Major Project Report

On

PREDICTION OF PHYSICIANS FOR PATIENT DIAGNOSIS

Submitted by

Bharath Simha - 16IT146

Bharat Sharma - 16IT210

D. Praneetha - 16IT215

Under the Guidance of

Prof. Ananthanarayana V. S.

Department of Information Technology

NITK Surathkal

Date of Submission: June 17, 2020

Department of Information Technology

National Institute of Technology Karnataka, Surathkal

2019-2020

Department of Information Technology, NITK Surathkal

Major Project

End Semester Evaluation Report (June 2020)

Course Code : IT499

Course Title : Major Project

Project Title : Prediction of Physicians For Patient Diagnosis

Project Group:

Name of the Student	Register No.	Signature with Date
Bharath Simha	16IT146	
Bharat Sharma	16IT210	
D. Praneetha	16IT215	

Place:

Date: *(Name and Signature of Major Project Guide)*

Abstract

This work provides a predictive model for selecting the most appropriate health care practitioners nearby who can diagnose a patient. First, identification of the doctors who can diagnose a patient is done. Second, probabilities are used to provide a ranking of each physician. Then the top physicians with higher probability to diagnose the disease are picked. For each physician, the top five specialists in each field in the nearby location from Practo's user are identified. These specialists are filtered through web scraping on Practo. These specialists are ranked using sentimental analysis on the reviews from the patients who previously visited them. In order to evaluate our results, random forest and logistic regression models are used. Then the construction of a basic user interface to suggest select Practo specialists based on the results is done. In conclusion, it is asserted that all selected specialists are able to diagnose the patient to an extent and that some specialists have a greater ability to diagnose the disease than others.

Keywords- *Open data, Logistic Regression, Random Forest, Practo, Sentiment Analysis, Flask*

Contents

List of Figures

1 Introduction

Medicine has developed a lot over the years and with technology also evolving, many online resources have come into the picture. Web-based applications that help people find the right doctor for them by providing doctors' details in their localities as well as fixing online appointments. But a major problem prevails to this day. With improving technology and exponentially increasing online information, it is as easy as asking a search engine a question to find the answers to what you're looking for. And this applies to when one's trying to find a cause for the weird symptoms that one might be facing whenever one feels sick. People start looking up online about the possible causes and tend to self-diagnose a lot. By self-diagnosing, they may be missing something that they cannot see. Another danger of self-diagnosis is that they may think that there is more wrong with themselves than there actually is. Self-diagnosis is also a problem when they are in a state of denial about their symptoms.

There are also cases where they approach their family doctor. But most of the times, they refer the patient to another doctor as the sickness might not be their area of expertise, and this might go on and on. There might also be the case of patients thinking that a particular physician is who they should be consulting based on their experience or their current symptoms. But the same case of deflection might happen. Now, both of these cases waste a ton of both the patient's as well as the doctors' valuable time. Given the technological advancement and availability of healthcare data, it can be used for finding patterns and extracting knowledge to provide better patient care and effective diagnostic capabilities. This project tries to achieve exactly that. It tries to fill in this gap of situations that lead to a patient self-diagnosing.

2 Literature Survey

2.1 *Related Work*

When considering Machine Learning models applied in the medical field, this particular paper [1] uses Naive Bayes to predict the disease based on the symptoms. And then suggests the details of the disease specialist based on their success rates. The evaluation metrics, accuracy, is found to be above 80%. Here, the doctors can be chosen based on success rate or fare. The only disadvantage here is that they have very few testing and training symptoms and use limited filters. In another paper by Qiwei Han [2], they essentially focus on patient-doctor matchmaking by considering a hybrid technique which finds the doctors visited by similar patients who visit the same doctor. They generate a list of doctors for each patient ranked by the predicted trust values. The disadvantage here is that they've collaborated with actual patients and have taken their previous records and considered those for the matchmaking process instead of considering only the current symptoms i.e. they have a personalized doctor recommendation system. And this is not for the public (individual centralized) which is what this project aims to achieve.

In this other paper by Disha Mahajan [3], they classify the data according to the requirements and then by applying association rules on it, they predict the diseases. With given symptoms of the patients, predicting disease and recommendation of the prescription of the obtained disease is done. But this prediction is done only for the prescription and not the actual recommendation of the doctors, and also, the prescription is done for very few diseases. The other paper is again a recommendation system [4]. The goal is to develop a recommendation system for identifying KOLs for any specific disease with unsupervised learning models. Now, this system can actually make recommendations to pharmaceutical companies and patients but isn't directly connected to doctors or pharmaceuticals. The last paper [5] by Abhaya and Chittaranjan, gives a proposed Intelligent HRS using Convolutional Neural Network (CNN) deep learning methods. The system finds recommended hospitals by calculating the similarity of patients choices. But here, the execution time is very high. If the similarity between the patients' choices is very less, the data would lose the vital information, and furthermore, the output doesn't really deal with actual doctors' recommendation, but instead, gives out the best possible hospitals according to the patient's choice history.

2.2 *Outcome of Literature Survey*

As seen in the last section, it has been noticed that a lot of research work and recommendation systems have been made but they've been done either with disease prediction or with only hospitals in mind. And the few papers which have actually considered a direct relation between a patient and a doctor have considered only a single doctor to be recommended, instead of a list of ranked doctor fields or doctors. Now, when considering the availability factor of a physician, the recommendation system would have to consider that aspect of a doctor as well. Another thing that was observed was that most of these papers collaborated with various hospitals and used their data for prediction. Now, this really puts a break in bench-marking this project's results, and hence, brings us to the part where the decision to use multiple models to compare the results is made. Part of the goal is to consider this lack of unavailability of a doctor recommended by a recommendation system and provide trusted alternatives who can diagnose the patient properly.

2.3 *Problem Statement*

An efficient predictive model for selecting an appropriate group of healthcare practitioners based on patient details.

2.4 *Objectives*

- Determining the type of doctor required to diagnose the patient based on patient details.
- Construct an eligible group of type of doctors to which a patient can refer to, in order.
- Collecting doctors' details from open-source data and using a model to construct their ranking based on online reviews.
- Mapping the Practo profiles with the predicted type of doctors.

3 Detailed Design

3.1 *System Architecture*

The first objective is achieved by combining the initial data-sets into a final data-set which has all the required patient data. This data is then used to train the logistic regression and random forest models. Upon testing with patient cases, the models' outputs are evaluated in such a way that the final output produced is a group of doctors ranked by probability of diagnosing that patient. This is done using what we call Beta values. Then the set of doctors who can successfully diagnose the patient based on a threshold are selected. After that, the generated output; doctor fields/types, is used to filter Practo doctors' data.

The Practo data is scraped before hand and is kept in a separate file, from where textblob is applied on it for sentiment analysis.

The trained models are stored in pickle files and are used for testing through the user interface. Finally, the patient is given the recommendation of these top Practo doctors with their name with a link to their Practo profile through the user interface constructed.

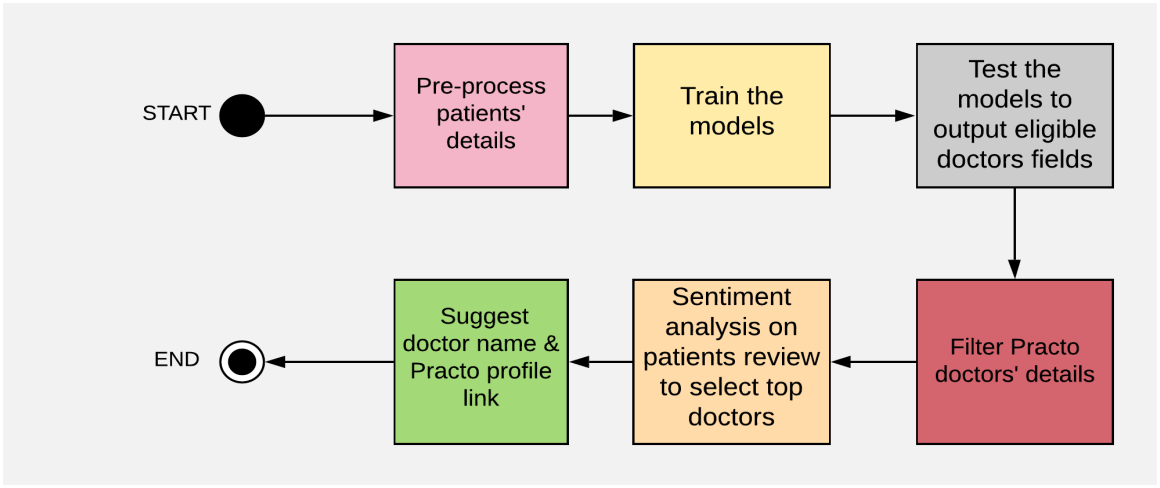

Figure 3.1: System Architecture.

3.2 *Datasets Analysis*

Two different data sets are used to combine a final data set for a group of suitable doctors' prediction. The patient data set is the SPARCS one and new data set is manually built to obtain proper symptoms for diseases mentioned in the SPARCS data set and the fields of doctors who can properly diagnose it.

3.2.1 *Patient Data Pre-processing*

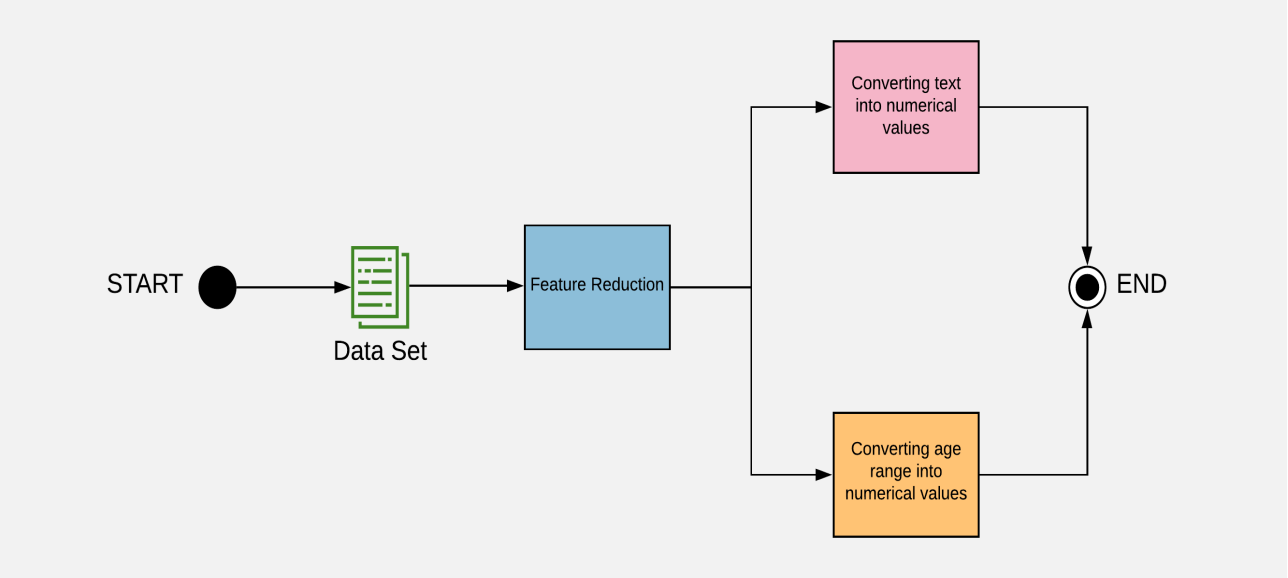

Figure 3.2: SPARCS data pre-processing.

The SPARCS data set used in this work has more than 1,000,000 observations and 39 variables in its raw state, which includes many missing, redundant and irrelevant values. All variables not correlated with the objectives of the study were removed. A total of 37 diseases were selected, with more than 50 different symptoms, and 11 types of physicians corresponded with the diagnoses of the above diseases. Then this SPARCS data set is combined with the manually built data set using the column *CCS Diagnosis Description* of the SPARCS data which has the disease the patient is diagnosed with.

3.2.2 *Practo Data Collection and Pre-processing*

As shown in the figure below, this project makes use of *BeautifulSoup* and *urllib* module's request functions to access the website and scrape the doctor's details and the corresponding patient reviews. Then the data is pushed into a file for further analysis.
Now, it needs to be determined what exactly a patient's review conveys i.e. sentiment analysis. This way, when given a test case, the best possible Practo doctor would be filtered not only based on the ratings, experience, etc. but also, based on the overall sentiment of the patients who were examined by them. This is done using textblob. The final doctors' data set made consists of doctor's name, field, location and profile link.

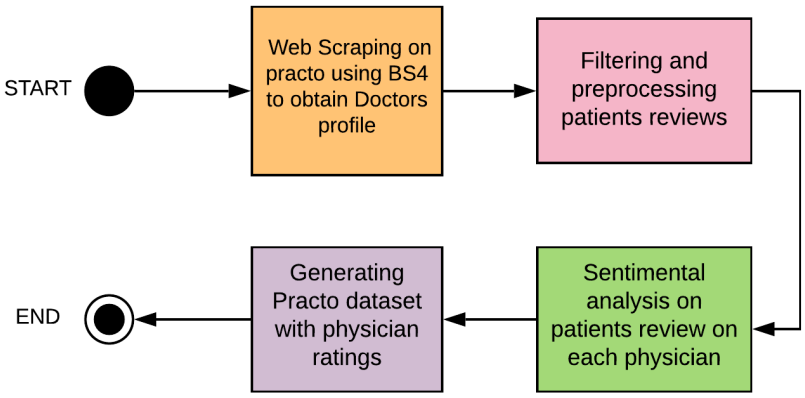

Figure 3.3: Practo data pre-processing.

3.3 *Predictive Models Used*

The goal is to predict the physicians who are likely to provide an appropriate diagnosis for a patient based on the patient's profile. This can be done using two predictive models. These are available in python's Sklearn library.

- Logistic Regression: Logistic regression is estimating the parameters(Beta values) of a logistic model through training and predicting the probabilities of physicians. Beta values get relative physician probabilities, physicians are ranked based on these probabilities. These can be extracted using sklearn's coefficient function available for its model.

- Random Forest: Pick random patients data for the training set and build the decision tree associated with those data. And then build a tree for every column picked randomly. For a new data point, make each one of your trees predict the value of Y for the data point, and assign the new data point the average across all of the predicted Y values. The basic idea behind this is to combine multiple decision trees in determining the final output rather than relying on individual decision trees.

Both these models are used in a One vs Rest (OvR or one-vs-all, OvA or one-against-all, OAA) manner. This strategy involves training a single classifier per class, with the samples of that class as positive samples and all other samples as negatives. These models are trained on

the final data set. Their accuracies are checked and they're tested on a few test cases before finally storing them into pickle files for further usage.

3.4 *Practo Details Mapping*

When testing the models, the obtained list of doctors and the patient address is considered for filtering doctors in Practo. This enables us to filter Practo doctors by locality and their field.

Now, within these filtered doctor's list, the reviews for each doctor are analyzed with the help of TextBlob. TextBlob is a python library and offers a simple API to access its methods and perform basic NLP tasks. The sentiments of the sentences can be obtained using textblob.sentiment() method. The average sentiment of all the reviews for each doctor are taken along with average rating, overall experience, and overall patients diagnosed to decide the top five doctors to recommend to the patient.

So, for every doctor's field generated by the predictive models i.e. the list of doctors' fields, maximum top five doctors are selected.

3.5 *Webinterface*

In order to make use of the model, a web interface is designed which recommends user a list of doctors based on the details filled. The interface is implemented using Flask. Flask is a micro web framework written in Python. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions.
Since, trained model is already saved in a pickle file, users' input is taken via HTML page and then, queried to the model which returns the type of physician. Then, the type is queried to a CSV file which contains the doctor's details and returns the top physicians (max 5) based on the ratings and location. Then, the final result is displayed on the web page as a list of doctors. Each doctors name contains a hyperlink which redirects to their Practo profile.

4 Results and Discussion

The final data set obtained consists of eleven columns and around 60,000 rows. The first SPARCS data set is shown in figure 4.1 after feature extraction. The second manually built data set is shown in figure 4.2. The data set obtained after combining both of these is shown in figure 4.3.

30 to 49	F	197	Minor
50 to 69	F	146	Minor
18 to 29	M	50	Minor
18 to 29	F	154	Minor
18 to 29	F	124	Minor
70 or Older	F	123	Minor
70 or Older	M	122	Major
30 to 49	M	123	Minor

Figure 4.1: Data extracted from SPARCS.

gastroesophageal reflux disease	pain	pain chest	burning sensation	hyponatremia	gastroenterologist
dehydration	fever	diarrhea	vomiting	hypotension	general physician
cardiomyopathy	shortness of breath	orthopnea	hypokinesia	jugular venous distention	cardiologist
chronic kidney failure	vomiting	orthopnea	hyperkalemia	oliguria	nephrologist

Figure 4.2: Symptoms and diseases data.

18 to 29	F	124	Minor	cough	fever	decreased translucency	shortness of breath	pulmonologist
70 or Older	F	123	Minor	cough	wheezing	shortness of breath	chest tightness	Allergist
70 or Older	M	122	Major	wheezing	cough	shortness of breath	chest tightness	Allergist

Figure 4.3: Data set after combining data from both 4.1. and 4.2.

After pre-processing, each patient record contains 7 features, as shown in figure 4.4:

	Age	Gender	S1	S2	S3	S4	Doctor
0	1	1	35	41	11	1	9
1	4	0	7	19	9	41	0
2	3	0	33	35	4	24	0
3	3	1	29	17	48	20	4
4	0	1	19	10	50	25	10

Figure 4.4: Final data after pre-processing.

- Patients age range: This variable describes the age range of the patient. There are 5 ranges: 0-17, 18-29, 30-49, 50-69 and 70+.
- Gender: This variable indicates the sex of the patient, male or female.
- Symptoms: A set of 4 variables that describe the manifestation of a disease. Each variable, Symptom1, Symptom2, Symptom3 and Symptom4 contains a set of clinical signs.
- Class label: The doctor's field is the class label for the data set.

Precision and accuracy for all the 11 classes for logistic regression model using One vs Rest Classifier is as shown in figure 4.5

	Precision	Accuracy
0	0.82	0.86
1	0.82	0.86
2	0.82	0.86
3	0.82	0.86
4	0.83	0.87
5	0.82	0.86
6	0.82	0.86
7	0.87	0.90
8	0.82	0.86
9	0.83	0.87
10	0.83	0.86

Figure 4.5: Precision and accuracy for LR model.

Probabilities of all 11 classes to be able to diagnose the patients for some 3 test cases for logistic regression model using One vs Rest classifier. are shown in the figure 4.6.

	Allergist	Endocrinologist	General Physician	Cardiologist	Gastroenterologist	Nephrologist	Neurologist	Pediatrician	Psychiatrist	Pulmonologist	Rheumatologist
0	0.905516	0.897717	0.921901	0.913620	0.913422	0.920392	0.897668	0.999573	0.919471	0.912018	0.912724
1	0.912265	0.935740	0.920992	0.939244	0.927671	0.903604	0.899341	0.745244	0.905352	0.933433	0.931883
2	0.881476	0.906254	0.919435	0.912892	0.907052	0.891713	0.883865	0.999990	0.905735	0.908178	0.920634

Figure 4.6: Probabilities of each type of physician.

Precision and accuracy for all the 11 classes for random forest model using One vs Rest Classifier are shown in figure 4.7

	Precision	Accuracy
0	0.84	0.86
1	0.82	0.86
2	0.82	0.86
3	0.82	0.86
4	0.82	0.86
5	0.82	0.86
6	0.82	0.86
7	0.89	0.90
8	0.82	0.86
9	0.82	0.86
10	0.82	0.86

Figure 4.7: Precision and accuracy for RF model.

Probabilities of all 11 classes to be able to diagnose the patients for some 3 test cases for random forest model using One vs Rest classifier are shown in figure 4.8

	Allergist	Endocrinologist	General Physician	Cardiologist	Gastroenterologist	Nephrologist	Neurologist	Pediatrician	Psychiatrist	Pulmonologist	Rheumatologist
0	0.931355	0.938899	0.933221	0.814891	0.944916	0.934478	0.883007	1.000000	0.932431	0.731708	0.946216
1	0.969033	0.904769	0.898911	0.966588	0.854188	0.905392	0.972800	0.638555	0.937987	1.000000	0.939697
2	0.979752	0.854084	0.855233	0.887698	0.864322	0.849877	0.882552	1.000000	0.958358	0.979421	0.905205

Figure 4.8: Probabilities of each type of physician.

Both these trained models were stored in pickle files; to be accessed while building the user interface for testing the user input data.

The basic code for scraping Practo doctor's data was written as well. This was done in python using *BeautifulSoup* and the *urllib*'s request module. This code allows us to scrape the doctors' data with respect to their field of specialization. The basic code for fetching the profile data for these 11 types of doctors was written. And the data obtained is as shown in figures 4.9 and 4.10.

The fields are as shown in figure 4.9 i.e. name, specialization, location and profile link.

The user interface constructed takes the inputs as shown in figure 4.11. And the output is produced as the names of the doctors with a hyperlink to their Practo profiles as shown in figures 4.12 and 4.13, respectively.

	Name	Specialization	Location	Link
2	Dr. Sheela Chakravarthy	Internal Medicine	bangalore	https://www.practo.com/bangalore/doctor/sheela-chakravarthy-internal-medicine?specialization=Internal Med
3	Dr. Sheetal Kamat	Internal Medicine	bangalore	https://www.practo.com/bangalore/doctor/sheetal-kamat-internal-medicine?specialization=Internal Medicine&
4	Dr. B Rajashekar	General Physician	bangalore	https://www.practo.com/bangalore/doctor/dr-rajashekhar-general-physician?specialization=General Physician&
5	Dr. Raja Selvarajan	Diabetologist	bangalore	https://www.practo.com/bangalore/doctor/dr-raja-selvarajan-diabetologist?specialization=Diabetologist&prac
6	Dr. Sharat Honnatti	general physician	bangalore	https://www.practo.com/bangalore/doctor/dr-sharat-honnatti-general-physician?specialization=&practice_id=
7	Dr. Ashok M N	General Physician	bangalore	https://www.practo.com/bangalore/doctor/dr-ashok-m-n-cardiologist?specialization=General Physician&practi
8	Dr. Tharanath S	General Physician	bangalore	https://www.practo.com/bangalore/doctor/dr-tharanath-s-general-physician?specialization=General Physician&
9	Dr. Pankaj Singhai	Internal Medicine	bangalore	https://www.practo.com/bangalore/doctor/dr-pankaj-singhai-1-internal-medicine?specialization=Internal Medi
10	Dr. Shalini Joshi	Internal Medicine	bangalore	https://www.practo.com/bangalore/doctor/shalini-joshi-internal-medicine?specialization=Internal Medicine&p
11	Dr. Mohan Badagandi	General Physician	bangalore	https://www.practo.com/bangalore/doctor/dr-mohan-badagandi-diabetologist?specialization=General Physicia
12	Dr. Renu Saraogi	General Physician	bangalore	https://www.practo.com/bangalore/doctor/dr-renu-saraogi-general-physician?specialization=General Physician
13	Dr. Tharangini S R	General Physician	bangalore	https://www.practo.com/bangalore/doctor/dr-tharangini-s-r-general-physician?specialization=General Physicia
14	Dr. Sachin	General Physician	bangalore	https://www.practo.com/bangalore/doctor/sachin-56-general-physician?specialization=General Physician&prac
15	Dr. Sudha Menon	Internal Medicine	bangalore	https://www.practo.com/bangalore/doctor/dr-sudha-menon-internal-medicine?specialization=Internal Medicin
16	Dr. Dinesh V Kamath	General Physician	bangalore	https://www.practo.com/bangalore/doctor/dr-dinesh-v-kamath-diabetologist-1?specialization=General Physici
17	Dr. R Manjunath	Internal Medicine	bangalore	https://www.practo.com/bangalore/doctor/dr-r-manjunath-internal-medicine?specialization=Internal Medicine
18	Dr. Manohar K N	Internal Medicine	bangalore	https://www.practo.com/bangalore/doctor/dr-manohar-kn-physiotherapist?specialization=Internal Medicine&p
19	Dr. Ambanna Gowda	Internal Medicine	bangalore	https://www.practo.com/bangalore/doctor/dr-ambana-general-physician?specialization=Internal Medicine&pra
20	Dr. S G Puranik	General Physician	bangalore	https://www.practo.com/bangalore/doctor/s-g-puranik-general-physician?specialization=General Physician&pr
21	Dr. Sheetal Reddy Desai	Diabetologist	bangalore	https://www.practo.com/bangalore/doctor/dr-sheetal-reddy-desai-internal-medicine?specialization=Diabetolo
22	Dr. Jayasree Kailasam	Internal Medicine	bangalore	https://www.practo.com/bangalore/doctor/dr-jayasree-kailasam-internal-medicine?specialization=Internal Med
23	Dr. Ramesh S.	Internal Medicine	bangalore	https://www.practo.com/bangalore/doctor/dr-ramesh-s-internal-medicine?specialization=Internal Medicine&p
24	Dr. Ramesh Kumar D	Internal Medicine	bangalore	https://www.practo.com/bangalore/doctor/dr-ramesh-kumar-d-dentist?specialization=Internal Medicine&prac

Figure 4.9: Raw Practo data after initial scraping.

```
11 10 False
I recommend the doctor ::: She is my favorite doctor visited multiple times for different issues. She is very friendly and removes our worry
I recommend the doctor ::: She has very friendly and patient nature. Listened to all my problems and gave ****** treatment. I feel that much
I recommend the doctor ::: It was a very nice experience with the doctor.
I do not recommend the doctor ::: *** ******** ** ****** ********* **** ********** *** was able detect the cause .She was very polite in beh
I recommend the doctor ::: My mother had a post spine surgery follow up with Dr. Sheela and was asked to come on a particular date to the ho
I recommend the doctor ::: I have been consulting Dr. Sheela Chakravarthy from last two months. She is very **** *** approachable. She liste
I recommend the doctor :::
I recommend the doctor ::: Only one session has taken place until now and it was satisfactory. The diagnosis was executed very well and the
I recommend the doctor ::: I am writing this on behalf of my father, he was complained with low platelet count. As soon as we got to know ab
I recommend the doctor ::: Dr. Sheela Chakravarthy is an excellent doctor. Her treatment was good. Staff also co operated well. Consultation
```

Figure 4.10: Snapshot of feedback from initial scraping.

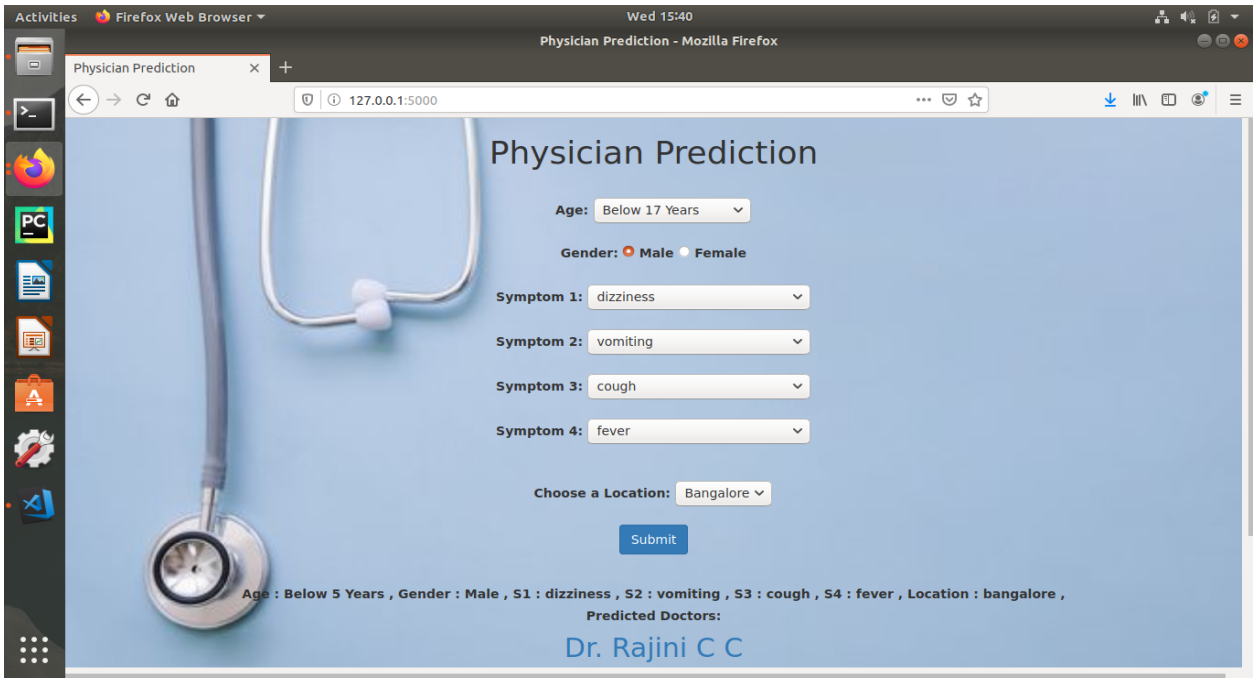

Figure 4.11: Web interface for physicians prediction.

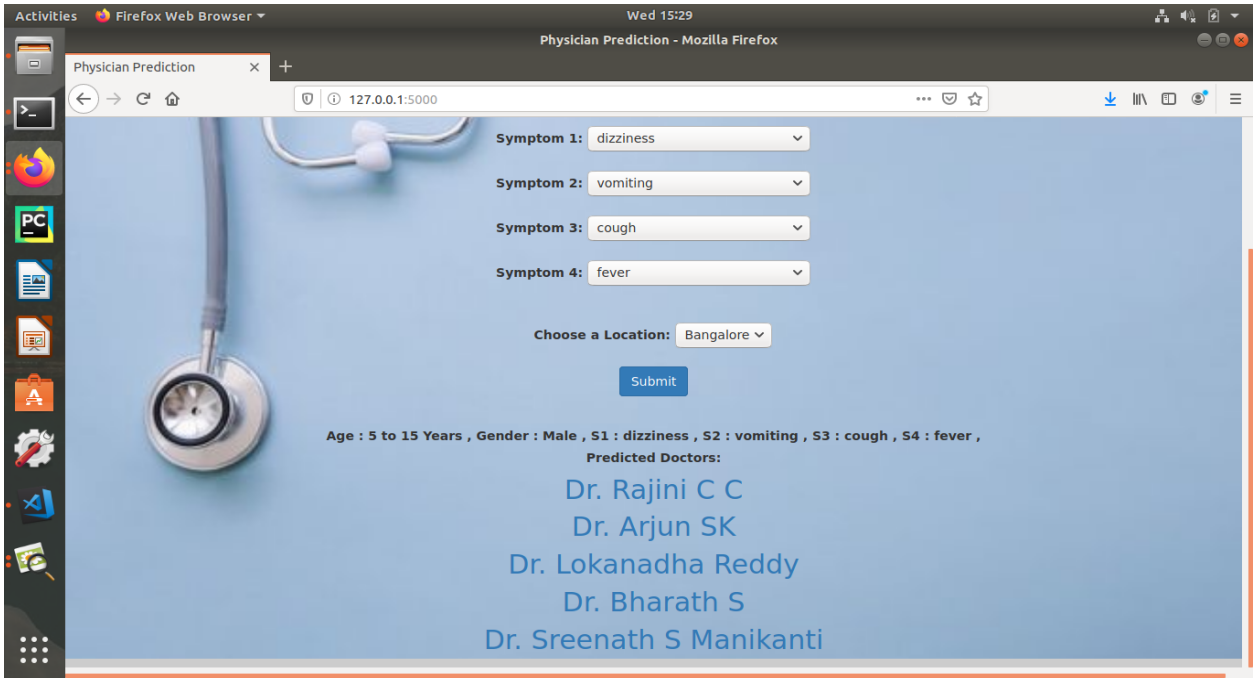

Figure 4.12: List of recommended doctors based on the given input.

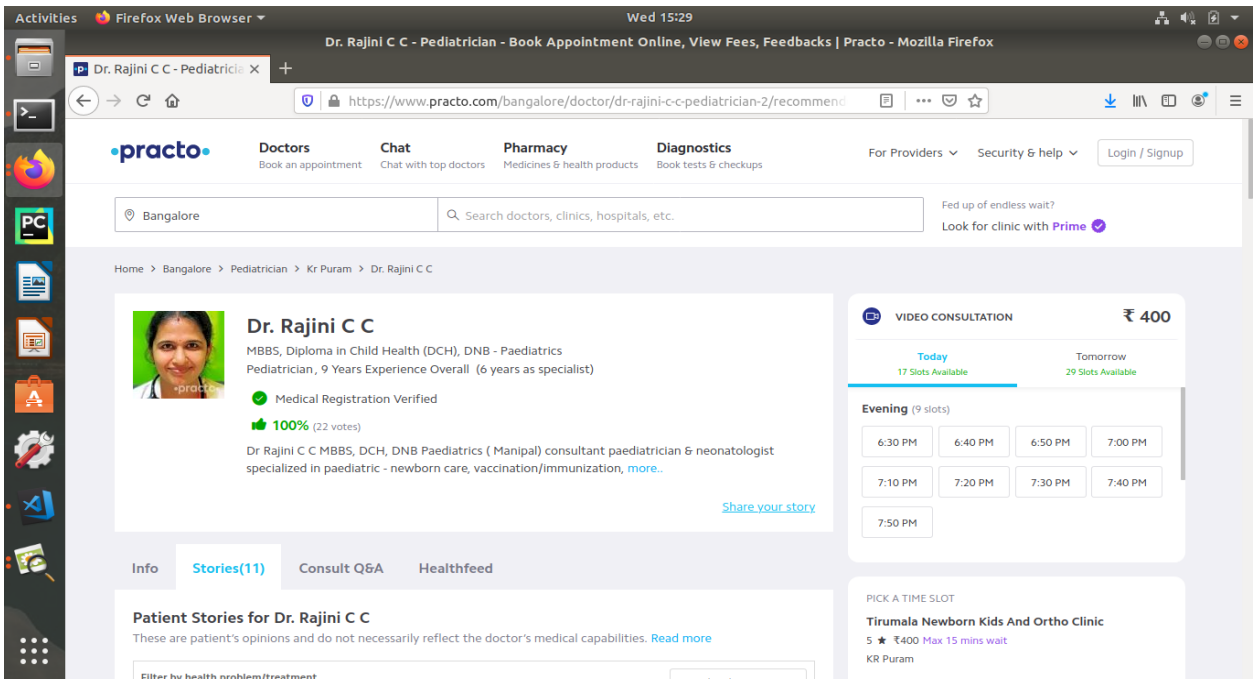

Figure 4.13: Redirection to doctor's profile on practo website.

5 Conclusion and Future Work

In this work, pre-processing of the overall data-set was done, as mentioned in the previous sections. And then, manual collection of the symptoms, diseases and type of doctors data was done. The two data sets were mapped together and were finally pre-processed. The structure for web is made for scraping, as mentioned, for collecting the required Practo data of doctors. The objective to predict the doctor's field is successfully achieved. Sentiment analysis is successfully done on the Practo comments for the respective doctors. And finally, a basic user interface is constructed to bring together both the parts of this project i.e. predicting doctor's field and performing sentiment analysis on Practo data.

The future work is to improve the user interface as well as convert this into an android application. Data from other websites along with Practo can be considered as well. More disease symptoms and type of doctors can be added.

6 Timeline of the Project

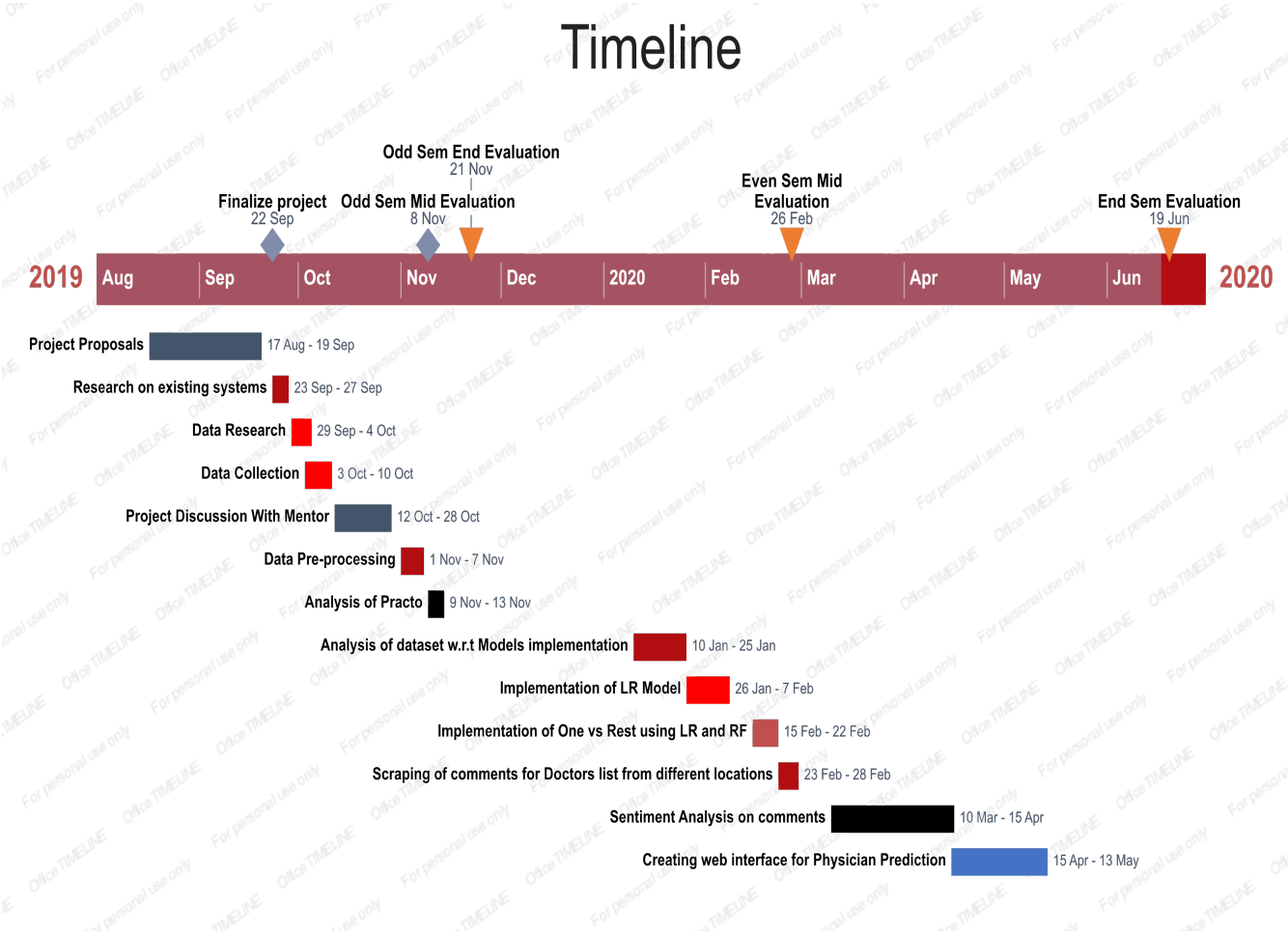

Figure 6.1: Timeline

References

[1] Tejaswita P. Vaidya ,Dhanashri Gujar and Rashmi Biyani, Disease Prediction and Doctor Recommendation System.", International Research Journal of Engineering and Technology, Volume 05, Issue 03, March 2018

[2] Ramakrishnan, Naren Keller, Benjamin Mirza, Batul Grama, Ananth Karypis, George. (2001). When being Weak is Brave: Privacy in Recommender Systems. Computing Research Repository - CORR.

[3] Disha Mahajan , Mrudula Phalak , Saniya Pathan," Prediction System for Diseases and Suggestion of Appropriate Medicines",International Journal of Innovations Advancement in Computer Science, Volume 6, Issue 11, December 2017

[4] Guo, Li Jin, Bo Yao, Cuili Yang, Haoyu Huang, Degen Wang, Fei. (2016). Which Doctor to Trust: A Recommender System for Identifying the Right Doctors. Journal of Medical Internet Research. 18. e186. 10.2196/jmir.6015.

[5] Sahoo, A.K.; Pradhan, C.; Barik, R.K.; Dubey, H. DeepReco: Deep Learning Based Health Recommender System Using Collaborative Filtering. Computation 2019.

[6] Federico Cesconi, Textblob - Natural language processing sentiment analysis explained to business people, Available:https://textblob.readthedocs.io/en/dev/quickstart.html, Visited: 28th October 2019

PREDICTION OF PHYSICIANS FOR PATIENT DIAGNOSIS

by Bharath Simha, Bharath Sharma, D. Praneetha 16it146, 16it210,16it215

Submission date: 26-Jun-2020 05:37PM (UTC+0530)
Submission ID: 1349980429
File name: Major_project-Bharat.pdf (3.23M)
Word count: 3439
Character count: 17979

Major Project Report

On

PREDICTION OF PHYSICIANS FOR PATIENT DIAGNOSIS

Submitted by

Bharath Simha - 16IT146

Bharat Sharma - 16IT210

D. Praneetha - 16IT215

Under the Guidance of

Prof. Ananthanarayana V. S.

Department of Information Technology

NITK Surathkal

Date of Submission: June 17, 2020

Department of Information Technology

National Institute of Technology Karnataka, Surathkal

2019-2020

Department of Information Technology, NITK Surathkal

Major Project

End Semester Evaluation Report (June 2020)

Course Code : IT499

Course Title : Major Project

Project Title : Prediction of Physicians For Patient Diagnosis

Project Group:

Name of the Student	Register No.	Signature with Date
Bharath Simha	16IT146	
Bharat Sharma	16IT210	
D. Praneetha	16IT215	

Place:

Date:

(Name and Signature of Major Project Guide)

Abstract

This work provides a predictive model for selecting the most appropriate health care practitioners nearby who can diagnose a patient. First, identification of the doctors who can diagnose a patient is done. Second, probabilities are used to provide a ranking of each physician. Then the top physicians with higher probability to diagnose the disease are picked. For each physician, the top five specialists in each field in the nearby location from Practo's user are identified. These specialists are filtered through web scraping on Practo. These specialists are ranked using sentimental analysis on the reviews from the patients who previously visited them. In order to evaluate our results, random forest and logistic regression models are used. Then the construction of a basic user interface to suggest select Practo specialists based on the results is done. In conclusion, it is asserted that all selected specialists are able to diagnose the patient to an extent and that some specialists have a greater ability to diagnose the disease than others.

Keywords- *Open data, Logistic Regression, Random Forest, Practo, Sentiment Analysis, Flask*

Contents

List of Figures

1 Introduction

Medicine has developed a lot over the years and with technology also evolving, many online resources have come into the picture. Web-based applications that help people find the right doctor for them by providing doctors' details in their localities as well as fixing online appointments. But a major problem prevails to this day. With improving technology and exponentially increasing online information, it is as easy as asking a search engine a question to find the answers to what you're looking for. And this applies to when one's trying to find a cause for the weird symptoms that one might be facing whenever one feels sick. People start looking up online about the possible causes and tend to self-diagnose a lot. By self-diagnosing, they may be missing something that they cannot see. Another danger of self-diagnosis is that they may think that there is more wrong with themselves than there actually is. Self-diagnosis is also a problem when they are in a state of denial about their symptoms.

There are also cases where they approach their family doctor. But most of the times, they refer the patient to another doctor as the sickness might not be their area of expertise, and this might go on and on. There might also be the case of patients thinking that a particular physician is who they should be consulting based on their experience or their current symptoms. But the same case of deflection might happen. Now, both of these cases waste a ton of both the patient's as well as the doctors' valuable time. Given the technological advancement and availability of healthcare data, it can be used for finding patterns and extracting knowledge to provide better patient care and effective diagnostic capabilities. This project tries to achieve exactly that. It tries to fill in this gap of situations that lead to a patient self-diagnosing.

2 Literature Survey

2.1 *Related Work*

When considering Machine Learning models applied in the medical field, this particular paper [1] uses Naive Bayes to predict the disease based on the symptoms. And then suggests the details of the disease specialist based on their success rates. The evaluation metrics, accuracy, is found to be above 80%. Here, the doctors can be chosen based on success rate or fare. The only disadvantage here is that they have very few testing and training symptoms and use limited filters. In another paper by Qiwei Han [2], they essentially focus on patient-doctor matchmaking by considering a hybrid technique which finds the doctors visited by similar patients who visit the same doctor. They generate a list of doctors for each patient ranked by the predicted trust values. The disadvantage here is that they've collaborated with actual patients and have taken their previous records and considered those for the matchmaking process instead of considering only the current symptoms i.e. they have a personalized doctor recommendation system. And this is not for the public (individual centralized) which is what this project aims to achieve.

In this other paper by Disha Mahajan [3], they classify the data according to the requirements and then by applying association rules on it, they predict the diseases. With given symptoms of the patients, predicting disease and recommendation of the prescription of the obtained disease is done. But this prediction is done only for the prescription and not the actual recommendation of the doctors, and also, the prescription is done for very few diseases. The other paper is again a recommendation system [4]. The goal is to develop a recommendation system for identifying KOLs for any specific disease with unsupervised learning models. Now, this system can actually make recommendations to pharmaceutical companies and patients but isn't directly connected to doctors or pharmaceuticals. The last paper [5] by Abhaya and Chittaranjan, gives a proposed Intelligent HRS using Convolutional Neural Network (CNN) deep learning methods. The system finds recommended hospitals by calculating the similarity of patients choices. But here, the execution time is very high. If the similarity between the patients' choices is very less, the data would lose the vital information, and furthermore, the output doesn't really deal with actual doctors' recommendation, but instead, gives out the best possible hospitals according to the patient's choice history.

2.2 *Outcome of Literature Survey*

As seen in the last section, it has been noticed that a lot of research work and recommendation systems have been made but they've been done either with disease prediction or with only hospitals in mind. And the few papers which have actually considered a direct relation between a patient and a doctor have considered only a single doctor to be recommended, instead of a list of ranked doctor fields or doctors. Now, when considering the availability factor of a physician, the recommendation system would have to consider that aspect of a doctor as well. Another thing that was observed was that most of these papers collaborated with various hospitals and used their data for prediction. Now, this really puts a break in bench-marking this project's results, and hence, brings us to the part where the decision to use multiple models to compare the results is made. Part of the goal is to consider this lack of unavailability of a doctor recommended by a recommendation system and provide trusted alternatives who can diagnose the patient properly.

2.3 *Problem Statement*

An efficient predictive model for selecting an appropriate group of healthcare practitioners based on patient details.

2.4 *Objectives*

- Determining the type of doctor required to diagnose the patient based on patient details.
- Construct an eligible group of type of doctors to which a patient can refer to, in order.
- Collecting doctors' details from open-source data and using a model to construct their ranking based on online reviews.
- Mapping the Practo profiles with the predicted type of doctors.

3 Detailed Design

3.1 *System Architecture*

The first objective is achieved by combining the initial data-sets into a final data-set which has all the required patient data. This data is then used to train the logistic regression and random forest models. Upon testing with patient cases, the models' outputs are evaluated in such a way that the final output produced is a group of doctors ranked by probability of diagnosing that patient. This is done using what we call Beta values. Then the set of doctors who can successfully diagnose the patient based on a threshold are selected. After that, the generated output; doctor fields/types, is used to filter Practo doctors' data.

The Practo data is scraped before hand and is kept in a separate file, from where textblob is applied on it for sentiment analysis.

The trained models are stored in pickle files and are used for testing through the user interface. Finally, the patient is given the recommendation of these top Practo doctors with their name with a link to their Practo profile through the user interface constructed.

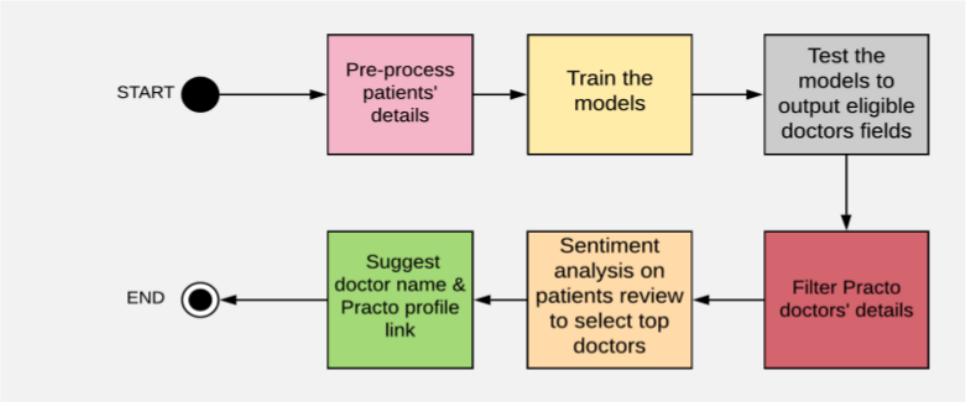

Figure 3.1: System Architecture.

3.2 *Datasets Analysis*

Two different data sets are used to combine a final data set for a group of suitable doctors' prediction. The patient data set is the SPARCS one and new data set is manually built to obtain proper symptoms for diseases mentioned in the SPARCS data set and the fields of doctors who can properly diagnose it.

3.2.1 *Patient Data Pre-processing*

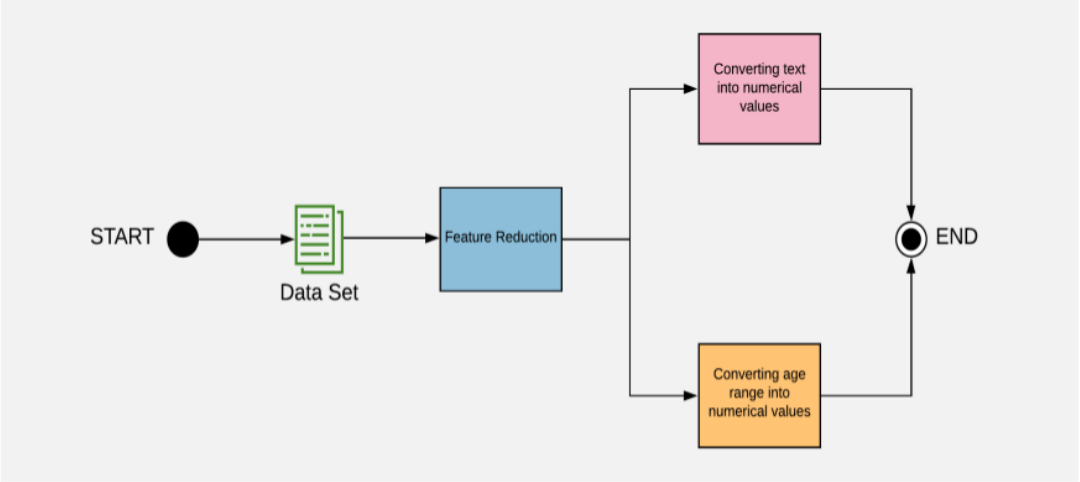

Figure 3.2: SPARCS data pre-processing.

The SPARCS data set used in this work has more than 1,000,000 observations and 39 variables in its raw state, which includes many missing, redundant and irrelevant values. All variables not correlated with the objectives of the study were removed. A total of 37 diseases were selected, with more than 50 different symptoms, and 11 types of physicians corresponded with the diagnoses of the above diseases. Then this SPARCS data set is combined with the manually built data set using the column *CCS Diagnosis Description* of the SPARCS data which has the disease the patient is diagnosed with.

3.2.2 *Practo Data Collection and Pre-processing*

As shown in the figure below, this project makes use of *BeautifulSoup* and *urllib* module's request functions to access the website and scrape the doctor's details and the corresponding patient reviews. Then the data is pushed into a file for further analysis.

Now, it needs to be determined what exactly a patient's review conveys i.e. sentiment analysis. This way, when given a test case, the best possible Practo doctor would be filtered not only based on the ratings, experience, etc. but also, based on the overall sentiment of the patients who were examined by them. This is done using textblob. The final doctors' data set made consists of doctor's name, field, location and profile link.

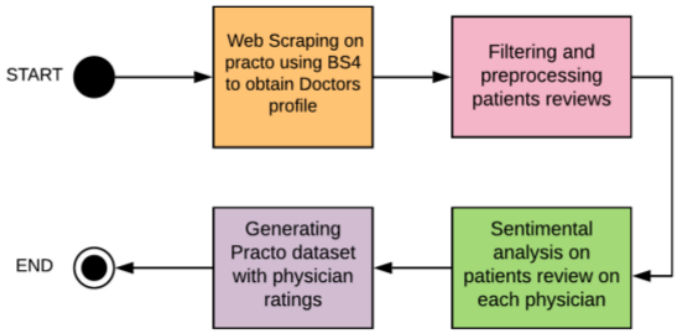

Figure 3.3: Practo data pre-processing.

3.3 *Predictive Models Used*

The goal is to predict the physicians who are likely to provide an appropriate diagnosis for a patient based on the patient's profile. This can be done using two predictive models. These are available in python's Sklearn library.

- Logistic Regression: Logistic regression is estimating the parameters(Beta values) of a logistic model through training and predicting the probabilities of physicians. Beta values get relative physician probabilities, physicians are ranked based on these probabilities. These can be extracted using sklearn's coefficient function available for its model.

- Random Forest: Pick random patients data for the training set and build the decision tree associated with those data. And then build a tree for every column picked randomly. For a new data point, make each one of your trees predict the value of Y for the data point, and assign the new data point the average across all of the predicted Y values. The basic idea behind this is to combine multiple decision trees in determining the final output rather than relying on individual decision trees.

Both these models are used in a One vs Rest (OvR or one-vs-all, OvA or one-against-all, OAA) manner. This strategy involves training a single classifier per class, with the samples of that class as positive samples and all other samples as negatives. These models are trained on

the final data set. Their accuracies are checked and they're tested on a few test cases before finally storing them into pickle files for further usage.

3.4 *Practo Details Mapping*

When testing the models, the obtained list of doctors and the patient address is considered for filtering doctors in Practo. This enables us to filter Practo doctors by locality and their field.

Now, within these filtered doctor's list, the reviews for each doctor are analyzed with the help of TextBlob. TextBlob is a python library and offers a simple API to access its methods and perform basic NLP tasks. The sentiments of the sentences can be obtained using textblob.sentiment() method. The average sentiment of all the reviews for each doctor are taken along with average rating, overall experience, and overall patients diagnosed to decide the top five doctors to recommend to the patient.

So, for every doctor's field generated by the predictive models i.e. the list of doctors' fields, maximum top five doctors are selected.

3.5 *Webinterface*

In order to make use of the model, a web interface is designed which recommends user a list of doctors based on the details filled. The interface is implemented using Flask. Flask is a micro web framework written in Python. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions.
Since, trained model is already saved in a pickle file, users' input is taken via HTML page and then, queried to the model which returns the type of physician. Then, the type is queried to a CSV file which contains the doctor's details and returns the top physicians (max 5) based on the ratings and location. Then, the final result is displayed on the web page as a list of doctors. Each doctors name contains a hyperlink which redirects to their Practo profile.

4 Results and Discussion

The final data set obtained consists of eleven columns and around 60,000 rows. The first SPARCS data set is shown in figure 4.1 after feature extraction. The second manually built data set is shown in figure 4.2. The data set obtained after combining both of these is shown in figure 4.3.

30 to 49	F	197	Minor
50 to 69	F	146	Minor
18 to 29	M	50	Minor
18 to 29	F	154	Minor
18 to 29	F	124	Minor
70 or Older	F	123	Minor
70 or Older	M	122	Major
30 to 49	M	123	Minor

Figure 4.1: Data extracted from SPARCS.

gastroesophageal reflux disease	pain	pain chest	burning sensation	hyponatremia	gastroenterologist
dehydration	fever	diarrhea	vomiting	hypotension	general physician
cardiomyopathy	shortness of breath	orthopnea	hypokinesia	jugular venous distention	cardiologist
chronic kidney failure	vomiting	orthopnea	hyperkalemia	oliguria	nephrologist

Figure 4.2: Symptoms and diseases data.

18 to 29	F	124	Minor	cough	fever	decreased translucency	shortness of breath	pulmonologist
70 or Older	F	123	Minor	cough	wheezing	shortness of breath	chest tightness	Allergist
70 or Older	M	122	Major	wheezing	cough	shortness of breath	chest tightness	Allergist

Figure 4.3: Data set after combining data from both 4.1. and 4.2.

After pre-processing, each patient record contains 7 features, as shown in figure 4.4:

	Age	Gender	S1	S2	S3	S4	Doctor
0	1	1	35	41	11	1	9
1	4	0	7	19	9	41	0
2	3	0	33	35	4	24	0
3	3	1	29	17	48	20	4
4	0	1	19	10	50	25	10

Figure 4.4: Final data after pre-processing.

- Patients age range: This variable describes the age range of the patient. There are 5 ranges: 0-17, 18-29, 30-49, 50-69 and 70+.
- Gender: This variable indicates the sex of the patient, male or female.
- Symptoms: A set of 4 variables that describe the manifestation of a disease. Each variable, Symptom1, Symptom2, Symptom3 and Symptom4 contains a set of clinical signs.
- Class label: The doctor's field is the class label for the data set.

Precision and accuracy for all the 11 classes for logistic regression model using One vs Rest Classifier is as shown in figure 4.5

	Precision	Accuracy
0	0.82	0.86
1	0.82	0.86
2	0.82	0.86
3	0.82	0.86
4	0.83	0.87
5	0.82	0.86
6	0.82	0.86
7	0.87	0.90
8	0.82	0.86
9	0.83	0.87
10	0.83	0.86

Figure 4.5: Precision and accuracy for LR model.

Probabilities of all 11 classes to be able to diagnose the patients for some 3 test cases for logistic regression model using One vs Rest classifier. are shown in the figure 4.6.

	Allergist	Endocrinologist	General Physician	Cardiologist	Gastroenterologist	Nephrologist	Neurologist	Pediatrician	Psychiatrist	Pulmonologist	Rheumatologist
0	0.905516	0.897717	0.921901	0.913620	0.913422	0.920392	0.897668	0.999573	0.919471	0.912018	0.912724
1	0.912265	0.935740	0.920992	0.939244	0.927671	0.903604	0.899341	0.745244	0.905352	0.933433	0.931883
2	0.881476	0.906254	0.919435	0.912892	0.907052	0.891713	0.883865	0.999990	0.905735	0.908178	0.920634

Figure 4.6: Probabilities of each type of physician.

Precision and accuracy for all the 11 classes for random forest model using One vs Rest Classifier are shown in figure 4.7

	Precision	Accuracy
0	0.84	0.86
1	0.82	0.86
2	0.82	0.86
3	0.82	0.86
4	0.82	0.86
5	0.82	0.86
6	0.82	0.86
7	0.89	0.90
8	0.82	0.86
9	0.82	0.86
10	0.82	0.86

Figure 4.7: Precision and accuracy for RF model.

Probabilities of all 11 classes to be able to diagnose the patients for some 3 test cases for random forest model using One vs Rest classifier are shown in figure 4.8

	Allergist	Endocrinologist	General Physician	Cardiologist	Gastroenterologist	Nephrologist	Neurologist	Pediatrician	Psychiatrist	Pulmonologist	Rheumatologist
0	0.931355	0.938899	0.933221	0.814891	0.944916	0.934478	0.883007	1.000000	0.932431	0.731708	0.946216
1	0.969033	0.904769	0.898911	0.966588	0.854188	0.905392	0.972800	0.638555	0.937987	1.000000	0.939697
2	0.979752	0.854084	0.855233	0.887698	0.864322	0.849877	0.882552	1.000000	0.958358	0.979421	0.905205

Figure 4.8: Probabilities of each type of physician.

Both these trained models were stored in pickle files; to be accessed while building the user interface for testing the user input data.

The basic code for scraping Practo doctor's data was written as well. This was done in python using *BeautifulSoup* and the *urllib*'s request module. This code allows us to scrape the doctors' data with respect to their field of specialization. The basic code for fetching the profile data for these 11 types of doctors was written. And the data obtained is as shown in figures 4.9 and 4.10.

The fields are as shown in figure 4.9 i.e. name, specialization, location and profile link.

The user interface constructed takes the inputs as shown in figure 4.11. And the output is produced as the names of the doctors with a hyperlink to their Practo profiles as shown in figures 4.12 and 4.13, respectively.

	Name	Specialization	Location	Link
2	Dr. Sheela Chakravarthy	Internal Medicine	bangalore	https://www.practo.com/bangalore/doctor/sheela-chakravarthy-internal-medicine?specialization=Internal Med
3	Dr. Sheetal Kamat	Internal Medicine	bangalore	https://www.practo.com/bangalore/doctor/sheetal-kamat-internal-medicine?specialization=Internal Medicine&
4	Dr. B Rajashekar	General Physician	bangalore	https://www.practo.com/bangalore/doctor/dr-rajashekhar-general-physician?specialization=General Physician&
5	Dr. Raja Selvarajan	Diabetologist	bangalore	https://www.practo.com/bangalore/doctor/dr-raja-selvarajan-diabetologist?specialization=Diabetologist&pract
6	Dr. Sharat Honnatti	general physician	bangalore	https://www.practo.com/bangalore/doctor/dr-sharat-honnatti-general-physician?specialization=&practice_id='
7	Dr. Ashok M N	General Physician	bangalore	https://www.practo.com/bangalore/doctor/dr-ashok-m-n-cardiologist?specialization=General Physician&practi
8	Dr. Tharanath S	General Physician	bangalore	https://www.practo.com/bangalore/doctor/dr-tharanath-s-general-physician?specialization=General Physician&
9	Dr. Pankaj Singhai	Internal Medicine	bangalore	https://www.practo.com/bangalore/doctor/dr-pankaj-singhai-1-internal-medicine?specialization=Internal Medi
10	Dr. Shalini Joshi	Internal Medicine	bangalore	https://www.practo.com/bangalore/doctor/shalini-joshi-internal-medicine?specialization=Internal Medicine&pr
11	Dr. Mohan Badagandi	General Physician	bangalore	https://www.practo.com/bangalore/doctor/dr-mohan-badagandi-diabetologist?specialization=General Physicia
12	Dr. Renu Saraogi	General Physician	bangalore	https://www.practo.com/bangalore/doctor/dr-renu-saraogi-general-physician?specialization=General Physician
13	Dr. Tharangini S R	General Physician	bangalore	https://www.practo.com/bangalore/doctor/dr-tharangini-s-r-general-physician?specialization=General Physicia
14	Dr. Sachin	General Physician	bangalore	https://www.practo.com/bangalore/doctor/sachin-56-general-physician?specialization=General Physician&prac
15	Dr. Sudha Menon	Internal Medicine	bangalore	https://www.practo.com/bangalore/doctor/dr-sudha-menon-internal-medicine?specialization=Internal Medicin
16	Dr. Dinesh V Kamath	General Physician	bangalore	https://www.practo.com/bangalore/doctor/dr-dinesh-v-kamath-diabetologist-1?specialization=General Physici
17	Dr. R Manjunath	Internal Medicine	bangalore	https://www.practo.com/bangalore/doctor/dr-r-manjunath-internal-medicine?specialization=Internal Medicine
18	Dr. Manohar K N	Internal Medicine	bangalore	https://www.practo.com/bangalore/doctor/dr-manohar-kn-physiotherapist?specialization=Internal Medicine&p
19	Dr. Ambanna Gowda	Internal Medicine	bangalore	https://www.practo.com/bangalore/doctor/dr-ambana-general-physician?specialization=Internal Medicine&pra
20	Dr. S G Puranik	General Physician	bangalore	https://www.practo.com/bangalore/doctor/s-g-puranik-general-physician?specialization=General Physician&pr
21	Dr. Sheetal Reddy Desai	Diabetologist	bangalore	https://www.practo.com/bangalore/doctor/dr-sheetal-reddy-desai-internal-medicine?specialization=Diabetolo
22	Dr. Jayasree Kailasam	Internal Medicine	bangalore	https://www.practo.com/bangalore/doctor/dr-jayasree-kailasam-internal-medicine?specialization=Internal Med
23	Dr. Ramesh S.	Internal Medicine	bangalore	https://www.practo.com/bangalore/doctor/dr-ramesh-s-internal-medicine?specialization=Internal Medicine&p
24	Dr. Ramesh Kumar D	Internal Medicine	bangalore	https://www.practo.com/bangalore/doctor/dr-ramesh-kumar-d-dentist?specialization=Internal Medicine&prac

Figure 4.9: Raw Practo data after initial scraping.

```
11 10 False
I recommend the doctor ::: She is my favorite doctor visited multiple times for different issues. She is very friendly and removes our worry
I recommend the doctor ::: She has very friendly and patient nature. Listened to all my problems and gave ****** treatment. I feel that much
I recommend the doctor ::: It was a very nice experience with the doctor.
I do not recommend the doctor ::: *** ******** ** ****** ********* **** ********** *** was able detect the cause .She was very polite in beh
I recommend the doctor ::: My mother had a post spine surgery follow up with Dr. Sheela and was asked to come on a particular date to the ho
I recommend the doctor ::: I have been consulting Dr. Sheela Chakravarthy from last two months. She is very **** *** approachable. She liste
I recommend the doctor :::
I recommend the doctor ::: Only one session has taken place until now and it was satisfactory. The diagnosis was executed very well and the
I recommend the doctor ::: I am writing this on behalf of my father, he was complained with low platelet count. As soon as we got to know at
I recommend the doctor ::: Dr. Sheela Chakravarthy is an excellent doctor. Her treatment was good. Staff also co operated well. Consultatior
```

Figure 4.10: Snapshot of feedback from initial scraping.

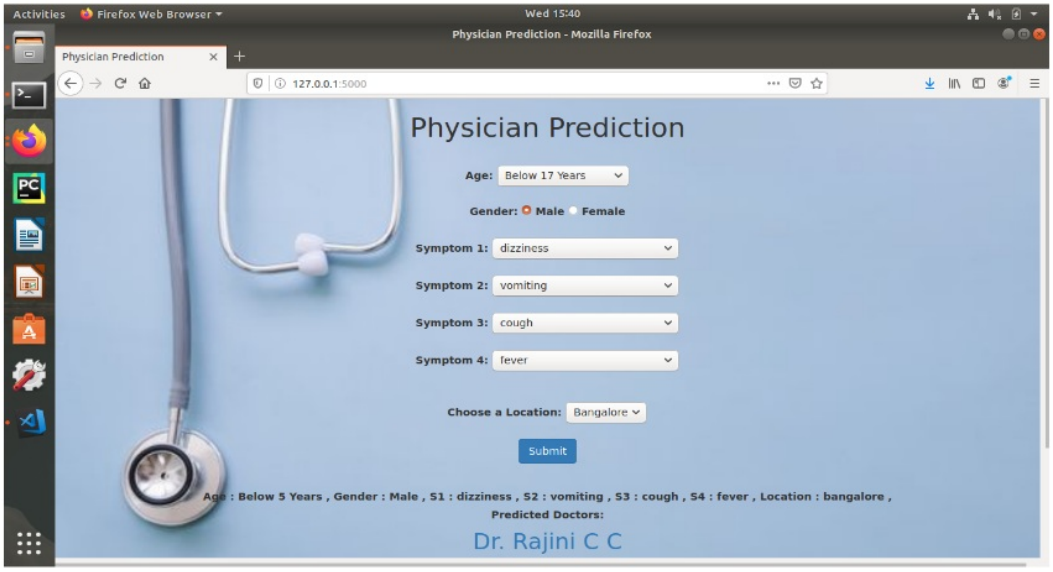

Figure 4.11: Web interface for physicians prediction.

Figure 4.12: List of recommended doctors based on the given input.

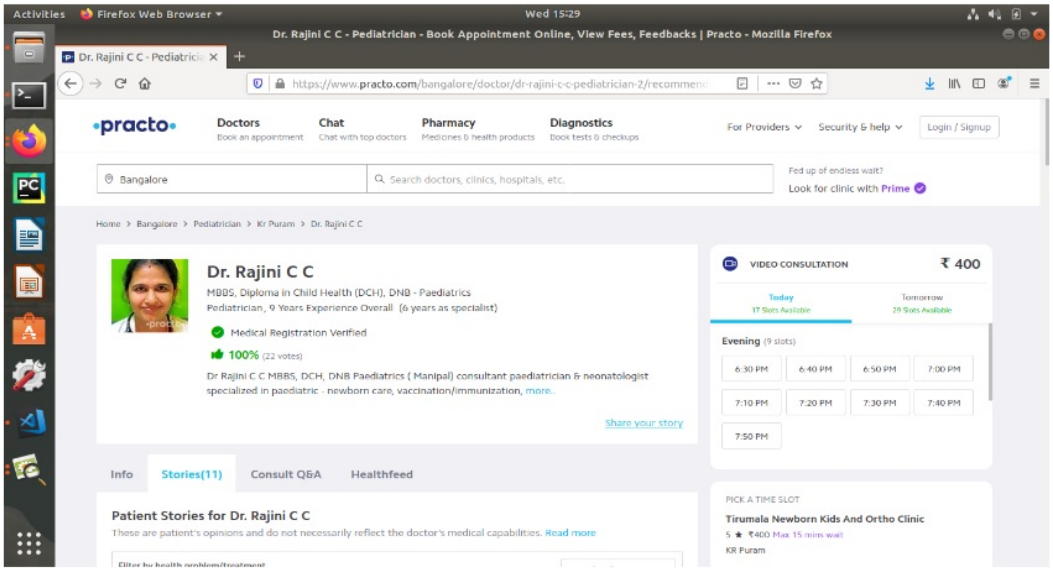

Figure 4.13: Redirection to doctor's profile on practo website.

5 Conclusion and Future Work

In this work, pre-processing of the overall data-set was done, as mentioned in the previous sections. And then, manual collection of the symptoms, diseases and type of doctors data was done. The two data sets were mapped together and were finally pre-processed. The structure for web is made for scraping, as mentioned, for collecting the required Practo data of doctors. The objective to predict the doctor's field is successfully achieved. Sentiment analysis is successfully done on the Practo comments for the respective doctors. And finally, a basic user interface is constructed to bring together both the parts of this project i.e. predicting doctor's field and performing sentiment analysis on Practo data.

The future work is to improve the user interface as well as convert this into an android application. Data from other websites along with Practo can be considered as well. More disease symptoms and type of doctors can be added.

6 Timeline of the Project

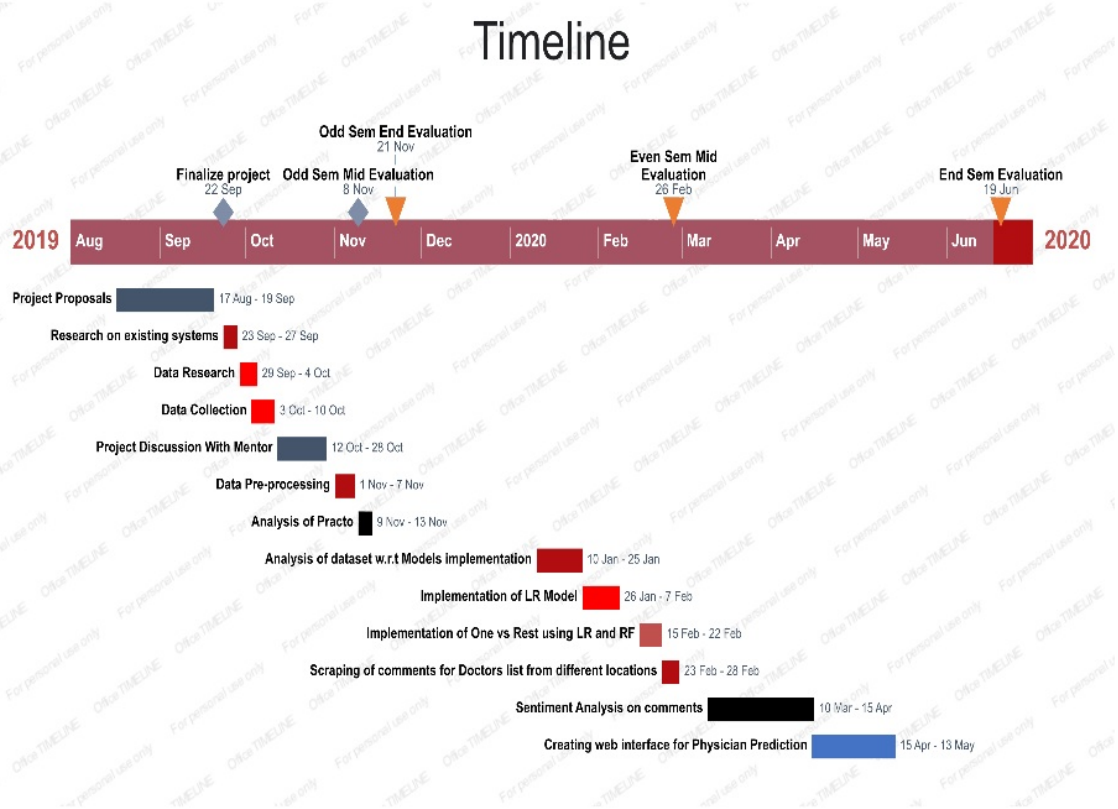

Figure 6.1: Timeline

References

[1] Tejaswita P. Vaidya ,Dhanashri Gujar and Rashmi Biyani, Disease Prediction and Doctor Recommendation System.", International Research Journal of Engineering and Technology, Volume 05, Issue 03, March 2018

[2] Ramakrishnan, Naren Keller, Benjamin Mirza, Batul Grama, Ananth Karypis, George. (2001). When being Weak is Brave: Privacy in Recommender Systems. Computing Research Repository - CORR.

[3] Disha Mahajan , Mrudula Phalak , Saniya Pathan," Prediction System for Diseases and Suggestion of Appropriate Medicines",International Journal of Innovations Advancement in Computer Science, Volume 6, Issue 11, December 2017

[4] Guo, Li Jin, Bo Yao, Cuili Yang, Haoyu Huang, Degen Wang, Fei. (2016). Which Doctor to Trust: A Recommender System for Identifying the Right Doctors. Journal of Medical Internet Research. 18. e186. 10.2196/jmir.6015.

[5] Sahoo, A.K.; Pradhan, C.; Barik, R.K.; Dubey, H. DeepReco: Deep Learning Based Health Recommender System Using Collaborative Filtering. Computation 2019.

[6] Federico Cesconi, Textblob - Natural language processing sentiment analysis explained to business people, Available:https://textblob.readthedocs.io/en/dev/quickstart.html, Visited: 28th October 2019

PREDICTION OF PHYSICIANS FOR PATIENT DIAGNOSIS

ORIGINALITY REPORT

20%	9%	12%	12%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Nfongourain Mougnutou Rémy, Tekinzang Tedondjio Martial, Tayou Djamegni Clémentin. "The prediction of good physicians for prospective diagnosis using data mining", Informatics in Medicine Unlocked, 2018 Publication	5%
2	en.wikipedia.org Internet Source	2%
3	Submitted to National Institute of Technology Karnataka Surathkal Student Paper	2%
4	www.geeksforgeeks.org Internet Source	2%
5	Abhaya Kumar Sahoo, Chittaranjan Pradhan, Rabindra Kumar Barik, Harishchandra Dubey. "DeepReco: Deep Learning Based Health Recommender System Using Collaborative Filtering", Computation, 2019 Publication	1%
	Submitted to Indiana University	

6	Student Paper	1%
7	Qiwei Han, Mengxin Ji, Inigo Martinez de Rituerto de Troya, Manas Gaur, Leid Zejnilovic. "A Hybrid Recommender System for Patient-Doctor Matchmaking in Primary Care", 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), 2018 Publication	1%
8	www.analyticsvidhya.com Internet Source	1%
9	Submitted to K Ramakrishna College of Engineering Student Paper	1%
10	Submitted to CSU, San Jose State University Student Paper	<1%
11	www.ijert.org Internet Source	<1%
12	Submitted to Colorado Technical University Online Student Paper	<1%
13	epdf.tips Internet Source	<1%
14	eprints.kfupm.edu.sa Internet Source	<1%

15	www.science.gov Internet Source	<1%
16	Submitted to Sandra Day O'Connor High School Student Paper	<1%
17	www.tandfonline.com Internet Source	<1%
18	"Intelligent Computing, Networking, and Informatics", Springer Science and Business Media LLC, 2014 Publication	<1%
19	Submitted to Asian Institute of Technology Student Paper	<1%
20	Submitted to Auckland Institute of Studies at St. Helens Student Paper	<1%
21	Submitted to Stockhom University & The Royal Institute of Technology Student Paper	<1%
22	Submitted to University of Oxford Student Paper	<1%
23	Submitted to University of Sydney Student Paper	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		