

Machine Learning Project

Disease Symptom Prediction Chatbot

Team members:

- Harshavardhan Manne
- Devi Sumanth Pasunoori
- Mohammad Abrar Baig
- Neeraj Reddy Aluka

Agenda

Abstract

Problem Specification

Data Overview

Methods

Project Design and Milestones

Programming and APIs

Related Projects

Conclusion



Abstract

- ▶ The "Disease Symptom Prediction Chatbot" project represents a significant advancement in the application of machine learning in healthcare. This innovative system leverages natural language processing (NLP) and predictive analytics to interpret user symptoms and provide potential disease diagnoses. The chatbot engages with users in a conversational manner, gathering information about their symptoms and health concerns. Using a comprehensive database of diseases and their associated symptoms, the system employs advanced algorithms to analyze the input and predict possible ailments.
- ▶ At the core of this project is a machine learning model trained on a vast array of medical data, including symptoms, diagnoses, and patient histories. This model is designed to recognize patterns and correlations between various symptoms and diseases, allowing for accurate and timely predictions. Because it can process natural language, the chatbot can comprehend and interpret user inputs even when they are given in informal or non-standard medical terminology. As a result, a wide variety of users can access and easily use the system.
- ▶ The potential impact of the Disease Symptom Prediction Chatbot in the healthcare industry is substantial. It offers an accessible preliminary diagnosis tool, reducing the burden on healthcare professionals and assisting in early disease detection. However, the project also raises important considerations regarding accuracy, privacy, and the need for professional medical consultation. The chatbot is intended as a supplementary tool, and its predictions are not substitutes for professional medical advice.

Problem Specification

- The dataset we used is acquired from kaggle. It is in the format of .json.
- Our data is comprised of english words. Dataset is present in the format of key-value pairs.
- Each datapoint has a "tag" property which gives us the name of the disease and a "patterns" property which consists of array of sentences which are basically symptoms.
- For each tag property in the dataset, there will be atleast one element present in the patterns property. This way symptoms of patient will be associated with disease of patient.
- Since there are tags associated with data, and tags are nothing but labels, this is supervised machine learning model.

Dataset Overview

- Our dataset encompasses diverse diseases, their associated symptoms, and precautionary measures. It amalgamates textual descriptions of symptoms and precautions, along with a binary encoding scheme for disease representation, enabling structured disease classification for machine learning models.

Methods

- ▶ We focus on employing unsupervised learning algorithms in tandem with NLP techniques for data preprocessing. Our approach involves using K-Nearest Neighbors (KNN) and Decision Trees for prediction tasks and enhancing accuracy through ensemble learning methods.

KNN in disease symptom prediction chatbot

- ▶ Data Collection: Gather a diverse dataset comprising patient symptoms and corresponding disease labels.
- ▶ Preprocessing: Cleanse and preprocess the dataset by handling missing values, encoding categorical variables, and splitting into training and testing sets.
- ▶ Decision Tree Model: Implement a decision tree algorithm (e.g., ID3, C4.5) using libraries like scikit-learn, feeding it the training data.
- ▶ Training & Validation: Train the decision tree model on the training set and validate its performance using cross-validation techniques to fine-tune hyperparameters.
- ▶ Prediction and Chatbot Integration: Develop a chatbot interface allowing users to input symptoms, leveraging the trained decision tree to predict potential diseases and provide relevant information or recommendations.

Decision Tree for disease symptom prediction chatbot

- ▶ Data Collection: Gather a diverse dataset comprising patient symptoms and corresponding disease labels.
- ▶ Preprocessing: Cleanse and preprocess the dataset by handling missing values, encoding categorical variables, and splitting into training and testing sets.
- ▶ Decision Tree Model: Implement a decision tree algorithm (e.g., ID3, C4.5) using libraries like scikit-learn, feeding it the training data.
- ▶ Training & Validation: Train the decision tree model on the training set and validate its performance using cross-validation techniques to fine-tune hyperparameters.
- ▶ Prediction and Chatbot Integration: Develop a chatbot interface allowing users to input symptoms, leveraging the trained decision tree to predict potential diseases and provide relevant information or recommendations.

Project Design and Milestones

- ▶ Data Collection and Preprocessing: This phase involves cleaning and preparing the dataset for training the ML model.
- ▶ Machine Learning Model Development: Development of the ML model for symptom prediction, assessing accuracy via cross-validation techniques.
- ▶ NLP Integration: Implementing the NLP component for effective symptom extraction and interpretation in the chatbot.
- ▶ User Interface Development: Designing an intuitive interface for user interaction and symptom input.
- ▶ Testing and Validation: Thorough testing to ensure accurate disease identification and addressing ethical and privacy concerns.
- ▶ Documentation and Optimization: Summarizing lessons learned, creating guidelines, and ensuring the chatbot's accuracy and usability.

Programming and APIs

- ▶ Python serves as our primary programming language. We utilize Scikit-Learn, TensorFlow, and GitHub API for model development and version control. Matplotlib, seaborn, Pandas, and Numpy are key modules employed for data visualization and manipulation.

Related Projects

- ▶ We draw inspiration from previous research, such as "Symptoms Based Multiple Disease Prediction Model using Machine Learning Approach" and "The Role of Machine Learning Algorithms for Diagnosing Diseases," among others. These projects have contributed significantly to our understanding and approach in developing our chatbot.

Conclusion

- Our disease symptom prediction chatbot amalgamates cutting-edge machine learning techniques, NLP, and a comprehensive dataset to provide accurate predictions based on user-input symptoms. The project's milestones ensure a systematic approach to development, validation, and optimization, aiming to create a reliable and user-friendly tool for disease prediction.

Thank You

Harsha Vardhan Manne
Brag Mohammad Abrar
Neeraj Reddy Aluka
Devi Sumanth Pasunoori