

Prediction of Anti-Retroviral Drug Consumption for HIV Patient in Hospital Pharmacy using Data Mining Technique

Dr. Patrick D. Cerna

Assistant Professor, College of Computing and Informatics, Haramaya University, Ethiopia
P.O. Box 335 Dire Dawa, Ethiopia
Contact No: +251921608860
E-mail: dr.patrickcerna@gmail.com

Thomas Jemal Abdulahi

Department of Information Science College of Computing and Informatics, Haramaya University, Ethiopia
P.O. 138, Dire Dawa, Ethiopia
Contact No: +251933258788
E-mail: thomasjemal2005@gmail.com

Abstract—Pharmacy handles all the medicine needed in the hospital that consists of vast amount of records. These produce large scale of datasets that are complex to manage and thereby need tools and technique to easily process, interpret, forecast and predict future consumption. Due to this, the method of predicting and forecasting stock consumption using Data Mining technique in hospital pharmacy is not be a surprising issue. Thus, this research investigated the potential applicability of data mining technology to predict the Anti-Retroviral drugs consumption for pharmacy based up on patient's history datasets of Jugal hospital, Harar, Ethiopia. The methodology used for this research is based on Knowledge Discovery in Database which had mostly relied on using the decision tree algorithms specifically M5P model tree. WEKA software, a data-mining tool were used for interpreting, evaluating and predicting from large datasets. Result with the data set suggests that tree based modeling approach can effectively be used in predicting the consumption of ARV drugs.

Index Terms—Data Mining, Anti-Retroviral Drugs, Pharmacy, Knowledge Discovery.

I. INTRODUCTION

1.1 Background

Ethiopia is home to approximately 800 000 patients with HIV/AIDS and the prevalence of HIV/AIDS in the general population is estimated to be 1.5% [8]. However, studies conducted in the general population in Ethiopia show that the situation is worse. There are 1.3 million adults living with HIV/AIDS in the country [9].

Started in 2005, the free antiretroviral therapy (ART) programme in Ethiopia, has decreased mortality and morbidity and improved the quality of life of patients [16].

According to the World Health Organization [19], standard antiretroviral therapy (ART) consists of the combination of at least three antiretroviral (ARV) drugs to maximally suppress the HIV virus and stop the progression of HIV disease.

The Jugal hospital ARV pharmacy is one of region hospital that provides services of ARV drugs for patients in the Harari Regional State. The ARV pharmacy was established with the purpose of giving different types of services, like, ARV drugs, other supplementary drugs, and ART services for HIV/AIDS patients. However, the ARV pharmacy had several challenges, like stock management issues, updating of drugs and ART treatment issues, skilled manpower issues, effectively digitizing of patients record issues and other issues that influence the provision of ARV drugs to patients.

Data mining is about extraction of interesting (non-trivial, implicit, previously unknown and potentially useful) information or patterns from data in large databases. Similarly, data mining, fondly called patterns analysis on large sets of data, uses tools like association, clustering, segmentation and classification for helping better manipulation of the data to help the pharma firms compete on lower costs while improving the quality of drug discovery and delivery methods [11].

1.2 Statement of the Problem

Countries or programs must identify the number of patients for whom they can guarantee a full supply of ARV drugs for multiple years. In supply chain terms, this identification means that countries will treat ARV drugs as full-supply commodities, although the medications will be in full supply only for a limited number of patients [18].

Hospital pharmacies in Ethiopia faces some challenges and problems on controlling the drug supply and consumption of ARV drugs. Since these drugs are coming from AID and governmental organizations for the treatment of HIV/AIDS patients, its importance on the

management of stock is given priority. This result to the wastage of resources or inadequate resource utilization problem even malfunctioning of drugs and some reaches they're expiry date.

By processing and digesting stock record it can extract supply and demands pattern that could be invaluable in the ARV consumption of Pharmacy. In fact, tools and techniques are prerequisite to process and interpret large volume of data so as to make use of them for decision-making purpose. However, in the case of Jugal Hospital there have been no techniques and tools that can support in managing ARV records efficiently. With this, almost all the decision-making processes of the hospital are not supported by tools and techniques that could extract patterns from previous stock records.

Therefore, the study will investigate on using data mining tool for the prediction and forecasting of ARV drugs based on daily and monthly consumption of Harari Jugal Hospital ARV pharmacy. Thus, as well as determining and insuring of how likely to be the pharmacy stock level by considering the minimum and maximum standardized stock level.

The data mining techniques will be used to predict the consumption rate of ARV drugs in the Jugal hospital pharmacy using patient's history of records as data sets. It will employ decision tree algorithm specifically MSP model tree and linear regression for forecasting and prediction future ARV drugs consumption.

1.3 Research Objectives

1.3.1 General Objectives

The general objective of the study is to examine the potential of data mining tools and techniques in developing models that could help predict the future pharmacy ARV consumption in Jugal hospital.

1.3.2 Specific Objectives

1. To assess the potential data mining techniques in prediction of ARV consumption in Jugal hospital.
2. To extract and prepare the dataset required for data mining from the patient's database or records of Jugal hospital.
3. To suggest a prediction method based on Knowledge Discovery in Database (KDD) process in the stock management of the most prescribed drugs.
4. To predict the consumption amount of ARV drugs in the Jugal hospital pharmacy using patient's history of records.

1.3.3 Research Questions

This study is guided by the following research questions:

1. What are the potential of data mining techniques in prediction of ARV consumption in Jugal Hospital?
2. What is the process of extraction to prepare dataset required for data mining?

3. What will be the essential usage of Knowledge Discovery in Databases (KDD) process for stock management of prescribed drugs?
4. What is the predicted consumption amount of ARV drugs in the Jugal Hospital?

II. RELATED WORK AND STUDIES

2.1 Data Mining and Knowledge Discovery

Data mining refers to extracting or "mining" knowledge from large amounts of data. The term is actually a misnomer. Remember that the mining of gold from rocks or sand is referred to as gold mining rather than rock or sand mining. Thus, data mining should have been more appropriately named "knowledge mining from mining from data," which is unfortunately somewhat long [12].

Data mining is the exploration and analysis of large quantities of data in order to discover meaningful patterns and rules. It is to allow a corporation to improve its marketing, sales, and customer support operations through a better understanding of its customers [15].

Further, data mining can provide intelligence in terms of: drug positioning information, patent population characteristics, indications of what the drug is being used for, prescribing physician characteristics, regional preferences, prevalence of diseases, preferred drug for diseases, procedures being performed, disease-related information [17].

Data mining is concerned with discovering new, meaningful information, so that decision makers can learn as much as they can from their valuable data assets. Using advanced information technologies, knowledge discovery in databases (KDD) can uncover veins of surprising and golden insights in a mountain of factual data [14].

Knowledge Discovery in Databases is the process of identifying a valid, potentially useful and ultimately understandable structure in data. This process involves selecting or sampling data warehouse, cleaning or preprocessing it, transforming or reducing it (if needed), applying a data mining component to produce a structure, and then evaluating the derived structure. Thus, the structures that are the outcome of the data mining process must meet certain conditions so that these can be considered as knowledge. These conditions are validity, understandability, utility, novelty and interestingness [3].

2.2 Decision Trees: Definitions and Applications

Decision trees are powerful and popular tools for classification and prediction. In data mining, a decision tree is a predictive model which can be used to represent both classifiers and regression models [7]. Decision trees are a simple, but powerful form of multiple variable analyses. They provide unique capabilities to supplement, complement, and substitute for traditional statistical forms of analysis (such as multiple linear regression), a variety of data mining tools and techniques (such as neural networks), and recently developed

multidimensional forms of reporting and analysis found in the field of business intelligence ^[1].

The decision tree technology can be used for exploration of the dataset and business problem. This is often done by looking at the predictors and values that are chosen for each split of the tree. Often times these predictors provide usable insights or propose questions that need to be answered ^[2].

2.3 M5P Tree Model

Model trees are essentially decision trees with linear models at the leaves. M5P is a reconstruction of Quinlan's M5 algorithm for inducing trees of regression models. M5P combines a conventional decision tree with the possibility of linear regression functions at the nodes ^[5].

The algorithm known as the M5P algorithm is used for inducing a model tree, the aim is to construct a model that relates a target value of the training cases to the values of their input attributes. The quality of the model will generally be measured by the accuracy with which it predicts the target values of the unseen cases ^[5].

2.4 ART in Hospitals

With the advent of ARV drugs, HIV/AIDS has become a treatable chronic disease. Currently, more than 4 million people are receiving ART in low and middle - income countries. In Ethiopia, HIV was first reported in 1984. In 2008, an estimated 1.03 million people were living with HIV and more than 125 000 people get newly infected every year, and an estimated 290,000 people need ART. Ethiopia started fee - based ART in 2003 and launched free ART in January 2005 and as of March 2010, 511 health facilities (142 hospitals and 369 health centers) provide ART service throughout the country ^[10].

2.5 Applications of Data Mining Techniques in the Pharmaceutical Industry

The application of data mining techniques in pharmaceutical industry, clinical data analysis – clinical data analysis evaluates and streamlines from large amount of information ^[17]. Data mining helps to see trends, irregularity, and risk during product development and launch. Marketing and sales analysis can be the identification of the most profitable product and allocation of marketing funds. Data mining here helps to examine consumer behavior in terms of prescription renewal and product purchases.

Data mining discovers and extracts useful patterns of this large data to find observable patterns. That is to demonstrate the ability of data mining in improving the quality of the decision making process in pharma industry. Issues in the pharma industry are adverse reactions of the drugs ^[11].

2.6 Applications of Data Mining Techniques in Ethiopia

Classification techniques such as decision trees, naive Bayes, and logistic regression have been used to predict nonpayment for Ethiopian telecom, but due to space limitations, only report our results from the decision tree models. The data required for prediction should be a

significant amount of cleaning. The sample, which was built for the models, was all customers who stopped service in August 2009, plus a random sample of 1992 users who had not stopped service ^[13].

Another research was conducted through the application of data mining for HIV status prediction in order to support the scaling up of HIV status testing in Addis Ababa. He used the CRISP – DM methodology, J48 and ID3, with a priority algorithm for HIV Status predictive modeling and discovering Association rules between HIV status and selected attributes ^[6].

2.7 Application of Data Mining in Other Sector

A propose a data mining based algorithm to tackle the problem, adopting a co-training styled algorithm to discover pattern in software development environment. Decision tree is trained as based learners and a majority voting procedure is then launched to determine labels of unlabeled data. Based learners are then trained again with newly labeled data and such iteration stops when a consistent state is reached. Our method is naturally semi-supervised which can improve generalization ability by making use of unlabeled data ^[20].

Data mining techniques were also used in forecasting maximum temperature, rainfall, evaporation and wind speed. This was carried out using Artificial Neural Network and Decision Tree algorithms and meteorological data collected between 2000 and 2009 from the city of Ibadan, Nigeria. The performances of these algorithms were compared using standard performance metrics, and the algorithm which gave the best results used to generate classification rules for the mean weather variables. The results show that given enough case data, Data Mining techniques can be used for weather forecasting and climate change studies ^[21].

III. METHODOLOGY

The study was conducted on Harari region Jugal hospital ARV pharmacy which is found in Harari region that is the ARV pharmacy of the hospital also found inside the hospital compound. The overall study period was covered a one-year time of interval or 12 months. The data was taken from the existing records which is stored in the database on the list of patient's records that have taken the ARV drug program.

The study used criterion sampling in which the study set criteria and pick all cases that met criterion. Additionally, as data mining works on large amounts of data's in the database the study collected patient's history of records data's from the period of 2009 to 2013 for the effective and efficient prediction of stock management system in ARV pharmacy. Out of the 7500 databases, 1100 data sets were selected as a sample using simple random sampling. This was selected based on the basis on their drug consumption amount rate compare to others within five years.

Knowledge Discovery in Database (KDD) was the primary methodology in interpreting, evaluating and predicting the ARV drugs consumption. This

methodology was selected for this specific study because of three reasons. First, the KDD methodology is best suited for academic researches. Second, as an outsider for the domain using KDD methodology reduces the skill required for the knowledge discovery. Third, the KDD methodology is also independent from any tools and techniques, so one can follow any desired technique during the study

Data were extracted from the database using Microsoft Excel and imported in the Data Mining tool called WEKA. WEKA was developed at the University of Waikato in New Zealand, and the name stands for *Waikato Environment for Knowledge Analysis*. To select the best algorithms performance of the algorithms and the structure of the data that was taken into consideration. After thoroughly checking the available algorithms in WEKA machine learning software, M5P model tree was selected which was one of the algorithms for predictions by Decision Tree.

IV. RESULTS AND DISCUSSION

4.1 Distribution of Findings

In this section, an attempt has been made to review the literature on the concepts and techniques of M5P algorithm in general to discuss about the results that were found during data mining process. M5P algorithm and Linear Regression were also compared for meeting assessment of potential data mining technique for Jugal hospital ARV pharmacy.

The information as shown Table 1.1 explains precisely the distributions of sampled records based on the gender attribute. Due to this, out of 1100 sampled datasets 522 are males and 578 are females that are 47.5% of the sampled population are males and 52.5% are females, respectively.

Table 1. Respondents Rate by Gender

Gender	Number	Percentage (%)
Male	522	47.5%
Female	578	52.5%
Total	1100	100%

Table 2. Respondents Rate by TB Status

TB Status	Number	Percentage (%)
Positive	205	18.6%
Negative	895	81.4%
Total	1100	100%

The above table explains precisely the distributions of sampled records based on the Tb status attribute. Out of 1100 sampled datasets 205 are positives and 895 are negatives that are 18.6% of the sampled population are positives and 81.4% are negatives, respectively. TB status of a patient is also a parameter for prediction of the drugs consumption as it directly related with HIV/AIDS compare to other diseases.

Thirteen attributes were selected from patient’s history database which has an impact for prediction of ARV consumption, namely; Gender, Age, TB status, One A Dispense Dose, One B Dispense Dose, One C Dispense Dose, One D Dispense Dose, One E Dispense Dose, One F Dispense Dose, Two E Dispense Dose, Four A Dispense Dose, Four B Dispense Dose, and Four D Dispense Dose, respectively.

The ARV drugs is consist of three drug combination which is can be composed of d4T (Estavudinine), 3TC (Lamuvudinide), NVP (Neverapine), AZT (Zidovudine), and TDF (Terafovine), EFV (Efavernaze). Further, ARV drugs can be categorized either for adult regimens and child regimens.

The combination of these drugs is as follows:

- One A = d4T + 3TC + NVP
- One B = d4T + 3TC + EFV
- One C = AZT + 3TC + NVP
- One D = AZT + 3TC + EFV
- One E = TDF + 3TC + EFV
- One F = TDF + 3TC + NVP
- Two E = d4T + 3TC + AZT
- Four A = d4T + 3TC + NVP
- Four B = d4T + 3TC + EFV
- Four D = AZT + 3TC + EFV

The prescription of drugs is either as single prescription or a combination thereof. It could be One A, One B, or One A and One E based on the patient’s condition. Because of this, the study used TB status as a parameter for prediction of the drugs in the stock and TB also directly related with HIV/AIDS than other diseases. Additionally, the prescription drugs for those who have HIV/AIDS plus TB is different with those who have HIV/AIDS only. Due to this, the study analyzed the data’s based on the above parameters.

Table 3. Rate of ARV Drugs based on Minimum, Maximum and Mean Standard Deviation

Variables	Minimum	Maximum	Mean Standard Deviation
Age	5	60	33.6 10.246
One A Dispense Dose	30	3872	1392.315 1020.485
One B Dispense Dose	18	3538	1290.247 1043.649
One C Dispense Dose	36	3878	957.288 826.033
One D Dispense Dose	42	4986	1160.809 1066.693
One E Dispense Dose	32	3486	1180.116 1066.57
One F Dispense Dose	60	2834	1293.113 834.877
Two E Dispense Dose	34	1532	953.5 620.889
Four A Dispense Dose	60	4148	1642.059 1464.91
Four B Dispense Dose	670	1226	855.333 287.117
Four D Dispense Dose	268	530	442.667 135.296

The data was collected on the period from year 2009 – 2013 in which includes 1100 sample out of 7500 total population. The dataset on the span of 5 years only includes active patient as the basis for consumption.

Table 3 shows the minimum, maximum, and the mean amount of ARV drugs that were consumed with in the last five years. The values were an aggregate value or total sum of dispensed drugs per individuals within five years intervals. For example on One A, it clearly implies that the minimum amount of age as 5, maximum amount as 60, the mean amount of 34, and the standard deviation value of 10.246 The standard deviation values that were displayed on the table also used in further analysis of the prediction process in M5P decision tree algorithm.

4.2 Simulating of M5P Model using WEKA

Data’s were first fetched from the Jugal Hospital ARV pharmacy patient’s history database, then it was prepared in table form using Micro soft excel for adjusting an extension format for the WEKA software.

The WEKA software mostly uses the ARFF (attribute – relation file format) or CSV (comma separated value) file extension format for processing. So this study used CSV file format for saving the files and further processing using M5P algorithm in WEKA software. Due to this, the following table shows the attributes and their respected values in CSV file format. The study also used CSV file format for Linear Regression in order to make a comparison with M5P algorithm for meting assessment of potential data mining technique which is described under specific objective of the study.

Gender	TB Status	Age	One A Dispense Dose	One B Dispense Dose
Nominal	Nominal	Numeric	Numeric	Numeric
M	Negative	50.0	0.0	1310.0
M	Negative	27.0	0.0	0.0
F	Negative	28.0	0.0	0.0
F	Positive	45.0	1818.0	0.0
M	Positive	38.0	0.0	0.0
F	Positive	30.0	0.0	0.0
F	Negative	40.0	398.0	0.0
M	Negative	20.0	0.0	0.0
M	Negative	29.0	0.0	0.0
F	Negative	35.0	0.0	0.0
M	Negative	34.0	440.0	0.0
M	Negative	41.0	1190.0	1972.0

Fig.1. Source Data Encoded in WEKA Explorer for Analysis

The CSV file format table shows the attributes with their respective datatypes that is Gender and TB status both have the nominal data type and age, One A, One B, One C, One D, One E, One F, Two E, Four A, Four B, and Four D dispensed dose all have a numerical data type. And the data’s that were entered from One A to Four D dispensed dose were the total sum or aggregate values of the ARV drugs respectively. Because all values that were sumed can be assigned either a drug which was prescribed to the patient or combination of drugs which were prescribed to the patient. The rows that were assigned with the number zero also considered as a missing values

which the WEKA software adjusted by the user to insert zeros as a missing values. Therefore, the table below shows the option tab for adjusting the missing values in WEKA software in third option as “Set Missing Values to...”. So the number zero explains the ARV drugs were not dispensed to patient’s during the five year period.

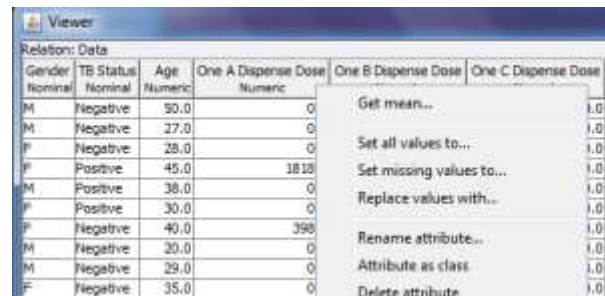


Fig.2. Demonstration of the Analysis Process

After the data’s were inserted to the WEKA software the data should be analyzed using a decision tree M5P – M20 algorithm. The algorithm is a best tool kit for predicting purpose therefore this study used the algorithm for every drug to be predicted in order to maintain the minimum and maximum stock level. So the diagram below shows the M5P – algorithm analyses phase from initial.

This diagram mostly known as a preprocess phase which includes all attributes with their minimum, maximum, mean, and standard deviation values and with their graphical demonstration. Therefore, before the data’s moved to the second phase for analysis in the classify phase all data’s should be explained here clearly.

After preprocess phase then the classify phase which is the second option should be clicked for analysis with different algorithms. So this study in classify phase used the M5P – M20 algorithm as shown in the diagram below.

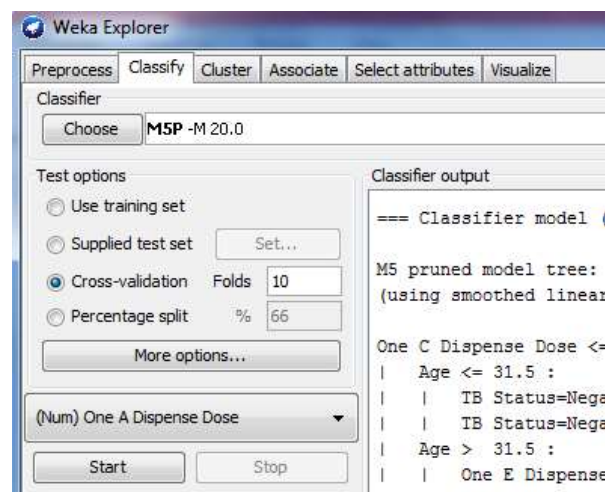


Fig.3. The WEKA Classify Phase

In classify phase the priority thing should be selecting of the algorithms from the choose button as demonstrated in the diagram. So for this study the M5P algorithm were selected and the number next to the algorithm indicates the minimum number of instances which will be found in

the leaf nodes. The minimum number of instances which was set to 20 because to simplify the tree. The other main point in classify phase would be specifying of the cross - validation value. So in this study the cross validation values were set by default to 10 because it is a standard way of predicting the error rate of a learning technique given a single, fixed sample of data was to use stratified 10 – fold cross – validation. The data was also divided randomly into 10 parts in which the class is represented in approximately the same proportions as in the full dataset. And according to Ian and Eibe the main reasons for using of cross – validation 10 value would be it is the right number of folds to get the best estimate error (Ian H. and Eibe F., 2005). After setting the cross – validation folds to 10 the next step should be pressing the button start on the WEKA software interface leads to the following out put for class One A dispense dose.

Table 4. Distributions of Records through Dispensed Class and Priority Class

No	Class of Dispensed Dose	Priority Class for the Node
1	One A Dispense Dose	One C Dispense Dose
2	One B Dispense Dose	One E Dispense Dose
3	One C Dispense Dose	One E Dispense Dose
4	One D Dispense Dose	One E Dispense Dose
5	One E Dispense Dose	One C Dispense Dose
6	One F Dispense Dose	One E Dispense Dose
7	Two E Dispense Dose	One E Dispense Dose
8	Four A Dispense Dose	Age
9	Four B Dispense Dose	One A Dispense Dose
10	Four D Dispense Dose	One A Dispense Dose

This is means the nodes of the tree are chosen over the attribute that maximizes the expected error reduction as a function of the standard deviation of output parameter. And also the splitting process will terminate if the output values of all the instances that reach the node vary only slightly or only a few instances remain.

Therefore, for determining classes One A dispense dose and One E dispense dose class One C dispense dose got a priority based on the above suggestions. And for Classes One B, One C, One D, One F and Two E Dispense Dose class One E Dispense Dose got a priority. Class One A Dispense Dose got a priority for determining Four B and Four D dispense dose classes. And only the age attribute got a priority class for determining Four A Dispense Dose class. So during the splitting processes for determining each classes One E Dispense Dose got a priority for five Classes of dispensed dose, One A and One C dispense dose got priority for two classes of dispensed dose respectively, and the age attribute got priority only for one class dispensed dose.

As a result of the above, One A, One C, One E, and Age class attributes reduces the expected error that would reach to the nodes. Due to this One A, One C, One E dispensed dose and Age attribute used as a best attributes for splitting criterion to split the training data that reaches to particular nodes. So when considering all the outputs of the pruned model trees for all classes the splits were

done by taking all attributes. Due to this, One A and One D classes have done splits on eight different positions of the thirteen attributes in the data respectively, One B done splits on seven different positions, One C and One F have done splits on six different positions respectively, One E done splits on four different positions, Two E done split once in a position, Four A done splits on three different positions, and finally Four B and Four D have done their splits on two different positions.

Due to this, the linear equation should be followed to every rule of all drugs in order to compute their predicted values to the end. That is, the number of rules of the linear model should be associated with the corresponded logical algorithm of all drugs.

Finally, the MSP algorithm used three steps to predict the values of all drugs dose, like, first generating tree using the training data for calculating a linear model for each node of the tree generated. Secondly, this step tries to simplify the regression tree generated in the previous search (first post – pruning) deleting the nodes of the linear models whose attributes do not increase the error. And thirdly, this step reduces the size of the tree without reducing the accuracy (second post – pruning).

Therefore, this study used visualizing classifier error for comparison of the predicted value with respect to actual value.

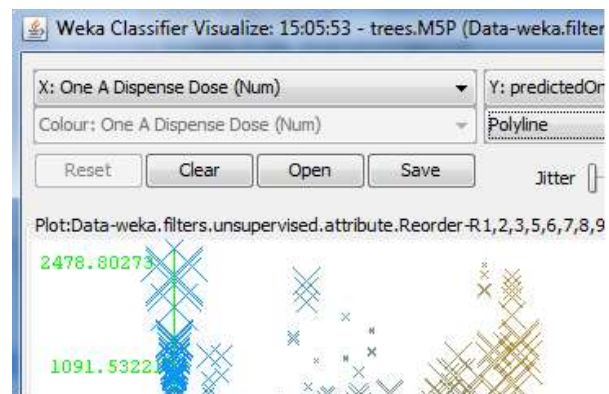


Fig.4. Predicted One a Dispense Dose using M5P Algorithm

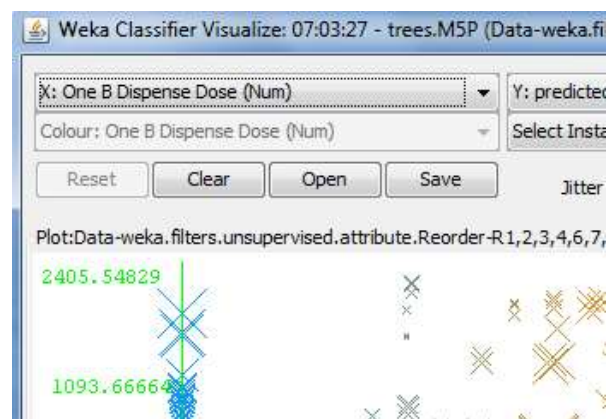


Fig.5. Predicted One b Dispense Dose using M5P Algorithm

The predicted value was plotted on the Y axis because it was a dependent variable and has a minimum value of -295.7383, mean value of 1091.53221, and finally the

maximum value of 2478.80273. The “X” sign with orange color indicates the ideal prediction and the “X” sign with blue indicates the test compounds. Therefore, the minimum value has a negative sign because the weighing value of the predicted attribute has a minus sign. Finally, the average predicted value of the One A Dispense Dose should be taken as a predicting component. Due to this, One A can be predicted as 1091.53221/1926 or 0.56% and this figure states that One A Dispensed Dose consumption was predicted by 0.56% on the actual consumption. Therefore, in the future One A should be needed or stocked an additional amount of 0.56% on the actual dispensed dose.

Table 5. Predicted ARV Drugs Consumption using M5P Algorithm

Variables	Combination	Actual	Predicted (%)
One A Dispense Dose	d4T + 3TC + NVP	1296 dose	56%
One B Dispense Dose	d4T + 3TC + EVP	1769 dose	62%
One C Dispense Dose	AZT + 3TC + NVP	1969 dose	15%
One D Dispense Dose	AZT + 3TC + EFV	2493 dose	37%
One E Dispense Dose	TDF + 3TC + EFV	1743 dose	49%
One F Dispense Dose	TDF + 3TC + NVP	1417 dose	57%
Two E Dispense Dose	d4T + 3TC + AZT	766 dose	55%
Four A Dispense Dose	d4T + 3TC + NVP	2074 dose	43%
Four B Dispense Dose	d4T + 3TC + EFV	613 dose	16%
Four D Dispense Dose	AZT + 3TC + EFV	265 dose	19%

Table 5 describes each of the ARV drugs predicted percentile value for the next five years that is an additional amount of each percentage value on the actual value.

Finally, the study predicted all the ARV drugs that were stated in table 1d, as follows, so One A, One B, One C, One D, One E, One F, Two E, Four A, Four B, and Four D drugs type have an average consumption amounts of for One A 1091.53221 or 0.56%, One B 1093.66664 or 0.62%, One C 295.9863 or 0.15% , One D 926.4922/2493 or 0.37%, One E 868.0255/1743 or 0.49%, One F 811.8969 or 0.57%, Two E 423.4734 or 0.55%, Four A 896.4711 or 0.43%, Four B 101.5889 or 0.16%, and Four D 52.744/265 or 0.19% , respectively for the next five years. So with the rate of consumption after getting a ratio with the above actual or previous average consumption amounts of each drugs therefore in the future the Jugal hospital ARV pharmacy should stock an additional amounts of 56/100, 62/100, 15/100, 37/100, 49/100, 57/100, 55/100, 43/100, 16/100, and 19/100 dosage amount of drugs for One A, One B, One C, One D, One E, One F, Two E, Four A, Four B, and Four D, respectively in their stock for the treatment of the patients for the next five years.

V. SUMMARY, CONCLUSION AND RECOMMENDATION

5.1 Summary

The potential of data mining techniques always relies on uncovering and un - hiding of information from large data bases by adopting the KDD (Knowledge Discovery in Data bases) process. Based from the literature and the result of this study, data mining techniques through the adoption of the nine steps of the KDD (Knowledge Discovery in Data bases) process, reveal important and critical information that will help ARV pharmacy to make good decisions.

Using M5P decision tree algorithm the study reveal the most prescribed drugs that need to be in stock for future consumption in the pharmacy.

5.2 Conclusion

This study predicts ARV drugs consumption for Jugal hospital ARV pharmacy using data mining technique. Result with the data set suggests that tree based modeling approach can effectively be used in predicting the consumption of ARV drugs. By utilizing the KDD process as primary methodology, the study met the prediction process through extraction and preparation of data sets. Finally, the M5P model tree was a better technique for development of prediction method for Jugal hospital ARV pharmacy.

5.3 Recommendation

The study also recommends further work by using different methodology or alternative data mining techniques that can be a basis for comparison. Such will enhance the reliability and validity of the predicted result.

ACKNOWLEDGEMENT

The author would like to acknowledge the people in the Pharmacy Department of Harar Jugal Hospital for the hospitality and kindness in sharing relevant data used in this study.

REFERENCES

- [1] Barry de Ville. 2006. Decision Trees for Business Intelligence and Data Mining: Using SAS Enterprise Miner. SAS Institute Inc., Cary, NC, USA.
- [2] Berson A, Smith S, and Thearling K. 2000. Building Mining Applications for CRM. McGraw-Hill.: USA.
- [3] Bharat A. and Sumit T. 2009. Data Mining and Data Warehousing. University Science Press, New Delhi-110002.
- [4] Bhattacharya B, Solomatine DP. 2003. Neural Networks and M5 model trees in modeling water level discharge relationship for an Indian river. In European Symposium on Artificial Neural Networks
- [5] Deepa, C., Sathiyakumari, K. and Pream Sudha, V. 2010. Prediction of the Compressive Strength of High Performance Concrete Mix Using Tree Based Modeling. International Journal of Computer Applications vol. 6 – No.5
- [6] Elias L. 2011. HIV Status Predictive Modeling Using

Data Mining Technology, Master of Science Thesis, School of Information Science and School of Public Health, Addis Ababa University: Addis Ababa, Ethiopia.

- [7] Ehsan H., Hamed Davari Ardakani and Jamal Shahrabi. 2010. Applications of data mining techniques in stock markets: A survey. *Journal of Economics and International Finance* vol. 2(7), pp. 109-118.
- [8] Ethiopian Federal Ministry of Health HIV/AIDS Prevention and Control Office. Country progress report in HIV/AIDS response. Addis Ababa: Ethiopian Federal Ministry of Health HIV/AIDS Prevention and Control Office, 2012.
- [9] Federal ministry of health of Ethiopia. Proceedings of the symposium on management of latent tuberculosis in HIV infection. Ethiopian Federal Ministry of Health in collaboration with ICAP-Ethiopia and U.S CDC- Ethiopia, Addis Ababa, Ethiopia, 13-14 November 2006.
- [10] Hailu Tadege and Negussu, 2010. Trends in Antiretroviral Drugs Prescribing at Public Health Facilities in Ethiopia: Compliance to Treatment Guidelines.
- [11] Jayanthi R., 2009. Data Mining in Pharma Sector: benefits. *International Journal of Health Care Quality Assurance*. Vol. 22 No. 1, 2009 pp. 82-92.
- [12] Jiawei H. and Micheline K. 2006. *Data Mining: Concepts and Techniques*, 2nd Edition. Morgan Kaufmann Publishers.
- [13] Mariye Yizgaw, Shawndra Hill, Anita Banser, Lemma Lessa. 2010. Using Data Mining to Combat Infrastructure Inefficiencies: The Case of Predicting Non-payment for Ethiopian Telecom. Masters Thesis. AAU.
- [14] Michael L. G. and Bel G. R. 1999. Data mining – a powerful information creating tool. *OCLC Systems and services*, Volume 15. Number 2. 1999 pp. 81-90., MCB University Press.
- [15] Michael J.A.B and Gordon S. L. 2004. *Data Mining Techniques for Marketing, Sales, and Customer Relationship Management*, 2nd Edition. Wiley Publishing, Inc.
- [16] Reniers G, Araya T, Davey G, et al Steep declines in population-level AIDS mortality following the introduction of antiretroviral therapy in Addis Ababa, Ethiopia. *AIDS* 2009; 23: 511–18.
- [17] Sandhya Joshi and Hanumanthachar Joshi. 2013. Applications of Data Mining in health and pharmaceutical industry. *International Journal of Scientific and Engineering Research*, Volume- 4, Issue – 4, April-2013,
- [18] USAID (United States Agency for International Development. May 2006. Supply Chain Management of Antiretroviral Drugs, Considerations for initiating and expanding national chains. Deliver Report
- [19] WHO (2014). HIV/AIDS Treatment and Care. <http://www.who.int/hiv/topics/treatment/en/>. Accessed on July 14, 2014.
- [20] Gang Zhang, Caixian Ye, Chunru Wang, Xiaomin He, "Data Mining based Software Development Communication Pattern Discovery", *IJMECS*, vol.2, no.2, pp.25-31, 2010.
- [21] Folorunsho Olaiya, Adesesan Barnabas Adeyemo, "Application of Data Mining Techniques in Weather Prediction and Climate Change Studies", *IJIEEB*, vol.4, no.1, pp.51-59, 2012.

Authors' Profiles



Dr. Patrick D. Cerna is an Assistant Professor in Department of Information Technology, College of Computing and Informatics in Haramaya University, Ethiopia. He is a member of some of the prestigious IS/IT international organization e.g. Association of Computing Machinery (ACM), Association of Information Systems (IS), International Association of Computer Science and Information Technology (IACSIT) and Institute of Advanced Engineering and Science (IAES). He is also a reviewer as member of scientific committee and editorial board in the journals and conference proceedings namely IIR Virtual Conference on Science, Engineering and Technology Applications, *International Journal of Computing, Communication and Instrumentation Engineering*, 2nd International Conference on Computing Paradigms and International Scholarly and Scientific Research & Innovation.



Thomas Jemal Abdulahi: is in the Department of Information Science, College of Computing and Informatics in Haramaya University, Ethiopia.

Manuscript received 07 05, 2015; revised 08 26, 2015; accepted 09 30, 2015.

How to cite this paper: Patrick D. Cerna, Thomas Jemal Abdulahi, "Prediction of Anti-Retroviral Drug Consumption for HIV Patient in Hospital Pharmacy using Data Mining Technique", *International Journal of Information Technology and Computer Science (IJITCS)*, Vol.8, No.2, pp.52-59, 2016. DOI: 10.5815/ijitcs.2016.02.07