**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 model’s 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 user’s (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 |

# 