International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 10 (2018) pp. 154-163

Received: 23 Dec, 2017, Accepted: 2 March, 2018, Published: 11 July, 2018

Analyzing Children's Data Using Machine Learning: A Case Study in Ethiopia

Gebremedhin Gebreyohans[1] and Niketa Gandhi[2]

[1] IT Doctoral Program, Information Systems track, Addis Ababa University,
Addis Ababa, Ethiopia
gereisms12@gmail.com

[2] Machine Intelligence Research Labs (MIR Labs),
Scientific Network for Innovation and Research Excellence,
P.O. Box 2259, Auburn, Washington 98071, USA
niketa@gmail.com

***Abstract*: This research focuses on classification of children into four classes namely orphan, single orphan, vulnerable and safe. The aim of this classification is to help the outside donors of the Love for Children Organization and further to get full information about each child for internal purpose of the organization. To achieve this three classification techniques were used which are Decision tree, Bayesian learning and Neural network within the framework of KDD (Knowledge Discovery in Databases) data mining model. The children dataset was collected, cleaned, transformed and integrated for experimenting with the classification model. The final dataset consists of 17044 records that have been experimented and evaluated against their performances. The collection of dataset was experimented with the 10-fold cross-validation and splitting the datasets in to 70/30%, 66/44% and 50/50% for training/and for testing respectively. Additionally, a comparison of decision tree (98.83 %,), bayesian learning (98.32%) and neural network (98.86%) model in terms of the overall classification accuracy and their advantage was made. The research concludes that decision tree (98.83%) should be selected as a model because it gives better results than Bayesian learning (98.32%) and better advantage over Neural Network (98.86) for the classification of organizations children.**

***Keywords*:** application of data mining, children data set, data mining, data mining techniques, decision tree, ethiopia, KDD.

I. Introduction

It is true that every child needs safety to learn, play and grow and a time to develop into the adults who will one day care for and lead our country, our world, and our future. Yet hundreds of millions children around the world are in need of help. According to the World Health Organization (WHO) [1], the issues that children in Ethiopia face are some of the most challenging in the world. Even in an average year, the education, health and economic situation for millions of Ethiopian children can only be described as a crisis. More than 99.3 Million people lives, out of this about 35% of girls and boys are out of school 59 out of 1000 children die before their 5th birthday [1].

Similarly, according to UNICEF report on issues of child mortality, a child born in developing country is over 13 times more likely to die within the first five years of life as compared to economically advanced countries [2]. Sub-Saharan Africa accounted for about half of these deaths in the developing world. Surprisingly the causes of illnesses and admissions to hospitals that utilized the scarce resource in the region were diseases that can be easily prevented. Child illnesses and deaths were higher for children from rural and poor families and whose mother lack basic education. An Ethiopian child is 30 times more likely die by his or her fifth birth day than a child in Western Europe [2], [3].

For this, there are a number of organizations who are providing help for the needy children. Including Love for Children Organization (LCO) is non-governmental Organization. Currently, performing child selection by consulting psychologists and social workers to decide whether the child needs help or not. This creates tremendous human involvement, ambiguity, and unnecessary cost of the consulting personnel to the organization. This in turn results in the possibility of allocating fund for non-needy children. Hence, an automated system can help facilitate and simplify the process of identifying and selecting children who are in need of help.

In the Information Technology era information plays vital role in every sphere of the human aspects. Thus, to efficiently inspire information, it is very important to generate information from massive collection of data. The data can range from simple numerical figures and text documents to more complex information such as spatial data, multimedia data, and hypertext documents [4]. However, the huge size of these data sources makes it impossible for a human experts to come up with interesting information or patterns that will help in the proactive decision making process. Therefore, to take complete advantage of data; the data retrieval is not enough. It requires a tool for automatic summarization of data, extraction of the essence of information stored, and the discovery of patterns in raw data. This tool is data mining (DM) [4].

In the past lot of work has been done to assess the application of DM in different sectors like Airlines, Banking, Insurance, HealthCare, and to the knowledge of the researcher there are only few attempts in our country that have been done so far towards the application of DM in the children related field [5] [6]-[11]. Researchers conducted the application of DM in predicting child mortality in Ethiopia as a case study in the Butajira Rural Health Project. Researcher employed the classification technique, neural network and decision tree algorithms to develop the model for predicting child mortality. Another research by [12] also tried to study the application of DM technology to identify significant patterns in census or survey data as a case of the 2001 child labor survey in Ethiopia. Researchers applied the association rule DM technique and the Apriori algorithm for identifying relationships between attributes within the 2001 child labor survey database that was used to clearly understand the nature of child labor problem in Ethiopia. Another research undertaken by [13], tried to apply the DM techniques for street children of Ethiopia. Apriori algorithm was used for Association rule mining. These researchers [5], [12], [13] applied the different DM techniques in support of children classification. As per the knowledge of the researcher till now there is no work that have been done so far regarding the application of DM in orphan and vulnerable children in a children related organizations in Ethiopia. Hence this study has a great contribution in applying DM technology for the purpose of children classification to different categories in children support organization to get funds from different donors to help those vulnerable children. The aim of the study is to classifying children dataset with incorporated data mining and it will be useful for decision making purpose of the organization. Classifying children as orphan or vulnerable based on the children's health condition and status of parents to give information for the organization provides funds for children from different segments of the community who are assumed to be needy. Having this, the main significance of the study lies in making a contribution on the concept of DM in LCO and to create a general awareness among the Organization members. Finally, it will show how DM can implement using classification techniques has a direct impact on children classification problems.

Although, this research complements with the research work discussed above done by other researchers, but it differs in the area of application, variables used for model building, the theme it contains, and data mining methods.

Thus, in the present research an attempt is made to address the following research questions:-

- What knowledge is suitable for DM in classification of children to acquire and modeled data mining algorithms?
- Which DM techniques (Decision tree, neural network or Bayesian Classification) produce a better classification model for mining children data set?

To address these research questions, the overall objective for this study is, to acquire domain knowledge using interview and document analysis, i.e., to extract the data, clean and transform the data into the format suitable for mining and to develop a model that helps the organization identify different patterns of children data through the application of classification techniques so that identification of different classes of children can be automated (i.e., to decide whether a child is vulnerable, orphan, single orphan or safe based on the given data). Depending on the result of the classifier, the organization can decide whether to provide or deny fund to a particular child that applies for getting fund from the organization.

According to WHO [1], the definition of orphans and vulnerable children are 'children who are compromised as a result of the illness or death of an adult who contributed to the care and/or financial support of the child'. This could be said with the children that are found in LCO compound as those who have lost either one of the parents. Vulnerability is stated according to the poverty levels in LCO compound. LCO takes its activities in collaboration with local administration (Kebele, Sub-city), community based organization, volunteers and other likeminded organizations. For this purpose, the researchers have tried to build a model that classifies and segments the data items in a way that is suitable for decision making. As a scope this research limits to: Three different approaches of classification techniques which are: Naïve Bayes, Decision Tree and Neural Network. Clustering and association rule mining schemes were not used.

II. The Knowledge Discovery Process

The term Data Mining or Knowledge Discovery in databases has been adopted for a field of research dealing with the automatic discovery of implicit information or knowledge within databases [2], [14], [15], [16]. Knowledge Discovery in Databases (KDD) is the process of identifying useful information in data [16], [17]. A widely accepted formal definition of data mining is given subsequently. According to this definition, data mining is the non-trivial extraction of implicit previously unknown and potentially useful information about data [18].

In order to define and analyze the business problem properly, the primary data was collected by interviewing concerned officers (Experts) in the organization. The offices that the researcher has been conducted the survey are the database of the LCO. Then based on the information obtained from these attempts, the overall children classification process has been done. The model employed in this research is KDD process which consists of five phases - data selection, preprocessing, transformation, data mining and interpretation/evaluation as shown in Figure 1.

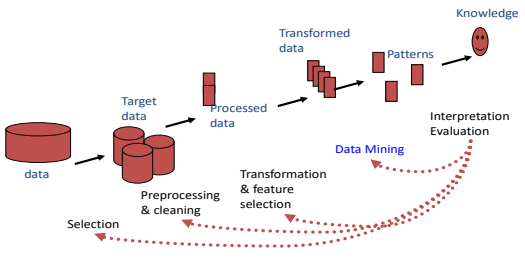

Figure 1. KDD process [Adopted from (Fayyad et. al. (1996)]

KDD is often used as a synonym for DM. KDD is the process of finding useful information and patterns in data .KDD can be define as the whole process involving: data selection; data pre-processing: cleaning; data transformation; mining; result evaluation and visualization [15], [19]. Data mining, on the

other hand, refer to the modeling step using the various techniques to extract useful information/pattern from the data. Therefore, DM is a one step in KDD which is the use of algorithms to extract hidden patterns & knowledge in data. KDD process steps which are employed in this study are summarized as follows:

III. Understanding the data

After getting familiar with the problem domain, the data used for the paper is obtained from LCO found in Addis Ababa-Ethiopia, sub city Mexico. The office keeps records of Children on a centralized manner using MS-Access. The data to be mined was collected and arranged into a new database to make it suitable for the experiment and for the selected data mining tool. A new database was prepared by analyzing the collected data using data preprocessing tasks like data cleaning, data reduction and data transformation.

The initial dataset is collected from the LCO the data set selected comprises child data the organization has been using from 2000 to 2011. The data has been imported from the Microsoft Access database of the organization to Microsoft Excel for processing. Several preprocessing methods have been applied to get the relevant data for the system. The data has more than 20 attributes like (Child _Name', 'House _Number', 'Telephone _Number', Sex, Age, Status _of _ Natural _Parents, Died _Parent, Guardian _Occupation, Cause _of _Parent _Death, Guardian _Relation _with _child, Child_ Duties _at_ Home, Child _Health_ Condition, Child _Grade_ Level, Child _Type) and are total of more than 20000 records. The records were stored in 12 different tables and later, merged into one large table. The data set includes not only the children who have got fund but also whose application has been rejected. But, the data which are used for analyzing the specified case are not complete. Some data has been removed from the database because it is not that much significant in the final result. Such as, children phone_ number, house_ number and child _name since some of them are redundant or irrelevant. In addition, the researcher tried to do the data cleaning such as handling missing attributes values and noise removal before deriving the attributes that are used to build the classification model. Accordingly, an attribute 'Died_ Parent', was not complete for those children whose both parents are alive. It was handled by replacing with the user-defined value 'None'. Finally, a total of 11 attributes out of the original 20 that best suit the objectives and out the original 20,000 instances 17044 records have taken for analysis; remaining non-relevant records have been removed. These are: Sex, Age, Status _of _ Natural _Parents, Died _Parent, Guardian _Occupation, Cause _of _Parent _Death, Guardian _Relation _with _child, Child_ Duties _at_ Home, Child _Health_ Condition, Child _Grade_ Level, Child _Type. This dataset is sufficient enough to perform some data mining techniques to get the desired result.

After this the researcher has selected appropriate tool and dataset formats. Even though there are several data mining tools that may fulfill the objectives, techniques and tasks the researcher used for this, Weka 3.9.2 data mining and knowledge discovery tool was used. The obvious advantage of a package like Weka is that a whole range of data preparation, feature selection and data mining algorithms are integrated [20]. This means that only one data format is needed, and trying out and comparing different approaches becomes really easy [21]. Since the data mining software used to generate classification (WEKA) accepts data only in ARFF format, the researcher first converted the data on MS Excel file into comma separated text format (.csv) format which is a format where commas are placed between values in adjacent columns.

IV. Experimental Results and Analysis

This part shows the steps and procedures followed during experimentations and discovering regularities for predicting children classification within LCO dataset.

The algorism selected for classification purposes were J48 decision tree, naive bayes and neural network those can classify an instance in to already identify classes. The researcher tested the algorism with different algorisms and record numbers to improve the classification accuracy. Accuracy refers to the percentage of correct predictions made by the model when compared with the actual classifications. Finally, compared and selected the best classification model from the three algorisms. The classification accuracy of each of these models is reported and their performance is compared in classifying new instances of records.

A. J48 Decision tree Model building

Decision tree was introduced by J. Ross Quinlan, who is a researcher in machine learning and developed a decision tree algorithm known as ID3 [22]. Decision tree method is widely used in data mining and decision support system [2], [14], [15], [16], [18], [23]. This algorithm is fast and its representation is easy to understand for rule generation as well as for classification problems. It is an excellent tool for decision representations [20]. The accuracy of a classifier refers to the ability of a given classifier to correctly predict the class label of new or previously unseen data. The indicators in Decision tree are algorithms that are built automatically from a given dataset. Typically the goal is to find the optimal decision tree by minimizing the generalization error [20], [24]. The decision is a flow chart like structure, where each internal node denotes a test on an attribute, each branch of the tree represents an outcome of the test and each leaf node holds a class label [2], [25], [26].

At each node it will be sent either left or right according to some test. Eventually, it will reach a leaf node and be given the label associated with that leaf. As described before, the J48 algorithm is used for building the decision tree model. J48 algorithm contains some parameters that can be changed to further improve classification accuracy. Initially the

Table 1. J48 algorithm parameters and their default values

Parameter	Description1	Default Vakue
Confidence Factor	The confidence factor used for the pruning (smaller values incur more pruning)	0.25
minNumObj	The minimum number of instance per leaf	2
Unpruned	Whether pruning is performed	False

classification model is built with the default parameter values of the J48 algorithm. Table 1 summarizes the default parameters with their values for the J48 decision tree algorithm.

By changing the different default parameter values of the J48 algorithm, the experimentations of the decision tree model-building phase are carried out.

1) Experiment 1

The first experimentation is performed with the default parameters. The default 10-fold cross validation test option is employed for training the classification model. Using these default parameters the classification model is developed with a J48 decision tree having 20 numbers of leaves and 27 tree size. Table 2 depicts the resulting confusion matrix of this model.

As shown in the confusion matrix in table 2, the J48 learning algorithm scored an accuracy of 98.7972%, which indicates that, out of the total number of records supplied, 16839 (98.79 %) records are classified correctly and 205 (1.21%) are misclassified or incorrectly classified. Furthermore, the resulting confusion matrix of this experiment has shown 100% of the records are correctly classified in the orphans and safe which indicates, the algorithm classified the entire orphan and safe in their respective class and out of the 7323 vulnerable children, who are described in vulnerable, 7186 (98.13%) of them are classified correctly in their designated class, i.e. vulnerable, while only 39 (0.0053 %) of them are misclassified in single _orphans and 98(0.013%) of them are misclassified in safe. In addition to this out of the 4615 single orphan children, 4545 (98.48 %) of them are classified correctly in their designated class, i.e. single orphan, while only 70(0.015 %) of them are misclassified in vulnerable. As described before, the size of the tree and the number of leaves produced from this training was 27 and 20 respectively. Therefore, to make ease the process of generating rule sets or to make it more understandable, the researcher attempted to modify the default values of the parameters so as to minimize the size of the tree and number of leaves. With this objective, the minNumObj (minimum number of instances in a leaf) parameter was tried with 25, 20, 15, 10 and 5. But the minNumObj set to these it doesn't give a better tree size and accuracy compared with the other trials. Means their value has not that much difference compared with the first one. That means the complexity of the decision tree to generate rules is the same in both the experiments. So, since there is no a tangible difference in the tree size and number of leaves in this experiments the accuracy of the model is 98.79% so, in this experiment with the default minNumObj parameter value is taken as the J48 decision tree model.

2) Experiment 2

This experiment is performed, by changing the default testing option (the 10-fold cross validation). In this learning scheme a percentage split is used to partition the dataset into training and testing data. The purpose of using this parameter was to assess the performance of the learning scheme by increasing the proportion of testing dataset if it could achieved a better classification accuracy than the first experimentation. First this experiment has run with the default value of the percentage split (66%). The result of this learning scheme is summarized and presented in Table 3.

Out of the 17044 total records 11249(66%) of the records are used for training purpose while 5794(44%) of the records are used for testing purpose. As we can see from the confusion matrix of the model developed with this proportion, out of the 5794 testing records 5723 (98.7575%) of them are correctly classified. Only 72 (1.2425%) records are incorrectly classified.

3) Experiment 3

This experiment is performed, by changing the default testing option (the 66% for training and 44% for testing). So the percentage split parameter set to 70, which is to mean 70% for training and 30% for testing, resulted with a better accuracy. The result of this learning scheme is summarized and presented in Table 4.

In this experiment out of the 17044 total records 11931 (70%) of the records are used for training purpose while 5113 (30%) of the records are used for testing purpose. As we can see from the confusion matrix of the model developed with this proportion, out of the 5113 testing records 5051(98.7874 %) of them are correctly classified. Only 62 (1.2126%) records are incorrectly classified.

4) Experiment 4

In the fourth experiment on this is by changing the 70/30% to the default 50/50% parameters. The result of this is shown in Table 5.

In this experiment out of the 17044 total records 8522 (50%) of the records are used for training purpose while 8522 (50%) of the records are used for testing purpose. The confusion matrix of the model developed with this proportion shows, out of the 8522 testing records 8422 (98.83%) of them are correctly classified. Only 100 (1.17%) records are incorrectly classified.

In the previous four experiments when the testing data is increased the performance of the algorithm for predicting the newly coming instances is also diminished as well when compare the 10-fold cross validation, 70/30% and 66/44% which scores an accuracy of 98.797%, 98.787% and 98.757% respectively but the 4th experiment changes this flow that is while using 50% of the records for testing 50% for training this scores highest accuracy of 98.83%. Though this experiment is conducted by varying the value of the training and the testing datasets, this shows that the experiment 50/50% (98.83%), 10-fold cross validation (98.80%), 70/30% (98.79%), 66/44 % (98.76%) conducted, is better experiment from highest to lowest scoring respectively.

Generally, from the four experiments conducted before, the model developed with the 50/50% parameter values of the J48 decision tree algorithm test option gives a better classification accuracy of predicting of the children classification in their respective class category. Therefore, among the different decision tree models built in the foregoing experimentations, the fourth model, with the 50/50% parameter values, has been chosen due to its better overall classification accuracy.

B. Naïve Bayes Model Building

Naïve Bayes Classifier algorithm is widely used by many researchers [2], [18]. It is a simple statistical Bayesian Classifier based on Bayes' theorem with strong class conditional independence assumption for classifying the data.

Table 2. Confusion matrix output of the J48 algorithm with default value

Actual	Predicted2				Total	Correctly classified (accuracy rate)
	Vulnerable	Orphan	Single_Orphan	Safe		
Vulnerable	7159	0	66	98	7323	98.13%
Orphan	0	1983	0	0	1983	100%
Single_Orphan	41	0	4574	0	4615	98.48%
Safe	0	0	0	3123	3123	100%
Total	7256	1983	4584	3221	17044	98.7972%

Table 3. Confusion Matrix output of the J48 algorithm with the percentage – split set to 66%

Actual	Predicted3				Total	Correctly classified (accuracy rate)
	Vulnerable	Orphan	Single _Orphan	Safe		
Vulnerable	2401	0	31	41	2473	97.088%
Orphan	0	711	0	0	711	100%
Single_Orphan	0	0	1569	0	1569	100%
Safe	0	0	0	1042	1942	100%
Total	2401	711	1600	1083	6695	98.7575%

Table 4. confusion Matrix output of the J48 algorithm with the percentage – split set to 70%

Actual	Predicted4				Total	Correctly classified (accuracy rate)
	Vulnerable	Orphan	Single _Orphan	Safe		
Vulnerable	2123	0	28	34	2185	97.16
Orphan	0	630	0	0	630	100%
Single_Orphan	0	0	1383	0	1383	100%
Safe	0	0	0	915	915	100%
Total	2123	630	1411	949	5113	98.7874%

Table 5. Confusion Matrix output of the J48 algorithm with the percentage – split set to 50%

Actual	Predicted				Total	Correctly classified (accuracy rate)
	Vulnerable	Orphan	Single _Orphan	Safe		
Vulnerable	3603	0	0	54	3657	98.52%
Orphan	0	1039	0	0	1039	100%
Single_Orphan	46	0	2261	0	2307	98.00%
Safe	0	0	0	1519	1519	100%
Total	3649	1039	2261	1573	8522	98.8266%

Table 6. Confusion matrix output of the naïve bayes simple algorithm

Actual	Predicted				Total	Correctly classified (accuracy rate)
	Vulnerable	Orphan	Single_Orphan	Safe		
Vulnerable	7113	0	112	98	7323	98.13%
Orphan	0	1983	0	0	1983	100%
Single_Orphan	97	0	4518	0	4615	97.68%
Safe	0	0	0	3123	3123	100%
Total	7210	1983	4630	3221	17044	98.1988%
10-fold cross validation						

It works based on the presence (or absence) of a particular feature of a class [20].

The same attributes that are used to build the decision tree models, are also used in this Naïve Bayes modeling experiments. With all preprocessing in place, the experiment proceed with the different naïve bayes models by changing the default parameter values. The 10-fold cross validation, which is set by default, the percentage split with 66/44%, 70/30% and

50/50% for training and testing the model test options are employed. Naïve Bayes makes predictions using Bayes' Theorem, which derives the probability of a prediction from the underlying evidence. Naïve Bayesian classifiers assume that the effect of an attribute value on a given class is independent of the values of the other attributes [18], [27].

1) Experiment 1

The first experiment of the Naïve Bayes model building is performed using the Naïve Bayes Simple algorithm with the default10-fold cross validation test option. Table 6 shows the resulting confusion matrix of the model developed using the Naïve Bayes Simple algorithm with the default 10-fold cross validation test option.

The result from this experiment shows that out of the 17044 total records 16737 (98.1988%) of them are correctly classified and 307 (1.8012%) records are incorrectly classified. In addition to this the resulting confusion matrix has shown that out of 7323 vulnerable records 7113 (9713%) of them are correctly classified while 112 (0.015%) of the records are misclassified in single orphan and 98(0.0134%) of the records are misclassified in safe. And out of 4615 single orphan records 4518 (97.89) of them are correctly classified and 97 (0.021%) of them are incorrectly classified as vulnerable. Furthermore, the confusion matrix of this experiment shows that 100% of the records are correctly classified in the orphan and safe of their class. This shows that the model correctly classified those children data's in their respective class. The model developed with Naïve Bayes Simple Algorithm is poor in the accuracy of classifying new children dataset to the respected class, compared with the decision tree model that is developed before.

1) Experiment 2

The second experiment of the Naïve Bayes model building is performed using the Naïve Bayes Simple algorithm with the 66/44% training and testing percentage split test option. Though different experiments are conducted by changing the size of the training and testing datasets, the one with 66/44% training and testing dataset scored better classification accuracy and it is presented here. The result of this experiment is shown in Table 7.

The confusion matrix shows that resulted from the model developed by the Naïve Bayes Simple Algorithm with the 66/44% percentage split, the model scored an accuracy of 98.22%. This shows that from the total 5795 test data, 5692 (98.22%) of the records are correctly classified, while 103 (1.777%) of them are misclassified. Compared with the second classification the first classification is better in classifying children's data correctly.

2) Experiment 3

The third experiment of the Naïve Bayes model building is performed using the Naïve Bayes Simple algorithm with the 70/30% training and testing percentage split test option. Though different experiments are conducted by changing the size of the training and testing datasets, the one with 66/44% training and testing dataset scored better classification accuracy and it is presented here. The result of this experiment is shown in Table 8.

The result from this experiment shows that out of the 5113 total records 5024 (98.26%) of them are correctly classified and 89 (1.74 %) records are incorrectly classified. Similarly the other results are achieved. The model developed with third experiment of the Naïve Bayes Simple Algorithm scores higher accuracy of classifying new children dataset to the respected class, compared with the experiment that is developed before.

3) Experiment 4

The fourth experiment of the naïve bayes model building is performed using the Naïve Bayes Simple algorithm with the 50/50% training and testing percentage split test option. Though different experiments are conducted by changing the size of the training and testing datasets, the one with 50/50% training and testing dataset scored better classification accuracy and it is presented here. The result of this experiment is shown in Table 9.

The confusion matrix of the model developed with this proportion shows that out of the 8522 testing records 8379 (98.32%) of them are correctly classified. Only 143 (1.68%) records are incorrectly classified. Generally, the model developed with fourth experiment of the Naïve Bayes Simple Algorithm scores higher accuracy of classifying new children dataset to the respected class, compared with the experiment that is developed before in naïve bayes algorithm.

C. Neural Network Model Building

This Naïve Bayes Classifier algorithm is widely used by many researchers [23]. Artificial neural network is an abstract computational model of the human brain. It has the ability to learn from experiential knowledge expressed through inter unit connection strengths, and can make such knowledge available for use [2]. A neural network starts with an input layer, where each node corresponds to a predictor variable. These input nodes are connected to a number of nodes in a hidden layer. The nodes in the hidden layer may be connected to nodes in another hidden layer, or to an output layer. The output layer consists of one or more response nodes [18].

To build the neural network model that classifies the data into the given classes based on the given data, it worked on finding the appropriate number of iterations that would result a maximum accuracy. The same attributes that are used to build the decision tree and naïve bayes models, are also used in this neural network modeling experiments. With all preprocessing in place, the experimentation proceeded with the different neural network models by having the default parameter values. The 10-fold cross validation, which is set by default, and the percentage split with 66-44%, 70/30% and 50/50% for training and testing the model test options are employed.

1) Experiment 1

This experiment is performed, by using the default testing option (the 10- fold cross validation testing option). The result of this learning scheme is summarized and presented in Table 10.

The result from this experiment shows that out of the 17044 total records 16854 (98.81%) of them are correctly classified and 190 (1.1848%) records are incorrectly classified. In addition to this the resulting confusion matrix has shown that out of 7323 vulnerable records 7229 (98.716%) of them are correctly classified while 40 (0.00546%) of the records are misclassified in single orphan and 54 (0.00737%) of the records are misclassified in safe. Further out of 4615 single_orphan records 4563 (98.87) of them are correctly

Table 7. Confusion matrix output of the naïve bayes simple algorithm

Actual	Predicted					Correctly classified (accuracy rate)
	Vulnerable	Orphan	Single_Orphan	Safe	Total	
Vulnerable	2400	0	32	41	2473	97.05
Orphan	0	711	0	0	711	100%
Single_Orphan	30	0	1539	0	1569	98.089
Safe	0	0	0	1042	1042	100%
Total	2430	711	1571	1083	5795	98.2226%
Percentage split (66/44% Training and Testing) test option						

Table 8. confusion matrix output of the naïve bayes simple algorithm

Actual	Predicted					Correctly classified (accuracy rate)
	Vulnerable	Orphan	Single _Orphan	Safe	Total	
Vulnerable	2123	0	28	34	2185	97.16%
Orphan	0	630	0	0	630	100%
Single_Orphan	27	0	1356	0	1383	98.05%
Safe	0	0	0	915	915	100%
Total	2150	630	1384	949	5113	98.2593%
Percentage split(70/30% Training and testing)test option						

Table 9. Confusion matrix output of the naïve bayes simple algorithm

Actual	Predicted					Correctly classified (accuracy rate)
	Vulnerable	Orphan	Single _Orphan	Safe	Total	
Vulnerable	3560	0	43	54	3657	97.35%
Orphan	0	1039	0	0	1039	100%
Single_Orphan	46	0	2261	0	2307	98.00%
Safe	0	0	0	1519	1519	100%
Total	3606	1039	2304	1573	8379	98.322%
Percentage split(50/50% Training and testing)test option						

Table 10. Confusion matrix output of the neural network algorithm

Actual	Predicted					Correctly classified (accuracy rate)
	Vulnerable	Orphan	Single_Orphan	Safe	Total	
Vulnerable	7229	0	40	54	7323	98.716%
Orphan	0	1983	0	0	1983	100%
Single_Orphan	52	0	4563	0	4615	98.87%
Safe	44	0	0	3079	3123	98.59%
Total	7325	1983	4603	3133	17044	98.8152%
10-fold cross validation						

Table 11. Confusion matrix output of the neural network algorithm

Actual	Predicted					Correctly classified (accuracy rate)
	Vulnerable	Orphan	Single_Orphan	Safe	Total	
Vulnerable	2432	0	0	42	2474	98.30%
Orphan	0	711	0	0	711	100%
Single_Orphan	30	0	1539	0	1569	98.088%
Safe	0	0	0	1042	1042	100%
Total	2462	711	1539	1084	5796	98.7748%
Percentage split (66/44% Training and testing) test option						

Table 12. Confusion matrix output of the neural network algorithm

	Predicted					Correctly classified
Actual	**Vulnerable**	**Orphan**	**Single_Orphan**	**Safe**	**Total**	**(accuracy rate)**
Vulnerable	2151	0	0	34	2185	98.44%
Orphan	0	630	0	0	630	100%
Single_Orphan	27	0	1356	0	1383	98.05%
Safe	0	0	0	915	915	100%
Total	2178	630	1356	949	5113	98.807%

Percentage split(70/30% Training and testing)test option

Table 13. Confusion matrix output of the neural network algorithm

	Predicted					Correctly classified
Actual	**Vulnerable**	**Orphan**	**Single _Orphan**	**Safe**	**Total**	**(accuracy rate)**
Vulnerable	3560	0	43	54	3657	97.35%
Orphan	0	1039	0	0	1039	100%
Single_Orphan	0	0	2307	0	2307	100%
Safe	0	0	0	1519	1519	100%
Total	3560	1039	2350	1573	8522	98.86%

Percentage split(50/50% Training and testing)test option

classified and 52 (0.01126%) of them are incorrectly classified as vulnerable. In addition out of 3123 safe children 3079 (98.59%) of them are correctly classified and 44 (0.0140%) are misclassified in vulnerable. Furthermore, the confusion matrix of this experiment shown, that 100% of the records are correctly classified in the orphan of their class. This shows that the model correctly classified those children data's in their respective class.

1) Experiment 2

This experiment is performed, by changing the default testing option (the 10 cross validation testing option). So the percentage split parameter set to 66/44%, which is to mean 66% for training and 44% for testing, the result of this learning scheme is summarized and presented in Table 11.

The result from this experiment shows that out of the 5796 total records 5724 (98.77%) of them are correctly classified and 71 (1.22 %) records are incorrectly classified. Both of the two neural net models used have generally shown very good classification accuracy. However, the first model built using the default parameters and 10-fold cross validation excels both in overall accuracy. Hence, it is chosen as best neural net model.

2) Experiment 3

This experiment is performed, by changing the default testing option (the 66% for training 44% for testing). So the percentage split parameter set to 70, which is to mean 70% for training and 30% for testing, the result of this learning scheme is summarized and presented in Table 12.

The result from this experiment shows that out of the 5113 total records 5052 (98.81%) of them are correctly classified and 61 (1.18%) records are incorrectly classified.

3) Experiment 4

The fourth experiment is performed, by changing the testing option (70/ 44%). So the percentage split parameter set to 50/50%, which is to mean 50% for training and 50% for testing, the result of this learning scheme is summarized and presented in Table 13

The confusion matrix of the model developed with this proportion shows that out of the 8522 testing records 8425 (98.86 %) of them are correctly classified. Only 97 (1.14 %) records are incorrectly classified. In comparison of the neural network model experiments the fourth experiment that splitting to 50/50% scores highest.

The neural network models are considered as black box. This is due to the fact that it does not really explicitly show why a certain records are segmented/ classified in to a certain class. Besides it does not generate rules like decision trees.

V. Comparison and Selection of the best model

Selecting a better classification technique for building a model, which performs best in handling the prediction of children classification, is one of the aims of this study. For that reason, the decision tree (particularly the J48 algorithm), the bayes (the Naïve Bayes Simple algorithm in particular) and neural network classification methods were applied for conducting experiments to build the best model. Summary of experimental result for the three classification algorithms that scores higher accuracy from each is presented in table 14.

Table 14. Accuracy of the J48 decision tree, Naïve Bayes and neural network

Classification Model	**Overall accuracy (17044 records)**	
	Correctly classified	Misclassified
Decision Tree	8422(98.83 %)	100(1.17%)
Naïve Bayes	8379 (98.32%)	143 (1.68%)
Neural Network	8425(98.86%)	97 (1.14 %)

The results of the three algorithms are compared with each other for their overall classification accuracy (performance). It is clearly shown in table 14, the overall performance of the decision tree model was 98.83% with 8422 data sets. However,

the classification accuracy of the naïve bayes model with this data size and parameter was 98.32%. In naïve bayes classifier the highest classification accuracy was achieved in the same datasets with decision tree and 50/50% test split option. Furthermore the classification accuracy of the neural network model was 98.86%.

The J48 decision tree has shown second better classification performance after neural network. Hence, it is really reasonable to conclude that the J48 decision tree model is the best classifier model for implementing of children classification applications in the LCO. The reason for the J48 decision tree to perform better than neural network is because of the linearity nature of the dataset. That means there is a clear demarcation point that can be defined by the algorithm to predict the class for a particular children datasets. Regarding the Naïve bayes, scoring a lower accuracy than the J48 decision tree is due to its assumption that each attribute is independent of other attributes, which is not true in reality especially in such NGO's like LCO. Moreover, in terms of ease and simplicity to the user the J48 decision tree is more self-explanatory. It generates rules that can be presented in simple human language.

Therefore, it is plausible to conclude that the J48 algorithm is more appropriate to this particular case than the Naïve bayes and neural network method. So, the model that is developed with the J48 decision tree classification technique is taken as the final working classification model.

VI. Conclusion and Recommendation

Recent advances in communication technologies, on the one hand, and computer hardware and database technologies, on the other, have made it all the more easy for organizations to collect, store and manipulate massive amounts of data. Having concentrated on the accumulation of data, the question is what to do next with this valuable resource? Indeed, the data contains and reflects activities and facts about the organization. The increase in data volume causes great difficulties in extracting useful information and knowledge for decision support. It is to bridge this gap of analyzing large volume of data and extracting useful information and knowledge for decision making that the new generation of computerized methods known as Data Mining or Knowledge Discovery in Databases (KDD) has emerged in recent years.

The discovery task was run on the children related database that consists of 17044 records in more than 20 tables describing a total of 11 attributes. In general, encouraging results were obtained by employing Bayesian network, neural networks and decision tree approaches. Although Bayesian network, neural Network and decision trees showed comparable accuracy and performance in predicting the children condition, the decision tree approach seems more applicable and appropriate to the problem domain since it provides additional features such as rules that can be expressed in human language so that anyone can easily understand how and why a classification of children is made.

This research work is conducted mainly for academic purpose. However, it is the researcher's belief that the findings of the research will help Governmental and non- governmental organizations to work on the application of data mining techniques to gain competitive advantage in their organization. Moreover, the research work can contribute a lot towards a comprehensive study in this area in the future.

In the course of doing this study and on the basis of the findings of the research work, the researcher has come up with the following recommendations:

The predictive model, which is developed in this research, generated various patterns. For the company to use it effectively there is a need to design a knowledge base system, which can provide advice for the domain experts.

The model building process in this investigation was carried out in three algorithms of classification model that are J48 decision tree, neural network and Bayesian network algorithm. Though, the results were encouraging, further investigation needs to be done using other classification techniques such as Support Vector Machine and other data mining techniques such as clustering and association rule to see if they could be more applicable to the problem domain.

References

[1] WHO, UNICEF, and C. Mathers. "Global strategy for women's, children's and adolescents' health (2016-2030)." *Organization* 2016, no. 9 (2017).

[2] Hailemariam, Tesfahun. "Application of data mining for predicting adult mortality." *Master's thesis. Addis Ababa, Ethiopia: Addis Ababa University* (2012).

[3] UNICEF. *The State of the World's Children 2011: Adolescence-an Age of Opportunity*. UNICEF, (2011).

[4] Deshpande, S. P., and V. M. Thakare. "Data mining system and applications: A review." *International Journal of Distributed and Parallel systems (IJDPS)* 1, no. 1 (2010): 32-44.

[5] Shegaw, A. "Application of data mining technology to predict child mortality patterns: the case of Butajira Rural Health Project (BRHP)." *Master of Science Thesis, School of Information Science, Addis Ababa University: Addis Ababa, Ethiopia* (2002).

[6] Mohamed Taha, Hamed Nassar, Tarek Gharib and Ajith Abraham, An Efficient Algorithm for Incremental Mining of Temporal Association Rules, *Data and Knowledge Engineering, Elsevier Science, Netherlands*, 69:800-815, 2010.

[7] Swagatam Das, Sambarta Dasgupta , Arijit Biswas, Ajith Abraham and Amit Konar, On Stability of the Chemotactic Dynamics in Bacterial Foraging Optimization Algorithm, *IEEE Transactions on Systems Man and Cybernetics - Part A, IEEE Press, USA*, 39(3): 670-679, 2009.

[8] Tibebe Tesema, Ajith Abraham and Crina Grosan, Rule Mining and Classification of Road Accidents Using Adaptive Regression Trees, *International Journal of Simulation Systems, Science & Technology, UK,* 6(10-11):80-94, 2005.

[9] Hongbo Liu, Ajith Abraham and Maurice Clerc, Chaotic Dynamic Characteristics in Swarm Intelligence, *Applied Soft Computing Journal, Elsevier Science*, 7(3):1019-1026, 2007.

[10] Lizhi Peng, Bo Yang, Yuehui Chen and Ajith Abraham, Data Gravitation Based Classification, *Information Sciences, Elsevier Science, Netherlands,* 179(6):809-819, 2009.

[11] Hesam Izakian, Behrouz Ladani, Kamran Zamanifar and Ajith Abraham, A Particle Swarm Optimization Approach for Grid Job Scheduling, *Third International Conference on Information Systems, Technology and Management (ICISTM-09), Communications in Computer and Information Science, Springer Verlag, Germany*, ISBN 978-3-642-00404-9, pp. 100-109, 2009.

[12] Helen, T. "Application of data mining technology to identify significant patterns in census or survey data: the case of 2001 child labor survey in Ethiopia." *Master of Science Thesis, School of Information Science, Addis Ababa University: Addis Ababa, Ethiopia* (2003).

[13] Kifle, W. "Application of kdd on crime data to support the advocacy and awareness raising program of forum on street children in Ethiopia." *PhD diss., Master's thesis, Addis Ababa University*, 2003.

[14] David, Julie M., and Kannan Balakrishnan. "Machine learning approach for prediction of learning disabilities in school age children." *Int. J. of Computer Applications, ISSN-0975-8887* 9, no. 10 (2010).

[15] Garg, Sonu Bala, Ajay Kumar Mahajan, and T. S. Kamal. "An Approach for Diabetes Detection using Data Mining Classification Techniques." (2017).

[16] Chowdary, B. V., and Y. Radhika. "A Survey on Applications of Data Mining Techniques." *International Journal of Applied Engineering Research* 13, no. 7 (2018): 5384-5392.

[17] Sharma, Sumana, and Kweku-Muata Osei-Bryson. "Toward an integrated knowledge discovery and data mining process model." *The Knowledge Engineering Review* 25, no. 1 (2010): 49-67.

[18] Ambili K. and Afsar P. "A Prediction Model for Child Development Analysis using Naive Bayes and Decision Tree Fusion Technique – NB Tree." *International Research Journal of Engineering and Technology (IRJET)*. Volume: 03 Issue: 07, (2016).

[19] Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth. "From data mining to knowledge discovery in databases." *AI magazine* 17, no. 3 (1996): 37.

[20] Senthilkumar, D., and S. Paulraj. "Prediction of low birth weight infants and its risk factors using data mining techniques." In *Proceedings of the 2015 International Conference on Industrial Engineering and Operations Management*, Dubai, United Arab Emirates (UAE), March 3 – 5,(2015):pp. 186-194.

[21] Witten, Ian H., Eibe Frank, Mark A. Hall, and Christopher J. Pal. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, (2016).

[22] Quinlan, J. Ross. "Induction of decision trees." *Machine learning* 1, no. 1 (1986): 81-106.

[23] Kale TD. "Application of Data Mining Techniques to Discover Cause of Under-Five Children Admission to Pediatric Ward: The Case of Nigist Eleni Mohammed Memorial Zonal Hospital." *J Health Med Informant* 6(2015): 178.

[24] Parashar, Hem Jyotsana, Singh Vijendra, and Nisha Vasudeva. "An efficient classification approach for data mining." *International Journal of Machine Learning and Computing* 2, no. 4 (2012): 446.

[25] Julie, M. David, and Balakrishnan Kannan. "Prediction of learning disabilities in school age children using decision tree." In *Recent Trends in Networks and Communications*, pp. 533-542. Springer, Berlin, Heidelberg, 2010.

[26] David, Julie M., and Kannan Balakrishnan. "Machine learning approach for prediction of learning disabilities in school age children." *Int. J. of Computer Applications, ISSN-0975-8887* 9, no. 10 (2010).

[27] García, Salvador, Julián Luengo, and Francisco Herrera. *Data preprocessing in data mining*. Springer, 2016.