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

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

Computer Education in Sri Lanka began in 1982. At that time a selected number of schools received a computer each to teach basic programming language to GCE (A/L) science students. The first batch was a set of Sinclair microcomputers, and a television set is to be used as a monitor. From time to time lots of selected schools received computers. The last batch was given IBM XT compatible with a pair of 5 1/4 540KB disk drivers as secondary storage. The software used was WordStar 2000, Lotus 123, Dabase III+ and GW Basic.

The success of this programme was evaluated around 1992 and it was found to be almost a failure. The lack of motivation from the students because it was not an exam able subject, the teachers were not competent enough to teach the subject and the maintenance being a big problem were attributed to failure.

In 1994, a new concept emerged instead of providing computers to each school, a computer center per educational zone (cluster of schools) was given. These centers were called the Computer Resources Centers (CRC). A non-core curriculum was implemented, a separate staff of instructors was employed, and a special examination was conducted. There were two levels for this course namely the National Certificate in Computer Application basic level and intermediate level. The initial syllabus, which included introduction to Computer Hardware and Software, Operating Systems, Word Processing, Presentations, Spread Sheet, Databases and Programming. Students had to pay a nominal payment for a 120-hour course that was used to maintain the center.

The objective of this course was to provide basic computer skills for daily work, to give practical skills in basic computer applications, to give students a competitive edge in the labour market and to form a solid foundation for higher education in IT

The lack of physical resources was considered a major disadvantage of this program. A CRC has a laboratory of 10 computers and a lecture room with an overhead projector. The average computer / student ratio is 1: 3. This has reduced the hands-on experience of students, which is vital for them to acquire skills. Taking a course on computer applications can sometimes be a bit demanding for most naive learners. Simply learning in the form of a manual, an instruction manual or an instructor is not effective in many cases. Those who have access to computers have an advantage over those who do not.

For most students in CRCs, this was considered a luxury. Typically, students taking a limited course in application packages often struggle to achieve the expected mastery of course objectives. Sometimes they understand the functions of the application of the application software, but their conservation seems to be a problem. This is the position of the students in the CRCs. Their problem worsened because they had to wait a few months before their practical knowledge could be assessed. During this period, no access to a computer would have been possible. Later, a directive from the Ministry of Education (MoE) removed this review, but the CRCs continue to operate.

The national policy on ICT education in schools was prepared for fully utilize and improve ICT in Education system in 2001. A remarkable quantitative increase in ICT education was initiated in Sri Lanka in 2002. The Ministry of Education had launched a programme to establish ICT laboratories in the school system and the fund was granted by the Asian Development Bank and the World Bank. General Education Project II (GEP II) was funded by World Bank in 2001-2004, As a result of that 400 ICT centres were established in schools. The Asian Development Bank (ADB) was funded to secondary education modernization project (SEMP) it was established 1006 Computer Learning centres (CLCs) in 2006. It was extended SEMP II project. The CRC program has not been expanded. The purpose of so called project was to encourage ICT as a subject as well as it will be used as a tool in the teaching and learning process. Although ICT resources were distributed quickly, the professional development programme lagged behind. As a result, ICT resources have been underutilized for some time. Subsequently, professional development programs were conducted by both the public and private sectors. The CAL materials provided by the Ministry of Education were not sufficient to be used in the learning and teaching process.

The Ministry of Education continued to provide schools with ICT resources, including free internet access. The internet connection was under the control of the School net managed at the University of Moratuwa. School net provided a platform for distributing CAL materials and hosted school websites. However, despite the considerable investments made by the government in ICT resources, teacher training and internet infrastructure, the effective use of these facilities has not reached a satisfactory level.

In laptops were presented to Sri Lankan education under 'One Laptop per Child project (OLPC)' for World Bank-funded primary schools and PC classmate provided by Intel as part of the EKSP project. Recently introduced projects such as "development of 1,000 secondary schools “and” the nearest school is the best school” as well. The objectives of these projects to the possibility of using ICTs as a tool and subject too. However, these projects were pilot and limited to a few schools. Guaranteeing 13 years of Education program was introduced in 2017 for the selected schools.

ICT Education in school, Sri Lanka

The Information Technology industry was transmitted rapidly in 2001, As a result of that, there was a need to introduce Information Technology to the education system. Accordingly, they have focused on the following.

- The education system should be developed to fulfil the requirement beyond 21st century’s challenges and competitions. For that the most effective and important tool is information technology.
- Accordingly, it is evident that every student in the school system should be provided with some degree of information learning.
- However, as the field of ICT is rapidly evolving, it is controversial to determine the level of learning in the subject of ICT for each grade in the school system but it should be made available to students of all levels of technology as much as possible.
- ICT education or ICT education is used in many countries.
- Decide how much ICT education should be provided to the school system. Students should consider the job market, tertiary and higher education opportunities.

General Information Technology (GIT) Grade 12 has been introduced in 2004 to focus on school leavers. It focuses on providing students with basic theoretical knowledge on the application and application of Information Technology in the field of ICT rather than use of ICT in higher education.

In addition, Information & Communication Technology in GCE (O/L) syllabus in the year 2006 was also given as a technical subject. The subject of Information &

Communication Technology has been introduced to the GCE (A/L) syllabus in 2009. Three ICT subjects as such web design, software development and graphic design have been introduced according to the “Guaranteeing 13 years of Education program” in 2017. Information & communication Technology in 6-9 syllabus in the year 2018 was introduced as an optional subject.

A number of teacher training programmes was conducted by Ministry of Education since 2006 to date but no appropriate data collection method to maintain the progress. Therefore, it is unable to take present status of the teacher training programme.

Although the funds were provided by the government or foreign organization such projects, it seems that Sri Lankan ICT education system is still behind while compare with the other countries. There are number of modules functioning for ICT as such O/L, A/L, and GIT in the Sri Lankan General Education System. Among that most successful ICT module is O/L ICT in Sri Lanka.

Objectives of the GCE (O/L) ICT curriculum

The main objectives of introducing ICT for GCE (O/L) are as follows:

(i) Improve basic computer literacy and to develop a foundation for further studies in ICT.

(ii) Develop understanding of ICT applications and their practices.

(iii) Develop the concepts and principles related to ICT.

(iv) Improve the skills for ICT based solutions in practically.

(v) Understand advantage and disadvantage of usage of ICT.

The ICT resources in the school are limited, therefore limited number of students are selected for this subject. Student were selected according to marks of Mathematics & English paper in Grade 9 or as well as another evaluation criteria was created by the school. However different methodologies were applied by schools. The ICT subject can be taught in three major languages (Sinhala, Tamil, and English) but the technical terms are taught in English. Eventually examination language can be selected by the student as prefer. ICT Curriculum for O/L has partially contained in the UNESCO literacy model. It seems that the main section of ICT literacy model in UNESCO was included social, ethical aspects, whereas in Sri Lankan aspects was ICT and society. Further, the Sri Lankan curriculum consists of Programming, web development, systems, and a Group project which are not included in the UNESCO literacy model. In Sri Lanka, flexible selection criteria were proposed to the student to select ICT subjects.

1.2 STATEMENT OF RESEARCH PROBLEM

After the introducing ICT subject for the school system in 2004, Ministry of Education has taken several actions to improve the ICT subject.

- Provided physical resources for the ICT laboratories.
- Recruited the ICT teachers.
- Continuous teacher training programme for ICT teachers.
- Changed the syllabus whenever necessary.
- Conducted students' seminars.
- Conducted remedial teacher training for lower performance schools.

When introducing ICT subject, there were not ICT appointed teachers. So English, Mathematics and Science teachers who have some ICT skills were appointed as an ICT teacher. Apart from that Ministry of Education (MoE) and National Institute of Education (NIE) have taken necessary action to give further knowledge to those teachers.

After that introduce ICT subject for the National Colleges of Education (NCoE) and produced ICT teachers using those places. So, those teachers were placed necessary schools by the Ministry of Education. But at that time there was a huge demand on ICT subject, so further taken service of previous appointed ICT teachers. After some years, graduate ICT teachers were recruited to the schools. The purpose for this recruiting is promoting A/L ICT subject, but their time table was adjusting and took their service in O/L ICT subject as possible. According to that, there were lot of subject expertise in the Education system now.

Students studied ICT subject first time in 2006 and those students sat for the O/L ICT paper on 2008.

Table 1.2.1 Number of students sat for the O/L ICT exam and passed rate up to 2018.

Year	Number Sat	Number Passed	%
2008	38759	33930	87.54
2009	40413	35081	82.13
2010	41490	36645	82.33
2011	39797	33875	71.39
2012	36437	30140	82.62
2013	35869	32652	91.03
2014	35032	31693	90.47
2015	43781	40653	92.86
2016	51144	46708	91.33
2017	52014	50226	96.56
2018	54677	52882	96.72

Data Source: Department of Examination report

Considering above facts, I decided my topic. Apart from that there is a great need at present to recognize the aspects that are effected to the students' performance and to introduce new methods accordingly. According to that my research is valuable for below partners.

- Ministry of Education, Provincial Department of Education and Zonal Education Office
- National Institute of Education
- Department of Examination
- Department of Publication
- Schools which are functioning O/L ICT subject
- Students who are selecting this subject in future

When taking action for below activities, it is good to pay attention on my research results.

- Policy decision related to ICT subjects.
- Revision of ICT subject
- Preparation of exam paper on ICT subject
- Preparation of ICT reading book

- Check whether continue current method or not. If decide to change then how to do it.

1.3 AIMS AND SPECIFIC OJECTIVES

1.3.1 AIM OF THE STUDY

Predict the student's performance of Information & Communication Technology Subject in GCE (O/L) Examination.

1.3.2 SPECIFIC OBJECTIVES

1. Examine the reasons for selecting Information and Communication Technology subject for GCE (O/L).
2. Examine the factors that affect the student's performance of Information and Communication Technology subject for GCE (O/L).
3. Predict the student's performance of GCE (O/L) Information and Communication Technology subject.
4. Provide recommendations for the improvement of the Information & Communication Technology subject in the GCE O/L.

CHAPTER 2

LITRETURE REVIEW

2.1 INTRDUCTION

Data mining can be defined as the process to take new aspects and patterns in huge set of data. Various method can be used for data mining as such crossing of statistics, machine learning and database systems. It is also a field of knowledge discovery in databases (KDD), which is the area of finding the different and potentially valuable information from huge set of data. (Fayyad et al. 1996)

Educational Data Mining (EDM) is special data mining system in Education area. EDM consist of tools, techniques, and research designs utilized to gain data from educational resource, online logs, and result of examination, after that analyses this collected data to required conclusions. EDM is theory-oriented and efforts on the link to pedagogical theory (Berland et al., 2014).

Given in the current situation there is a huge variety of different learning contexts, they define the analytical methods utilized by EDM. Hence EDM can be advantage in the current educational practices, as demonstrated in the research, could be crucial.

“Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique and increasingly large-scale data obtained from educational settings, and uses those methods to better understand students and the settings in which they learn” (International Educational Data Mining Society, 2011).

In accordance with International Educational Data Mining Society (2011), information in any educational context is generally consisted of numerous hierarchical levels, which cannot be decided prior to but must be confirmed by properties found in the data source. Important factors for study of educational data are as such sequence, context and time. In this circumstantial learning behaviors of students like students’ participation, login frequency, chat messages, and the type of problems raised to instructor accompany with their final scores can be analyzed. (Abdous et al. 2012)

2.2 REVIEW OF LITERATURE

This section discusses a brief review of papers published in EDM from 2006 to 2017 of different data mining techniques used for educational data mining and previous work done to explore the distinct factors affecting performance of students.

2.2.1 DIFFERENT DATA MINING TECHNIQUES USED FOR EDUCATIONAL DATA MINING

H. Mousa, A. Maghari predicted the performance of students using DM classification techniques. Three classifiers (Naïve Bayes, Decision Tree and K-NN) were used by them and found that the best results are given on Decision Tree classifier when used with students' data (social an academic attributes). Further, they identified that the social factors are effected to the student's performance is very miner and huge outcome is effected from the academic features as such result of previous year and first term. [5]

A.K.Pal,S.Pal have achieved goal, student's performance evaluated by him based on three selected parameters in classification algorithm using Weka. The result is ID3 the classifier shows lowest average error compared to other. In accordance with results it shows that among the machines the learning algorithm is verified, the ID3 classifier has substantial increase the generally accepted classification methods of taken for execution. The decision tree classifiers are observed and investigate are being conducted to taken the appropriate classifier for predicting student performance on BCA exam. According to accuracy of the classifiers indicate that the correct positive norm the model for the FAIL class is 0.84 for ID3 and C4. Five decision trees that mean the model is correctly identified. It is shown the students who are most probably to fail. On the other hand, the result can be increased through proper guiding to students. Five the tree-based algorithm can study actual and predictive models from student data collected over previous years. They could produce miner, but correctly list of forecasts for the student by using predictive models for attributes of new students. This can be recognized the students who required special attention. [6]

B.K.Baradwaj, S.Pal , Student data base was taken to predict the student division from early data base. The decision tree which is most suitable method is used for data classification. The information as such assignment marks, attendance, seminar and class test have been taken from previous database, to predict the performance. Above prediction can be used to increased performance of students as well as teachers. This exercised can be used to recognized students who wanted special guidance to minimize of the failure. [7]

A.A.Saa used number of data mining assignment to generate proper predictive models. Accordingly, student's grade can be predicted effectively and efficiently from the set of data set. Initially social, personal and academic data have been collected from university students. After that, the collected data set were pre-processed and scrutinized to get appropriate data mining tasks. Then the classification model was develop based on data mining task and it was presented. Eventually the result has been taken from classification model accordingly he observed that Naive Bayes model is more interested method. Now it is observed that academic performance of student is not totally depend on academic facts, despite the fact that there are many other factors that also have a greater impact. This research can be used to motivate and guide to students in the university. [8]

P.N.W.A.L.K.Premarathne who identified some data mining patterns from the examination results by using association rule. Thus, the study leads to the discovery of any predominant relationship between the mathematics results of the G.C.E. (O/L). Further she found that there is no relationship for GCE A/L ICT result whether students who followed GCE O/L ICT or not. Initially, she has concentrated on determining the level of difficulty of each and every question in the examination using the answers were provided in the questionnaire. According to the selected questions, a study was started. The data mining Association rule has used to recognized suitable pattern in survey. According to the analysis it appears that the result of Mathematics at the usual level of student were directly affected on their result in the ICT subject in GCE (A/L). Further it was demonstrated that the subsequent ICT subject at the regular exam level also has a direct impact on the result of its GCE A/L ICT subject. This mean that the better the performance of mathematics and ICT, the greater probability of success at the GCE (A/L) ICT subject. He has observed that there is an impact of GCE (A/L) ICT result between genders. Accordingly, it appears that the performance of female students is better than male students in the GCE (A/L) ICT. According to the analysis Mathematics and ICT subject in GCE O/L is most essential subjects to success G. C. E. (A/L) ICT. Even though analysis showed that it was not arising in every time. It was noticed that, the student did not perform well selected questions as such logic gates and networking. However male students were well achieved above said question better than female students. As a result of other factors quantitative analysis cannot be covered. Further notice that some other few students who achieved very well in GCE O/L exam but they did not meet above criteria. Finally, it was revealed that fulfil of basic requirement of the students is more likely to show higher results in GCE A/L ICT. [9]

2.2.2 DIFFERENT FACTORS AFFECTING THE PERFORMANCE OF INFROMATION & COMMUNICATION TECHNOLOGY SUBJECT IN THE GCE (O/L)

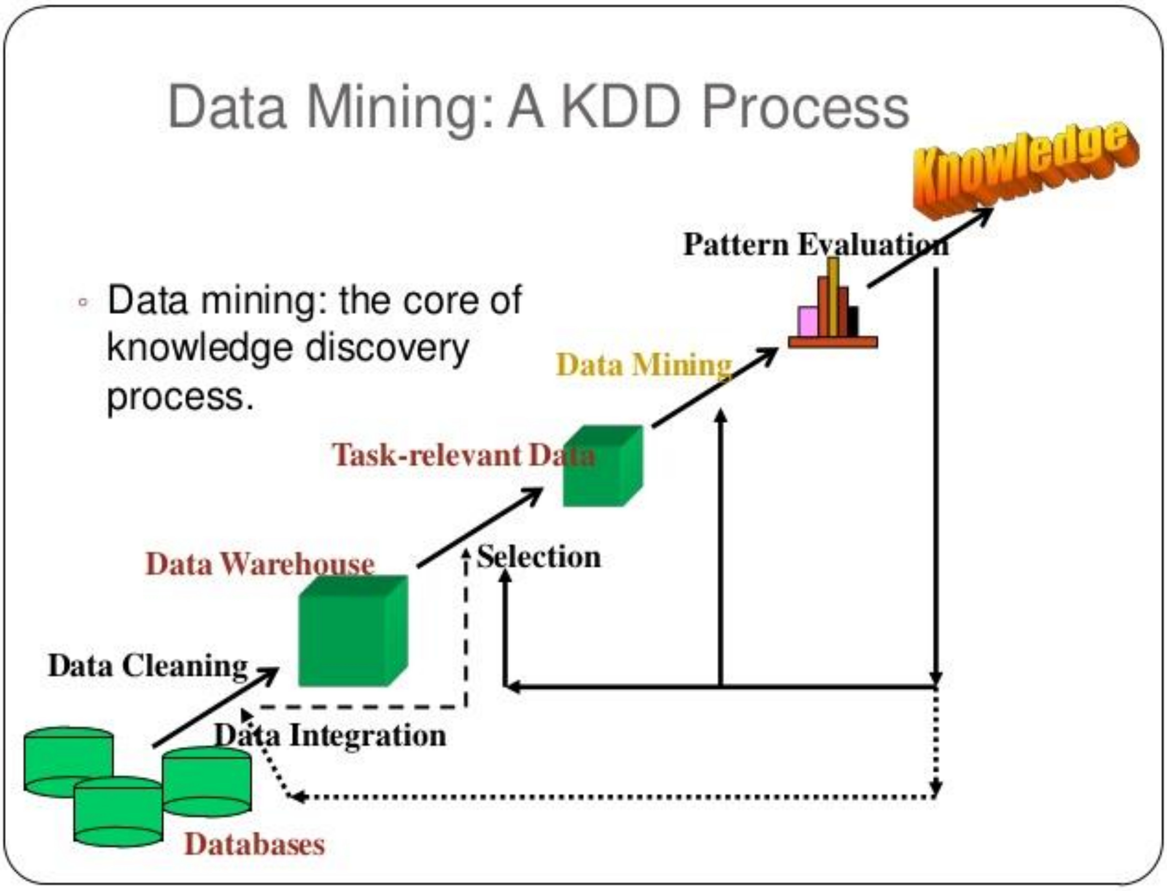
M.G.N.A.S.Fernando, M.B.Ekanayake proposed that all infrastructure and computer related resources for ICT education should be implemented immediately and at the same time it is necessary to identify the cadre requirements and design suitable training method to cater to the present curriculum and its future expansions. They suggested that there are several challenges to overcome when transforming the existing into a blended technology-based ICT education. At the same time content revising, updating and many changes according to implementation is another challenge. To overcome all above challenges, it is necessary to have proper leadership, guidance and a well-planned vision for the future. [10]

A.Ilmudeen Proposed that physical resource allocation and usage human resources, revised ICT training programmes increasing the practical component, using methods of distance learning for improving ICT saturation among students, continuous and a permanent training programme is important for Introduction of ICT in Sri Lankan school. Sri Lanka is in initial stage of ICT implementation with lower evaluation, not in this curriculum ask for any ICT knowledge as input requirement. Thus, this program is designed for the introduction of ICT as a technical tool the subject will be given in GCE (O/L). That the main concentration is development capabilities for used ICT tools as well as give theoretical knowledge for the students to continue their higher education in the field of ICT. It is indicated that main negative points of ICT education as such absence of a computer lab, less qualified ICT teachers, less motivation for ICT subject, high cost implementation for maintenance computer labs, language literacy, no ICT accreditation at the national level. Resources have to be develop in the schools in Sri Lanka as a requirement of ICT industry through Government in collaboration with other organization such as NGOs, privet and public partnership projects. [11]

CHAPTER 3

TECHNOLOGIES USED FOR THE STUDY

3.1 KDD PROCESS



Data Mining: A KDD Process

- Data mining: the core of knowledge discovery process.

3.1.1 Data Cleaning: It is called the deduction of noisy and unconnected data from collected data set. Following process has to be done when the data cleaning process;

- Clean **Missing values**.
- Clean **noisy** data
- Clean with **Data inconsistency discovery** and **Data transformation tools**.

3.1.2 Data Integration: It is stated as diverse data from numerous sources and combined into common source (Data Warehouse). Data integration can be done with the different tools as follows;

- **Data Migration tools**.
- **Data Synchronization tools**.
- **ETL** (Extract-Load-Transformation) process.

3.1.3 Data Selection: The process where data relevant to the analysis is decided and retrieved from the data collection. Data selection method are as follows;

- **Neural network**.
- **Decision Trees**.
- **Naive bayes**.
- **Clustering**, **Regression**, etc.

3.1.4 Data Transformation: The process of transforming data into proper form required by mining procedure. The process can be given as follows;

- **Data Mapping**.
- **Code generation**

3.1.5 Data Mining: It called as the clever techniques that are used to extract patterns potentially valuable.

- The Task relevant data have transformed into the **patterns**.
- The Purpose of model is decided using **classification** or **characterization**.

3.1.6 Pattern Evaluation: It is stated that the recognizing firmly increasing patterns representing knowledge based on given measures.

- Identify **interestingness score** of each pattern.
- Take **summarization** & **Visualization** to produced data understandable

3.1.7 Knowledge representation: The technique which utilizes visualization tools to represent data mining results. As such;

- **Reports**.
- **Tables**.
- **Discriminant rules**, **classification rules**, **characterization rules**, etc.

3.2 DATA MINING

Data mining has taken such as valid, hidden, potentially valuable patterns in large set of data. Unanticipated/ earlier unidentified relationships are recognized from the set of data.

3.2.1 Data Mining task

Data mining tasks are as follows: –

- Descriptive: describe the general properties & characteristic of data.
- Predictive: Understanding of forecast from present set of data

3.2.2 Functionalities of Data Mining

Data mining functionalities is required to recognize various type of patterns from data set.

3.2.2.1. Classification:

Important and applicable information can be recovered from data set using classification method. Further this method can be classified data into different sectors.

3.2.2.2. Clustering:

Clustering analysis is required to recognize relationship between each other as such similarities, difference….

3.2.2.3. Regression:

Regression analysis is applied to recognize and analysis the connection with variables. This can be identified the possibility of a definite variable, given the occurrence of other variables.

3.2.2.4. Association Rules:

It is used to identify the association between two and more items and it was generated a concealed pattern in data set.

3.2.2.5. Outer detection:

In this scenario, it is used to identify the objects in the set of data those are not matched either an expected pattern or performance. It could be used in different kind of domains, such as intrusion, detection, fraud or fault detection …etc. The outer detection is also called as Outlier Analysis or Outlier mining.

3.2.2.6. Sequential Patterns:

Sequential patterns, this is taken to define or recognize similar patterns or trends of transaction data in specific time bar.

3.2.2.7. Prediction:

This is a part of combination with the other data mining techniques such as classification, sequential patterns, trends and clustering…etc. Accordingly, future event can be predicted using analyzed of past events or instances.

3.2.3 Challenges of Data mining:

- Skillful experts are wanted to convey the data mining queries.
- Over fitting: If data set is small, it is no suitable for future states.
- Large databases are difficult to manage in the mining process.
- It is difficult to manage large database in the mining process.
- Business practices are need to be change to identify uncovered data.
- Data mining process may not be accurate If set of data is not diverse
- Integration information is required from various data bases then Global information system could be complicated.

3.2.4 Advantages of Data Mining:

- Deliver knowledge-based data.
- Increase profit in production as well as operation.
- Cost comparison between other applications.
- Use decision-making process.
- Prediction of trends, behaviors and automated discovery of concealed patterns.

- New system and existing platforms are implemented.
- It is the quick process. User can be analyzed large size of data in short time.

3.2.5 Disadvantages of Data Mining

- Company may transfer important information to gain economic benefits
- Many software is difficult to operate and need more practical training.
- There are different types of tools which are operate in various manner. Therefore, various type of algorithms was created in the design. So, it is difficult to select suitable data mining tool.
- Generally, the data mining techniques are not accurate therefore in this circumstance it can be caused difficult conditions

3.3 CURRENT DATA MINING TOOLS

- Rapid Miner
- WEKA
- R-Programming Tool
- Python based Orange and NTLK
- Knime

CHAPTER 4

METHODOLOGY

4.1 PROLEM STATEMENT

The main objectives of this research is to predict performance of student for Information & Communication Technology Subject in GCE (O/L) Examination.

4.2 DATA COLLECTION TECHNIQUES

In this research, action will be taken to collect data from 30 secondary schools in Colombo Districts.

4.2.1 Population data set:

Table 4.2.1: No of schools functioning O/L ICT subject in 2018

Province	National school				Provincial School				Total
	1 AB	1 C	Type 2	Total	1 AB	1 C	Type 2	Total	
Western	64	0	1	65	120	204	160	484	549
Central	43	8	0	51	62	196	79	337	388
Southern	59	1	0	60	77	168	132	377	437
Northern	19	1	0	20	79	79	82	240	260
Eastern	29	1	0	30	56	101	50	207	260
North western	30	3	0	33	74	177	79	330	363
North central	9	0	0	9	48	96	70	214	223
Uva	27	6	0	33	45	123	62	230	263
Sabaragamuwa	25	0	0	25	75	129	130	334	350
Total	**305**	**20**	**1**	**326**	**636**	**1273**	**844**	**2753**	**3079**

Data Source: - School census data 2018

Table 4.2.2: No of students who studied ICT subject in 2018

Province	No of students in National school	No of students in Provincial school	Total No of students
1. Western	5147	7157	12304
2. Central	1697	3233	4930
3. Southern	5817	3687	9504
4. Northern	996	1969	2965
5. Eastern	1306	1890	3196
6. North Western	2041	4070	6111
7. North Central	1009	2791	3800
8. Uva	1345	2306	3651
9. Sabaragamuwa	1872	2916	4788
Sri Lanka	**21230**	**30019**	**51249**

Data Source: - School census data 2018

4.2.2 Sample data set:

I was selected 30 schools which is functioning GCE (O/L) ICT subject in Colombo district. Then I was selected 10 students between in each schools and total sample is 301 students.

4.2.3 Data collection method:

The questionnaire method is used for data collection and google forms was used for that. In addition, examination results of the Department of Examinations will be used as secondary data in this research.

4.2.4 Limitation of the data set: Although GCE O/L subject is functioning in all the districts in Sri Lanka, I was selected only 30 schools in Colombo district. Factors are affected for it as follows.

- Limitation of the research time
- Difficulties of travelling all the districts for collect data

4.2.5 Data Set Attributes

Table 4.2.5.1 Description of attributes and possible values are as follows;

Attribute	Description	Possible Values
Gender	Student's Gender	{F, M}
WhSel	Why selected ICT	{Consent, A Request of parents, B Request of teachers, C The social demand, D Other, E}
ReMaths	Grade obtained for O/L Mathematics	{A, B, C, S, F}
ReEnglish	Grade obtained for O/L English	{A, B, C, S, F}
ReICT	Grade obtained for O/L ICT	{A, B, C, S, F}
WhDiff	Why do you think that ICT paper is the hard?	{missed the topic, A couldn't understand some topic, B no expert ICT teacher, C Lack of preparation for the exam, D}
ParTution	Did you participate tution classes	{yes, no}
PrKnow	Did you get practical training for the required topics	{yes, Sometimes, never }
ExRea	Did you get external readings in addition to using the textbook?	{yes, no}
ReWrk	Have you participated in the GCE O / L Examination Review Workshop on Information Technology?	{yes, no}
PraTest	Can you increase your score if you have a practical test in addition to the written test?	{yes, no}
ICTTea	Is there an ICT teacher at the school?	{yes, no}
FaIn	Family monthly income	{more than 200000, A 200000-100000, B 100000-50000, C 50000-25000, D less than 25000, E}

PaEdL	Parents Education level	Post graduate, A degree, B diploma, C A/L, D O/L, E}
Lap/Com	Have a computer/laptop in your home?	{yes, no}
IntFac	Have an internet facility in your home?	{yes, no}
HSup	Get home support to learn ICT	{parents or other adults at home express the theory part, A experience in observing people in the ICT field or further education, B parents or other adults at home giving practical training, C provide physical recourses, D no support, E}

4.3 APPLYING DM CLASSIFICATION ALGORITHMS

The classification is mostly used data mining technique. It can be handled simply and easy. Where the data mining, the Classification which is a prediction method for a data object class or a category based on before studied classes from a training dataset which knows as an object classes. K-nearest neighbor (K-NN), Decision trees, Neural networks of the network, Naive Bayes are the several classification methods available in the data mining.

K-nearest neighbor (K-NN), ID3 Decision Tree and Naïve-Bayes classifier have been used for my experiments. My data set was small size, includes 301 records and 17 features. So, I have selected above classifiers.

4.4 RESULTS EVALUATION

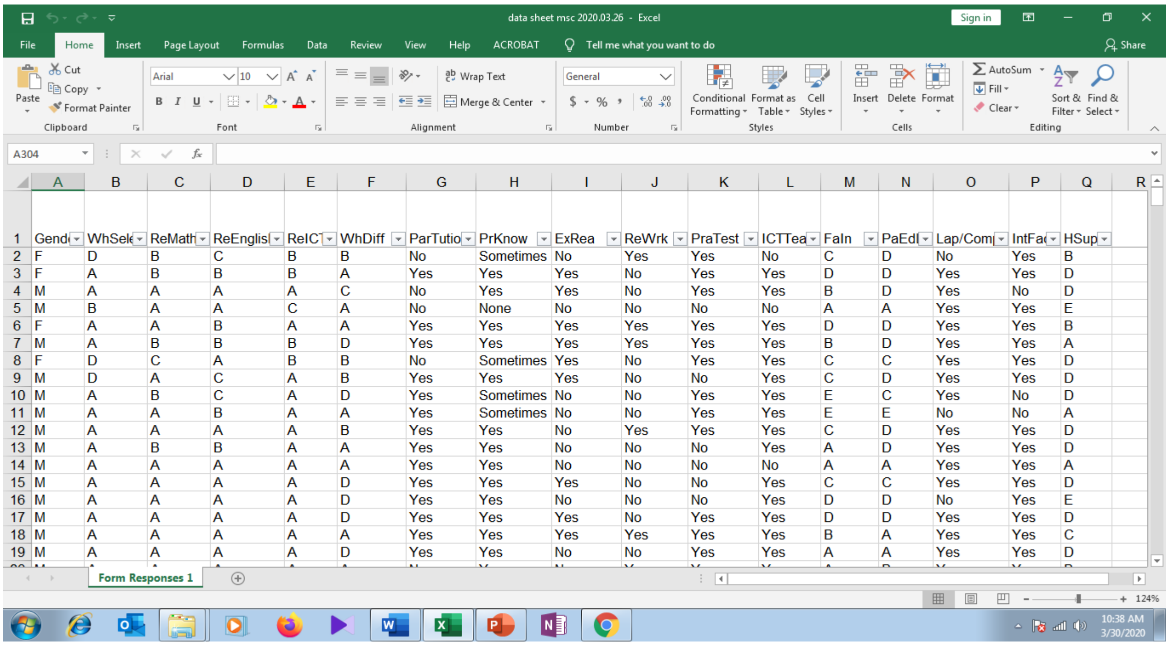
Considering project objectives, I used evaluation techniques and methods to evaluate results. The result and evaluation are excluded in the fifth chapter. In the evaluation process we used precision and recall measures.

CHAPTER 5

RESULTS EVALUATION

5.1 Collected data from the student questioners listed as follows:

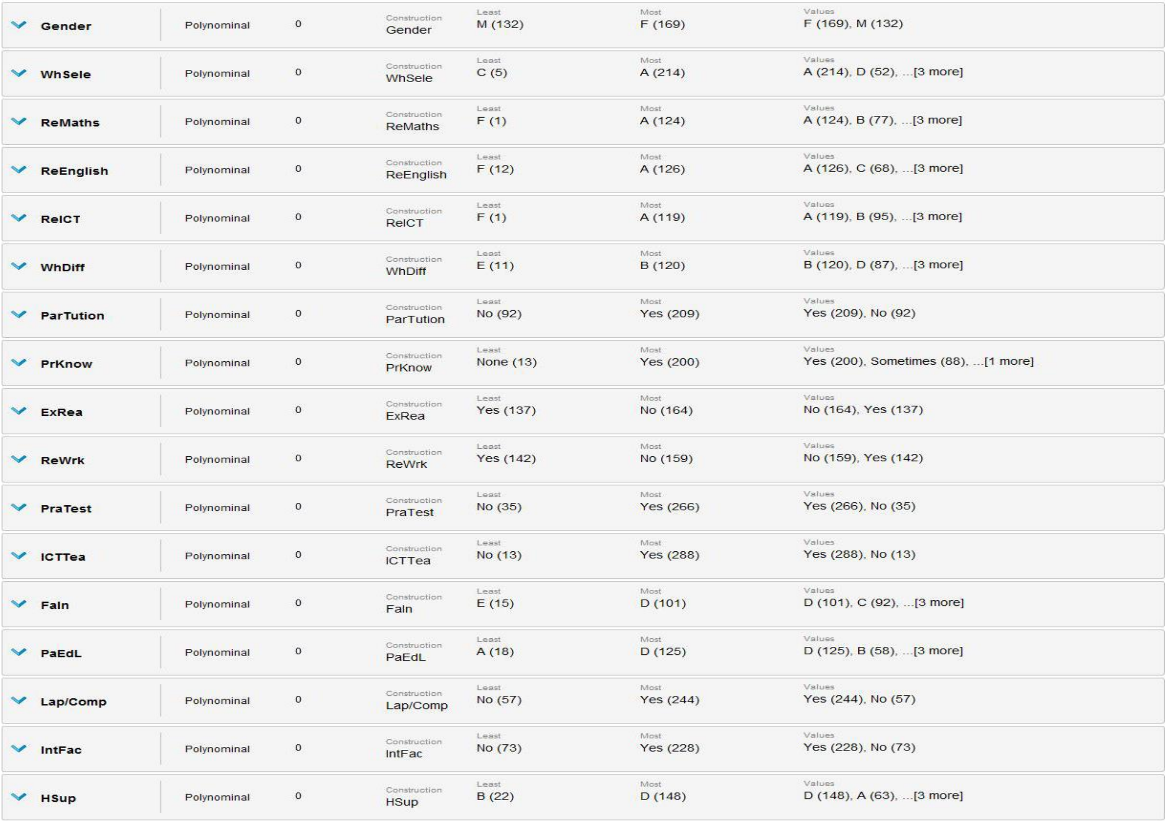
Figure 5.1 Collected data sheet



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Gend	WhSel	ReMath	ReEnglis	ReIC	WhDiff	ParTutio	PrKnow	ExRea	ReWrk	PraTest	ICTTea	FaIn	PaEdI	Lap/Com	IntFa	HSup
2	F	D	B	C	B	B	No	Sometimes	No	Yes	Yes	No	C	D	No	Yes	B
3	F	A	B	B	B	A	Yes	Yes	Yes	No	Yes	Yes	D	D	Yes	Yes	D
4	M	A	A	A	A	C	No	Yes	Yes	No	Yes	Yes	B	D	Yes	No	D
5	M	B	A	A	C	A	No	None	No	No	No	No	A	A	Yes	Yes	E
6	F	A	A	B	A	A	Yes	Yes	Yes	Yes	Yes	Yes	D	D	Yes	Yes	B
7	M	A	B	B	B	D	Yes	Yes	Yes	Yes	Yes	Yes	B	D	Yes	Yes	A
8	F	D	C	A	B	B	No	Sometimes	Yes	No	Yes	Yes	C	C	Yes	Yes	D
9	M	D	A	C	A	B	Yes	Yes	Yes	No	No	Yes	C	D	Yes	Yes	D
10	M	A	B	C	A	D	Yes	Sometimes	No	No	Yes	Yes	E	C	Yes	No	D
11	M	A	A	B	A	A	Yes	Sometimes	No	No	Yes	Yes	E	E	No	No	A
12	M	A	A	A	A	B	Yes	Yes	No	Yes	Yes	Yes	C	D	Yes	Yes	D
13	M	A	B	B	A	A	Yes	Yes	No	No	No	Yes	A	D	Yes	Yes	D
14	M	A	A	A	A	A	Yes	Yes	No	No	No	No	A	A	Yes	Yes	A
15	M	A	A	A	A	D	Yes	Yes	Yes	No	No	Yes	C	C	Yes	Yes	D
16	M	A	A	A	A	D	Yes	Yes	No	No	No	Yes	D	D	No	Yes	E
17	M	A	A	A	A	D	Yes	Yes	Yes	No	Yes	Yes	D	D	Yes	Yes	D
18	M	A	A	A	A	A	Yes	Yes	Yes	Yes	Yes	Yes	B	A	Yes	Yes	C
19	M	A	A	A	A	D	Yes	Yes	No	No	Yes	Yes	A	A	Yes	Yes	D

5.2 Collected data has been uploaded to the rapid miner software and analyzed.

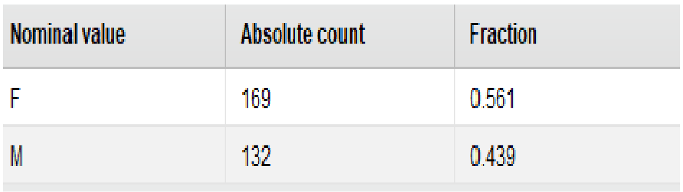
Figure 5.2 least value and most value in the field



			Construction	Least	Most	Values
Gender	Polynominal	0	Gender	M (132)	F (169)	F (169), M (132)
WhSele	Polynominal	0	WhSele	C (5)	A (214)	A (214), D (52), ...[3 more]
ReMaths	Polynominal	0	ReMaths	F (1)	A (124)	A (124), B (77), ...[3 more]
ReEnglish	Polynominal	0	ReEnglish	F (12)	A (126)	A (126), C (68), ...[3 more]
ReICT	Polynominal	0	ReICT	F (1)	A (119)	A (119), B (95), ...[3 more]
WhDiff	Polynominal	0	WhDiff	E (11)	B (120)	B (120), D (87), ...[3 more]
ParTution	Polynominal	0	ParTution	No (92)	Yes (209)	Yes (209), No (92)
PrKnow	Polynominal	0	PrKnow	None (13)	Yes (200)	Yes (200), Sometimes (88), ...[1 more]
ExRea	Polynominal	0	ExRea	Yes (137)	No (164)	No (164), Yes (137)
ReWrk	Polynominal	0	ReWrk	Yes (142)	No (159)	No (159), Yes (142)
PraTest	Polynominal	0	PraTest	No (35)	Yes (266)	Yes (266), No (35)
ICTTea	Polynominal	0	ICTTea	No (13)	Yes (288)	Yes (288), No (13)
FaIn	Polynominal	0	FaIn	E (15)	D (101)	D (101), C (92), ...[3 more]
PaEdL	Polynominal	0	PaEdL	A (18)	D (125)	D (125), B (58), ...[3 more]
Lap/Comp	Polynominal	0	Lap/Comp	No (57)	Yes (244)	Yes (244), No (57)
IntFac	Polynominal	0	IntFac	No (73)	Yes (228)	Yes (228), No (73)
HSup	Polynominal	0	HSup	B (22)	D (148)	D (148), A (63), ...[3 more]

5.3 After applying the preparation methods, we analyze the data visually and figure out the distribution of values.

Table 5.3.1 Collected data group by gender



Nominal value	Absolute count	Fraction
F	169	0.561
M	132	0.439

Figure 5.3.1 Histogram of Gender attribute

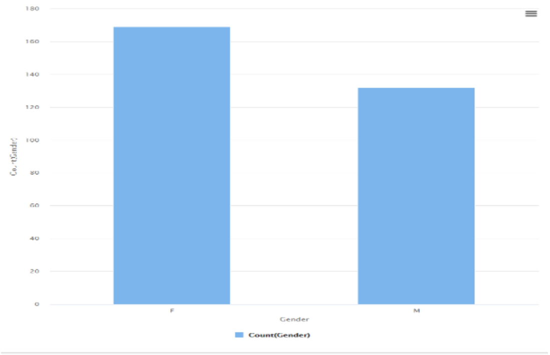
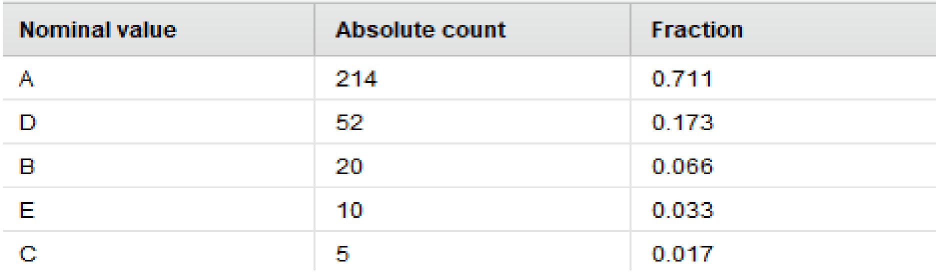


Table 5.3.2 Why did you select ICT subject for the O/L Examination



Nominal value	Absolute count	Fraction
A	214	0.711
D	52	0.173
B	20	0.066
E	10	0.033
C	5	0.017

Figure 5.3.2 Histogram of Whsel attribute

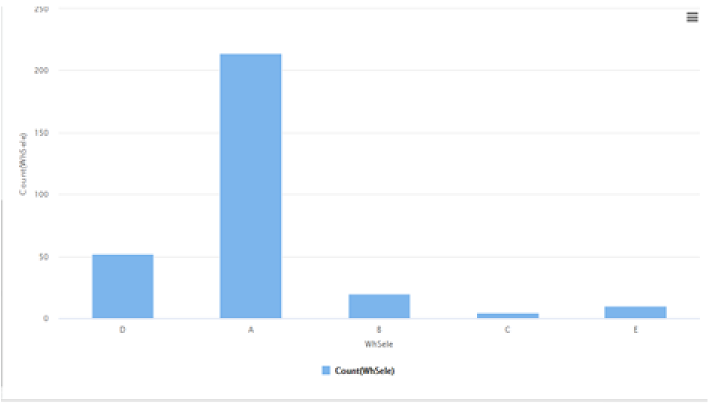
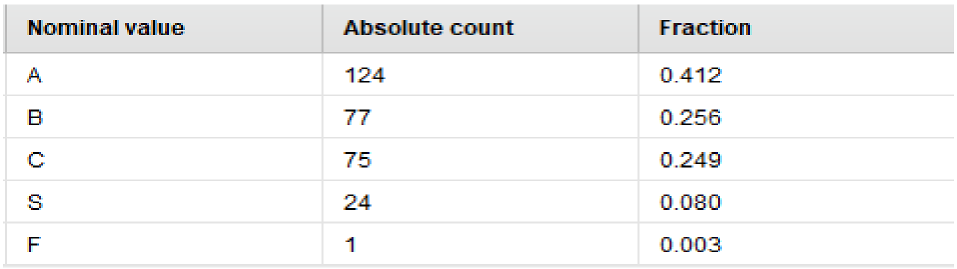


Table 5.3.3 What was the grade obtained for O/L Mathematics



Nominal value	Absolute count	Fraction
A	124	0.412
B	77	0.256
C	75	0.249
S	24	0.080
F	1	0.003

Figure 5.3.3 Histogram of ReMaths attribute

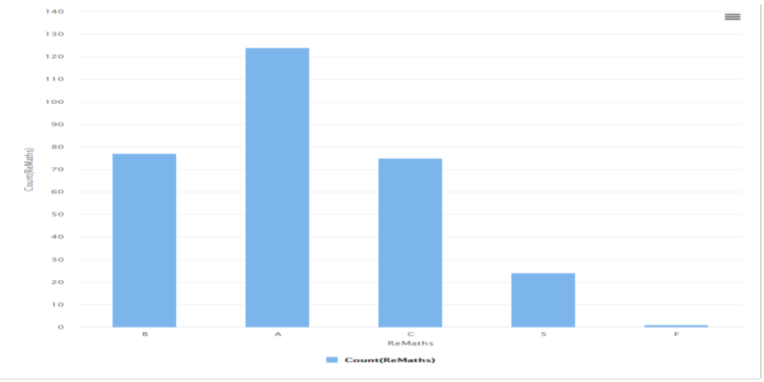
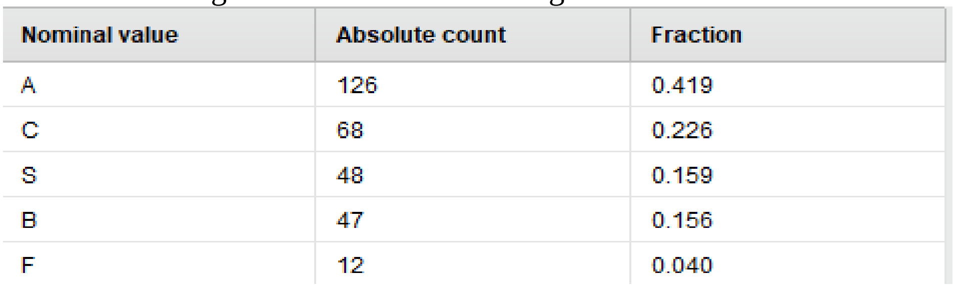


Table 5.3.4 What was the grade obtained for O/L English?



Nominal value	Absolute count	Fraction
A	126	0.419
C	68	0.226
S	48	0.159
B	47	0.156
F	12	0.040

Figure 5.3.4 Histogram of ReEnglish attribute

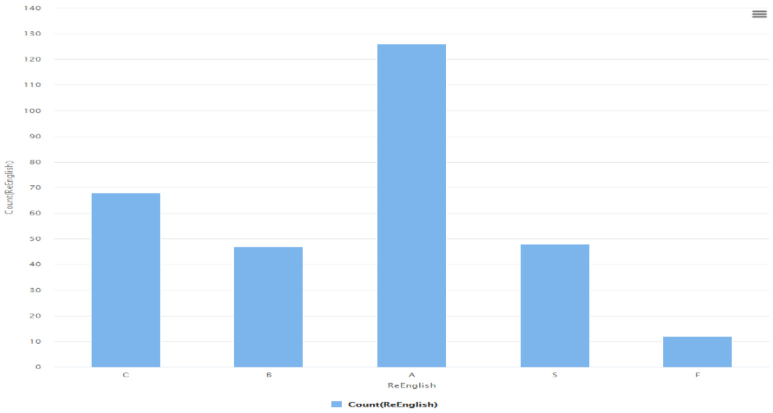
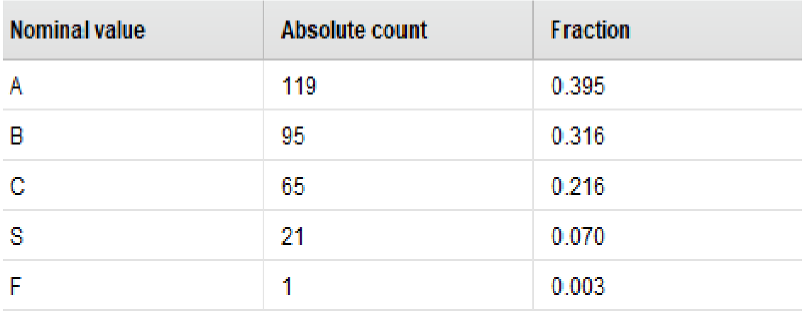


Table 5.3.5 What was the grade obtained for O/L ICT?



Nominal value	Absolute count	Fraction
A	119	0.395
B	95	0.316
C	65	0.216
S	21	0.070
F	1	0.003

Figure 5.3.5 Histogram of ReICT attribute

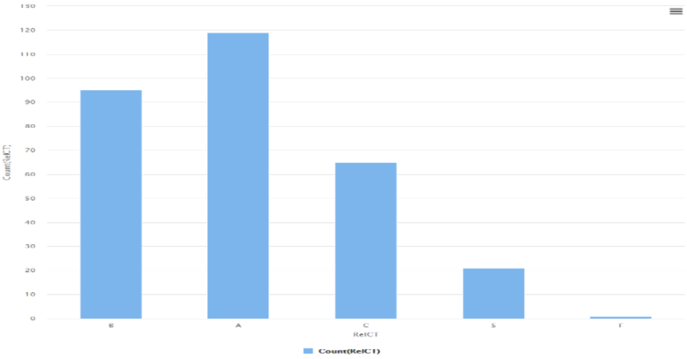
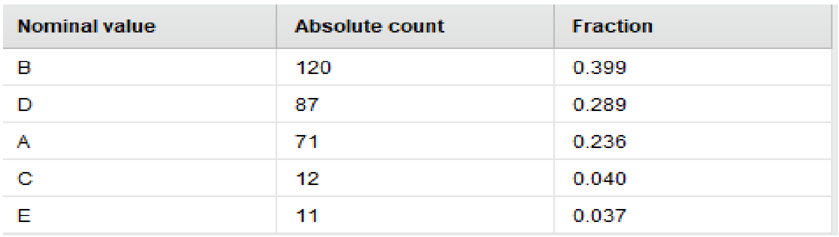


Table 5.3.6 Why do you think that ICT paper is hard?



Nominal value	Absolute count	Fraction
B	120	0.399
D	87	0.289
A	71	0.236
C	12	0.040
E	11	0.037

Figure 5.3.6 Histogram of WhDiff attribute

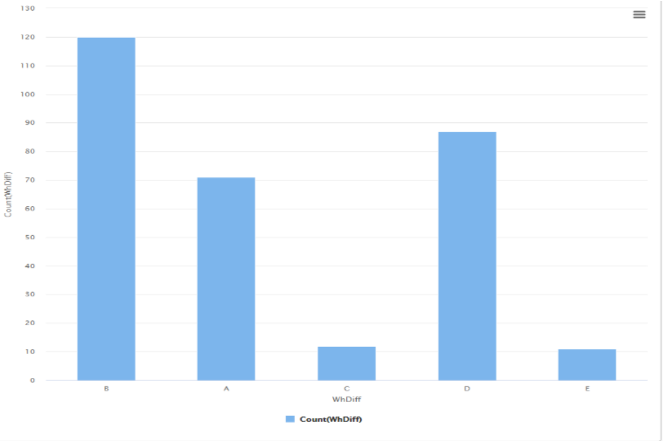
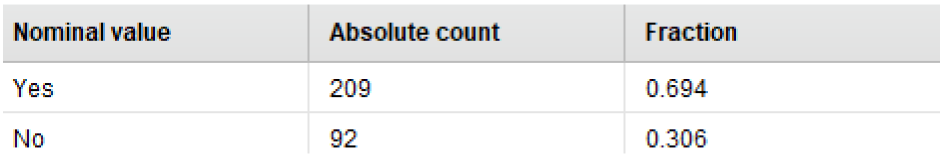


Table 5.3.7 Did you participate ICT tution classes?



Nominal value	Absolute count	Fraction
Yes	209	0.694
No	92	0.306

Figure 5.3.7 Histogram of ParTution attribute

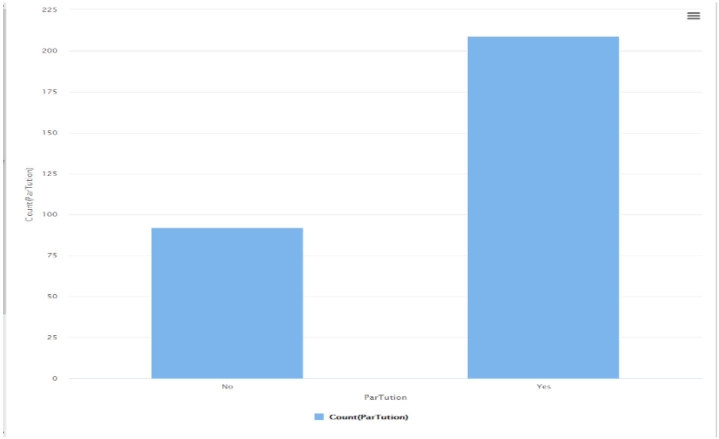
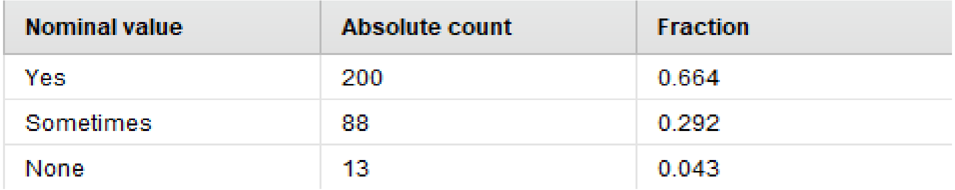


Table 5.3.8 Did you get practical training for the required topics?



Nominal value	Absolute count	Fraction
Yes	200	0.664
Sometimes	88	0.292
None	13	0.043

Figure 5.3.8 Histogram of PraKnow attribute

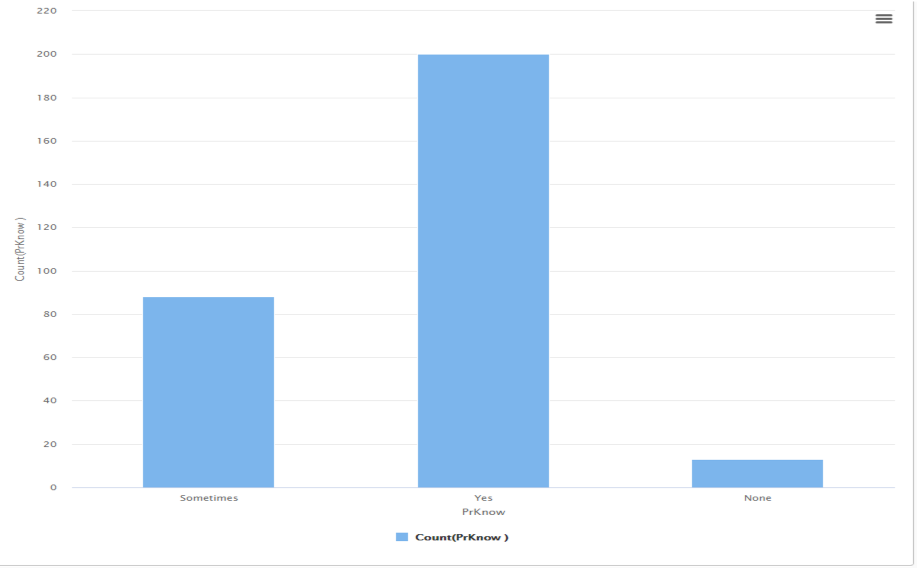
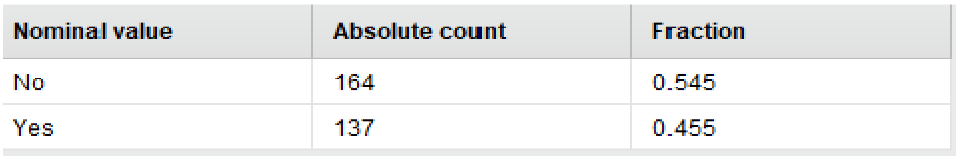


Table 5.3.9 Did you get external readings in addition to using the textbook?



Nominal value	Absolute count	Fraction
No	164	0.545
Yes	137	0.455

Figure 5.3.9 Histogram of ExRea attribute

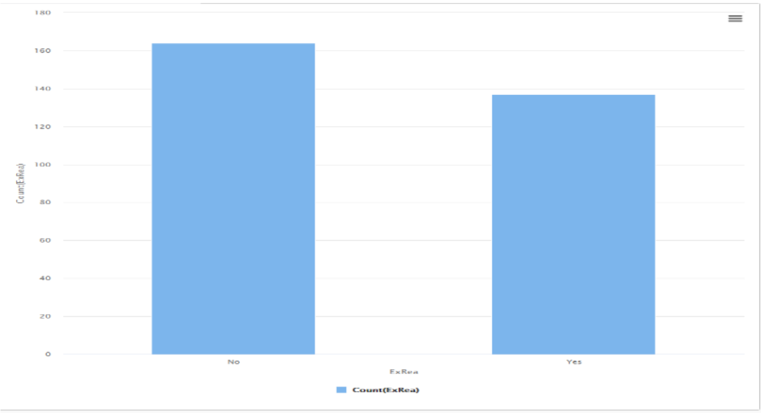
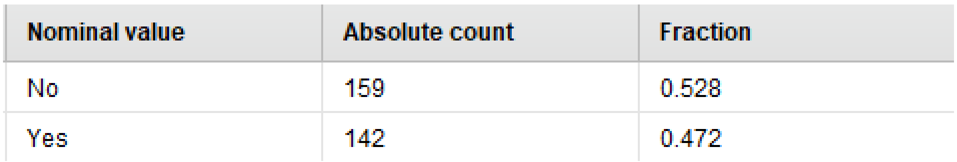


Table 5.3.10 Have you participated in the GCE O / L Examination Review Workshop on Information Technology?



Nominal value	Absolute count	Fraction
No	159	0.528
Yes	142	0.472

Figure 5.3.10 Histogram of ReWrk attribute

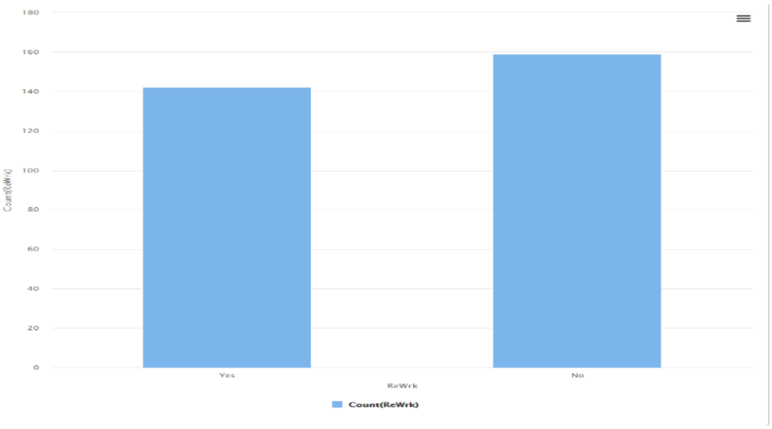
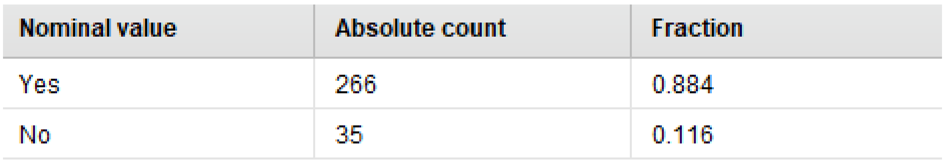


Table 5.3.11 Can you increase your score if you have a practical test in addition to the written test?



Nominal value	Absolute count	Fraction
Yes	266	0.884
No	35	0.116

Figure 5.3.11 Histogram of PraTest attribute

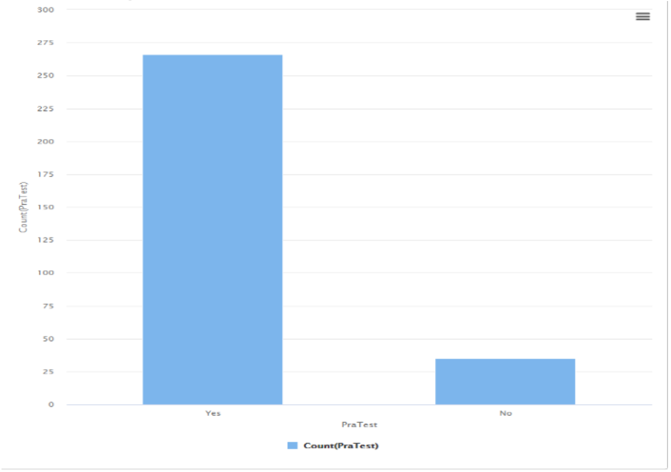
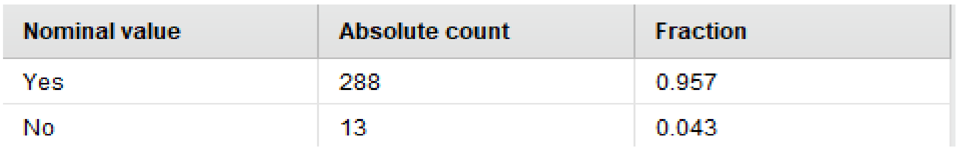


Table 5.3.12 Is there an ICT teacher at the school?



Nominal value	Absolute count	Fraction
Yes	288	0.957
No	13	0.043

Figure 5.3.12 Histogram of ICTTea attribute

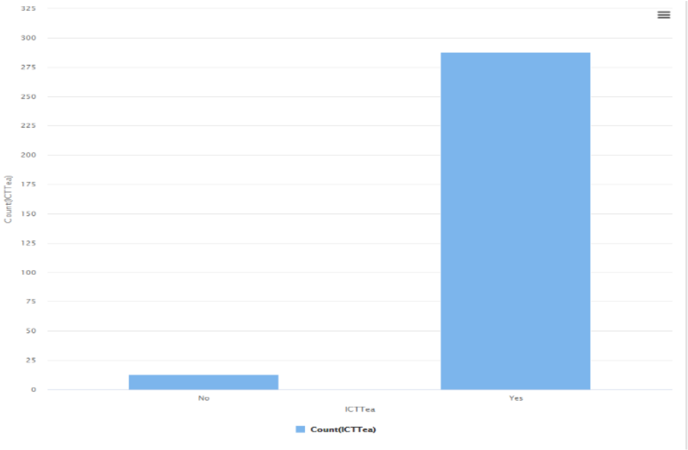
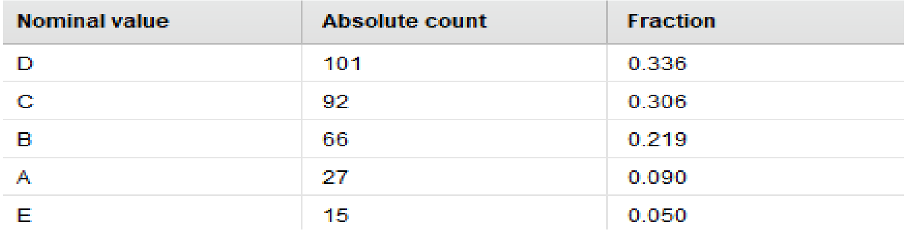


Table 5.3.13 What is your monthly family income?



Nominal value	Absolute count	Fraction
D	101	0.336
C	92	0.306
B	66	0.219
A	27	0.090
E	15	0.050

Figure 5.3.13 Histogram of FaIn attribute

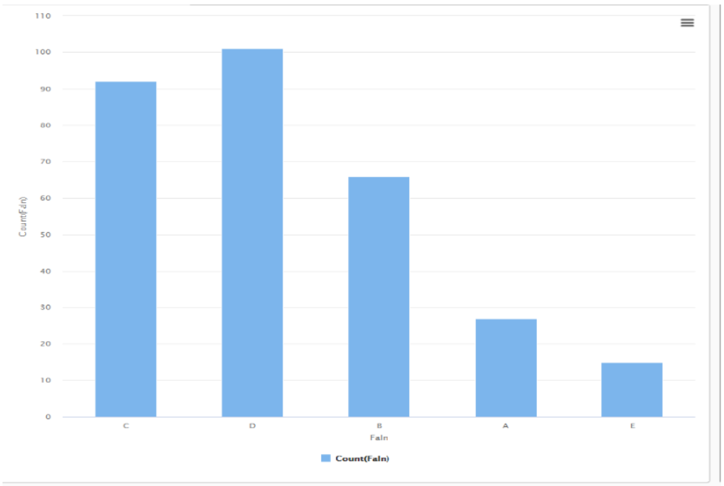
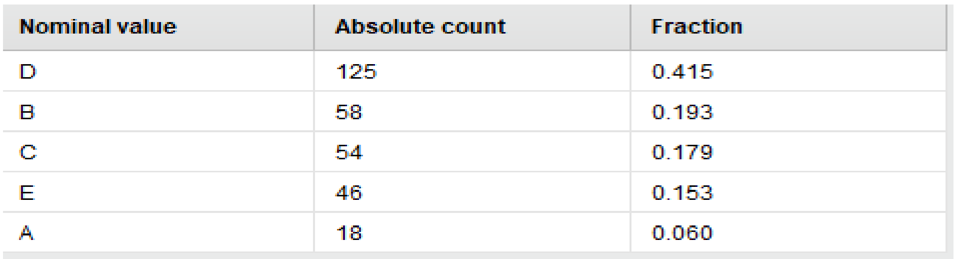


Table 5.3.14 What is the level of your parent’s education?



Nominal value	Absolute count	Fraction
D	125	0.415
B	58	0.193
C	54	0.179
E	46	0.153
A	18	0.060

Figure 5.3.14 Histogram of PaEdL attribute

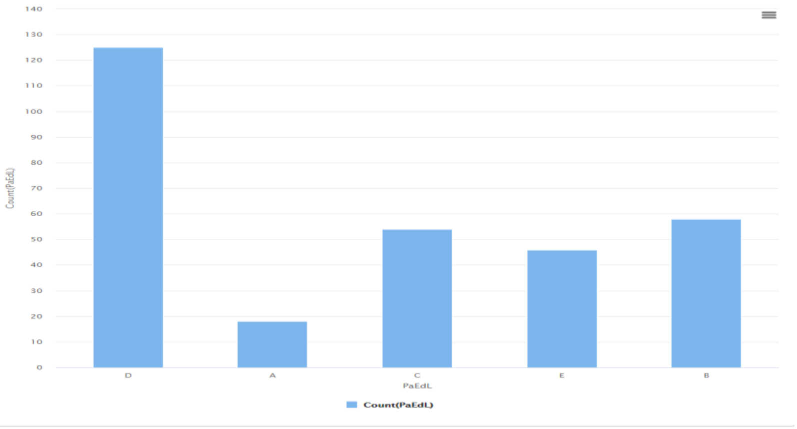
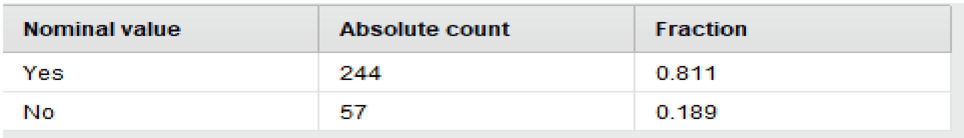


Table 5.3.15 Have a computer/laptop in your home?



Nominal value	Absolute count	Fraction
Yes	244	0.811
No	57	0.189

Figure 5.3.15 Histogram of Lap/Com attribute

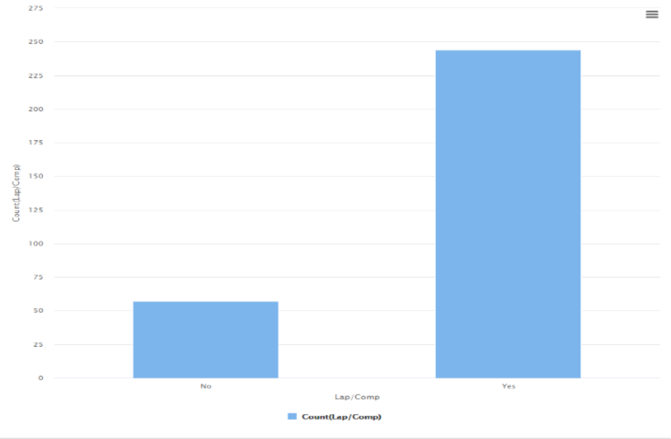
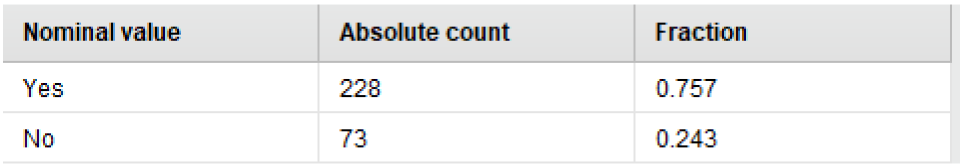


Table 5.3.16 Have an internet facility in your home?



Nominal value	Absolute count	Fraction
Yes	228	0.757
No	73	0.243

Figure 5.3.16 Histogram of IntFac attribute

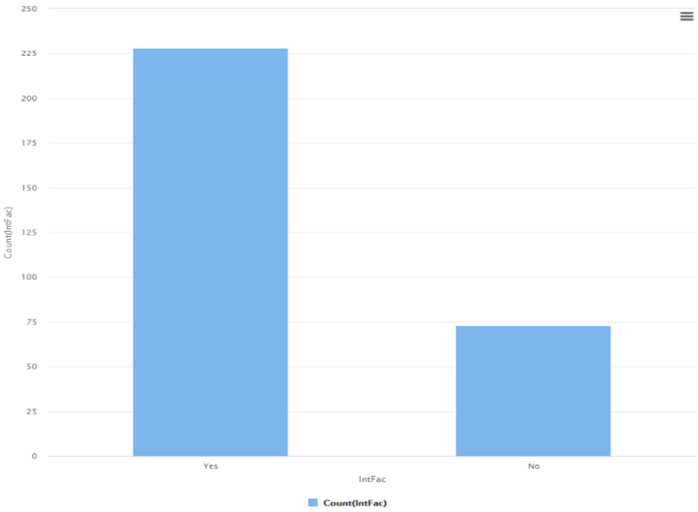
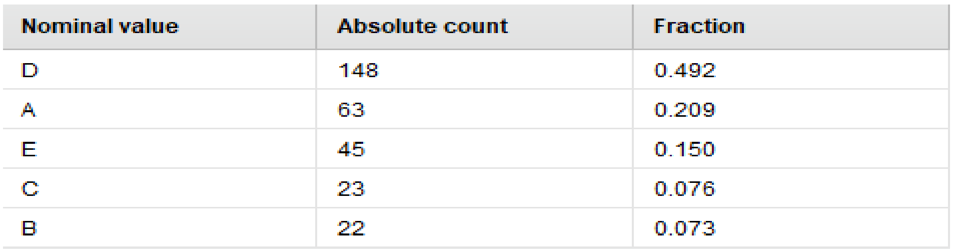
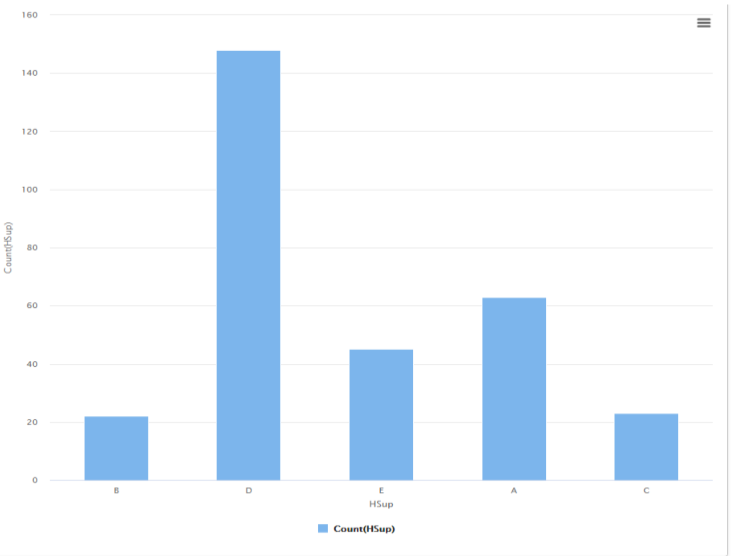


Table 5.3.17 How to get support in the home to learn this subject?



Nominal value	Absolute count	Fraction
D	148	0.492
A	63	0.209
E	45	0.150
C	23	0.076
B	22	0.073

Figure 5.3.17 Histogram of Hsup attribute

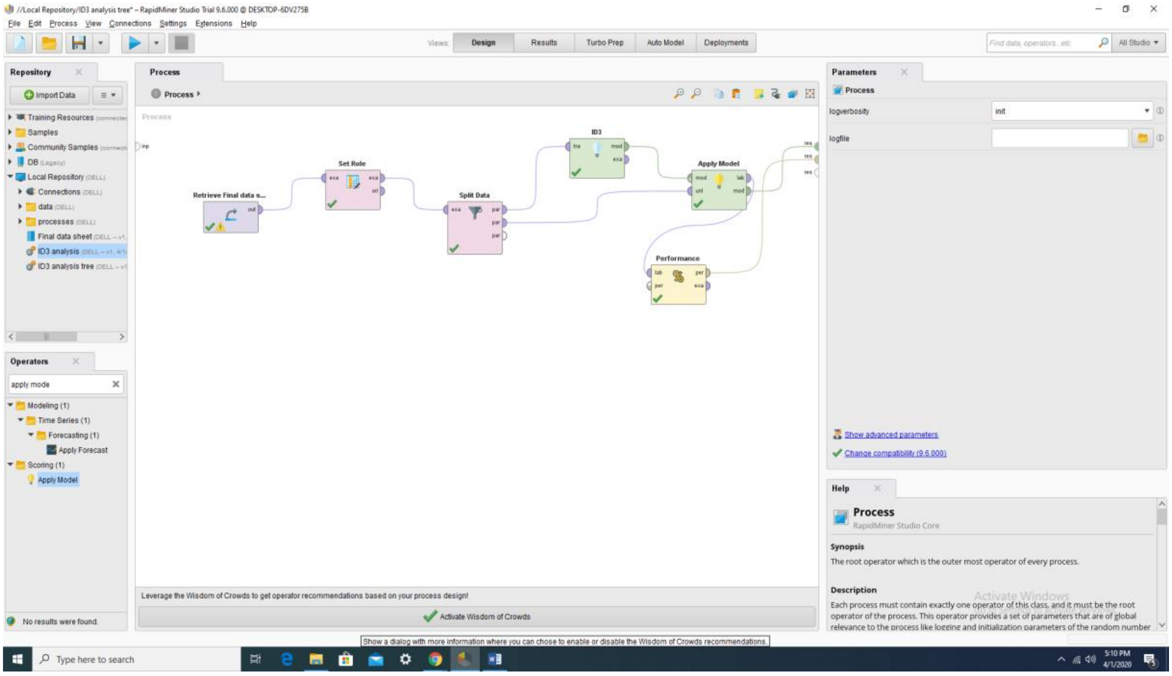


ID3

In Decision Tree learning, one of the most popular algorithms is the **ID3 algorithm** or the **Iterative Dichotomiser 3 algorithm.** J. Ross Quinlan, ID**3 algorithm was developed** in 1975, It is commonly taken to create a decision tree from a collected set of data by using a top-down, greedy search, to test each attribute at every node of the tree. The subsequent tree can be taken to classify future samples.

In this occasion set of data is divided into two parts and 70% of data was taken to create ID3 model. Then another 30% of data was used to apply the model.

A decision tree was generated from a dataset visualized as follows:



logverbosity: init

logfile:

Leverage the Wisdom of Crowds to get operator recommendations based on your process design!

Activate Wisdom of Crowds

Show a dialog with more information where you can chose to enable or disable the Wisdom of Crowds recommendations.

Synopsis

The root operator which is the outer most operator of every process.

Description

Each process must contain exactly one operator of this class, and it must be the root operator of the process. This operator provides a set of parameters that are of global relevance to the process like logging and initialization parameters of the random number

Following settings are considered with the ID3 operator to produce the decision tree;

- Splitting criterion =information gain ratio
- Minimal size of split =4
- Minimal leaf size =2
- Minimal gain =0.01

When the ID3 decision tree algorithm was run, Confusion matrix was created and it can be shown as follows. According to the ID3 algorithm, it could be predicted the class of 60 objects out of 89, and accuracy value is 67.42%.

Table 5.3.18 Confusion matrix of ID3 model

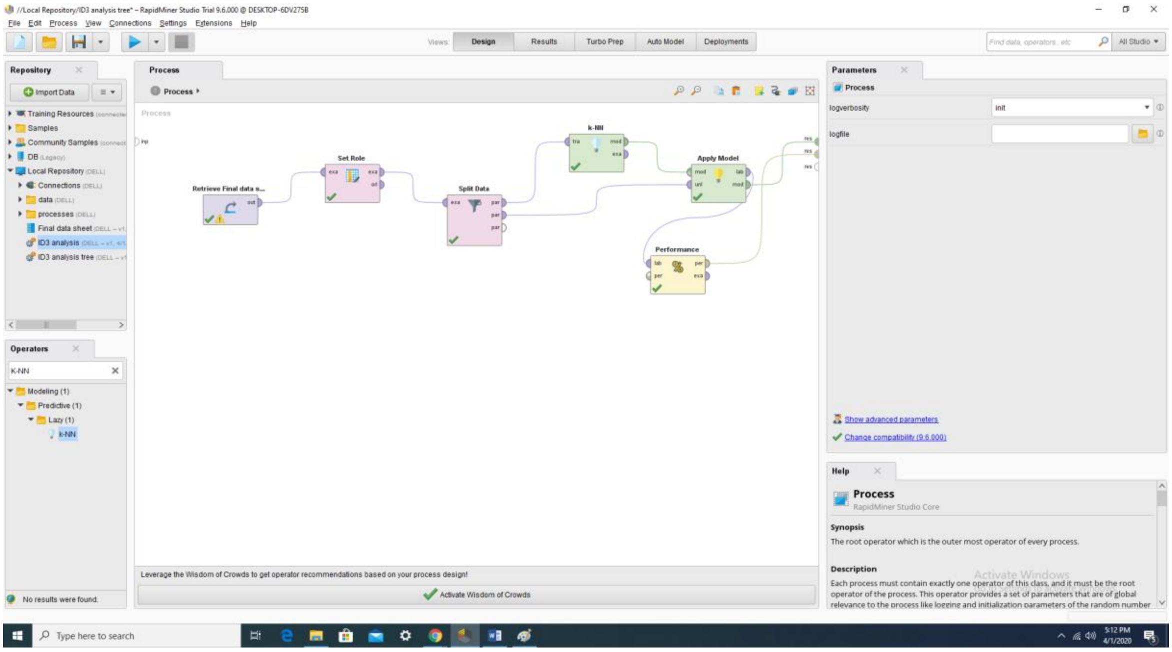
ID3	Actual						Class Precision
		B	A	C	S	F	
Prediction	B	**17**	6	6	1	0	56.67%
	A	8	**29**	1	0	0	76.32%
	C	1	1	**10**	1	0	76.92%
	S	1	0	2	**4**	0	50.00%
	F	0	0	0	0	**0**	0.00%
Class Recall		60.71%	80.56%	52.63%	66.67%	0.00%	

K-NN

K-nearest neighbors (KNN) algorithm is a type of supervised machine learning algorithm. It can be applied for both classification and regression predictive problems.

In this case set of data is separated into two fragments and 70% of data has been taken to create K-NN model. Then another 30% of data was used to apply the model.

K-NN algorithm was generated from a data set visualized as follows:


//Local Repository/ID3 analysis tree* – RapidMiner Studio Trial 9.6.000 @ DESKTOP-6DV275B
File Edit Process View Connections Settings Extensions Help
Views: Design Results Turbo Prep Auto Model Deployments
Find data, operators, etc
All Studio
Repository
Import Data
Training Resources
Samples
Community Samples
DB (Legacy)
Local Repository (DELL)
Connections (DELL)
data (DELL)
processes (DELL)
Final data sheet (DELL – v1
ID3 analysis (DELL – v1, 4/1
ID3 analysis tree (DELL – v1
Operators
K-NN
Modeling (1)
Predictive (1)
Lazy (1)
k-NN
No results were found.
Process
Retrieve Final data s...
Set Role
Split Data
k-NN
Apply Model
Performance
Parameters
Process
logverbosity
init
logfile
Show advanced parameters
Change compatibility (9.6.000)
Help
Process
RapidMiner Studio Core
Leverage the Wisdom of Crowds to get operator recommendations based on your process design!
Activate Wisdom of Crowds
Type here to search
5:12 PM
4/1/2020

Synopsis

The root operator which is the outer most operator of every process.

Description

Each process must contain exactly one operator of this class, and it must be the root operator of the process. This operator provides a set of parameters that are of global relevance to the process like logging and initialization parameters of the random number

When K-NN algorithm was run, confusion matrix was generated and it can be shown as follows. According to the K-NN algorithm, it could be predicted the class of 59 objects out of 89, and accuracy value is 66.29%.

Table 5.3.19 Confusion matrix of K-NN model

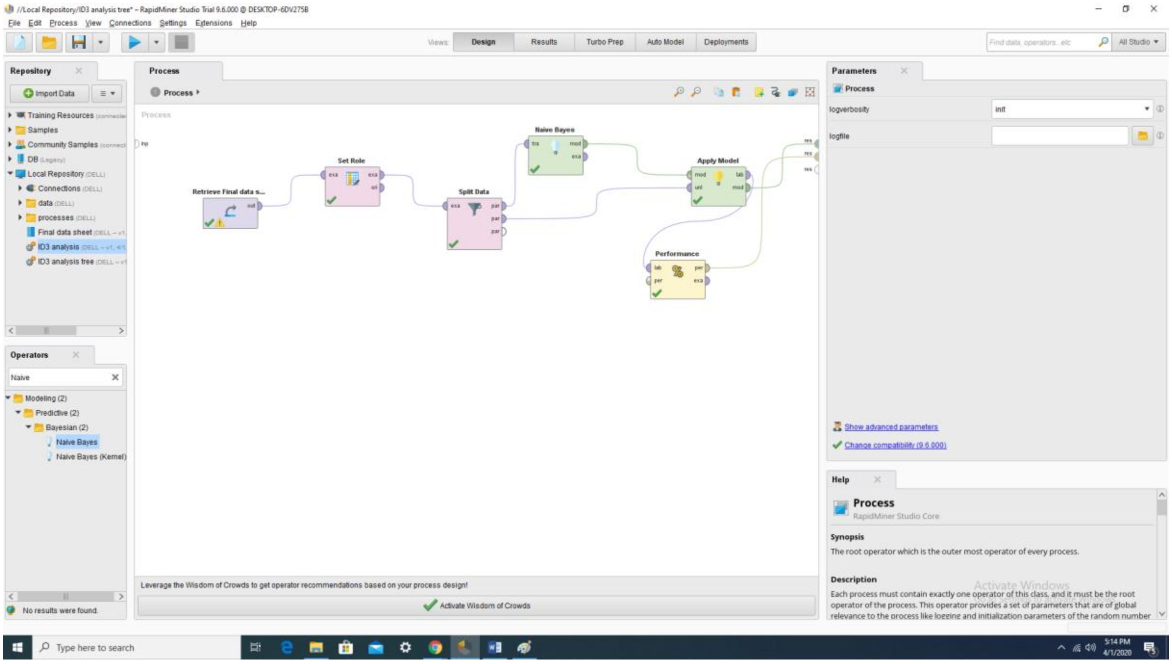
K-NN		Actual					Class Precision
		B	A	C	S	F	
Prediction	B	**14**	7	3	0	0	58.33%
	A	10	**29**	4	1	0	65.91%
	C	4	0	**12**	1	0	70.59%
	S	0	0	0	**4**	0	100.00%
	F	0	0	0	0	**0**	0.00%
Class Recall		50.00%	80.56%	63.16%	66.67%	0.00%	

Naive Bayes

A classifier is a machine learning model that can be used to separate different objects based on certain features.

In this situation data set has distributed into two portions and 70% of data has been taken to create Naïve Bayes model. Then another 30% of data was used to apply the model.

A Naïve Bayes was generated from a dataset visualized as follows:



Leverage the Wisdom of Crowds to get operator recommendations based on your process design!

The root operator which is the outer most operator of every process.

Each process must contain exactly one operator of this class, and it must be the root operator of the process. This operator provides a set of parameters that are of global relevance to the process like logging and initialization parameters of the random number

After running the Naïve Bayes algorithm, confusion matrix was generated and it can be shown as follows. According to the Naïve Bayes algorithm, it could be predicted the class of 54 objects out of 89, and accuracy value is 60.67%.

5.3.20 Confusion matrix of Naïve Bayes model

Naïve Bayes		Actual					Class Precision
		B	A	C	S	F	
Prediction	B	**16**	3	3	2	0	66.67%
	A	10	**28**	2	0	0	70.00%
	C	2	5	**7**	1	0	46.67%
	S	0	0	7	**3**	0	30.00%
	F	0	0	0	0	**0**	0.00%
Class Recall		57.14%	77.78%	36.84%	50.00%	0.00%	

Accuracy or sample accuracy of a model was taken based on “total correct predictions” divided by “total number of instances”. Altimetry it is observed that some algorithm predicted is better than others. Therefore, it appears that the highest accuracy is shown in ID3 classifier compared with others.

Table 5.3.21 Accuracy of the models taking all the attributes

Classification model	Accuracy
ID3	67.42%
K-NN	66.29%
Naïve Bayes	60.67%

I wanted to check that whether "Gender" attribute is directly impact to the performance of the students. So I removed "Gender" attribute from the data set and analyzed. At that time accuracy is as follows:

Table 5.3.22 Accuracy of the models when removing “Gender” attribute

Classification model	Accuracy
ID3	62.92%
K-NN	65.17%
Naïve Bayes	59.55%

The accuracy level was decreased in ID3, K-NN and Naïve Bayes models compared with the accuracy using whole data set. It seems that "Gender" attributes is direct affect to performance of the ICT subject.

I wanted to check that whether "reason of selecting ICT subject" attribute is directly impact to the performance of the students. So I removed “reason of selecting ICT subject” attribute from the data set and analyzed. At that time accuracy is as follows:

Table 5.3.23 Accuracy of the models when removing “reason of selecting ICT subject" attribute

Classification model	Accuracy
ID3	65.17%
K-NN	62.92%
Naïve Bayes	59.55%

The accuracy level was decreased in ID3, K-NN and Naïve Bayes models compared with the accuracy using whole data set. It seems that “reason of selecting ICT subject “attribute is direct affect to performance of the ICT subject.

I wanted to check that whether "Mathematics result of the GCE (O/L)" attribute is directly impact to the performance of the students. So I removed “Mathematics result of the GCE (O/L)" attribute from the data set and analyzed. At that time accuracy is as follows:

5.3.24 Accuracy of the models when removing "Mathematics result of the GCE (O/L)"

Classification model	Accuracy
ID3	61.80%
K-NN	52.81%
Naïve Bayes	51.69%

The accuracy level was decreased in ID3, K-NN and Naïve Bayes models compared with the accuracy using whole data set. It seems that performance of the Mathematics is direct affect to performance of the ICT subject.

I wanted to check that whether "English result of the GCE (O/L)" attribute is directly impact to the performance of the students. So I removed “English result of the GCE (O/L)" attribute from the data set and analyzed. At that time accuracy is as follows:

5.3.25 Accuracy of the models when removing "English result of the GCE (O/L)"

Classification model	Accuracy
ID3	66.29%
K-NN	57.03%
Naïve Bayes	52.81%

The accuracy level was decreased in ID3, K-NN and Naïve Bayes models compared with the accuracy using whole data set. It seems that performance of the English subject is direct affect to performance of the ICT subject.

I wanted to check that whether "hard of the paper” attribute is directly impact to the performance of the students. So I removed “hard of the paper “attribute from the data set and analyzed. At that time accuracy is as follows:

5.3.26 Accuracy of the models when removing” hard of the paper” attribute

Classification model	Accuracy
ID3	61.80%
K-NN	66.29%
Naïve Bayes	61.80%

The accuracy level was decreased in ID3 model compared with the accuracy using whole data set. The accuracy level was not changed in K-NN model compared with the accuracy using whole data set. The accuracy level was increased in Naïve Bayes model compared with the accuracy using whole data set. It seems that reason for the ICT paper is hard is affect to performance of the ICT subject when ID3 models was used. But we cannot state exactly whether reason for the ICT paper is hard is affect or not to performance of the ICT subject according to K-NN model. Reason for the ICT paper is hard is not affect to performance of the ICT subject when Naïve Bayes models was used.

I wanted to check that whether "participate tuition class" attribute is directly impact to the performance of the students. So I removed “participate tuition class “attribute from the data set and analyzed. At that time accuracy is as follows:

5.3.27 Accuracy of the models when removing” participate tuition class” attribute

Classification model	Accuracy
ID3	65.17%
K-NN	67.42%
Naïve Bayes	57.3%

The accuracy level was decreased in ID3 and Naïve Bayes model compared with the accuracy using whole data set. The accuracy level was increased in K-NN model compared with the accuracy using whole data set. It seems that participation of tuition class is affect to performance of the ICT subject when ID3 and Naïve Bayes models were used. But participation of tuition class is not affect to performance of the ICT subject when K-NN model was used.

I wanted to check that whether "getting practical knowledge" attribute is directly impact to the performance of the students. So I removed “getting practical knowledge “attribute from the data set and analyzed. At that time accuracy is as follows:

5.3.28 Accuracy of the models when removing” getting practical knowledge” attribute

Classification model	Accuracy
ID3	66.29%
K-NN	66.29%
Naïve Bayes	60.67%

The accuracy level was decreased in ID3 model compared with the accuracy using whole data set. The accuracy level was not changed in K-NN and Naïve Bayes models compared with the accuracy using whole data set. It seems that practical ICT knowledge is affect to performance of the ICT subject when ID3 models was used. But we cannot state exactly whether practical ICT knowledge is affect or not to performance of the ICT subject according to K-NN and Naïve Bayes models.

I wanted to check that whether "getting external reading" attribute is directly impact to the performance of the students. So I removed “getting external reading “attribute from the data set and analyzed. At that time accuracy is as follows:

5.3.29 Accuracy of the models when removing” getting external reading” attribute

Classification model	Accuracy
ID3	68.54%
K-NN	62.92%
Naïve Bayes	60.67%

The accuracy level was increased in ID3 model compared with the accuracy using whole data set. The accuracy level was decreased in K-NN model and the accuracy level was not changed in Naïve Bayes model compared with the accuracy using whole data set. It seems that getting external reading is not affect to performance of the ICT subject when ID3 models was used. It seems that getting external reading is affect to performance of the ICT subject when K-NN models was used but we cannot state exactly whether getting external reading is affect or not to performance of the ICT subject according to Naïve Bayes model. Finally, it is difficult to say whether getting external reading in addition to using the text book is affect or not to performance.

I wanted to check that whether "participated ICT review workshop" attribute is directly impact to the performance of the students. So I removed "participated ICT review workshop "attribute from the data set and analyzed. At that time accuracy is as follows:

5.3.30 Accuracy of the models when removing "participated ICT review workshop" attribute

Classification model	Accuracy
ID3	66.29%
K-NN	65.17%
Naïve Bayes	62.92%

The accuracy level was decreased in ID3 and K-NN models compared with the accuracy using whole data set. The accuracy level was increased in Naïve Bayes model compared with the accuracy using whole data set. It seems that participated in ICT review workshop is affect to performance of the ICT subject according to ID3 and K-NN models. But it is not affect according to Naïve Bayes model.

I wanted to check that whether "practical test" attribute is directly impact to the performance of the students. So I removed "practical test "attribute from the data set and analyzed. At that time accuracy is as follows:

5.3.31 Accuracy of the models when removing “practical test” attribute

Classification model	Accuracy
ID3	66.29%
K-NN	61.80%
Naïve Bayes	60.67%

The accuracy level was decreased in ID3 and K-NN models compared with the accuracy using whole data set. The accuracy level was not changed in Naïve Bayes model compared with the accuracy using whole data set. It seems that practical test in addition to the written test is affect to performance of the ICT subject according to ID3 and K-NN models. That mean if students faced a practical test in addition to the written test then they could increase their score. But it is not affect according to Naïve Bayes model.

I wanted to check that whether "ICT teacher” attribute is directly impact to the performance of the students. So I removed “ICT teacher “attribute from the data set and analyzed. At that time accuracy is as follows:

5.3.32 Accuracy of the models when removing “ICT teacher” attribute

Classification model	Accuracy
ID3	62.92%
K-NN	67.42%
Naïve Bayes	64.04%

The accuracy level was decreased in ID3 compared with the accuracy using whole data set. The accuracy level was increased in K-NN and Naïve Bayes models compared with the accuracy using whole data set. It seems that there is an ICT teacher at the school is affect to performance of the ICT subject according to ID3 model. But there is an ICT teacher at the school is not affect to performance of the ICT subject according to K-NN and Naïve Bayes models.

I wanted to check that whether "Family monthly income” attribute is directly impact to the performance of the students. So I removed “Family monthly income “attribute from the data set and analyzed. At that time accuracy is as follows:

5.3.33 Accuracy of the models when removing “Family monthly income” attribute

Classification model	Accuracy
ID3	62.92%
K-NN	67.42%
Naïve Bayes	62.92%

The accuracy level was decreased in ID3 compared with the accuracy using whole data set. The accuracy level was increased in K-NN and Naïve Bayes models compared with the accuracy using whole data set. It seems that monthly income of the family is affect to performance of the ICT subject according to ID3 model but it is not affected to performance of the ICT subject according to K-NN and Naïve Bayes models.

I wanted to check that whether "parent education” attribute is directly impact to the performance of the students. So I removed “parent education “attribute from the data set and analyzed. At that time accuracy is as follows:

5.3.34 Accuracy of the models when removing “parent education” attribute

Classification model	Accuracy
ID3	69.66%
K-NN	67.42%
Naïve Bayes	57.03%

The accuracy level was increased in ID3 and K-NN models compared with the accuracy using whole data set. The accuracy level was decreased in Naïve Bayes models compared with the accuracy using whole data set. It seems that parent’s education level is not affect to performance of the ICT subject according to ID3 and K-NN models. But parent's education level is affect to performance of the ICT subject according to Naïve Bayes models.

I wanted to check that whether "Laptop/Computer" attribute is directly impact to the performance of the students. So I removed "Laptop/Computer "attribute from the data set and analyzed. At that time accuracy is as follows

5.3.35 Accuracy of the models when removing "Laptop/Computer" attribute

Classification model	Accuracy
ID3	65.17%
K-NN	67.42%
Naïve Bayes	62.92%

The accuracy level was decreased in ID3 model compared with the accuracy using whole data set. The accuracy level was increased in K-NN and Naïve Bayes models compared with the accuracy using whole data set. It seems that a computer/laptop in a home is affect to performance of the ICT subject according to ID3 model. But a computer/laptop in a home is not affect to performance of the ICT subject according to K-NN and Naïve Bayes models.

I wanted to check that whether "an internet facility in the home" attribute is directly impact to the performance of the students. So I removed "an internet facility in the home" attribute from the data set and analyzed. At that time accuracy is as follows:

5.3.36 Accuracy of the models when removing "an internet facility in a school" attribute

Classification model	Accuracy
ID3	65.17%
K-NN	67.42%
Naïve Bayes	64.04%

The accuracy level was decreased in ID3 model compared with the accuracy using whole data set. The accuracy level was increased in K-NN and Naïve Bayes models compared with the accuracy using whole data set. It seems that an internet facility in the home is affect to performance of the ICT subject according to ID3 model. But an internet facility in the home is not affect to performance of the ICT subject according to K-NN and Naïve Bayes models.

I wanted to check that whether "Home support" attribute is directly impact to the performance of the students. So I removed "Home support "attribute from the data set and analyzed. At that time accuracy is as follows:

5.3.37 Accuracy of the models when removing "Home support" attribute

Classification model	Accuracy
ID3	67.42%
K-NN	66.29%
Naïve Bayes	60.67%

The accuracy level was not changed in ID3, K-NN and Naïve Bayes models compared with the accuracy using whole data set. It seems that we cannot state exactly whether getting home support is affect or not to performance of the ICT subject according to ID3, K-NN and Naïve Bayes model.

5.3.38 Summary of result after removing each attributes in the data set

Factors affect to students performance of the ICT subject	Affect or not to performance			Final decision
	ID3	K-NN	Naïve Bayes	
Gender	Affect	Affect	Affect	Affect
Result of Mathematics	Affect	Affect	Affect	Affect
Result of English	Affect	Affect	Affect	Affect

Factors affect to students performance of the ICT subject	Affect or not to performance			Final decision
	ID3	K-NN	Naïve Bayes	
Reason for the ICT paper is hard	Affect	Cannot decide	Not Affect	Cannot decide
Participated for the ICT tuition	Affect	Not Affect	Affect	Affect
Got practical training for the required topics	Affect	Cannot decide	Cannot decide	Cannot decide
Got external readings in addition to using the textbook	Not Affect	Affect	Cannot decide	Cannot decide
Participated in the examination review workshop on IT	Affect	Affect	Not Affect	Affect
Have a practical test in addition to the written test	Affect	Affect	Cannot decide	Affect
There is an ICT teacher at the school	Affect	Not Affect	Not Affect	Not Affect
Parents income	Affect	Not Affect	Not Affect	Not Affect
Parents Education level	Not Affect	Not Affect	Affect	Not Affect
Computer/Laptop facilities in the home	Affect	Not Affect	Not Affect	Not Affect
Internet facility in the home	Affect	Not Affect	Not Affect	Not Affect
Home support to learn ICT	Cannot decide	Cannot decide	Cannot decide	Cannot decide

CHAPTER 6

CONCLUSION AND RECOMENDATIONS

6.1 CONCLUSION

After the interviewed with officers of Ministry of Education, National Institute of Education, Department of Examination and Department of Publication, it was highlighted that below mentioned points were affected to the student's performance of GCE (O /L) ICT subject.

- Student's Gender
- Reason of selecting ICT subjects
- Student's performance for the Mathematics and English subject
- Reason for the ICT paper is hard
- Participation for the ICT tuition
- Participation in the examination review workshop on IT
- There is no practical test in addition to the written test
- Parent's Education level
- Family background
- School ICT teacher

After that a survey was done and data as such related academic personal and social were collected from GCE (O/L) ICT students. Questioner was created targeting above mentioned points. Then, the collected dataset was converted to suitable data mining tasks. Then data mining tasks was implemented to create classification model and testing. Appropriate results have been shown from the classification models.ID3 Decision Tree, K-nearest neighbor (K-NN) and Naïve-Bayes classifier have been used for my experiments. According to created model below mentioned points are most affect to the student's performance of ICT subject.

- Student's Gender
- Reason of selecting ICT subjects
- Student's performance for the Mathematics and English subject
- Participation for the ICT tuition
- Participation in the examination review workshop on IT
- Have a practical test in addition to the written test

Social effects such as parent's income, parent's education level and home support to learn ICT subject are not highly affects to the student’s performance of GCE (O /L) ICT subject according to this study.

I could not decide to below mentioned reasons are affect or not to ICT paper is hard.

- Missed the topics in ICT syllabus while in the teaching learning process
- Difficult to understand some topics to students
- No expert ICT teachers
- Students are lack of preparation for the exam

6.2 RECOMMENDATIONS

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

REFERENCES

[1] Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P., (1996). From data mining to knowledge discovery in databases. AI magazine, 17(3), p.37.

[2] Berland, M., Ryan|Blikstein,Paulo,(2014). Educational Data Mining and Learning Analytics: Applications to Constructionist Research. Technology, Knowledge and Learning, 19(1–2), pp.205–220.

[3] International Educational Data Mining Society, (2011). International Educational Data Mining Society.

[4] Abdous, M., Wu, H. & Yen, C.-J., (2012). Using data mining for predicting relationships between online question theme and final grade. Journal of Educational Technology & Society, 15(3), p.77.

[5] Mousa, H. & Maghari, A., School Students' Performance Predication Using Data Mining Classification. International Journal of Advanced Research in Computer and Communication Engineering, Vol. 6: issue 8, 2017

[6] Pal, A.K. & Pal, S., Analysis and Mining of Educational Data for Predicting the Performance of students. International Journal of Electronics Communication and Computer Engineering, Vol.4: issue 5.

[7] Baradwaj, B.K. & Pal, S., Mining Educational Data to Analyze Student's Performance. (IJACSA) International Journal of advanced Computer Science and Applications, Vol.2, No.6, 2011

[8] Saa, A.A., Educational Data Mining & Student's Performance Prediction. (IJACSA) International Journal of Advanced Computer Science and Applications, Vol.7, No.5, 2016

[9] Premarathne, P.N.W.A.L.K., Using Data Mining Techniques for Investigating of Performance in ICT subject at G.C.E Advanced Level. (ICTer) International Conference on Advances in ICT for Emerging Regions, 2018

[10] Fernando, M.G.N.A.S., Ekanayake M.B., Sustainable Quality Improvement of ICT Education in secondary school curriculum of Sri Lanka (2018).

[11] Ilumudeen, A., Importance of Information and Communication Technology (ICT) curriculum in Government school of Sri Lanka: A critical Review of Educational Challenges and Opportunities. South Eastern University of Sri Lanka

APPENDICES

Appendix- A

Analysis and Mining of Educational Data for Predicting the Student's Performance of Information & Communication Technology Subject in GCE (O/L) Examination

(Research conduct for MSc in Information Technology ,Faculy of Information Technology,University of Moratuwa)

01. Gender: Female □ Male □

02. Why did you select ICT subject for the O/L Examination?

a. Consent □
b. Request of parents □
c. Request of teachers □
d. The social demand □
e. Other □

03. What was the grade obtained for O/L Mathematics?

a. A □ d. S □
b. B □ e. F □
c. C □

04. What was the grade obtained for O/L English?

a. A □ d. S □
b. B □ e. F □
c. C □

05. What was the grade obtained for O/L ICT?

a. A □ d. S □
b. B □ e. F □
c. C □

06. Why do you think that ICT paper is hard?

a. missed the topic □
b. couldn't understand some topic □
c. no expert ICT teacher □
d. lack of preparation for the exam □

07. Did you participate ICT tution classes?

a. yes □ b. no □

08. Did you get practical training for the required topics?

a. yes □

b. sometimes □

c. never □

09. Did you get external readings in addition to using the textbook?

a. yes □ b. no □

10. Have you participated in the GCE O / L Examination Review Workshop on Information Technology?

a. yes □ b. no □

11. Can you increase your score if you have a practical test in addition to the written test?

a. yes □

b. no □

12. Is there an ICT teacher at the school?

a. yes □ b. no □

13. What is your monthly family income?

a. more than 200000 □

b. 200000 - 100000 □

c. 100000 - 50000 □

d. 50000 – 25000 □

e. less than 25000 □

14. What is the level of your parent's education?

a. post graduate □

b. degree □

c. diploma □

d. A/L □

e. O/L □

15. Have a computer/laptop in your home?

a. yes □ b. no □

16. Have an internet facility in your home?

a. yes □ b. no □

17. How to get support in the home to learn this subject?

a. parents or other adults at home express the theory part □
b. experience in observing people in the ICT field or further education at the home
c. parents or other adults at home giving practical training □
d. provide physical recourses □
e. no support □