Injoy Challenge*

Challenge Evaluation

All challenges will be evaluated in 2 stages. A first selection stage, and a final presentation stage with the top 4 teams for each challenge.

The first stage will involve a quick abstract of up to 200 words and a short video of up to 1min. The abstract will contain the information that you deem important to justify your approach and explain your solution. The video will contain an oral description of your main results with key figures of your choice, and a screen-recorded "demo run" of your code on a randomly sampled subset of the data (method described below). The purpose of the demo run is to validate that your results originated from your code, but you won't be judged on the performance from the demo run. The demo run can be very quick and take little space in your video, but you are free to incorporate it within the oral description of your results. The purpose of the oral description is to highlight your results and convince the judges to send you to the final presentation stage, in addition to your abstract.

Demo run:

1. Shuffle an array of indices with length equal to the test dataset.
2. Use its first N elements to create the demo dataset (N is challenge-specific).
3. Run your model with it and produce the demo plots (one of your key figures).

Note: In the case of cross-validation figures, feel free to reduce the number of folds. The only reason to down-sample the data is to fit the demo run in the 1min limit.

The final presentation stage will involve a 5-6min presentation with an additional 4 min allocated to questions from the judges (10min total). You are free to decide what you do with the presentation.

You will submit your abstract, video, code, and final presentation slides on the *devpost* before the deadline which is **Sunday at Noon**.

*This challenge does not have any associated prizes.

Challenge Description

Background

Irritable Bowel Syndrome (IBS) is a common functional gastrointestinal disorder that affects a significant portion of the population. The primary symptoms of IBS include abdominal pain, bloating, constipation, and diarrhea. The precise cause of IBS is not yet fully understood, but it is believed to be related to a combination of genetic, environmental, and lifestyle factors.

One of the most significant contributors to IBS symptoms is food. For people with IBS, certain foods can trigger symptoms such as abdominal pain, bloating, and diarrhea. Some common trigger foods include dairy products, FODMAPs, gluten, alcohol, and artificial sweeteners. Conversely, some foods, such as fiber-rich fruits, vegetables, and whole grains, may help alleviate IBS symptoms.

The impact of food on IBS symptoms varies greatly from person to person, making it challenging to determine the most effective diet for managing IBS. As a result, many people with IBS rely on a trial-and-error approach to identify their trigger foods and develop a personalized diet to manage their symptoms.

Therefore, determining the effect of food on a person's gastrointestinal symptoms is essential for people with IBS to effectively manage their condition and improve their quality of life.

The Problem

We are trying to figure out the link between food and gastrointestinal symptoms for individuals with IBS. To achieve this, we need to accurately predict the symptom score and be able to explain the contribution of each food to the final prediction for each individual user. Although each user records their diet and symptoms, the dataset is limited with only a few data points per user. To simplify the data, the food data is combined into broader categories like FODMAP, gluten, and wheat, which can lead to multicollinearity.

The Project

Develop a method to determine how much each food category contributes to the modelâs symptom score prediction for every individual user.

The Dataset

The dataset consists of paired food (columns beginning with F) and symptom data (symptom_value column) for 50 users (indicated by user_number column). For each user there are between 5 and 50 data points. Each row corresponds to the food consumed prior to the user experiencing the symptom (symptom_value column). Three files are provided for this project:

1. data.csv - contains the data points for all 50 users
2. food_symptom_relationship_strength.csv - contains how each food (food column) contributes (contribution) to the userâs (user_num) symptom
3. food_hierarchy.csv - contains information about which food groups are related to each other

Scoring

The performance of your approach will be evaluated by the Spearman correlation of your predicted contribution of each food and to the symptom score for each user and the true contribution (contribution column in food_symptom_relationship_strength.csv).

Marking Rubric

Category	Description	Score
Innovation	How unique or creative is the model design implemented?	1 2 3 4 5
Performance	How good is the correlation between the true and predicted effect of food on the symptom?	1 2 3 4 5
Presentation	Was the presentation well-organized, clear, and easy to follow?	1 2 3 4 5

Table 1: Marking Rubric

Data

You can download the data by clicking on the link here. It contains all files submitted to us by Injoy.

For the demo dataset, use 500 elements of the data.

Resources

You can find a list of resources for many ML and Data Science topics here.

Good luck!