Predictive Modeling for HIV-Testing: Applications of Data Mining

Tesfay Gidey Hailu, MSc, Addis Ababa Science and Technology University, Addis Ababa

Abstract

In this paper, we compared the prediction power of different data mining techniques used to develop the HIV testing prediction model. Four popular data mining algorithms including: Decision tree, Naive Bayes, Neural network, and Logistic regression were used to build the model that predicts whether an individual was being tested for HIV among adults-based EDHS 2011. Experimental results revealed that the decision tree (random tree algorithm) performed the best with accuracy of 96%, the decision tree induction method (J48) came out to be the second best with a classification accuracy of 79%, followed by neural network (78%). In this study, CRISP-DM method was used to predict the model for HIV testing and explore association rules between HIV testing and the selected attributes. Besides, data preprocessing was also performed and missing values for the categorical variable were replaced by the modal value of the attribute. The target dataset contained 30,625 study participants. Out of which 16,515 (54%) participants were women while the rest 14,110 (46%) were men. The age of the participants in the dataset ranged from 15 to 59 years old with modal age of 15 - 19 years old. Among the study participants, 17,719 (58%) have never been tested for HIV while the rest 12,906 (42%) had been tested. Residence, educational level, wealth index, HIV related stigma, knowledge related to HIV, region, age group, risky sexual behavior attributes, knowledge about where to test for HIV and knowledge on family planning through mass media were found to be significant predictors for HIV testing. Results have shown that data mining is so crucial in extracting relevant information for the effective utilization of HIV testing services which has significant importance at all levels of clinical, community and public health. Thus, it is paramount to apply different data mining techniques for the same settings and compare the model performances given experts knowledge-domain into account. Furthermore, this study would also invite interested researchers to explore more on the application of data mining techniques to healthcare industry or else in related and similar settings as a future work.

Keywords

Data Mining, Comparative Analysis, Predictive Modeling, HIV Testing

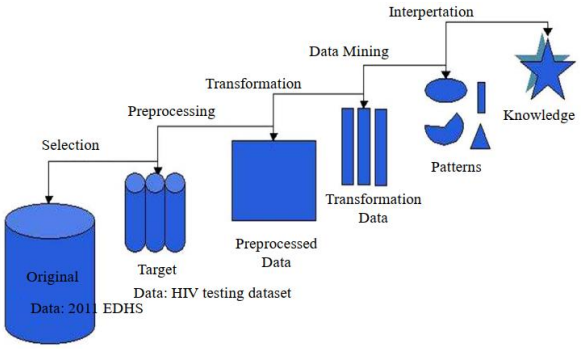


Figure 1. Overview of the steps constituting the KDD process.

Recent Publications (minimum 5)

1. Hailu TG (2016) A Multilevel Modeling Analysis of the Determinants and Cross-regional Variations of HIV Testing in Ethiopia: Ethiopian DHS2011. J Biom Biostat 7: 277. doi:10.4172/2155-6180.1000277
2. Tesfay Gidey (2015): Determinants and Cross-Regional Variations of Contraceptive Prevalence Rate in Ethiopia: A Multilevel Modeling Approach: American Journal of Mathematics and Statistics: 95-110 DOI: 10.5923/j.ajms.20150503.01
3. Hailu, T.G. (2015) Comparing Data Mining Techniques in HIV Testing Prediction. *Intelligent Information Management*, **7**, 152-179. http://dx.doi.org/10.4236/iim.2015.73014
4. Berhe, L. and Gidey, T. (2016) Assessing the Awareness and Usage of Quality Control Tools with Emphasis to Statistical Process Control (SPC) in Ethiopian Manufacturing Industries. Intelligent Information Management, 8, 143-169. http://dx.doi.org/10.4236/iim.2016.86011
5. A book published in LAP Lambert Academic Publishing : Predictive Modeling for HIV testing using Data Mining Techniques: https://www.morebooks.de/store/us/book/predictive-modeling-for-hiv-testing-using-data-mining-techniques/isbn/978-3-659-19878-6



Paste your Logo here

Biography (150 word limit)

Tesfay Gidey Hailu received the BSc. degree from Addis Ababa University, Ethiopia, in 2006, MSc degree from Mekelle University, College of Health Sciences, Tigray, Ethiopia, in 2013, and MSc degree in Software Engineering from HILCoE School of Computer Science and Technology, Ethiopia, in 2018. He is working toward the Ph.D. degree in the department of Information and Communication Engineering, UESTC. His research interests include indoor localization, machine learning, transfer learning, data mining, Health Informatics, Epidemiology, and Bio/statistical modeling.

Email: tesfaygidey21@gmail.com or
tesfaygdey@yahoo.com

Notes/Comments:

Presenting author details
Full name: Tesfay Gidey Hailu
Contact number: 13258103244
Twitter account:
Linked In account: https://www.linkedin.cn/injobs/in/tesfay-gidey-hailu-aaa12b102
Session name/ number:
Category: (**Oral presentation**/ Poster presentation)