What is the relation between the results of the classification and instances of diet or nutrition-related diseases?

(Proposal)

The research is for creating a new food and nutrition classification. What is more, it can be used to discover the relationship between the consumption and the diseases which are caused by it.

In this dissertation, I would like to classify food categories by nutrition, connect food consumption with geographic information in London (by a specific radius of area, or borough), get the diet or nutrition-related diseases data in the area/borough, linked the food consumption and nutrition-related diseases.

As the previous research result, there are 7 major parts in nutrient classifications which are Carbohydrates, Proteins, Fats, Vitamins, Minerals, Dietary fibre, and Water. And two well-known diet or nutrition-related diseases are obesity and diabetes. Research also shows 50% fat intake with 50% Carbohydrates intake will more easily absorb by the body and turn to fat then causes obesity. The food which includes the such Carbohydrates-Fat ratio will be consider as the high-risk food of causing obesity. What is the definition of obesity? Body mass indices (BMIs) are used to define it. The indices are calculated as the weight in kilograms divided by the square of height in meters. The BMIs of the subjects are compared with the federal standards for BMI categorization: less than 18.5 = underweight, 18.5 to 24.9 = normal, 25.0 to 29.9 = overweight, and 30.0 and greater = obese. And Highly Processed Carbohydrates, Sugar-Sweetened Drinks, Saturated and Trans Fats, and Red and Processed Meats will greatly increase Diabetes Risk. Highly Processed Carbohydrates are also known as the refined carbs which have two main types: Sugars and Refined grains. Trans fats appear in packaged baked goods and fried foods, while saturated fats can be found in fatty meats, butters, and full-fat milk and cheese. The food contains such things above will be considered as high-risk food for diabetic.

By the previous result of the way that the body absorbs fat and carbohydrates, the food has a similar Carbohydrates-Fat ratio (which is close to 1) can be selected out as the main consumption that causes obesity. The range of ratios can be adjusted depends on the amount of selected food. After finding the classification, the next step is to find out the consumption of these kinds of food in different Tesco market locations. By assuming a part of the local people consumes their daily food in Tesco, comparing the nutrition-related diseases which are obesity in this case. Using linear regression on the data to find if the relationship is significant. The definition of high-risk food for a diabetic is more complicated, for example, a standard of Highly Processed Carbohydrates needs to be defined. By investigating the consumers' data, compare the consumption of Highly Processed Carbohydrates and the diabetic in the same area, and using the same method for the rest of the other 3 kinds of food. In the end, also by linear regression, the relationship can be found out between the consumption and the number of diabetics.

The extensive LSOA-level Tesco grocery dataset recently published by Aiello et al. <https://figshare.com/collections/Tesco_Grocery_1_0/4769354/2> - the data contain detailed information derived from some 420m food items purchased from 411 Tesco stores in London in 2015. Information includes average nutrient profiles from baskets of goods purchased as well as categories of goods purchased.

The data from customers in Tesco has to be informed consent and voluntary participation. While there will be data of nutrition-related diseases patients who might not want personal information to be exposed. Therefore, confidentiality and anonymity are especially important to keep patients or consumers personal information. Based on this, the research and investigation are going to only assess relevant components.