Journal of
Health & Medical Informatics
ISSN: 2157-7420

Obed, J Health Med Informat 2019, 10:4

Research Article Open Access

A Predictive Model for Demand for First-Line Antiretroviral (ARV) Drugs Using Data Mining Techniques

Asasira Obed*

Mbarara University of Science andTechnology, Uganda

Abstract

One of the most important indices of defining general welfare and quality-of-life of people in the world is physical and mental health of individuals.

Health care managers and planners therefore must make future demand for healthcare services and the need for medicines to achieve fully and reliable supply. With the introduction of the test and treat methodology of managing HIV patients, first line Antiretroviral drugs (ARVs) must be in adequate availability to enable facilities implement this strategy of HIV eradication. Discontinuation of antiretroviral therapy Antiretroviral drugs (ART) due to shortages may result into viral rebound, immune decomposition, and clinical progression of the virus, therefore there is need to plan ahead of time to avail the most required stock for ARV drugs.

There are no proper forecasting and anticipation mechanisms of future demand for first line Antiretroviral drugs (ARV) and this is a cross cutting problem for all the public health facilities in Mbarara District and this has led to overstocking and understocking of these drugs leading to shortages and wastage related to expiry

This study aimed at designing a predictive model for demand of first-line ARV drugs in Mbarara district, using data mining techniques. Using the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology, the objectives of this study were to extract and prepare dataset required for data mining, examine different methods used in demand prediction, Design a model for predicting demand and evaluate this model.

The model was trained under the Waikato Environment for Knowledge Management (WEKA) which is a data mining environment and predicted the demand for first line ARVs in various health facilities Mbarara district Uganda. The test results showed that the forecasting in time series approach was more suitable and efficient for drug cycles ahead demand forecasting. Forecast results demonstrated that the model performed remarkably well with increased number of actual data and iterations. A regression model gave more accurate forecast results with 7.3% Mean Percentage Error as compared to alternative methods of demand forecasting whose error was above 30%.

Keywords: Data mining; Prediction; First line ARVs; Decision making; WEKA; Linear regression; Time series

Introduction

Developing countries, where budgets for medicines are often tight, the supply cycle needs to be well-managed to prevent all types of wastage, including pilferage, misuse and expiry. This wastage reduces the quantity of medicines available to patients and therefore the quality of health care they receive. At least US $550 000 worth of antiretroviral drugs expired in Uganda. National drug authority has incinerated 1500 tonns of expired essential medicines and ARV expiry is at 40% of the drugs incinerated [1].

The Ugandan government has launched the UNAIDS approach of "test and treat" where every individual testing HIV positive must start Antiretroviral therapy(ART), meaning more HIV positive patients are likely to be registered across all health facilities, forcing already overburdened clinics to double their caseload. If the influx of new patients isn't matched with resources available health facilities are likely to face shortages of ARV drugs [2]. Discontinuation of antiretroviral therapy (ART) due to shortages may result into viral rebound, immune decomposition, and clinical progression of the virus [3].

Stock-outs, also known as shortages or complete absence of a particular inventory, in public health facilities have become a hallmark in Uganda's health system making the notions of persistent doubt in access to healthcare [4].

There are no proper forecasting and anticipation mechanisms of future demand for first line ARV drugs and this is a cross cutting problem for all the public health facilities in the District and this has led to overstocking and under stocking of these drugs leading to shortages and wastage related to expiry.

The available method relies on consumption alone as a factor affecting demand for ARV drugs and does not consider other factors like change in HIV prevalence, new HIV positive clients starting ART in the facility, HIV positive clients on ART transferring in from other health facilities, HIV positive clients on ART who transferring out to other health facility, HIV positive clients on ART who have spent a cycle without picking ART drugs, HIV positive clients on ART who die during the cycle, HIV positive clients on first-line ART switching to Second-line ART and seasonality factors that would bear significant value in determining demand for ARV medicines. There is need to find out the factors contributing to an increase or a decrease in demand for first-line ARV drugs and develops a model that predicts need for ARV drugs for Mbarara District health facilities [5].

***Corresponding author:** Asasira Obed, Healthcare Data Analyst, Mbarara University of Science and Technology, Uganda, E-mail: ruberaobed@gmail.com

Received April 30, 2019; **Accepted** May 24, 2019; **Published** June 05, 2019

Citation: Obed A (2019) A Predictive Model for Demand for First-Line Antiretroviral (ARV) Drugs Using Data Mining Techniques. J Health Med Informat 10: 335.

General objective

The general objective of this study was to design a predictive model for demand of first-line ARV drugs in Mbarara district, Uganda.

Specific objectives

- To extract and prepare the datasets required for data mining from (Health Management Information System) HMIS databases (DHIS2).
- To identify the most suitable demand forecasting methods for first-line ARV drugs and determine factors that affect demand for ARV drugs in Mbarara District.
- To design a factor based model for predicting demand for first-line ARV drugs in Mbarara District.
- To test and validate the developed model.

Materials and Methods

The CRISP-DM reference model for data mining provides an overview of the life cycle of a data mining project. It contains the phases of a project, their respective tasks, and their outputs. The study used WEKA (Waikato Environment for Knowledge Analysis) which is free and open source software used for machine learning algorithms and data mining tasks. WEKA implements different algorithms which include data pre-processing, classification, Decision Trees, Artificial Neural Networks, data analysis, predictive modeling and logic regression as well as contains a collection of visualization tools. This was being the data mining software tool of choice for this project. This study focused on prediction of first-line ARV drugs demand in Mbarara district. The study relied on secondary data collected by various health facilities in the district from six ART high volume sites with more than 1000 clients on ART. These sites were known to have started ART clinics before 2016.

Data preparation

The data set used for the building of the model was collected through the extraction, formatting and organization of the first line ARV drugs consumption data for Mbarara district for each health facility from the Health Information System (DHIS2). The data collected included ARV regimen consumption time series data for high volume ART facilities for three years (2016, 2017 and 2018) (Figure 1).

Data processing and analysis

This started in the field, with checking for completeness of the data and performing quality control checks, while sorting and grouping data. The plan involved;

- Sorting data,
- Performing quality-control checks,
- Data processing,
- Simple Data analysis.

Key algorithms

This solution implemented a few algorithms to realize the desired functionalities. These algorithms include the following:

1. WEKA Add on package for Time Series data Forecasts,
2. Importing of prediction data into the model from WEKA,
3. Decision making using WEKA forecast package.

Period	Organisation unit	Regimen	Bi-monthly Art_consumption	Newly started	Transfer in	Transfer Out	Dead	Lost to fol	Switched to 2nd line
Jan to Feb	Bugamba HC IV	106-6 Tenofovir/Lamivudine (TDF/3TC) 300mg	111	0	2	0	0	0	5
Jan to Feb	Bugamba HC IV	106-6 Tenofovir/Lamivudine/Efavirenz (TDF/3T	820	15	3	1	0	0	2
Jan to Feb	Bugamba HC IV	106-6 Zidovudine /Lamivudine/Nevirapine (AZ	226	0	2	2	0	0	4
Jan to Feb	Bugamba HC IV	106-6 Zidovudine/Lamivudine (AZT/3TC) 300m	59	5	2	3	0	0	3
Jan to Feb	Bugamba HC IV	106-6 Efavirenz (EFV) 600mg	27	5	0	4	0	0	2
Jan to Feb	Bugamba HC IV	106-6 Abacavir/Lamivudine (ABC/3TC) 60mg/	0	0	0	5	0	0	3
Jan to Feb	Bugamba HC IV	106-6 Nevirapine (NVP) 200mg	74	0	0	0	0	0	2
Jan to Feb	Bugamba HC IV	106-6 Nevirapine (NVP) 50mg	4	0	0	0	0	0	4
Jan to Feb	Bugamba HC IV	106-6 Cotrimoxazole 960mg tablet	25 040	25	9	0	0	0	2
Jan to Feb	AIC Mbarara Main Branch CLINIC	106-6 Tenofovir/Lamivudine (TDF/3TC) 300mg	71 340	27	0	0	0	0	3
Jan to Feb	AIC Mbarara Main Branch CLINIC	106-6 Tenofovir/Lamivudine/Efavirenz (TDF/3T	37 470	38	2	0	1	0	3
Jan to Feb	AIC Mbarara Main Branch CLINIC	106-6 Zidovudine /Lamivudine/Nevirapine (AZ	5 780	23	2	2	0	0	2
Jan to Feb	AIC Mbarara Main Branch CLINIC	106-6 Zidovudine/Lamivudine (AZT/3TC) 300m	173	23	1	1	0	0	3
Jan to Feb	AIC Mbarara Main Branch CLINIC	106-6 Efavirenz (EFV) 600mg	730	52	1	0	0	0	3
Jan to Feb	AIC Mbarara Main Branch CLINIC	106-6 Abacavir/Lamivudine (ABC/3TC) 60mg/	0	0	0	1	1	0	4
Jan to Feb	AIC Mbarara Main Branch CLINIC	106-6 Nevirapine (NVP) 200mg	24 630	2	0	2	2	0	3
Jan to Feb	AIC Mbarara Main Branch CLINIC	106-6 Nevirapine (NVP) 50mg	2 160	0	0	3	3	0	2
Jan to Feb	AIC Mbarara Main Branch CLINIC	106-6 Cotrimoxazole 960mg tablet	37475	52	6	0	4	0	3
Jan to Feb	Kakoba HC III	106-6 Tenofovir/Lamivudine (TDF/3TC) 300mg	391	14	2	2	1	0	2
Jan to Feb	Kakoba HC III	106-6 Tenofovir/Lamivudine/Efavirenz (TDF/3T	1 590	25	3	1	3	0	3
Jan to Feb	Kakoba HC III	106-6 Zidovudine /Lamivudine/Nevirapine (AZ	2 216	0	2	0	2	0	4
Jan to Feb	Kakoba HC III	106-6 Zidovudine/Lamivudine (AZT/3TC) 300m	296	41	2	1	1	0	2
Jan to Feb	Kakoba HC III	106-6 Efavirenz (EFV) 600mg	99	23	2	0	2	0	3
Jan to Feb	Kakoba HC III	106-6 Abacavir/Lamivudine (ABC/3TC) 60mg/	16	1	0	2	1	0	4
Jan to Feb	Kakoba HC III	106-6 Nevirapine (NVP) 200mg	465	0	0	0	3	0	8
Jan to Feb	Kakoba HC III	106-6 Nevirapine (NVP) 50mg	72	1	0	1	1	0	2
Jan to Feb	Kakoba HC III	106-6 Cotrimoxazole 960mg tablet	139 300	98	11	1	2	0	2

Figure 1: Raw data for first line ARVS for years 2016, 2017, 2018.

Comparison of accuracy for different algorithms

Algorithms have been compared using Mean absolute percentage error (MPE) and Relative absolute error at first step prediction. The results show that Linear Regression has higher accuracy compared to other algorithm and therefore it was the algorithm of choice for this study (Table 1).

Results and Discussion

Model evaluation

The evaluation of the model involved testing of the model that was developed and the results from WEKA (Unit testing). This level of testing aimed at verifying the functionalities of different subsystems in reading and manipulating data in WEKA [6].

The tasks in unit testing

The dataset was trained under WEKA with linear regression algorithm.

There are 6 ARV drug supply cycles in a year and the model used time series data for 18 cycles from years 2016, 2017, 2018 and predictions results for 6 cycles prediction in 2019.

We looked at the broad view of the totality of cases reported for each factor in each health facility from year 2016 to 2018. This data aided the forecasting of the demand for first line ARVS in the next 6 cycles.

The study considered three combinations of first-line ARV drugs since they are the most supplied and recommended by National medical stores and National drug authority for managing HIV positive patients in Mbarara District (Table 2) [7].

Algorithm	Mean percentage error (MPE)	Relative absolute error (RAE)
Artificial Neural Network (Multilayer Perceptron)	16.8601	0.0237
Linear Regression	8.1573	0.0171
Decision trees (Random Forest)	40.0509	35.8365

Table 1: Comparison of accuracy for different algorithms.

Predictions for AZT/3TC/NVP (Lamivudine ZIdovudine Nevirapine)

Weka prediction for AIC Mbarara Main Branch for Zidovudine/Lamivudine/Nevirapine (AZT/3TC/NVP) is given in Figures 2 and 3.

The model showed at instance 19 which is the first cycle of 2019 that, there were no registered new clients starting ART on AZT/3TC/NVP. There were 2 clients transferring in from another facility on same regimen, no clients transferred to other facilities, there were no cases of death in the cycle, no cases of switching to second line and 11 clients did not to pick their drugs within the cycle. The results have been compared with the already known data for this ARV regimen in the HMIS database for January to February cycle 2019 considering all factors that affect demand for first line ARV drugs as shown below [8].

Testing the accuracy of the prototype forecast results

For the forecast results to be dependable, they must be seen to be accurate and realistic. To test this on this model, the following procedures were done;

Performing the forecast for available data

Since from data collection and simulation I had data from 2016 to 2018, I set out to forecast the results for newly enrolled clients in care in cycle 1 2019. The model forecasted the real results of cycle 1 with a 7.3% marginal error (Table 3) [9].

Comparing Weka and other data mining tools (GMDH shell DS)

Grouped Method of Data Handling (GMDH shell) is a tool for extraction of knowledge from real-world datasets. It is easy-to-use but powerful solution for analysis of multivariate datasets obtained from different research fields and business areas. GMDH Shell can accurately forecast time series, create classifiers and regression models. Based on

Task	Description
1	Running WEKA and Forecasting Package
3	Reading Data into WEKA from a CSV file
4	Performing a Forecast
6	Running of Decision Support Reports

Table 2: Tasks in unit testing.

=== Future predictions from end of training data ===

inst#	Newly started on ART	Transfer in	Transfer Out	Dead	Lost to follow
1	0	6	0	0	12
2	0	3	2	0	14
3	0	4	0	0	18
4	0	2	0	0	19
5	0	0	0	0	17
6	0	8	2	0	19
7	0	6	0	0	14
8	0	5	0	0	15
9	0	4	0	0	16
10	0	5	0	0	14
11	0	3	0	0	16
12	0	7	3	0	17
13	0	3	0	0	13
14	0	2	0	0	15
15	0	4	0	0	16
16	0	3	0	0	15
17	0	4	0	0	15
18	0	8	1	0	17
19*	0	2.074	-2.8379	0	10.5499
20*	0	2.3421	0.9755	0	16.2282
21*	0	6.4575	-0.857	0	18.6684
22*	0	5.985	-1.3905	0	15.1336
23*	0	1.2373	-0.7208	0	11.8668
24*	0	10.2487	3.118	0	19.3005

Figure 2: Prediction results for AZT/3TC/NVP, AIC Mbarara main branch.

Figure 3: Learning and prediction line showing prediction results for AZT/3TC/NVP AIC Mbarara.

Cycle	Organization Unit	Newly started on ART actual	WEKA Forecast	WEKA%Error
C1 2019	Bugamba HC IV	5	5.7465	14.9
C1 2019	AIC Mbarara Main Branch CLINIC	11	10.1796	7.5
C1 2019	TASO Mbarara CLINIC	2	2.0452	2.3
C1 2019	Bwizibwera HC IV	21	22.3978	6.7
C1 2019	Kinoni HC IV	5	5.0388	3.9
C1 2019	Mbarara Municipal Council HC IV	9	9.7573	8.4
Mean Percentage Error (MPE)=7.3%.				

Table 3: Comparing actual values and results from Weka.

Health Facility	Actual reported in DHIS2	Prediction For Weka	% Error for Weka	Prediction for GMDH	% Error for GMDH
AIC Mbarara	11	10.1796	7.5	15	36.4
Bugamba HCIV	5	5.7465	14.9	6	20
Bwizibwera HCIV	21	22.3978	6.7	14	33.3
Kinoni HCIV	5	5.0388	3.9	4	20
MMC HCIV	9	9.7573	8.4	8	11.1
TASO Mbarara	2	2.0452	2.3	6	200
MPE		7.30%			53.50%

Table 4: Comparing results for GMDH and WEKA.

artificial neural networks, it allows you easily create predictive models, as well as preprocess data with dead simple point-and-click interface [10].

The results from this comparison shown that WEKA is a better tool for time series modeling because it can allow upload of big data compared to GMDH tool that allows importation of chunks of data for analysis. The results from predictions using both WEKA and GMDH are represented below in a table considering the actual values reported at first cycle in 2019 in DHIS2. The results have shown that WEKA results are more reliable than GMDH (Table 3) [11].

Comparing Weka and other data mining tools (GMDH shell DS)

Grouped Method of Data Handling (GMDH shell) is a tool for extraction of knowledge from real-world datasets. It is easy-to-use but powerful solution for analysis of multivariate datasets obtained from different research fields and business areas. GMDH Shell can accurately forecast time series, create classifiers and regression models. Based on artificial neural networks, it allows you easily create predictive models, as well as preprocess data with dead simple point-and-click interface [12].

The results from this comparison shown that WEKA is a better tool for time series modeling because it can allow upload of big data compared to GMDH tool that allows importation of chunks of data for analysis. The results from predictions using both WEKA and GMDH are represented below in a table considering the actual values reported at first cycle in 2019 in DHIS2. The results have shown that WEKA results are more reliable than GMDH (Table 4) [13,14].

Conclusion

The big and complex data in Mbarara district HMIS databases has not adequately been used by the facilities for analysis and forecasting of demand and hence aiding of decision making process. Most of the health facilities in the district do not use any forecasting mechanism to assist fore planning of their services but continue grappling with shortage and expiry of ARV drugs due to poor forecasting methods. For those who use forecasting, they use simple linear mechanism mainly based on counts made from registers and does not consider the many other variables that affect the demand for their services by the clientele. They only look at the numbers and project future numbers which rarely happens to be realistic forecast and hence end up not using them for any decision making.

A Data mining approach therefore for forecasting of demand for health services is the most accurate and ideal way of building a forecasting model that employs multiple variables. This was the main objective of this project the use of data mining approach to build a demand forecasting model in Mbarara District. The project proposed and validated a prototype based on Data mining to forecast demand for predicting demand for ARV drugs in health facilities..

References

1. Candan G, Taskin MF, Yazgan HR (2014) Demand forecasting in pharmaceutical industry using artificial intelligence: Neuro-fuzzy approach. J Military Info Sci 2: 41-49.
2. Zakumumpa H, Dube N, Damain RS, Rutebeemberwa E (2018) Understanding the dynamic interactions driving the sustainability of ART scale-up implementation in Uganda. Glob Health Res Policy 3: 23.
3. Holkmann Olsen C, Mocroft A, Kirk O, Vella S, Blaxhult A, et al. (2007) Interruption of combination antiretroviral therapy and risk of clinical disease progression to AIDS or death. HIV Med 8: 96-104.
4. Muyinda H, Mugisha J (2015) Stock-outs, uncertainty and improvisation in access to healthcare in war-torn Northern Uganda. Soc Sci Med 146: 316-323.
5. Cerna PD, Abdulahi TJ (2016) Prediction of anti-retroviral drug consumption for HIV patient in hospital pharmacy using data mining technique. IJITCS 8: 52-59.
6. Ghousi R, Mehrani S, Momeni M (2012) Application of data mining techniques in drug consumption forecasting to help pharmaceutical industry production planning. Proceedings of the 2012 IEOM. 1162-1167.
7. Kaur H, Krishan Wasan S (2006) Empirical study on applications of data mining techniques in healthcare. J Comp Sci 2: 194-200.
8. Kaur N, Kaur A (2016) Predictive modelling approach to data mining for forecasting electricity consumption. 6th International Conference-Cloud System and Big Data Engineering (Confluence). IEEE. 331-336.
9. Milovic B, Milovic M (2012) Prediction and decision making in health care using data mining. Int J Pub Health Sci 12: 1380.
10. Muriithi IA (2014) A data mining approach to private healthcare services demand forecast in Nairobi County. Kenya.
11. Ramos MI, Cubillas JJ, Feito FR (2016) Improvement of the prediction of drugs demand using spatial data mining tools. J Med Syst 40: 1-9.
12. Soni J, Ansari U, Sharma D, Soni S (2011) Predictive data mining for medical diagnosis : An overview of heart disease prediction. Int J Comp Appl 17: 43-48.
13. Soyiri IN, Reidpath DD (2013) An overview of health forecasting. Environ Health Prev Med 18: 1-9.
14. Windisch R (2011) Scaling up antiretroviral therapy in Uganda: using supply chain management to appraise health systems strengthening. Global Health 7: 25.