Capstone Project on Improving Nutrition Using Data Science

Vishnupriya Venkateswaran

MS Student in Data Science, The George Washington University

Nima Zahadat, Ph.D.

Professor of Data Science, The George Washington University

**Abstract**

Proper nutrition is one of the keys to a healthy life. Nutrition focuses on how diseases, conditions, and problems can be prevented or reduced with a healthy diet. Similarly, nutrition involves identifying how certain diseases and conditions may be caused by dietary factors, such as poor diet (malnutrition), food allergies, food intolerances, and overeating that causes obesity. You can improve your health by keeping a balanced diet.

The purpose of this project and paper is to analyze the different essential nutrients required for a person of a certain age and what are the various micronutrients, vitamins, and minerals a person needs daily and to analyze the different popular diets and determine if they are fit for a person and which diet is the healthiest one. Since fast foods are the go-to meals for many people, I do an extensive exploratory data analysis on fast-food restaurants to find foods that have a high nutritional value.

A food recommender system is built using machine learning models to recommend food items that are healthier than the person’s initial choice of food. Using different data science techniques and tools, data is extracted, cleaned, and used to deep dive into improving nutrition based on derived facts.

**Introduction**

Recent years have seen an impending global pandemic of obesity or malnutrition. Diets in the 2000s began to shift toward increased reliance upon processed foods, raised away from home intake, and greater use of edible oils and sugar-sweetened beverages. Reduced physical activity and increased sedentary time were also seen. Diets and activity patterns had changed drastically in the U.S. By the time we understood that the dietary quality of U.S. diets was worsening, physical activity was drastically reduced, and obesity was rising across the U.S. and Europe. At that point, only the U.S. was considered a country with an obesity problem: more than half of adults in some age-gender-race-ethnic specific subpopulations were overweight or obese. (Now and then: The Global Nutrition Transition,2012)

Home economics is dying as a taught art in the schools and workplaces. Processed food and pre-processed meals are increasing, in particular, fast food is becoming a significant part of our lives, and we become increasingly concerned about a more comprehensive array of health conditions related to nutrition.

In this paper, we empirically examine the impact of nutrition on individuals of different ages and gender and study the emerging trends on different diets and health quotient of many food items. The goal is to build a model that would suggest a healthier food choice for the overall well-being of a person. The data is collected from many different sources like Kaggle, health.gov, Kaiser Permanente, and world population review. Exploratory Data Analysis is performed, and many data mining techniques are used to analyze the data sets to understand why nutrition is essential and what are the essential nutrients required and how to improve the quality of life by improving the quality of food we eat.

**Literature Review and Research**

The purpose of the literature review is to establish the need for study and the foundation of the research key findings. Literature review and research are organized into four different sections.

* The first section outlines the background of the study: What are the essential nutrients required for a person's well-being? And a survey of the current status of food and nutrient intake and health.
* The second section outlines the different popular diets and their effects on the human body.
* The third section examines the fast food industry and its impact on the daily nutritional intake. These studies are vital as they are among the few empirical types of research done on the subject.
* The fourth section includes the measures taken to improve nutrition and how machine learning and artificial intelligence plays an integral part in this.

**Food and Nutrient Intakes and Health: Current Status and Trends**

Humans require a wide range of essential micronutrients and macronutrients for average growth and development and to support healthy aging throughout the life cycle. Essential nutrients, including most vitamins, minerals, amino acids, and fatty acids, water, and fiber, must be obtained through foods and beverages. Understanding the extent to which the U.S. population and various age, sex, and racial/ethnic groups within the population achieve nutrient intake requirements through available food and beverage intake, including foods and beverages that are enriched or fortified. These tasks are conducted by the Dietary Guidelines Advisory Committee (DGAC).

The DGAC considers that the primary source of nutrients should come from foods and beverages. Nutrient-dense forms of foods (those providing substantial amounts of vitamins, minerals, and other nutrients and relatively few calories) are recommended to ensure optimal nutrient intake without exceeding calorie intake or reaching excess or potentially toxic levels of certain nutrients (Dietary Guidelines Advisory Committee,2015). In the process of evaluating the adequacy of nutrient intake of the U.S. population, the DGAC identified two levels of concern.' Shortfall nutrients are those that may be under-consumed relative to the Estimated Average Requirement or Adequate Intake Overconsumed nutrients are those that are consumed in amounts above the Tolerable Upper Limit of Intake or other nationally recognized standards. ( Dietary Guidelines Advisory Committee,2015)

Currently, in the United States, Nutrient intake data from a representative sample of the U.S. population ages two years and older indicate that: vitamin A, vitamin D, vitamin E, folate, vitamin C, calcium, and magnesium are under-consumed relative to the Estimated Average Requirement. Iron is under-consumed by adolescents and premenopausal females, including women who are pregnant. Potassium and fiber are under-consumed relative to the Adequate Intake. Sodium and saturated fat are overconsumed relative to the Upper Limit of Intake or other standards for maximal intake.

The rising prevalence of obesity around the globe places an increasing burden on the health of populations, on healthcare systems, and overall economies. Obesity represents a prominent risk factor in multiple disease conditions as well as for global mortality (Wang et al. 2011). Obese patients are at high risk for developing comorbidities such as metabolic syndrome, cardiovascular disease, and endocrine disorders. Adult obesity is anticipated to ascend by 33 % in the coming two decades, with excessive obesity incidence rising by 130 %. Coalesce by the aging, and these trends entail that there will be as numerous as 65 million further adults obese in 2030 than in 2010, of whom 24 million will be over the age of 60 years( Wang et al. 2011). Obesity is a global pandemic affecting 2.1 billion people worldwide, with economic costs expected to exceed $125 billion in the United States by 2030. Having a balanced and healthy diet is, therefore, very important.

Taking nutritional supplements are increasing in recent times. Supplementation has many potential advantages over fortification and dietary approaches for improving micronutrient intake. Past experience with vitamin A supplementation in young children has proved to be remarkably successful. Demonstrated efficacy of vitamin A supplements for improved child survival in many settings and a technical consensus on how to implement interventions were significant factors in achieving this success. The programmatic approaches available for increasing micronutrient intake include dietary modification, food fortification, and supplementation. While each method has its inherent advantages and disadvantages, these different approaches for improving micronutrient status of populations are more often complementary options than competing options hence the use of supplements to meet the nutrient intake for a person (Shrimpton, R., & Schultink, W, 2002). Supplements can play an essential role in meeting the micronutrient needs of the developing world, but their full potential is still far from being realized. Multiple micronutrient supplements under program conditions in developing countries thus assume considerable importance for furthering the use of supplements to improve micronutrient status in developing countries.

**Popular Diets**

On researching the popular diets, the study conducted by The Continuing Survey of Food Intake by Individuals data was used to examine the relationship between popular prototype diets and diet quality as measured by the healthy eating index (HEI), consumption patterns, and body mass index (BMI). The prototype diets included vegetarian (no meat, poultry, or fish on the day of the survey) and non-vegetarian. The non-vegetarian group was further subdivided into low carbohydrate (less than 30% of energy from carbohydrate), medium (30% to 55%), and high (greater than 55% of energy). Within the top carbohydrate group, participants were classified as having Pyramid or non-Pyramid eating patterns. The Pyramid group was defined as 30% or less of energy from fat and at least one serving from the five major food groups in the USDA Food Guide Pyramid. Finally, the non-Pyramid group was further subdivided into low fat (less than 15% of energy from fat) and moderate fat (15% to 30% of energy from fat). Energy intakes were low for the vegetarians (1,606 kcals) and high carbohydrate/low-fat group (1360 kcals). BMIs were lowest for women in the vegetarian group (24.6) and the high carbohydrate/low fat group; for men, the weakest BMIs were observed for vegetarians and the high carbohydrate Pyramid group. A review of the literature suggests that weight loss is independent of diet composition. (Popular Diets Kennedy, Eileen T, et al., 2011).

Worldwide, an estimated 2 billion people live primarily on a meat-based diet, while an estimated 4 billion live mainly on a plant-based diet. A comparison of the calorie and food consumption of a lactovovegetarian diet and a meat-based diet is provided in Table 1. In the lactoovovegetarian diet, the meat and fish calories were replaced by proportionately increasing most other foods consumed in Table 1 in the vegetarian diet except sugar and sweeteners, fats, and vegetable oils. The total weight of food consumed was slightly higher (1002 kg per year) in the lactoovovegetarian diet than in the meat-based diet (995 kg per year). The most food calories consumed in both diets were associated with food grains, and the second-largest number of calories consumed was from sugar and sweeteners. The number of feed grains used to produce the animal products (milk and eggs) consumed in the lactoovovegetarian diet was about half (450 kg) the number of feed grains fed to the livestock (816 kg) to produce the animal products consumed in the meat-based diet (Table 1). (David Pimentel,2003)

A screenshot of a social media post

Description automatically generated

The diet that's gaining more popularity is the vegan diet. A vegetarian diet is associated with many health benefits because of its higher content of fiber, folic acid, vitamins C and E, potassium, magnesium, and many phytochemicals and fat content that is more unsaturated. Compared with other vegetarian diets, vegan diets tend to contain less saturated fat and cholesterol and more dietary fiber (Winston J Craig,2009). What then, is the nutritional and health status of those who follow a vegan diet? Compared with other vegetarians (e.g., lactoovovegetarians). The study conducted suggests that vegans should regularly consume vitamin B-12–fortified foods, such as fortified soy and rice. To ensure adequate calcium in the diet, calcium-fortified plant foods should be daily consumed in addition to consuming the traditional calcium sources for a vegan. To ensure an adequate vitamin D status, especially during the winter, vegans must regularly consume vitamin D–fortified foods such as soymilk, rice milk, orange juice, breakfast cereals, and margarines that are fortified with vitamin D. The vegan diet has it benefits when done right or it would lead to severe health conditions (Winston J Craig,2009).

Another popular diet is the Ketogenic diet. The ketogenic diet (K.D.) is a very high-fat, low-carbohydrate diet that restricts carbohydrate to ≤10% of consumed energy, starkly different from the 50–65% of energy from carbohydrate of the typical American diet. An experiment was conducted for Alzheimer's disease. Fifteen A.D. patients participated in the study, and the Dietary intake was measured via multiple self-reported 3-d food records completed by participant study partners at home, in real-time. The diet’s principles can be as shown in Table 2.

A screenshot of a cell phone

Description automatically generated

The results from the study showed Ten participants adhered to the K.D. Study partners provided complete food records for 6 KD-adherent individuals. Micronutrient intake was similar between diets, meeting Dietary Reference Intakes for most nutrients. Between diets, the K.D. was associated with increased consumption of choline and vitamin K and decreased intake of manganese and fiber. (*Current* Developments in Nutrition,2019)

**Impact of Fast Foods on Health**

Foods consumed outside the home are typically less healthy than those consumed at home. Fast foods (e.g., pizza, burgers, and fried chicken) tend to be high in saturated fat and salt, energy-dense, nutrient-poor and served in large portions. The regular consumption of meals from fast-food outlets and the increasing fast-food meal consumption frequency over time have been associated with adult weight gain. Increased patronage of fast-food outlets has also been associated with excess weight gain over time. The consumption of fast foods appears to be influenced by both individual- and neighborhood-level factors. One individual-level factor is socioeconomic status, characterized in terms of income, occupation, or educational attainment. Adults with lower levels of education, in particular, have been reported to consume unhealthy fast foods more frequently and to visit fast-food outlets more often than those with higher levels of education.

A study conducted by the University of Cambridge School of Clinical Medicine was to test whether observed differences in fast-food consumption and obesity by fast-food outlet exposure are moderated by educational attainment. They used linear and logistic regression models to examine associations between educational attainment, fast-food consumption, BMI, and odds of being obese. The results showed Greater fast-food consumption, BMI, and odds of obesity were associated with greater fast-food outlet exposure and a lower educational level. Fast-food consumption and BMI were significantly different across education groups at all levels of fast-food outlet exposure (P-value < 0.05). High fast-food outlet exposure amplified differences in fast-food consumption across levels of education (Does neighborhood fast-food outlet exposure increase inequalities in diet and obesity,2016).

Another study conducted among children showed Half of U.S. children consumed fast food: 39.5% low consumers and 10.5% high consumers. The top consumer was associated with overweight/ obesity and had stronger associations with inadequate total intake. (Fast food consumption with poor dietary outcomes,2019)

Kid's meals are popular in Fast Food chains. A study conducted by the National School Lunch Program conducted analysis compared the nutrient values of meals offered by major fast-food companies with restaurants. For each meal combination, the following were analyzed: total energy, percentage of energy from fat, total fat, saturated fat, sodium, total carbohydrates, dietary fiber, added sugars, protein, vitamin A, vitamin C, calcium, iron, energy density. The results showed that Meals that did not meet the criteria were more than 1.5 times more energy-dense. (Nutrient quality of fast-food kids’ meals,2019)

**Measures taken to improve nutrition**

With the rise of machine learning and artificial intelligence. Many studies and models have been used to take steps in improving nutrition, and one such model is the food recommender system. The personalized food recommender system aims to assist the users in daily diet selections based on some nutrition guidelines. The research paper used focuses on the development of a customized food recommender system that can provide dietary recommendations, which are based on both individual diet needs and preference. The design of the system uses a knowledge-based framework. Correctly, the knowledge engineering approach is used in modeling the relevant user profile as well as food and nutrition knowledge in an ontology form. The ontology and some rule-based knowledge are used as the basis for constructing a knowledge base in providing the user recommendations. Both the ontology and system development processes are elaborated and exemplified. (Knowledge-based Framework,2010)

An important concept used in the food recommender system is the nutrient profiling of foods. Nutrient profiling of foods is defined as the science of ranking foods based on their nutrient composition. Measures of nutrient density, previously applied only to total diets, are now being adapted for use with individual foods. Assigning foods into categories based on their nutrient content has many potential applications, ranging from consumer education and dietary guidance to nutrition labeling and the regulation of health claims (Nutrient profiling of foods,2010). Multiple efforts to develop, validate and test nutrient profile models are now underway. Reports issued by the Food Standards Agency (FSA) in the United Kingdom, which is the equivalent of the Food and Drug Administration (FDA) in the United States, are now available online.

Nutrient profile models can be based on 1) qualifying nutrients are known to be beneficial to health, mostly vitamins and minerals; 2) disqualifying nutrients, mostly fats, sugars, and sodium; or 3) some combination of both. The content of fruits, vegetables, nuts, or whole grains in a portion of food can also be taken into account. The nutrition index can be calculated using the formula provided in the paper. These indices are used to profile food items as healthy and unhealthy and can be grouped into clusters that can be used to recommend healthy food items from the data.

**Data and datasets**

1. **World population Review**

This dataset contains data on the healthiest countries in the world. The factors considered in this study are Health risks (tobacco use, high blood pressure, obesity), Availability of clean water, Life expectancy, Malnutrition, and Causes of death.

1. **Daily Dietary Nutrients**

The daily dietary Nutrients dataset is from Kaiser Permanente. The data contains information on Micronutrients, Minerals, and Vitamins for Age-Sex Groups Based on Dietary Reference Intakes and Dietary Guidelines Recommendation.

1. **Diet Analysis**

The data for Keto, Vegan, and Low carb diet is web scrapped from nutritionvalue.org. The data set contains information on the daily dietary value in each of the diets.

1. **Fast Food Data**

The data for fast food analysis is collected from Kaggle. The nutritional facts for Mc Donald’s menu provide a nutrition analysis of every menu item on the US McDonald's menu, including breakfast, beef burgers, chicken and fish sandwiches, fries, salads, soda, coffee, and tea, milkshakes, and desserts. The nutritional facts for the Starbucks menu provide nutritional information for Starbucks’ food and drink menu items.

1. **Food Items with its nutrients**

Two datasets are collected for Machine Learning. The first is the USDA National Nutrient Database for Standard Reference from the U.S. Department of Agriculture's Agricultural Research Service. The dataset contains the primary source of food composition data in the United States with more than 8,000 data points with over 50 variables of micronutrients in each food (such as carbohydrate, protein, etc.). The second data set is the Instacart data from Kaggle. The dataset for this competition is a relational set of files describing customers' orders over time. The dataset is anonymized and contains a sample of over 3 million grocery orders from more than 200,000 Instacart users.

**Data Analysis**

The datasets are extracted and stored for data analysis. The data is analyzed using Data Mining techniques in Python. The first process is to clean the data and remove all null and NaN values which are not required for the analysis. After the data is ready, Exploratory data analysis is performed for various aspects of the study. Data is analyzed and visualized using Tableau, D3.js, and plotly. Machine learning supervised, and unsupervised models are used using scikit-learn packages in Python.

The **first** portion of the analysis was to determine the healthiest countries in the world. This is based on a variety of factors that contribute to healthy, happy countries. Typically, the healthiest countries are those that are developed. These countries have lower rates of pollution, access to quality health care, and access to clean, safe drinking water. The Bloomberg Global Health Index takes a look at several of these factors to rank the healthiest countries in the world. The factors considered in this study are Health risks (tobacco use, high blood pressure, obesity), Availability of clean water, Life expectancy, Malnutrition, and Causes of death.

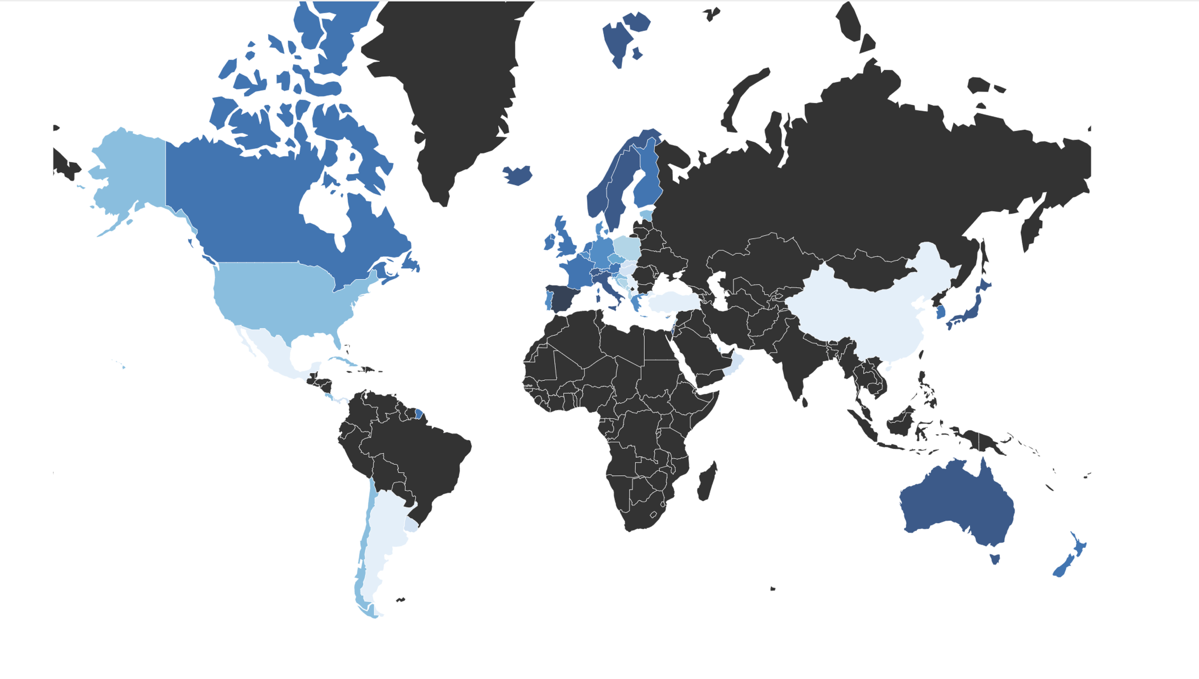


Figure 1: Healthiest Countries in the world.

Using these factors, each country is given a rating out of a top score of 100. According to the 2019 ranking, [Spain](http://worldpopulationreview.com/countries/spain-population/) is considered to have the healthiest people in the world with a score of 92.75. Other countries that ranked high for health include Italy, Iceland, Japan, and Switzerland. This can be seen in Figure 1. The top 10 healthiest countries, according to the Bloomberg Global Health Index, are Spain, Italy, Iceland, Japan, Switzerland, Sweden, Australia, Singapore, Norway, Israel.

The **second** portion of the analysis involves a deep dive into the essential nutrients that are required daily for a person of a certain age. The nutrients are divided into three categories, such as micronutrients, vitamins, and minerals. The percentage of the daily nutrient value is as represented below:

A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated

Figure 2: Daily Requirements of Vitamins, Minerals, and Macronutrients

The **third** portion of the analysis is to compare the most popular diets in America. The three most popular diets are the ketogenic diet, low carb diet, and Vegan diet.

The ketogenic diet is a very low-carb, high-fat diet. It involves drastically reducing carbohydrate intake and replacing it with fat. This reduction in carbs puts your body into a metabolic state called [ketosis](https://www.healthline.com/nutrition/what-is-ketosis). The graph1 shows a comparison between how much nutrients you need in a day and how much you get while on the Keto diet. It requires a person to eat how amounts of fiber and proteins.

A picture containing screenshot

Description automatically generated

Graph 1: Keto Diet Facts

Some of the side effects of keto diets are Vitamin deficiency, Dehydration, Kidney damage, Headache, Fatigue, Brain fog, Irritability, Nausea, and Stomachache.

The low carb diet limits carbohydrates — such as those found in grains, starchy vegetables, and fruit and emphasizes foods high in protein and fat. The graph2 shows a comparison between how much nutrients you need in a day and how much you get while on the Low Carb diet. As seen, none of the daily nutrient value is met.

A screenshot of a social media post

Description automatically generated

Graph 2: Low Carb Diet Facts

Some of the side effects of the low carb diets are Headache, Weakness, Muscle cramps, Fatigue, skin rash, constipation or diarrhea and long term side effects include kidney damage, increased cancer risk, impairment of physical activity, and lipid abnormalities can all be linked to long-term restriction of carbohydrates in the diet.

The vegan diet is a type of vegetarian diet that excludes meat, eggs, dairy products, and all other animal-derived ingredients. This diet based on a wide variety of whole plant foods such as fruits, vegetables, whole grains, legumes, nuts, and seeds.

Graph 3 shows a comparison between how much nutrients you need in a day and how much you get while on a Vegan diet.

A screenshot of a social media post

Description automatically generated

Graph 3: Vegan Diet Facts

This diet is most beneficial when done the right way. It provides a balanced diet with a proper amount of carbs, fiber, and protein. Some of the side effects of a Vegan diet is that it leads to vitamin deficiency, which can be solved by changing the vegan diet accordingly. This diet is the best compared to other diets to be healthy.

The **fourth** portion of the analysis is to dive into the fast food industry in America. The top five fast-food chains in the USA based on their revenue is Mc Donald's, Starbucks, Dunkin Donuts, Taco Bell, and Burger King. To understand the nutritional value of fast food, I considered Mc Donald's and Starbuck's data.

Mc Donald's data consists of all the food items on the menu and their daily nutritional value. I considered the popular meal choices at Mc Donald's, which is Big Mac, Chicken Nuggets, Egg Muffin, Fries, and Frappe. One thing that was common among all these food items was the total fat content, which was higher than what a person is supposed to consume daily. The charts below the nutritional content for all the items mentioned. A person can judge if they need to eat this to be healthy.

A close up of a device

Description automatically generatedA picture containing screenshot

Description automatically generatedA close up of a logo

Description automatically generatedA screenshot of a cell phone

Description automatically generated

Figure 3: Nutritional Value of Mc Donald’s Menu

Moving on to Starbucks. The data contained a nutritional value for all food items and drinks on the menu. Upon analysis the items with the daily required nutrients, I picked the top healthy food items on the menu. This can be as seen below:

A screenshot of a computer

Description automatically generated

Figure 4: Healthy Food options in Starbucks

In the Starbucks drinks menu. I picked the highest calorie drinks, which are unhealthy. They are as seen below:

A screenshot of a computer

Description automatically generated

Figure 5: Unhealthy Drinks in Starbucks

The **fifth** portion is Machine Learning. Supervised Machine Learning is used to predict the micronutrients found in common foods using the Random Forest Model. Unsupervised Learning is used to build a Food Recommender system using K means model.

**Predicting Micronutrients Using Random Forest**

In this part of the study, I used supervised machine learning to predict micronutrients within certain foods. In this case, the input is records of the amount of nutrient composition within each meal, and the task is to predict the micronutrients based on other nutrients and compare how far are the predicted values from the actual values.

Before building the model. I wanted to check the correlation among the variables. It can be as seen in the heat map below:

A screenshot of a cell phone

Description automatically generated

We can see that some of the nutrients are either positively correlated or show no correlation with other nutrients, while only a couple is negatively correlated.

Now we chose the labels. I decided to predict Calcium, Iron, Zinc, Vitamin A, Vitamin D, Folate. First, we need to split the input data and labels in a training and test set. Scikit-learn's train\_test\_split function makes this process convenient. Now, we need to standardize our data before coding the model. This is important because some of the nutrients use different units. For example, Vitamin A uses the U.I. unit, while carbohydrate uses the gram.

We will be using Scikit-learn’s RobustScalerto standardize our data so that all of them are robust against the differences in units. Then we initialize the model and train it using the train data. Since we have a multiple output variable, we need to use MultiOutputRegressor to the random forest model.

The model has an accuracy of 94% on the training data and 62% on the test data. The model does not overfit or underfit the data.

A sample output for cheese is as follow:

True value: Vitamin A- 650, Iron-0.31, Zinc-2.70, Calcium -500, Vitamin D-30, Folate- 16

Predicted value: Vitamin A- 721, Iron-0.31, Zinc-2.66, Calcium -528, Vitamin D-21, Folate- 36

**Healthy Food Recommendation System**

To build a Health Food Recommendation System, I used unsupervised Learning. "Nutrient profiling of foods, described as the science of ranking foods based on their nutrient content, is fast becoming the basis for regulating nutrition labels, health claims, and marketing and advertising to children. Several nutrient profile models have now been developed by research scientists, regulatory agencies, and the food industry. Whereas some of these models have focused on nutrients to limit, others have emphasized nutrients known to be beneficial to health, or some combination of both. Although nutrient profile models are often tailored to specific goals, the development process ought to follow the same science-driven rules. These include the selection of index nutrients and reference amounts, the development of an appropriate algorithm for calculating nutrient density, and the validation of the chosen nutrient profile model against healthy diets. Nutrient profiles must be validated rather than merely compared to prevailing public opinion.

The food industry's concern has been that some profiling approaches, notably the "traffic light" system tend to separate foods into "good" and "bad," such that whole category of foods may be penalized. One way to deal with this issue is to create nutrient profiles that are category-specific, as opposed to an across-the-board approach.

The definition of healthy or nutrient-dense foods has taken many forms over the years:

More nutrients, fewer calories bucketing certain food groups as healthy: fruits, veggies, lean meats, etc. and bucketing certain foods as unhealthy: high fat (including nuts, olives, coconuts, and avocados). It is only recently that the FDA has explored the feasibility of allowing health claims if the food had a favorable nutrient profile, as an alternative to current measures based on grams of nutrients per serving."

"Nutrient profile models can be based on 1) qualifying nutrients are known to be beneficial to health, mostly vitamins and minerals; 2) disqualifying nutrients, mostly fats, sugars, and sodium; or 3) some combination of both. The content of fruits, vegetables, nuts, or whole grains in a portion of food can also be taken into account."

There are many algorithms used for nutrient profiling. They can be summarized, as shown in the table below:

Table 1: Nutrient Profiling Algorithms

A close up of a newspaper

Description automatically generated

I used the algorithm Nutrient Rich Food NRFNn.3 (NRFNn - LIM) = Unweighted arithmetic mean of %DV for n nutrients minus 3 negative nutrients.

Nutrients to be used can be:

Nutrients beneficial to health selected macronutrients (protein, fiber, essential fatty acids)

vitamins (vitamins A and C), and minerals (calcium and iron), omega-3 fatty acids, B-vitamins and folate, and additional minerals, typically potassium, zinc, and magnesium

Harmful nutrients FDA: total fat, saturated fat, cholesterol, sodium

The next step is to clean the data and remove unwanted columns and null values. Then the nutrient profile is calculated using the above algorithm. The nutrients are classified for all the food categories. Then I used the K-Means algorithm to cluster all the food groups with similar nutrient density. After the model is deployed, the food recommender is used to recommend healthier options.

Input: Cupcakes. Recommendation: Donuts (More Calcium, Lesser Sugar)

Input: Pinto Beans. Recommendation: Black Beans. (More potassium, protein, calcium, and iron)

Input: Cheddar Cheese. Recommendation: Italian Cheese (More Vitamin A, less sodium)

The results turned out to be very good as the recommender suggested a healthier food alternative when compared to the initial choice of food.

**Key Findings**

Focusing on nutrition, helped me understand how it was essential to a person's well-being and what are the essential nutrients required to be healthy. According to the research and data analysis, the vital nutrients required are vitamins, minerals, and micronutrients, and what are the specific ones needed for each category.

From the research conducted, it can be found that in the united states, the most popular diets are the Ketogenic diet, Low carb, and Vegan diet. An in-depth analysis of each diet sheds light on the advantages and disadvantages of each. Diets do not provide the daily required nutrition a body needs.

In this fast-paced era, everyone is looking to grab a quick bite to eat, especially from fast-food chains. The analysis and research helped in finding healthier options in this fast-food menu so we can make smarter choices if we have no choice than to consume fast foods.

Machine learning can be used to make a smarter food choice, hence building a healthy food recommendation system can suggest more robust alternatives to unhealthy cravings. The model would make an effort to scan through different food options to recommend healthy food, so you don’t have to put in the effort and time in researching nutrition in food items.

**Recommendations**

Based on the research, the results can be improved by studying nutrition for a broader range of demography. Studies for a particular nationality can help understand and modify the daily required nutrition. Also, research can be improved by analyzing the diet trends among kids and teenagers to improve overall nutrition. The analysis of fast food can be enhanced by doing an in-depth analysis of each nutrient on each item on the menu to find the unhealthy food items on the menu.

The random forest model built to predict the nutrients can be improved by increasing the performance accuracy of the model. This can be done by adding more data points to the input and also hyper tuning the parameters to increase the overall accuracy of the model. An API can be created to predict the nutrients of particular food items to make it more user-friendly.

The healthy food recommender system can be improved by refining the tagging process of food items into categories. Other algorithms can be used to calculate the nutrition profiling index to create more accurate food cluster groups. The users can be allowed to customize the nutrient index formula according to their needs. A user interface can be designed to make this application more user-friendly.

**Conclusion**

This project used data gathered from many different sources to study the overall nutrition required to live healthy. Different nutrients were analyzed to find the amount of nutrition needed daily, depending on the age of a person. A detailed analysis of popular diets was studied to determine the advantages and disadvantages of each diet.

Fast food analysis was conducted to see if we can find healthy meals at the biggest fast-food chains in America. A detailed study on the menu items at Mc Donald’s and Starbucks determined the nutrition level of many foods and drinks. This can help solve the problem of choosing a good food item when eating fast food.

A Random Forest classifier was used to predict the nutrient content in a food item from a list of many common foods we eat. The accuracy of the model can be increased by adding more data points and tuning the parameters according to the user’s requirement.

A healthy food recommender system helps in solving the problem of choosing healthy food over an unhealthy food choice. This can be used to suggest and recommend alternative food choices across many food groups and beverages. Overall this project helped in improving nutrition using different data science tools, methods, and techniques.

**Biography**

**Vishnupriya Venkateswaran**is a Data Science graduate student at the George Washington University. She has worked as a Software Engineer in India for three years and worked as a Data Analyst Intern at Kaiser Permanente in Maryland while pursuing her Master's. Her research interests span both computer science and data science. Much of her work has been on improving the understanding, design, and performance of data systems, mainly through the application of data mining, data visualization, statistics, machine learning, artificial intelligence, and performance evaluation.

**Dr. Nima Zahadat** is a professor of data science, information systems security, and digital forensics. His research focus is on studying the Internet of Things, data mining, information visualization, mobile security, security policy management, and memory forensics. He has been teaching since 2001 and has developed and taught over 100 topics. Dr. Zahadat has also been a consultant with the federal government agencies, the U.S. Air Force, Navy, Marines, and the Coast Guard. He enjoys teaching, biking, reading, and writing.

**References**

1. Now and then: Global nutrition transition,2012: Popkin, Barry M et al. “Global nutrition transition and the pandemic of obesity in developing countries.” *Nutrition reviews* vol. 70,1 (2012): 3-21. doi:10.1111/j.1753-4887.2011. 00456.x
2. Dietary Guidelines Advisory Committee. 2015. Scientific Report of the 2015 Dietary Guidelines Advisory Committee: Advisory Report to the Secretary of Health and Human Services and the Secretary of Agriculture. U.S. Department of Agriculture, Agricultural Research Service, Washington, DC. (https:// <https://health.gov/dietaryguidelines/2015-scientific-report/>)
3. Shrimpton, R., & Schultink, W. (2002). Can supplements help meet the micronutrient needs of the developing world? Proceedings of the Nutrition Society, 61(2), 223-229. doi:10.1079/PNS2002163
4. Wang et al. 2011: Health and economic burden of the projected obesity trends in the USA and the UK Wang, Y Claire et al. The Lancet, Volume 378, Issue 9793, 815 - 825
5. Popular Diets Kennedy, Eileen T et al. Journal of the American Dietetic Association, Volume 101, Issue 4, 411 - 420
6. David Pimentel, Marcia Pimentel, Sustainability of meat-based and plant-based diets and the environment, *The American Journal of Clinical Nutrition*, Volume 78, Issue 3, September 2003, Pages 660S–663S, <https://doi.org/10.1093/ajcn/78.3.660S>
7. Winston J Craig, Health effects of vegan diets, *The American Journal of Clinical Nutrition*, Volume 89, Issue 5, May 2009, Pages 1627S–1633S, <https://doi.org/10.3945/ajcn.2009.26736N>
8. *Current Developments in Nutrition* ,2019 : Matthew K Taylor, Russell H Swerdlow, Jeffrey M Burns, Debra K Sullivan, An Experimental Ketogenic Diet for Alzheimer Disease Was Nutritionally Dense and Rich in Vegetables and Avocado, *Current Developments in Nutrition*, Volume 3, Issue 4, April 2019, nzz003, <https://doi.org/10.1093/cdn/nzz003>
9. Does neighborhood fast-food outlet exposure amplify inequalities in diet and obesity,2016: Thomas Burgoine, Nita G Forouhi, Simon J Griffin, Søren Brage, Nicholas J Wareham, Pablo Monsivais, does neighborhood fast-food outlet exposure amplify inequalities in diet and obesity? A cross-sectional study, *The American Journal of Clinical Nutrition*, Volume 103, Issue 6, June 2016, Pages 1540–1547, <https://doi.org/10.3945/ajcn.115.128132>
10. Fast food consumption with poor dietary outcomes,2019: Jennifer M Poti, Kiyah J Duffey, Barry M Popkin, The association of fast food consumption with poor dietary outcomes and obesity among children: is it the fast food or the remainder of the diet?, *The American Journal of Clinical Nutrition*, Volume 99, Issue 1, January 2014, Pages 162–171, <https://doi.org/10.3945/ajcn.113.071928>
11. Nutrient quality of fast food kids meals,2019 : Sharon I O'Donnell, Sharon L Hoerr, Jason A Mendoza, Eugenia Tsuei Goh, Nutrient quality of fast food kids meals, The American Journal of Clinical Nutrition, Volume 88, Issue 5, November 2008, Pages 1388–1395, <https://doi.org/10.3945/ajcn.2008.26197>
12. A Knowledge-based Framework,2010: Knowledge-based Framework for Development of Personalized Food Recommender System Napat Suksom, Marut Buranarach, Ye Myat Thein, Thepchai Supnithi, Ponrudee Netisopakul
13. Nutrient profiling of foods,2010 : Adam Drewnowski, Victor Fulgoni, Nutrient profiling of foods: creating a nutrient-rich food index, *Nutrition Reviews*, Volume 66, Issue 1, 1 January 2008, Pages 23–39, <https://doi.org/10.1111/j.1753-4887.2007.00003.x>
14. Leah M Lipsky, Tonja R Nansel, Denise L Haynie, Danping Liu, Kaigang Li, Charlotte A Pratt, Ronald J Iannotti, Katherine W Dempster, Bruce Simons-Morton, Diet quality of U.S. adolescents during the transition to adulthood: changes and predictors, *The American Journal of Clinical Nutrition*, Volume 105, Issue 6, June 2017, Pages 1424–1432, <https://doi.org/10.3945/ajcn.116.150029>
15. Adam Drewnowski, Victor Fulgoni, Nutrient profiling of foods: creating a nutrient-rich food index, *Nutrition Reviews*, Volume 66, Issue 1, 1 January 2008, Pages 23–39, <https://doi.org/10.1111/j.1753-4887.2007.00003.x>
16. Joly, Arnaud, Pierre Geurts, and Louis Wehenkel. “Random Forests with Random Projections of the Output Space for High Dimensional Multi-Label Classification.” Lecture Notes in Computer Science (2014): 607–622. Crossref. Web
17. Popkin, Barry M et al. “Global nutrition transition and the pandemic of obesity in developing countries.” *Nutrition reviews* vol. 70,1 (2012): 3-21. doi:10.1111/j.1753-4887.2011. 00456.x
18. Prevalence of Micronutrient Malnutrition Worldwide 2002 Usha Ramakrishnan, Ph.D.
19. World Population Review (https:// worldpopulationreview.com/countries/healthiest countries/)
20. The Nutrient Rich Foods Index helps to identify healthy, affordable foods. Drewnowski A1. (https://academic.oup.com/ajcn/article/91/4/1095S/4597206)