## Inspiration/Abstract

The idea behind the process is that we first analyze the data provided from which we can make the assumption for the data points where the symptom values are zero, the person had no issues and felt healthy for that specific data point values as well. It is important to note that when it comes to food types, a higher value represents a higher quantity of food consumed of that particular type. We then went ahead and tried to find the correlation between different food types and the severity of the symptoms of Irritable Bowel Syndrome (IBS) and see which foods might possibly contribute to the development of IBS in the long run. Using the data that was provided we first tried to find a correlation between the food group and if symptoms will be seen. After we filtered the food which could cause IBS for the individual, we filtered the data and divided it into x and y segments, where x is the food value and y is the symptom value, and then used them to train a multilinear regression using XGBoost to predict the symptom value from the food values. We used SHAP values to determine how each food group contributes to the model’s prediction for each user which is given by XGBoost.It was from here that we tried to determine which of the most dangerous foods to consume should be avoided by running these gained values for the whole available dataset to know the food with common to cause IBS and hence should be avoided.

What It Does and How to Use It

It first shows consumption of which food types show symptoms of IBS symptom specific to the individual when you enter the specific user number.

## Challenges We Ran Into

The first major issue was with the description of the dataset provided. I would say our process and the documentation provided for the dataset could have been a little clearer. For example, the values of the dataset and what it corresponds to aren't clearly mentioned. A fake scenario of a particular patient depicting the story and the process would have helped to understand the data and land on a solution faster. We went on tried different approaches for finding the impact of a singular type of food on the symptom value, we tried linear regression, and random forest however they did not provide us with satisfactory results. We landed on multilinear regression to know the symptom value when eating a specific food related to quantity. What experimentation could have been done instead of having a mixed diet, a singular diet after a fast could have helped to reveal better results. We were very confused with the dataset provided and the problem we were trying to solve. There were a lot of assumptions that we were required to make when using the data leading to it not being a quality dataset. One of the challenges we had when trying to make an iteration of the model was to include the parent-child relationship between groups. However, upon implementing the parent-child relationship, it was unclear how this relationship affected the effects of food groups on the IBS symptom value, thus for this project we ended up deciding not to include the parent-child relationship in order to avoid false assumptions.

## Accomplishments We're Proud of

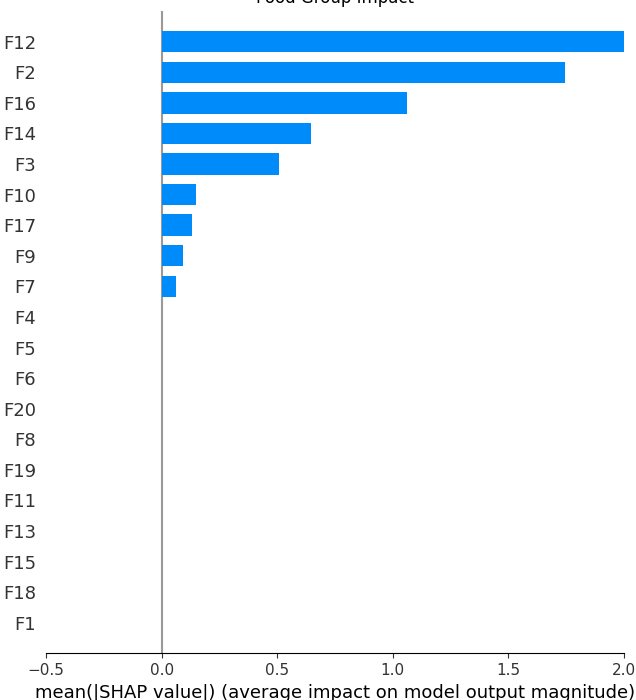
The different kinds of algorithms we learned and applied to find the actual cause was fun to find, we did find contrasting results to the ones provided and concluded with what food is harmful specific to the individual and up to what level. We are proud that we not only went ahead to find the food leading to IBS but also went ahead to find quantitative values for symptomatic values specific to one food group.

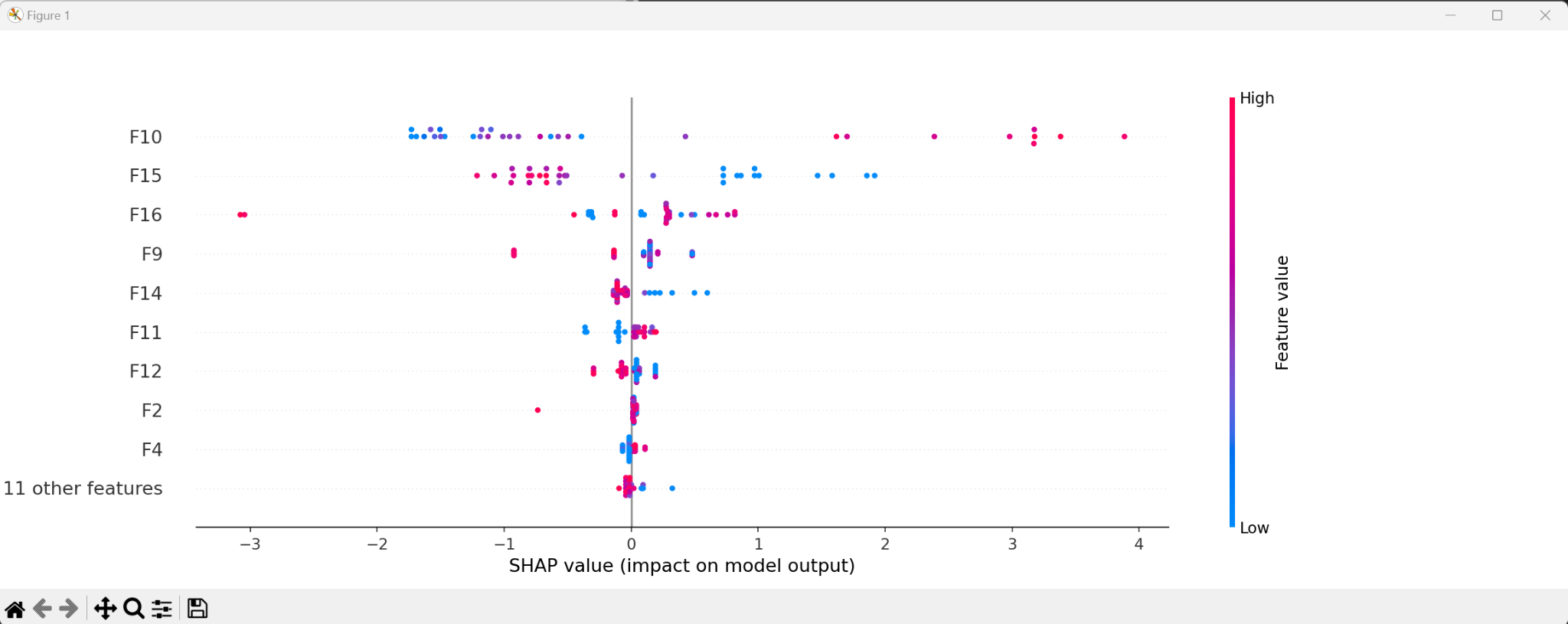
## What we learned

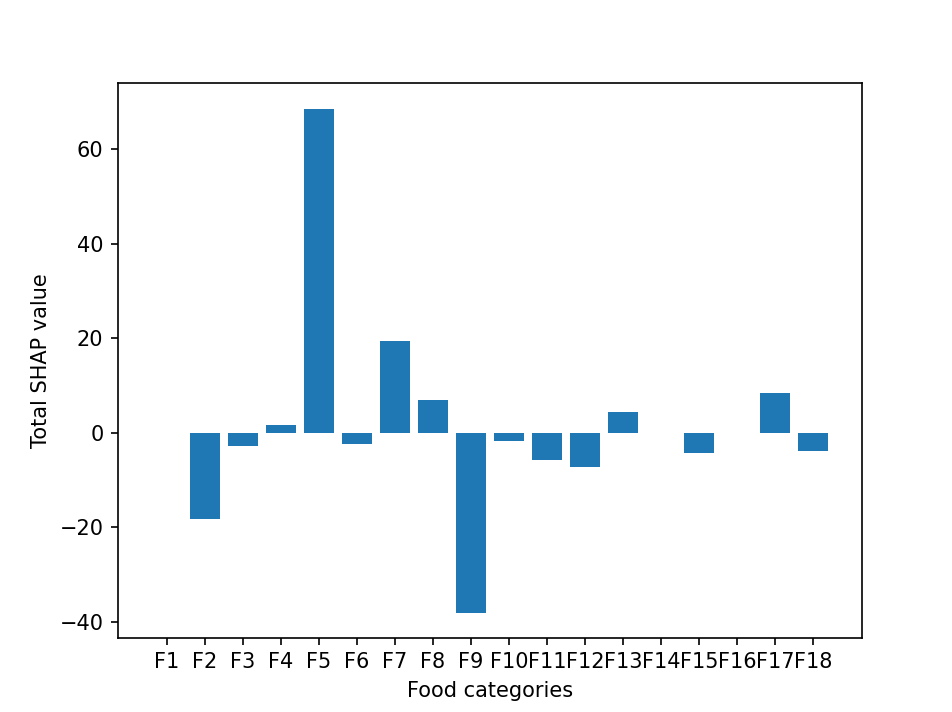
After running several tests and then doing a cumulative test to find the most IBS-causing food type we concluded that food type F5 has the most IBS-causing tendencies and hence a diet with a higher F5 food type will promote IBS. F9 has the most calming effect on IBS not exactly leading us to the conclusion but we believe this could mean it could help us recover from the caused effects.

## Built With

Python, Sklearn, XGBoost







We are building an XGBoost model to predict the severity of Irritable Bowel Syndrome (IBS) symptoms based on the consumption of different food types and their quantities. We start by importing the necessary libraries including pandas, numpy, xgboost, statsmodels, and matplotlib. Next, we read in the data from a CSV file and filter it based on the user's input. We then select the columns that have a high correlation with the IBS symptom value by feature selection techniques and drop any rows with null values thereafter.

We then split the dataset into training and test data, create a DMatrix for XGBoost, specify XGBoost parameters, and train the model. We create an explainer object for SHAP and calculate SHAP values for each feature for each instance. We summarize and plot the SHAP values for each feature using the SHAP library.