Healthcare Symptom Checker Bot

By Gopalakrishnan Kumar, MTech IIT-Bombay,

Freelance Data Science Consultant

LinkedIn: Profile Link :
https://www.linkedin.com/in/gopalakrishnankumar-a73301110/

Github:https://www.github.com/Gopalakrishnan-Kumar/

Kaggle URL- https://www.kaggle.com/gopalkk2

Google Colab URL

https://colab.research.google.com/drive/1ZqtZ8AaGCJ2Em-MRwoj9hT66vChoePB?usp=sharing

Project Overview

The Healthcare Symptom Checker Bot is designed to help users input symptoms and receive possible medical conditions along with urgency recommendations. The goal is to make healthcare guidance more accessible while clearly differentiating between mild, moderate, and severe conditions.

Dataset Description

The dataset healthcare_symptoms_dataset.csv contains the following fields:

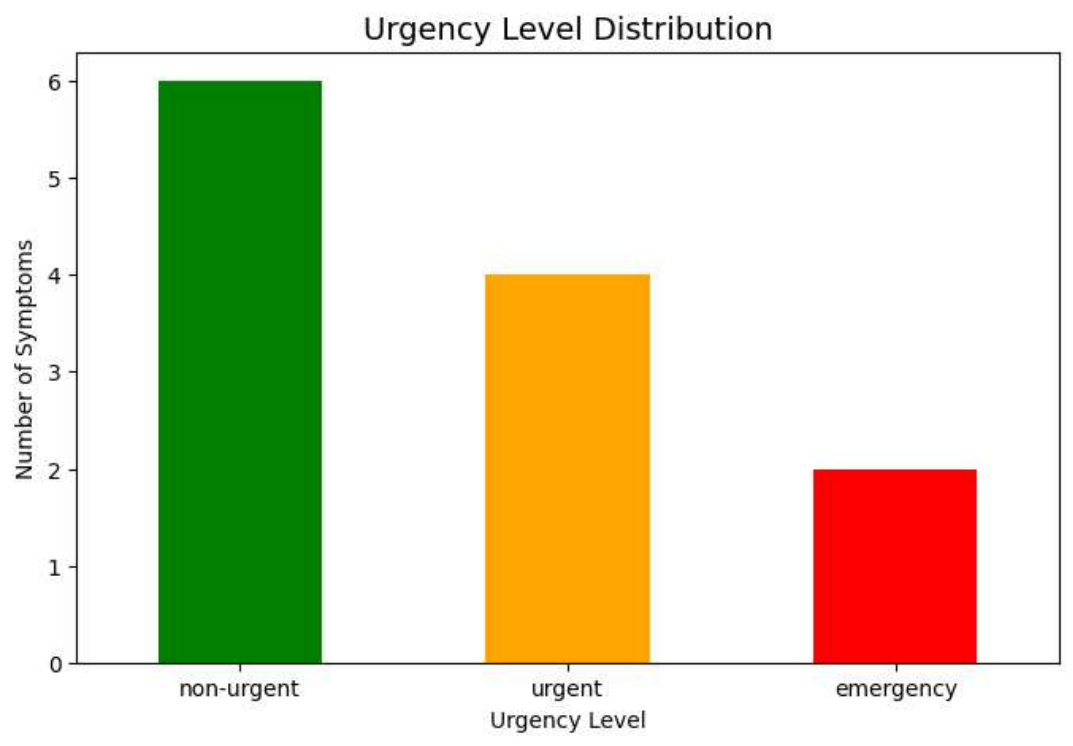
- **symptom** – Name of the symptom reported by a user (e.g., "fever", "chest pain").
- **possible_conditions** – List of potential health conditions related to the symptom, separated by semicolons.
- **urgency** – Urgency level of the symptom (e.g., "low", "medium", "high").

symptom	possible_conditions	urgency
fever	flu; malaria; dengue; COVID-19	high
headache	tension headache; migraine; sinus infection	medium
cough	common cold; bronchitis; pneumonia	medium

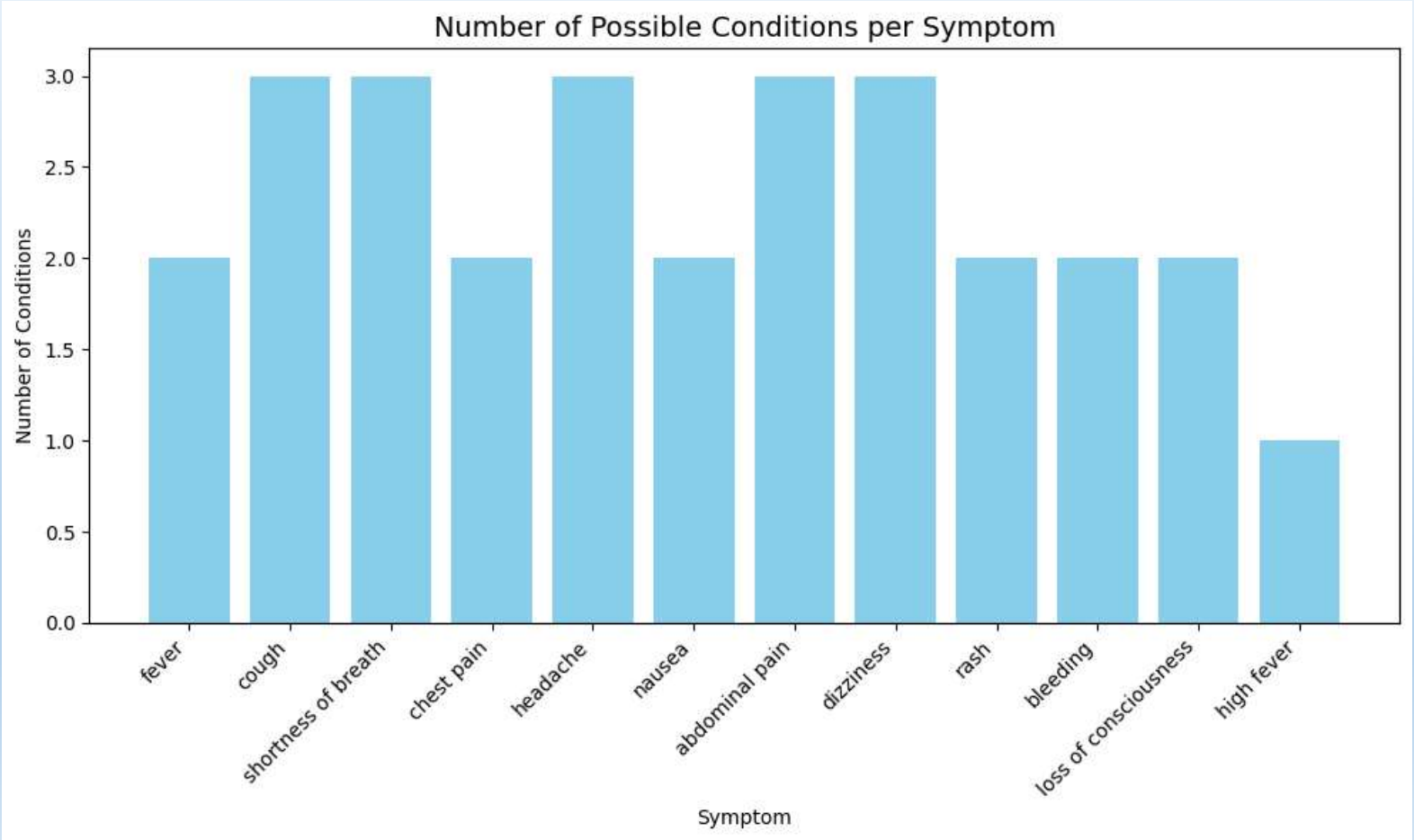
Data Processing

- **Loading Data** – The CSV file is read into a pandas DataFrame for processing.
- **Condition Count** – We calculate how many possible conditions are linked to each symptom.
- **Urgency Mapping** – We group symptoms by urgency level for visualization.
- **Text Processing** – All possible conditions are merged into one text block for the Word Cloud.

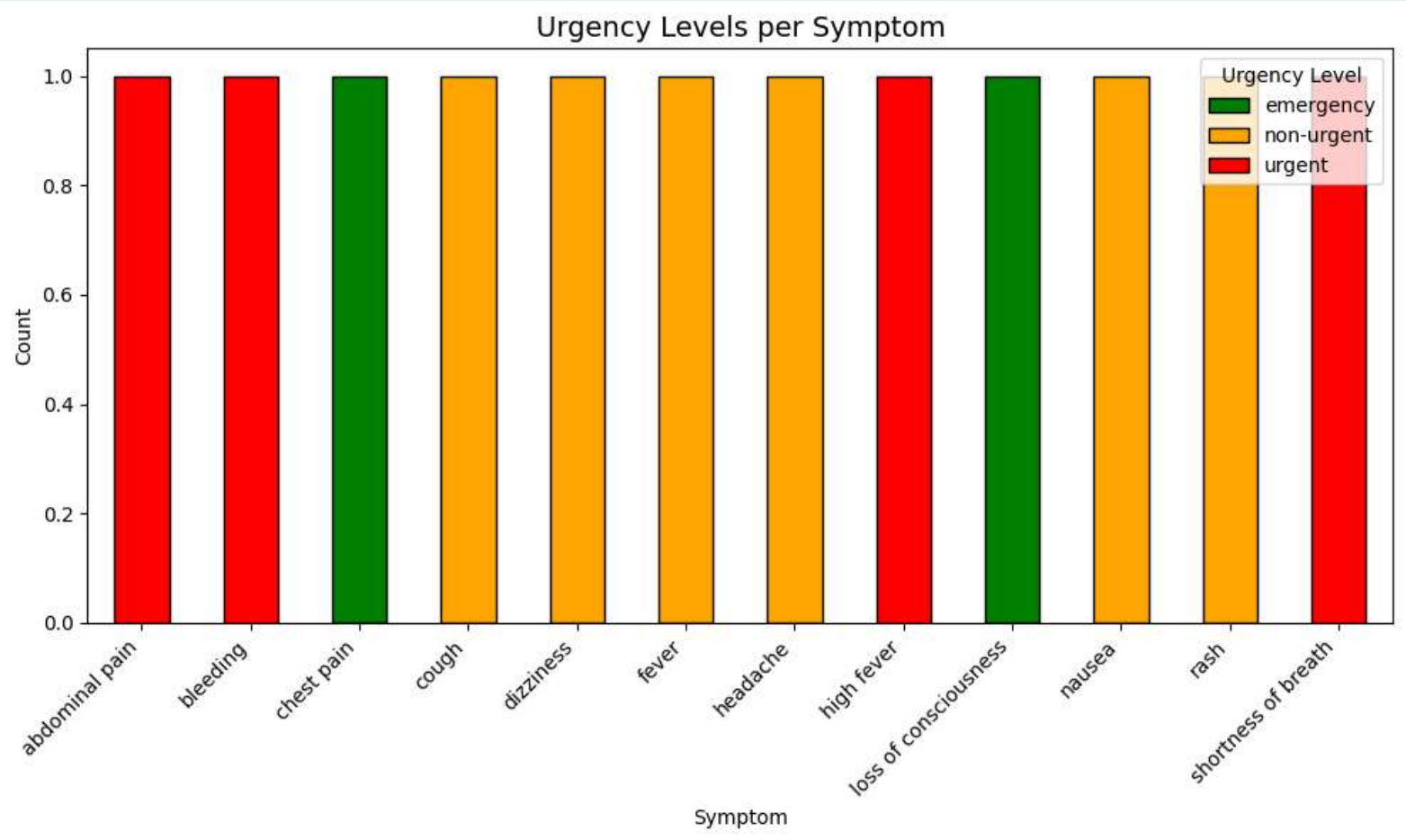
Data Visualization



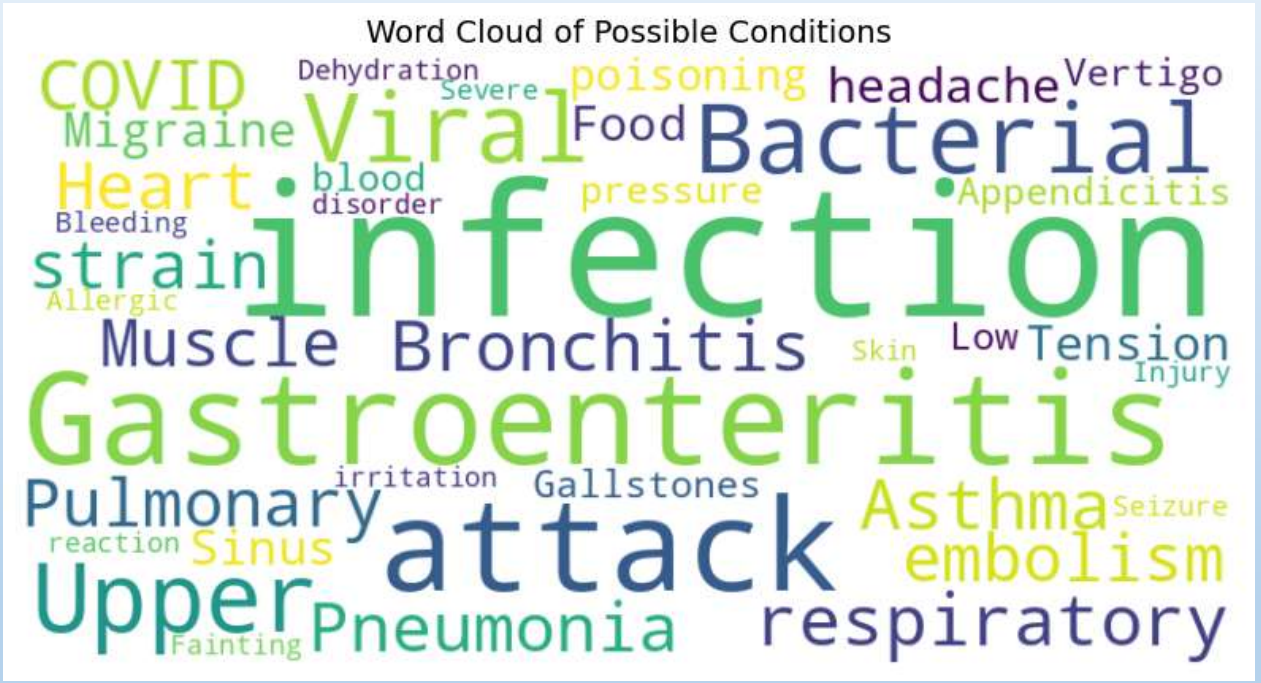
Data Visualization



Data Visualization



Data Visualization



Visualizations & Insights

4.1 Bar Chart – Number of Possible Conditions per Symptom

•A bar chart was created to show how many different conditions are linked to each symptom.

•**Insight:**

- Fever, cough, and chest pain are linked to the highest number of possible conditions.
- Some symptoms are linked to only one or two conditions, making them easier to diagnose.

4.2 Stacked Bar Chart – Urgency Distribution

•Displays urgency levels (low, medium, high) per symptom.

•**Insight:**

- Chest pain, shortness of breath, and severe headache have a high percentage of “high urgency” labels.
- Most mild symptoms fall under low or medium urgency.

Visualizations & Insights

4.3 Word Cloud – Common Medical Conditions

A word cloud was generated from all possible medical conditions.

Insight:

Conditions like “flu”, “migraine”, “bronchitis”, and “pneumonia” appear most frequently.

Shows the common illnesses people search for in symptom checkers.

Key Findings

- High urgency symptoms such as chest pain and shortness of breath require immediate attention.
- Fever is one of the most common symptoms and has multiple potential causes, making diagnosis dependent on additional factors.
- The dataset covers a mix of common and rare medical conditions, enabling broad applicability for chatbot responses.

Potential Improvements

- Integrate severity scoring based on **symptom combination patterns**.
- Add **AI-based condition prediction** using machine learning.
- Expand dataset with more **rare diseases** to improve coverage.
- Allow **multi-symptom input** rather than single symptom lookup.

Conclusion

•The Healthcare Symptom Checker Bot effectively maps symptoms to conditions and categorizes them by urgency. The combination of stacked bar charts, bar charts, and word clouds provides valuable insights for improving the bot’s diagnostic accuracy and prioritization.