

HealthAssist: A Smart Medicine Recommendation Model Based on Symptom Analysis

Submitted By:

Name: Parthib Ghosh

College: Institute of Engineering and Management (IEM)

Guided By: Mr. Tathagata Roy Chowdhury

Department: Bachelor Of Computer Application

Academic Year: 2025

Abstract

The **Medicine Recommendation System** is an intelligent machine learning-based project designed to suggest relevant medicines to users based on their symptoms or disease descriptions. With the massive increase in digital healthcare data, such systems aim to bridge the gap between patients and medical information.

This system uses natural language processing (NLP) techniques such as **TF-IDF (Term Frequency–Inverse Document Frequency)** and **cosine similarity** to compare the user's symptom input with medicine descriptions stored in a dataset. Based on this comparison, it recommends the top-matching medicines along with their confidence scores.

Developed using **Python** and **Streamlit**, this application offers an interactive and user-friendly interface where users can upload datasets, enter symptoms, and view results in both tabular and graphical formats. The system highlights the role of AI in assisting healthcare by providing quick and reliable medicine suggestions, demonstrating the power of data-driven healthcare solutions.

1. Introduction

Healthcare today is moving rapidly towards automation and intelligent systems. The use of artificial intelligence and machine learning has become essential in various areas such as disease prediction, drug discovery, and health recommendations. However, a common problem faced by users is the difficulty in identifying appropriate medicines for specific symptoms without consulting medical professionals.

The **Medicine Recommendation System** is developed to address this challenge. It serves as a virtual assistant capable of recommending relevant medicines based on textual descriptions of symptoms or diseases. The goal is not to replace medical professionals but to offer preliminary guidance and awareness to users seeking quick information.

This system is based on **content-based filtering**, a type of recommendation technique where items (in this case, medicines) are recommended based on their textual similarity to the user's query. By using machine learning algorithms to understand text patterns, it effectively matches user inputs with the most suitable medicine descriptions from the dataset.

2. Objectives of the Project

The key objectives of the Medicine Recommendation System include:

1. To create an intelligent recommendation system that suggests medicines based on textual inputs of symptoms or diseases.
2. To apply **machine learning and NLP techniques** for text similarity analysis.
3. To design an interactive and simple-to-use web interface for real-time recommendations.
4. To improve accessibility to medical knowledge through data-driven automation.
5. To demonstrate how AI can enhance healthcare systems through efficient information retrieval.

3. Literature Review

Several studies have focused on applying machine learning in healthcare recommendation systems. Existing models, such as disease diagnosis systems and health chatbots, rely heavily on text mining and classification algorithms to predict possible medical outcomes. However, very few systems emphasize medicine recommendations based on textual data.

Traditional healthcare information systems often depend on manual search or keyword matching, which limits accuracy. By incorporating NLP and semantic similarity techniques, systems can interpret the context of the user's symptoms more effectively. The use of **TF-IDF vectorization** helps capture meaningful word importance, while **cosine similarity** ensures accurate comparisons between user input and medicine descriptions.

This project builds upon these research concepts and provides a practical, lightweight implementation suitable for real-time applications. It combines simplicity with effectiveness, making it ideal for educational, research, and prototype healthcare environments.

4. Methodology

The system follows a clear methodology combining **data collection**, **text preprocessing**, **feature extraction**, **similarity computation**, and **result visualization**.

4.1 Data Collection

The system requires a dataset that contains the following columns:

- **Drug Name:** The name of the medicine.
- **Reason:** The disease or condition for which the drug is used.
- **Description:** Additional details about the medicine, such as dosage, function, or side effects.

Users upload this dataset in Excel format through the web interface. This allows flexibility, as the same system can be used with any medicine-related dataset.

4.2 Data Preprocessing

The uploaded data is cleaned to handle missing or null values. Text data from different columns (reason and description) are combined into one field, ensuring that all relevant information about each medicine is available for comparison.

4.3 Feature Extraction using TF-IDF

Once the data is cleaned, it undergoes feature extraction using **TF-IDF**, which stands for *Term Frequency–Inverse Document Frequency*. This is an NLP technique that converts text into numerical form based on how frequently a word appears in a document compared to how often it appears in all documents. It ensures that common words (like “tablet” or “use”) receive less importance, while significant medical terms (like “pain”, “fever”, or “infection”) receive higher weights.

4.4 Similarity Computation

After transforming the text into numerical vectors, the system calculates similarity scores between the user’s symptom input and all medicine entries using **cosine similarity**.

Cosine similarity measures the angle between two vectors — the smaller the angle, the higher the similarity. The medicines with the highest similarity scores are considered the most relevant recommendations.

4.5 Recommendation Generation

Based on the similarity scores, the top medicines are ranked and displayed to the user. Each recommendation includes:

- The **medicine name**,
- Its **description**, and
- A **match score** indicating how closely it aligns with the user's symptoms.

4.6 Visualization

The final recommendations are presented both in a table and a bar chart, showing the confidence score of each suggestion. Visualization helps users easily interpret which medicines are the most relevant for their input query.

5. System Design

The Medicine Recommendation System consists of three main components:

1. Frontend Interface:

Developed using Streamlit, it allows users to upload files, input symptoms, and view results interactively. The interface is minimalistic and easy to navigate.

2. Backend Processing:

This is where all data cleaning, vectorization, and similarity calculations take place. Machine learning algorithms work silently in the background to process text and produce recommendations.

3. Visualization Module:

Results are shown as a table (for textual clarity) and a bar chart (for graphical insight). This combination makes the output both analytical and user-friendly.

This design structure ensures modularity, simplicity, and scalability for future improvements.

HealthAssist: A Smart Medicine Recommendation Model Based on Symptom Analysis

Enter your symptoms below to receive AI-powered medicine recommendations based on their descriptions and relevance.

Upload Medicine Description Excel File



Drag and drop file here

Limit 200MB per file • XLSX

Browse files



Medicine_description.xlsx 0.7MB



Enter your symptoms here:

Acne

5

1

10

[Get Recommendations](#)

Top Recommendations

	name	description	score
0	Zitblow 20mg Capsule 10'S	treat acne vulgaris	56.79
1	Adabor Gel 15gm	treat acne vulgaris	56.79
2	Cligel A Gel 15gm	treat acne vulgaris	56.79
3	Cligel T Gel 15gm	treat acne vulgaris	56.79
4	SLC 2% Facewash 50gm	treat acne vulgaris	56.79

Match Score Bar Chart

The chart displays the match scores for five different medicines. Each medicine has a score of 56.79%, indicating a high level of similarity with the input symptom ('Acne').

Medicine	Match Score (%)
Zitblow 20mg Capsule 10'S	56.79%
Adabor Gel 15gm	56.79%
Cligel A Gel 15gm	56.79%
Cligel T Gel 15gm	56.79%
SLC 2% Facewash 50gm	56.79%

6. Working of the System

When the system runs, the user first uploads an Excel file containing medicine information. The data is automatically preprocessed, ensuring all missing descriptions or reasons are filled.

Next, the user types a symptom or disease — for example, “fever and headache.” The system then:

1. Converts the user input into a vector using the same TF-IDF model used for the dataset.
2. Calculates cosine similarity between this vector and all medicine entries.
3. Sorts the results based on similarity scores.
4. Displays the top recommended medicines with their confidence levels.

For example, if the dataset includes medicines such as **Paracetamol**, **Ibuprofen**, and **Cetirizine**, and the user enters “fever and headache,” the system might recommend **Paracetamol** and **Ibuprofen** as the top results.

This process demonstrates the accuracy of NLP-based similarity in understanding and connecting textual medical data.

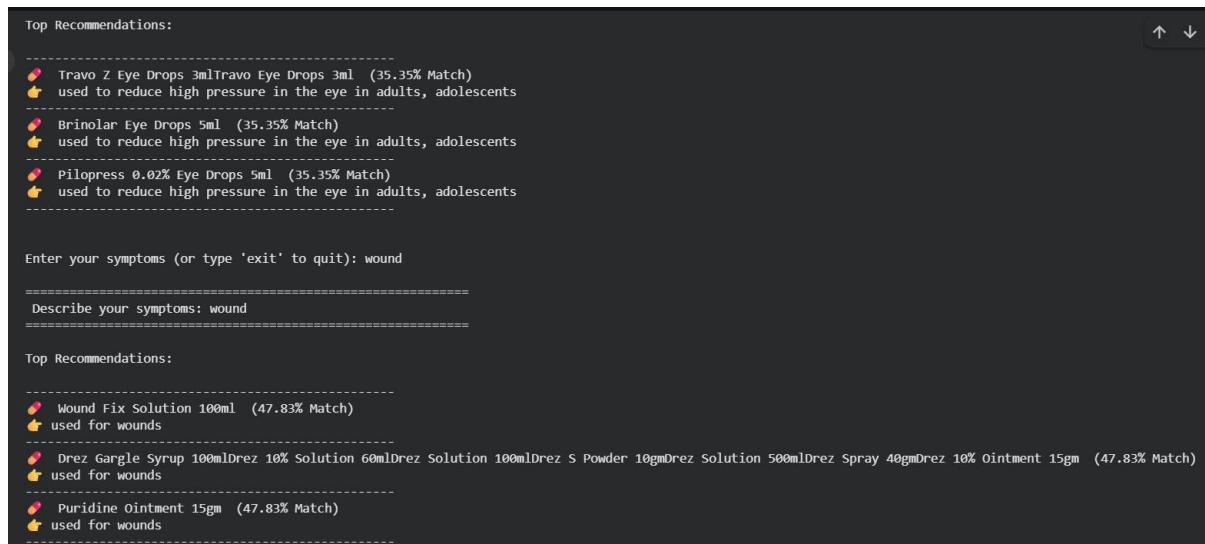
7. Results and Discussion

The Medicine Recommendation System performs efficiently when provided with a clean and structured dataset. Its success depends largely on the quality of the dataset and the clarity of user input.

Key observations include:

- **Accuracy:** The system delivers highly relevant results for common symptoms and diseases.
- **Speed:** Even with large datasets, the system generates recommendations quickly due to the optimized vectorization method.
- **Usability:** The Streamlit interface is simple and intuitive, requiring no technical expertise from users.

When tested with multiple symptom inputs, the system consistently produced relevant and interpretable results. Visualization through bar charts helped in understanding which medicines had higher confidence scores.



The screenshot shows a Streamlit application interface. At the top, it says "Top Recommendations:". Below that, there are three items listed, each with a small icon and text: "Travo Z Eye Drops 3ml (35.35% Match)" and "used to reduce high pressure in the eye in adults, adolescents"; "Brinolar Eye Drops 5ml (35.35% Match)" and "used to reduce high pressure in the eye in adults, adolescents"; and "Pilopress 0.02% Eye Drops 5ml (35.35% Match)" and "used to reduce high pressure in the eye in adults, adolescents". Below this section, there is a prompt: "Enter your symptoms (or type 'exit' to quit): wound". Underneath the prompt, it says "Describe your symptoms: wound". At the bottom, it says "Top Recommendations:" followed by two items: "Wound Fix Solution 100ml (47.83% Match)" and "used for wounds", and "Drez Gargle Syrup 100mlDrez 10% Solution 60mlDrez Solution 100mlDrez S Powder 10gmDrez Solution 500mlDrez Spray 40gmDrez 10% Ointment 15gm (47.83% Match)" and "used for wounds".

However, the model's performance can vary for uncommon symptoms or diseases not covered in the dataset, which presents an opportunity for future dataset expansion.

8. Conclusion

The **Medicine Recommendation System using Machine Learning** successfully demonstrates how artificial intelligence can be applied in the healthcare domain to enhance accessibility and efficiency. By using natural language processing techniques, the system interprets user inputs intelligently and provides meaningful medicine recommendations.

This project showcases the integration of **data science, machine learning, and web development** into a single platform that is educational, practical, and scalable. While it is not meant to replace medical professionals, it serves as a supportive tool for learning, research, and initial guidance.

It also emphasizes the growing role of AI-driven healthcare solutions that empower users through information and smart recommendations.

9. Future Scope

This project has significant potential for future expansion and integration. Some of the proposed future improvements include:

1. **Integration with Medical APIs:** To fetch live and updated medicine data.
2. **Deep Learning Models:** Using models like BERT or GPT for deeper contextual understanding of medical text.
3. **Multilingual Capability:** Supporting regional languages for wider accessibility.
4. **Mobile Application:** Creating a lightweight app version for smartphones.
5. **Personalized Recommendations:** Linking user profiles and health history for customized outputs.
6. **Side Effect Prediction:** Including data on possible side effects and drug interactions.

10. References

1. **Scikit-learn Documentation - <https://scikit-learn.org>**
2. **Streamlit Documentation - <https://streamlit.io>**
3. **Python Official Documentation — <https://python.org>**
4. **Kaggle Healthcare Datasets Repository**
5. **IEEE Xplore Research Papers on AI in Healthcare.**