Symptom-Based Medicine Recommendations Used in Natural Language Processing

Submitted By

Hridoy Chowdhury

ID: 183-35-2605

Department of Software Engineering

Daffodil International University

Supervised By

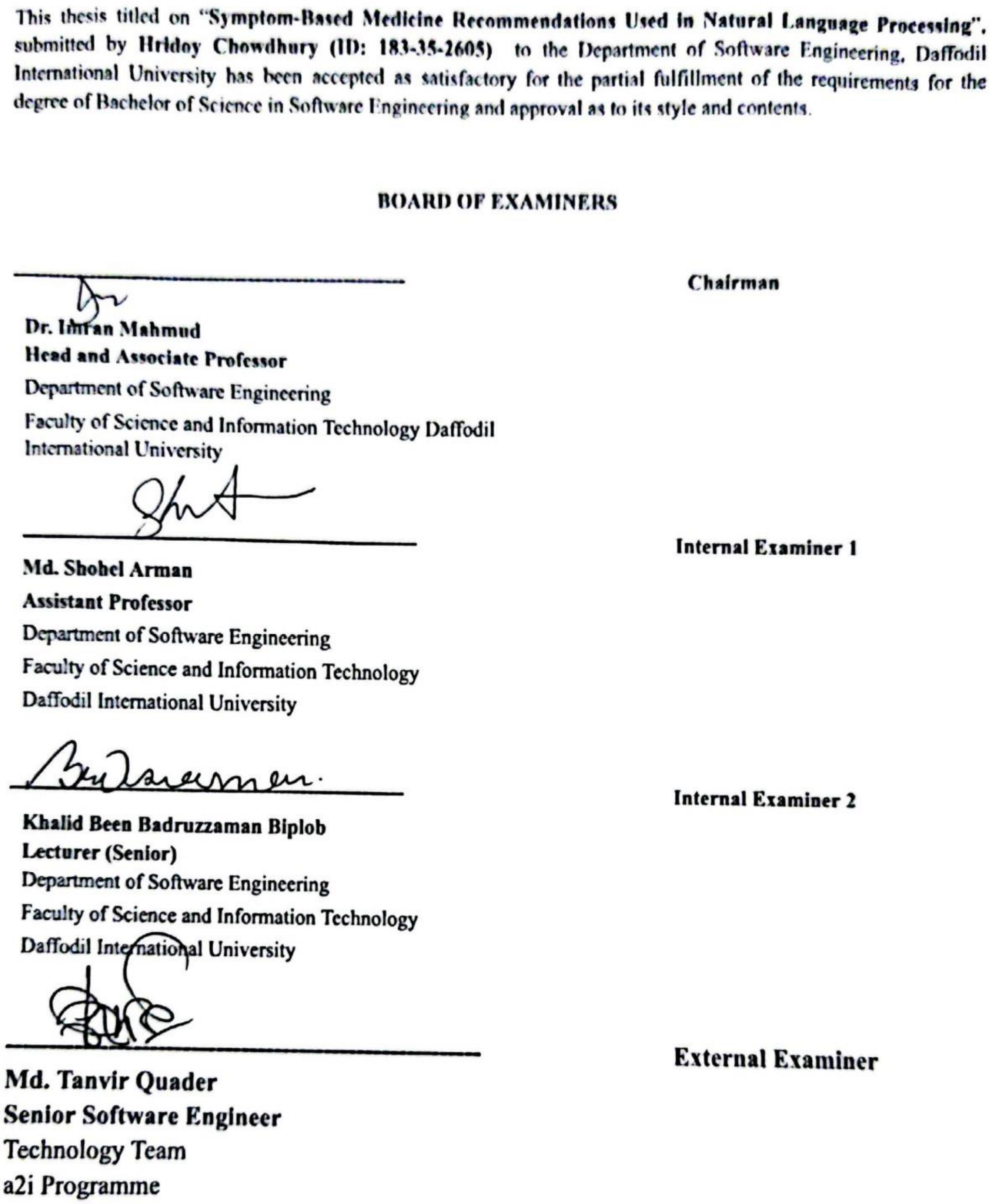
Md. Shohel Arman

Assistant Professor

Department of Software Engineering

Daffodil International University

Approval



This thesis titled on "Symptom-Based Medicine Recommendations Used in Natural Language Processing", submitted by **Hridoy Chowdhury (ID: 183-35-2605)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

Chairman

Dr. Imran Mahmud
Head and Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology Daffodil International University

Internal Examiner 1

Md. Shohel Arman
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 2

Khalid Been Badruzzaman Biplob
Lecturer (Senior)
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

External Examiner

Md. Tanvir Quader
Senior Software Engineer
Technology Team
a2i Programme

DECLARATION

I hereby declare that I have taken this project under the supervision of **Md. Shohel Arman, Assistant Professor, Department of Software Engineering, and Daffodil International University.** I also declare that this thesis doesn't have been submitted elsewhere for the award of any degree.

Submitted by:

Hridoy

……………………………………….

Hridoy Chowdhury

ID: 183-35-2605

Department of Software Engineering

Faculty of Science & Information Technology

Daffodil International University

Certified By:

………………………….

Md. Shohel Arman

Assistant Professor

Department of Software Engineering

Faculty of Science & Information Technology

Daffodil International University

ACKNOWLEDGEMENT

First, to Almighty God, I express my heartiest thanks and gratitude for His divine blessing in making it possible to successfully complete the final year thesis.

I am grateful and wish our profound indebtedness to **Md. Shohel Arman, Assistant Professor,** Department of Software Engineering, and Dhaka. Deep Knowledge & keen interest of our supervisor in the field of “Machine Learning” to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

I would like to express my heartiest gratitude to Dr. Imran Mahmud, Head In-Charge of the Software Engineering faculty for his kind help in finishing our project and also to other faculty members and the staff of the SWE department of Daffodil International University.

Finally, I must acknowledge with due respect the constant support and patients of our parents.

Table of Contents

LIST OF FIGURES

ABSTRACT

Since ancient times, viral illnesses have been at war with people. However, every living thing in the world—including tiny viruses—constantly strives for survival, according to the idea of evolution. Consequently, the burden of sickness and mortality caused by the transmission of infection by viruses to people is significant. Viruses are quickly developing on a daily basis. Although we have numerous cutting-edge methods for the detection, prevention, and treatment of infectious illnesses today, the introduction of new diseases continues to pose a severe threat to the health of the entire world's population. The new virus COVID-19 is a recent example. Because doctors were not aware of this illness at the time, many individuals perished. There are also many people who are financially strapped and unable to visit a doctor. We developed this system after studying these. Through this, we provide fundamental medical advice depending on the patient's numerous symptoms. In addition, we have developed a concept that can advance the pharmaceutical sector. We can use machine learning techniques to implement this strategy. yet, medical professionals are known for having sloppy cursive writing. About 7,000 fatalities per year in the United States are attributed to the inability to comprehend doctors' handwritten prescriptions. The issue ought to be worse because more doctors in Bangladesh and other least-developed countries write their prescriptions by hand. Because of this, both patients and pharmacists have trouble reading, and they frequently give the wrong drugs. In order to make it simpler to read prescriptions written by doctors, this study provides an offline handwritten prescription recognition system. For this study, samples of prescription' and medicine information were gathered from the Medex website and medical representatives in Noakhali city.

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

We are now in the 20th century, and people are now very advanced and innovative. But there are some places where people have not seen the touch of modernity yet.

The system I'm referring to is truly highly beneficial and useful for everyone.

One of the main uses of machine learning is the recommendation of symptoms-based medications. This technique is being used in several nations. Our primary role in it is to see the doctor, describe our problems, and receive a prescription for medication.

As a result, going to the doctor usually takes a long time and is expensive. When a system does this task, it will be completed much more quickly and precisely.

When we enter our ailment or symptoms into this system, the machine will identify them and then propose medications in line with those findings. This approach is being tested on a group of mad people by a Canadian research team. Disease Prediction using Machine Learning is the system that is used to predict diseases from the symptoms which are given by the patients or any user. The system processes the symptoms provided by the user as input and gives the output as the probability of the disease. Bangladesh has a doctor-to-population ratio of 0.304:1000, but the World Health Organization (WHO) recommends a doctor-to-population ratio of 1:1000.

In Bangladesh, doctors spend 48 seconds each visit providing basic treatment, compared to 22.5 minutes in Sweden, according to a worldwide survey. The consultation procedure includes hearing patients' problems, understanding test results, writing prescriptions, and outlining remedies. They would rather assist another patient than spend time writing a prescription. As a result of studies on illness prediction and medication prescription deformations, various diseases, and incomplete strokes present with continuous characters, ligatures, and noise make it challenging to distinguish cursive characters. Since the stroke of cursive letters varies, one of the most frequent identification issues arises when the stroke resembles the curves of some alphabet letters. Creating a model that can identify diseases and numbers in an input picture of doctors' cursive handwriting is the aim of this article. Additionally, this discovery will make it simpler for people in the medical and non-medical fields to read unreadable. Building a Deep Convolutional Recurrent Neural System is the general objective of this project. It is challenging to identify cursive characters due to deformations, various illnesses, partial strokes, continuous characters, ligatures, and noise, as well as study disease prognosis and medication recommendations. One of the most frequent issues with recognition in cursive writing is when the stroke resembles the curves of some alphabet letters. The objective of this study is to create a model that can identify the sickness and its progression.

digits in a cursive handwritten input picture by a doctor. Additionally, this research will make it simpler to read unreadable for both medical professionals and non-professionals. Making a Deep Convolutional Recurrent Neural System is the overarching objective of this project.

1.2 MOTIVATION OF THE RESEARCH

There are several benefits to using remote diagnosis systems, including cost-effectiveness, quick and accurate decision support for medical diagnoses, and the treatment and prevention of disease, illness, accidents, and other physical and mental impairments in humans.

1.3 PROBLEM STATEMENT

As we all know, giving medical advice may be challenging. In the realm of pattern recognition, illness recognition has recently emerged as one of the most fascinating and challenging study fields. In many applications, it may enhance human-machine interaction and makes a significant contribution to the enhancement of automated operations. Numerous research have concentrated on creating novel approaches and procedures that will speed up recognition while reducing processing time. There have previously been several works on this subject. Deep Convolutional Neural Networks, CNN, CRNN, and other methods are used by numerous researchers. They employed a number of methodologies, including as traditional deep learning methods and cutting-edge, recently announced OCR of NLP models and architecture. Patient illness symptoms and medication recognition methods are often the two most popular sorts. The finished writing is made available as a medication after this medicine recognition is often collected optically by a scanner in offline recognition. In the online system, the two-dimensional coordinates of succeeding sites are displayed as a function of time, and the writer's stroke order is also provided. Online techniques have been shown to be more effective than offline versions in identifying handwritten characters because of the temporal information the former provides.

From many sources, they obtained information. Others used publicly accessible data, while some researchers employed bespoke databases. In our efforts, we will leverage information gathered locally. The information gathered is distinct from that of other researchers. Our dataset includes several English and Bangla letters. Therefore, reforming those is challenging.

1.4 RESEARCH QUESTIONS

The research questions are,

- Q1:Which is a more efficient way to Symptom-based medicine recommendations?
- Q2: What type of drugs name and doses are there?
- Q3: is that possible to make a medicine recommendation tool??

1.5 RESEARCH OBJECTIVE

This study's main objective is to extract drug names from prescriptions. and collected all medical data into a dataset. Additionally, I wanted a better outcome so that the model would be more beneficial. Those of the thesis are

- To acquire more precise findings,
- to establish a low-cost,
- affordable solution,
- and to automate the process

1.6 RESEARCH SCOPE

- Sort symptoms by prescription.
- Classify various medicine names and keep them on hand for present and future usage.
- Get a better outcome than your past efforts.

CHAPTER 2

LITERATURE REVIEW

2.1 PREVIOUS LITERATURE

The basic concept was to provide medicine recommend to a Patient from their disease. recherche and the general public time and effort. Many researchers have previously studied and applied many sorts of deep learning algorithms to solve this

problem. With the help of a medical dictionary, several used recognition and medicine suggestions. Rather than designing applications, I concentrated my efforts on improving this recognition.

This work uses a variety of machine learning techniques[1] including Naive Bayes, Decision Tree, and Random Forest, to construct an illness prediction and medication recommendation system. The machine learning approach used to train the system involved mapping the numerous illness symptoms from the dataset A Computer-Based Disease Prediction and Medicine Recommendation System.

They create a universal medicine recommender system, [2]which comprises of a database system module, a data preparation module, a recommendation model module, a model assessment model, and a data visualization module, that uses data mining technologies for the medical diagnosis. Finally, because of its high accuracy, good efficiency, and scalability in this open dataset, SVM is chosen for the pharmaceutical recommendation model. We also suggested a mistake-check process that guarantees the safety and caliber of service with regard to patient safety. They want to develop a recommendation model in their subsequent work in order to further increase the model's efficacy and accuracy. standards they set.

In this study, a medical chatbot that may be used to diagnose illnesses and offer appropriate treatments is explained.[3] A chatbot can perform medical duties. In its role as a user application, the chatbot. By recognizing the symptoms that patients have described, making an accurate diagnosis, and offering the best treatment options for the ailment, a smart medical chatbot can be helpful to patients. People hardly regularly visit hospitals for check-ups in today's busy world. In these circumstances, chatbots are crucial since they quickly and easily offer diagnostic support. their next work is The chatbot's job may occasionally be out of bounds, and a user may need to see a doctor before completing any health-related tests.

There are three analyses of tongue comprehension or the thoroughness of finding the key linguistic linkages for analyzing the issue in the subject of sentences.[4] After then, the representation of the writings is complete. Knowing what a word means is essential for semantic comprehension. their future work is By inventing and creating a tool like a medical chatbot for health utilizing machine learning algorithms and NLP, the system's long-term goal is to create an alternative approach for these conventional types of hospital visits and appointments for doctor consultations to obtain diagnoses (Natural Processing language).

This overall situation prevents the person from having time to visit a doctor and from staying informed about his or her health. [5]Additionally, it frequently occurs that a person has free time but that their primary care physician is unavailable owing to a variety of obligations. The availability of excellent doctors is limited in rural regions, and the village people must go extremely far for treatment, therefore many villagers avoid seeing doctors even if their health condition is not serious in order to bridge this gap and offer communication between patient and doctor, even if they are situated at two different places and distant from each other but they may connect and the patient can obtain consultation from doctors are also flocking toward the city.

For the best patient care, doctors must apply the findings of big clinical intervention studies to the care of specific patients.[6] The averages of treating a diverse set of patients are reported in trials as relative risks or hazard ratios now. Trials only give one estimate of impact, which is an average group-level estimate that implicitly assumes each participant has average risk and average response to treatment. and their foreseeable future. The ability to identify the patients who will benefit from therapy the most, decrease the number of wasteful treatments, and save healthcare expenditures, this will aid in improving individual patient management.

The task of categorizing input patterns and allocating them to the proper entities falls on illness recognition systems. Entities vary from one system to the next. Systems for text recognition that use character categorization are known as character recognition systems. Character recognition systems take into account the kind of characters as a key component. These are programs that can read written, typewritten or printed characters.[7]

Recognizing a drug is more crucial, but it can be difficult and confusing when working with mixed cursive and cursive writing.[8] The challenge of identifying characters is increased by the variety of styles. Character recognition is essential for data processing and pattern recognition. The basic objective of character recognition is to transform readable characters into machine-processable configurations. CR converts the text in a picture into machine-readable text and provides results in ASCII or UNICODE formats.

Software known as optical character recognition (OCR) makes use of a more intricate matrix strategy known as pattern matching. It is a technique for turning readable characters from optically scanned and digitalized text into characters that computers can read.[9]

OCR is very helpful and practical when data that can be understood by both people and machines is needed and multiple data sources cannot be assumed.[10] It is possible to make text recognition easier by carefully choosing the optical character recognition system's capabilities through script identification.

A data modeling technique that can capture and depict complex input/output interactions is the neural network algorithm.[11] The creation of an artificial system that can carry out cognitive functions like to those carried out by the human brain is the aim of neural network technology.

Recurrent Neural Networks (RNNs) are a form of neural network that are employed to produce sequences in a number of different domains, such as music, text, and motion capture data.[9] RNNs may be trained to create sequences by anticipating each action and going one step at a time. It can simulate complex structures and have numerous layers, making it particularly effective in sequence modeling. It may store representations of earlier input events in the form of activations utilizing recurrent connections.[12]

An RNN architecture called Long Short-Term Memory (LSTM) was created to store data more quickly than standard RNNs.[13] LSTM has produced cutting-edge outcomes for a variety of sequence processing applications, including voice and handwriting recognition.

To solve this issue, Peilun Wu and colleagues introduced an integrated multi-classifier approach based on CNN and KNN.[14] They utilized three single classifiers and the suggested integrated multi-classifier identification approach for experiments on handwritten Chinese medicinal prescription recognition. 13 sets of handwritten Chinese medicine prescriptions with 112 medication names and dosages written in Chinese characters, English letters, and numbers served as the test participants.

In order to identify the pharmaceutical name from the medication website, Pritam S. Dhande and Reena Kharat employ the Convex hull approach for feature extraction and SVM for classification.[15] The convex hull method is used for feature extraction, while SVM is used for classification and recognition. On a sample of 50 distinct diseases, horizontal and vertical projection approaches produced word and text segmentation accuracy of 92% and 95%, respectively. Anita Pal and Dayashankar Singh use a Convolutional Neural Network alone to detect handwritten English characters as a novel strategy. ConvNet was utilized for pre-processing since it is a lot more effective than traditional classification methods. To solve this issue, Peilun Wu and colleagues introduced an integrated multi-classifier method based on CNN and KNN.

Word boundaries can frequently be jumbled up, resulting in completely distinct sentence comprehensions. The language's syntax helps us determine how words are being joined to create more complex meanings at the next level. A simple sentence is made of a subject and a predicate.[16] Nouns, pronouns, adjectives, verbs, adverbs, prepositions, conjunctions, and interjections are the components of speech used in English sentences. their upcoming work As it deals with new technology hurdles and pressure from the market to develop more user-friendly solutions, NLP's future will be redefined. The effect of the market is causing current NLP-based enterprises to compete more fiercely. Additionally, it is encouraging NLP to develop using open-source software. If the community for Natural Language Processing adopts open-source development, NLP systems will become less proprietary and consequently less expensive.

A conceptual model that describes the system's operations and structure are called system architecture. The connections between the system's parts, which describe how they interact to operate the whole system, are also included[17]. Within the block diagram In the illustration above, the user asks inquiries in the chatbot since the chatbot's front end is made up of a user interface that accepts user input requests for patient inquiries. To build and create a tool like a medical chatbot for health utilizing machine learning algorithms and NLP to provide an alternative approach for this conventional sort of hospital visits and appointments for doctor consultations to obtain diagnoses (Natural Processing language).

Different segmentation methods and degrees of segmentation were covered by Louloudisa et al.[18] The Hough transform method is utilized for text-line and word segmentation of associated cursive letters.In order to attain great efficiency in off-line recognition systems, neural networks have developed as rapid and trustworthy algorithms for categorization.[19] Recurrent neural networks (RNNs) are a form of neural network that are used to produce sequences in a range of domains, including music, text, and motion capture data. RNNs may be taught to create sequences by anticipating each action and moving one step at a time.Machine learning (ML) is a term for a collection of techniques that automatically spot patterns in data and then use those patterns to forecast future data or support decision-making in the face of uncertainty (1). AI is a subset of machine learning (AI) [20] .There are generally three methods to AI: symbolism (rule-based, like IBM Watson), connectionism (network and connection-based, like deep learning or artificial neural networks), and Bayesian (based on the Bayesian theorem). The most defining feature of ML is that decisions are made with the fewest possible human involvement and are data-driven. A prediction may be made when fresh data is introduced after the algorithm has learned from evaluating training data. A difficult challenge has always been finding a new drug. A new drug requires years of research and development. The overall number of candidate compounds was expected to be between 1060 and 10200 in the initial stages of drug development for any illness [20]. This is the reason it takes so long to identify the appropriate ingredients for producing a new drug.

The medical industry did not previously have any resources for using machine learning techniques to investigate potential treatments.

Since the advent of artificial intelligence (AI), the field of computer applications has seen a substantial increase. Artificial intelligence is nothing more than a computer-processed imitation of human intellect. When all of its citizens are healthy, a society is affluent. If one wants to be happy, maintaining one's health is crucial. A healthy body is necessary for a healthy mind, and it improves people's performance. People today are less concerned with their health. They neglect to manage their health and are less conscious of their health state due to their busy lives. People seem to place little value on their health, and they find it time-consuming to visit hospitals for check-ups, according to the most recent news from TOI [21]. The health has no place in The The hectic pace of life. The main goal of the work is to close the communication gap between the client and organizations that provide wellness services by responding quickly to the client's inquiries. In order to fully determine the key linguistic linkages for analyzing the topic in the subject of sentences, there are three analyses of tongue comprehension. The representation of the writings is complete at this stage.[22] Semantic understanding draws on word knowledge to infer that a chatbot is a substance that replicates human banter in a satisfying context when used in conjunction with a book or phonetic language (NLP).In recent years, the pharmaceutical industry has been regarded as one of the most lucrative. Because of the steadily increasing number of pharmacies, we can see that it is desirable. Large and small, independent, or a part of major pharmacy retail chains, they are opening up more and more often [23]. Today, every mall has a pharmacy kiosk or shop, often many, and they all have various looks and pricing that are frequently significantly different. Only a few decades ago, pharmacies were considered to be fairly unpleasant places.Only a third of pharmacy clients, according to statistics, fully understand what the medication they came for. 10% of them reportedly learned about the medication via an advertising, 40% received a doctor's prescription, and the remaining 80% relied on a pharmacist or friend's advice [4]. Therefore, a typical pharmacy customer requires information resources to support his decision to buy medication in a quick and impartial manner, taking also into account customer location and his means.[24] Many times, people only require medicine when they truly need it. They frequently are unable to access the drugstore network website in its whole. Such websites often don't offer the option of advising customers to buy from another network or a nearby drugstore. Knowledge discovery in data bases is a technique that includes data mining. It enables us to detect relationships between objects' attributes, details of which are contained in databases, and their interactions, as well as the essence of hidden connections in the data. It also helps us to draw attention to patterns present in a given collection of data. The broad practical and commercial usage of intellectual analysis systems attests to the pressing nature of the data exploration and processing challenge [25].The majority of hospitals are now overloaded and lack efficient patient queue management. Because each patient may need different phases or procedures, such as a check-up, various tests, such as a sugar level or blood test, X-rays or a CT scan, minor surgeries, throughout treatment,[26] patient queue management and wait time forecast constitute a difficult and intricate task. In this study, we refer to each of these phases or procedures as a treatment task or task. Time prediction and advice are extremely difficult since each patient's time needs for each therapy job may differ.Software that can have natural language conversations with people is known as a chatbot or conversational agent. The modeling of dialogue is one of the key challenges in artificial intelligence and natural language processing. Creating a successful chatbot has proven to be the most difficult task in artificial intelligence to yet. Although chatbots are capable of many different functions, their main job is to comprehend human speech and reply to it effectively.[27] In the past, chatbot architectures were built using straightforward statistical techniques or manually written templates and rules. End-to-end neural networks have replaced these models as learning capabilities have gotten better starting about 2015.A machine is given the utmost ability to replicate human thought and behavior thanks to artificial intelligence.Computer programs known as chatbots engage with users in natural language. The goal of this technology, which first emerged in the 1960s, was to determine whether chatbot systems could deceive users into thinking they were dealing with actual people. Chatbot platforms, however, are not just designed to amuse users and simulate human speech. Because chatbots primarily rely on artificial intelligence, we have chosen to contribute to the field of health informatics utilizing this technology[28].Aiming to give users precise information that is easy to understand, human computer interaction (HCI) is a new field. Users often utilize Google, Yahoo, and other information retrieval systems to find information on the internet. These systems' information retrieval output includes documents or connections to other web sites or documents. These information retrieval methods do not always provide students with the knowledge they need to solve their problems, and as a result, students' learning capacities are not improved. Such issues give rise to the demand for natural language dialog systems that enable students to naturally express their domain-specific issues and to quickly and satisfactorily get responses.[29]When artificial intelligence (AI) technologies were first put into use, information technology and communication were still developing. Systems like decision support systems, robots, natural language processing, expert systems, etc. are getting closer to mimicking human actions. Even in the disciplines of artificial intelligence, hybrid and adaptive approaches can be used to create more sophisticated techniques.In addition, there are currently hybrids of natural language and intelligent systems that can comprehend human natural language. These systems have the ability to educate themselves and update their expertise by reading all electronics-related publications that have ever been published online. Users of the system can ask questions, just as they would normally ask another person. Internet answering-engines are a common name for these systems.[30]

2.2 CONCLUSION

Using elegance, clustering, sequential pattern mining, affiliation rule mining, and analysis, information mining supports a variety of unique methodologies for knowledge discovery and prediction. In order to deliver the greatest services in a way that resonates with contemporary clients, the latest advancements in artificial intelligence and new ways of thinking have the potential to completely transform the customer experience. One well-known method of decision explanation is through using examples. In contemporary software teams, code reviews are one of the most cooperative procedures and rely heavily on communication between the reviewer and developer. and The neural network algorithm is a data modeling tool that can capture and depict complex input/output interactions.

CHAPTER 3

RESEARCH METHODOLOGY

The study uses the five (5) steps of the following methodology, which is illustrated in Fig no :1.there are data collection, data cleaning, Data pre-processing, lemmatization, and topic modeling algorithm.

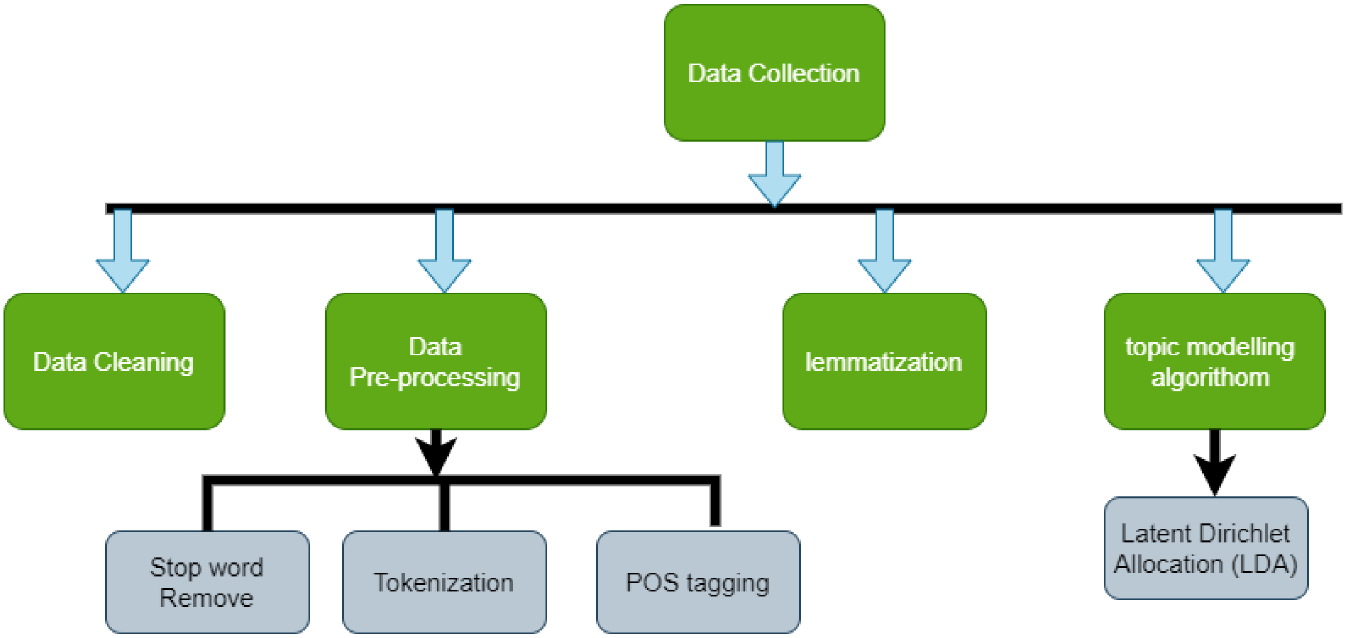
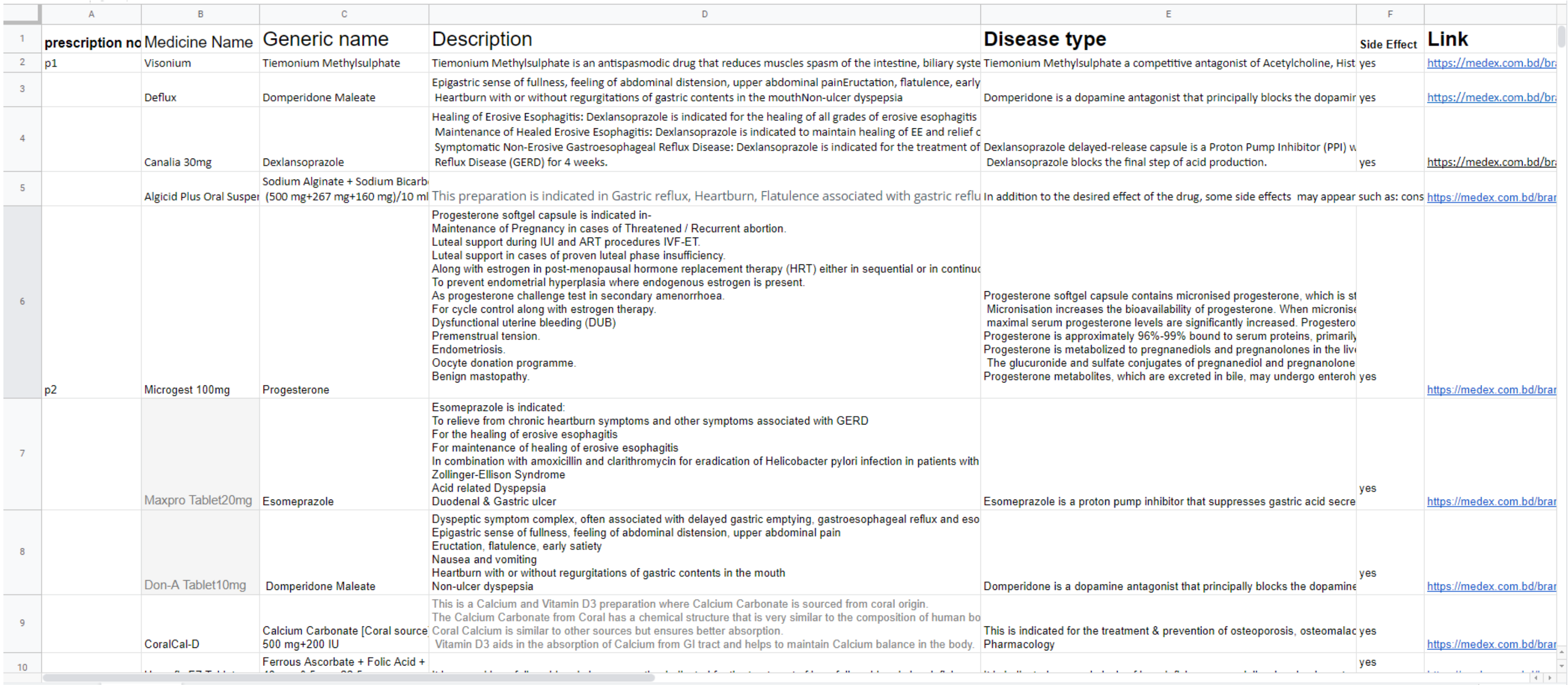


Figure 1: Methodology Phases

3.1 DATA COLLECTION

Data collection is the act of acquiring and analyzing information on relevant variables in a predetermined, methodical way so that one may respond to specified research questions, test hypotheses, and assess results. Many such data are available on the internet but they are from foreign countries. That's why I am working with the data of Bangladesh, I have taken pictures from the prescriptions of different patients and then I have created a dataset by collecting all the drug names, drug details, actions, side effects of the patient from the Medex website.



	A	B	C	D	E	F	
1	prescription no	Medicine Name	Generic name	Description	Disease type	Side Effect	Link
2	p1	Visonium	Tiemonium Methylsulphate	Tiemonium Methylsulphate is an antispasmodic drug that reduces muscles spasm of the intestine, biliary syste	Tiemonium Methylsulphate a competitive antagonist of Acetylcholine, Hist	yes	https://medex.com.bd/br
3		Deflux	Domperidone Maleate	Epigastric sense of fullness, feeling of abdominal distension, upper abdominal painEructation, flatulence, early Heartburn with or without regurgitations of gastric contents in the mouthNon-ulcer dyspepsia	Domperidone is a dopamine antagonist that principally blocks the dopamir	yes	https://medex.com.bd/br
4		Canalia 30mg	Dexlansoprazole	Healing of Erosive Esophagitis: Dexlansoprazole is indicated for the healing of all grades of erosive esophagitis Maintenance of Healed Erosive Esophagitis: Dexlansoprazole is indicated to maintain healing of EE and relief c Symptomatic Non-Erosive Gastroesophageal Reflux Disease: Dexlansoprazole is indicated for the treatment of Reflux Disease (GERD) for 4 weeks.	Dexlansoprazole delayed-release capsule is a Proton Pump Inhibitor (PPI) w Dexlansoprazole blocks the final step of acid production.	yes	https://medex.com.bd/br
5		Algicid Plus Oral Susper	Sodium Alginate + Sodium Bicarb (500 mg+267 mg+160 mg)/10 ml	This preparation is indicated in Gastric reflux, Heartburn, Flatulence associated with gastric reflu	In addition to the desired effect of the drug, some side effects may appear such as: cons		https://medex.com.bd/brar
6	p2	Microgest 100mg	Progesterone	Progesterone softgel capsule is indicated in- Maintenance of Pregnancy in cases of Threatened / Recurrent abortion. Luteal support during IUI and ART procedures IVF-ET. Luteal support in cases of proven luteal phase insufficiency. Along with estrogen in post-menopausal hormone replacement therapy (HRT) either in sequential or in continuc To prevent endometrial hyperplasia where endogenous estrogen is present. As progesterone challenge test in secondary amenorrhoea. For cycle control along with estrogen therapy. Dysfunctional uterine bleeding (DUB) Premenstrual tension. Endometriosis. Oocyte donation programme. Benign mastopathy.	Progesterone softgel capsule contains micronised progesterone, which is st Micronisation increases the bioavailability of progesterone. When micronise maximal serum progesterone levels are significantly increased. Progestero Progesterone is approximately 96%-99% bound to serum proteins, primarily Progesterone is metabolized to pregnanediols and pregnanolones in the liv The glucuronide and sulfate conjugates of pregnanediol and pregnanolone Progesterone metabolites, which are excreted in bile, may undergo enteroh	yes	https://medex.com.bd/brar
7		Maxpro Tablet20mg	Esomeprazole	Esomeprazole is indicated: To relieve from chronic heartburn symptoms and other symptoms associated with GERD For the healing of erosive esophagitis For maintenance of healing of erosive esophagitis In combination with amoxicillin and clarithromycin for eradication of Helicobacter pylori infection in patients with Zollinger-Ellison Syndrome Acid related Dyspepsia Duodenal & Gastric ulcer	Esomeprazole is a proton pump inhibitor that suppresses gastric acid secre	yes	https://medex.com.bd/brar
8		Don-A Tablet10mg	Domperidone Maleate	Dyspeptic symptom complex, often associated with delayed gastric emptying, gastroesophageal reflux and eso Epigastric sense of fullness, feeling of abdominal distension, upper abdominal pain Eructation, flatulence, early satiety Nausea and vomiting Heartburn with or without regurgitations of gastric contents in the mouth Non-ulcer dyspepsia	Domperidone is a dopamine antagonist that principally blocks the dopamine	yes	https://medex.com.bd/brar
9		CoralCal-D	Calcium Carbonate [Coral source] 500 mg+200 IU	This is a Calcium and Vitamin D3 preparation where Calcium Carbonate is sourced from coral origin. The Calcium Carbonate from Coral has a chemical structure that is very similar to the composition of human bo Coral Calcium is similar to other sources but ensures better absorption. Vitamin D3 aids in the absorption of Calcium from GI tract and helps to maintain Calcium balance in the body.	This is indicated for the treatment & prevention of osteoporosis, osteomalac Pharmacology	yes	https://medex.com.bd/brar
10			Ferrous Ascorbate + Folic Acid +			yes	

Figure 2: Sample of the dataset

3.2 DATA CLEANING

Data cleaning is the process of correcting or deleting inaccurate, corrupted, improperly formatted, duplicate, or insufficient data from a dataset. There are several potential for data duplication or labeling errors when merging different data sources. Even though they may appear to be right, bad data makes outcomes and algorithms untrustworthy. The specific procedures in the data cleaning process cannot be prescribed in a single, universal fashion since they differ from dataset to dataset. But in order to ensure that you are performing data cleaning in the proper manner each time, it is essential to create a template for your procedure.

One of the crucial components of machine learning is data cleaning. It is crucial to the process of creating a model.We have done a lot of work in data cleaning. As we have removed puncution. Let us get a clean text and Don't get any garbage. Then we get the data to lowercase so that we can easily understand why we did it. Then we have removed the white space so that the words are not too blank. Then I made the data from sentence to word.

3.3 DATA PRE-PROCESSING

An essential element in the NLP process is data pre-processing. Pre-processing uses methods to change unstructured data into something that is easier to grasp. I will do 3 tasks in data pre-processing. named for stopword removal, tokenization, and course tagging.

3.3.1 STOPWORDS REMOVE

The most frequent words in any natural language are stopwords. These stopwords might not significantly contribute to the meaning of the content when used for text data analysis and NLP model construction. The most frequent terms in a text are often "the," "is," "in," "for," "where," "when," "to," and "at,"

3.3.2 TOKENIZATION

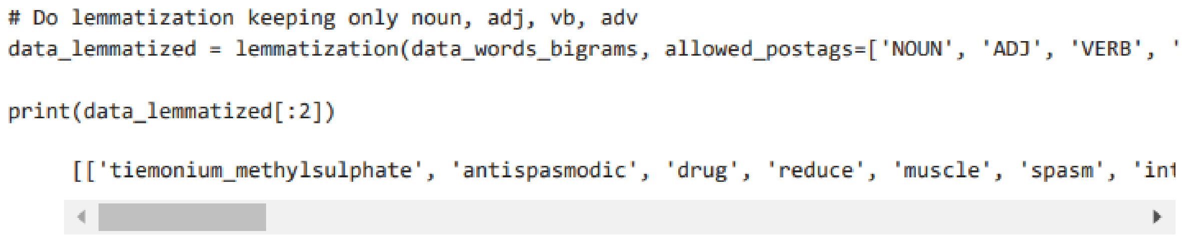
In natural language processing, tokenization is used to break down phrases and paragraphs into simpler language-assignable parts. Gathering the data (a sentence). In my work i used this model

3.3.3 POS TAGGING

Part-of-speech (POS) According to the definition of the word and its context, tagging is a common Natural Language Processing technique that involves classifying words in a text (corpus) in accordance with a certain component of speech.

3.4 LEMMATIZATION

Part-of-speech (POS) The practice of categorizing words in a text (corpus) in accordance with a certain part of speech, depending on the word's definition and context, is known as tagging in natural language processing. I am using 4 perspectives here. There are Noun, adj, verb, and adverbs.



```
# Do lemmatization keeping only noun, adj, vb, adv
data_lemmatized = lemmatization(data_words_bigrams, allowed_postags=['NOUN', 'ADJ', 'VERB', '

print(data_lemmatized[:2])

    [['tiemonium_methylsulphate', 'antispasmodic', 'drug', 'reduce', 'muscle', 'spasm', 'in
```

Figure 3: lemmatized model

3.5 TOPIC MODELING

The process of removing necessary features from a collection of words is known as topic modeling. This is significant because NLP treats each word in the corpus as a feature. In order to avoid spending time combing through all of the words in the data, feature reduction lets us concentrate on the important information. since this data is unsupervised data and unlevel data that's why i used a topic modeling algorithm. In this case, I use LDA in the topic model. And I found 10 topics.

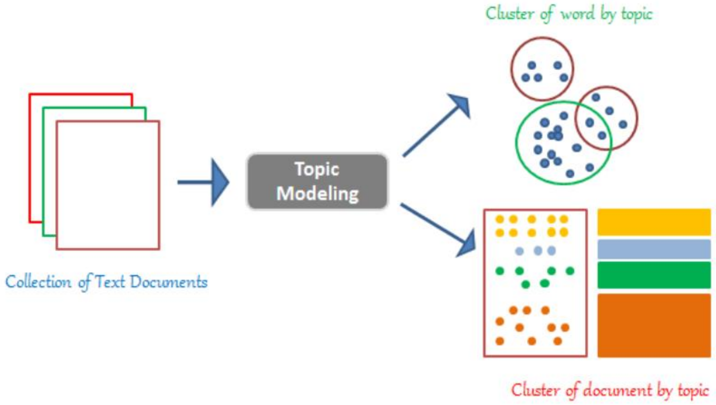


Figure 4: Topic model Diagram

3.5.1 Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation (LDA), a generative statistical model used in natural language processing, explains a collection of observations through unseen groups, with each group explaining why certain portions of the data are similar.

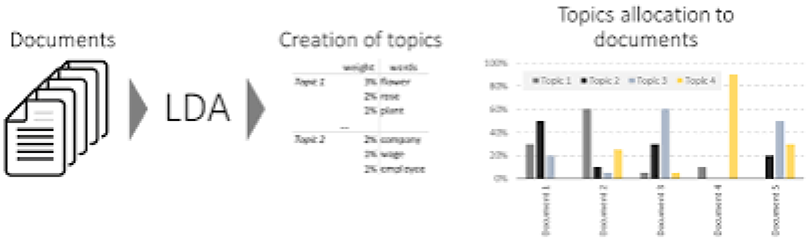


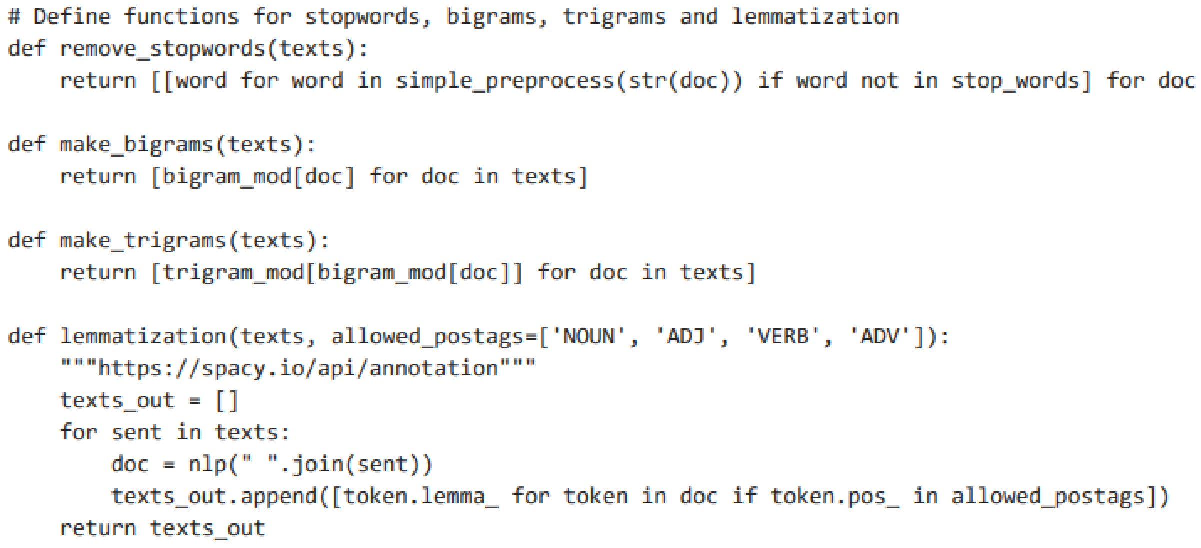
Figure 5: LDA Diagram

CHAPTER 4

RESULT AND DISCUSSION

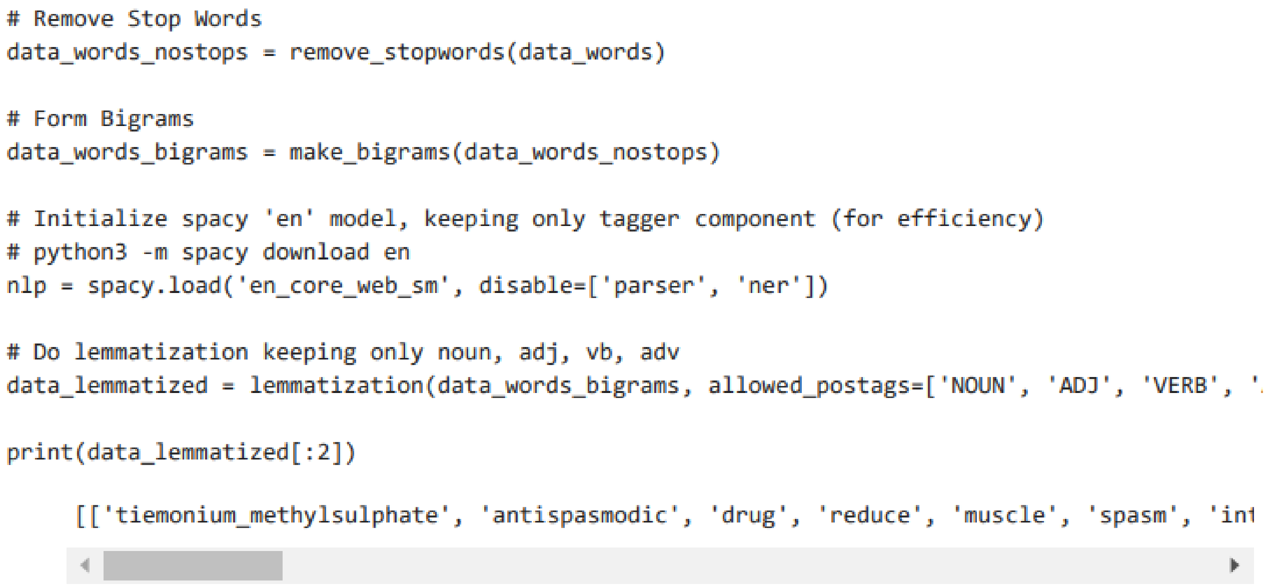
4.1 INTRODUCTION

First I cleaned the data from the dataset. There we have removed punctuation, brought the lower case, and removed white space so that there is no empty room. After that, I am doing tokenization to make words from data sentences. Because of that the words are smaller then I am removing stop word. After course tagging. Then I made trigram from bigram Then I used topic modeling algorithm er LDA. After that, our topic is created. Which topic will be selected, which is the best topic and has been finalized, that is topic number 38. In this, we can bring the best features.



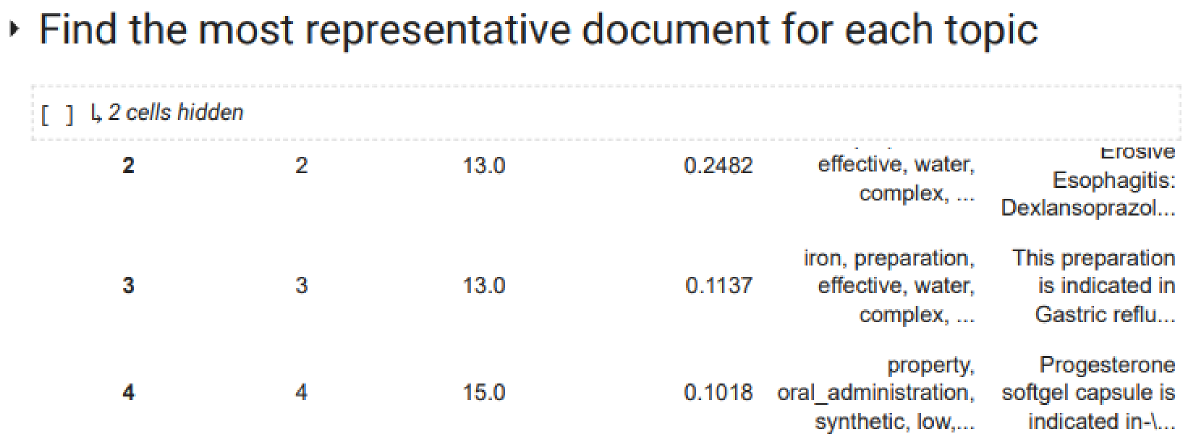
```
# Define functions for stopwords, bigrams, trigrams and lemmatization
def remove_stopwords(texts):
    return [[word for word in simple_preprocess(str(doc)) if word not in stop_words] for doc

def make_bigrams(texts):
    return [bigram_mod[doc] for doc in texts]

def make_trigrams(texts):
    return [trigram_mod[bigram_mod[doc]] for doc in texts]

def lemmatization(texts, allowed_postags=['NOUN', 'ADJ', 'VERB', 'ADV']):
    """https://spacy.io/api/annotation"""
    texts_out = []
    for sent in texts:
        doc = nlp(" ".join(sent))
        texts_out.append([token.lemma_ for token in doc if token.pos_ in allowed_postags])
    return texts_out
```

Figure 6: Bigrams, trigrams



```
# Remove Stop Words
data_words_nostops = remove_stopwords(data_words)

# Form Bigrams
data_words_bigrams = make_bigrams(data_words_nostops)

# Initialize spacy 'en' model, keeping only tagger component (for efficiency)
# python3 -m spacy download en
nlp = spacy.load('en_core_web_sm', disable=['parser', 'ner'])

# Do lemmatization keeping only noun, adj, vb, adv
data_lemmatized = lemmatization(data_words_bigrams, allowed_postags=['NOUN', 'ADJ', 'VERB', '

print(data_lemmatized[:2])

    [['tiemonium_methylsulphate', 'antispasmodic', 'drug', 'reduce', 'muscle', 'spasm', 'in
```

Figure 7: Stop Word

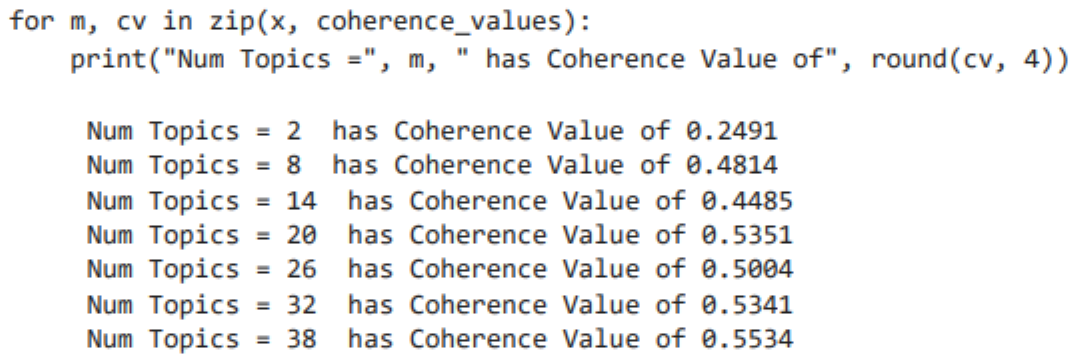


Find the most representative document for each topic

[ ] ↳ 2 cells hidden

2	2	13.0	0.2482	effective, water, complex, ...	Erosive Esophagitis: Dexlansoprazol...
3	3	13.0	0.1137	iron, preparation, effective, water, complex, ...	This preparation is indicated in Gastric reflu...
4	4	15.0	0.1018	property, oral_administration, synthetic, low,...	Progesterone softgel capsule is indicated in-\...

Figure 8: Document for each topic



```
for m, cv in zip(x, coherence_values):
    print("Num Topics =", m, " has Coherence Value of", round(cv, 4))

     Num Topics = 2  has Coherence Value of 0.2491
     Num Topics = 8  has Coherence Value of 0.4814
     Num Topics = 14  has Coherence Value of 0.4485
     Num Topics = 20  has Coherence Value of 0.5351
     Num Topics = 26  has Coherence Value of 0.5004
     Num Topics = 32  has Coherence Value of 0.5341
     Num Topics = 38  has Coherence Value of 0.5534
```

Figure 9: Finding Topic

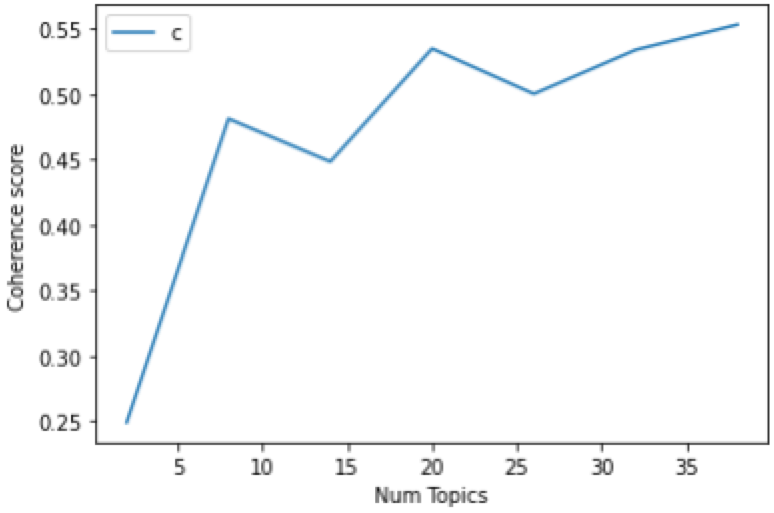
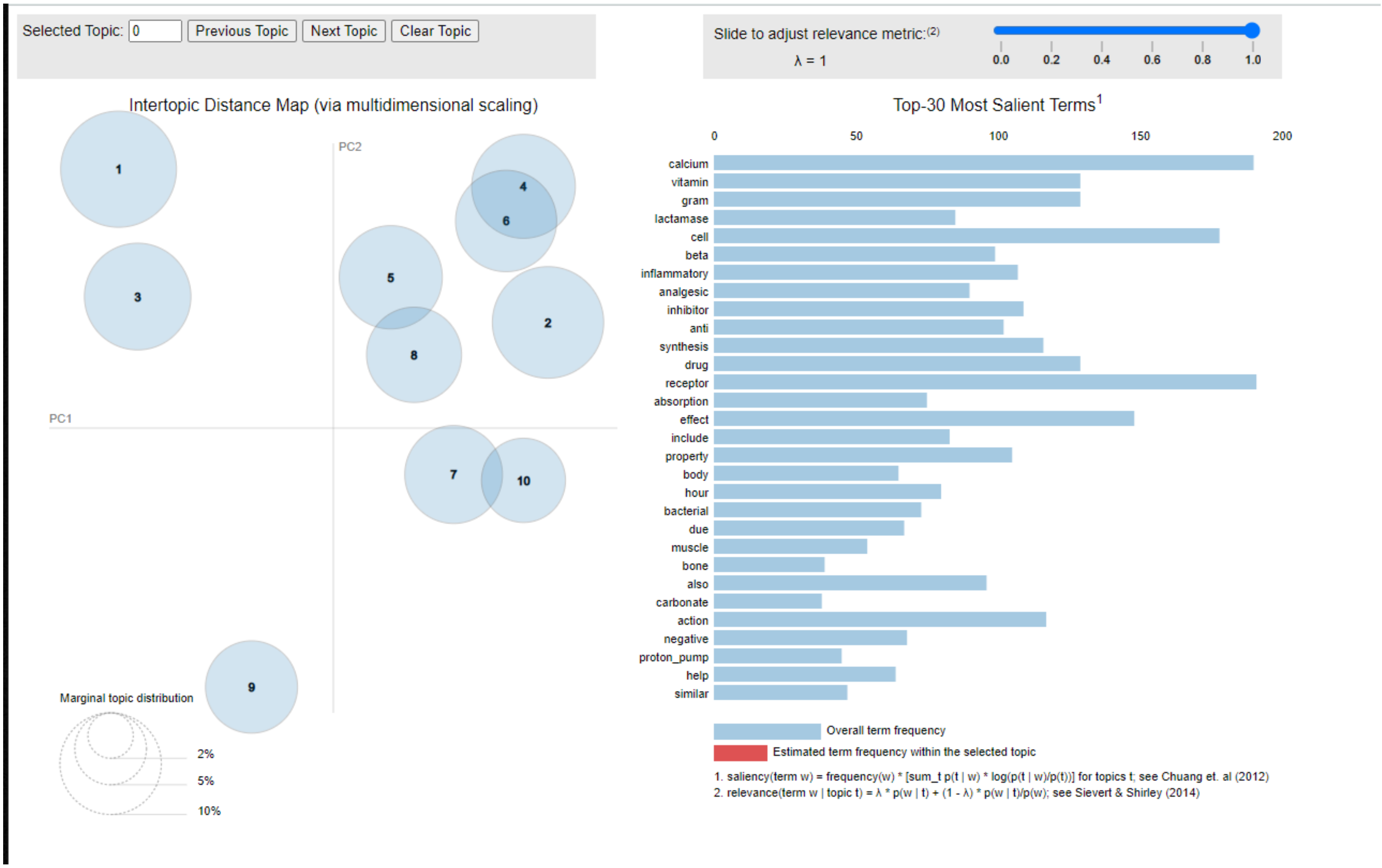


Figure 10: Topic modeling graph

4.2 RESULT DISCUSSION

When topic find out is done using LDA, for that we take the number of topics as 10 and the number of words as 10 are clustered like this. Created a dashboard for visualization. How to find valuable words, and topic features in any cluster has been shown.



Selected Topic: 0

Slide to adjust relevance metric:(2)

λ = 1

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Figure 11: Topic model result

CHAPTER 5

CONCLUSION AND LIMITATIONS

Symptom-based medicine recommendation is actually a very complex and challenging task. Every year many people die due to lack of treatment. so to solve this problem proposed a methodology for directing symptoms and recommending medicine. About 500 prescriptions and 650 drug names have been started. Among them, the first phase of feature extraction work has been done, topic modeling algorithm has been used, from there we get a value which is, the value of topic number 38 is the highest, the coherence value of which is 0.5534. These results are derived from the model. One of the well-known methods of decision explanation is by using examples. In contemporary software teams, code reviews are one of the most collaborative procedures and rely heavily on communication between the reviewer and developer. We created the EDRE bot in partnership with a business partner to lower the barriers to communication, make it easier to provide feedback, and shorten review times. In order to clarify a murky code review, EDRE uses an example. The two major components of EDRE are I finding unclear code reviews using text characteristics and I collecting a prioritized list of pertinent instances using analogical reasoning.

Limitations: For future work, I will start working on the features I got. I will classify the disease with these features, then recommend the medicine. After this, another thing can be done in the future, which is online-based software. And this project could be implemented as a mobile application accessible to a variety of use