IOSR Journal of Nursing and Health Science (IOSR-JNHS)
e-ISSN: 2320–1959.p- ISSN: 2320–1940 Volume 9, Issue 1 Ser. IX. (Jan - Feb. 2020), PP 01-06
www.iosrjournals.org

Describing Infant Immunization Status in Ethiopian

Hiwot Abebe[1], Million Meshesha (PhD) [2]
[1]*Department of Health Informatics, Addis Ababa University, Ethiopia.*
[2]*School of Information Science, Addis Ababa University, Ethiopia.*

Abstract:
***Background**: Immunization is one of the most cost effective and efficient interventions saving the lives of many millions of infants and children from dying of infectious and preventable diseases. New data from WHO and UNICEF (2018), 20 million children miss out on lifesaving vaccines such as measles, diphtheria and tetanus and 2–3 million children are dying annually from easily preventable diseases and many more fall ill. This research attempts to construct a predictive model using data mining technology that helps to describing the infants' immunization status in Ethiopia.*
***Methodology**: This study is guided by a Hybrid-data mining model which is a six step knowledge discovery process model such as problem understanding, data understanding, data preparation, data mining as well as evaluation and use of the discovered knowledge. The study has used 8,210 instances, 12 describing and one outcome variables to run the experiments. Classification data mining algorithms such as J48 decision tree, support vector machine, neural network and rule induction are used in all experiments due to their popularity in recent related works.*
***Results:** The J48 decision tree has given the best classification and a better describing accuracy of the infant immunization status in Ethiopia. The experiment has generated a model with accuracy of 62.5%, weighted precision of 62.5% and weighted ROC area of 67.6% for the J48 decision tree. Place of delivery, region, mother-education-level, wealth-status, listening-to-radio, mother-age and parity are identified as the main variables determining immunization status of infants. Therefore, increase awareness creation among women in communities' pastoralist, so as to enhance vaccine coverage.*
***Conclusion:** The results achieved from this research indicate that data mining is useful in describing infant immunization status for decision making. Ministry of Health, Zonal and District health officials at different levels are recommended to create awareness on infant immunization for the pastoralists for effectiveness of vaccine and its coverage. Health professionals need to intensively teach families about the vaccine type and their uses and side effects.*
***Key Word**: - Data Mining, Hybrid-Data Mining Process Model, Classification Algorithm, J48 Decision Tree, Support Vector Machine, Artificial Neural Network, Infant Immunization*

Date of Submission: 17-01-2020 Date of acceptance: 05-02-2020

I. Introduction

Immunization is one of the most cost effective and efficient intermediations saving in the lives of many millions of infants and children from dying of infectious and preventable diseases [1]. Infants and children are therefore needed to get immunization service not to be affected by communicable, but preventable diseases. Due to this fact, World Health Organization (WHO) launched the Expanded Programmed on Immunization (EPI) in 1974 with the aim of immunizing the world's infant against the six major communicable diseases such as: Diphtheria, Pertussis, Tetanus, Tuberculosis, Measles and Poliomyelitis their first year of life.[1].

Furthermore, the combined effort of WHO, United Nation Development Program (UNDP),United Nation Children's Fund (UNICEF),the World Bank, other developmental agencies, and national programs resulted in the achievement of the global goal of 80% immunization in 1990 [2]. After 1990, global immunization coverage for infants under one year of age was maintained at 80% for the recommended three doses of Diphtheria Pertussis Tetanus (DPT) and polio by many countries both from developing and developed nations [2].

However, there is a wide difference among regions and DPT3 coverage ranged from 91% in East Asia and pacific to 62% in East Africa and 42% in West and Central Africa [2]. In Ethiopia vaccine preventable communicable diseases are also major public health problems. Accordingly, the prevention and control of these communicable diseases have received high priority. EPI was started in Ethiopia in 1980 with the aim of reducing morbidity and mortality of children from vaccine preventable diseases [3, 4]. According to a 2013 immunization data report, vaccine coverage was 75 %; and Ethiopia has the second largest number of partially vaccinated children from the region, next to Nigeria [5-7] Factors associated with child full immunization included socio-demographic characteristics (mother and father education level, wealth-status, listening-to-radio, mother-age, parity and region), health service delivery (place of delivery) [8, 6, 9, 10].

According to the World Health Organization guidelines, children are considered fully immunized when they have received one dose of Bacillus Calmette Guerin (BCG), three doses of DPT, three doses of polio vaccines, and one dose of measles vaccination by the age of 9–12 months [11, 9]. Ethiopia has incorporated Haemophilus influenza type B (HiB) and hepatitis-B (HepB) antigens to the previous DPT vaccines and replaced as Pentavalent vaccine (DPT plus Hep B and Hib) [8, 5, 12, 13].Variety of vaccines, of which the Pneumococcal conjugate vaccine (PCV), Rota and Human papilloma (HPV) vaccines were the most recent, which have been introduced into the national EPI service overtime. Different findings showed that the proportion of full immunization coverage in the country ranged from 36.6% in Somalia region to 100% in Addis Ababa [8, 6, 9, 10–14].

Accordingly, infants are more likely to be vaccinated the first doses of vaccination than the third and the fourth doses in which 60% of children received BCG and from these only 35% of them receive measles vaccine which is the last vaccine dose in EPI program of Ethiopian [15]. This shows that there is a drop out of from vaccination. Therefore, the aim of this study was to predict infant immunization status in Ethiopia using data mining techniques and to generate data that could be used for better planning and strengthening of full immunization coverage.

II. Review of the related works

Adebayo et al [16] used Mathematical Model (MM) for predicting immunizes able diseases that affect infant up to 5 years of age. The model was adapted and deployed for use in six selected localized areas within Osun State in Nigeria. The DM techniques provided the means by which hidden information were discovered for detecting trends within databases, and thus facilitate the prediction of future disease occurrence in the tested locations. Results obtained showed that diseases have peak periods depending on their epdemicity, hence the need to adequately administer immunization to the right places at the right time. Therefore, this paper argues that using this model would enhance the effectiveness of routine immunization in Nigeria.

A study was conducted by Selam [17] on predicting the occurrence of measles outbreak in Ethiopia using DM techniques. For the prediction the key attributes used include Name of the reporting health facility, Date district sent record, Date record received at national level, name of the patient, Date of birth of the patient, Age in years of the patient, Vaccine status, whether the patient is vaccinated, unvaccinated or unknown, and Record status.

The methodology used to achieve the goal of building predictive model using DM technique for this research was a hybrid six-step Cios KDP. The required data was collected from WHO measles surveillance database covering the period 2006-2011. Naïve Bayes and decision tree DM techniques were employed to build and test the models using a dataset of 15,631 records.

To get a better insight in choosing which model produced sound prediction and higher accuracy, 12 experiments were done with J48 algorithm and naïve Bayes classifier, by inputting all the records with a 10-fold cross-validation mode, and percentage split (70%) for training and then remaining 30% of the record for testing the performance of the model. Experimental results show that J48 decision tree algorithm register better prediction accuracy.

Hemalatha and Megale [18.] briefly examine the potential use of classification based DM techniques such as decision tree, Artificial Neural Network to massive volume of Immunization data. In their study data analysis of infant with Immunization and vaccination have been used as an upstream, from protecting infant, against such infections and infectious diseases as BCG, DPG, Polio and Measles. After preliminary results were analyzed, the program projected that over three million cases of deaths would be prevented and it has been resulted in a statistically significant in table survey. There is still, however, much that can be done. Through the use of DM algorithms in order to verify the improvement of quality

Up to the knowledge of the researcher, no previous researches have been done to predict the routine infant immunization status by applying DM techniques in Ethiopia. Hence, this research contributes a lot to generate patterns that help in planning a better strategy for routine infant immunization status using DM technique.

III. Research Methodology

Research design

A hybrid-data mining process model was applied. This process model is developed based on Cross-Industry Standard Process (CRISP) models by adopting in to suit for academic researches

The Hybrid DM process model presented in figure 1 consists of a six-step Knowledge Discovery Process; i.e. understanding the problem domain, understanding the data, preparation of data, data mining, evaluation and use of the discovered knowledge

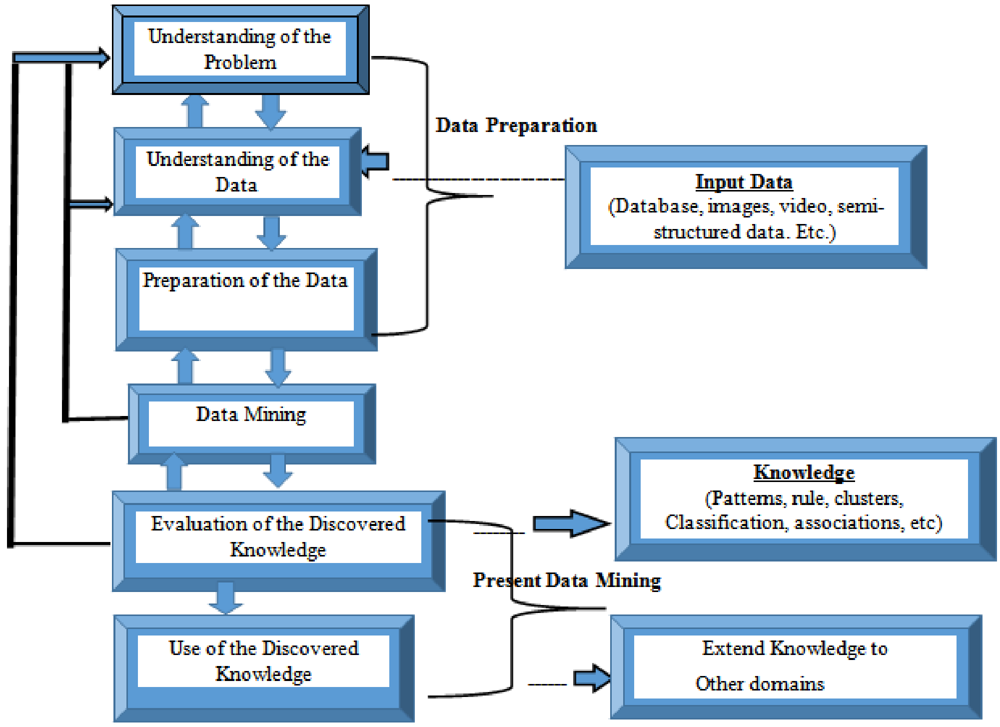

Figure 1 Hybrid-DM process model

Understanding of the problem

A model was needed to predict the most probable region with dropout rate of the EPI preventive processes before it causes additional harm to the society. In Ethiopia vaccine preventable communicable diseases are the major health problem. It is necessary to ensure all infants access to routinely recommended vaccines; that is, BCG, three doses of DTP, three doses of Polio and measles vaccination by the age of 12 months. To understand the problem domain of infant immunization status in Ethiopia, the researcher used secondary data in EDHS manual report and also discussion with domain expert health professionals from Government hospital and Health Center. A model is necessary to predict which region has partial or unimmunized infant dropout rate the EPI to take preventive processes before it causes additional harm to the society.

When we come to the result of the problem understanding of the discussion held with the domain area expert, one of the health professionals explained that, there are no methods as well as system of communication with family to remind about the continuation of the immunization program used in the health center to control if the child is fully or partially immunized. Recently in the health center, manual registration form was prepared to follow and remind the family about their infant status of immunization. The second health professional explained her ideas saying, in the referral hospital, there is no control system of the dropout of immunization infant and not having family communication system for the routine of infant immunization indicated so far in the health center

Understanding of the data

The data source for this research is taken from 2011 EDHS dataset and the children immunization coverage data. This study used a total of 8,210 records. The data was collected on vaccination coverage in two ways: first from vaccination in cards shown to the interviewer and from mothers' verbal reports.

Discussions were conducted with domain area experts and review of related research was made before selecting the target attributes from EDHS dataset. Based on the discussion and literature review, the researcher selected twelve (12) attributes.

TABLE 1: Selected Attributes with their Description from EDHS2011 dataset

No	Variable Name	Description	Data Types	Values
1	Age	Mother's age Group	Categorical	1. 15-19, 2. 20-24, 3. 25-29, 4. 30-34, 5. 35-39, 6. 40-44, 7. 45-49
2	Marital status	Current Marital status of the mother	Nominal	1. Single, 2.Married, /Living to gather, 3.Widowed, 4.Divorced /Separated
3	Mother Education	Mother's level of education	Nominal	0No education, 1. Primary, 2 Secondary, 3. Higher
4	Place of delivery	Access of health facility	Nominal	1. Home, 2. Government health facility, 3. Private health facility, 4.NGO health facility, 5. Other health facility (96, 12,26 …etc
5	Distance from health facility	Problem for getting medical help for self:	Nominal	0. No problem, 1. Big Problem 2. Not a big problem
6	Father Education	Father level of education	Nominal	0. No education, 1. Primary, 3, Secondary, 4. Higher education
7	Wealth status	The living standard of the mother	Nominal	1. Poor, 2. Middle 3. Rich
8	Residence	Place of Residence	Nominal	1. Urban, 2. Rural
9	Region	The 11 administrative region of the country where live	Nominal	1. Tigray 2. Affar 3. Amhara 4. Or Omiya 5. Somali 6. Benishangul Gumz 7. Southern. National Nationality People (SNNP), 12. Gambela 13. Harare 14. Addis Ababa 15. Dire Dawa
10	Frequency of listening to radio	How often do a mother listen to radio	Categorical	0. Not at all, 1 less than once a week, 2. At least once a week
11	Sex	Sex of child	Nominal	1. Male, 2. Female
12	Parity	Number of children of a mother including the child	Nominal	1, 1, 2. 2- 3. 3, 4-5, 4. 6-7 5,7-8 5,7-8 , 6..9-10 , 7.11-12
Dependent variable				
13	Immunization Status	The extent to which the child received all vaccines or not	Categorical	Full, Partial , No

The dataset has been described and visualized using Microsoft Excel version 2010 and SPSS version 22. Examining the properties of the dataset by frequency distribution shows there are missing values that need data preprocessing for preparing the data and conduct experiment.

Preparation of the data

A model was needed to predict which region most probable would have dropout rate the EPI preventive processes before it causes additional harm to the society. In Ethiopia vaccine preventable communicable disease are major health problem ensure all infant had access to routinely recommended vaccines that is BCG, three doses of DTP, three doses of Polio and measles vaccination by the age of 12 months. To understand the problem domain of infant immunization status in Ethiopia, the researcher used secondary data, the understanding of the detail the problem of infant immunization drop out in EDHS manual report and also discussion three health professionals from Yikatite hospital, Black lion hospital and kasanche Health Centers have been discuss the existing EPI problems. The researcher had 1 hour discussion with each of the professionals regarding the problem of the existing problem of EPI.

Understanding of the data

High quality data is a prerequisite for and data mining technique. The source of data for this study is 2011 EDHS dataset available from CSA or http://www.measuredhs.com. Descriptive summarization and visualizations of data conducted using statistical software of SPSS is an acronym for Statistical Package for Social Science. This application software was used to create a database for infant immunization dataset. Thus, for this research, total amounts of 8210 dataset utilized. The datasets for this study have a scale of measurement of nominal ten (10 attributes), categorical two (2 attributes) and one dependent categorical attribute. This datasets partitioned and used for training the model and testing the model accuracy.

Data Pre-processing

Data cleaning routines work to "clean" the data by filling in missing values, smoothing noisy data, identifying or removing outliers, and resolving inconsistencies [30].Generally data cleaning decreases errors and increases the data quality. In the dataset, seven attributes namely frequency listening to radio, distance to health facility, father's education level, BCG, DPT, Polio, and Measles have missing values. We replaced all of them with the most common value cross checking with value. In this study 8,210 instances and 12 attributes are ready for experimentation after data preprocessing.

III. Predictive model

The most important objectives of the study is identifying which DM algorithm performs best in predicting the infant immunization status. Therefore, the experiments in the study were carried out with J48 decision tree, Sequential Minimal Optimization (SMO) support vector machine; MLP artificial neural network and PART rule induction algorithms. The dataset used for each experiment in the study has all attributes or best selected attributes. Model comparison was performed using performance evaluation matrix like true prediction rate, false prediction rate, recall, precision, ROC area and accuracy of the model. The detail result of the best model selected from each classification category is shown in table below.

TABLE 2: The performance of the selected model created by classification algorithm

	Accuracy	WTP Rate	WFP Rate	W Precision	WF- Measure	W Recall	WROC Area
J48 pruned with all attributes	**62.497**	**0.625**	**0.355**	**0.621**	**0.608**	**0.625**	**0.676**
SMO with all attribute	61.2911	0.613	0.372	0.605	0.591	0.613	0.685
PART pruned with best attribute selected	62.0097	0.62	0.345	0.618	0.607	0.62	0.694
MLP with best selected attribute	61.0962	0.611	0.337	0.607	0.601	0.611	0.695

In this study, J48 classifier has achieved relatively the highest accuracy of 62.5% as compared to MLP, SMO and PART algorithms. Therefore, the model generated by J48 classifier with all 12 attributes was selected as the model that can predict the infant immunization status.J48 decision tree generated 470 rules. Based on the discussion with the domain expert, 11 rules selected from the total rules created using J48 decision tree concerning the three types of infant Immunization [8].

1. **Full immunization** coverage is defined as a child that has received one dose of BCG, three doses of pentavalent, pneumococcal conjugate (PCV), oral polio vaccines (OPV); two doses of Rota virus and one dose of measles vaccine for which he/she was eligible by age..
2. **Partially immunized**: When the child had received some but not all vaccines for his/her age as per schedule.
3. **Not immunized**: When the child had not received any of the vaccine for the age, though eligible age.

Rules created that describe the pattern of Partial immunization.

1. If place of delivery = Home and Region = Amhara and Listening-To-Radio = Not-At-All then Partial (537.0/160.0)
2. If please of delivery = Home and Region = Amhara AND Listening-To-Radio = Less-Than-Once-A-Week then Partial (227.0/75.0)
3. If Place-Of-Delivery = Government-Health-Facility and Residence = Rural and Region = Amhara and Parity = 2-3 and Distance-to-health-facility = Big-Problem then Partial (4.0/1.0)
4. If Place of delivery = Home and Region = Tigray and Mother Age = 30-34 and Listening radio= Not at all and Mother-Education-level = No-Education and Parity = 4-5 and Marital-Status = Married and Mother-Education-level = No-Education: Partial (36.0/15.0)
5. If Place-Of-Delivery = Private-Health-Facility and Residence = Urban and Region = SNNP then Partial (3.0/1.0)

Rules created that describe the pattern of Unimmunized

6. If Place of delivery = Home and Region = Affar and Mother-Education-level = No-Education and Wealth-Status = poor and Listening-To-Radio = Not-At-All and Mother Age = 25-29 and Parity =4- 5 and Distance-to-health-facility = Big-Problem: Unimmunized (39.0/11.0)
7. If Place of delivery = Home and Region = Affar and Mother-Education-level = No-Education and Mother-Age = 25-29 AND Wealth-Status = Rich and Father-Education = No-Education and Listening-To-Radio = Not-At-All then Unimmunized (17.0/7.0)
8. If Place of delivery = Home and Region = Somali and Wealth-Status = Poor and Mother-Education-level = No-Education and Listening-To-Radio = Not-At-All and Father-Education = No-Education: Unimmunized (214.0/83.0)

Rules created that describe the pattern of Full immunization

9. If Place of delivery = Home, Region = Dire Dawa, Father Education = No Education, Mother Age = 30-34, Mother Education level = No Education, and Residence = Rural: Full (47.0/16.0)
10. Place of Delivery = Government Health Facility, Residence = Rural, Region = Tigray: Full (47.0/12.0)
11. Place of Delivery = Government-Health Facility, Residence = Rural , Region = Addis Ababa: Full (178.0/30.0)

IV. Conclusion

In this study J48 decision tree achieved better accuracy for constructing a predictor model. The results achieved from this research indicate that data mining is useful in bringing relevant information from large and complex EDHS dataset. We can use this information for predicting infant immunization status. The most important attributes that determine infant immunization status were place of delivery, region, mother's educational level, listening to radio, father education level, residence, mother age, wealth status, parity, distance to health facility and marital status. Ministry of Health, Zonal and District health officials at different levels are recommended to create awareness on infant immunization for the pastoralists for effectiveness of vaccine and its coverage. Health professionals need to intensively teach families about the vaccine type and their uses and side effects. The application of the prototype is believed to minimize the infant immunization drop out. The knowledge of data mining is also recommended for public health professional because it extracts hidden knowledge. There is a need for the development of knowledge based system for infant immunization status with domain experts.

Reference

[1]. Grant J.P. The state of the world children: Oxford University Press. UNICEF, New York, 1991 "1"https://apps.who.int/iris/bitstream/handle/10665/63027/WHO_GPV_96.04.pdf?sequence=1
[2]. Maurice, John M., and Sheila Davey. State of the World's Vaccines and Immunization. World Health Organization, 2009
[3]. UNICF. UNICEF annual report, Ethiopia, Addis-Ababa, 1996
[4]. Ministry of health. Communicable diseases control Division report, Addis Ababa, 1976.
[5]. Plan GVA. Decade of vaccines global vaccine action Plan GVAP secretariatreport 2015, vol. 22: SAGE; 2015.
[6]. LaFond A, et al. Drivers of routine immunization coverage improvement inAfrica: findings from district-level case studies. Health Policy Plan. 2014; 30(3):298–308.
[7]. Teferi E. Factors influencing coverage and key challenges to achievingtargets of routine immunization in Africa: a systematic review. Ethiop JPediatrChild Health. 2016; 12(2):31–42.
[8]. World Health Organization, Immunization, vaccines, and biologicals: implementation research in immunization. 2017.
[9]. Dessie DB, Negeri MA. Determining factors of full immunization of childrenamong 12-23 months old in rural Ethiopia. Am J Public Health. 2018;6(3):160–5.
[10]. Asmamaw A, et al. Determinants of full valid vaccine dose administrationamong 12-32 months' children in Ethiopia: evidence from the Ethiopian2012 national immunization coverage survey. Ethiop J Health Dev. 2016; 30(3):135–41.
[11]. Animaw W, et al. Expanded program of immunization coverage andassociated factors among children age 12–23 months in Arba Minch town andZuria District, Southern Ethiopia, 2013. BMC Public Health. 2014; 14(1):464.
[12]. Fedral Ministry of Health, Ethiopia National Expanded Programme onImmunization. 2015: Addis Ababa.
[13]. Cherian T, Okwo-Bele J-M. The decade of vaccines global vaccine actionplan: shaping immunization programmes in the current decade: Taylor &Francis; 2014
[14]. Mohamud AN, et al. Immunization coverage of 12–23 months old childrenand associated factors in Jigjiga District, Somali National Regional State,Ethiopia. BMC Public Health. 2014; 14(1):865.
[15]. Ethiopia Demographic and Health Survey Addis Ababa, Ethiopia 2011.
[16]. Adebayo Peter Idowu, Bernard Ijesunor Akhigbe, Olajide Olusegun Adeosun, Aderonke Anthonia Kayode, and Adekemi Faidat Osungbade. Data Mining Techniques for Predicting Immunize-able Diseases: Nigeria as a Case Study, Volume 5– No.7, May 2013.
[17]. Assamnew, Selam. "Predicting the occurrence of Measles Outbreak in Ethiopia Using data mining Technology." Addis Ababa University, 2011.
[18]. Hemalatha, M., and S. Megala. "Mining techniques in health care: a survey of immunization." Journal of Theoretical & Applied Information Technology 25, no. 2 (2011)