iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** V **Month of publication:** May 2021

DOI: https://doi.org/10.22214/ijraset.2021.34748

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue V May 2021- Available at www.ijraset.com

Android Application for Automation of Health Consultancy

Ch.Vijayalakshmi[1], G. Kavita[2], G. Sravanthi[3]
[1, 2]*Computer Science Department, Chaitanya Bharathi Institute of Tenchnology(Autonomous)-Hyderabad.*
[3]*Computer Science Department, GITAM (Deemed to be University)-Hyderabad*

Abstract: ***The main aim of this work is to predict the drugs based on symptoms and other details provided by user using various machine-learning techniques, technologies and drug- development and clinical-trial data from various web sites. The outcome this work is an interactive application where the user can utilize the features provided. As part of this work, data is extracted first from various web sites using Data Scraping Technique. Then predicts the disease and drugs based on the details available in the data set. With the help of Support Vector Machine Algorithm the disease and drugs are predicted. We suggest these to the users through application. The main aim is to provide the user with most suitable disease and recommend best drug, through user-friendly application. In this work, first we find the data from drugs and clinical data present in online web pages using web scraping and developed Android application to communication with the users. Based on the symptoms provided by user, the system predicts the disease they have and the medication.***
Keywords: Web scraping, classification, Feature Extraction, Drug and Disease Prediction.

I. INTRODUCTION

As India started modernizing at a steady pace, the country has been slowly catching up with the rest of the world. One of the fruits of advancement is the surge in the use of internet in India. But, most of the internet usage is being spent on entertainment. Consequently, this resulted into a situation where the users are unaware of facilities that the internet can provide, such as medical assistance. Since there is poverty in the country mostly in the rural areas, technology needs to be introduced to counter the issue where doctors are not properly qualified or not entirely present in some areas. Lack of medical personnel can lead the general public taking the matter into their own hands, like taking medications that are recommended by person who is not aware of the dosage levels and the possible side effects. This can be avoided with the online medical assistance.

Developing an application that provides medical support to the users by mentioning their disease and suggesting medicines required to cure the disease. The predictions are based on the personal details, symptoms and extent of the symptoms. It provides the suggestions from drug and clinical data present in the online web pages. Drugs can be detrimental if they are used without prior knowledge of dosage levels. The dosage also varies with regarding to age, sex and various other factors. This issue could be countered if the individual is taught about the possibilities of the symptom which can motivate the patient to take the extra mile and get proper help immediately. This study provides the user with enough support regarding their diseases and recommends drugs and aid in case of inevitable self-medication. This is done through a user-friendly application that can understand the issues of the user accurately.

II. LITERATURE SURVEY AND METHODOLOGIES

Web Scraping [2] is important technique used for extracting unstructured data from the websites and transforming that data into structured. Web scraping is a form of data mining. The basic and important aim of the web scraping process is to mine information from a different and unstructured website and transform it into an comprehensible structure like spreadsheets, database or a comma-separated values (CSV) file. Data like item pricing, stock pricing, different reports, market pricing and product details, can be gathered through web scraping. Extract targeted information from websites contributions to take effective decisions in business process.

BeautifulSoup is a Python library used for parsing documents (i.e. mostly HTML or XML files). Using Requests to obtain the HTML of a page and then parsing whichever information you are looking for with BeautifulSoup from the raw HTML is the quasi-standard web scraping "stack" commonly used by Python programmers for easy tasks.

Selenium Python bindings provides a simple API to write functional/acceptance tests using Selenium WebDriver. Through Selenium Python API you can access all functionalities of Selenium WebDriver in an intuitive way.Selenium Python bindings provide a convenient API to access Selenium WebDrivers like Firefox, Ie, Chrome, Remote etc. The current supported Python versions are 2.7, 3.5 and above.

©IJRASET: All Rights are Reserved

HTTrack is a free and open-source Web crawler and offline browser. HTTrack allows users to download World Wide Web sites from the Internet to a local computer. By default, HTTrack arranges the downloaded site by the original site's relative link-structure. The downloaded (or "mirrored") website can be browsed by opening a page of the site in a browser. HTTrack can also update an existing mirrored site and resume interrupted downloads. HTTrack is configurable by options and by filters (include/exclude), and has an integrated help system. There is a basic command line version and two GUI versions (WinHTTrack and WebHTTrack); the former can be part of scripts and cron jobs.

Another Simple and user-oriented tool for data scraping is Web Content Extractor (WCE) which is developed by Newprosoft. It has good wizard that guide user to setup scraper. User can scrape data from website with few clicks and WCE is self-intelligent for putting data into different formats like Excel, text, HTML formats, Microsoft Access database, Structured Query Language(SQL) Script File, MySQL Script File, Extensible Markup Language (XML) file, HTTP submit form and Open Database Connectivity (ODBC) Data source.

It is a classification technique based on Bayes'[3] Theorem with 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.

In the fields of computational linguistics and probability, an *n*-gram [5] is a contiguous sequence of *n* items from a given sample of text or speech. The items can be phonemes, syllables, letters, words or base pairs according to the application. The *n*-grams typically are collected from a text or speech corpus.

Fig. 1 describes the block diagram of the proposed system. This android application uses java language in software known as android studio. The application has two login pages one for user login(user interface for patient) and admin login(login for the admin).Here the machine learning code which is written in python would be uploaded to the cloud and the symptoms entered by the user in the application would be sent to the cloud in json format and the output will be retrieved in the same format and the most probable disease will be predicted. Also full page information about the particular disease will be presented. The next activity page will contain the top 5 average rating drugs which are relevant to the disease. A feedback form will also be provided, so that if the patient checks with a real doctor then how accurate was the disease prediction. This would help improve the application in future.

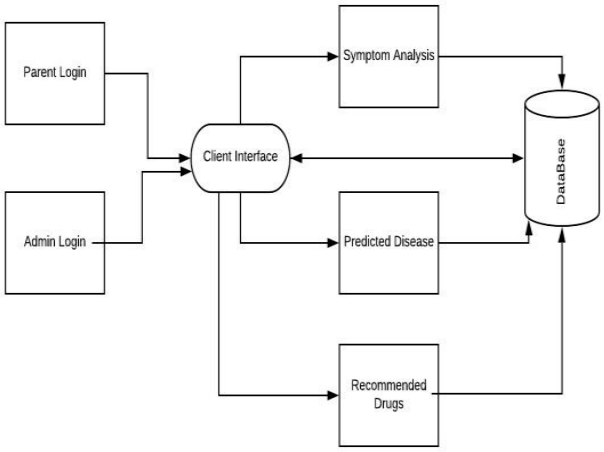


Fig.1 Data Flow Architecture

III.IMPLEMENTATION OF PROPOSED SYSTEM

As part of the implementation of this work, the following phases involved in both for Web scraping of Diseases and for Web scraping of Drugs.

A. Initializing the Web Sites

We consider the disease related web site which provide disease related data. Then we initialize such web sites in the python code. In the considered the web site we have the disease names and their related hyperlinks.

B. Generating Links

Form the sites; we extract the links related to the disease names. We do that using various tools such as Beautiful soap and pandas so on. Here we present the web sites where the disease name only starting with a. Finally, we present those websites.

©IJRASET: All Rights are Reserved

C. Organizing the Drugs

Here we extract all the disease names and produce the link which has the information related to that particular disease. Here we extract the links which have drugs related information using the html.parser. Which ask the web sites to perform action and extract certain information from the web sites. We use this in beautiful soap and create a model out of it. Finally, we create a set which have the disease name and their description link in a sequential form in logographical order.

D. Extraction of Descriptions

Here we extract the descriptions form the form the available web sites present in the previous set which are related to a particular disease. Then we produce the same description in the place of same link in set which have disease name of it. Finally, we will have a set which will have the disease name and description in the form of website.

E. Processing the Description

Here we process the description of the diseases which has the symptoms in it. We basically extract the symptoms form it. We first process it in the form of the symptoms we have and then we again process them to remove all the unnecessary scrap it contains with it. So, we do 2 operations on it, one for extracting the disease symptoms and other for cleaning the data for easy operation.

F. Formatting and Storing the Data

Here we modify the data into a data frame which have particular columns. There columns have disease name, symptoms and number of occurrences in it. Then we store the data into a csv file. If the file is already present it will override or if not present it will create a new one and store this data in it.

IV. RESULT AND DISCUSSIONS

Retrieved data from web scrapping into a .csv file. Then establish relation between the diseases and drugs which can be done by the reviews present. Get reviews to know which drugs are better and to know how many users have positive reviews. The structure of the data is that a patient with a unique ID purchases a drug that meets his condition and writes a review and rating for the drug he/she purchased on the date. Afterwards, if the others read that review and find it helpful, they will click useful Count, which will add 1 for the variable. Here we get the symptoms form the user. Based on the input the scroll symptoms produced as shown in Fig.2. User selects the symptoms provided by the list. These symptoms are just the ones present or found in the process of web scrapping. After selecting the symptoms user gets disease information as shown in Fig.3. After selecting the disease user gets the suggested drugs list as shown in Fig.4.

Fig .2 Symptoms Page

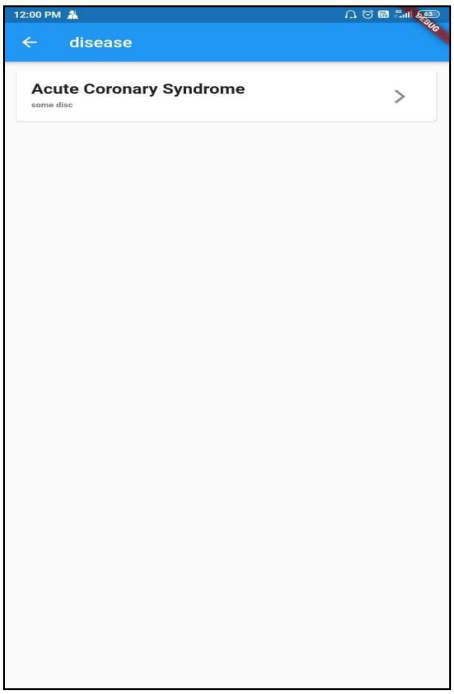


Fig.3 Disease Page

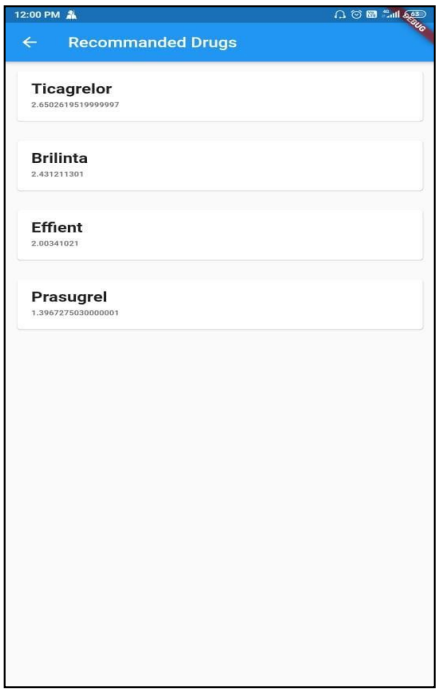


Fig.4 Drug Page

©IJRASET: All Rights are Reserved

V. CONCLUSION & FUTURE ENHANCEMENT

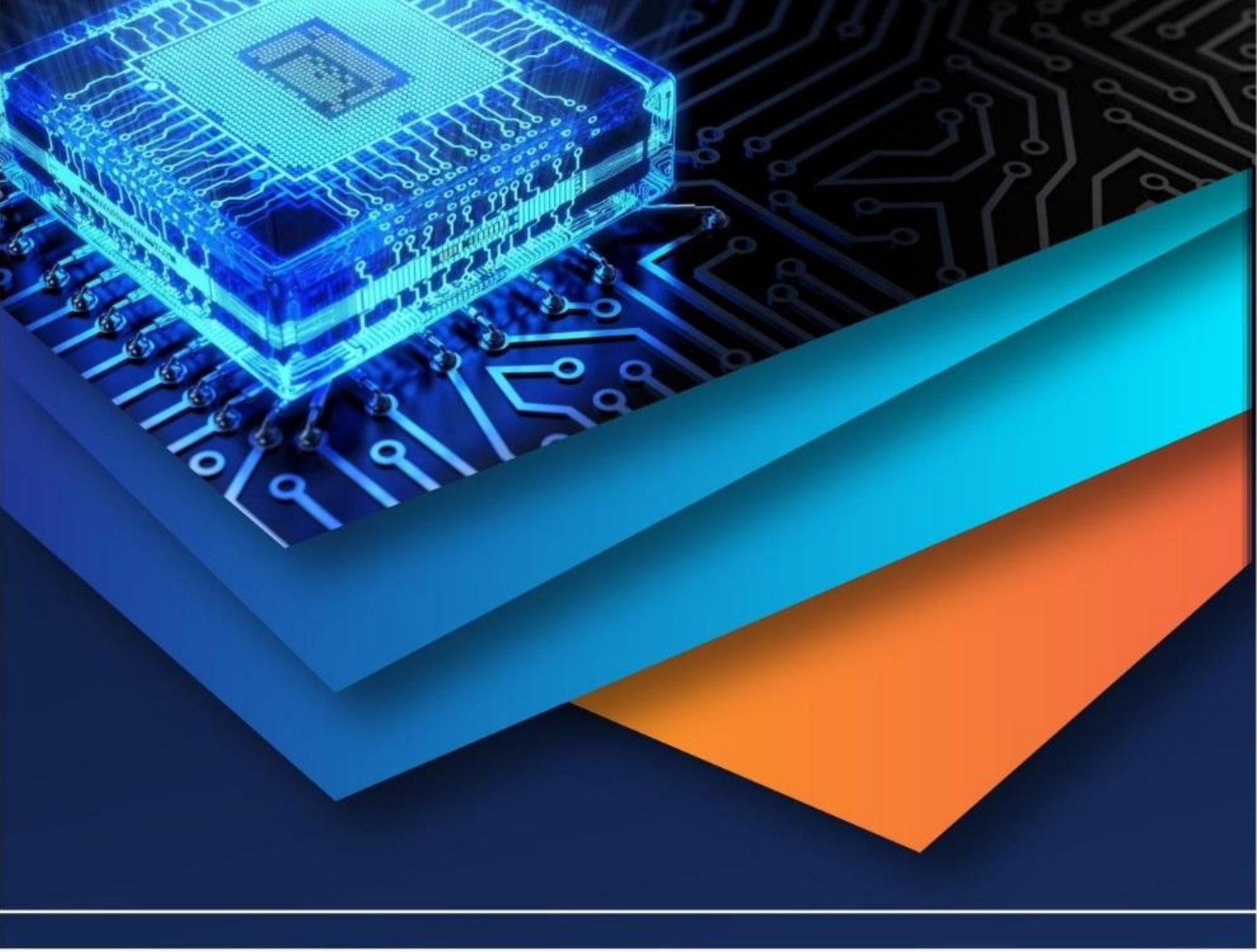
This application is capable of determining the diseases from a group of symptoms using multinomial naive byes algorithm, the diseases that are generated are used as input for the neural network model to determine the appropriate drugs. The validity of the drugs are then determined by the mean value of each drug. These predictive models are encapsulated into a flask application that is interconnected to the flutter framework which displays the results in a proper manner. The current limitations of the application are evident when determining a drug to a disease that is unique in nature and shares no similar traits to any other diseases. the results usually come empty or incorrect. Possible solution is to inject large amounts of data into the model, in order for the model to understand the nuances and possibly determine the correct drug.

This Study has noticed that there are different types of symptoms that can be found on the body. Where a patient wouldn't know how to accurately describe the condition that's forming on their body the symptom could range from specific type of acne, different types of lumps and the types of patches (which could sometimes can be an early sign of cancer). For these scenarios we can implement a image recognition system that can scan through a camera and attempt to determine the type of condition that is forming on the skin

REFERENCES

[1] F. Hutton Barron, Bruce E. Barrett “IEEE Transactions on Systems, Man, and Cybernetics: Systems” Volume: 49 , Issue: 11 , Nov. 2019.
[2] Felipe Jordão Almeida Prado Mattosinho, “Master Thesis, Mining Product Opinions and Reviews on the Web”, TU Dresden, July. 2010.
[3] Benelallam, A., Gómez, A., Sunyé, G., Tisi, M., Launay, D, ”A Scalable Persistence Layer for EMF Models”. In: Modelling Foundations and Applications, Lecture Notes in Computer Science, vol. 8569, pp. 230–241. Springer 2014.
[4] Youssef OUNEJJAR, Adil TANNOUCHE, “International Journal of Innovative Research in Advanced Engineering (IJIRAE)” ISSN: 2349-2163 Issue 1, Volume 2, Page – 178, January 2015.
[5] Samina Khalid, Tehmina Khalil, Shamila Nasreen,” A survey of feature selection and feature extraction techniques in machine learning”, Science and Information Conference IEEE, 2014.
[6] Matthew L. Jockers, Daniela M., “Literary and Linguistic Computing”, Volume 25, Issue 2, Pages 215–223, 12 April 2010.
[7] Jerome H Friedman. “Greedy function approximation: a gradient boostingmachine Annals of statistics" , pages 1189–1232, 2001.

©IJRASET: All Rights are Reserved



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)