Generalized Medicine Recommendation

Soumik Banerjee
Department of Mathematics,
School of Advanced Sciences,
Vellore Institute of Technology,
Vellore, India, 632014
soumik.banerjee2023@vitstudent.ac.in

Soumya Basu
Department of Mathematics,
School of Advanced Sciences,
Vellore Institute of Technology,
Vellore, India, 632014
soumya.bas2023@ vitstudent.ac.in

***Abstract*—Our "Generalized Medicine Recommendation" project is an innovative endeavor aimed at improving healthcare delivery by harnessing the power of data-driven insights. In the ever-evolving landscape of medical treatment, this project seeks to develop a versatile recommendation system that aids healthcare providers in making informed and personalized medication choices for their patients. Rather than focusing on a single medical condition or patient demographic, our approach is "generalized," designed to cater to some extent of broad spectrum of healthcare scenarios. We aim to create a user-friendly interface that empowers healthcare professionals with evidence-based insights, facilitating more precise and patient-centric treatment decisions. Leveraging advanced machine learning algorithms and a diverse dataset which contains information about, BMI, height, weight, age, blood groups, patient's medical history, allergies, and current health conditions to provide comprehensive medication recommendations.**

Keywords—Generalized Medicine Recommendation, Medical treatment, Healthcare providers, Patient demographic, Evidence-based insights, Diverse dataset, BMI, Medical history, Current health conditions, Medication recommendations.

Introduction

The project focuses on the health care data analysis and medication recommendation, as we are trying to automate the process of getting medications for common disease such as fever, cough/cold and gastric problems etc., for the common people who live in rural areas or have less/ no knowledge about the medicines. There are several recommendation systems available in the literature, few of them have been considered in this project as per the relevance of the topic. In [1], review of the state of art in medicine recommendation system have been considered with various types of recommendations systems, their advantages and disadvantages and open challenges. Varun et. al [2] proposed a medicine recommendation system based on collaborative filtering and content-based filtering techniques using patient medical records and drug information to recommend medicines. A system that can suggest care with selecting an exact disease to the patient has been discussed in [3]. In [4] authors used a stacked ANN model to improve the fairness and safety of treatment for infectious diseases. [5] proposes a medical recommender system based on continuous-valued logic and multi-criteria decision operators, using interpretable neural networks. This project aims to develop a medication recommendation system and an interactive dashboard for users which will suggest medication.

The rest of the article is presented as follows. Introductory section is followed by section 2 data collection and processing of data. Section 3 consists of designing and modelling of the project, section 4 consists of data extraction followed by model development in section 5 and testing in section 6.

1. Requirements gathering

1.1 Data Preprocessing

Data preprocessing serves as a crucial cornerstone for the subsequent stages of data analysis and modelling. It is imperative to have data that is both clean and well-organized, and where encoding has been appropriately applied. This foundational step is indispensable when dealing with healthcare data analysis. It encompasses the necessary procedures to get raw medical data ready for in-depth analysis and modelling, which encompasses steps like data cleaning and encoding.

1.1.1 Data Cleaning

The initial stage involves rectifying data irregularities and errors. In this project, this is accomplished by excluding rows that contain missing information in vital columns, such as 'Height,' 'Age,' 'Weight,' 'Gender,' 'Blood Group,' and 'Blood Pressure.' By eliminating these rows, the dataset is guaranteed to be free from incomplete or questionable data.

1.1.2 Data Encoding

Healthcare datasets commonly consist of categorical attributes like gender or blood group. To make them suitable for efficient processing by machine learning models, it is necessary to transform these attributes into numerical representations. The provided code illustrates the utilization of label encoding as one method for translating categories into numeric values.

1.2 BMI Calculation

We compute the Body Mass Index (BMI) for individuals, which provides a numeric assessment of body fat derived from height and weight. This calculated BMI acts as a significant health metric, allowing for a more detailed evaluation of weight-related health consequences.

1.3 Data Saving

Following these preprocessing procedures, the refined data is stored as 'new_med.csv,' providing a stable foundation for future analysis and modeling, eliminating the necessity for re-preprocessing.

1.4 Statistical Overview

Basic statistical metrics, encompassing mean, standard deviation, and quartiles, are computed and presented to provide an initial dataset understanding. These summaries serve as a foundation for in-depth analysis, guiding subsequent insights into patient demographics and health metrics.

2. Design and Modelling

In the "Design and Modeling" phase, the dataset undergoes comprehensive analysis to unearth valuable insights and build predictive models for medical conditions and medications. This phase is pivotal for deriving insights, including patterns related to BMI, height, gender, blood groups, allergies, age, and common health issues, which can inform healthcare decisions, public health policies, and future predictive modeling.

2.1 BMI category analysis

2.1.1 BMI Calculation: BMI is calculated for each person based on their weight and height, categorizing them as underweight, normal weight, overweight, or obese, providing essential insights into their weight-related health.

$$BMI = \frac{Weight\ (in\ kgs)}{Height^2 (in\ m)}$$

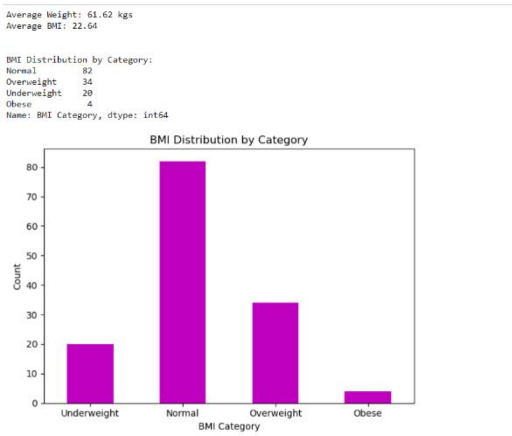


Fig 2.1. BMI Distribution by Category

2.1.2 Analysis: The analysis examines how individuals are distributed across BMI categories, offering insights into the population's overall weight-related health status. It's valuable for identifying obesity or underweight prevalence, aiding healthcare professionals and policymakers in addressing weight-related health concerns.

2.2 Height category analysis

2.2.1 Height categorization: Height values are grouped into categories like "Very Short," "Short," "Medium," "Tall," and "Very Tall" to evaluate the distribution of these height categories in the dataset.

2.2.2 Analysis: The analysis visualizes the distribution of individuals across these height categories, offering insights into the height diversity in the population and potential correlations with health conditions. This can help identify associations between height categories and health risks or advantages.

2.3 Gender and blood group analysis

2.3.1 Gender Distribution: The dataset is scrutinized to assess the gender distribution among participants, aiming to uncover potential gender-related health patterns or variations.

2.3.3 Blood group distribution: Likewise, the project investigates and visually represents the distribution of blood groups. This analysis is essential for evaluating potential genetic or blood group-related health attributes or risks within the dataset.

2.3 Allergies analysis

.3.1 Common allergies: The project detects and illustrates prevalent allergies among participants. This analysis aids in comprehending common sensitivities within the population, improving healthcare management, and providing recommendations, particularly for individuals with known allergies.

2.4 Age distribution analysis

2.4.1 Age distribution: The project scrutinizes and visually represents age distribution within the dataset, offering insights into participant demographics across various age groups.

2.5 Symptoms and health issues analysis

2.5.1 Common symptoms: The project pinpoints and visually displays prevalent symptoms and health issues encountered by participants, supporting the comprehension of widespread health concerns. This information aids in evaluating healthcare needs, trends, and priorities, facilitating effective healthcare resource allocation.

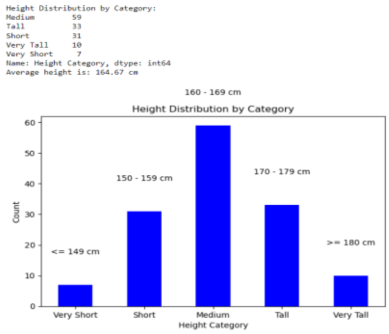


Fig 2.2. Height Distribution Category

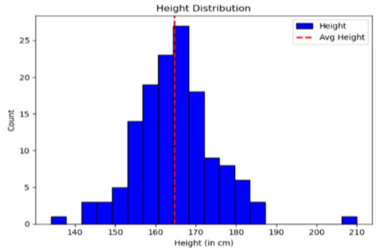
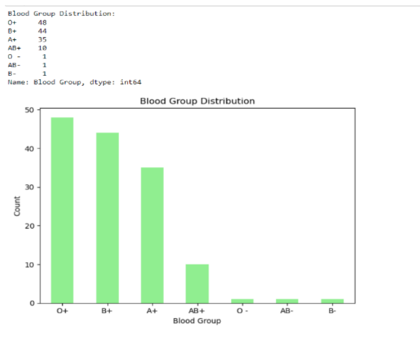


Fig 2.3. Average Height Distribution

Fig 2.4. Gender Distribution

.3. Data Extraction and Cleaning

Medication recommendation model- A dedicated dataset, "Dataset 2," is crafted specifically for conditions and associated medications. The "Data Extraction and Cleaning" stage for the Medication Recommendation Model concentrates on extracting pertinent data, purifying it to remove irregularities, and storing the refined dataset. This dataset plays a pivotal role in crafting a precise medication recommendation model, offering healthcare professionals and patients a valuable tool for making informed medication decisions based on distinct health conditions.



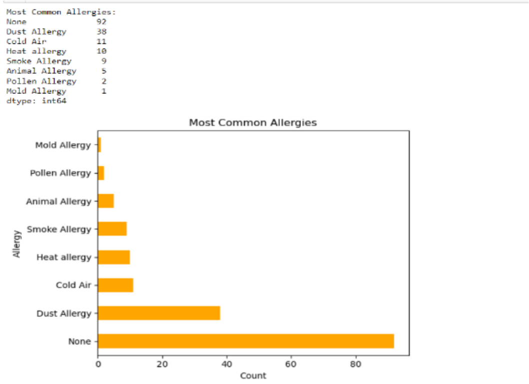


Fig 2.6..Most Common Allergies

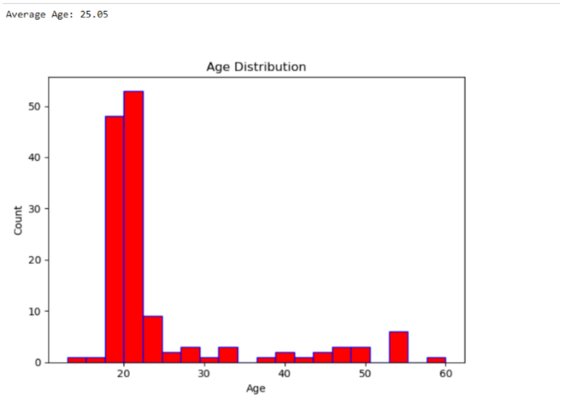


Fig 2.7. Age distribution

3.1 Data Extraction

3.1.1 Objective: The main aim is to extract pertinent information for the medication recommendation model from the original dataset, focusing on individuals' health conditions and the associated medications.

3.1.2 Extraction process: Extracting information about conditions and their medications simplifies the dataset by retaining only essential variables needed for training the medication recommendation model.

3.2 Data Cleaning

3.2.1 Objective: Data cleaning is crucial for refining the newly extracted dataset, eliminating irrelevant entries, inconsistencies, or errors. This step ensures the dataset used for modeling is accurate and effective in making medication recommendations.

3.2.2 Cleaning process: Redundant and irrelevant entries are removed, ensuring the dataset is streamlined for optimal performance of the medication recommendation model. Inconsistencies in data formats or values are also addressed to maintain data integrity.

3.3 Dataset Saving

3.3.1 Objective: After extraction and cleaning, the refined dataset is saved as 'conditions_and_medications.csv.' This facilitates convenient access and efficient data input for the subsequent modeling phase.

3.3.2 Importance: Saving the cleaned dataset is crucial to minimize repetitive data extraction and cleaning. It maintains dataset consistency for model development and facilitates its use in subsequent iterations and updates.

4. Model Development

The "Model Development" phase involves creating and evaluating a Decision Tree model for medication recommendation. Integrated into a user-friendly Dash framework, this facilitates interactive user input and real-time medication predictions, offering a valuable tool for healthcare guidance.

4.1 Decision Tree Model

4.1.1 Model creation:

4.1.1.1 Dataset loading and encoding: The saved dataset is loaded, and encoding techniques are applied to convert categorical data into a machine-learning-friendly format for model training, ensuring effective handling of diverse dataset features.

4.1.1.2 Decision tree class: The Decision Tree Classifier is selected to learn patterns in the dataset, specifically connections between health conditions and recommended medications. This algorithm enables the model to make predictions based on the acquired knowledge.

4.1.2 Model evaluation: Once the Decision Tree model is trained, its accuracy is evaluated by testing its predictions on the same dataset. This assessment gauges the model's effectiveness in recommending medications for specific health conditions, providing insights into its learning performance.

4.2 Medication Recommendation Interface

4.2.1 User interaction: An interactive and user-friendly interface is developed using the Dash framework, enabling easy interaction with the model. This interface acts as a bridge between users and the model, providing straightforward access to the model's capabilities without requiring technical expertise.

4.2.2 User input: The interface allows users to select their health symptoms or conditions from a dropdown list, a crucial step in providing the model with the context needed for accurate medication recommendations.

4.2.3 Real-time prediction: The interface excels in real-time medication predictions, offering immediate recommendations based on user-selected symptoms and conditions. This capability enhances the practicality of the interface, providing users with prompt and personalized medication guidance.

5. Testing

Testing is integral to the medication recommendation system's development, ensuring the Decision Tree model's accuracy and the user interface's user-friendliness. The process comprises two main aspects: validating the model's efficacy in recommending medications and assessing the interface's ease of use and accessibility. This thorough testing guarantees the system's functionality and aligns with user expectations.

5.1 Model accuracy: In the testing phase, the emphasis is on evaluating the accuracy of the Decision Tree model, responsible for medication recommendations based on health conditions. Rigorous testing involves exposing the model to diverse scenarios, assessing its predictive accuracy by comparing recommendations to a known set of medications for specific conditions. The process scrutinizes the model's effectiveness and reliability across various health conditions, ensuring consistent and dependable recommendations. This validation step confirms the model's capability to effectively provide accurate medication suggestions to users.

5.2 User-Friendly Interface: Equally crucial in the testing phase is evaluating the user-friendly interface, the interactive dashboard for medication recommendations. Usability testing involves real users navigating the interface, providing feedback on its user-friendliness and accessibility. The focus is on ensuring effortless navigation, assessing clarity in user instructions, the interface layout, and the ease of symptom selection. This evaluation aims to confirm that users can smoothly access medication suggestions without encountering significant obstacles or confusion.

Future Scope

Getting accreditation form the medical council of India (MCI), so that we can deploy out system in rural area hospitals, where there are less doctors and the ratio patient is to doctor is much higher.

Conclusion

``Generalized Medicine Recommendation" project aims to develop a versatile recommendation system that aids healthcare providers in making informed and personalized medication choices for their patients. The project focuses on data-driven insights and leverages advanced machine learning algorithms and a diverse dataset to provide comprehensive medication recommendations. The project also includes data preprocessing, BMI calculation, statistical overview, design and modelling, data extraction and cleaning, model development, and testing. The future scope includes getting accreditation from the medical council of India to deploy the system in rural area hospitals.

References

1. Benjamin Stark, Constanze Knahl, Mert Aydin and Karim Elish, “A Literature Review on Medicine Recommender Systems” International Journal of Advanced Computer Science and Applications(IJACSA), 10(8), 2019. http://dx.doi.org/10.14569/IJACSA.2019.0100802

2. Goyal, Varun A., Dilip J. Parmar, Namaskar I. Joshi, and Komal Champanerkar. "Medicine recommendation system." *Medicine (Baltimore)* 7, no. 3, 2020.

3. N. Varshney, S. Ahuja and Kanishk, "An Intelligence System for Medicine Recommendation: Review," *2021 5th International Conference on Information Systems and Computer Networks (ISCON)*, Mathura, India, pp. 1-5, 2021.

4. Bhimavarapu, U.; Chintalapudi, N.; Battineni, G. “A Fair and Safe Usage Drug Recommendation System in Medical Emergencies by a Stacked ANN”. Algorithms, 15, 186, 2022. https://doi.org/10.3390/ a15060186

5. Ochoa, J.G.D., Csiszár, O. & Schimper, T. “Medical recommender systems based on continuous-valued logic and multi-criteria decision operators, using interpretable neural networks”. *BMC Med Inform Decis Mak* **21**, 186, 2021.