



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 Issue: IV Month of publication: April 2018

DOI: <http://doi.org/10.22214/ijraset.2018.4612>

www.ijraset.com

Call: ☎ 08813907089

E-mail ID: ijraset@gmail.com

A Disease Prediction by Human Habits Using Modify Naive Bayes Classifier

Ramesh Piprotar¹, Prof. Daxa Vekariya²

¹Computer Department, Noble Engineering College - Junagadh

²Prof. Noble Engineering College - Junagadh Gujarat Technological University

Abstract: With development of technology human habits change with technology changes. Changing human habits its produces some diseases by specific habits. For example habits of eating sweet causes diabetes. Some habits likes drinking some drinks, chewing gutka, eating some foods, taking drugs, working styles, types of job, types of works, sleep time etc., responsible for specific disease. Now problem is how to predict human diseases by human habits? We propose different machine learning techniques to solve this problem. In this work includes many patients with various diseases. Dataset collects from many doctors. This dataset consist patient's diseases and its habits. This work will useful for diseases prediction causes by human habits. This work useful for decreasing diseases by changing human habits. **Keywords:** Human diseases, Human habits, Naïve Bayes classifier, questionnaires, prediction.

Keywords: Human diseases, human habits, Naïve Bayes classifier, weighted naïve Bayes classifier, prediction, probability.

I. INTRODUCTION

In 21st Century, with development of technology human lives drastically change. This changes changed human habits. Because of changes of human habits its produce some specific diseases. There for it is important to know identify these habits which are responsible for diseases. So, we can stay safe from those diseases. There are lots of techniques for classification of classifying data and finding useful information. Data mining is used to discover patterns and relationships in the data in order to help make better decisions. Data mining process is very useful for finding knowledge and solution of many problems. Nowadays classification used in many applications like sentiment analysis, disease analysis, weather analysis and so on. Using data mining process we can find some human habits, which causes various diseases. After analysis we can aware and be safe for our health. It's helpful to our society. There are lots of data of human health but its needs to structured and find useful information from unstructured data. I try to structured human diseases data and human habits. After that I apply classification algorithm to classifying data and find results. I used Naïve Bayes classifiers for classifications of data.

Naive Bayes (NB) classifiers were introduced many years ago as a practical simplification of the Bayes inference theorem. Despite simplified assumptions they proved quite effective in solving numerous problems, particularly in such cases where the data amount is large and where fast learning is needed.

Their efficient use in voice recognition, weather forecasting, medical applications etc. Naïve Bayes Classifiers compare training set with dataset and calculate probability. Naive Bayes classifiers working based on probability but sometime probability is zero than its difficult to prediction. This limitation negated Using Modify Naive Bayes classifiers. Naive Bayes is a family of probabilistic algorithms that take advantage of probability theory and Bayes' Theorem to predict the category of a Training Set (like a piece of a customer review).

They are probabilistic, which means that they calculate the probability of each category for a given Training Set, and then output the category with the highest one. There are many classification algorithms available like Space Vector Machine Learning (SVM), Naïve Bayes Classifiers (NB Classifiers), K-Means Algorithm, C4.5 Algorithm and ID3 Algorithm.

Modify Naïve Bayes Classifier is more suitable for mining in medical data. We have used weighted Naïve Bayes Classifier in our proposed system.

Using weighted Naïve Bayes Classifier we can predict diseases caused by human habits and we can identify habits. Generating database about diseases and patients habits we can find disease cause by specific human habits. I classified some habits likes exercise habits divided into regular, irregular and no, eating habits divided into light, heavy, late night etc. I have collected patient's data from various patients. I have also collected habits related data from patients. It's useful for finding probability of diseases caused by specifics habits. Naïve Bayes Algorithm compares training set with patient's data and calculates diseases probability. If we are aware of our habits then we will be care our health. It's useful for human health. That's give us to inspiration to do research on that direction.

II. OBJECTIVES

A. The Main Objective of This Research is Shown Below

- 1) Find habits these responsible for specific disease
- 2) Classify diseases caused by some habits
- 3) Try to reduce bad assumption
- 4) Classifying human habits and diseases.
- 5) Improve human health

III. MOTIVATION

The motivation behind this approach is that, using data mining technique predict diseases causes by human habits. Using this approach we can aware about our health and reduce bad habits. The common motivation is that the structure of documents may contain useful information for classification. We show how we can improve the accuracy of classification. One of the goals of Data Mining is to develop the methods for people based on patient's data. Thus, to develop a user adaptable system, the patients habits that occurring before disease. In this chapter, we introduce a data mining techniques and natural language processing approaches for automated detection patients habits and behaviours like eating foods, drinking, times spending, sleeping, exercise etc.

IV. NEED OF RESEARCH

Need of research is to develop a more efficient and effective use of Naïve Bayes Classifier in medical. Using Naïve Bayes algorithm in predict disease causes by human habits. Due to change of technology, human habits are change. Because of some habits we are getting some diseases. There for it needs to identify habits, these are responsible for diseases. Using modify Naïve Bayes Clarifier algorithm we can easily identify habits these are responsible and we can reduce our bad habits.

V. METHODOLOGY

A. Introduction Of Naive Bayes Classifier

The Naive Bayes Classifier technique is based on Bayesian theorem. The Naive Bayes Classifier particularly used when the dimensionality of the inputs is high. It is a classification technique based on an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. It is easy and fast to predict class of test data set. It also performs well in multi class prediction. Bayes theorem provides a way of calculating the posterior probability, $P(c|x)$, from $P(c)$, $P(x)$, and $P(x|c)$. Naive Bayes classifiers assume that the effect of the value of a predictor (x) on a given class (c) is independent of the values of other predictors. This assumption is called class conditional independence.

B. Basic Concept of Modify Naïve Bayes Classifier

The traditional Naïve Bayes Classifier was designed considering the fact that attribute have same importance and in the database simply their presence or absence is mentioned. In the medical field all the symptoms does not equally contribute in predicting a particular disease. For example in medical domain predicting the probability of lung cancer, the smoking attribute is having more impact than the exercise attribute. In the proposed framework called Weighted Naive Bayes Classifier (WNBC) that assigns different weights to different attributes according to their predicting capabilities by consulting the expert doctors. We have given weight to each attribute as per doctor's experience. Domain based weights are used to assign weight of each attribute using expert knowledge. Experiments have been given more accurate result compare to traditional Naïve Bayes Classifier. We have given weight 1 to 5 to the each attribute as per expert knowledge. Different diseases have different weight.

C. Naive Bayesian Approach in Disease prediction.

The endeavour of the classification is to build classifier model that work on Bayes theorem. Then, the classifier is used to predict attributes of new training set. This prediction is based on probability. Unlabelled data compared with dataset and find out probability of various diseases based on human habits.

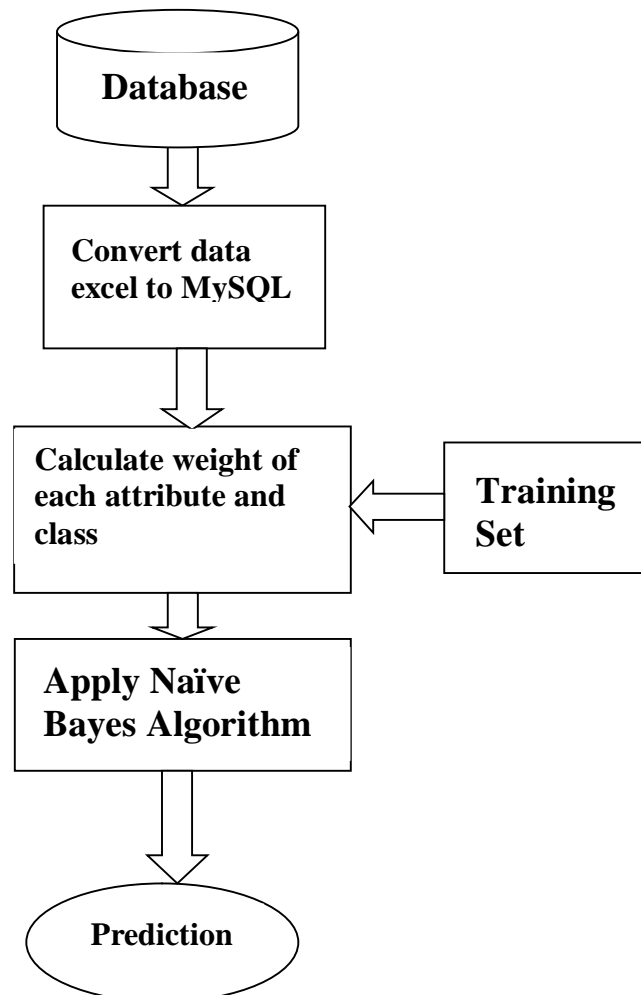


Figure. Steps of Prediction using Weighted Naive Bayes Classifiers

Above fig. describe the steps of prediction. Enter attribute values in dataset in excel format. After it convert excel data into MySQL so that we can apply some query easily. When we enter training set it is calculate weight of each attribute and find individual probability of each attributes. Training set has set of habits. Then, apply naïve Bayes classifier algorithm that predict diseases.

D. Basic Algorithm Proposed System

Basic Architecture Proposed System Here W_i is a weight of habits. Each habit has 1 to 5 weights. W is a total weight of habits. $P(H)$ is a hypothesis of all attributes. $P(E)$ is a evidence. $P(H/E)$ is gives probabilities of hypothesis. In our proposed algorithm, first insert training set and that training set compare with dataset. In next step that calculate weight of each attribute and then, find probability of each attributes of training sets. In last step it calculates probability of diseases using Naïve Bayes algorithm. Attribute weight Attribute weight is assigned depending upon the domain. For example in case of diabetes, exercise assigned highest weight based on the symptoms caused diabetes and smoking has low weight.

$$P(C_w) = C_i * W_i / W$$

E. Algorithm

- 1) Enter Habits 1, 2,.....
- 2) s.execute(query)
- 3) Attribute Weight (W)
- 4) $P(C_w) = C_i * W_i / W$
- 5) $P(E) = P(C_i/X) / P(X)$
- 6) $P(H|E)$
- 7) Print Probability of diseases

VI. RESULTS ANALYSIS

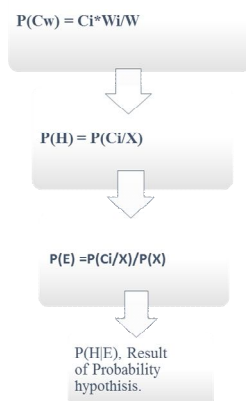


Fig. Steps of Algorithm

TABLE 6.1 ACCURACY COMPARISONS

HABITS	WEIGHT
EXERCISE	5
SLEEPING	3
EATING	3
WORKING_AREA	2
WORKING_STYLE	1
ALCOHOLIC	2
COLD_DRINKS	2

TABLE 6.2 WEIGHT OF HABITS FOR DIABETES

sR_No	DATASETS	ALGORITHMS	PERCENTAGE OF ACCURACY
1	SMALL	NAÏVE BAYES	81.50%
2	SMALL	WEIGHTED NAÏVE BAYES	83.23%
3	LARGE	NAÏVE BAYES	82.15%
4	LARGE	WEIGHTED NAÏVE BAYES	85.54%

VI. CONCLUSIONS

There are lots of techniques for Classification and find out interesting results and solutions from large data. We proposed a new mechanism of weighted Naïve Bayes Classification to predict diseases from human habits. We used Weighted Naïve Bayes Classifier for calculate probability of diseases. In Weighted Naïve Bayes Classifier each attribute assign some specific weight because all symptoms of disease are not equal for caused disease. The Weighted Naïve Bayes Classifier gave better accuracy compare to existing Traditional Naïve Bayes Algorithm. We tried to achieve more efficiency and high utilization of Naïve Bayes Classifier for human habits and find habits which responsible for some specific diseases. Our Proposed work we applied Weighted Naïve Bayes Classifier on diabetes. It can apply all diseases for find out more accurate result. It is future work of this proposed work.

REFERENCES

- [1] Divya Tomar and Sonali Agarwal, "A survey on Data Mining approaches for Healthcare", 2013IJBSBT.
- [2] Shweta Kharya, Sunita Soni, "Weighted Naive Bayes Classifier: A Predictive Model for Breast Cancer Detection" January 2016 International Journal of Computer Applications.
- [3] Manav Masrani and Poornalatha G, "Twitter Sentiment Analysis Using a Modified Naïve Bayes Algorithm", 2 sep2017, Proceedings of 38th International Conference on Information Systems Architecture and Technology – ISAT 2017.
- [4] Xianshan Li, Fengda Zhao, Lingfu Kong and Peiliang Wu, "Learning Human Daily Behavior Habit Patterns Using EM Algorithm for Service Robot" 2007 IEEE International Conference on Robotics and Biomimetic December 15 -18, 2007, Sanya, China.
- [5] Sang-Bum Kim, Kyoung-Soo Han, Hae-Chang Rim, and Sung HyonMyaeng, "Some Effective Techniques for Naive Bayes Text Classification", Nov 2006 IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING.
- [6] Dr. Devendra K. Tayal, Amita Jani, Kanak Meena "Development of Anti-Spam Technique Using Modified K-Means & Naïve Bayes Algorithm", 31 October 2016 Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on.
- [7] Karl-Michael Schneider, "Techniques for Improving the Performance of Naive Bayes for Text Classification", 6th International Conference, CICLing 2005, Mexico City
- [8] Marcella Dimaggio, Francesco Leotta, Massimo Mecella, Daniele Sora, "Process-Based Habit Mining: Experiments and Techniques," 2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress.
- [9] Han-joon Kim¹ and Jae-young Chang², "Improving Naive Bayes Text Classifier with Modified EM Algorithm", International Symposium on Methodologies for Intelligent Systems.
- [10] Kanupriya Sharma, Amanpreet Kaur, "Personality prediction of Twitter users with Logistic Regression Classifier learned using Stochastic Gradient Descent", IOSR Journal of Computer Engineering (IOSR-JCE).
- [11] Dejan Markovikj, Sonja Gievska, Michal Kosinski, David Stillwell, "Mining Facebook Data for Predictive Personality Modeling", AAAI Technical Report WS-13-01 Computational Personality Recognition.
- [12] Pedro Domingos, Michel Pazzani, "On the Optimality of the Simple Bayesian Classifier under Zero-One Loss", 1997 Kluwer Academic Publishers.
- [13] Radu Dogaru, "A Modified Naïve Bayes Classifier for Efficient Implementations in Embedded Systems" IEEE 2011 Signals, Circuits and Systems (ISSCS), 2011 10th International Symposium on.
- [14] WEBSITE:
- [15] http://www.sersc.org/journals/IJBSBT/vol5_no5/25.pdf
- [16] <https://machinelearningmastery.com/better-naive-bayes/>



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)