**Health Tips Recommendation System using K-Nearest**

**Neighbours (KNN)**

**Introduction**

Personalized health recommendations are essential for improving individual well-being by addressing specific health concerns based on demographic and medical profiles. In this report, we present the design and implementation of a K-Nearest Neighbors (KNN) or Cosine Similarity based recommendation system for health tips. The goal is to predict personalized health advice based on a user’s profile, which includes age, gender, and medical condition.

**Dataset**

The dataset used for building the recommendation system consists of the following features:

Age: The age of the user (numerical).

Gender: The gender of the user (categorical: Male or Female).

Medical Condition: The medical condition of the user (categorical: Asthma, Cholesterol, Diabetes).

Health Tips: Textual health recommendations (target variable for classification).

BMI: Body Mass Index of individuals.

Smoking: Smoking status of individuals.

Sleep Time: Average sleep duration

The target variable in this dataset is the `Health\_Tips` column, which contains personalized health advice. The features are used to predict relevant health tips for users with similar profiles.

**Preprocessing**

***Encoding Categorical Variables***

Since KNN works with numerical data, we converted the categorical variables `Gender`, `Medical\_Condition`, and `Health\_Tips` into numerical values using Label Encoding. This technique assigns a unique integer to each category.

***Standardization of Numerical Features***

To ensure that all features are on a similar scale, the `Age` column was standardized using Standard Scaler. This step is crucial for KNN, as distance-based algorithms like KNN are sensitive to the magnitude of features.The standardized dataset was then split into training and test sets using an 80-20 ratio.

**Model Building**

***K-Nearest Neighbors Classifier***

The K-Nearest Neighbors (KNN) algorithm was selected for this task due to its effectiveness in finding similarities between user profiles. KNN is a non-parametric and instance-based learning algorithm, which means that it doesn’t assume any specific distribution of the data and makes predictions by finding the most similar training examples (neighbors).

***Prediction for New User Profiles***

A function was implemented to predict health tips for new users based on their age, gender, and medical condition. The function takes in a new user profile, encodes the inputs, and returns the predicted health tip by identifying the closest profiles using KNN.

**Model Evaluation**

The KNN model was evaluated using the test set, and the following evaluation metrics were calculated:

***Confusion Matrix:***

The confusion matrix allows us to understand how well the model performed in predicting each class. Misclassifications were noted between tips like "Maintain a healthy weight" and "Avoid known allergens and triggers."

***Classification Report:***

The classification report showed the precision, recall, and F1-scores for each class of health tips. The model performed better for common health tips, while tips with fewer samples in the training data had lower performance metrics.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | **Recall** | **Precision** | **F1**-**Score** | **Accuracy** |
| **KNN** | 0.40 | 0.39 | 0.39 | 0.40 |

***Accuracy Score :***

The overall accuracy of the model was calculated by comparing the predicted health tips with the actual tips in the test set.

**Methodology**

The implementation of cosine similarity involved the following steps:

***Data Preparation***: The dataset was structured into a DataFrame, allowing for efficient filtering and processing.

***Vectorization***: The TfidfVectorizer from sklearn was employed to convert the text data in the "Health Tips" column into numerical vectors, capturing the importance of each word in the context of the tips.

***Recommendation Function***: A function was created to recommend health tips based on user-provided attributes:

* Input parameters included age, gender, medical condition, and optional filters for BMI, smoking status, and sleep time.
* The function filtered the dataset based on the provided parameters to find relevant health tips.
* Cosine similarity was calculated between the input health tip and all available health tips in the dataset.
* Recommendations were sorted by similarity score, and duplicates were avoided to ensure distinct outputs.

***Testing***: The function was tested with various combinations of parameters to validate its effectiveness and ensure it returns appropriate health tips.

**Conclusion**

The K-Nearest Neighbors based recommendation system for health tips provides a simple yet effective solution for generating personalized health advice based on user profiles. The system leverages similarity between users in terms of age, gender, and medical condition to recommend relevant health tips.

Cosine similarity is a powerful technique for recommending health tips tailored to individual user attributes. By refining the filtering process and providing clear parameter options, the recommendation system can enhance its utility, helping individuals receive relevant health advice based on their unique situations. Future work could involve integrating machine learning techniques to further improve recommendation accuracy and user satisfaction.