**Report: 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 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**

* **Handling Categorical Data**:
  + **One Hot Encoder** was used to transform categorical variables such as "Gender," "Medical Condition," and "Test Results" into numerical values.
  + **Label Encoder** was applied to variables like "Doctor" and "Hospital" for user comparisons.
* **Handling Numerical Data**:
  + **Standard Scaler** normalized features like "Age" and "Billing Amount" to ensure comparability and avoid bias in similarity calculations.

These steps were essential for preparing the data for **Cosine Similarity** computation.

**3. Model Choice and Rationale**

We selected a **content-based recommendation** system using **Cosine Similarity** due to its ability to measure the angle between two vectors, effectively capturing the similarity between user profiles. This method is particularly suitable for high-dimensional data such as user demographics and medical history, allowing the system to focus on relationships rather than absolute values. Given the structured user profile data, this approach effectively identifies similar users and recommends relevant health tips.

**4. Model Implementation**

The **Cosine Similarity matrix** was calculated between user profiles to find similar users. For each user, the system:

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

**5. Model Evaluation**

The system's performance was evaluated by how well the recommended tips aligned with users' original health advice. The system successfully recommended relevant medications like **Ibuprofen** and **Paracetamol** but encountered issues with **duplicate recommendations**.

**Suggested Improvements:**

* **Avoid Duplicate Recommendations**: Implement a more advanced filtering system.
* **Incorporating Collaborative Filtering**: Enhance 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 the recommendations.

**6. Theoretical Explanation**

**Cosine Similarity** measures the cosine of the angle between two vectors, ranging from -1 (complete dissimilarity) to 1 (complete similarity). User profiles were represented as vectors, with each feature (e.g., age, medical history) forming an element. Cosine Similarity's focus on vector orientation rather than magnitude made it ideal for this recommendation task.A screenshot of a white background

Description automatically generated

Cosine Similarity was chosen because it is well-suited for **high-dimensional, normalized data** like user demographics. It focuses on the *orientation* of vectors rather than their magnitude, which makes it effective for recommendation tasks based on profile similarity.

**7. Performance Metrics**

* **Cosine Similarity Score**: The system identifies the most similar users based on their profile attributes and suggests the top 3 health tips.
* **Real-Life Health Evaluation**: Recommendations generally align with the user's original health tips, though further evaluation against medical guidelines is needed.

**8. Conclusion**

This project successfully developed a content-based recommendation system using **Cosine Similarity** to suggest health tips tailored to user profiles. While the model effectively identified similar users and provided relevant recommendations, improvements are needed to enhance evaluation methods and recommendation diversity. Future work will focus on expanding the dataset, incorporating collaborative filtering, and refining the evaluation process to ensure practical, real-world applicability of the health tips provided.