International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8, Issue-1, May 2019

Design of Disease Prediction System using Bayes Network with Android Application

S.Rajaprakash, S. Muthuselvan, K.karthik, Vikrant Pradhan, Abhay kumar

Abstract*: Health is a major concern in the current era. It is more evident in a developing and dense country like India. The number of patients highly exceeds the treatment facilities that are available at any given instance. Moreover the growing medical expenses for checks up are ever rising. Due to this the lower-class masses which comprises of major portion of the population do not get the care they deserve. Due to the population density the incoming patients are more in general hospital, so it difficult to manage out patients which leads lot issue. Bayes network in machine learning one of the important methods to prediction method and also good output over vagueness data. To overcome the above issue, diseases prediction system are framed using Naive bayes network are framed with help of android application. It will be useful for the out patients and doctors. Android development as the front end that serves the user with the basic UI for the input of symptoms and outputting the predicted output.

***Index Terms*: Bayes Network, Android application, Diseases Prediction.**

I. INTRODUCTION

Machine Learning as we know it is widely trending in the recent technological sector. And rightfully it has proven to be a major topic due to its capabilities. This capability has been the major motivation factor that gave the idea for this project. We have incorporated machine learning with yet another technology that brought about a sort of revolution in the mobile phone world android ecosystem. The main objective of this work has been to show how these technologies combined can help people in positive ways. Health Care being one of the major areas of concern has therefore been the target area of this project. Due to wide exposure of android devices it makes it easier for people to receive medical diagnosis on the tips of their finger. The current situation has been that for any occurrence of symptoms the patient has to visit a doctor or diagnostician to find out what disease someone might be having. This is time consuming as well as money consuming. The healthcare expenses are rising day by day and it is being very difficult for common people to get good healthcare with the current economy. The project shows how we can apply different technologies to address such problems and provide cheaper and better healthcare to the common public.

II. CONDITION PROBABILITY

Conditional probabilities arise naturally in the investigation of experiments where an outcome of a trial may affect the outcomes of the subsequent trials.

We try to calculate the probability of the second event (event B) given that the first event (event A) has already happened. If the probability of the event changes when we take the first event into consideration, we can safely say that the probability of event B is dependent of the occurrence of event A.

We can write the conditional probability as, the probability of the occurrence of event A given that B has already happened.

$$P\left(\frac{A}{B}\right) = \frac{P(AandB)}{P(B)} = \frac{Probabilityoftheoccuranceofboth AandB}{ProbabilityofB}$$

III. PRIOR PROBABILITY

Earlier probability, in Bayesian statistical inference, is the possibility of an occasion before new statistics is gathered. This is the pleasant rational assessment of the probability of an outcome based totally at the cutting-edge expertise earlier than a test is achieved.

IV. POSTERIOR PROBABILITY

The posterior probability of a random event or an uncertain proposition is the conditional probability that is assigned after the relevant evidence or background is taken into account. Similarly, the posterior probability distribution is the probability distribution of an unknown quantity, treated as a random variable, conditional on the evidence obtained from an experiment or survey. "Posterior", in this context, means after taking into account the relevant evidence related to the particular case being examined.

Revised Manuscript Received on 30 May 2019.

***** Correspondence Author

Dr. S. Rajaprakash*, Dept. of Computer Science and Technology, Aarupadai Veedu Institute of Technology, Vinayaka Missions Research Foundation Chennai, India.

S. Muthuselvan, Dept. of Computer Science and Technology, Aarupadai Veedu Institute of Technology, Vinayaka Missions Research Foundation Chennai, India.

K. Karthik, Dept. of Computer Science and Technology, Aarupadai Veedu Institute of Technology, Vinayaka Missions Research Foundation Chennai, India.

Dr.K.Somasundaram Dept. of Computer Science and Technology, Aarupadai Veedu Institute of Technology, Vinayaka Missions Research Foundation Chennai, India.

Vikrant Pradhan Dept. of Computer Science and Technology, Aarupadai Veedu Institute of Technology, Vinayaka Missions Research Foundation Chennai, India.

Abhay kumar Dept. of Computer Science and Technology, Aarupadai Veedu Institute of Technology, Vinayaka Missions Research Foundation Chennai, India.

V. BAYES NETWORK

Bayesian networks are a kind of probabilistic graphical version that makes use of Bayesian inference for opportunity computations. Bayesian networks intention to model conditional dependence, and therefore causation, via representing conditional dependence with the aid of edges in a directed graph. Via these relationships, you can successfully conduct inference on the random variables inside the graph thru the use of things.

A Bayesian network is a directed acyclic graph in which every facet corresponds to a conditional dependency, and every node corresponds to a unique random variable. Officially, if an edge (a, b) exists inside the graph connecting random variables A and B, it means that P(B|A) is an aspect in the joint probability distribution, so we should understand P(B|A) for all values of B and A in order to conduct inference. Within the above example, when you consider that rain has an area going into WetGrass, it means that P(WetGrass|Rain) will be a thing, whose probability values are unique subsequent to the WetGrass node in a conditional chance desk.

Bayesian networks satisfy the local markov belongings, which states that a node is conditionally independent of its non-descendants given its mother and father. Inside the above example, this means that P (Sprinkler|Cloudy, Rain) = P (Sprinkler|Cloudy) since Sprinkler is conditionally independent of its non-descendant, Rain, given Cloudy. This belongings lets in us to simplify the joint distribution, acquired inside the preceding phase the usage of the chain rule, to a smaller shape. After simplification, the joint distribution for a Bayesian network is identical to the made of P (node|parents(node)) for all nodes, stated below:

$$P(X_1,\dots,X_n) = \prod_{i=1}^{n} P(X_i|X_1,\dots,X_{i-1}) = \prod_{i=1}^{n} P(X_i \mid Parents(X_i))$$

In large networks, this property lets in us to significantly lessen the quantity of required computation, because commonly, maximum nodes will have few mother and father relative to the general size of the network.

VI. ANDROID

Android is a mobile operating system developed by Google. It is based on a modified version of the Linux kernel and other open source software, and is designed primarily for touchscreen mobile devices such as smartphones and tablets. Initially developed by Android Inc., which Google bought in 2005, Android was unveiled in 2007, with the first commercial Android device launched in September 2008. The operating system has since gone through multiple major releases, with the current version being 9 "Pie", released in August 2018. Android has been the best-selling OS worldwide on smartphones since 2011 and on tablets since 2013. As of May 2017, it has over two billion monthly active users, the largest installed base of any operating system, and as of December 2018, the Google Play store features over 2.6 million apps.

VII. PAST WORK

Machine Learning has been a significant research topic that has been incorporated within the heath sector. From monitoring health, analyzing x-rays, CT-scans to making advancement in genomics. Prediction of health issues is one of the major areas where classification or regression algorithms are run on data to give a probable result. One of the major areas has been in identification of tumors and cancers at earlier stages using supervised learning approach from large number of past data. This would have been impossible to be performed by doctors alone with high accuracy rates.

In the year 2012 Rafe Torabi et al. proposed a model in bayeian network using the maximum likelihood algorithm to predict the student' scores based on the students attributes and educational history and he tested system using the 500 different student in various information technology university levels [1]. In the year 2012 Eisuke Kita, Yi Zuo proposed a model to prediction stock prices by using a baysian network model. First, the network is determined from the daily stock price and then, it is applied for predicting the daily stock price which was already observed. NIKKEI stock average and Toyota motor corporation stock price were considered as the numerical examples. In case of NIKKEI stock average prediction, the average and the maximum errors of the present algorithm are 6% and 30% below them of the time-series prediction algorithms, respectively. [2] In the year 2013 Ehsan Nazerfard proposed a model to predict daily activity using Bayesian network. They proposed a novel two-step inference process to predict the next activity features and then to predict the next activity label. They also proposed an approach to predict the start time of the next activity which is based on modeling the relative start time of the predicted activity using the continuous normal distribution and outlier detection. [3] In the year 2015, Shiva Asadianfam proposed a paper that reviewed Number of high school students in the fields of study: mathematics, Experimental Sciences, humanities, vocational, business and science were studied were compared. The purpose of this research was to predict the academic major of high school students using Bayesian networks. The effective factors were used in academic major selection for the first time as an effective indicator of Bayesian networks [4] In the year 2008, N Fenton proposed a model that uses of Bayesian networks (BNs) in predicting software program defects and software program reliability. The approach allowed analysts to comprise causal procedure elements as well as integrate qualitative and quantitative measures, consequently overcoming a number of the standard obstacles of conventional software program metrics methods.[5]In the year 2012 Russell J Kennet proposed a paper to examine the use of Bayesian networks for improving weather prediction, applying them to the problem of predicting sea breezes. The Bayesian network performed on par with those generated automatically by data mining. [6]
In the year 2015 S Karthika proposed a classification methodology utilizing the benchmark Naïve Bayesian classification algorithm for the classification of persons into different classes based on several attributes representing their educational qualification. The performance results prove that the classical Naïve Bayesian classification algorithm performs well when the attributes are non-numerical. [7]

In the year 2016 Selvani Deepthi Kavila proposed a framework which facilitates a researcher to recognize a document based on its domain. This technique is capable of identifying the main research area as well as the sub area by creating a word list for each area of paper and updating it as new domains are added. [8]

In the year 2012 Shadab Adam Pattekari proposed a System this research is to broaden an wise gadget using facts mining modeling approach, specifically, naive Bayes. It is implemented as web based utility in this consumer answers the predefined questions. It retrieves hidden data from saved database and compares the person values with educated statistics set.[9]

In the year 2017 Thomas I. Aruleba proposed a system this research has developed hypertension predictive system using data mining modelling technique, namely, Naïve Bayes. Medical profiles such as age, sex, blood pressure, chest pain and blood sugar it can predict the likelihood of patients getting a hypertension.[10]. In the year 2018 Arun R proposed a system to predict heart diseases based on Bayesian Network. For the given symptoms it was able to predict the type and criticality of the disease to a very good degree. [11]

In the year 2018 Santos A shinde proposed a paper that reviews automation the process of predicting diseases more accurately using machine learning methods. [12]

In the year 2013 Zhiqiang Cai proposed a failure prediction system that models automatically from historical failure dataset. The paper proposes an effective algorithm to build the failure prediction Bayesian Network model with data mining technology. [13]

VIII. LIMITATION

What we have tried to address in this work that is difference is the easy usability. It is meant as a general system by incorporate everyday symptoms and diseases instead of having a narrow scope that is limited to only a certain critical disease. Of course, this means the project restricts itself in providing detailed diagnosis for these critical diseases but it also means it can be used by a larger scope of users who can just as easily get basic diagnosis results. Moreover, building it over the android and having algorithms run in server makes it a scalable and robust architecture and provides user good experience.

IX. METHODOLOGY

Step 1: The user Logs In to the app. (login is required to store symptoms search history for the user)

Step 2: In the search input provided in the android app, symptoms are entered and submitted.

Step 3: This list of symptoms are sent to the backend server using the http service.

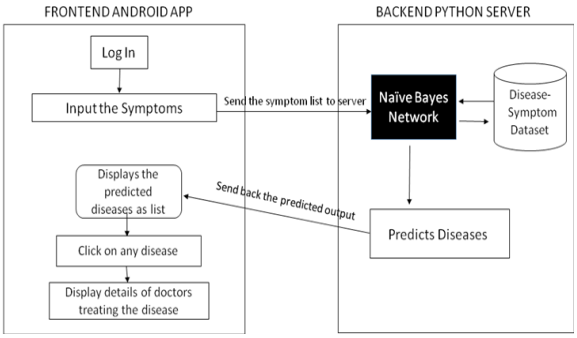

Figure-1 Architecture

Step 4: In the backend, the list is fed to a machine learning algorithm that has been trained to predict diseases based on a large dataset of symptoms-diseases.

Step 5: The algorithm outputs the prediction as a list of diseases with their corresponding probability.

Step 6: This output of prediction is then sent back to the app.

Step 7: The app reads the response and presents to the user as a list of diseases with the corresponding probability of having that disease.

Step 8: The user can further click on the disease to get more details on the disease and the doctors associated the disease.

Dataset: The dataset used for the machine learning model was taken from the website http://people.dbmi.columbia.edu/~friedma/Projects/DiseaseSymptomKB/index.html. The dataset is represented as an html table in the website. Python Pandas was used to extract the data from the html table. The data was then cleaned and all Null values removed in order to pass it as input to the machine learning model.

X. IMPLEMENTATION

The work is implemented in two different parts. The Frontend has been implemented as an android app. We have used React Native (a JavaScript Library for building apps for iOS and Android) to build the app. It is a simple app where a user can enter the symptoms and he/she will get the predicted diseases as output with the corresponding probability.

The main implementation of the algorithm is done in the server side using python as the programming language. We have used Bayes Network Classification algorithm for the machine learning process. To implement the prediction system a dataset is used which was available in the internet as an html table with each row corresponding to a disease and a symptom. The data was extracted and converted into a csv file and the data was cleaned and processed using pandas (a python library for data science). This was necessary for the machine learning algorithm to work correctly and to provide the input to the model in the correct input format. For backend we used flask (a python framework to create backend applications) to implement the server-side application.

The way it works is that after the user provides the list of symptoms and submits, the list is sent via post request to the backend app.

In the server side the program processed the symptoms names and fetches code for each symptom from a MySQL database table. The corresponding codes are need to make the machine learning process efficient. This is then converted into standard input for the Bayes Classification algorithm with pandas and passed to the model. The model outputs prediction set with disease codes and prediction. This disease code is then looked up in the MySQL table for the corresponding Disease Name and finally sent back to the android application as a JSON format. The App then reads the JSON file and displays it as a list.

XI. RESULT AND DISCUSSION

As a sample input for the system we used a symptom list of the following.

1) Shortness of breath, chest pain, nausea, sweat, vomiting.

The prediction system outputs the following prediction.

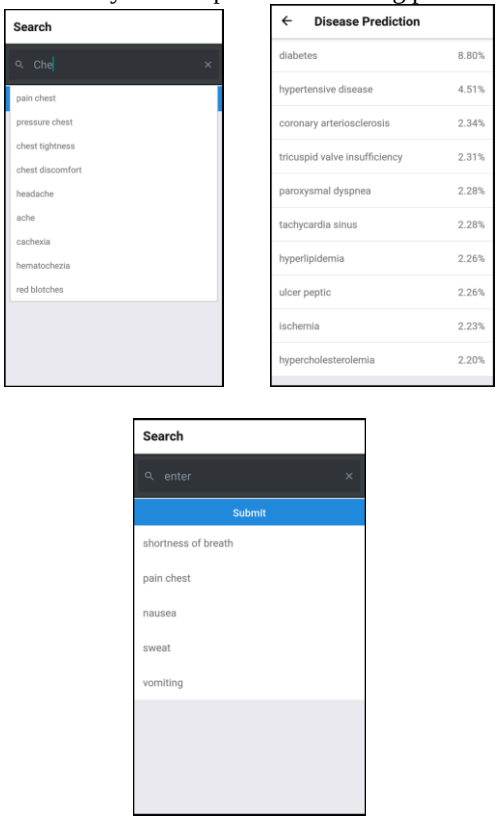

Figure-2 Screen shots in Android Application

We get the following prediction:

Diabetes - 8.80%

Hypertensive disease (Heart condition due to High Blood Pressure) - 4.51%

Coronary arteriosclerosis – 2.34%

For this set the prediction coincides with good prediction of the disease as these symptoms reflect the main symptoms for diabetes and High Blood Pressure.

2) Abdominal pain, vomiting, nausea, decreased body weight, dizziness

The prediction system outputs major chance of having gastritis (inflammation in the lining of stomach). This also outputs good prediction for the given symptoms.

XII. CONCLUSION:

We need to address that the project only reflects a very small part of the disease set and is only accurate to a factor depending on the diseases and symptoms given on the dataset. The model however is scalable and with more diseases added the model can become more accurate. This project can only be used for general diseases and symptoms and for more specific diagnosis the user must visit the doctor for physical examination. For example, by just giving symptoms it not not possible to predict the occurrence of cancer or more specifically is it a benign or a malignant one. The project is not meant to be a replacement for diagnoses but it helps to show how we machine learning can be applied to various fields and can provide better guidance and assistance to people in the digital age.

REFERENCES

1. G. O. Young, “Synthetic structure of industrial plastics (Book style with paper title and editor),” in Plastics, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.
2. W.-K. Chen, Linear Networks and Systems (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.
3. H. Poor, An Introduction to Signal Detection and Estimation. New York: Springer-Verlag, 1985, ch. 4.
4. B. Smith, “An approach to graphs of linear forms (Unpublished work style),” unpublished.
5. E. H. Miller, “A note on reflector arrays (Periodical style—Accepted for publication),” IEEE Trans. Antennas Propagat., to be published.
6. J. Wang, “Fundamentals of erbium-doped fiber amplifiers arrays (Periodical style—Submitted for publication),” IEEE J. Quantum Electron., submitted for publication.
7. C. J. Kaufman, Rocky Mountain Research Lab., Boulder, CO, private communication, May 1995.
8. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interfaces(Translation Journals style),” IEEE Transl. J. Magn.Jpn., vol. 2, Aug. 1987, pp. 740–741 [Dig. 9th Annu. Conf. Magnetics Japan, 1982, p. 301].
9. M. Young, The Techincal Writers Handbook. Mill Valley, CA: University Science, 1989.
10. (Basic Book/Monograph Online Sources) J. K. Author. (year, month, day). Title (edition) [Type of medium]. Volume(issue). Available: http://www.(URL)
11. J. Jones. (1991, May 10). Networks (2nd ed.) [Online]. Available: http://www.atm.com
12. (Journal Online Sources style) K. Author. (year, month). Title. Journal [Type of medium]. Volume(issue), paging if given. Available: http://www.(URL)
13. R. J. Vidmar. (1992, August). On the use of atmospheric plasmas as electromagnetic reflectors. IEEE Trans. Plasma Sci. [Online]. 21(3). pp. 876—880.
14. S.Rajaprakash , R.Ponnusamy “Location Prediction System for Banking Loans UsingBelief Network Approach” International Journal of Modern Engineering Research, Pp-15-18.2013

AUTHORS PROFILE

Dr.S.Rajaprakashis M.sc, M.Phil M.E Ph.D. currently working as Associate professor of CSE in Aarupadai Veedu Institute of Technology an ambit institution of Vinayaka Missions Research Foundation (Deemed to be University), Tamil Nadu, India. He has 17 years of experience in academics, research, and development activities. Published 20 research papers in referred Journals and Conferences. His area of Interest Artificial Intelligence, Computational Intelligence, Discrete Mathematics and Automata theory. .Received grants from Tamil Nadu State Council for Science and Technology .He has peer Reviewed Manuscripts in reputed international Journals and Conferences.

He is a member in following professional societies: CSI and ISTE and Ramajunam Mathematicsl Society.

Mr.S.Muthuselvan MCA ME (PH.D) currently working as Assistant professor Aarupadai Veedu Institute of Technology an ambit institution of Vinayaka Missions Research Foundation (Deemed to be University), Tamil Nadu, India published more than 13 national and international journal and conference and organizing committee for 4 international conference, 2 national conference and 11 years of teaching experience with 4 years of research experience. He is a member in following professional societies: CSI and ISTE.

K.Karthik ME (PH.D) currently working as Assistant professor Aarupadai Veedu Institute of Technology an ambit institution of Vinayaka Missions Research Foundation (Deemed to be University), Tamil Nadu, India published more than 7 national and international journal and conference and organizing committee for 4 international conference ,2 national conference and 15 years of teaching experience with 4 years of research experience. He is a member in following professional societies: CSI and ISTE.

Vikrant Pradhan IV year CSE Aarupadai Veedu Institute of Technology an ambit institution of Vinayaka Missions Research Foundation (Deemed to be University), Tamil Nadu, India.

Abhay kumar A IV year CSE Aarupadai Veedu Institute of Technology an ambit institution of Vinayaka Missions Research Foundation (Deemed to be University), Tamil Nadu, India