



RIGHT FOOD FOR HEALTHY LIFE

USING MACHINE LEARNING TECHNIQUES TO CHOOSE RIGHT FOOD
CATEGORY BASED ON NUTRIENTS PRESENT IN THE DATA

ASHUTHOSH CHANDRAMOULI



MOTIVATION TO DO THE PROJECT



HEALTH IS WEALTH. A HEALTHY MAN IS MORE PROACTIVE AND EFFICIENT AT WORK. HE WILL BE AN ASSET TO AN ORGANIZATION AND TO THE SOCIETY.



IN ORDER TO BE HEALTHY, A PERSON MUST EAT RIGHT AND EXERCISE REGULARLY. FOOD PLAYS A MAJOR ROLE IN A PERSON'S HEALTH.



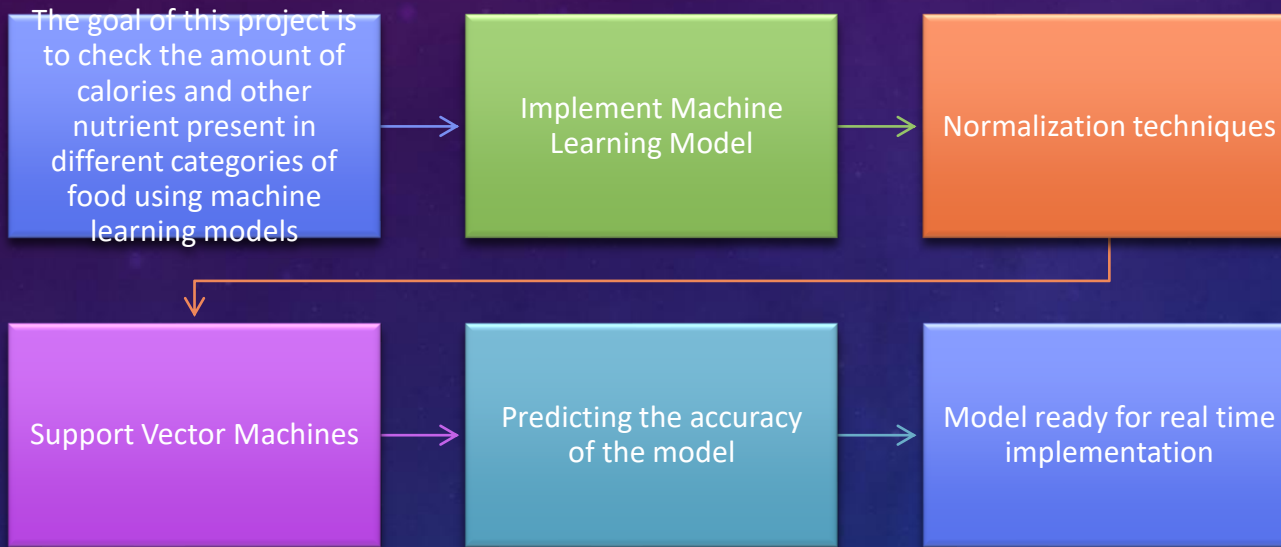
RIGHT FOOD IS THE KEY FOR HEALTHY LIFE.



WE OFTEN SHOP GROCERIES OR READY-TO-EAT PRODUCTS WITHOUT EVEN LOOKING AT THE NUTRIENT PRESENT IN IT. REALIZING IT, I WANTED TO DO AN ANALYSIS WHICH WILL HELP ME PICK RIGHT FOOD PRODUCTS BASED ON THE NUTRIENT LEVELS I AM LOOKING AT.



OBJECTIVES



NUTRIENTS CHART

WE SEE THIS CHART ON MOST OF THE PRODUCTS. BUT NOT ALL OF US CARE MUCH ABOUT THE DATA GIVEN ON THOSE CHARTS. CHOOSING THE RIGHT FOOD IS THE KEY FOR A HEALTHY LIFE

1. Serving Information →

2. Calories →

3. Nutrients →

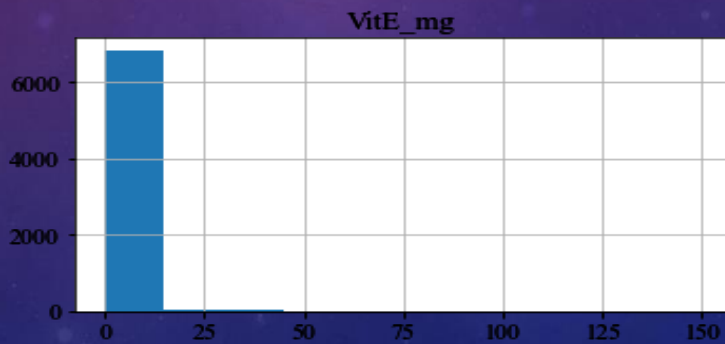
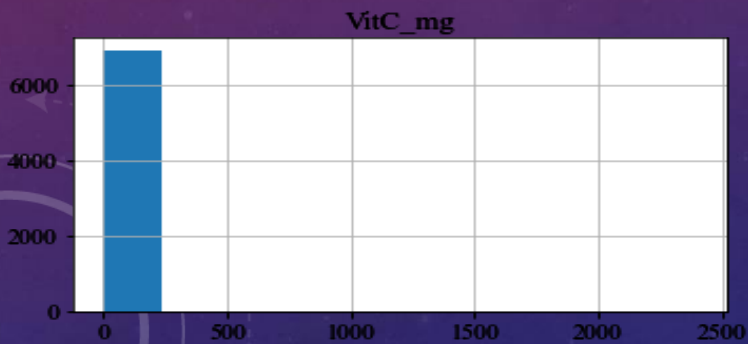
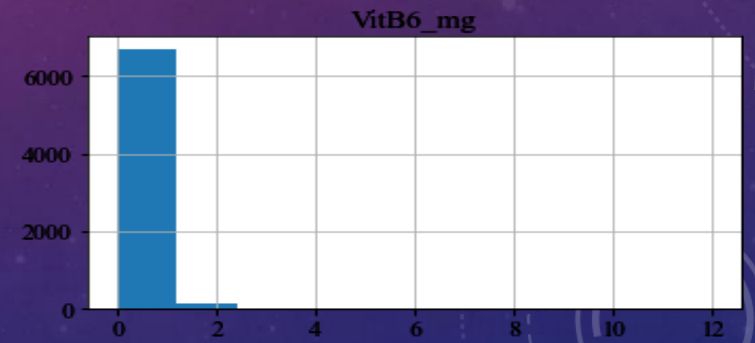
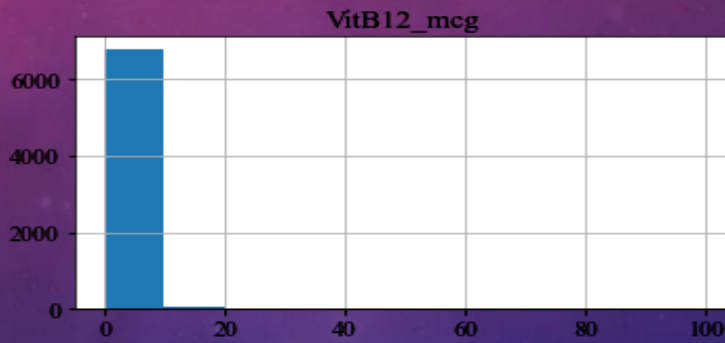
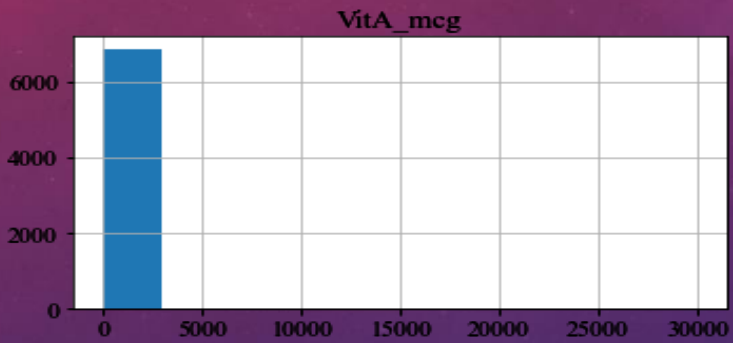
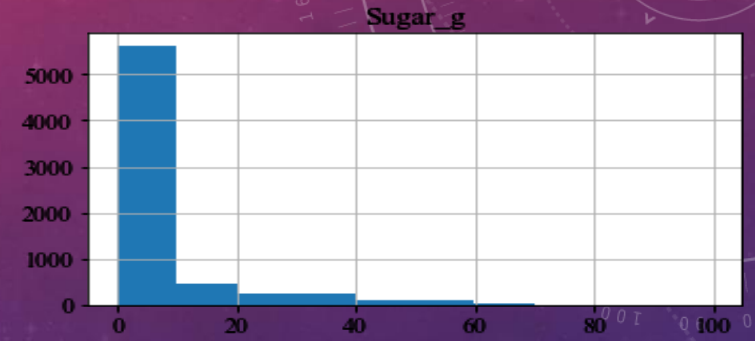
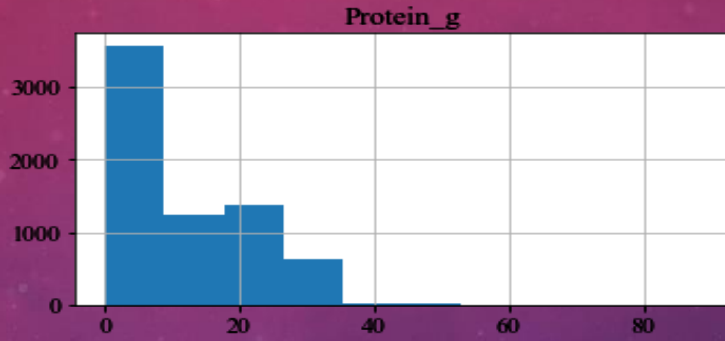
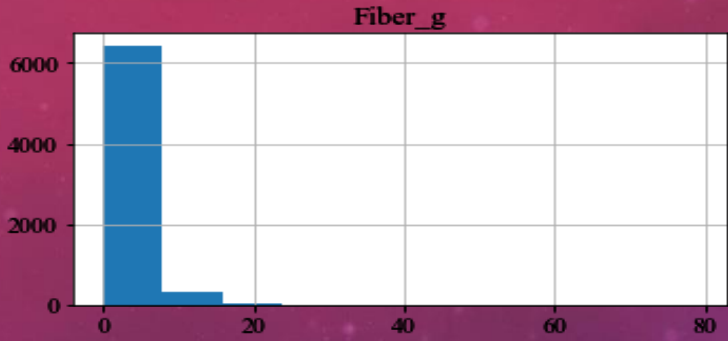
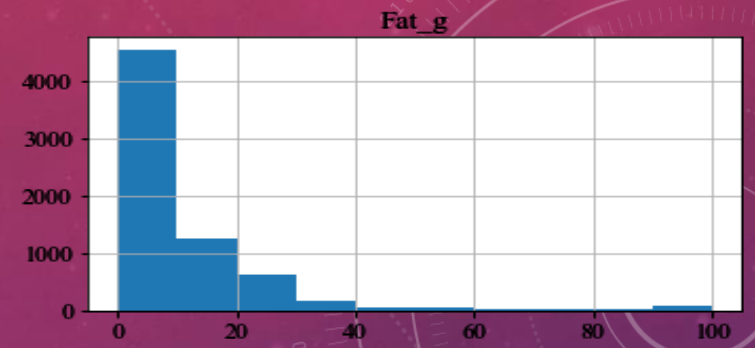
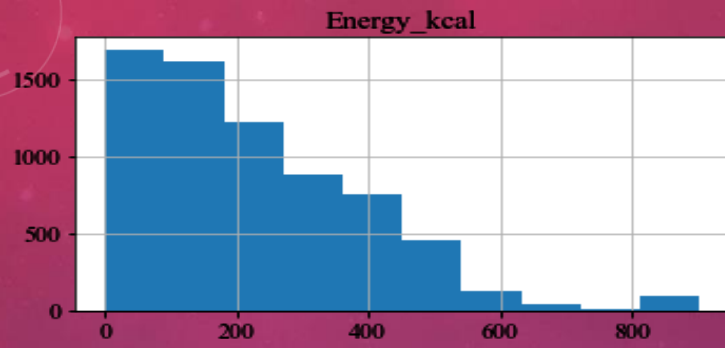
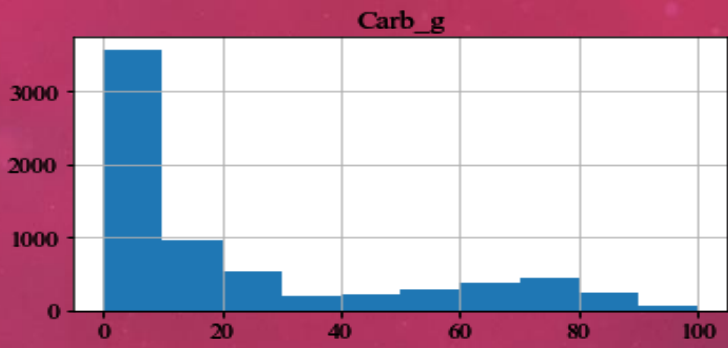
4. Quick Guide to percent Daily Value (%DV) →

- 5% or less is **low**
- 20% or more is **high**

Nutrition Facts	
4 servings per container	
Serving size	1 cup (227g)
Amount per serving	
Calories	280
% Daily Value*	
Total Fat 9g	12%
Saturated Fat 4.5g	23%
Trans Fat 0g	
Cholesterol 35mg	12%
Sodium 850mg	37%
Total Carbohydrate 34g	12%
Dietary Fiber 4g	14%
Total Sugars 6g	
Includes 0g Added Sugars	0%
Protein 15g	
Vitamin D 0mcg	0%
Calcium 320mg	25%
Iron 1.6mg	8%
Potassium 510mg	10%

* The % Daily Value (DV) tells you how much a nutrient in a serving of food contributes to a daily diet. 2,000 calories a day is used for general nutrition advice.

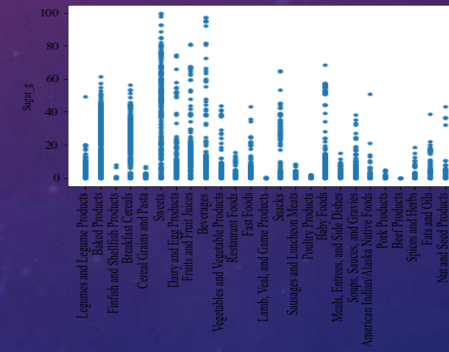
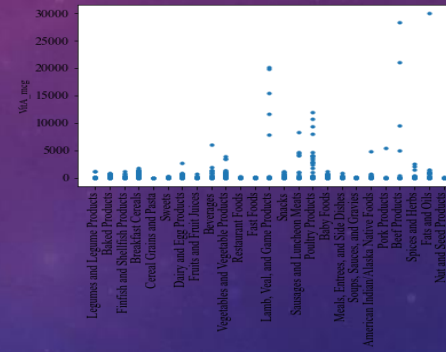
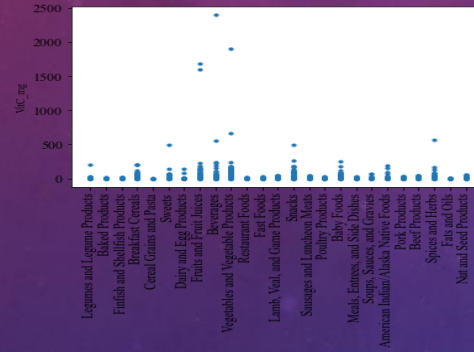
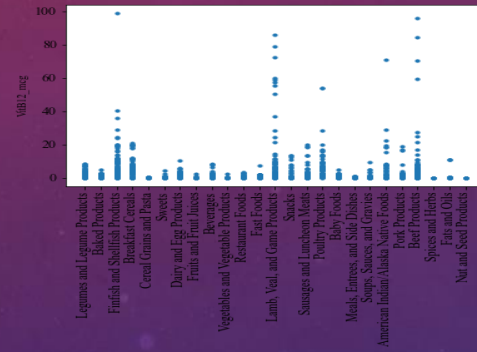
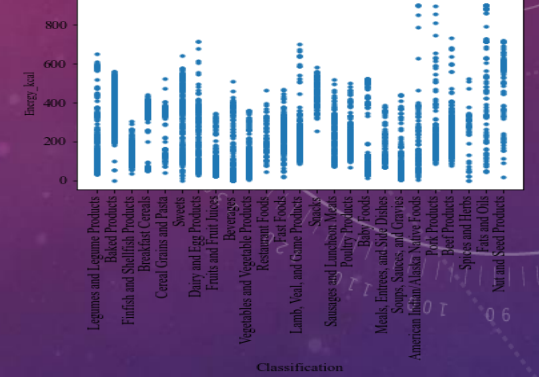
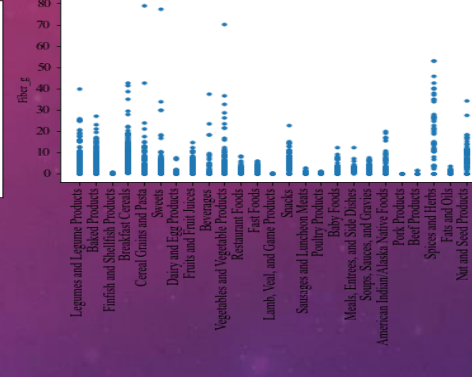
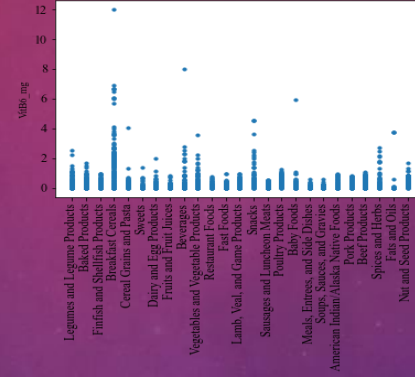
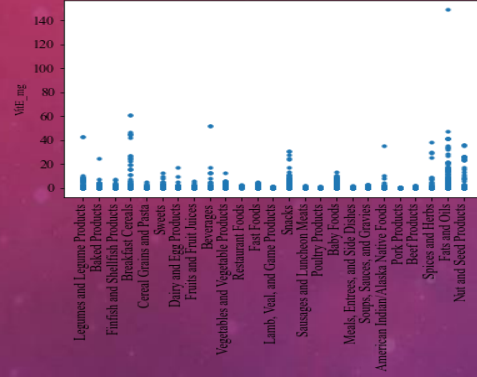
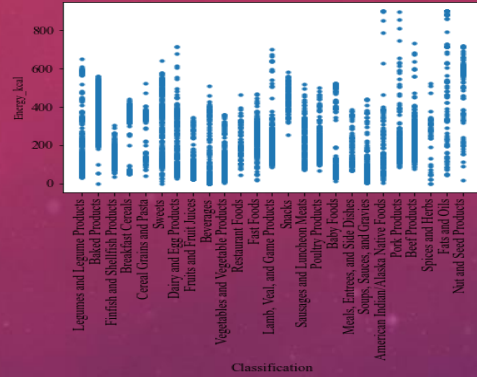




DATA EXAMINATION



CHECKING VALUE DISTRIBUTION IN EACH NUTRIENT



Correlation Matrix for cleaned dataset

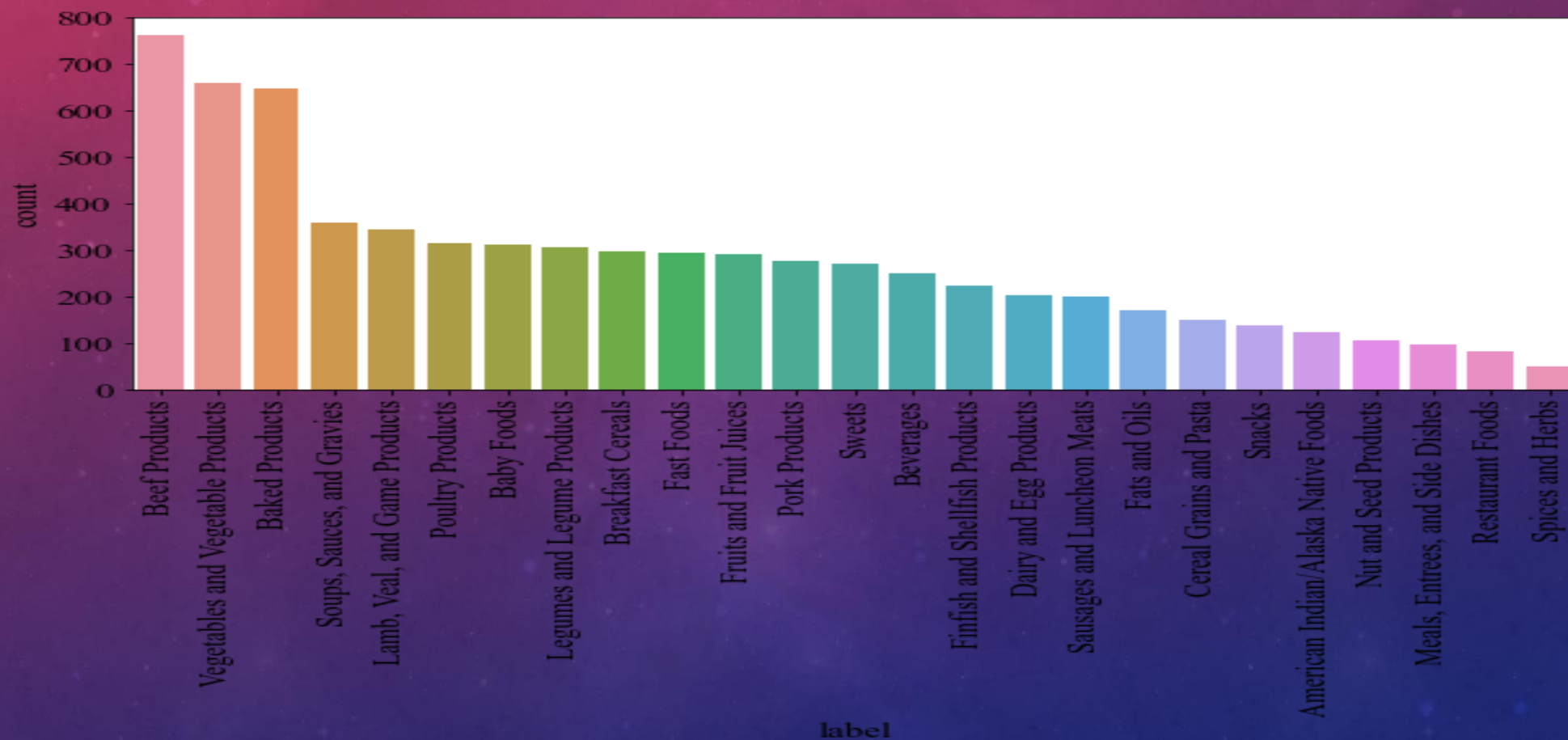
	Energy_kcal	Protein_g	Fat_g	Carb_g	Sugar_g	Fiber_g	VitA_mcg	VitB6_mg	VitB12_mcg	VitC_mg	VitE_mg
Energy_kcal	1	0.11	0.81	0.49	0.31	0.2	0.029	0.12	-0.0094	-0.04	0.3
Protein_g	0.11	1	0.058	-0.31	-0.27	-0.075	0.024	0.23	0.26	-0.075	-0.024
Fat_g	0.81	0.058	1	-0.054	-0.0061	-0.027	0.028	-0.048	-0.02	-0.065	0.33
Carb_g	0.49	-0.31	-0.054	1	0.61	0.45	0.0021	0.2	-0.096	0.072	0.074
Sugar_g	0.31	-0.27	-0.0061	0.61	1	0.11	0.012	0.095	-0.05	0.068	0.072
Fiber_g	0.2	-0.075	-0.027	0.45	0.11	1	0.0079	0.24	-0.051	0.071	0.16
VitA_mcg	0.029	0.024	0.028	0.0021	0.012	0.0079	1	0.12	0.55	0.096	0.045
VitB6_mg	0.12	0.23	-0.048	0.2	0.095	0.24	0.12	1	0.28	0.22	0.29
VitB12_mcg	-0.0094	0.26	-0.02	-0.096	-0.05	-0.051	0.55	0.28	1	0.0099	0.063
VitC_mg	-0.04	-0.075	-0.065	0.072	0.068	0.071	0.096	0.22	0.0099	1	0.068
VitE_mg	0.3	-0.024	0.33	0.074	0.072	0.16	0.045	0.29	0.063	0.068	1

CORRELATION MATRIX

- Fat and Energy have highest correlation
- There is a great variation of correlation



CATEGORY_COUNT



RESULT



CLASSIFICATION_REPORT

	precision	recall	f1-score	support
American Indian/Alaska Native Foods	1.00	0.08	0.15	24
Baby Foods	0.49	0.37	0.42	62
Baked Products	0.70	0.91	0.79	129
Beef Products	0.40	0.99	0.57	152
Beverages	0.50	0.04	0.07	50
Breakfast Cereals	0.93	0.73	0.82	59
Cereal Grains and Pasta	0.49	0.57	0.52	30
Dairy and Egg Products	0.29	0.05	0.08	41
Fast Foods	0.63	0.69	0.66	58
Fats and Oils	0.86	0.71	0.77	34
Finfish and Shellfish Products	0.33	0.34	0.33	44
Fruits and Fruit Juices	0.55	0.64	0.59	58
Lamb, Veal, and Game Products	0.67	0.03	0.06	68
Legumes and Legume Products	0.58	0.57	0.58	61
Meals, Entrees, and Side Dishes	0.00	0.00	0.00	19
Nut and Seed Products	0.73	0.52	0.61	21
Pork Products	0.00	0.00	0.00	55
Poultry Products	0.60	0.05	0.09	63
Restaurant Foods	0.00	0.00	0.00	16
Sausages and Luncheon Meats	0.63	0.60	0.62	40
Snacks	1.00	0.11	0.20	27
Soups, Sauces, and Gravies	0.29	0.19	0.23	72
Spices and Herbs	0.50	0.40	0.44	10
Sweets	0.60	0.65	0.62	54
Vegetables and Vegetable Products	0.46	0.87	0.61	132
accuracy			0.52	1379
macro avg	0.53	0.40	0.39	1379
weighted avg	0.53	0.52	0.45	1379

- There is great variation between the precision and recall values. This is because of data imbalance
- This data imbalance can be fixed using data balance techniques
- I have used SVM model, but other models can be used to improve the accuracy.
- The accuracy can also be improved by considering more x parameters i.e. nutrient values in this case.



LIMITATIONS AND LATER WORK

Limitations:

- Apart from the 12 popularly known nutrient values I have not considered other nutrient values
- Since I have limited the nutrient consideration the accuracy of the model is just 52%
- Not all the SVM parameters has been used.
- Since the data is imbalanced the precision is varying

Later work:

- Other Machine Learning models can be used to improve the accuracy.
- This model can be implemented on real time machine to identify the food products
- Data Balance technique can be employed.





THANK YOU!

[HTTPS://GITHUB.COM/ASHUTHOSHC/RIGHTFOOD_HEALTHYLIFE](https://github.com/ashuthoshc/rightfood_healthyLife)

