

**ANALYSIS AND MINING OF EDUCATIONAL DATA
FOR PREDICTING THE STUDENT'S PERFORMANCE
OF INFORMATION AND COMMUNICATION
TECHNOLOGY SUBJECT IN GCE (O/L)
EXAMINATION**

NAME: P.S. SANDAMALI

INDEX NUMBER: 169332V

Degree of Master of Science/ Information Technology

Department of Information Technology

University of Moratuwa

Sri Lanka

September 2020

**ANALYSIS AND MINING OF EDUCATIONAL DATA
FOR PREDICTING THE STUDENT'S PERFORMANCE
OF INFORMATION AND COMMUNICATION
TECHNOLOGY SUBJECT IN GCE (O/L)
EXAMINATION**

NAME: P.S. SANDAMALI

INDEX NUMBER: 169332V

Thesis submitted in partial fulfillment of the requirements for the
Degree of Master of Science/ Information Technology

Department of Information Technology

University of Moratuwa

Sri Lanka

September 2020

DECLARATION

I declare this to be my own work. This dissertation does not incorporate material previously submitted to a degree or diploma from another University or higher education institution without approval, and as far as I know, does not contain any previous material published or created by someone else, except when a textual acknowledgment is given.

Name of student

Signature of student

P.S.Sandamali

.....

Date

Supervised by

Name of supervisor

Signature of supervisor

S.C.Premaratne

.....

Date

ACKNOWLEDGEMENT

To Mr.S.C.Premaratne, lecturer of the Faculty of Information Technology, University of Moratuwa and all the Professors, Lecturers of the Faculty of Information technology for providing the necessary guidance and guidance for this independent thesis....

To Mr.P.N.Illapperuma, Director of Education of the Information and Communication Technology Branch of the Ministry of Education and all the staff who assisted me in gathering studies and information.....

To Provincial, Zonal Information and Communication Technology Officers, Schools Principals, Teachers and Students who assisted in conducting the sample survey.....

To all my batch 10 brothers and sisters who supported me in various ways.

I thank sincerely.

ABSTRACT

Educational Data Mining (EDM) is find out interesting patterns and knowledge in educational organizations. This study is concerned with the student's performance of the GCE (O/L) ICT subject.

This study was investigated multiple factors affect student's performance of the GCE (O/L) ICT subject.

In this study, the classification method is used to predict the student's performance. ID3, K-NN and Naïve Bayes method are used here. Those methods were designed using Rapid miner tool.

After that qualitative model was generated and performance of student's based on personal, social and academic factors can be predicted using this.

According to this study, it may be used to predict the performance of students for the GCE (O/L) ICT subject and teachers can give special attentions and advise to students who need special attentions.

TABLE OF CONTENT

DECLARATION	I
ACKNOWLEDGEMENT	II
ABSTRACT	III
TABLE OF CONTENT	IV
LIST OF FIGURES	V
LIST OF TABLES	VI
LIST OF APPENDICES	VIII
CHAPTER 1: INTRODUCTION	1
1.1 Background	2
1.2 Statement of Research Problem	6
1.3 Aim and Specific Objectives	8
CHAPTER 2: LITERATURE REVIEW	9
2.1 Introduction	10
2.2 Review of Literature	10
CHAPTER 3: TECHNOLOGIES USED FOR THE STUDY	14
3.1 KDD Process	15
3.2 Data Mining	17
3.3 Popular Multi-Purpose Data Mining Tools That Are Leading the Trends	19
CHAPTER 4: METHODOLOGY	20
4.1 Problem Statement	21
4.2 Data Collection Techniques	21
4.3 Applying DM Classification Algorithms	24
4.4 Results and Evaluation	24
CHAPTER 5: RESULTS AND EVALUATION	25
CHAPTER 6: CONCLUSION AND RECOMENDATIONS	51
6.1 Conclusion	52
6.2 Recommendations	53
REFERENCES	54
APPENDICES	56
Appendix- A	56

LIST OF FIGURES

LIST OF FIGURES	Page
Figure 5.1 Collected data sheet	26
Figure 5.2 Least value and most value in the field	26
Figure 5.3.1 Histogram of Gender attribute	27
Figure 5.3.2 Histogram of Whsel attribute	27
Figure 5.3.3 Histogram of ReMaths attribute	28
Figure 5.3.4 Histogram of ReEnglish attribute	28
Figure 5.3.5 Histogram of ReICT attribute	29
Figure 5.3.6 Histogram of WhDiff attribute	29
Figure 5.3.7 Histogram of ParTution attribute	30
Figure 5.3.8 Histogram of PraKnow attribute	30
Figure 5.3.9 Histogram of ExRea attribute	31
Figure 5.3.10 Histogram of ReWrk attribute	31
Figure 5.3.11 Histogram of PraTest attribute	32
Figure 5.3.12 Histogram of ICTTea attribute	32
Figure 5.3.13 Histogram of FaIn attribute	33
Figure 5.3.14 Histogram of PaEdL attribute	33
Figure 5.3.15 Histogram of Lap/Com attribute	34
Figure 5.3.16 Histogram of IntFac attribute	34
Figure 5.3.17 Histogram of Hsup attribute	35

LIST OF TABLES

LIST OF TABLES	Page
Table 1.2.1 Number of students sat for the O/L ICT exam and passed rate Up to 2018.	7
Table 4.2.1 No of schools functioning O/L ICT subject in 2018	22
Table 4.2.2 No of students who studied ICT subject in 2018	23
Table 4.2.5.1 Attributes Description and Possible values	24-25
Table 5.3.1 Collected data group by gender	27
Table 5.3.2 Why did you select ICT subject for the O/L Examination	27
Table 5.3.3 What was the grade obtained for O/L Mathematics	28
Table 5.3.4 What was the grade obtained for O/L English?	28
Table 5.3.5 What was the grade obtained for O/L ICT?	29
Table 5.3.6 Why do you think that ICT paper is hard?	29
Table 5.3.7 Did you participate ICT tuition classes?	30
Table 5.3.8 Did you get practical training for the required topics?	30
Table 5.3.9 Did you get external readings in addition to using the textbook?	31
Table 5.3.10 Have you participated in the GCE O / L Examination Review Workshop on Information Technology?	31
Table 5.3.11 Can you increase your score if you have a practical test in addition to the written test?	32
Table 5.3.12 Is there an ICT teacher at the school?	32
Table 5.3.13 What is your monthly family income?	33
Table 5.3.14 What is the level of your parent's education?	33
Table 5.3.15 Have a computer/laptop in your home?	34
Table 5.3.16 Have an internet facility in your home?	34
Table 5.3.17 How to get support in the home to learn this subject?	35
Table 5.3.18 Confusion matrix of ID3 model	37
Table 5.3.19 Confusion matrix of K-NN model	38
Table 5.3.20 Confusion matrix of Naïve Bayes model	40
Table 5.3.21 Accuracy of the models	40
Table 5.3.22 Accuracy of the models when removing "Gender" attribute	41

Table 5.3.23 Accuracy of the models when removing “reason of selecting ICT subject” attribute	41
5.3.24 Accuracy of the models when removing "Mathematics result of the GCE (O/L)"	42
5.3.25 Accuracy of the models when removing "English result of the GCE (O/L)"	42
5.3.26 Accuracy of the models when removing” hard of the paper” attribute	43
5.3.27 Accuracy of the models when removing” participate tuition class” attribute	43
5.3.28 Accuracy of the models when removing” getting practical knowledge” Attribute	44
5.3.29 Accuracy of the models when removing” getting external reading” Attribute	44
5.3.30 Accuracy of the models when removing “participated review workshop” attribute	45
5.3.31 Accuracy of the models when removing “practical test” attribute	46
5.3.32 Accuracy of the models when removing “ICT teacher” attribute	46
5.3.33 Accuracy of the models when removing “Family monthly income” Attribute	47
5.3.34 Accuracy of the models when removing “parent education” attribute	47
5.3.35 Accuracy of the models when removing “Laptop/Computer” attribute	48
5.3.36 Accuracy of the models when removing “an internet facility in the home” attribute	48
5.3.37 Accuracy of the models when removing “Home support” attribute	49
5.3.38 Summary of results after removing each attributes in the data set	49

LIST OF APPENDICES

Appendices	Page
Appendix-A (Student Questioners) Analysis and Mining of Educational Data for Predicting the Student's Performance of Information & Communication Technology Subject In GCE (O/L) Examination	56-58