

An Intelligence System for Medicine Recommendation: Review

Neeraj varshney

Department of Computer Engineering
and Applications GLA University
Mathura
neeraj.varshney@gla.ac.in,

Samarth Ahuja

Department of Computer Engineering
and Applications GLA University
Mathura
samarth.ahuja_cs17@gla.ac.in

Kanishk

Department of Computer Engineering
and Applications GLA University
Mathura
kanishk.gla_cs17@gla.ac.in

Abstract- A disease detection system may suggest care with selecting an exact disease to the patient. The technologies available today can assist us in developing such referral systems that can give us the right decisions. It is designed according to different algorithms. It is therefore important to understand the artistic developments of such systems, their benefits as areas that require more research. In this synthesis paper, a literature review of disease recommendation systems is conducted and described according to various characteristics.

Keywords: Machine learning, medicine recommendation, disease prediction, patient monitoring

I. INTRODUCTION

Nowadays, Technology increases rapidly and we all know that almost all the time we spend on gadgets. These gadgets changed our life a lot in a family from grandparents to children. due to which people become less caring about their health. Due to the new technology trend, people adopt new diseases or we can say that they are spoiling their health because of excessive use of gadgets. With the increasing trends of technology, People become so busy that they have no time to set the meeting with the doctor in a month or to take the treatment from the doctor with this busy life. That's why there is a need to adopt the new facility of recommender with the increasing technology [10]. A recommendation system is a system that recommends the disease based on the problems that the patient have or based on the patent review.

II. ABOUT RECOMMENDATION SYSTEM

A recommender system or we can say a filtering system is a type of system which suggests or predicts the most related features according to your class or type to which users are more concerned.

Three types of recommendation systems are as follows:

- (a) Collaborative Filtering.
- (b) Content-Based Filtering.
- (c) Hybrid Recommendation Systems.

A. Collaborative Filtering

Collaborative filtering uses the system's interactions and data from other users to filter information. It's focused on the

premise that people who agreed on a certain item's assessment would definitely agree again in the future.

The idea is simple: when we're looking for a new movie to watch, we always turn to our friends for suggestions. Naturally, we put more faith in suggestions from friends who have similar tastes to us.

The so-called similarity index-based methodology is used by the majority of collaborative filtering systems. A number of users are chosen based on their similarities to the active user in the neighbourhood-based approach. Calculating a weighted average of the ratings of the selected users is used to infer the successful user.

The relationship between users and objects is the subject of collaborative-filtering systems. The similarity of two things is measured by the similarity of their scores by users who have rated both of them.

B. Content-Based Filtering

Based on the user's previous activities or explicit reviews, content-based filtering uses item features to suggest other products that are close to what they want.

Let's hand-engineer some features for the Google Play store to illustrate content-based filtering. A feature matrix is shown below, with each row representing an app and each column representing a feature. Categories (such as Education, Casual, and Health), the app's publisher, and a variety of other features may be included. To make it easier, let's pretend this function matrix is binary: a non-zero value indicates that the app has that feature.

In the same feature space, you also represent the consumer. Any user-related functionality can be provided directly by the user. For example, in their profile, a user selects "Entertainment apps." Other features can be tacit, depending on the applications they have installed previously. The consumer, for example, installed a Science R Us-published app.

The model should make suggestions that are applicable to this consumer. To do so, first choose a similarity metric (for example, dot product). The framework must then be configured to score each candidate object based on this similarity metric. The model did not use any details about other users, so the suggestions are exclusive to this individual.

C. Hybrid Recommendation Systems

Recommender systems are automated tools that use a variety of techniques to create and provide recommendations for products and other individuals to users. Hybrid recommender systems incorporate two or more recommendation methods in a variety of ways to take advantage of their synergistic benefits. The current state of the art in hybrid recommender systems is presented in this systematic literature review.

It is the first comprehensive study devoted solely to hybrid recommenders. We discuss the most important issues that have been considered, as well as the data mining and suggestion methods that have been used to solve them. We also look at the hybridization groups that each hybrid recommender falls under, as well as the application domains, the assessment process, and possible research directions. According to our results, the majority of studies weight collaborative filtering with another strategy [9].

In addition, cold-start and data sparsity are the two traditional and top problems discussed in 23 and 22 reports, respectively, although most of the writers continue to use movies and movie datasets. Providing more credible and user-oriented assessments remains a common challenge because most studies are assessed through comparisons with similar approaches using accuracy metrics. Newer problems, such as reacting to changes in user background, changing user preferences, and providing cross-domain recommendations, have also been established [8]. Hybrid recommenders, as a hot subject, provide a strong foundation for responding appropriately by exploring newer opportunities such as contextualizing recommendations, using parallel hybrid algorithms, and processing larger datasets, etc.

Recommendation systems collect data and auto analyze this data to generate customized recommendations for users. These systems rely on explicit data such as inputs provided by the user.

The performance of the diagnosis and recommendation for treatment of the disease is at the center of this project. The efficiency of this project is disease recommendations and doctors contact for further concern, when the user entered their symptoms, the user will get exact and timely treatment of the disease. Patients would be highly obliged if they get a useful and appropriate diagnosis and doctor's contact which will save their lot of time and they will get initial knowledge about their diseases [7].

To get more accurate disease symptoms, the dataset must be good and more data should be used to train our model. Highly rich knowledge of experienced doctors is also very useful to predict accurate diseases. It is based on some core modules like: -

- (a) TKinter: -the GUI interface of python
- (b) Numpy and Pandas: - to perform mathematical computation and to read CSV files.
- (c) The machine learning decision tree algorithm is used and introduced to classify illness symptoms based on the given inspection reports.

III. LITERATURE SURVEY

Physical well-being is among the most commonly concerned issue on the internet. Large number of people are taking care of their health and health related issues. However, many people are losing their lives due to health hazards. Due to increase in the number of people visiting the health care centers, the health care centers have huge volume of data about patients and their respective health specifications. Thus, it is a necessity for reliable way for medical experts to use the data in a well-planned and logical manner. A case could be the permission to collect data from existing database on a specific problem at the point of supervision when it is necessary. Furthermore, there are more treatments, tests, drugs recommendations accessible for staff's day to day use. Hence, it becomes very challenging for them to figure out the cure to be given to a patient based on the tests, indications, past reports.

These type of recommender systems which are based on decision tree algorithms of machine learning and by using tkinter as its GUI suggests related diseases by symptoms from which patients are suffering and makes easier for doctors to give treatment and diagnosis advice according to patient history which is similar to other patients having same or related symptoms and that makes very less chance for doctors and paramedical staff to make any mistakes in giving treatment and it also give more broader way to think about patient disease.

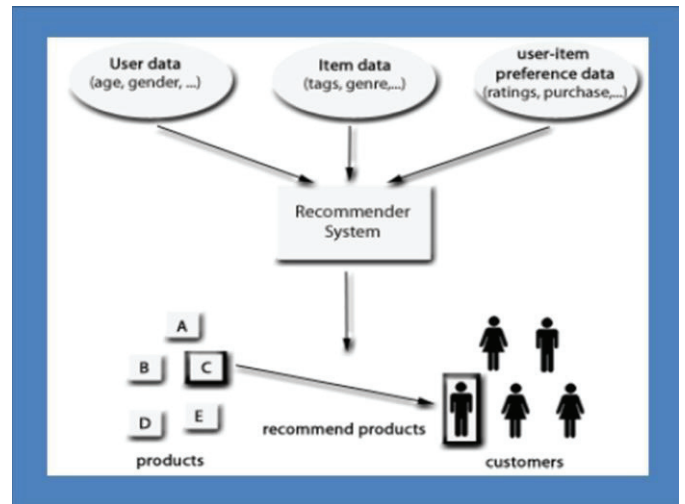


Fig. 1. Recommendation system

Deepbendu et al. [1] proposed medicine recommendation system using Hadoop and python. It is noticed that medical Science and digital world have created that can store huge amount of data. Everyone noticed this trend. Now, we can store huge amount of data in very cheap amount and access it very easily. We have also collected multiple knowledge for analyses. The goal of this type of invention to provide benefit to companies all over the world and to researchers and the world, it is increasingly important associated with the capacity for timely analysis of giant amount of knowledge with totally different areas. They used the easy way to fetch the data from the user sources like email, social media and the website that predict most probably disease.

The Methodology is in four parts.

- 1) Fetching – A huge amount of data has been collected from lots of people in integrated format in the form of Name and Symptoms which has been gathered from email, Social media and websites. By the help of Python, a huge amount of data has been fetched in words.
- 2) Refining – In this process they refine the data by storing it in HDFS by the help of Hadoop and python. After that they have arranged the symptoms words by the help of MapReduce technique.
- 3) Analyzing – In this step they search the outcome other than email and social media like Lab research database, Health Records.
- 4) Visualization – In last part for visualization they used Hadoop and python for linked the database by which user got the option of medication and also the surgery is required in patient's disease.

With the use of Hadoop and python they have developed an easy way to carry out people health issues. With the help of this project they can predict or confirm the disease just by symptoms. By this project patients get advice like medicines and diet chart. It is also the new way to connect two distinct fields i.e., computer and medicines.

Subasish et al [2] Created Smart Health Care Prediction through data mining. The major purpose of this project is to give predictions to user's three different health issues like liver, kidney and heart are in a healthy condition or not by comparing it with the dataset. It uses existing data from different databases to convert it into completely new results. The parameters on which this project is built are clusters, predictions, trajectory analysis and predictive analysis. The feature selection method helps to enhance the accuracy of the algorithm through the reduction of dimensional characteristics, health sectors have a lot of data performing data mining is quite easy and useful. The use of technology helps us to perform tasks automatically which means a reduction of manual work and easier extraction of data and help us to get better knowledge and early discovery of diseases. By using the concept of fitness number, the proposed method is presented to the end-user. Converting the data set into binary form particularly for the liver kidney and heart, through this process large data set is converted into a number of parts. Now, by performing the intersection computation of each part of the dataset now only select the minimum support fitness value system. This model is constructed using the KNN algorithm.

With the use of data mining techniques to make specific decisions, which will enhance the success rate of health care organizations and be beneficial for the health of the patients. This study allows us to apply methods of data exploration, function selection and classification algorithm to predict Chronic kidney disease, Heart disease, and Liver disease. It can be thoroughly analyzed and predicted through data mining and feature selection methods. In the future development of information technology, this business of data mining will realize its potential of hidden knowledge in medical data field. Anjali et al [3] proposed a system to eliminate the visits of patients to

hospitals and doctors for some common diseases such as cold, nausea, stomach pain, and many more. This application allows users to enter their symptoms and the system will provide users top 5 diseases that patients might suffer from. Furthermore, users can avail themselves of three different treatments namely ayurvedic, homeopathic, and natural home remedies according to their symptoms, this application is useful for those diseases which do not require immediate medical attention and can be treated at home, this application is not very secure but it is very useful as it does not waste time on searching on websites and to go and visit doctors. For gathering data, they have visited few homeopathic and Ayurvedic centers to gather knowledge about treatment and medicine prescriptions and for home remedies, they have consulted elders and also from medical journals recommended by health care workers and also from the internet to find the most effective remedies. By using the Mysql database to store information in the form of tables and also joined them according to requirements. They have used Python programming language to develop the system which is used to interact and manage database, Tkinter (Grayson 2000) is used as GUI for the user to interact with the system. The public view is shifting towards natural treatment due to the high cost of treatments, side effects of medicines, and only temporary relief from the disease. In the survey, the results are in favor of ayurvedic, homeopathic treatments, and home remedies for common diseases like common flu, etc, 75.6% of people do not trust allopathic treatment for common diseases, the output is of top 5 possible diseases according to the symptoms which helps them to recognize health condition, after that remedies of different treatments is provided so that user can take treatment accordingly. Nikhil et al [4] proposed 4 machine learning algorithms to reduce wrong predictions and provide accurate results to the users. It is very useful in those cases where two or more diseases have the same symptoms and cause difficulty for doctors to understand the disease and to avoid faulty decisions sometimes which causes the patient to serious condition. By using a single algorithm to predict disease we cannot get accurate results, So in this project, they have used 4 major machine learning algorithms for classification i.e. namely: -

1. Naive Bayes-It is based on Bayes rule of probability, useful when to know the probability of event A when the probability of event B is given, in other words, a feature of one event does not depend on another feature.
2. Random Forest- Collection of a various individual decision tree that acts together is called as random forest, each decision tree gives its result of class and that class which has a maximum number of output becomes the result of random forest, due to property of low co-relation there are more chances to get accurate results. It uses Bootstrapping techniques to construct trees.
3. Extra Tree- Extremely Randomized Tree, is an advanced version of random forest, An extra classifier makes use of the whole dataset for constructing trees.
4. SVM- Support vector machine, principle working of SVM is to find the best hyperplane which can

differentiate between different classes, having maximum margin is the best hyperplane to opt.

They have used the Tkinter library of python as GUI for interacting with the system.

They have tried different algorithms and the accuracy obtained from respected algorithms are as given: -

Naive Bayes: - 0.92

Random forest: -0.913

Extra tree: -0.936

SVM: -0.934

So, by combining these they get an accuracy of 92%. By using different algorithms this project is quite useful in avoiding faulty predictions.

And it gives a new way for researchers to think of combining different and major algorithms to give the model better accuracy and predict exact results.

Min et al [5] proposed a convolution neural network (CNN) to automatically extract text characteristics from unstructured data, yielding better results. They combine structured and unstructured data to forecast disease risk, gathering medical records from hospitals and constructing missing data using a latent factor model, and consulting doctors and medical experts to obtain useful features for structured data. A CNN algorithm is used to process unstructured text data. Finally, for structured and unstructured data, a CNN-based multimodal disease risk prediction (CNN-MDRP) is developed.

1. Data: - They used real-life patient data stored in the hospital data center; the data generated by the hospital included electronic health analysis, medical image data, and gene data, among other things. It comprises both structured and unstructured data. Structured data includes laboratory reports and basic patient records, while unstructured data includes the patient's explanation of the condition, physicians' conversations, and interrogation and diagnosis.
2. Imputation: - Introduction of data imputation for CNN based multimodal disease risk prediction algorithm, In hospitals, there is a large number of missing data that may be due to human errors or other such ways, so they have filled this missing data before data imputation, by modifying and deleting incomplete data so to improve data quality. After that using data integration for data processing, for imputation they use latent factor model which explains observing variables as a latent variable.

They designed (CNN MDRP) which is based on (CNN UDRP) for both structured and unstructured data structures. For structured data, they use 79 features in S-data and 100 features in T-data. This process is divided into two parts: -

- a. Training word embedding: - Pure corpus, collected from all patients in the hospital, is required for word training. We set these data as corpus sets after cleaning

them. The word2vec tool's n-skip gram algorithm trains the word vector using the ICTACLAS word segmentation tool.

- b. Training parameters of CNN-MDRP:- The risk evaluation of whether the patient suffers from chronic disease is reached by using the stochastic gradient approach for parameter training.

They proposed a new Convolutional neural network (CNN-MDRP) that uses both structured and unstructured data from hospitals, which is a very novel work in the field of big data that uses both forms of data in projects. Their prediction accuracy is 94.8 percent, and their prediction speed is faster than that of unimodal risk prediction (CNN UDRP).

Akash et al [6] proposed a Combination of structure and unstructured data in healthcare field. By using latent factor model for revised construction of unstructured data and for structured data with the help of paramedical staff, doctors and other hospital experts determined major chronic diseases with efficiency.

For unstructured and structured data, they have used k-mean algorithm for selection of features for disease automatically, k-mean is also working very efficiently with unstructured data.

None of the previous system have combined structured and unstructured data in this field of disease detection or big data field. Although much higher level of accurate data is obtained with 95% accuracy by combining structured and unstructured features. As a result, we have use Decision Tree classifier for predict the disease and also filter the similar symptoms which may cause due to that disease.

IV RESULT ANALYSIS

Paper	Method	Accuracy
Deepbendu et al. [1]	Using Big data and python	Malaria detection – 50%
Deepbendu et al. [1]	Using Big data and python	MELANOMA-70%
Deepbendu et al. [1]	Using Big data and python	TYPE-2 DIABETES – 80%
Deepbendu et al. [1]	Using Big data and python	ARTHRITIS -85%
Nikhil et al [4]	Random Forest	91.3%
Nikhil et al [4]	Extra Tree	93.6%
Nikhil et al [4]	Naive Bayes	92.7%
Nikhil et al [4]	SVM	93.4%
Min et al [5]	CNN Model	94.8%
Boshra et al. [8]	j48 decision tree	83.73%
Boshra et al. [8]	KNN	82.77%

IV. PROPOSED WORK

Disease detection has been accomplished by adopting Machine Learning techniques. We have used a dataset applied training of a model (train/test) on this dataset. After training the disease detector, we will load the disease detector performing disease detection, and then classifying each type of disease.

A. Predicting the disease using symptoms

There is also doctor's dataset where there is information about best doctors and their website. From this dataset we recommend the best doctors available for consultation. We also have training and test dataset where the first row of this contains all the symptoms which are used in this project, whenever there appears digit '1' then it shows that this symptom is included in the disease and digit '0' represents that this symptom is excluded from the disease.

B. Python language used for coding

We have used Decision Tree to train a classifier for predicting the disease from the symptoms. To accomplish this task, we had made small adjustments by making the GUI with the tkinter, a highly efficient GUI applied to embedded devices. Set Decision Tree imports allows for predicting the disease form the symptoms and also predicting the more symptoms for that disease.

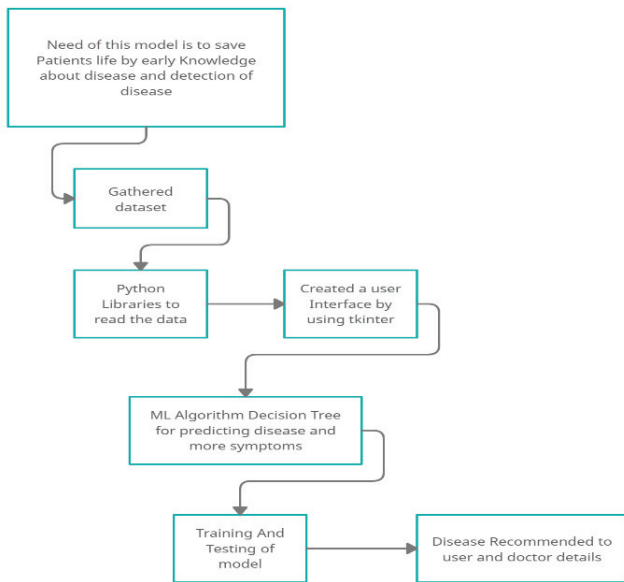


Fig. 2. Proposed Model

V. CONCLUSION

In conclusion, by utilizing supervised learning algorithms, Decision Tree Classifier in Machine Learning allowed us to developed two recommender systems that can give a medical recommendation with an excellent efficiency and accuracy based on diagnosis and symptoms. That is why Recommendation System is very useful in term of disease prediction and on the behalf of disease predict by the doctor.

REFERENCES

- [1] Das, Deepbendu, Rishabh Pandey, and Ankur Saxena. "Disease prediction using Hadoop with Python." Retrieved September 22 (2017): 2018.
- [2] Mohapatra, S., Patra, P.K., Mohanty, S. and Pati, B., 2018, December. Smart Health Care System using Data Mining. In 2018 International Conference on Information Technology (ICIT) (pp. 44-49). IEEE.
- [3] Jain, A., Manvitha Reddy, G., Kumar Tomar, Y. and Saxena, A., 2018. Evolution of Health Care Systems Based on Quantitative Investigation Using Python. International Journal of Information Systems & Management Science, 1(1).
- [4] Sharma, N., Singh, M. and Gupta, Y., DISEASE PREDICTOR: ALGORITHMS TO REDUCE THE FAULTY PREDICTIONS.
- [5] Chen, M., Hao, Y., Hwang, K., Wang, L. and Wang, L., 2017. Disease prediction by machine learning over big data from healthcare communities. Ieee Access, 5, pp.8869-8879.
- [6] Jamgade, A.C. and Zade, S.D., 2019. Disease prediction using machine learning. International Research Journal of Engineering and Technology, 6(5), pp.6937-6938.
- [7] Patel, S. and Patel, H., 2016. Survey of data mining techniques used in healthcare domain. International Journal of Information, 6(1/2), pp.53-60.
- [8] Bahrami, B. and Shirvani, M.H., 2015. Prediction and diagnosis of heart disease by data mining techniques. Journal of Multidisciplinary Engineering Science and Technology (JMEST), 2(2), pp.164-168.
- [9] Dahiwade, D., Patle, G. and Meshram, E., 2019, March. Designing disease prediction model using machine learning approach. In 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC) (pp. 1211-1215). IEEE..
- [10] Blockeel, H. and Struyf, J., 2002. Efficient algorithms for decision tree cross-validation. Journal of Machine Learning Research, 3(Dec), pp.621-650.