

**DISEASE PREDICTOR CHATBOT USING NAIVE -**

**BAYES AND NLP**

# A PROJECT REPORT

Submitted by

**MADHUSHA HARINI** **M G** (231801091)

**JISHANTHI R** (231801072)

# AI2331 FUNDAMENTALS OF MACHINE LEARNING

Department of Artificial Intelligence and Data Science

Rajalakshmi Engineering College, Thandalam

**BONAFIDE CERTIFICATE**

This is to certify that the Mini project work titled “**DISEASE PREDICTOR CHATBOT USING NAIVE - BAYES AND NLP**” done by , Madhusha Harini .M.G

(231801091),Jishanthi .R(231801072) is a record of bonafide work carried out by him/her under my supervision as a part of MINI PROJECT for the subject titled **AI23331 -FUNDAMENTALS OF MACHINE LEARNING** by Department of Artificial Intelligence and Data Science.

**SIGNATURE SIGNATURE**

**Dr.GNANASEKAR J M Dr.INDIRA PRIYA P**

**HEAD OF THE DEPARTMENT ASSISTANT PROFESSOR.**

Department of Artificial Intelligence Department of Artificial

Intelligence and Data Science, Intelligence and Data Science,

Rajalakshmi Engineering College, Rajalakshmi Engineering College,

Thandalam, Thandalam,s

Chennai-602 105. Chennai- 602 105.

Submitted for the project viva-voce examination held on .

**INTERNAL EXAMINER EXTERNAL EXAMINER**

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **S. No** | **Chapter** | **Page Number** |
| 1. | ABSTRACT | iv |
| 2. | INTRODUCTION | 1 |
| 3. | RELATED WORK | 3 |
| 4. | MODEL ARCHITECTURE | 5 |
| 5. | IMPLEMENTATION | 6 |
| 6. | RESULT | 8 |
| 7. | CONCLUSION | 10 |
| 8. | APPENDIX | 11 |
| 9. | REFERENCES | 15 |

# ABSTRACT

Our project is a disease prediction chatbot system designed to diagnose illnesses based on a person's symptoms. Leveraging advanced machine learning techniques, particularly the Naive Bayes algorithm, the model accurately predicts potential diseases by analyzing input symptoms using natural language processing (NLP). The chatbot not only greets users warmly but also enables quick and informed decisions regarding health concerns. In addition to diagnosing diseases, it provides users with personalized health precautions to help manage their well-being effectively. The system also includes functionality to offer healthcare numbers for various countries, ensuring users have immediate access to professional medical assistance when necessary. This innovative model represents a significant improvement over traditional diagnostic methods, providing both precise predictions and practical advice to empower users in managing their health. The integration of a chatbot interface enhances user engagement, making health management a collaborative and interactive process that contributes to a more informed and health-conscious society.

# CHAPTER 1 INTRODUCTION

The Disease Predictor project is a pioneering Python application designed to revolutionize how users seek medical advice based on reported symptoms. In an era of technological advancements, this project harnesses the power of machine learning to assist individuals in identifying potential medical conditions promptly. The primary focus lies in utilizing advanced machine learning techniques, specifically employing the Naive Bayes algorithm, which enables the disease predictor to make accurate predictions about potential diseases. Additionally, the chatbot provides relevant information on those diseases and suggests precautionary measures to enhance user understanding of their health situation. By integrating these functionalities, the Disease Predictor project represents a significant step forward in leveraging technology to improve healthcare access and decision-making.

As part of this project, a comprehensive dataset was employed to enable accurate disease predictions. The primary dataset consists of symptoms and corresponding diseases, essential for the Naive Bayes model to analyze user input effectively. The chatbot identifies potential medical conditions based on reported symptoms and provides healthcare numbers for various countries for immediate assistance. Additionally, it offers personalized health precautions related to identified diseases, ensuring users receive relevant advice. This structured approach ensures the chatbot delivers precise predictions and practical guidance for those seeking medical advice.

This report delves into the intricacies of the Disease Prediction Chatbot project, covering its machine learning model, data analysis, implementation details, and the integration of natural language processing (NLP) for an enhanced user experience. The subsequent sections will provide an in-depth exploration of the various components, methodologies, and outcomes achieved throughout the development of this innovative healthcare solution. By focusing on accurate disease predictions and personalized health precautions, this project represents a significant advancement in utilizing technology to improve healthcare access and decision-making.

The integration of machine learning in healthcare has emerged as a transformative force, significantly enhancing the accuracy of medical diagnosis and patient care. The importance of utilizing machine learning in this sector lies in its ability to analyze vast amounts of symptom and disease data, enabling the Disease Prediction Chatbot to provide timely and accurate predictions. By leveraging advanced algorithms like Naive Bayes, the chatbot facilitates personalized health advice and precautionary measures based on user-reported symptoms. This innovative approach not only improves diagnostic efficiency but also empowers users to make informed health decisions. Below are key aspects highlighting the significance of machine learning in the healthcare landscape:

## Predictive Analytics:

Machine learning algorithms can analyze large datasets to identify patterns and trends, enabling the prediction of potential health issues and disease outcomes. Early detection of diseases allows for timely intervention and improved patient outcomes.

## Disease Identification and Diagnosis:

Machine learning models, such as Naive Bayes and NLP, can be trained on medical data to assist in the identification and diagnosis of diseases based on symptoms. Improved accuracy in disease identification enhances the efficiency of healthcare professionals.

## Personalized Medicine:

Machine learning algorithms analyze patient data, genetic information, and treatment outcomes to tailor medical treatments to individual patients.The project utilizes two powerful machine learning models, the Naive Bayes and Natural language processing (NLP), to predict potential diseases based on reported symptoms. Each model has its strengths and characteristics that contribute to the accuracy and reliability of the ChatBot's predictions.

## NAIVE BAYES:

Naive Bayes is a probabilistic classification algorithm based on Bayes' theorem, which assumes that the presence of a particular feature in a class is independent of the presence of any other feature. This simplicity makes it a powerful tool for classification tasks, especially in text classification and disease prediction. In the Disease Prediction Chatbot, the Naive Bayes model is trained on a dataset that includes symptoms as features and corresponding diseases as labels. The algorithm calculates the probability of each disease given the input symptoms, allowing the model to make predictions by selecting the disease with the highest probability. This approach facilitates efficient and accurate disease identification, making Naive Bayes an ideal choice for the chatbot’s predictive capabilities.

Training Process:

Features: Symptoms reported by users.

Labels: Corresponding diseases:

Naive Bayes predicts diseases based on the independence of symptoms reported by users. It calculates disease likelihood using Bayes' theorem and symptom frequencies from the training dataset. Cross-validation assesses the model's performance, with the average score indicating its predictive accuracy, ensuring reliable disease predictions.

1. **NATURAL LANGUAGE PROCESSING (NLP) :**

Natural Language Processing (NLP) is utilized in the Disease Prediction Chatbot to interpret and analyze user-reported symptoms effectively. It converts user input into a structured format, allowing the chatbot to identify relevant symptoms and match them with corresponding diseases accurately. NLP also enables the chatbot to understand variations in language, enhancing the user experience by facilitating more natural interactions. Additionally, the chatbot greets users warmly, setting a friendly tone for the conversation and encouraging users to describe their symptoms for accurate disease predictions.

Training Process:

Features: Symptoms reported by users. Labels: Corresponding diseases.

NLP processes user input to extract symptoms from natural language descriptions, allowing the Disease Prediction Chatbot to understand various expressions of health concerns. By structuring this input, NLP enables accurate matching of symptoms to diseases. The effectiveness of NLP is assessed through user interactions, where successful symptom recognition enhances the chatbot's predictive capabilities.

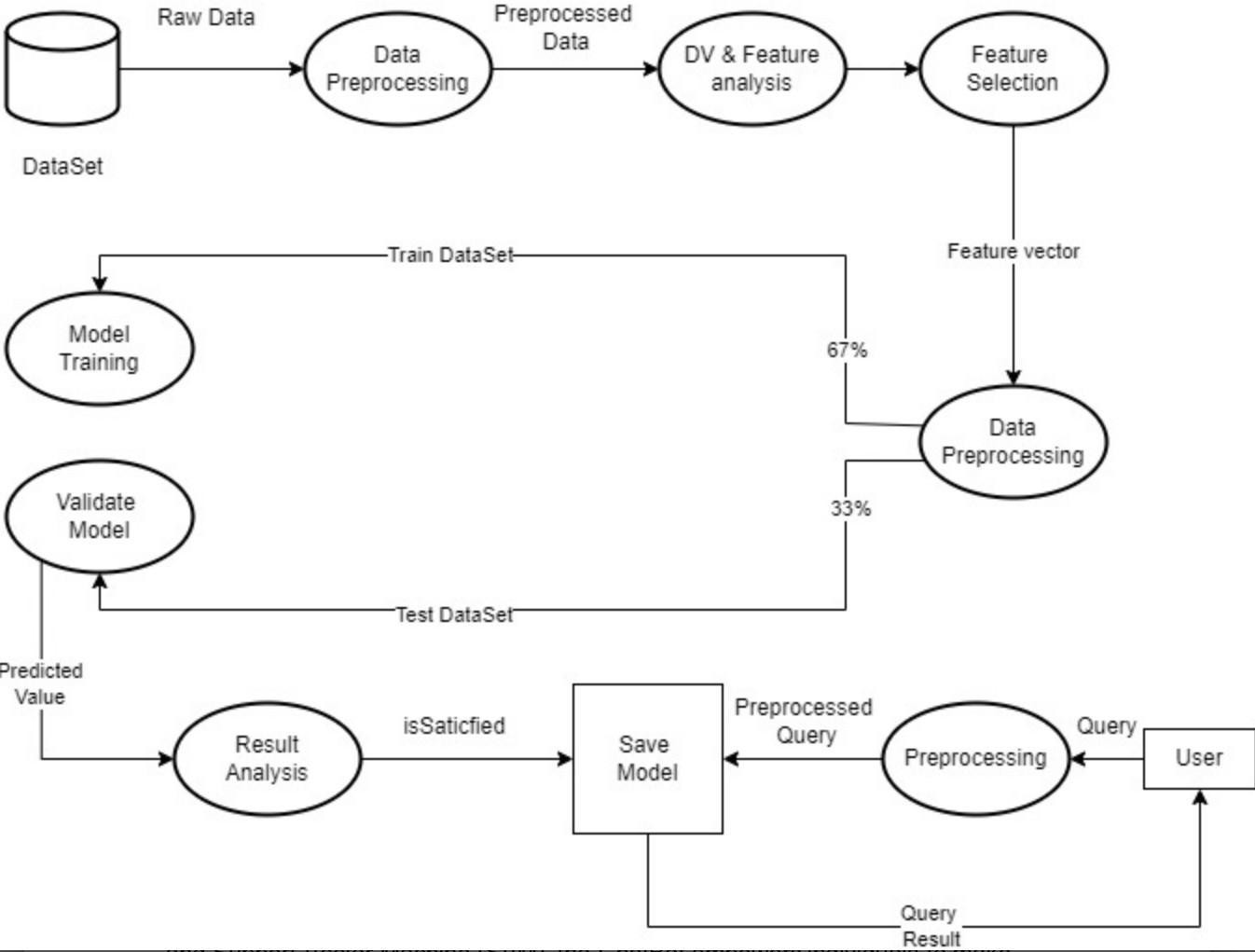
**CHAPTER 2 RELATED WORK**

* 1. The goal of this project is to develop a medical disease predictor using Artificial Intelligence that can diagnose potential illnesses based on reported symptoms and provide essential information about those diseases. This chatbot aims to improve healthcare accessibility and reduce costs by offering an interactive platform for users to describe their symptoms. Utilizing the Naive Bayes algorithm for machine learning, the chatbot processes user input through natural language processing (NLP) techniques to accurately identify relevant symptoms. The model is trained on a comprehensive dataset of symptoms and diseases, enabling it to make informed predictions. Additionally, the system provides users with tailored health precautions based on identified conditions, ensuring they receive relevant advice to manage their well-being effectively. By organizing and analyzing data efficiently, the chatbot delivers precise disease predictions while enhancing user understanding of their health situation.
  2. The authors use Natural Language Processing(NLP) algorithm.The system provides text-text assistance to communicate with bot in a user friendly manner.The disease predictor provides medical suggestions that can cure the disease based on user symptoms.If it is a severe health problem the user will be advised to consult a doctor for a better treatment.The disease predictor can also give you medical prescriptions for health problems.Along with the medicines,the disease predictor also provide you with ayurvedic remedies and homeopathy treatments for related health problems.
  3. The authors propose a disease predictor as a software application that facilitates interaction between users and a medical chatbot over the internet. Chatbots play a crucial role in today's healthcare landscape by providing patients with timely and relevant information. This project aims to create a medical disease predictor based on Artificial Intelligence and natural language processing (NLP). Through the chatbot, users can communicate using a text interface to describe their symptoms and receive instant responses. The chatbot assists potential patients by offering disease predictions, recommending precautions, and enhancing access to healthcare information. Available 24/7, these chatbots are user-friendly, efficient, and designed to provide comfortable

interactions, ultimately supporting users in managing their health effectively.

* 1. The authors propose a medical disease predictor called MedBot, designed to provide users with accurate health advice and support for maintaining a healthy lifestyle. The core concept is to develop a healthcare disease predictor based on Artificial Intelligence and Natural Language Processing, enabling the MedBot to identify potential illnesses and provide essential information before users consult a doctor. This approach enhances accessibility and aims to reduce healthcare costs. MedBot functions as a virtual medical assistant, educating users about their symptoms and encouraging healthier choices. Through a text-based interface, users engage in online conversations regarding their medical concerns, receiving personalized disease predictions based on the symptoms they provide. MedBot actively interacts with users, guiding them toward informed health decisions and improving their overall well-being.
  2. This disease predictor may be used by regular humans in any sort of emergency case, where it can advise people on primary care before seeing a doctor, or it can sometimes work as a doctor for small and short- term health issues such as a cold, headache, and so on. Along with this chatbot, there will be assistance for those in need who seek immediate solutions. A user can identify the true condition by reporting symptoms of it. The true goal of this job is to work on the user's symptoms and make medical advice based on them in order to decrease the time and expense associated with the procedure.

## CHAPTER 3 MODEL ARCHITECTURE



**User Input:** Users interact with the disease predictor by inputting their symptoms .

**Machine Learning Models**: The project employs two machine learning models, the Naive Bayes Classifier and Natural Language Processing , for disease prediction based on symptoms.

**Disease Information:** The disease classifier accesses datasets containing disease descriptions to provide comprehensive information about the predicted diseases.

**Precautionary Measures:** Users receive personalized precautionary measures based on the predicted disease, promoting preventive healthcare practices.

**Symptom Analysis and Guidance:** Reported symptoms are analyzed using the chatbot's primary dataset to provide users with relevant information on potential medical conditions. The chatbot also offers personalized health precautions and accessible health care numbers for various countries, ensuring that users receive both immediate guidance and the option to seek professional assistance if needed.

**Description of the datasets used :** The Disease classifier project relies on a set of datasets to facilitate accurate disease prediction, provide detailed information about diseases, and offer personalized precautionary measure.

## CHAPTER 4 IMPLEMENTATION

Disease disease predictor Implementation

The interaction with users in the Disease Prediction Chatbot is managed by the symptom\_input function. This function prompts users to enter their symptoms and utilizes the Naive Bayes Classifier to analyze the input, generating predictions about potential diseases along with relevant information and suggested precautions.

## Explanation:

**Symptom Input:**

The user is prompted to input symptoms they are experiencing.

The input is processed, and if there is ambiguity or multiple matches, the user is asked to choose the most relevant symptom.

## Primary Prediction:

The chatbot analyzes the user’s input symptoms using the **Naive Bayes algorithm**, evaluating the data against predefined symptom-disease relationships to generate accurate predictions.

## Follow-up Prediction:

A follow-up prediction is generated based on the initial symptoms provided by the user, enhancing the accuracy of the disease diagnosis through further analysis.

## Condition Calculation:

The severity of symptoms are used to calculate the condition and provide appropriate recommendations.

## Result Presentation:

## Naive Bayes Prediction Process:

**Model Training:**

The Naive Bayes model is trained using the dataset, learning the probability distributions of symptoms associated with various diseases. This involves calculating the likelihood of each symptom given a specific disease.

## User Interaction:

When the user inputs their symptoms, the disease predictor uses the trained Naive Bayes model to analyze the input data. The model evaluates the probability of each disease based on the provided symptoms.

## Probability Calculation:

The model calculates the conditional probabilities for each disease based on the user's symptoms. It applies Bayes' theorem to derive the most probable disease by considering the prior probabilities and the likelihood of symptoms.

## Prediction Output:

The disease with the highest calculated probability is identified as the predicted condition. The model presents this prediction to the user along with relevant information about the disease and suggested precautions.

## Continuous Learning:

As the system interacts with more users, it can update its model to improve accuracy and adapt to new data, enhancing its predictive capabilities over time.

## Follow-up Prediction:

A follow-up prediction is generated based on the initial symptoms provided by the user, enhancing the accuracy of the disease diagnosis through further analysis.

## Output Presentation:

The model outputs the predicted disease(s), associated descriptions, and recommended precautions to the user.

## Key Features of the Disease Predictor Project:

**User Interaction:**

The project features an interactive chatbot that prompts users to input their symptoms, ensuring effective communication and symptom reporting.

## Information Delivery:

Predictions are accompanied by detailed information about identified diseases, including descriptions and suggested health precautions for user awareness.

## Healthcare Resource Access:

The application provides contact information for healthcare professionals in various countries, facilitating immediate medical assistance.

## User Education:

It educates users on the severity of their symptoms, promoting better understanding and proactive healthcare management.

## CHAPTER 5

**RESULTS AND DISCUSSIONS**

**Challenges Faced**

* **Any challenges encountered during the development of the disease predictor:**
* Domain-Specific Challenges:

Healthcare terminology and domain-specific nuances can be complex. Ensuring that the disease predictor classifier understands and communicates medical information accurately is crucial for user trust.

* Generalization to Diverse Symptoms:

Ensuring that the model generalizes well to a diverse range of symptoms and conditions is important for the chatbot's effectiveness. Overfitting to specific patterns in the training data should be avoided.

## Solutions or workarounds implemented to overcome challenges :

* **Domain-Specific Challenges:**

Involve healthcare professionals in the development process to validate model outputs. Implement continuous learning mechanisms to adapt to evolving medical knowledge.

## Generalization to Diverse Symptoms:

Ensure a diverse and representative dataset during model training.

Regularly update the dataset to include new information and emerging medical conditions.

## CHAPTER 6 CONCLUSION AND FUTURE WORKS

**Conclusion**

In conclusion, the TalkBot.AI disease prediction project is a Python application designed to assist users in identifying potential medical conditions based on reported symptoms. The disease predictor utilizes machine learning techniques, specifically NAIVE BAYES and NATURAL LANGUAGE PROCESSING (NLP) to make predictions. The project aims to provide users with information on diseases, their descriptions, and suggested precautions.

**Future Enhancements**

* Ideas for improving the disease predictor:

Enhancing a healthcare disease prediction involves continuous improvement in various aspects to provide a better user experience and more accurate assistance. Here are some ideas for future enhancements to the Healthcare ChatBot:

* Integration with Telemedicine Services:

Enable the disease predictor to seamlessly integrate with telemedicine services, allowing users to schedule virtual consultations with healthcare professionals based on the chatbot's recommendations.

Personalized Health Profiles:

Implement user profiles to track individual health history, symptoms, and recommendations over time. This can enhance the chatbot's ability to provide personalized advice.

* Natural Language Understanding (NLU) Improvements:

Enhance the chatbot NLU capabilities to better understand user queries, including context-aware responses and the ability to handle more complex medical inquiries.

* Collaboration with Healthcare Professionals:

Establish partnerships with healthcare professionals to ensure the chatbot's recommendations align with current medical practices and guidelines.

* User Authentication and Data Security:

Implement secure user authentication mechanisms to protect user health data and ensure compliance with data protection regulations.

## CHAPTER 7 APPENDIX

**Appendix-1: CODE**

import pandas as pd

from sklearn.feature\_extraction.text import CountVectorizer from sklearn.naive\_bayes import MultinomialNB

from sklearn.pipeline import make\_pipeline import ipywidgets as widgets

from IPython.display import display

data = {

'Disease': ['Flu', 'Cold', 'COVID-19', 'Allergy'], 'Symptoms': [

'fever,cough,sore throat,runny nose,body aches,fatigue,chills', 'cough,runny nose,sneezing,sore throat,low-grade fever,mild headache',

'fever,cough,loss of taste,shortness of breath,body aches,fatigue,sore throat', 'sneezing,itchy eyes,runny nose,postnasal drip,cough,red eyes'

]

}

df = pd.DataFrame(data)

df['Symptoms'] = df['Symptoms'].apply(lambda x: x.split(','))

model\_data = []

for disease, symptoms in zip(df['Disease'], df['Symptoms']): for symptom in symptoms:

model\_data.append((symptom.strip(), disease))

model\_df = pd.DataFrame(model\_data, columns=['Symptom', 'Disease'])

X = model\_df['Symptom'] y = model\_df['Disease']

model = make\_pipeline(CountVectorizer(), MultinomialNB()) model.fit(X, y)

healthcare\_numbers = {

"India": "Aarogya Setu: 1075", "USA": "CDC: 1-800-232-4636",

"UK": "NHS: 111"

}

precautions = [

"Wash your hands frequently with soap and water.", "Avoid close contact with sick individuals.",

"Practice good respiratory hygiene by covering your mouth and nose with a tissue or your elbow when you cough or sneeze.",

"Stay home if you feel unwell."

]

def predict\_disease(symptoms\_input):

symptoms = [symptom.strip() for symptom in symptoms\_input.lower().split(',')] predicted\_diseases = model.predict(symptoms)

if predicted\_diseases.size > 0: unique\_diseases = set(predicted\_diseases)

return f"You may have: {', '.join(unique\_diseases)}. Please note that I can make mistakes, and it's best to consult a healthcare professional for a proper diagnosis."

else:

return "Sorry, I couldn't identify any disease based on your symptoms."

def get\_healthcare\_number(country):

return healthcare\_numbers.get(country, "Healthcare number not available for the specified country.")

def chat\_response(user\_input): user\_input = user\_input.lower()

if "hi" in user\_input or "hello" in user\_input:

return "Hello! How can I assist you today? You can describe your symptoms." elif "not well" in user\_input or "i am not well" in user\_input:

return "I'm sorry to hear that. Please take care of yourself and let me know how I can help." elif "helpline" in user\_input or "help line" in user\_input or "healthcare number" in user\_input:

return "Please tell me your country to provide the helpline number." elif "precautions" in user\_input:

return "\n".join(precautions) # Return precautions when asked else:

return None # No specific response for other inputs

def handle\_healthcare\_request(user\_input): words = user\_input.split()

country = words[-1].capitalize() # Assume the last word is the country name return get\_healthcare\_number(country)

input\_box = widgets.Textarea( description='Your Message:',

placeholder='Type your symptoms or greeting...', layout=widgets.Layout(width='100%', height='100px')

)

output\_box = widgets.Output()

submit\_button = widgets.Button(description='Send')

def on\_submit(button): with output\_box:

output\_box.clear\_output()

user\_input = input\_box.value.strip() chat\_reply = chat\_response(user\_input)

print(f"You: {user\_input}")

if not user\_input:

print("Please enter a message to proceed.") return

if chat\_reply:

if "helpline" in user\_input or "help line" in user\_input or "healthcare number" in

user\_input:

helpline\_info = handle\_healthcare\_request(user\_input) print(f"Bot: {helpline\_info}")

else:

print(f"Bot: {chat\_reply}")

else:

# If no specific chat response, proceed with disease prediction result = predict\_disease(user\_input)

print(f"Bot: {result}")

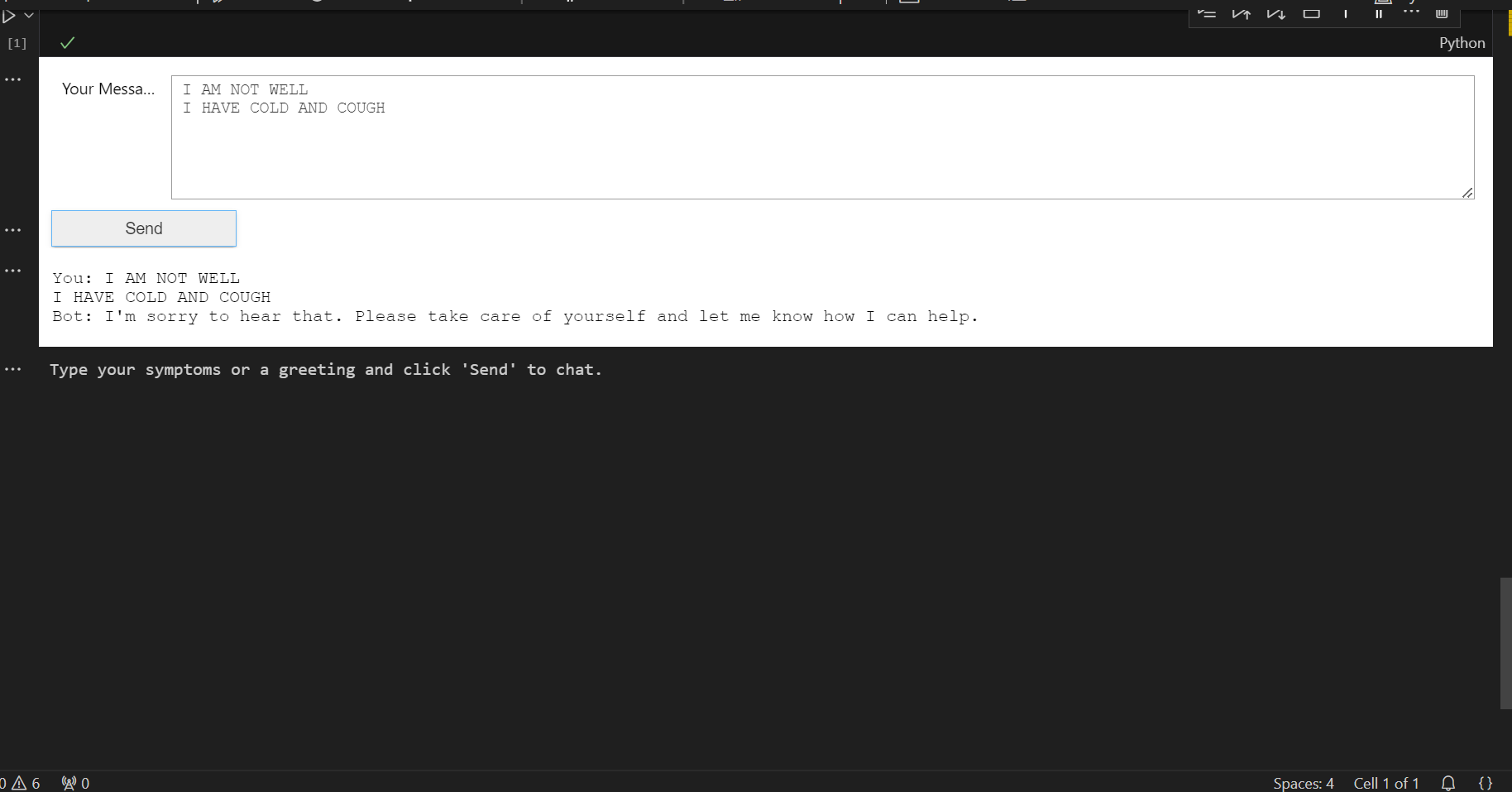
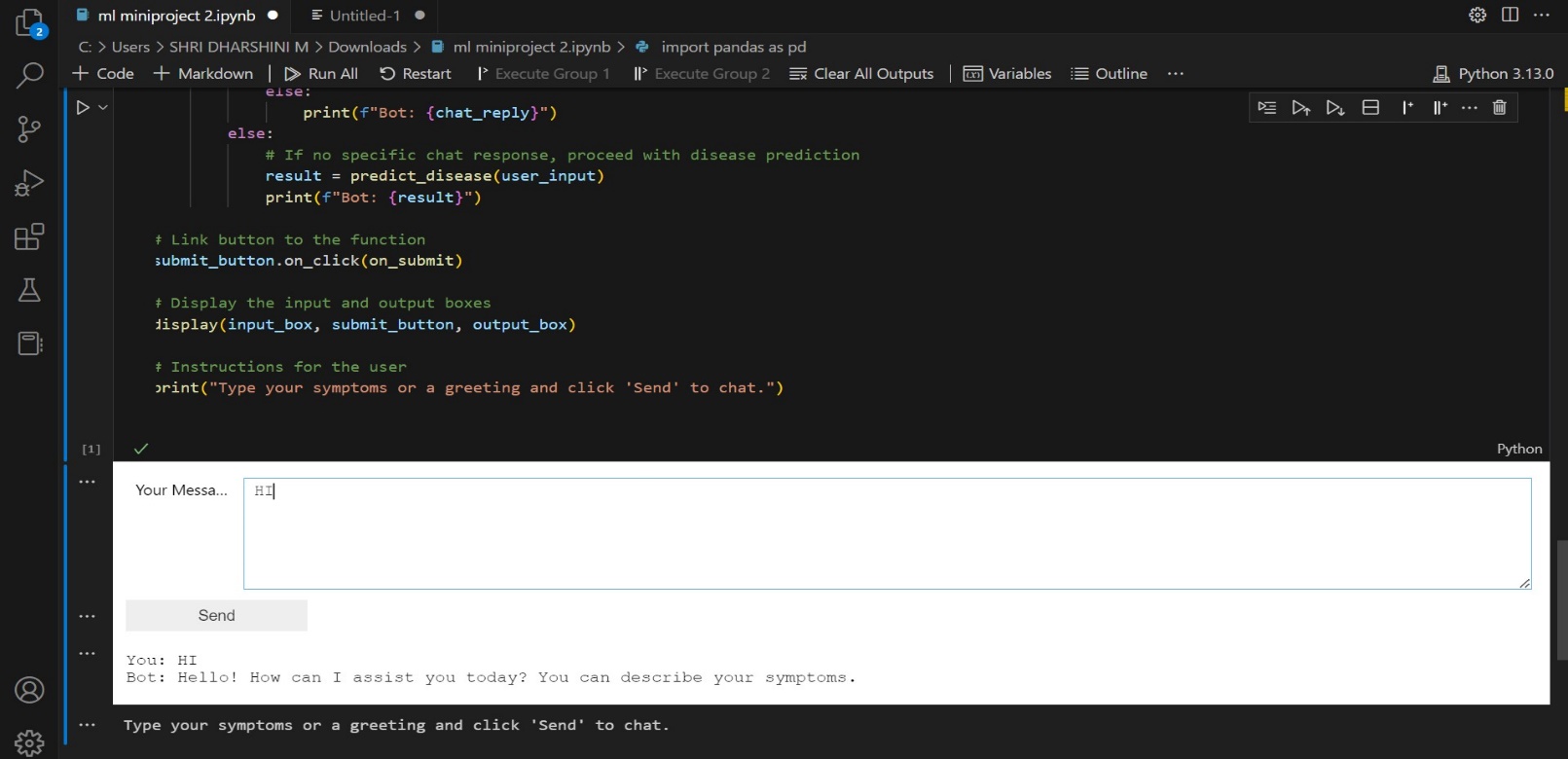
submit\_button.on\_click(on\_submit)

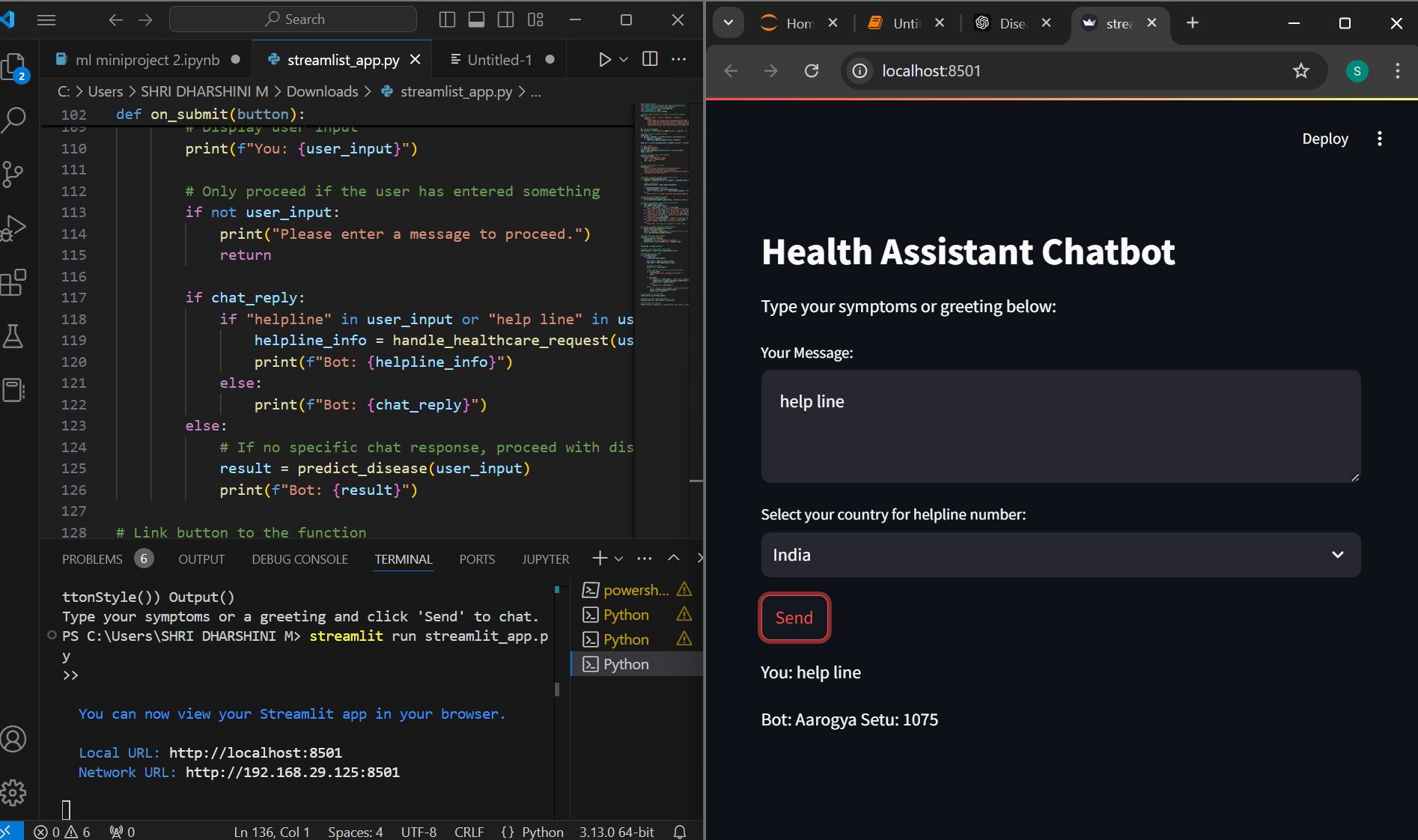
# Display the input and output boxes display(input\_box, submit\_button, output\_box)

# Instructions for the user

print("Type your symptoms or a greeting and click 'Send' to chat.")

## CHAPTER 8 OUTPUT





**CHAPTER 8 REFERENCES**

1. L. Athota, V. K. Shukla, N. Pandey, and A. Rana, "Disease Predictor for Healthcare System Using Artificial Intelligence," 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2020, pp. 619-622, doi: 10.1109/ICRITO48877.2020.9197833. Keywords: {Medical services; Databases; Expert systems; Feature extraction; Medical diagnostic imaging; Chatbot; Healthcare; Artificial Intelligence; Virtual Assistance; TFID; N-gram}. (https://ieeexplore.ieee.org/document/9197833)
2. Ramalingam, J., Dava, S., Nagappan, U., Ganesan, K., & Venkateswaran, N. (2023). "Personal Healthcare Disease Predictor for Medical Suggestions Using Artificial Intelligence and Machine Learning," Eur. 12, 6004-6012. doi: 10.31838/ecb/2023.12.s3.670. (https://[www.researchgate.net/publication/372658029\_Section\_A-Research\_paper\_Personal\_Healt](http://www.researchgate.net/publication/372658029_Section_A-Research_paper_Personal_Healt) hcare\_Chatbot\_for\_Medical\_Suggestions\_Using\_Artificial\_Intelligence\_and\_Machine\_Learning\_ Eur)
3. Shetty, R., Bhosale, A., Verma, P., & Phalke, A. (2022). "AI Based Healthcare Disease Predictor," Proceedings of the 7th International Conference on Innovations and Research in Technology and Engineering (ICIRTE-2022), organized by VPPCOE & VA, Mumbai-22, INDIA. Available at SSRN: (https://ssrn.com/abstract=4109100) or <http://dx.doi.org/10.2139/ssrn.4109100>
4. K. Anjum, M. Sameer, and S. Kumar, "AI Enabled NLP based Text to Text Medical Chatbot," 2023 3rd International Conference on Innovative Practices in Technology and Management (ICIPTM), Uttar Pradesh, India, 2023, pp. 1-5, doi: 10.1109/ICIPTM57143.2023.10117966. Keywords: {Training; Machine learning algorithms; Hospitals; Machine learning; Transforms; Oral communication; Chatbots}. (https://ieeexplore.ieee.org/document/10117966)
5. "Design and Implementation of Healthcare Chatbot Using Artificial Intelligence," International

Journal of Engineering Research and Technology, DOI: 10.17577/IJERTCONV10IS12016. https://[www.ijert.org/design-implementation-of-healthcare-chatbot-using-artificial-intelligence)](http://www.ijert.org/design-implementation-of-healthcare-chatbot-using-artificial-intelligence))