A DATA MINING APPROACH TO IDENTIFY THE FACTORS AFFECTING FOR THE ICT RESULTS IN G.C.E ORDINARY LEVEL EXAMINATION

P.G.P.S. Weerasooriya[1], K.K.P.L. Kumara[1], W.L.S.Kumasaru[1], K.B.G.S.G. Kariyawasam[1] and Niroshi. P[1]

[1]Department of ICT, Faculty of Technological Studies, University of Vavuniya

Abstract

he main aim of the paper is to identify the factors affecting the Information Communication Technology results in the G.C.E. Ordinary Level examination in Sri Lanka. The dataset was collected from the students who are schooling in the Northern and Western provinces and faced to the G.C.E. Ordinary Level examination in between year 2015 and 2019. The factors considered are gender, type of the school, availability of enough and proper human resources to teach ICT, availability of laboratory facilities, details of the family background, availability of computer facilities at home, results for the English subject, and attending for the private classes. The data set is clustered using the K-Means algorithm and applied the Apriori algorithm on both the entire data set and each cluster. The result of this study showed that poor English knowledge, family background, provinces, and the type of school affect the ICT results of the students.
Keywords: Apriori algorithm, Data Mining, K-Means algorithm

1 INTRODUCTION

Information Communication Technology (ICT) is the infrastructure and component which enables modern computing. It has become an important aspect of most countries and has huge importance for the present and as well as for the future. Upcoming lifestyles of people are bound with ICT as most of the activities in day-to-day life have been already automated. To develop ICT education, rapid, planned, and sustainable ICT education programs need to be implemented in the general education system. For this purpose, the Sri Lankan government has taken several steps to provide equipment, infrastructure facilities, and teacher training to schools. But, the result of students at the national level examinations shows that the supports provided by the government have not improved the quality of ICT education to expected standards. It has been noticed that not only the human resources and infrastructure facilities but also some other factors such as family background, English knowledge, etc. could be reasons for this performance difference. But according to the literature found and best of our knowledge, there are no any studies have been carried out to identify the factors affecting the students' performances of ICT subjects at the national level examinations by using advanced analysis techniques[1] . The aim of this research is to identify the causative factors and the association among those factors which may affect the students' performances in G.C.E. (O/L) ICT subject by scanning

the individual students' examination results, family background, and the facilities available in schools with regards to the infrastructure and human resources. Specific objectives of the study are suggesting the quality indicators which could improve the results of the ICT students, educating on how to identify the weak students in ICT, spot the students who are at risk of failing the ICT subject at the G.C.E. (O/L) examination. The research is based on the discipline of Data Mining and the data has been collected from two different provinces Northern and Western through a questionnaire. The analysis has been carried out against the students who faced the G.C.E. (O/L) ICT subject in Sri Lanka from the year 2015 to 2019.

2 LITERATURE REVIEW

A review of the literature regarding the study reveals the following findings. Premarathne and Premarathne[2] had investigated the performances in ICT subject at the G.C.E, (A/L) examination. The main objective of this research is to identify the problems relating to the G.C.E. (A/L) Examination in ICT subject and provide suggestions for enhancing students' performances and the teaching process by using the association rule mining technique. Fernando [1] studied the Quality of ICT education at G.C.E. (O/L) in Sri Lankan schools. This study focuses on improving the quality of ICT education in Sri Lankan schools and suggests the necessary enhancements to be taken to achieve international standards. Using the descriptive and inferential statistical technique. Ijmuiden[3] also investigated the performances in ICT subjects in ICT education in Sri Lanka. This research's main objective is to identify the opportunities for ICT implementation. This research highlights; lack of a computer laboratory, cost of devices, lack of ICT skilled teachers, less motivation on ICT among students, high cost of implementation and maintenance of computer laboratories, language barrier, unavailability of national-level certification for ICT as the challenges. Basri, Jehan. Alandejani, Feras, and Almadani[4] discussed the adoption of ICT by the universities and the impact that it makes on the university students' academic performances through customized Learning Management Systems (L.M.S.). They mainly focused on university students.

3 METHODOLOGY

Our primary data set for the analysis was collected through a questionnaire from the school students who faced G.C.E. (O/L) examination from the year 2015 to 2019 in Northern and Western provinces. The questionnaire was prepared according to past scholars' questions which are measured validity and reliability. The data set consists of the 1100 students' details which is contained the details of the attributes province, gender, type of the school, availability of enough and proper human resources to teach ICT, availability of laboratory facilities, details of the family background, availability of computer facilities at home, results for the English subject, attending the private classes for ICT and the G.C.E. (O/L) ICT results of the students for ICT. The data set contains only categorical data. The values of each attribute are given in Table 01. Since the students' performances in ICT subjects are affected by various elements, sophisticated analysis techniques are necessary to identify the factors and associations among those. The complexity associated with the content of the data sets is high as well. Therefore, it was decided to apply the Data Mining techniques using Weka Library.

   Data Preprocessing As the data set is manually entered it may consist of incomplete and missing values. Therefore, data preprocessing techniques need to be used before applying the Data Mining algorithms. Data Analyzing: Clustering Clustering is also called segmentation. Since the clustering algorithm can identify the different groups in a dataset, it was decided to conduct the cluster analysis

Table 1. Details of Attributes of the Data Set

Attribute	Values
Province	Northern Province, Western Province
Gender	Male, Female
Types of the school	MahaVidyala, Central College, National College, International School
Permanent ICT teachers	Yes, No
ICT Laboratory	Yes, No
Enough Computers available in the lab	Yes, No
Parents Education Level	Grade 1 to Grade 8, O/L and A/L, Graduate or Diploma, Other
Parents Occupation	Self-employed, Labor, Executive, Top-level employee
Computer available at home	Yes, No
Marks of English	Less than 20, 21 – 45, 46 – 65, 66 – 85, More than 86
Private classes	Yes, No
Grade	A, B, C, S, W

Figure 1. The Division of the Students' Results in the Entire Data Set
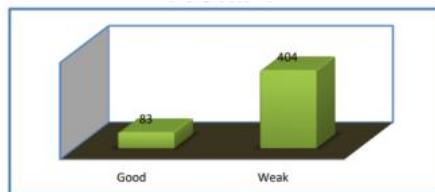

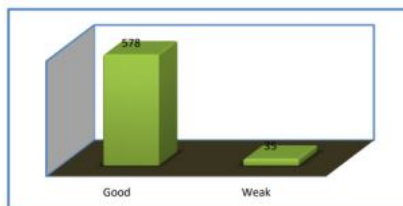**Figure 2.** The Division of the Students' Results in the Cluster 0

Figure 3. The Division of the Students' Results in the Cluster 1

to group the data set. Clustering based on the Simple K-Means clustering algorithm in WEKA is used. The biggest challenge with K-Means clustering is to find the most optimal number of clusters. The Elbow finding technique was decided to use for this purpose [4]. The elbow method uses to determine the number of clusters in a data set. To identify the associations between attributes of the data set, Apriori algorithm will be used on each cluster and the entire data set since if the Apriori algorithm is applied only on the entire data set, the most important rules can be hidden. Therefore, applying the Apriori algorithm on each cluster might be able to reveal additional hidden rules. As the next step, the K-Means algorithm is applied to the entire data set to divide the data set into two clusters. So, at the end Cluster 0 and Cluster 1 consists of 487 and 613 data respectively. To categorize the students based on their G.C.E. (O/L) ICT results, an additional column has been added to the data set as "Result" and calculate the result as follows. if Grade = "A" or "B" or "C" then Result = "Good" else Result = "Weak" Figure 01, Figure 02, and Figure 03 show the division of the student's results in the entire data set, Cluster 0 and Cluster 1. The graph clearly shows that the majority of Cluster 0 students are having "Weak" results while the entire data set and the Cluster 1 are having more "Good" results.

Data Analyzing: Association Rule Mining Association rule mining algorithm is a Data Mining technique used to extract associations among a large set of items. Association Rule Mining based on the Apriori algorithm in WEKA is used to discover the associations. The support for the analysis is varied from 1.0 to 0.1 and the Confidence for the analysis is set at 1. The Apriori algorithm is applied to the entire data set, Cluster 0 and Cluster 1 Separately. The obtained rules are shown in Table 02. The entire data set, Cluster 0 and Cluster 1 produced 7, 5, and 7 rules respectively.

Table 2. Cluster-wise Association Rules

Cluster No	Rule body	Support	Lift
Entire data set	66-85, Central College =¿ Good	0.1099001	1.665658
	66-85, Western Province =¿ Good	0.1407811	1.665658
	66-85, O/L And A/L =¿ Good	0.102634	1.665658
	Graduate Or Diploma,Western Province =¿ Good	0.1743869	1.665658
	Executive,More Than 86 =¿ Good	0.1171662	1.665658
	Female,More Than 86 =¿ Good	0.1226158	1.665658
	Central College,Female,Western Province =¿ Good	0.1462307	1.665658
Cluster 0	Central College,Labour =¿ Weak	0.1393443	1.207921
	Grade 1 To Grade 8, Less Than 20 =¿ Weak	0.1188525	1.207921
	Less Than 20, Self Employed =¿ Weak	0.1270492	1.207921
	46-65, Labour =¿ Weak	0.1372951	1.207921
	21-45, O/L And A/L =¿ Weak	0.1557377	1.207921
Cluster 1	Top Level Employee =¿ Good	0.1775244	1.062284
	National College =¿ Good	0.1986971	1.062284
	66-85 =¿ Good	0.2785016	1.062284
	Graduate Or Diploma =¿ Good	0.3289902	1.062284
	Female,More Than 86 =¿ Good	0.2198697	1.062284
	Central College,Female =¿ Good	0.2736156	1.062284
	Executive,More Than 86 =¿ Good	0.2100977	1.062284

4 RESULTS

Association rules for entire data set show the attributes which affect a "Good" result. According to the rules found, the students who earned more than 66 marks for English are having a higher possibility to earn a good result for ICT in O/L. And also, the students who are studying in a central college or western province have a chance to get a good result. Association rules for cluster 0 show the details of the students who could get a "Weak" pass for their ICT O/L result. The rules show that if the students' parent's occupation level is comparatively low (labor or self-employed.) or education level is poor then students' results could be low as well. Furthermore, poor English knowledge also affects the lower the students' results. 5 Association rules for cluster 1 show the details of the students who could get a "Good" pass for their ICT O/L result. The rules are scattered in almost all the attributes and the rules show that if the parents' occupation is comparatively high or parents' educational level is high or if the student is from a National or Central college or if the English result is more than 66, then the student may able to earn a good result.

5 DISCUSSIONS

The main factor identified through this study is that the impact of English knowledge on the results of the ICT subject. Good English knowledge would increase the ICT results and poor English knowledge would decrease it. ICT is mostly based on the English language and most of the ICT materials are also in English. Although English is teaching as a secondary language from grade 3 to 11, most of the students are not familiar with the language. This makes students uncomfortable learning ICT as well. The results show that Family background also has a huge impact on the ICT results. Family socioeconomic and economic status has an important impact on the quality of the educational opportunities that students have access to. The higher the family's status, the higher the qualities of the student's educational opportunities attend. If the parents' economic status is not good enough, they are not capable enough to buy ICT resources such as computers or laptops for the students. Furthermore, if the parents are not educated and if they have no knowledge about ICT then they are unable to direct their children towards new technologies. According to the results, the province and the type of the school highly affect the students' ICT results. The schools situated in rural areas do not fulfill their most essential needs for teaching ICT subjects such as IT laboratory facilities, teachers, and other resource persons. Not only in rural areas, as a country in Sri Lanka, there is also an issue with the lack of ICT human capital. Although the teachers are trained in many ICT programs, they were not properly utilized due to many reasons. One of the reasons is that there is no specific cadre assign for ICT in some schools. And also, in most cases, the teachers in rural areas are not regularly updated with the ongoing technologies related to ICT.

6 CONCLUSION

Discussion The main factor identified through this study is that the impact of English knowledge on the results of the ICT subject. Good English knowledge would increase the ICT results and poor English knowledge would decrease it. ICT is mostly based on the English language and most of the ICT materials are also in English. Although English is teaching as a secondary language from grade 3 to 11, most of the students are not familiar with the language. This makes students uncomfortable learning ICT as well. The results show that Family background also has a huge impact on the ICT results. Family socioeconomic and economic status has an important impact on the quality of the

educational opportunities that students have access to. The higher the family's status, the higher the qualities of the student's educational opportunities attend. If the parents' economic status is not good enough, they are not capable enough to buy ICT resources such as computers or laptops for the students. Furthermore, if the parents are not educated and if they have no knowledge about ICT then they are unable to direct their children towards new technologies. According to the results, the province and the type of the school highly affect the students' ICT results. The schools situated in rural areas do not fulfill their most essential needs for teaching ICT subjects such as IT laboratory facilities, teachers, and other resource persons. Not only in rural areas, as a country in Sri Lanka, there is also an issue with the lack of ICT human capital. Although the teachers are trained in many ICT programs, they were not properly utilized due to many reasons. One of the reasons is that there is no specific cadre assign for ICT in some schools. And also, in most cases, the teachers in rural areas are not regularly updated with the ongoing technologies related to ICT.

REFERENCES

[1] M. G. N. A. S. Fernando. *I Jesrt Quality of Information and Communication Technology (Ict) Education At Gce (O / L) in Sri Lankan Schools*. November 2015.

[2] P. N. W. A. Premarathne and S. C. Premaratne. *Using data mining techniques for investigating of performance in ICT subject at G.C.E advanced level*. 2019.

[3] I. Aboobucker. *Importance of Information and Communication Technology (ICT) Curriculum in Government School of Sri Lanka: A Critical Review of Educational Challenges and Opportunities*. December, 2019.

[4] Feras M.Basri W. S. Almadani and J. A. Alandejani. *ICT Adoption Impact on Students' Academic Performance: Evidence from Saudi Universities*. December, 2019.