**Building a Simple Recommendation System for Health Tips**

1. **Introduction**

This report details the development of a content-based recommendation system that provides personalized health tips based on user profiles (age, gender, medical history, etc.). The recommendation system uses Cosine Similarity to suggest the top 3 health tips for each user, with an evaluation of how well these tips align with real-life health advice.

2. **Key Preprocessing Steps**

Preprocessing was a critical part of the system. The following steps were taken:

* Handling Categorical Data: LabelEncoder was used to transform categorical variables such as "Reason" and "Description" into numerical values. This encoding is essential for computing similarity between profiles.

These preprocessing steps ensured that all features were in a format suitable for calculating Cosine Similarity.

3. **Data Cleaning Process**

To enhance the quality of the recommendations generated by the model, the Description column in the dataset was cleaned through the following steps:

* Counting Non-Zero Values: A pivot table was created to count non-zero entries for each medication description, identifying relevant attributes associated with each entry.
* Identifying Sparse Entries: A new DataFrame was formed to track the non-zero counts. Entries with a non-zero count of 1 were deemed insufficiently informative.
* Dropping Sparse Descriptions: These sparse entries were removed from the dataset, ensuring that only medications with adequate relevant information remained for analysis.
* Resetting the Index: The index of the cleaned DataFrame was reset for a clean and sequential structure.

These steps resulted in a refined dataset, which is expected to improve the accuracy and relevance of the medication recommendations provided by the system.

4. Model Choice and Rationale

We chose a content-based recommendation system using Cosine Similarity for the following reasons:

* Cosine Similarity measures the angle between two vectors (representing user profiles) and captures the degree of similarity between profiles regardless of magnitude.
* This method is ideal for the task as it focuses on the relationship between user attributes rather than absolute values, making it suitable for high-dimensional data like user demographics and medical history.
* Given the structured nature of the user profile data, this method provides an effective way to find similar users and suggest relevant health tips.

5. Model Implementation

We calculated the Cosine Similarity matrix between user profiles to identify similar users. For each user, the system:

* Identified the top 3 most similar users based on their profile attributes.
* Extracted the health tips (medication) associated with these similar users.
* Recommended these tips, ensuring a diversity of suggestions by filtering duplicates.

**6. Model Evaluation**

The evaluation of the system involved checking how well the recommendations align with the user’s original health tips. For instance, the system recommended tips such as Ibuprofen and Penicillin for users with relevant medical conditions. However, we noticed some duplicate recommendations, which need to be addressed.

To improve the evaluation, future iterations of the model will compare recommendations with actual medical guidelines to ensure that suggested health tips align with established best practices in healthcare.

**Suggested Improvements:**

* **Avoid Duplicate Recommendations**: The current model sometimes suggests the same health tip multiple times. A more advanced filtering system will be implemented to ensure diversity in recommendations.
* **Incorporating Collaborative Filtering**: In addition to the current content-based approach, collaborative filtering could be introduced to improve recommendations by analyzing patterns from multiple users with similar health conditions.
* **Using More Demographic Data**: Including additional user data (e.g., lifestyle, diet) could refine recommendations, making them more tailored to individual needs.

**7. Theoretical Explanation**

**Cosine Similarity**

Cosine Similarity measures the cosine of the angle between two vectors. It ranges from -1 to 1, where:

* 1 indicates complete similarity,
* 0 indicates no similarity (orthogonal vectors).

A screenshot of a white background

Description automatically generated