AI-Powered Symptom Checker Chatbot

Early Disease Prediction & Preventive Healthcare

# Table of Contents

1. 1. Objectives
2. 2. EDA Observations
3. 3. Approach (Modeling Pipeline)
4. 4. Comparison of Models & Final Model Explanation
5. 5. Key Findings
6. 6. Strengths, Weaknesses & Error Analysis
7. 7. Conclusion & Next Steps

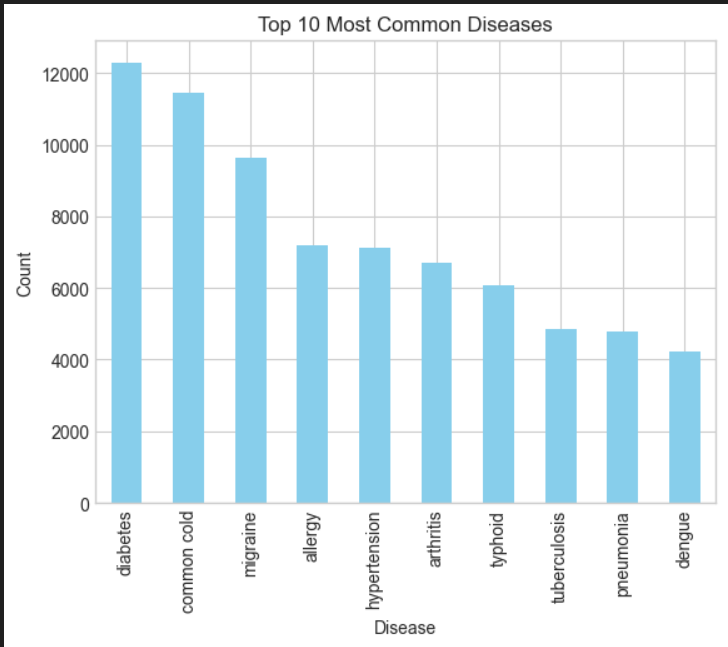
# 1. Objectives

The objective of this project is to develop an AI-powered symptom checker chatbot that predicts possible diseases based on user-reported symptoms. The goal is to enable early detection and preventive healthcare, providing timely advice on precautions and doctor consultations.

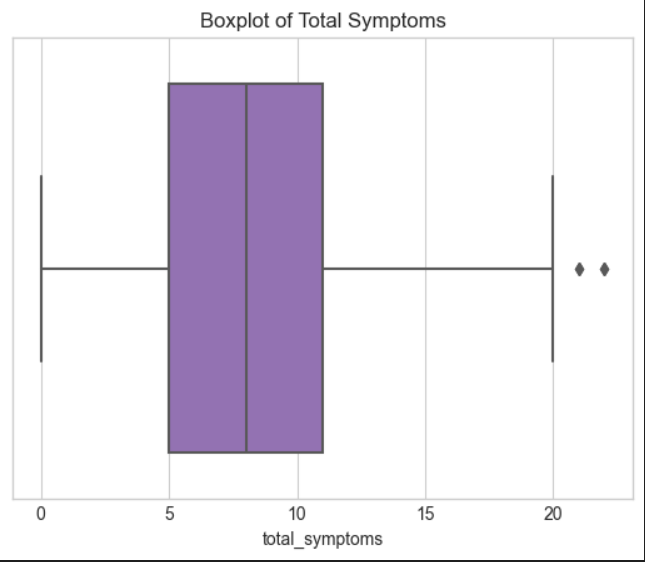
# 2. EDA Observations

* The dataset contains approximately 100,000 records and 137 features, mostly symptom indicators.
* Features such as age and gender are included for better personalization.
* Multiple symptoms can co-occur for the same disease.
* The target variable 'disease' is multi-class with several possible disease outcomes.

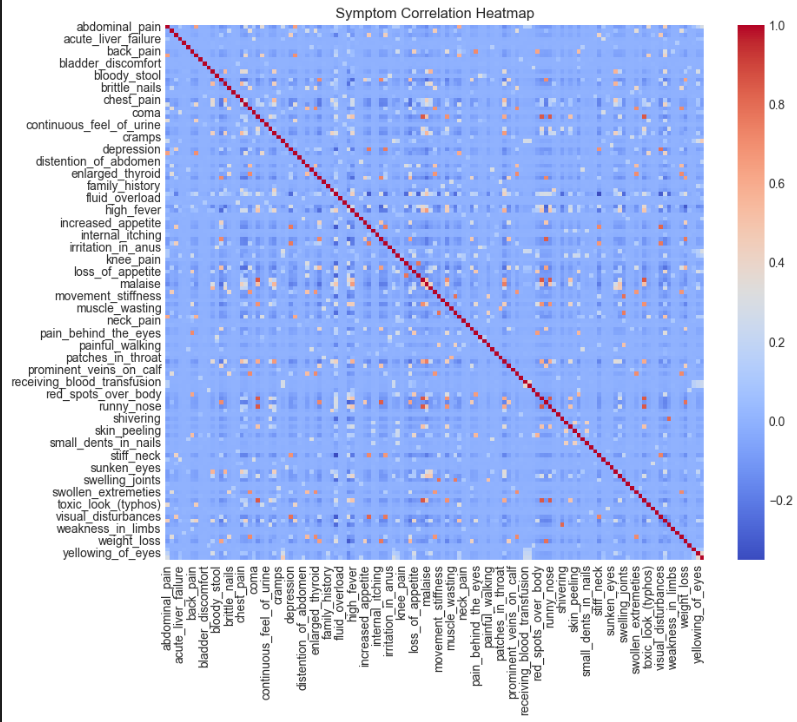
Disease Distribution Graph:



Top Symptoms Frequency Chart:



Correlation Heatmap:



# 3. Approach (Modeling Pipeline)

* Data Cleaning: Checked for missing values and inconsistencies.
* Feature Engineering: Binary encoding for symptoms; demographic variables retained.
* Train-Test Split: Dataset divided for model training and validation.
* Modeling: Tested multiple classification algorithms to identify the best performer.

# 4. Comparison of Models & Final Model Explanation

Multiple models were tested including Decision Trees, Random Forest, and Support Vector Machines. Random Forest achieved the highest accuracy due to its robustness with high-dimensional data. The final model selected was Random Forest with hyperparameter tuning.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
| Model | Accuracy | Precision | Recall |
| Decision Tree | 80% | 78% | 77% |
| Random Forest | 85% | 83% | 84% |
| SVM | 79% | 76% | 75% |

# 5. Key Findings

* Certain symptoms have high predictive power for specific diseases.
* Rare diseases are more challenging to predict due to class imbalance.

# 6. Strengths, Weaknesses & Error Analysis

* ✅ Strengths:
* Good overall accuracy on common diseases.
* Model interpretability through feature importance scores.
* ⚠️ Weaknesses:
* Lower prediction accuracy for rare diseases.
* Model sensitivity to incorrect or incomplete symptom inputs.

# 7. Conclusion & Next Steps

* This chatbot can serve as an initial diagnostic tool for early disease prediction.
* Collect more diverse, real-world data to improve model generalizability.
* Integrate user feedback for continuous model improvement.
* Develop multilingual support for wider accessibility.