Frontiers in Dentistry

Applicability of Data Mining and Predictive Analysis for Tobacco Cessation: An Exploratory Study

Kavita Rijhwani[1*], Vikrant R Mohanty[1], Aswini YB[1], Vaibhav Singh[2], Sumbul Hashmi[1]

1. Department of Public Health Dentistry, Maulana Azad Institute of Dental Sciences, Delhi, India
2. Department of Computer Science, Rameshwaram Institute of Technology and Management, Lucknow (U.P), India

Article Info

Article type:
Original Article

Article History:
Received: 22 Jan 2020
Accepted: 19 Aug 2020
Published: 20 Oct 2020

**** Corresponding author:***
Department of Public Health Dentistry, Maulana Azad Institute of Dental Sciences, Delhi, India

Email: dockavirijhwani@gmail.com

A B S T R A C T

Objectives: Predictive analysis can be used to evaluate the enormous data generated by the healthcare industry to extract information and establish relationships amongst the variables. It uses artificial intelligence to reveal associations not suspected by the healthcare professionals. Tobacco cessation is clearly beneficial; however, many tobacco users respond differently as it is based on multitude of factors. Our objectives were to assess the data mining techniques using the WEKA tool, evaluate its role in predictive analysis, and to predict the quit status of patients using prediction algorithms in tobacco cessation.

Materials and Methods: WEKA, a data mining tool, was used to classify the data and evaluate them using 10-fold cross-validations. The various algorithms used in this tool are Naïve Bayes, SMO, Random Forest, J-48, and Decision Stump to further analyze its role in determining the quit status of patients. For this, secondary data of 655 patients from a tobacco cessation clinic were utilized and described using 20 different attributes for prediction of quit status.

Results: The Decision Stump and SMO were found to be having the best prediction and accuracy for prediction of the quit status. Out of 20 attributes, previous quitting attempt, type of intervention, and number of years since the habit was initiated were found to be associated with early quitting rate.

Conclusion: This study concluded that data mining and predictive analytical models like WEKA tool will not only improve patient outcomes but identify variables or a combination of variables for effective interventions in tobacco cessation.

Keywords: Data Mining; Tobacco Use Cessation; Algorithms

➢ **Cite this article as:** Rijhwani K, Mohanty VR, YB A, Singh V, Hashmi S. Applicability of Data Mining and Predictive Analysis for Tobacco Cessation: An Exploratory Study. *Front Dent*. 2020; 17:24. doi: 10.18502/fid.v17i24.4624

INTRODUCTION

Tobacco use is a major cause of preventable death and disease in India. India is the second largest tobacco consumer, and third largest tobacco producer, in the world [1]. The current cost of tobacco use in India includes 1 million deaths per year (approximately 1/6 of all tobacco-related deaths worldwide), and billions of dollars of direct attributable health costs [2]. The problem is worsening, and tobacco use is believed to be the cause of 13% of deaths in India in 2020 [3].

A combination of approaches needs to be taken which targets at avoiding initiation of tobacco by non-users and cessation among the current users. Among the tobacco users, more than half will die from tobacco-related diseases if they do not quit. If the current use of tobacco among adults could have been reduced by half by the year 2020, 180 million deaths due to tobacco could be avoided [4].

Tobacco cessation is one of the critical activities under the National Tobacco Control Program in India. Tobacco cessation needs to be urgently expanded by training health professionals in providing routine clinical interventions, increasing the availability of pharmacotherapy and subsidy on it, and developing wide-reaching strategies such as quit lines, and cost-effective strategies [5].

Several cessation strategies like raising awareness about the harmful effects of tobacco, implementing and enforcing COTPA act, pharmacological interventions, motivational interviewing and behavioral counselling [6,7] have been developed and researched. However, the overall success rate of long-term tobacco use abstinence is still modest, even when intensive interventions are implemented. Many barriers are present against quitting tobacco [7]. Lack of awareness of harm, ingrained cultural attitudes, and lack of support for tobacco cessation in the community are the ones that dominate the literature [8]. The significant addictive property of nicotine makes quitting difficult and relapse common. Healthcare professionals often receive not much training in this respect, and among them, very few take necessary actions to assess and intervene tobacco consumption.

Data mining is the term used for sequential processing of selecting, exploring, and modelling huge data such that unknown patterns or relationships are discovered which may give useful and meaningful insight to the data analyst. The term was given in 1990 and has become a synonym for 'knowledge discovery in databases'. This method also forms a type of knowledge discovery technique such that data can be easily compiled and analyzed to have useful information. Data mining has two goals namely prediction and description. Description stresses on finding patterns that explain data to be interpreted by humans; whereas, prediction checks for attributes in the data to know the future state of other attributes [9].

The knowing of prediction in data mining is predictive analysis. This deals with underlying trends of patterns to understand the future probabilities and trends. It has various learning models to help clinicians diagnose, treat, and monitor patients. It uses artificial intelligence to reveal associations between the attributes that the healthcare professionals would not suspect [10]. Various studies have been done in the medical field where data mining techniques and prediction models are being developed for risk prediction (heart disease, breast cancer, renal disease, and dengue diseases), diagnosis, treatment planning, and healthcare resource management [11-14].

Several factors are known to indicate whether tobacco users are more or less likely to quit. Factors such as sociodemographic parameters, previous quitting attempts, nicotine dependence, and type of treatment given are some of them [15]. Prediction in tobacco cessation will help in assessing the patterns of tobacco usage and treatment outcomes, and create evidence-based resources for future research.

Many risk prediction models have been developed for dental diseases. Risk of developing dental caries, diagnosis, and treatment planning in orthodontic patients, and risk assessment of developing periodontal diseases [16-18] are areas researched. But as such no prediction model has been tested to understand the quit status in tobacco cessation. Hence, this novel approach aimed to understand the data mining techniques and predictive analysis using the WEKA tool and to assess the role of predictive analysis in tobacco cessation using different algorithms.

MATERIALS AND METHODS

Patient dataset:

The present study had a retrospective design and was carried out at a tobacco cessation clinic in a public tertiary care center utilizing the cohort dataset of patients who were tobacco users that were visited from 2015 to

2016. The participants were selected from retrospective data available at the tobacco cessation clinic which constituted our study setting. The patient information was collected using a pre-validated structured close-ended questionnaire used as a standard recording tool for all patients attending the tobacco cessation clinic. This ensured selection of subjects with similar sociodemographic and dependency characteristics to nullify the effect of any confounding factor on their quit status.
A total of 15 attributes which could influence the quit status of tobacco users were then identified by a single examiner. The various attributes were the sociodemographic details (including age, sex, level of education, level of income, and occupation), number of dependents, type of tobacco user (smoke form, smokeless form, dual user), number of years since habit initiated, previous attempts of quitting (yes/no, reasons for quitting and reasons for relapse), the Fagerstrom nicotine dependence scale (low, medium, or high), stage of behavior change (pre-contemplation, contemplation, preparation, action, or in maintenance) and frequency of tobacco use per day. The detailed description of the dataset is shown in Table 1. The patient details were then de-identified and subject details were anonymized to avoid selection bias. Ethical clearance was obtained from the Institutional Ethical Review Board of the Maulana Azad Institute of Dental Sciences at Delhi University. The study was conducted in full accordance with the World Medical Association Declaration of Helsinki. For data mining, various data mining tools were initially identified through a literature review, and then the WEKA tool was found to be suitable for the study purpose keeping in mind the study objectives.
Description of the Weka data mining tool:
Weka [19] is an open-source software written in Java (Waikato University, Department of Computer Science, New Zealand).

Table 1. Description of dataset used for predicting the quit status

Attributes/Variables	Description
Age	0-15 years., 15-25 years., 26-36 years., 37-47 years., 48-60 years
Sex	Male/Female
Education	Professor, Graduate or post-graduate, intermediate, high school diploma, middle school, primary school, illiterate
Marital status	Married/Unmarried
Religion	Hindu, Muslim, Sikh, Christian, others
Occupation	Profession, Semi-profession, clerical/shop owner/farmer, skilled worker, semi-skilled , unskilled, unemployed
Working hrs.	8-12 hrs., 12-16 hrs., 16-20 yrs.
No. of dependents	0-4, 4-8, 8-12, >12
Type of tobacco user	Smoke form, smokeless form
Frequency of use	10 or less, 11-20, 21-30, 31 or more
No. of yrs. since habit initiated	Less than 5 yrs., 5-10 yrs., 10-15 yrs., 15-20 yrs., more than 20 yrs.
Previous attempts	Yes /no
Stage of behavior change	Pre-contemplation, contemplation, preparation, action, maintenance
Nicotine dependence	High, medium, low
Type of intervention	Cold turkey, Behavioral counseling, Behavior counseling +nicotine replacement therapy
Outcome	Decreased the habit, quit, not quitting

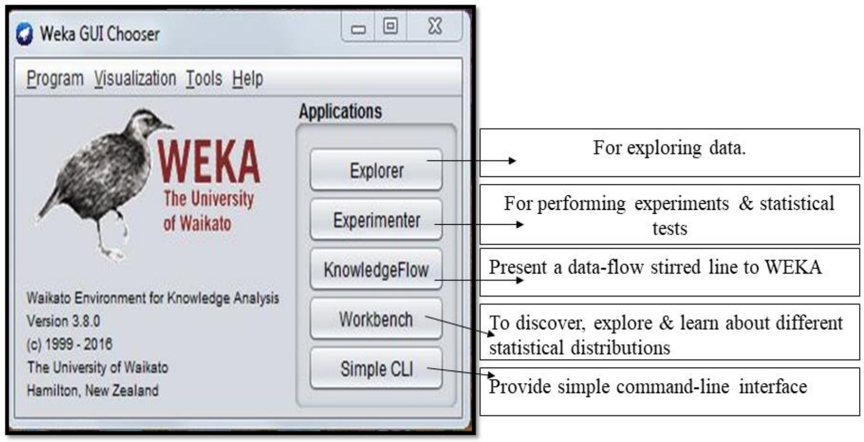

Fig. 1. Five different modes/applications in WEKA tool

It has various algorithms for classification and association of rule mining with graphical user interfaces and visualization adequacies for data exploration and evaluation. This tool has been used exhaustively in the data mining community for education and research purposes.

The tool now has made its roots in other academic fields as well as in various commercial settings. Free online availability with portability, graphical user interface, extensibility documentation, and support make it user-friendly.

The Weka tool incorporates inbuilt software for data pre-processing, classification, regression, clustering, association and visualization of cause-effect relationships.

Another way is to apply several different classifiers/attributes and compare their performance in order to choose one for prediction/outcome. In the present study, we used those classifiers which have been already researched to be effective predictive models for medical and dental diseases. The Weka has five different modes to do work in as shown in Figure 1. The pre-processing and classification in Weka can be done by different types of classifiers. Classification is the process of finding a model or function, which can describe and distinguish data classes or concept [19]. The intention of this process is to predict the end outcome of the intervention/any process based on the data attributes which in our case were the patient characteristics.

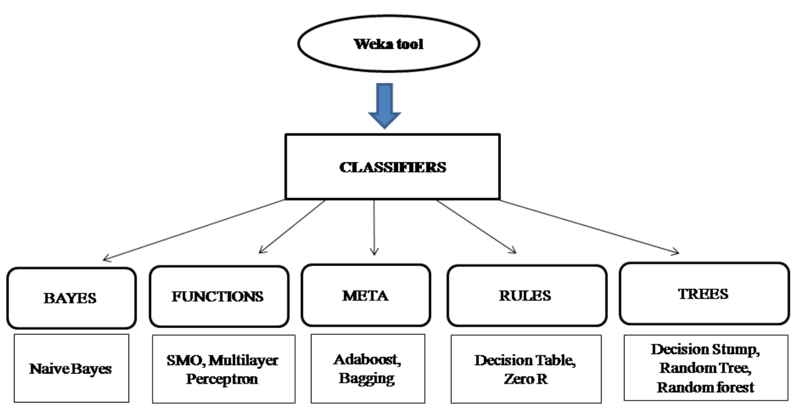

Fig. 2. Different types of classifiers in WEKA tool used for classification and prediction models

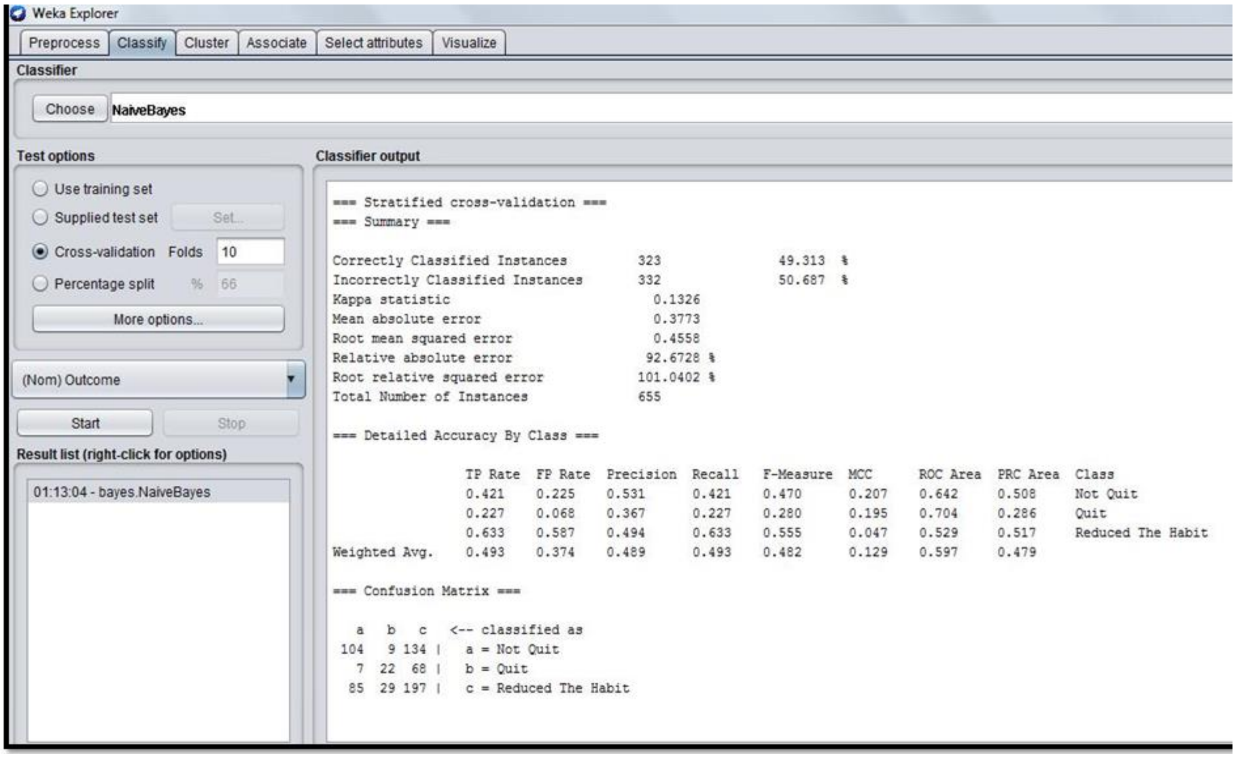

Fig. 3. An example of Naive Bayes algorithm after applying on dataset

There are 5 classifiers used for analysis in WEKA tool as shown in Figure 2. A stepwise approach was followed for predictive analysis. Prior to predictive analysis, Weka version 3.8 was downloaded from the software tool Waikato Environment for Knowledge Analysis (Weka), developed at the University of Waikato, New Zealand, and installed on a laptop computer. Orientation, sensitization and training of investigator for a period of 1 month were done for WEKA tool by a trained software engineer specialist in data mining techniques.

Retrospective data of 655 patients were obtained from the tobacco cessation clinic and entered into Excel sheets. Data to be loaded in WEKA should be in supported file formats like the comma supported value (CSV), attribute relation file format (arff), and C-4.5. Hence, our Excel data were converted to CSV format using free online software converter (convert XLS to CSV, convertio software online). The WEKA explorer was opened followed by the CSV data file. Then, the various classifier models from the classify tab option were explored and the appropriate test mode option was selected after which the start button was clicked to get the results in the form of best prediction model for tobacco cessation and predictors of quit status.

Data analysis:

Pre-processing of 655 patients' dataset was done and divided according to 15 different attributes. The next step was data undergoing different classifiers' algorithms.

In our study, comparative analysis of the classification techniques such as Random Tree, J-48, Decision Stump, Naïve Bayes and SMO was used for prediction of quit status of patients and to find a predictive model with the best prediction and accuracy. These algorithms were considered as they were used by a previous study [20] and taken as standard for comparison. The results presented were in the form of "Confusion Matrix" that denotes instances being classified as true positives, false positives and as misclassified. Thus, the WEKA tool enables to predict the outcome based on the combined effect of various classifiers/attributes. This could prove highly beneficial in the medical/dental field by predicting the outcome of a particular disease or intervention, thereby helping in decision making. The predictive accuracy of a particular classifier/algorithm is determined by the number of correctly classified instances as shown in Figure 3. This technique strongly suggests that data mining algorithms are able to predict a class for prediction.

This tool has an added advantage of classifying different attributes as per their ranking or info gain ratio by going into select attributes option and will help us to know which are the predictors of quit status. The dataset was applied with Info gain attribute evaluation. This would help us to know which attributes were the predictors of quit status as per their ranking list.

There are some other evaluation techniques to get ranking list of attributes but info gain is considered to be best as in a previous study [20]. The advantage is that it tells how much information this attribute has for outcome variable which may range from 0 (no information) to 1 (maximum information).

RESULTS

The retrospective data of 655 patients visiting a tobacco cessation clinic were used for prediction analysis with different algorithms of WEKA tool. A filtering approach was applied to our data set for computing the next level of prediction in the WEKA.

The different algorithms were applied on the data set. The tool uses stratified 10-fold validation in order to assess the performance of classification techniques for predicting a class. Predictive accuracy is determined by the tool using their confusion matrix or ability to correctly classify the dataset or instances. In our study, out of 5 classifiers, Decision Stump and SMO were found to be the correctly classifying instances. The Naive Bayes and Decision Stump took the least time in building the model (0.05 s) followed by Random Tree. The highest time was taken by SMO (1.77 s). Out of 655 instances, Decision Stump correctly classified 55.87% of instances followed by SMO (52.21%) as shown in Table 2; whereas, Random Tree showed maximum incorrectly classified instances followed by Naive Bayes.

Out of 20 attributes, 5 attributes were considered to be predictors of quitting status as per their ranking or info gain ration values. attempts of quitting (0.14) and the least belonged to the stage of behavior change (0.013). The 5 attributes are depicted in Table 3.

The highest info gain ratio value was for previous attempts of quitting (0.14) and the least belonged to the stage of behavior change (0.013). The ranking suggests that the higher the ranking position, the better the predictive accuracy of that attribute or variable for the quitting status would be.

Each of these models' confusion matrix was classified as a (not quitting), b (quit) and c (decreased the habit). This also depicts the ability of algorithms to correctly classify instances in these 3 quit outcome states. The result should be interpreted as row compared with column. Out of 655 instances, Decision Stump (366) correctly classified 91 instances as not quitting, 25 as quit and 250 as decreased the habit, and other numbers in that matrix were the misclassified instances. The same was followed for other algorithms. Decision Stump was followed by SMO (342), J-48 (338), Naive Bayes (313) and Random Tree (308) in accurately classifying the quit status.

The sensitivity and specificity of classifiers for correctly predicting the quit status and classifying them into quit, not quitting and decreased habit were also assessed. Decision Stump was found to be more accurate (55.87%) and having higher sensitivity and specificity (0.516, 0.559) as compared with other classifiers as shown in Table 4.

DISCUSSION

In India, early experiences with tobacco cessation occurred in the context of primary community education for cancer control. More recently, tobacco cessation clinics were established to develop models of intervention, and train health professionals in service delivery.

The tobacco cessation clinics need to be expanded at the primary, secondary, and tertiary care levels, and cost-effective community tobacco cessation models need to be developed.

Table 2. Comparison of different classifiers for prediction accuracy

Evaluation criteria	Classifiers				
	NAÏVE BAYES	SMO	Decision stump	J-48	Random tree
Timing to build a model (seconds)	0.05	1.77	0.05	0.17	0.06
Correctly classified instances	41.33%	52.21	55.87%	51.60%	47.02%
Incorrectly classified instances	50.68%	47.78%	44.12%	48.39%	52.97%

Table 3. Ranking of predictors for quit status

Attributes/Variables	Info Gain Ratio Values	Ranking
Previous attempt of quitting	0.14	1ST
Type of intervention	0.11	2ND
No. of yrs. since habit initiated	0.06	3RD
Reason for relapse	0.017	4TH
Stage of behavior change	0.013	5TH

Table 4. Sensitivity, specificity and accuracy of various classifiers

Classifier	Sensitivity	Specificity	Accuracy
NAIVE BAYES	0.487	0.492	41.33%
J-48	0.506	0.518	51.60%
SMO	0.515	0.518	52.21%
Decision Stump	0.516	0.559	55.87%
Random Tree	0.455	0.460	47.02%

The physicians' negative beliefs and attitudes can be resolved by emphasizing that the valuable predictors of tobacco cessation can be easily identified; thus, making an efficient use of their time [15]. Thus, our study was carried out to obtain the best prediction model using data mining techniques and WEKA tool, and assess its role in prediction of quit status in tobacco cessation.

The present study showed that prediction analysis is feasible for tobacco cessation patients. When different models were compared, Decision Stump and SMO showed the best prediction accuracy as they were able to correctly classify the predictors. This study also provided a unique opportunity to investigate individual-level predictors of tobacco cessation from 20 variables of database. Among these 20 variables, previous attempts of quitting, type of intervention and number of years since habit initiated were found to be higher predictors for quitting tobacco in the present study setup.

There are no studies of prediction models for tobacco cessation for comparison with the present study results. However, an attempt was made to discuss the results and the application of prediction models.

The different classifier models have correctly classified instances ranging from 41.33% to 57.87% and timing to build a model that ranged from 0.05 to 1.77 s, showing that it is time saving for analysis. However, some other studies [21-23] in different contexts not related to tobacco cessation have shown correctly classified instances better than our study results. These differences may be due to the use of retrospective data in our study which had been used for clinical purposes. Also, so many misclassified instances were present that might have affected the final analysis. However, date collection and entry modification to suit the prediction analysis have to be done and standardized.

Many data mining techniques have predicted the odds of developing heart disease, kidney disease, cataract, viral diseases such as dengue, sickle cell anemia, diabetes and for breast cancer survival analysis as stated in a study in 2013 [23]. According to a study, Naive Bayes algorithm plays a key role in predicting liver disease [24]; whereas, Random Tree was better for analysis of sickle cell anemia [21,22]. Joshi et al. [12] showed that LMT classifier and Bayes Network give more accurate diagnosis and prognosis for breast cancer. Algorithms such as Naïve, J48, KNN, and C4.5 were used for classification to diagnose diseases such as heart disease, AIDS, brain cancer, diabetes, kidney disease, and viral diseases such as dengue, and hepatitis C [23]. The comparative study analysis [12,21-23] revealed high accuracy i.e. 97.77% for cancer prediction.

In dentistry, studies of predictive analysis are sparse and are related to diagnosis, treatment planning, and risk assessment of dental caries, malocclusion or periodontal disease using Artificial Neural Networks and other risk models [16-18]. None of these studies have used the data mining techniques or WEKA tool for disease or risk prediction. Dental caries prediction models such as the Markov model [25] and Cariogram [26] were researched but not the data mining techniques. A properly trained neural network could effectively be used for periodontitis [18] and dental caries risk prediction [16]. A previous study analyzed the prediction of survivability of oral cancer patients using predictive models. They found that the Tree Boost model was marginally better than the Single Tree and Decision Tree Forest and had considered 18 predictors [27].

Tobacco cessation, which is presently a priority, has lots of road blocks for its applicability [28]. Hence, understanding the predictors will help in better treatment strategies and guide professionals to effectively intervene. The WEKA tool, which is being used for prediction, has advantages such as open source and being freely available, maintainable and fully implementable in JAVA and runs on almost any platform. However, the WEKA tool has a restricted application to small or medium-sized datasets, and sub-sampling is required for larger datasets. This study is the first of its kind on data mining techniques in tobacco cessation, hence comparison with other studies could not be done, and the predictors of other diseases in the literature were mentioned wherever deemed necessary. One inherent quality of WEKA is that it couples the ROC curve with the accuracy and confusion matrices Brier score, and other measures, such as the k index. Standardized, validated and complete clinical variables should be used which can be applicable for WEKA tool. Among the variables, the follow up periods were not included in the analysis so complete abstinence category could not be justified. More prospective studies with standardized variables should be considered to prove the accuracy, usefulness and precision of these prediction models with clinical outcomes.

Our study results cannot be generalized to other populations as the used data belonged to patients visiting a tertiary dental care center in India. However, the WEKA tool has an added advantage and can be used for prediction analysis and also help in solving the problems of clinical research.

CONCLUSSION

Our study results conclude that the prediction models can be applicable for tobacco cessation services, and Decision Stump as well as SMO had the best prediction accuracy to correctly classify the predictors. Previous attempts of quitting, type of intervention, and number of years since habit initiated were the best predictors of quitting status.

Data mining and predictive analytics will not only improve patient outcomes but understand variables or combination of variables for effective interventions in tobacco cessation. They also provide patient centric approach towards new and hidden patterns in data, from which the knowledge is being generated. This knowledge can help in providing medical and other services to the patients. Healthcare institutions that use data mining techniques have the opportunity to predict future requirements, needs, desires, and health status of the patients and to make adequate and optimal decisions about their treatments. Evidence-based care can be planned and executed better.

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES

1. World Health Organization, Research for International Tobacco Control. WHO report on the global tobacco epidemic, 2008: the MPOWER package. World Health Organization; 2008 Feb 11.
2. Jha P, Jacob B, Gajalakshmi V, Gupta PC, Dhingra N, Kumar R, et al. A nationally representative case–control study of smoking and death in India. N Engl J Med. 2008 Mar;358(11):1137-47.
3. Shimkhada R, Peabody JW. Tobacco control in India. Bulletin of the World Health Organization. 2003 Jan;81(1):48-52.
4. Thankappan KR. Tobacco cessation in India: A priority health intervention. Indian J. Med. Res. 2014 Apr;139(4):484.
5. Murthy P, Saddichha S. Tobacco cessation services in India: Recent developments and the need for expansion. Indian J Cancer. 2010 Jul;47(5):69.
6. World Health Organization. WHO report on the global tobacco epidemic, 2013: enforcing bans on tobacco advertising, promotion and sponsorship. World Health Organization; 2013.
7. Kaur J, Jain DC. Tobacco control policies in India: implementation and challenges. Indian J Public Health. 2011 Jul;55(3):220.
8. Persai D, Panda R, Gupta A. Examining implementation of tobacco control policy at the district level: a case study analysis from a high burden state in India. Adv Prev Med. 2016 Jan;2016.
9. Hand DJ, Mannila H, Smyth P. Principles of data mining. MIT Press; 2001. Retrieved from: https://rb.gy/ppygoo
10. Koh HC, Tan G. Data mining applications in healthcare. J Healthc Inf Manag. 2011 Jan;19(2):65.
11. Vijayarani S, Sudha S. Comparative analysis of classification function techniques for heart disease prediction. Int J Innov Res Comp Comm Eng. 2013 May;1(3):735-41.
12. Joshi J, Doshi R, Patel J. Diagnosis and prognosis breast cancer using classification rules. Int. J Eng. Res. Gen. Sci. 2014 Oct;2(6):315-23.
13. Vijayarani S, Dhayanand S, Phil M. Kidney disease prediction using SVM and ANN algorithms. Int J Comput. 2015 Mar;6(2):1-2.
14. Kumar MN. Alternating decision trees for early diagnosis of dengue fever. arXiv preprint arXiv:1305.7331. 2013 May.
15. Caponnetto P, Polosa R. Common predictors of smoking cessation in clinical practice. Respir Med. 2008 Aug;102(8):1182-92.
16. Kositbowornchai S, Siriteptawee S, Plermkamon S, Bureerat S, Chetchotsak D. An artificial neural network for detection of simulated dental caries. Int J Comput Assist Radiol Surg. 2006 Aug;1(2):91-6.
17. Bahaa K, Noor G, Yousif Y. The artificial intelligence approach for diagnosis, treatment and modelling in orthodontic. Principles in contemporary orthodontics. 2011 Nov 25:451.
18. Shankarapillai R, Mathur LK, Nair MA, Rai N, Mathur A. Periodontitis risk assessment using two artificial neural networks-a pilot study. Int. J Dent Clin. 2010 Dec;2(4).
19. Bouckaert RR, Frank E, Hall M, Kirkby R, Reutemann -9-P, Seewald A, Scuse D. WEKA manual for version 3kato, Hamilton, New Zealand. 1. The University of Wai2016 Dec 19. http://infochim.u-strasbg.fr/cgi-bin/Weka-3-9-1/WekaManual.pdf
20. Sunil C, Rama J. Comparison of Performances of different Svm implementations when used for automated evaluation of descriptive answers. ICTACT Journal of Soft Computing. 2015 Apr:971-78.
21. Solanki AV. Data mining techniques using WEKA classification for Sickle Cell Disease. Int J Inf Technol. 2014;5(4):5857-60.
22. Yasodha P, Kannan M. Analysis of a population of diabetic patients databases in WEKA tool. Int J Sci Eng Res. 2011 May;2(5):1-5.
23. Durairaj M, Ranjani V. Data mining applications in healthcare sector: a study. Int J Sci Tech Res. 2013 Oct;2(10):29-35.
24. Dhamodharan S. Liver Disease Prediction Using Bayesian Classification. In 4th National Conference on Advanced Computing, Applications & Technologies 2014 May.
25. Kopycka-Kędzierawski DT, Billings RJ. Application of nonhomogenous Markov models for analyzing longitudinal caries risk. Community Dent Oral Epidemiol. 2006 Apr;34(2):123-9.
26. Bratthall D, Hänsel Petersson G. Cariogram–a multifactorial risk assessment model for a multifactorial disease. Community Dent Oral Epidemiol. 2005 Aug;33(4):256-64.
27. Sharma N, Om H. Data mining models for predicting oral cancer survivability. Netw Model Anal Health Inform Bioinform. 2013 Dec;2(4):285-95.
28. Oberoi SS, Sharma G, Nagpal A, Oberoi A. Tobacco cessation in India: how can oral health professionals contribute. Asian Pac J Cancer Prev. 2014 Jan;15:2383-91.