MEDICINE RECOMMENDATION SYSTEM USING COSINE SIMILARITY

[1]Mr. A. Selvaraj, [2]Ms. P. Ponkalpana, [3]Ms. S. Nazeeha Sherin, [4]Ms. G. Sri Harini

[1]Assistant Professor (Sr. Gr), [2,3,4]Final Year Students
Department of Information Technology
K. L. N College of Engineering

Abstract- **With the ever-increasing complexity of healthcare data and the need for personalized treatment approaches, there is a growing high demand for efficient and accurate medicine recommendation systems. The approach involves representing medicines and symptoms as vectors in a high-dimensional space based on relevant features such as medical history, symptoms, and medication interactions. By calculating the cosine similarity between these vectors, we can effectively measure the similarity between medicines and patient's symptoms. Through experiments with real healthcare datasets, the system demonstrates superior performance compared to traditional methods. With scalability, interpretability, and adaptability to diverse healthcare contexts, this system holds promise for enhancing treatment outcomes and personalized medicine recommendation.**

Keyword: **Medicine Recommendation, Cosine Similarity, Vectorizer.**

I. INTRODUCTION

In recent years, the integration of technology into healthcare sector has led to the development of various innovative solutions aimed at improving patient care and treatment outcomes. [1] Traditional recommendation methods such as rule-based systems and collaborative filtering often rely on predefined rules or user-item interaction data to generate recommendations. [3] Rule-based systems typically involve manually crafted rules or heuristics that dictate which medications to recommend based on specific criteria or conditions. The solution is the Medicine Recommendation System, which utilizes similarity-based algorithms to provide personalized medication recommendations based on the patient's symptoms, medical history, and other relevant factors. This system leverages the power of machine learning and natural language processing techniques to analyze large datasets of medical records and recommend the most suitable medications using the similarities of the previous recommendations for individual patients. The primary goal of the Medicine Recommendation System is to assist healthcare professionals in making informed decisions regarding medication prescriptions, ultimately enhancing the quality of patient care and reducing the risk of adverse drug reactions. By analyzing similarities between patients' symptoms and medical profiles, the system can [2] identify patterns and associations that may not be immediately apparent to human clinicians. This enables the system to recommend medications that are not only effective in treating the patient's condition but also tailored to their specific needs and preferences for the medical students and professionals. Furthermore, the cosine similarity-based recommendation system offers a transparent and interpretable approach to medication recommendation. [5] By quantifying the similarity between medications based on their textual descriptions, healthcare professionals can understand the rationale behind each recommendation and make informed decisions based on their clinical expertise and patient needs. This transparency is essential in medical applications where trust, accountability, and interpretability are paramount.

II. RELATED WORKS

Recommendation systems are essential in the healthcare industry, especially when they are designed to handle the particular difficulties of medical applications like medication and therapy suggestions. Because it uses mathematics to measure the similarity of [2] vectors in a multidimensional space, cosine similarity stands out above other approaches for its efficacy in managing textual data, such as medical descriptions and patient records. Notwithstanding its widespread application in disease diagnosis and medication interaction prediction, the technique has certain drawbacks, such as sensitivity to high dimensionality and a lack of deeper semantic understanding. Using cutting-edge NLP approaches or hybrid models that [4] combine different recommendation algorithms could significantly improve the performance of cosine similarity-based systems. These developments may result in more precise, dependable, and context-aware.

EXISTING SYSTEM

On the basis of previous datasets, rule-based or [3] collaborative filtering techniques are used by traditional medical recommendation systems to produce recommendations. [6] Collaborative filtering may be hampered by data sparsity and the initial cold start issue, even if it might be useful in capturing user preferences and commonalities. Conventional medicine recommendation systems have played a key role in supporting medical practitioners in their decision-making when it comes to prescribing medications. [2] The filtering methods make use of the drug names' history database. Based on predetermined principles, the current system recommends [1] Chinese medications using the Generative Adversarial Technique. It was used to suggest Chinese medications and therapeutic approaches.

DRAWBACKS OF EXISTING SYSTEM

- Data Sparsity
- Complexity and Implementation Challenges
- Cold Start Problem
- Limited Personalization

PROPOSED SYSTEM

Medication recommendations using cosine similarity represent a new way to inform treatment [3] decisions based on a patient's symptoms and medication list. Compared to traditional methods such as law-based or integral methods, the cosine-like method has simplicity, efficiency and conciseness First, the cosine-like proposition works by measuring the similarity between different drugs and treatments based on descriptive terms. [2] By using cosine similarity to measure similarity between drugs, the system can identify drugs with similar effects or mechanisms of action. This allows the system to recommend medications that are not only effective in a particular treatment, but are also relevant to the patient's needs and preferences. Additionally, the cosine similarity-based approach to drug recommendation uses drug description or behavior to calculate the similarity between different treatments. The system uses cosine similarity, a mathematical technique to measure the similarity between two vectors to identify medications related to medical conditions. Drug recommendation using cosine similarity is a data-driven approach designed to help physicians and patients select appropriate medications based on similarity between drug and patient characteristics.

BENEFITS OF PROPOSED SYSTEM

- Support for Clinical Decision-Making
- It offers Personalized Recommendations
- The recommendation system is scalable and adaptable to diverse medical conditions

DATAFLOW DIAGRAM

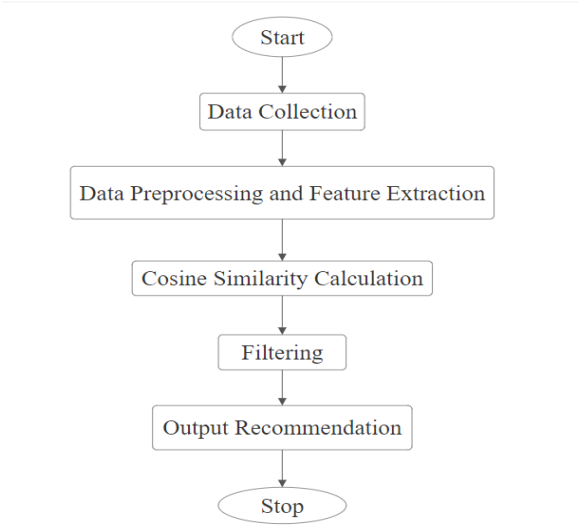

Fig1 Data Flow of the system

All processes in the dataflow diagram are continuous and clearly show how the user's input changes from various stages to the final recommendation using the mathematical process of cosine similarity, such as finding the best match.

III. METHODOLOGY

1. Drugs and patient profiles are represented as feature vectors in a high-dimensional semantic space, it captures the intrinsic properties of drugs and description, enabling meaningful comparisons and recommendations.
2. Cosine similarity is calculated between the feature vectors of drugs and patient profiles to quantify their similarity.
3. Cosine similarity scores are thresholded to filter out drugs with low similarity scores, ensuring that only relevant medications are included in the recommendations.

IV. MODULES

1. DATA COLLECTION

[1] Gathering pertinent data from various sources, such as drug databases, clinical trial data and medical databases, is the initial step in the process. These resources include details on medication characteristics, medical background, signs, and treatment results.

2. DATA PREPROCESSING

Preprocessing is done on the gathered data in order to [3] eliminate noise, standardize formats, and extract pertinent features. This comprises text processing methods that transform textual data into numerical representations appropriate for analysis, like vectorization, tokenization, and stemming. There are many different kinds of [2] vectorizers. The text input is transformed into a matrix of token counts by this system using the scikit-learn count vectorizer module. The text data is transformed into a numerical matrix that shows how frequently each word appears in each document.

3. COSINE SIMILARITY CALCULATION

After preprocessing, the [2] Count Vectorizer is used to convert the text into numerical vectors. Because it transforms text data into a structured, fixed-length numerical representation, this transformation is essential. In order to keep dimensionality manageable and concentrate on the most important attributes, Count Vectorizer counts the frequency of each word (token) that appears in the text. To make sure that the vectors represent meaningful material rather than frequently used but unnecessary words, common stop words are eliminated.

The fundamental step in computing cosine similarity is to take the dot product of the vectors that represent each drug and normalize it using the product of the vector norms (magnitudes). In terms of math, this is stated as:

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}$$

Here, A and B are the feature vectors of two drug descriptions. The dot product $A \cdot B$ measures the sum of the products of corresponding elements from the vectors, providing a scalar output that quantifies the degree of term overlap between the descriptions. The normalization factor $\|A\|\|B\|$, where $\|A\|$ and $\|B\|$ are the Euclidean norms of the vectors, adjusts for the length of the descriptions, ensuring that the similarity measurement is not biased by the length of the text but rather reflects the genuine content similarity.

The resulting cosine similarity scores fall into three categories: orthogonality (no similarity), diametrically opposite vectors (perfect dissimilarity), and identical direction vectors (perfect similarity), with a score of -1 denoting no similarity. In actuality, the scores for positive count vectors usually fall between 0 and 1. The recommendation algorithm in question generates a similarity matrix by comparing the vectors of each drug in the dataset to all other vectors.

Using the same text preparation and vectorization procedure, the system initially turns a given query—let's say a user input like "Headache Migraine Fever"—into a vector. Next, it compares the cosine similarity of each drug's vector in the dataset. The medications whose descriptions most closely match the user's input are recommended by the system, which chooses the medications with the highest similarity scores, suggesting their possible efficacy for the illness stated. Based on thorough text data analysis, this approach enables context-aware, nuanced medication suggestions that can greatly improve the user's capacity to choose appropriate therapies.

```
similarity

array([[1.        , 0.25197632, 0.43643578, ..., 0.        , 0.        ,
        0.        ],
       [0.25197632, 1.        , 0.25660012, ..., 0.        , 0.        ,
        0.        ],
       [0.43643578, 0.25660012, 1.        , ..., 0.        , 0.        ,
        0.        ],
       ...,
       [0.        , 0.        , 0.        , ..., 1.        , 0.1490712 ,
        0.1490712 ],
       [0.        , 0.        , 0.        , ..., 0.1490712 , 1.        ,
        1.        ],
       [0.        , 0.        , 0.        , ..., 0.1490712 , 1.        ,
        1.        ]])
```

Fig2 Cosine Similarity Calculation

4. DATA VISUALIZATION AND EVALUATION

Data visualization is essential to the pharmaceutical recommendation system project's ability to communicate results in a way that is both comprehensible and visually appealing. For instance, a [5] heatmap works especially well since it provides a visual representation of data matrices with color-coded systems. The cosine similarity scores of various medications can be shown in the context of this research using a heatmap. This visual aid makes it easier for stakeholders to identify possible substitutes or related medications by rapidly identifying which medications are most comparable based on their descriptions.

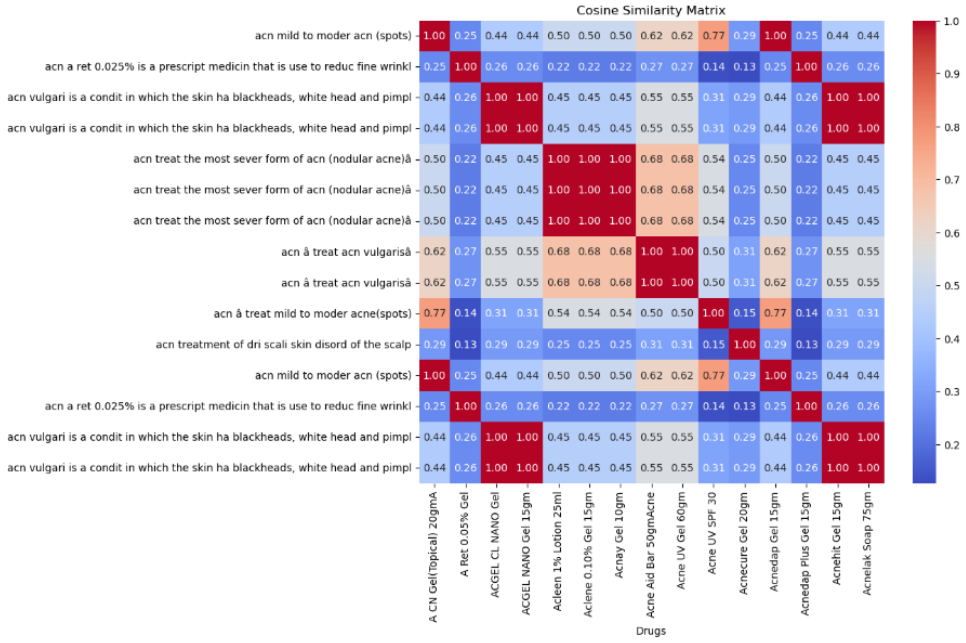

Fig3 Heatmap

5. RECOMMENDATION

The algorithm recommends a drug for a certain medical complaint based on the computed cosine similarity scores. [6] Recommendations give priority to drugs with greater cosine similarity scores since they are thought to be more comparable to the medicine under question. The advice was created in an intuitively designed online application. A place where everyone may submit their symptoms and independently determine the name of the medication, and where medical experts and students can utilize it for practice and study to provide a [3] more accurate medical description.

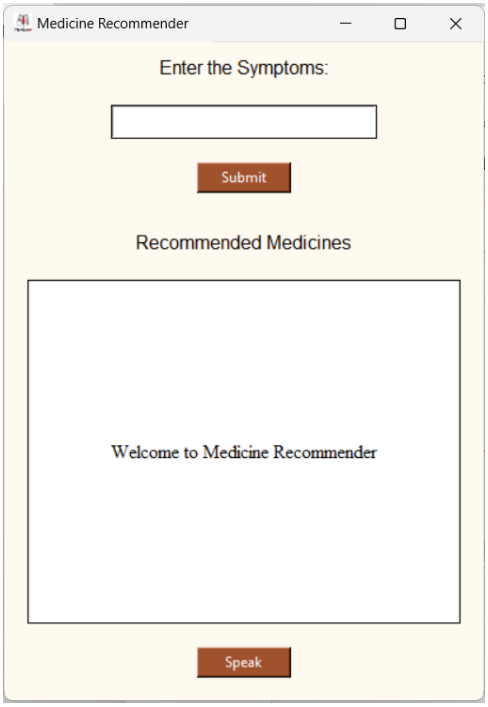

Fig4 Web Application

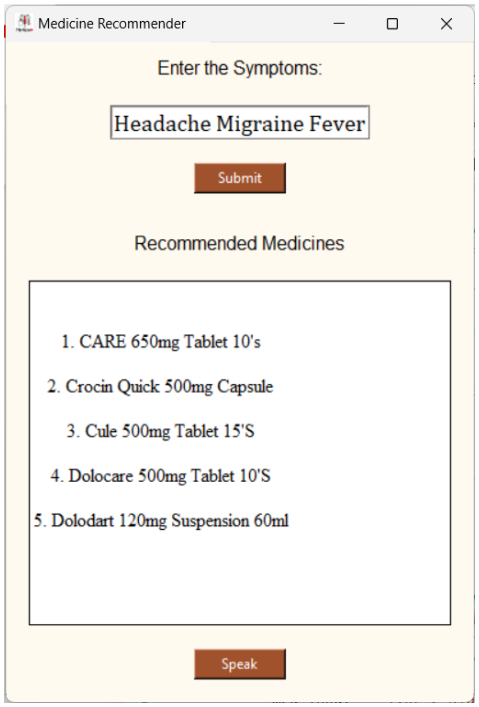

Fig5 Medicine Recommendation System

CONCLUTION

The medicine recommendation system employing cosine similarity offers a sophisticated solution for personalized drug selection.

This approach enhances treatment decision-making, improves patient outcomes, and fosters evidence-based medicine practices. By empowering healthcare professionals with data-driven insights, the system contributes to enhanced patient care, satisfaction, and overall healthcare efficacy. Its fusion of advanced computational methods with medical expertise marks a significant step towards personalized and optimized drug therapy management.

REFERENCES:

1. Chuitian Rong, Xueyan Li, Xuemei Sun, and Huabo Sun, “Chinese Medicine Prescription Recommendation Using Generative Adversarial Network” in Journals & Magazines IEEE Access Volume: 10, 14 January 2022.
2. Md Nadeem Noori, Jameel Ahamed, and Mumtaz Ahmed, “Matrix Factorization and Cosine Similarity Based Recommendation System for Cold Start Problem in E-Commerce Industries” in International Journal of Computing and Digital Systems, ISSN (2210-142X), Int. J. Com. Dig. Sys. 15, No. 1, Feb – 2024.
3. S. Gousiya Begum1, P. Kiran Sree, Drug Recommendations Using A “Reviews and Sentiment Analysis” by A Recurrent Neural Network in Indonesian journal of multidisciplinary science June 25th, 2023.
4. Luis Fernando Granda Morales, Priscila Valdiviezo-Diaz, Ruth Reategui, Luis Barba-Guaman, “Drug Recommendation System for Diabetes Using a Collaborative Filtering and Clustering Approach” in J Med Internet Res 2022, vol. 24, iss. 7, e37233.
5. Mahima Mohapatra, Mamata Nayak, Saswati Mahapatra, “A Machine Learning based Drug Recommendation System for Health Care” in Gratuate Research in Engineering and Technology, Volume 1, issue 6, Application of Intelligent Computing and Big Data Analytics in Healthcare, May 2022.
6. Ruiyang Wang and Tao Li, “Collaborative filtering recommendation algorithm based on user behavior and drug semantics” in J. Phys.: Conf. Ser. 1802 032005, 2021.