Results=B^C(A)=0.83^C(B)=0.12^C(C)=0.03^C(D)=0.01^CON(F)=0.01

47. CA=A ^ AL=Maths →
Results=A^C(A)=0.83^C(B)=0.12^C(C)=0.03^C(D)=0.01^CON(F)=0.01

48. CA=B ^ AL=Biology ^ District=Jaffna → Results=B^
C(A)=0.7^C(B)=0.2^C(C)=0.07^C(D)=0.00^CON(F)=0.03

49. CA=B ^ AL=Biology → Results=
A^C(A)=0.7^C(B)=0.2^C(C)=0.07^C(D)=0.00^CON(F)=0.03

50. CA=B ^ AL = Maths^ District= Jaffna → Results=B^
C(A)=0.68^C(B)=0.29^C(C)=0.03^C(D)=0.00^CON(F)=0.00

51. CA=B ^ AL=Maths ^ District=Vavuniya → Results=B^
C(A)=0.68^C(B)=0.29^C(C)=0.03^C(D)=0.00^CON(F)=0.00

52. CA=B ^ AL=Maths → Results= A
^C(A)=0.68^C(B)=0.29^C(C)=0.03^C(D)=0.00^CON(F)=0.00

53. CA=C → Results= B ^C(A)=0.00^C(B)=1.00^C(C)=0.00^C(D)=0.00^CON(F)=0.00

Set of Rules for subject 1[w]

Accuracy=62 to 78 %

54. CA=A ^ AL=Biology ^ District=Puttalam → Results=B^
^C(A)=0.7^C(B)=0.3^C(C)=0.00^C(D)=0.00^CON(F)=0.00

55. CA=A ^ AL= Biology ^ District=Rathnapura →
Results=B^C(A)=0.7^C(B)=0.3^C(C)=0.00^C(D)=0.00^CON(F)=0.00

56. CA=A ^ AL=Biology → Results=A ^
C(A)=0.7^C(B)=0.3^C(C)=0.00^C(D)=0.00^CON(F)=0.00

57. CA=A ^ AL=Maths → Results=A
^C(A)=0.68^C(B)=0.28^C(C)=0.04^C(D)=0.00^CON(F)=0.00

58. IF CA=B ^ AL=Biology ^ Disstrict=Ampara → Results=C
^C(A)=0.11^C(B)=0.74^C(C)=0.11^C(D)=0.05^CON(F)=0.00

59. IF CA=B ^ AL=Biology ^ Disstrict=Polonnaruwa → Results=C
^C(A)=0.11^C(B)=0.74^C(C)=0.11^C(D)=0.05^CON(F)=0.00

60. IF CA=B ^ AL=Maths ^ Disstrict=Colombo → Results=A
^C(A)=0.24^C(B)=0.66^C(C)=0.1^C(D)=0.00^CON(F)=0.00

61. IF CA=B ^ AL=Maths ^ Disstrict=Nuwa Eliya → Results=C
^C(A)=0.24^C(B)=0.66^C(C)=0.1^C(D)=0.00^CON(F)=0.00

62. CA=B ^ AL=Biology → Results=B ^
C(A)=0.11^C(B)=0.74^C(C)=0.11^C(D)=0.05^CON(F)=0.00

63. CA=B ^ AL=Maths → Results=B
^C(A)=0.24^C(B)=0.66^C(C)=0.1^C(D)=0.00^CON(F)=0.00

64. CA=C ^ AL=Biology ^ District=Colombo → Results=B
^C(A)=0.00^C(B)=0.38^C(C)=0.62^C(D)=0.00^CON(F)=0.00

65. CA=C ^ AL=Biology ^ District=Kurunegala → Results=B ^
C(A)=0.00^C(B)=0.38^C(C)=0.62^C(D)=0.00^CON(F)=0.00

66. CA=C ^ AL=Biology → Results=C ^
C(A)=0.00^C(B)=0.38^C(C)=0.62^C(D)=0.00^CON(F)=0.00

67. CA=C ^ AL=Maths ^ District=Mannar → Results=C
^C(A)=0.00^C(B)=0.47^C(C)=0.29^C(D)=0.24^CON(F)=0.00

68. CA=C ^ AL=Maths ^ District=Mulathivu → Results=C ^
C(A)=0.00^C(B)=0.47^C(C)=0.29^C(D)=0.24^CON(F)=0.00

69. CA=C ^ AL=Maths ^ District=Polonnaruwa → Results=C
^C(A)=0.00^C(B)=0.47^C(C)=0.29^C(D)=0.24^CON(F)=0.00

70. CA=C ^ AL=Maths Results=B
^C(A)=0.00^C(B)=0.47^C(C)=0.29^C(D)=0.24^CON(F)=0.00