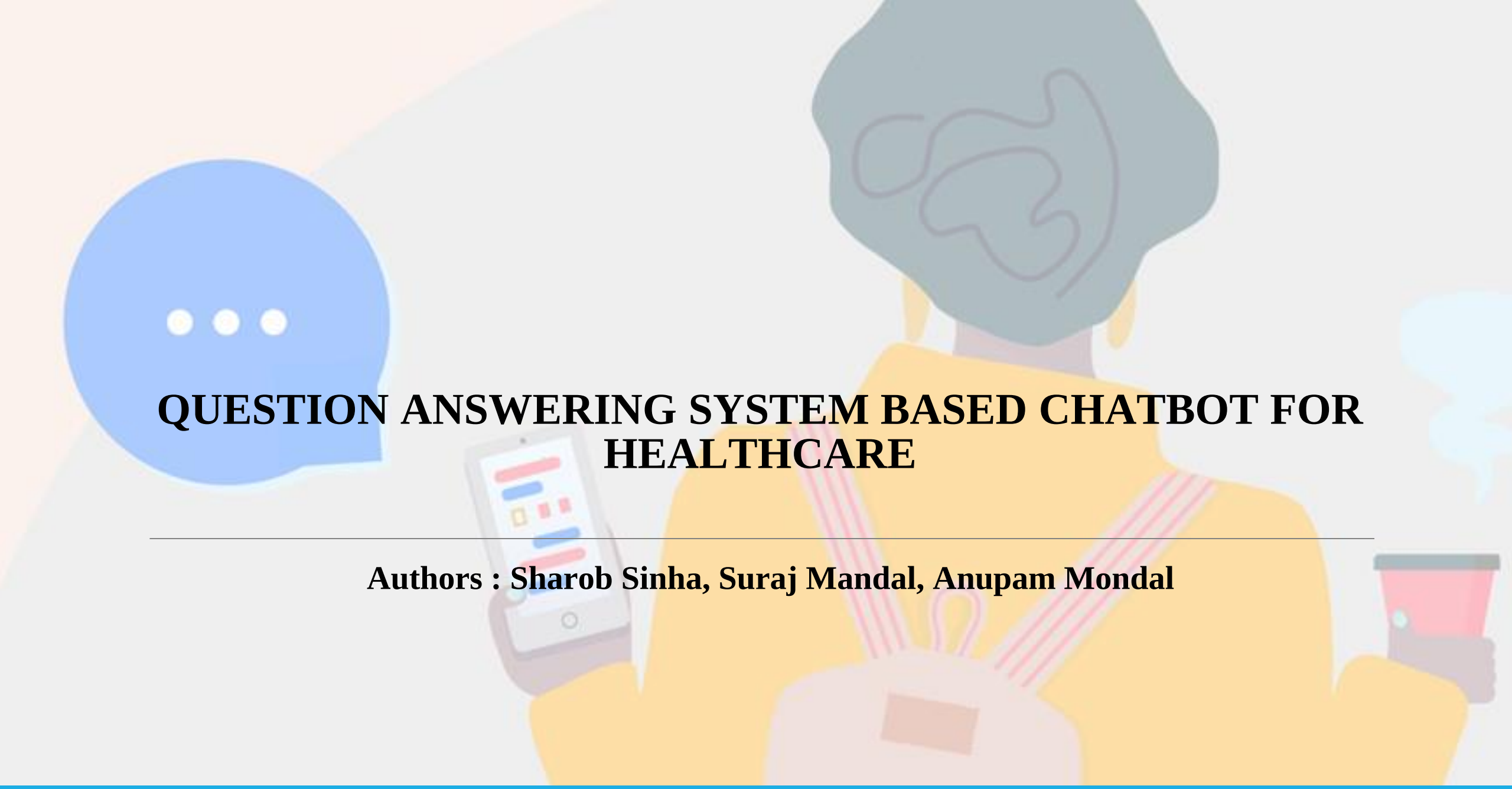


QUESTION ANSWERING SYSTEM BASED CHATBOT FOR HEALTHCARE

Authors : Sharob Sinha, Suraj Mandal, Anupam Mondal

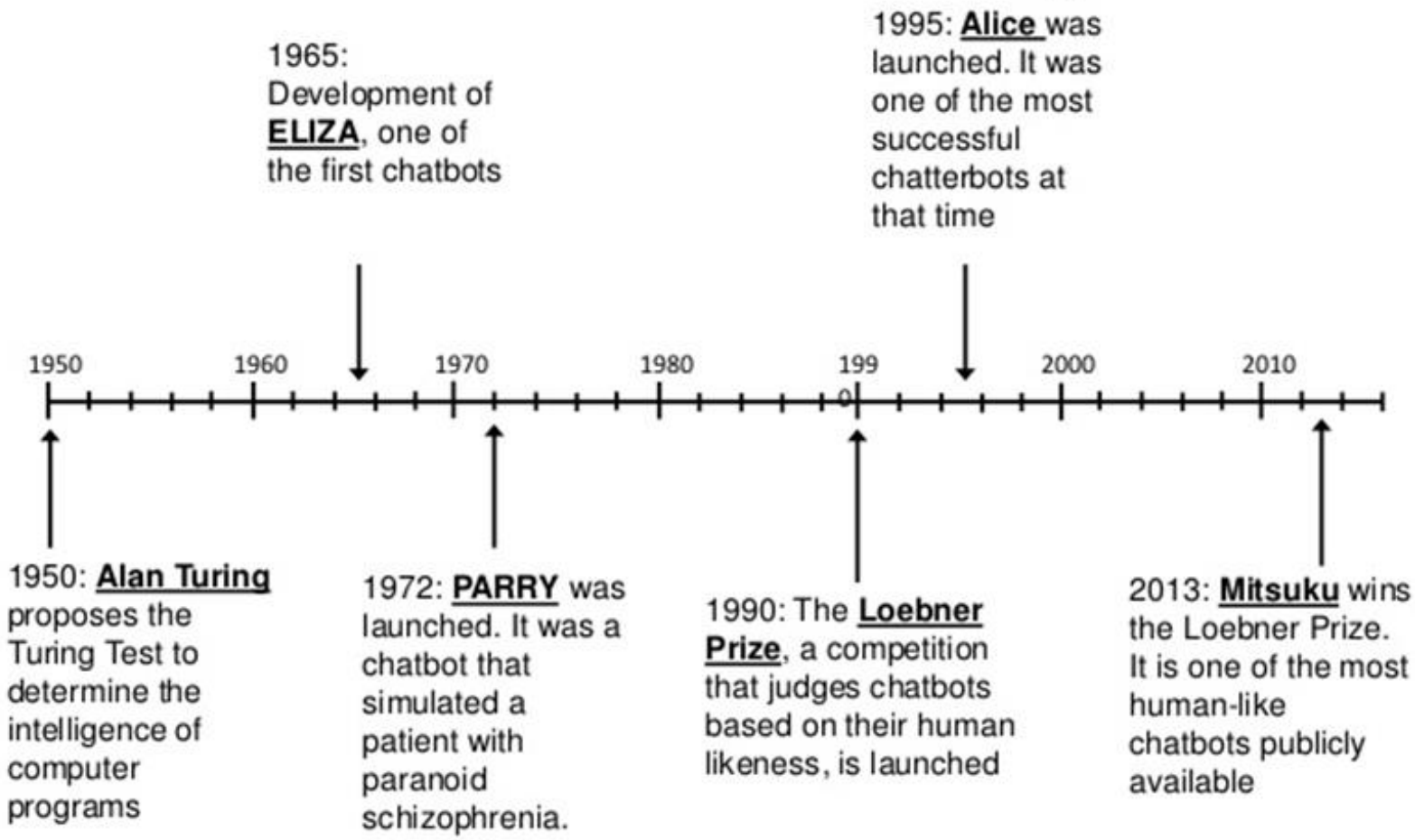
OVERVIEW OF THIS PRESENTATION

- What is a Chatbot?
- History of Chatbot
- Motivations
- Dataset
- Methodology
- Result Analysis
- Deployment
- Conclusions

What Is A Chatbot ?

A chatter robot (chatbot) is a type of conversational agent, a computer program designed to simulate an intelligent conversation with one or more human users in natural language via auditory or textual methods.

Chatbots: History



```
  Welcome to
                 EEEEEE  LL      IIII  ZZZZZZZ   AAAAA
                 EE      LL       II       ZZ   AA   AA
                 EEEEE   LL       II     ZZZ    AAAAAAA
                 EE      LL       II    ZZ      AA   AA
                 EEEEEE  LLLLLL  IIII  ZZZZZZZ  AA   AA

  Eliza is a mock Rogerian psychotherapist.
  The original program was described by Joseph Weizenbaum in 1966.
  This implementation by Norbert Landsteiner 2005.


ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```

	Your.MD	Sensely	Buoy Health	Infermedica	Florence
Funds raised	$17.3 MM	$11.8 MM	$9 MM	$5 MM	Unspecified
Year founded	2012	2013	2014	2012	2016
HQ location	London, UK	San Fran., CA	Boston, USA	Wroclaw, POL	Dortmund, DEU
Staff size	49	20	23	26	Unspecified
Target user	Patient, Clinician	Patient, Clinician	Patient	Patient, Clinician	Patient
Types of data processed	Chat (text)	Chat (text), image, video	Chat (text)	Chat (text), image, video	Chat (text)
Est. current users	Unspecified	Unspecified	Unspecified	Unspecified	2,000+?

Data collected from Crunchbase and LinkedIn, March 2018

techemergence

Motivations

- With the increase in population it has become very hard for doctors to assist every patient
- A patient could not get immediate help from doctors individually
- To develop a healthcare chatbot which will provide a virtual assistant
- To design a healthcare chatbot for interpreting patient’s symptoms

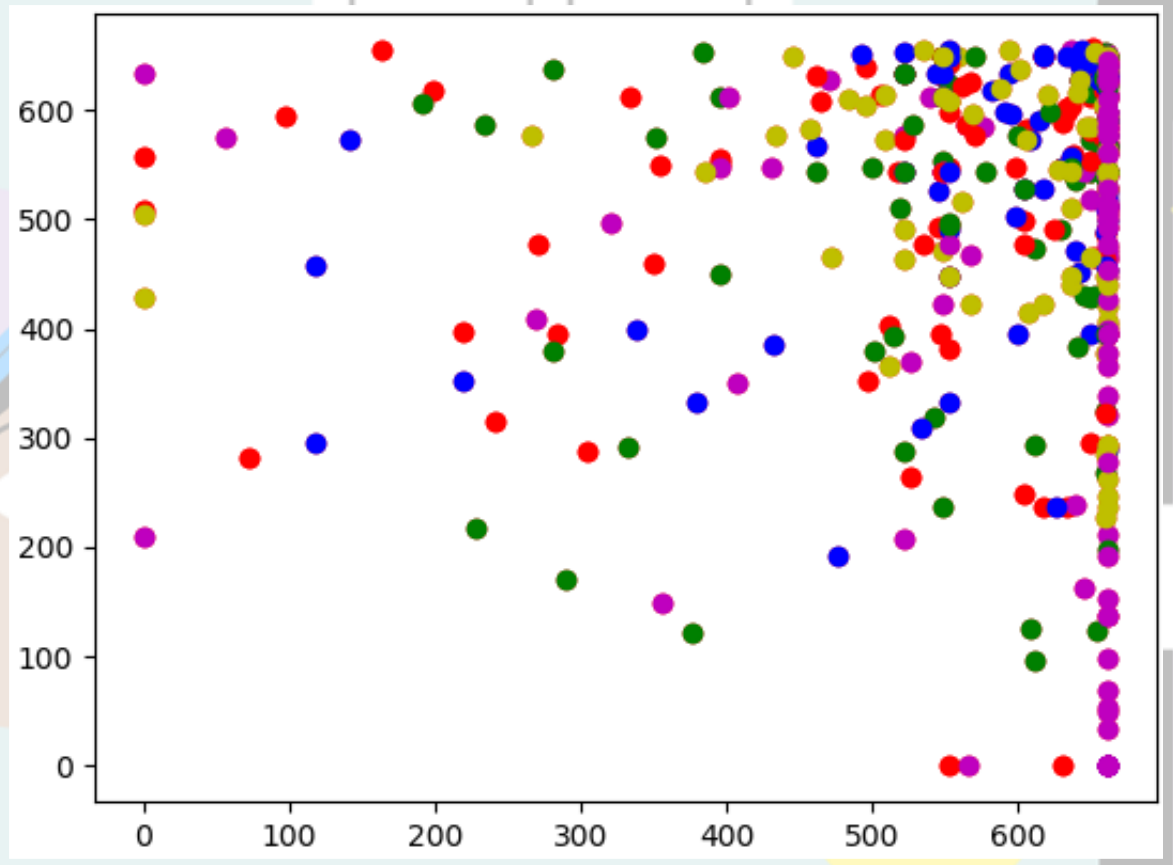
Dataset

- A proper dataset is essential in building an automated chatbot application.
- The objective of the project was to develop a chatbot which can recognize diseases for a set of symptoms. Hence, the required dataset was all existing diseases along with their symptoms.
- We have collected our dataset from the website https://www.medicinenet.com/
- The final dataset was containing 439 diseases and 1909 symptoms. The final dataset contains diseases followed by their symptoms.

Sample Dataset

	A	B	C	D
1	alcohol abuse	blackout	dizziness	shakiness
2	keratitis	Pain in the eyes	vision disorder	blurred vision
3	infectious mononucleosis	fatigue	fever	chills
4	fungal arthritis	heat	swelling	warmth
5	scarlet fever scarlatina	peeling	roughness	red spots
6	dissociative identity disorder	narcissistic personality disorder	posttraumatic stress disorc	dissociative identity disorde
7	osteoarthritis	pain in the joints	hands	hip
8	bocavirus infection	Cough	Wheezing	Fever
9	teen depression	suicide	Apathy	Complaints of pains
10	Excessive or inappropriate guilt	being late for classes	skipping school	Loss of interest in food or c
11	coxsackie virus	a poor appetite	and respiratory illness	including sore throat
12	staph infection	blister	boil	impetigo
13	dry eyes	dryness in eyes	eye discomfort	feeling of something in eye
14	salivary gland cancer	salivary gland cancer	A lump or swelling on or ne	Numbness in part of your fa
15	malaria	fever	headaches	and vomiting

Plot Diagram to Visualize Dataset



Methodology

The following algorithm illustrates the module building steps in detail.

STEP-1: Initially, we have generated the sparse matrix.
STEP-2: Thereafter, we have employed K-Means and Mini Batch K-Means approach on the processed sparse matrix to assign the proper groups of diseases.
STEP-3: After a detailed observation, we have set a threshold K value as 7 which was incremented up to K value 12 to extract the correct group from our dataset (D).

$$\mathbf{D = <S_1, S_2, \ldots., S_n : Di>}$$

Methodology

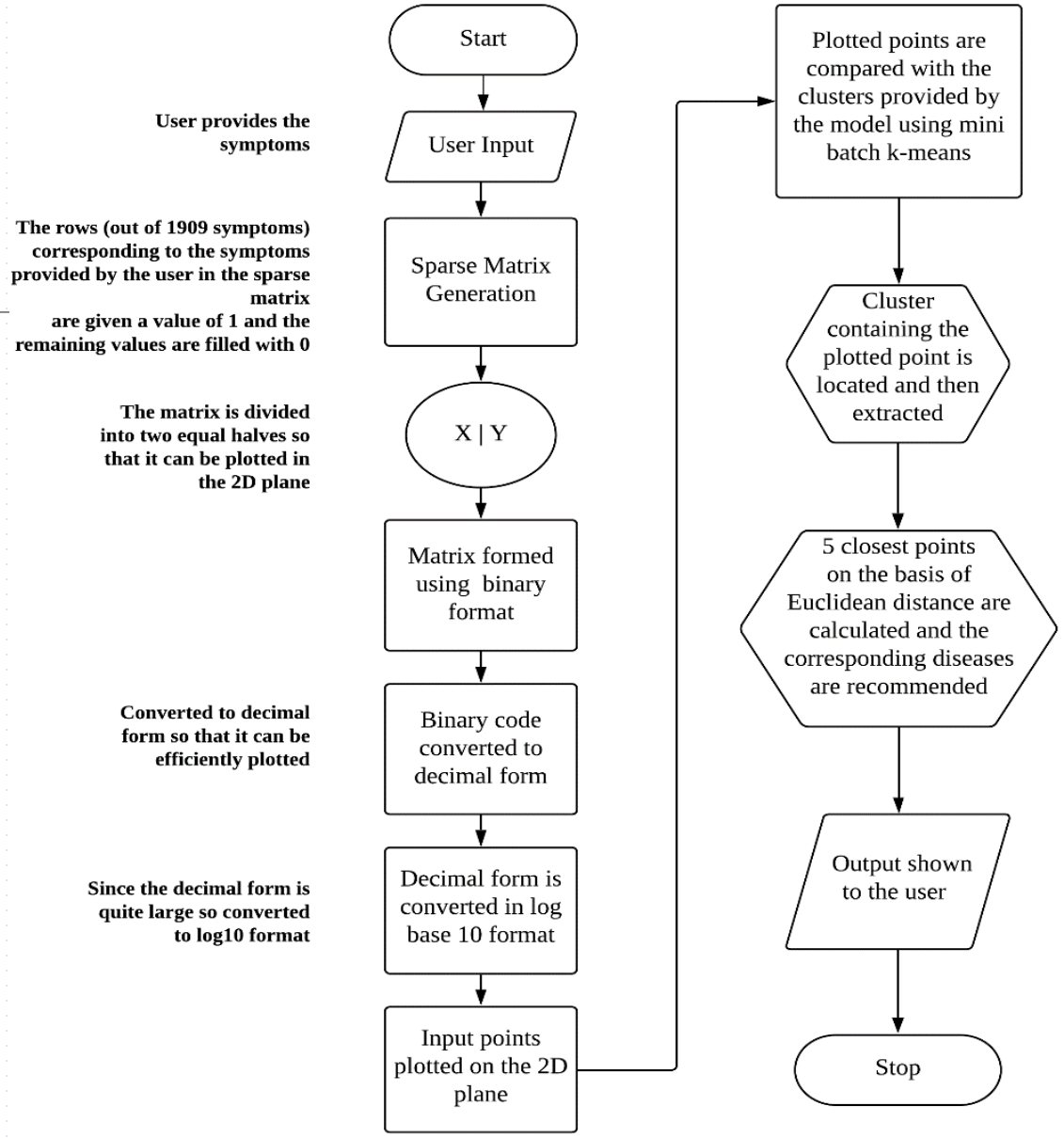
STEP-4: A user has provided a set of symptoms (Si) to identify common diseases. In that occasion, we have employed Euclidean distance technique to compute the similarity between STEP 3 generated group of diseases along with the input symptoms.

$$Si = \langle S_1', S_2', \ldots\ldots, S_n' \rangle$$

$$\textbf{Similarity} = \sqrt{(S_1 - S_1')^2 + (S_2 - S_2')^2 + \ldots\ldots + (S_n - S_n')^2}$$

STEP-5: In our case, we have set a threshold value > 0.6 for similarity measurement.

STEP-6: Finally, we have considered the similar diseases which scores more than 0.6. On the other hand, it is also applicable for predicting similar symptoms from a set of users given diseases under the proposed chatbot.



The Flowchart illustrating the steps of the working module

Result Analysis

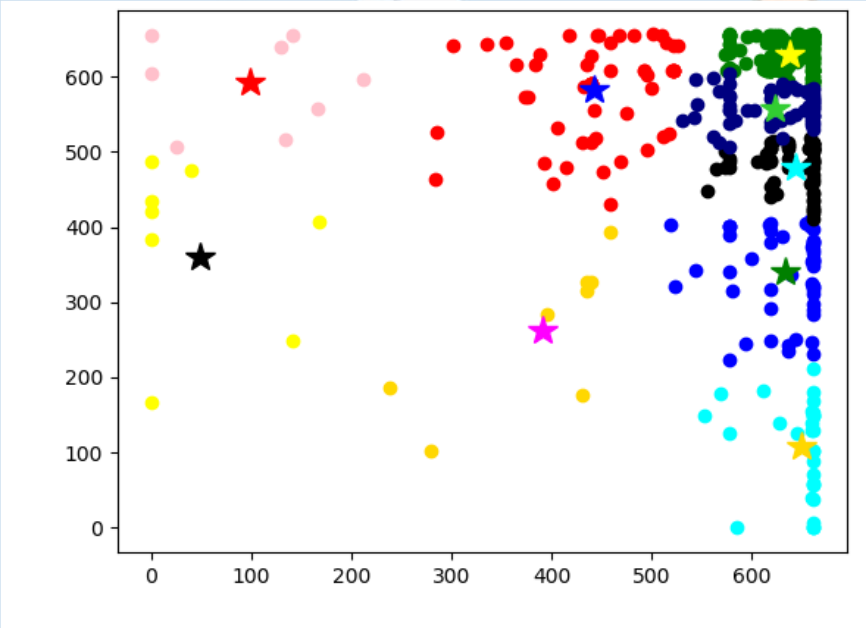
In order to validate both of the clusters of the proposed system, we have performed a comparison-based agreement between the annotated and predicted output. The output has been produced in the form of accuracy.

In this process, we have varied K values from 2 to 12.We have noticed that the range of K values 2 to 6 not provided an adequate output for the proposed system. So, we have presented rest of K values in the range of 7 to 12.

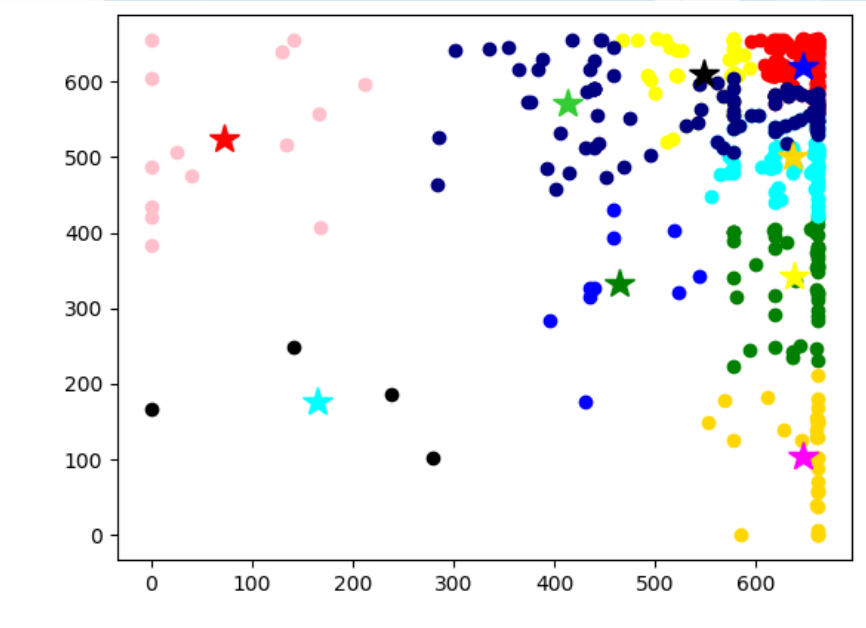
K-values	Mini-Batch K-Means	K-Means
7	80	74
8	83	78
9	87	84
10	81	76
11	80	72
12	82	75

A comparative result analysis for various K-values under Mini Batch K-Means and K-Means

When ‘K’ instance was taken as ‘9’

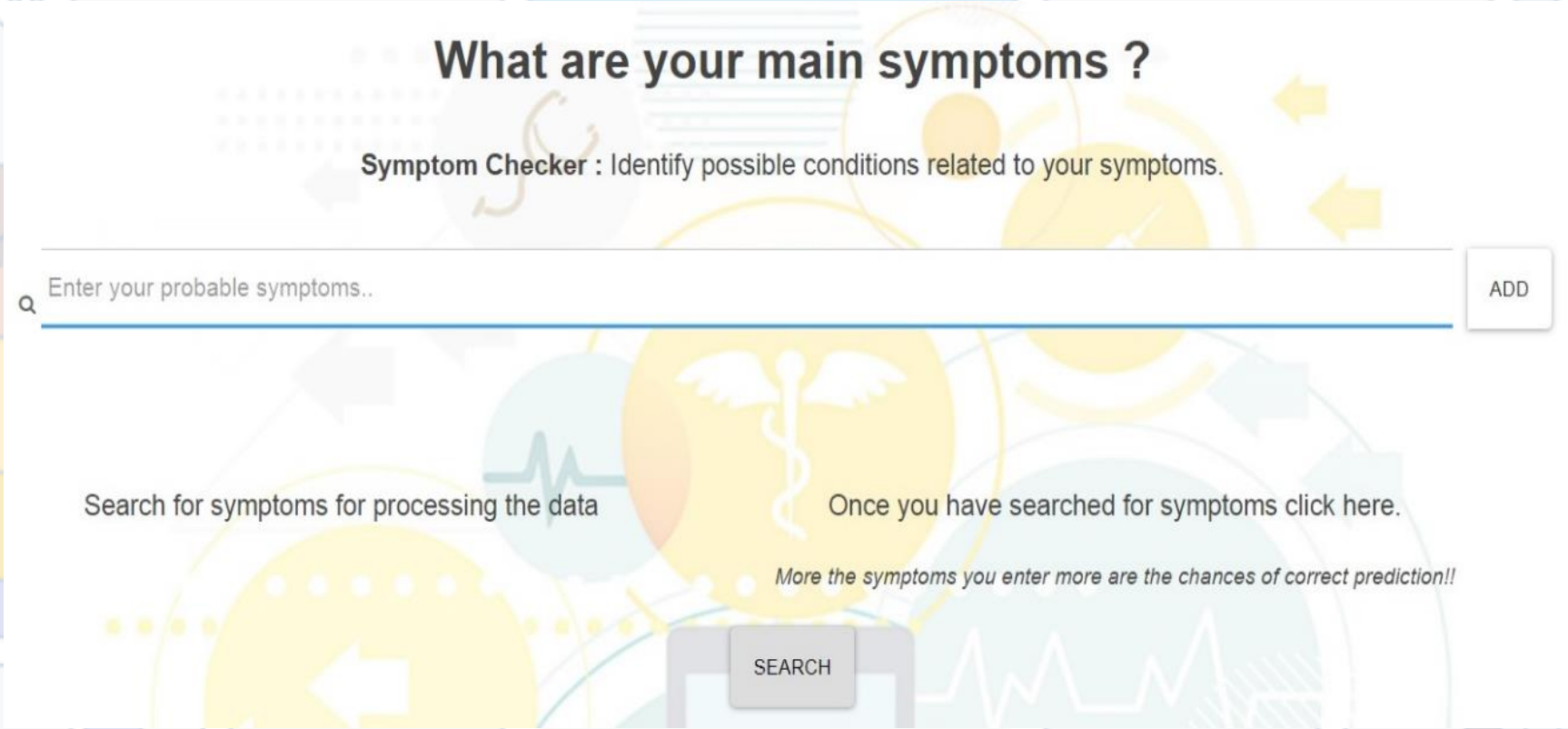


Mini Batch K Means (K=9)



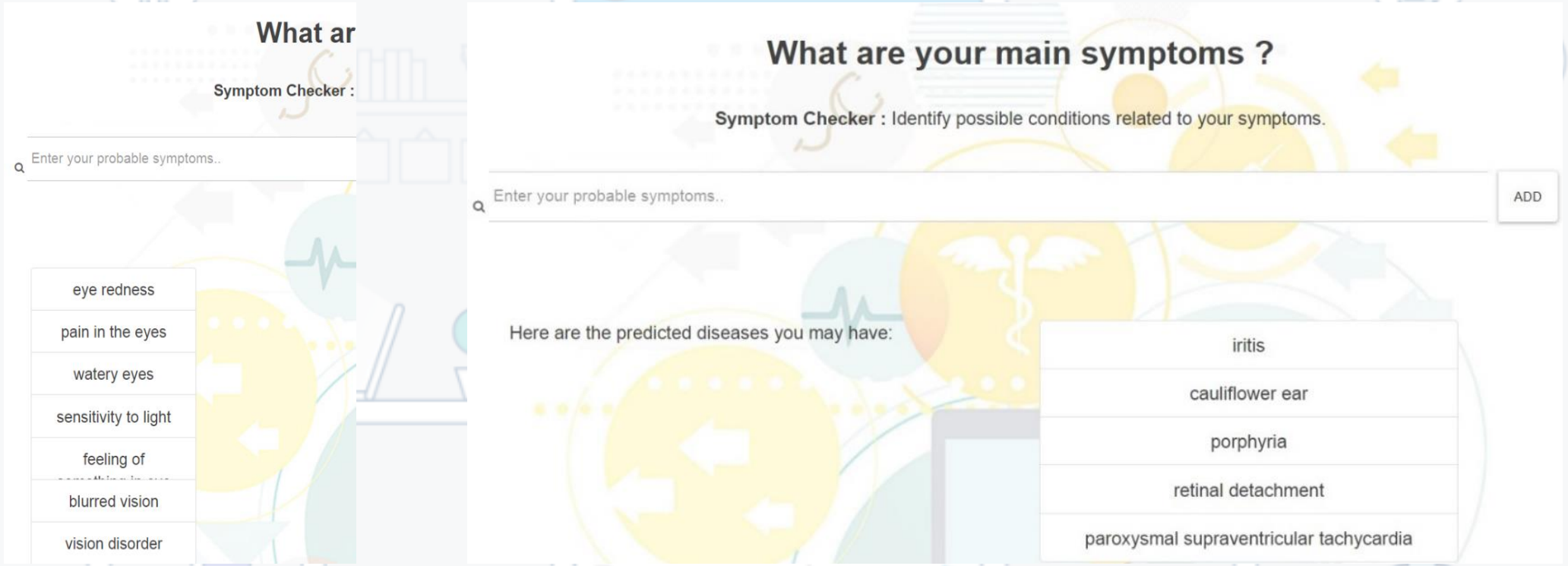
K Means (K=9)

Deployment



The Main UI of the Medical Chatbot

Deployment



The Symptoms of the disease iritis is given as input and the result predicted by the system have the disease iritis

Conclusions

➢ **The research was primarily focused on developing a chatbot in the healthcare domain.**

➢**To prepare the chatbot we have prepared an experimental dataset that contains 439 diseases and 1909 symptoms. Thereafter, we have converted in the form of sparse matrix and applied on two different clustering approaches namely, K-Means and Mini-Batch K-Means. We have taken a set of inputs in the form of symptoms and diseases and processed with Euclidean distance to identify a similar group of diseases and symptoms, respectively.**

➢**Finally, we have validated the proposed chatbot depend on K value selection where we have observed K value 9 provides 87% accuracy using Mini-Batch K-Means clustering.**

➢**Additionally, we have observed that this chatbot is unable to predict a set of diseases with unknown symptoms due to dependency on the sparse matrix. Hence, we have planned to manage this difficulty in our future research. The proposed chatbot may assists in designing various automated applications like lexicon design, concept identification, relationship extraction, recommendation system etc. in healthcare.**

References

•“ELIZA”, Weizenbaum, Joseph (1976). Computer Power and Human Reason: From Judgment to Calculation. New York: W.H. Freeman and Company. pp. 2,3,6,182,189. ISBN 0-7167-0464-1.

•Turing A.M. (2009) Computing Machinery and Intelligence. In: Epstein R., Roberts G., Beber G. (eds) Parsing the Turing Test. Springer, Dordrecht

•Mauldin, M.L., 1994, August. Chatterbots, tinymuds, and the turing test: Entering the loebner prize competition. In AAAI (Vol. 94, pp. 16-21).

•Ni, L., Lu, C., Liu, N. and Liu, J., 2017, November. Mandy: Towards a smart primary care chatbot application. In International Symposium on Knowledge and Systems Sciences (pp. 38-52). Springer, Singapore.

•Brindha, G. "Emerging trends of telemedicine in India." Indian Journal of Science and Technology 6.sup 5 (2013).


THANK YOU