

Recommender System for Home Remedy

Sahil Shaikh

IT Department

St. Francis Institute of Technology
Mumbai, India

Sneha Mannekkad

IT Department

St. Francis Institute of Technology
Mumbai, India

Lohit Poojary

IT Department

St. Francis Institute of Technology
Mumbai, India

Shikha Shah

IT Department

St. Francis Institute of Technology Mumbai, India

Joanne Gomes

IT Department

St. Francis Institute of Technology Mumbai, India

Abstract—In today's world, people are falling sick very frequently due to various types of reasons, such as drinking contaminated water, having allergies, being exposed to pollution, and seeing frequent weather changes. The elderly can recuperate from mild illnesses at home, as well as it may be difficult to visit the clinics during the pandemic situation. The solution to this problem is home remedy. Existing system like DIETOS, suggests nutritious diets on a daily basis based on the questionnaires filled by the users to improve the quality of life. This research paper presents a home remedy recommending system. The system suggests fruits and herbs as a home remedy to the user symptoms using different Data Mining algorithms. Based on the accuracy of these algorithms the user can then consume recommended fruits and herbs. The result of this research paper shows that among the four algorithms Decision Tree, Random Forest, Naive Bayes, and kNearest Neighbor used, Naive Bayes gives the highest accuracy of recommendation.

I. INTRODUCTION

These days, people are very commonly getting sick due to various types of reasons such as contaminated food and water, allergies, pollution, and weather changes. In the case of mild illnesses, people can recover at home and it is better to visit the clinics less often, especially during the pandemic situations. As traditional medicine is still recognized as the preferred primary health care system in many communities, with over 60% of the world's population and about 80% in developing countries depending directly on medicinal plants for their medical purposes [1], they can be used to cure the mild illnesses. This is due to a number of reasons including affordability, accessibility and low cost.

The authors Agapito G., Calabrese B., et.al of research paper [2], proposed a dynamic real-time questionnaire system that builds the user's health profile. It provides individualized nutritional recommendations based on the said health profiles. These questionnaires are prepared by medical doctors and compiled by the users. Another system, proposed by authors Pruthvish Desai, Shrey Dey et.al. in research paper [3] is comprised of an automatic medicine vending machine that dispenses drugs based on a prescription from a doctor. There is also an online portal for users to check his/her prescriptions.

Paper [4] by authors T.Venkat Narayana Rao and Anjum Unnisa, takes the patient review as input and performs sentiment analysis thus suggesting the medicines to the user. In research paper [5] by authors Kazuyuki Shimada, Hidekatsu Takada, et.al. there is a drug recommendation system for patients, which helps the doctors to judge the status of patients and also allows patients to choose a first-line drug appropriately. Lastly, in research paper [6] proposed by the authors Santosh Cheruku, Samantha Deokinanan et.al., user inputs the area of the body where they are experiencing the symptom and then the system recommends the drugs.

Some of the restrictions observed in the literature papers are, the access to the medicines in system [4] is only given on getting the prescription from the doctors. This paper recommends medicines instead of the home remedies. Age of person and the demographic information during the training phase is not included in paper [6]. Thus, there are opportunities to include many novel concepts to include home remedies.

The proposed system in this research paper, takes the user health symptoms and recommends the fruits and herbs that are beneficial in curing the illness. Data mining algorithms used are Decision Trees, Random Forests, Naive Bayes, and k Nearest Neighbors. Accuracy level of the algorithms will let the user know how effective the recommendation is, and based on that, they can use the home remedies. Furthermore, users can also see the details about how one can consume the fruits or herbs recommended by the system to cure the ailment.

Section II presents related works and existing systems that are close to the proposed system. Section III provides information regarding the data mining prediction algorithms Decision Tree, Random Forest, Naive Bayes and kNearestNeighbor that are used in the proposed system. Section IV presents the detailed information about the working of the proposed system. The section V discusses the user interface of the website. The section VI presents the observations and results. The comparative analysis is included in the section VIII concludes the paper.

II. REVIEW OF RELATED WORK

This section explains the related work of existing similar recommendation systems. Authors Agapito G., Calabrese B., et.al proposed a system in research paper [2], called DIETOS (DIET Organizer System) which is a recommender system for the adaptive delivery of nutrition contents to improve the quality of life of both healthy people and individuals affected by chronic diet-related diseases. The system here is able to build a user's health profile, and provides individualized nutritional recommendation according to the health profile. This profile is created through the use of dynamic real-time questionnaires prepared by medical doctors and compiled by the users, whereas our proposed system allows the user to enter the symptoms via the webpage.

All Time Medicine and Health Device presented by the authors Pruthvish Desai, Shrey Dey et.al in research paper [3] comprises of an automatic medicine vending machine which dispenses drugs as per a doctor's prescription. The vending mechanism is controlled by the raspberry pi which is a single board computer and the second aspect of it is the online portal for a user to check his prescriptions, and for the doctor to generate an e-prescription. The device dispenses out the prescribed medicines by the doctor when the user credentials of patients are validated from the database. The online portal is built on two fronts - a webpage and an android application which are linked to the same database. The patient can view his details and prescriptions through the android application or webpage by logging in with appropriate credentials. The system in this literature paper requires a prescription by the doctors for the required medication, whereas our proposed system need not require prescription.

Authors T. Venkat Narayana Rao, Anjum Unnisa and Kotha Sreni designed a Medicine Recommendation System Based On Patient Reviews in research paper [4]. This paper proposes a medicine recommendation system, which takes the patient review data and performs sentiment analysis on it to find the best medicine for a disease by using N-Gram model. An N-Gram model is a type of probabilistic language model for predicting the next item in the sequence. It is basically a contiguous sequence of n items from a given sample of text or speech. In order to increase the accuracy, a LightGBM model is used to perform medication analysis. LightGBM is a fast, distributed, high performance gradient boosting framework based on decision tree algorithms, used for ranking and classification. The system makes use of two algorithms N-Gram model and the LightGBM model, whereas our proposed system uses algorithms such as Decision Tree, Random Forest, Naive Bayes, and k-Nearest Neighbor.

In research paper [5], authors Kazuyuki Shimada, Hidekatsu Takada, et.al proposed a Drug-Recommendation System for patients with infectious diseases. It is a decision support system that helps doctors select appropriate first-line drugs. The system classifies patients' abilities to protect themselves from infectious diseases as a risk level for infection. It is a

system that helps doctors to judge the status of patients and allows them to choose a first-line drug appropriately. The system performs the following steps:

- Step 1: Input a patient's risk factors.
- Step 2: Confirm a risk level calculated by the system
- Step 3: Select a bacterium. After this, the system lists potential drugs in order of priority.

The system recommends the drugs based on the risk factor of the patient, whereas our proposed system recommends the home remedies using classification.

Medical Recommender System proposed by authors San-tosh Cheruku, Samantha Deokinanan et.al in research paper

[6] is a system that can give recommendation with an excellent efficiency and accuracy based on diagnosis and health symptoms. It is developed utilizing the Naive Bayesian Classifier method where users input the area of the body they are experiencing a symptom, the kind of symptom, and the system would then recommend highly effective drugs that can treat such illness. The system uses only one algorithm, whereas our proposed system shows the results for four algorithms.

III. PREDICTION ALGORITHMS

Predictive modeling algorithm is a statistical technique using Machine Learning and Data Mining to predict and forecast likely future outcomes with the aid of historical and existing data. It works by analyzing current and historical data and projecting what it learns on a model generated to forecast likely outcomes.

A. Decision Tree Algorithm:

Decision Tree algorithm falls under the category of supervised learning. They can be used to solve both regression and classification problems. Fig. 1 illustrates a basic flow of Decision Tree algorithm. The goal of using this algorithm is to create a training model that can use to predict the class or value of the target variable by learning simple decision rules inferred from prior data (training data).

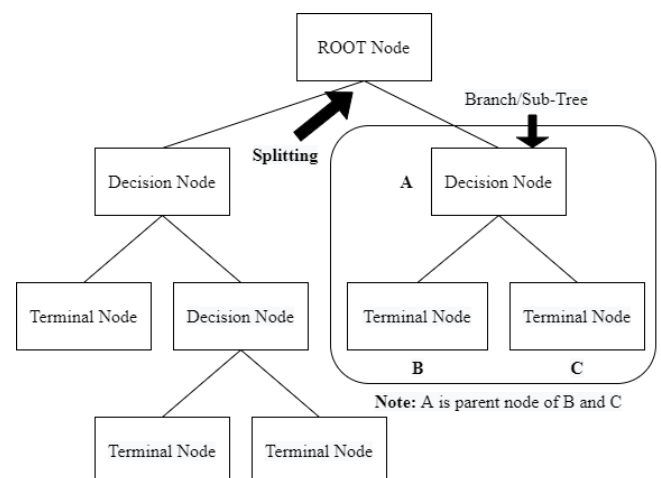


Fig. 1. Decision Tree Algorithm

Important Terminology related to Decision Trees from Fig. 1 are explained below:

- Root Node: It represents the entire population or sample and this further gets divided into two or more homogeneous sets.
- Splitting: It is a process of dividing a node into two or more sub-nodes.
- Decision Node: When a sub-node splits into further sub-nodes, then it is called the decision node.
- Leaf / Terminal Node: Nodes do not split is called Leaf or Terminal node.
- Pruning: When we remove sub-nodes of a decision node, this process is called pruning. You can say the opposite process of splitting.
- Branch / Sub-Tree: A subsection of the entire tree is called branch or sub-tree.
- Parent and Child Node: A node, which is divided into sub-nodes is called a parent node of sub-nodes where as sub-nodes are the child of a parent node.

B. Random Forest Algorithm:

Random Forest is a supervised learning algorithm which is used for both classification as well as regression. But however, it is mainly used for classification problems. As we know that a forest is made up of trees and more trees means more robust forest. Similarly, random forest algorithm creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting. It is an ensemble method which is better than a single Decision Tree because it reduces the over-fitting by averaging the result. Fig. 2 illustrates the working of Random Forest algorithm:

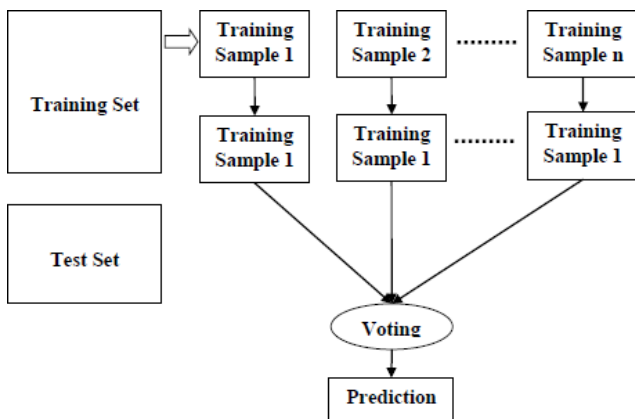


Fig. 2. Random Forest Algorithm

We can understand the working of Random Forest algorithm with the help of following steps:

- Step 1: First, start with the selection of random samples from a given dataset.
- Step 2: Next, this algorithm will construct a decision tree for every sample. Then it will get the prediction result from every decision tree.
- Step 3: In this step, voting will be performed for every predicted result.

Step 4: At last, select the most voted prediction result as the final prediction result.

C. Naive Bayes Algorithm:

Naive Bayes is a probabilistic machine learning algorithm based on the Bayes Theorem, used in a wide variety of classification tasks. Bayes' Theorem is a simple mathematical formula used for calculating conditional probabilities. Conditional probability is a measure of the probability of an event occurring given that another event has (by assumption, presumption, assertion, or evidence) occurred.

The formula is:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \tag{1}$$

Where in equation (1), A and B are events and $P(B) \neq 0$. $P(A)$ is the priori of A (the prior probability, i.e. Probability of event before evidence is seen). The evidence is an attribute value of an unknown instance (here, it is event B). $P(A|B)$ is a posteriori probability of B, i.e. probability of event after evidence is seen.

D. kNearest Neighbor Algorithm:

kNearest Neighbor is a simple algorithm that stores all the available cases and classifies the new data or case based on a similarity measure. It is mostly used to classify a data point based on how its neighbors are classified. KNN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suited category by using KNN algorithm. KNN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.

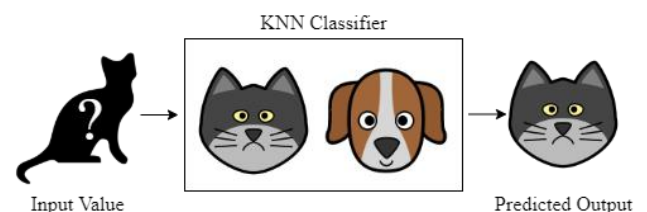


Fig. 3. kNearest Neighbor Algorithm

Example: Consider Fig. 3, we have an image of a creature that looks similar to cat and dog, but we want to know whether it is a cat or dog. So for this identification, we can use the KNN algorithm, as it works on a similarity measure. The KNN model will find the similar features of the new data set to the cats and dogs images and based on the most similar features it will put it in either cat or dog category.

IV. RECOMMENDER SYSTEM FOR HOME REMEDY

The proposed system is divided into three parts. Fig. 4 shows the overview of the basic flow of the proposed system. The user first selects exactly five health symptoms from a drop-down list. When the user submits their symptoms, the

system processes them using the four data mining algorithms. After that, the system recommends home remedies like fruits and herbs that can be consumed for treating the illness. Additionally, there's another module that shows nearby hospitals based on the user's current location.

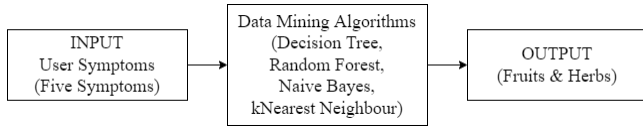


Fig. 4. Basic Working of System

Firstly, the user selects symptoms (exactly five) from the list displayed on the website, which are sent to the data mining algorithms. The train/test method is used on a manually created dataset to measure the accuracy of the model where 67% of the dataset is used for training, and 33% of the same is used for testing. The system then recommends home remedies such as fruits and herbs along with the accuracy of the algorithms used. A total of 45 symptoms are included in the system which includes: Diabetes, Diarrhea, Fever, Constipation, Colic, Cough, High Cholesterol, Parasite Infections, Toothache, Cold, Flu, Urine Odor, Ulcer, Sore Throat, Arthritis, Hair Loss, Digestion, Eczema, Insomnia, Blood Pressure, Anemia, Skin Damage, Weak Eyesight, Spasm, Gas, Skin Swelling, Blood Clotting, Asthma, Dehydration, Muscle Soreness, Nausea, Jaundice, Scurvy, Vomiting, Vertigo, Morning Sickness, Athletes Foot, Ringworm, Dental Plaque, Tooth Decay, Wounds, Itching, Over Weight, Anxiety, and Stress.

- **Decision Tree:** The algorithm here takes the user input, it classifies and arranges the corresponding data in the form of leaf nodes and internal nodes. It starts with all training instances associated with the root node. Use information gained to choose which attribute to label each node with. Recursively construct each subtree on the subset of training instances that would be classified down that path in the tree.
- **Random Forest:** Based on the user input, the algorithm here classifies and arranges the corresponding data into leaf nodes and internal nodes. The algorithm starts with all the training instances related to the root node. It uses the information gained to choose which attribute to label each node with. Finally, recursively constructs each subtree on the subset of training instances that would be classified down that path in the tree.
- **Naive Bayes:** This algorithm makes use of Bayes Theorem. It predicts membership probabilities for each class such as the probability that given record or data point belongs to a particular class. The class with the highest probability is then considered as the most likely class.
- **kNearest Neighbor:** KNN algorithm selects number of neighbors (let's say total k neighbors), and calculates the Euclidean distance of k number of neighbors. Then takes the k nearest neighbors as

per the calculated Euclidean distance. Among these k neighbors, it counts the number of the data points in each category and assigns the new data points to that category for which the number of the neighbor is maximum.

Based on the data mining algorithms the system displays the fruits and herbs according to the user health symptoms. The displayed result may include one or many (dependent on user input) of the fruits and herbs such as Apple, Guava, Papaya, Avocado, Coconut, Cranberry, Fig, Banana, Orange, Java Plum, Kiwi, Watermelon, Sweet Lime, Ginger, Garlic, Clove, Cinnamon, Turmeric, and Lemon Balm.

V. ONLINE WEB PORTAL FOR PROPOSED SYSTEM

The proposed system is developed using two core technologies HTML5 (Hypertext Markup Language) and CSS3 (Cascading Style Sheets) along with JavaScript. PHP scripting language is used to store the user details into the MySQL database server. Python language is used for the working of the algorithms. Finally, the connection between the frontend web development part and the backend algorithms is done using a CGI script via Wamp Server.

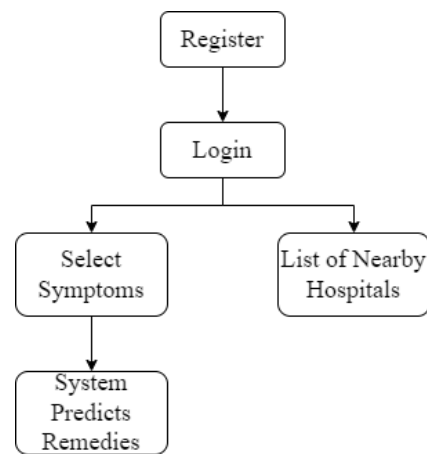


Fig. 5. Flowchart of Proposed System

Fig. 5 demonstrates the flowchart of the proposed system. The Register page consists of four input fields i.e. 'First Name', 'Last Name', 'Email', and 'Password' as shown in Fig. 6.

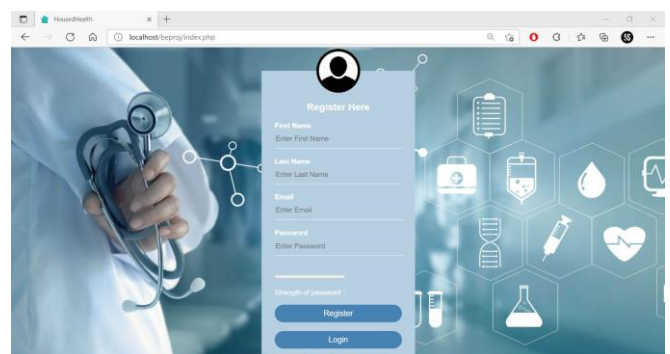


Fig. 6. Register Page

A 'Register' button is used to submit the above mentioned user data inputs into the database. A 'Login' button is also included in case the user has already registered on the website. Validation is done using JavaScript for the inputs, such as 'First Name' should consist of only alphabets and a minimum of three letters are required. The same validation is done for 'Last Name' input. A special regex (regular expression) is used for 'Email' pattern, thus only valid emails containing a domain will be allowed. For 'Password' validation, there should be at least six characters including an uppercase letter, a lowercase letter, a number, and a special character. The Loginpage consists of two input fields, 'Email' and 'Password'. The 'Login' button submits the email and password, which is then matched against the database. In the event of a matching record, the user is directed to the Homepage of the website. Otherwise, the access to the website is restricted. A 'Register' button is included in case the user has not yet registered.

The Homepage includes a search bar where the user can select the health symptoms from a drop-down menu. A total of 45 symptoms are included in the system. The user has to select five health symptoms from the list, which are then passed to the python file at the backend, where the algorithms from Sklearn libraries are used to recommend home remedies to the user. Fig. 7 shows the Homepage of the website.

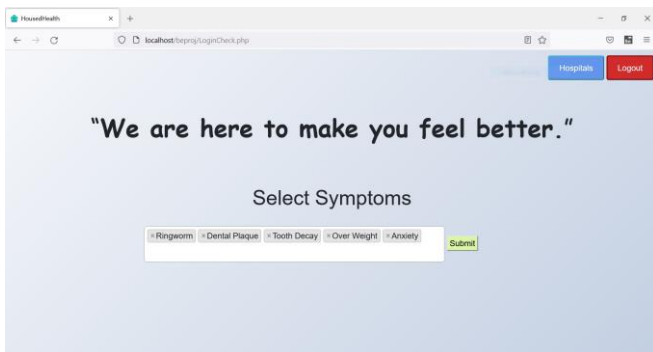


Fig. 7. Homepage

The predictions from each of the four algorithms are shown in Fig. 8. The recommendations are accompanied by their respective accuracy ratings and recommendation with highest accuracy is displayed as 'Strongly Recommended'.

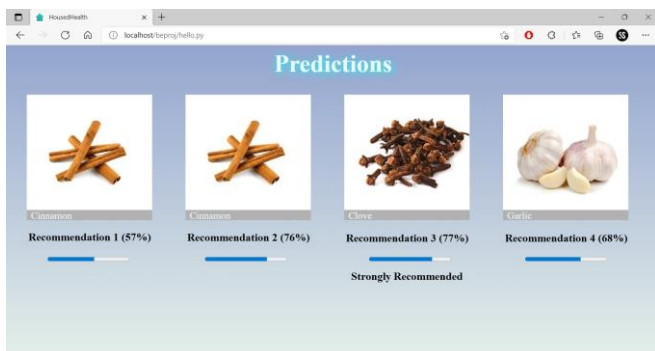


Fig. 8. Predictions Page

Additionally, when the user clicks on a particular fruit or herb that is recommended, more information about that particular home remedy is displayed in a new window. The user can also find instructions on how to use the product to cure his/her ailment. Fig. 9 displays the product details page.

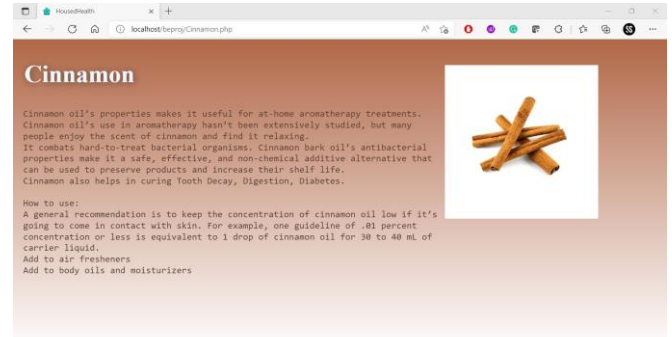


Fig. 9. Product Details

The proposed system also displays a list of nearby hospitals based on the user's real time location. This is done using Google Places API. Fig. 10 shows a list of nearby hospitals based on the location of the user.

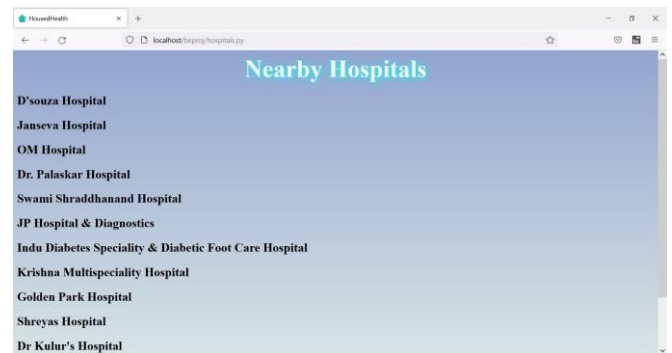


Fig. 10. List of Nearby Hospitals

VI. TEST CASES AND DISCUSSIONS

In this section, the comparison of different test cases and their results are discussed as well as the algorithms are compared for accuracy. The algorithm that recommends with the highest accuracy is displayed as 'Strongly Recommended'.

TABLE I TEST CASE 1

Symptom 1	Symptom 2	Symptom 3	Symptom 4	Symptom 5
Diarrhea	Fever	Colic	Cough	High Cholesterol

Table I includes the symptoms from Test Case 1. The result obtained from the above health symptoms shows that the system recommends Guava in Decision Tree & kNearest Neighbor algorithm, and Apple in Random Forest & Naive Bayes algorithm.

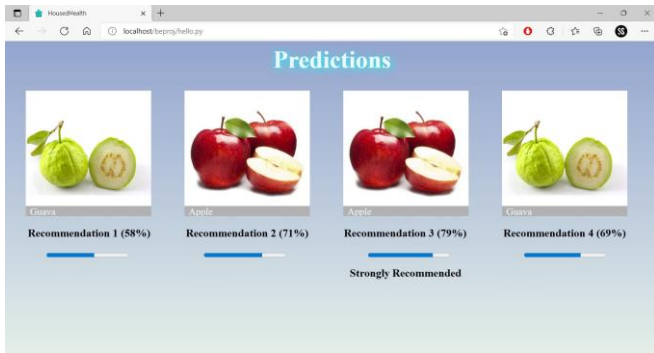


Fig. 11. Result of Test Case 1

From Fig. 11 it can be concluded that The accuracy of Naive Bayes comes out to be 79% which is the highest among all the algorithms. The result obtained from Naive Bayes algorithm recommends Apple.

TABLE II TEST CASE 2

Symptom 1	Symptom 2	Symptom 3	Symptom 4	Symptom 5
Ringworm	Dental Plaque	Tooth Decay	Over Weight	Anxiety

Table II indicates the symptoms from Test Case 2. The result obtained from the above health symptoms shows that the system recommends Cinnamon in Decision Tree and Random Forest Algorithm, Clove in Naive Bayes and Garlic in kNearestNeighbor Algorithm as shown in Fig. 12.

Also, Naive Bayes offers you two hyperparameters to tune for smoothing: alpha and beta. A hyperparameter is a prior parameter that are tuned on the training set

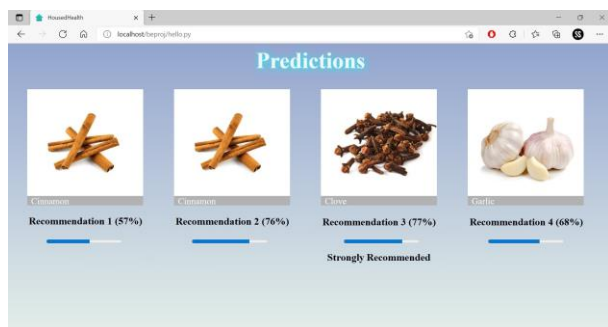


Fig. 12. Result of Test Case 2

From Fig. 12 it can be concluded that Naive Bayes yet again has the highest accuracy of 77% among all the algorithms. The result obtained from Naive Bayes Algorithm recommends Clove.

VII. COMPARATIVE ANALYSIS

Naive Bayes is a linear classifier while KNN is not; It tends to be faster when applied to big data. In comparison, KNN is usually slower for large amounts of data, because of the calculations required for each new step in the process. If speed is important, choose Naive Bayes over KNN. In general, Naive Bayes is highly accurate when applied to big data.

to optimize it. In comparison, KNN only has one option for tuning: the 'k', or number of neighbors. This method is not affected by the curse of dimensionality and large feature sets, while KNN has problems with both. Simple decision trees tend to over fit the training data more so that other techniques which means you generally have to do tree pruning and tune the pruning procedures. For Naive Bayes, the math is complex, but the result is a process that's highly accurate and fast—especially when you're dealing with Big Data. In our system, we have observed and analysed that amongst the four algorithms, Naive Bayes shows the highest accuracy when compared to the other three algorithms. It shows the accuracy in the range 75% to 85% for most of the test cases.

VIII. CONCLUSIONS

The project helps users in curing the disease by giving the home remedies like fruits and herbs that the user should consume in order to get rid of the illness. The main purpose of this project is to help the user to easily search for herbs and fruits that will be good for the health of the user depending on any health issue or disease that he/she is suffering from. This system helps the user to reduce its searching time to a great extent by allowing the user to enter its health problem and search accordingly. The admin can add fruits and herbs to the system and its information. This system also allows the user to view the selected fruit or the herb's description which describes the details about how one can consume the fruit or the herb. The system also includes a module where the user can get a list of nearby hospitals based on the user's real time location. Thus, this system helps to cure the user's disease to a great extent.

REFERENCES

- [1] Susana Oteng Mintah, Tonny Asafo-Agyei, Mary-Ann Archer, Peter Atta-Adjei Junior, Daniel Boamah, Doris Kumadoh, Alfred Appiah, Augustine Ocloo, Yaw Duah Boakye, and Christian Agyare, "Medicinal Plants for Treatment of Prevalent Diseases", 2019 IntechOpen, DOI: 10.5772/intechopen.82049.
- [2] G. Agapito et al., "DIETOS: A recommender system for adaptive diet monitoring and personalized food suggestion," 2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), New York, NY, USA, 2016, pp. 1-8, doi: 10.1109/WiMOB.2016.7763190.
- [3] P. Desai, B. Pattnaik, S. Dey, T. S. Aditya, K. Rajaraman and M. Aarthy, "All Time Medicine and Health Device," 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), Coimbatore, India, 2019, pp. 5-9, doi: 10.1109/ICACCS.2019.8728306.
- [4] T. Venkat Narayana Rao, Anjum Unnisa and Kotha Sreni, "Medicine Recommendation System Based On Patient Reviews," 2020 International Journal of Scientific Technology Research Volume 9, Issue 02, ISSN 2277-8616
- [5] Kazuyuki Shimada, Hidekatsu Takada, Satoshi Mitsuyama, Hideyuki Ban, Hitoshi Matsuo, Hitoshi Otake, Hiroyuki Kunishima, MD, PhD, Keiji Kanemitsu, MD, PhD, and Mitsuo Kaku, MD, PhD, "Drug- Recommendation System for Patients with Infectious Diseases," 2005 AMIA Annu Symp Proc.
- [6] Santosh Cheruku, Samantha Deokinanan, Rajwant Mishra and Priya Shaji, "Medical Recommender System," 2019 RPubS, DATA 607 Final Project