Nutrition

Hesham Al Kayed 8/13/2019

1 Introduction

This project aims to build a model that can predict which food is Raw and which is Processed by examining the food nutrition values and use them as predictors. We will use Nutrition data from U.S. DEPARTMENT OF AGRICULTURE website.¹.

In Section 2.1 we Build our data set and format the result as needed. In Section 2.2 we explore our data to find the difference between Raw and Processed food, this exploration will be based on food groups, which will help us make sense of the data at hand. In Section 3 we will build a tree based model and compare the performance of different models.

2 Exploratory data analysis

2.1 Data Structure

Downloaded data will contain a documentation file and data files. as mentioned in the documentation, data files are carrot '^i delimited. We will use and connect NUT_DATA, NUTR_DEF, FOOD_DESC and FD_GROUP tables as described in the documentation. then we will select needed fields and rename them as follow $ProdID = NDB_No$, $NutrVal = Nutr_Val$, $ShrtDesc = Shrt_Desc$, $FoodGroup = FdGrp_Desc$. (left of the equal sign is the new name)

Table 1: Data Set

ProdID	NutrVal	Units	NutrDesc	ShrtDesc	ManufacName	FoodGroup
01001	717	kcal	Energy	BUTTER,WITH SALT	NA	Dairy and Egg Products
01001	0	mg	Caffeine	BUTTER,WITH SALT	NA	Dairy and Egg Products
01001	0	mg	Theobromine	BUTTER,WITH SALT	NA	Dairy and Egg Products
01001	24	mg	Calcium, Ca	BUTTER,WITH SALT	NA	Dairy and Egg Products
01001	2	mg	Magnesium, Mg	BUTTER,WITH SALT	NA	Dairy and Egg Products
01001	24	mg	Phosphorus, P	BUTTER,WITH SALT	NA	Dairy and Egg Products

Table 2: Types and Missing Data

	ProdID	NutrVal	Units	NutrDesc	ShrtDesc	ManufacName	FoodGroup
Type	character	double	character	character	character	character	character
NA	0	0	0	0	0	426413	0
Empty	0	0	0	0	0	0	0

Table 3: Unique Values

ProdID	NutrVal	Units	NutrDesc	ShrtDesc	ManufacName	FoodGroup
7793	17446	5	97	7790	139	25

As we can see in *Table.2* all variables are complete except for ManufacName, which as per downloaded documentation is only available if possible. Our data set have 7 variables and 468669 records. we can see from *Table.3* that we have 7793 products and 7790 descriptions which indicates a duplicated product ID.

¹https://www.ars.usda.gov/ARSUserFiles/80400525/Data/SR-Legacy/SR-Leg ASC.zip

Table 4: Duplicated IDs

ProdID	ShrtDesc
04657	OIL,INDUSTRIAL,PALM KERNEL (HYDROGENATED),CONFECTION FAT
04658	OIL,INDUSTRIAL,PALM KERNEL (HYDROGENATED),CONFECTION FAT
13351	BEEF, CHUCK, UNDER BLADE CNTR STEAK, BNLESS, DENVER CUT, LN, 0" FA
13352	BEEF, CHUCK, UNDER BLADE CNTR STEAK, BNLESS, DENVER CUT, LN, 0" FA
25000	POPCORN,OIL-POPPED,LOFAT
25001	POPCORN,OIL-POPPED,LOFAT

by removing duplicated IDs we get the below table.

Table 5: Unique Values

ProdID	NutrVal	Units	NutrDesc	ShrtDesc	ManufacName	FoodGroup
7790	17443	5	97	7790	139	25

2.2 Data Description

As we want to build a model to classify raw foods from processed food we will start by exploring the difference between raw and processed food. we will separate raw foods from processed foods by searching for 'RAW', in the short description field using the following Reg. expression '(?<![:alpha:])RAW(?![:alpha:])', and we will consider every product in Sausages and Luncheon Meats as processed product and every product in Spices and Herbs as raw food, Available data have 6374 Processed products and 1416 raw foods.².

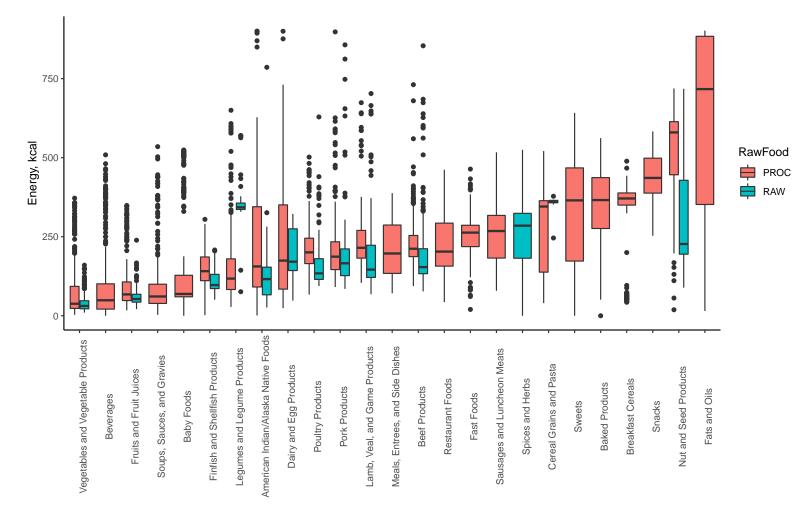


Figure 1: Comparing Raw and Processed Foods Energy Values

²We have to note that this method of differentiating between raw and processed food may not include all cases. but because we don't know how the data is constructed and their is no mention to this information in the documentation, we will assume that this method is correct

Fig.1 shows that Processed food on average have more kaloris per 100g. we can see that this does not apply to Legumes and Legume Product. Fig.2 Shows the median value for each of the energy sources in food, we can see that most raw foods do not have sugar in them and processed food does, raw Legumes and Legume Product have more protein and more carbohydrate than processed Legumes and Legume Product, which can explain why raw Legumes and Legume Product has more energy than processed Legumes and Legume Product and other food groups don't.

Fiber content per 100g does not deffer much between processed and raw foods, except for Legumes and Legume Products as shown in Fig.3. we can see in Fig.4 that all raw foods except Legumes and Legume Product have more water content.

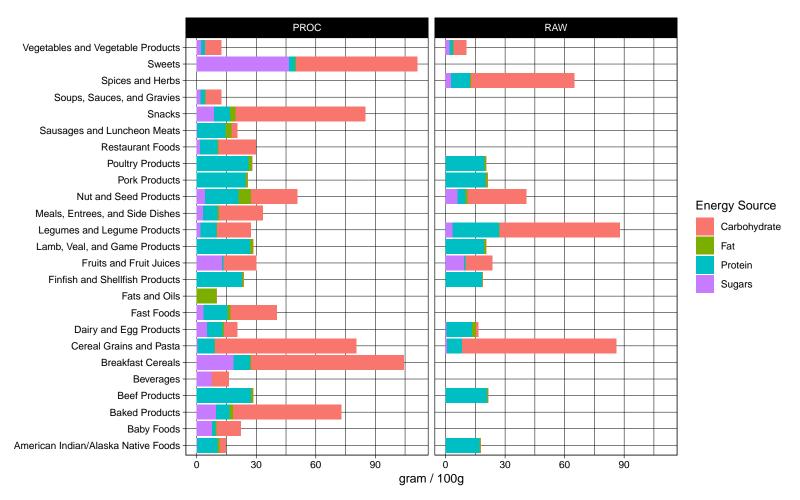


Figure 2: Energy Sources In Food

Fig. 5 and Fig. 6 show the content of Minerals and Vitamins in each food group and enable us to compare raw and processed foods in that group. scale used is based on the **median value** for each group. Numerical values shows **q75**, **median and q25** (Upper, meddle and lower).

At first glance we can see that raw foods rarely contains any Fluoride, F and most processed foods contain Fluoride, F. Sodium, Na will range higher in processed products compared to raw foods, but again Legumes and Legume Products values are higher in its raw form compared to its processed form.

Calcium, Ca to my surprise scours the highest in Spices and Herbs in its raw form, with more than double the value compared to Dairy and Egg Products in its processed form. Spices and Herbs have the highest Iron, Fe, Manganese, Mn in all food groups, and high content value for other Minerals, although you cant eat much Spices and Herbs in one day, mixing them with food seems like a nutritious idea.

we can find another difference in Vintamin C as we can see that it rangers higher in raw foods, and its main source are plants sourced food, Vegetables, Fruits, Nuts/Seeds and Spices/Herbs.

above are a small part of our data. we can further explore the ratio of fat and carbohydrates different types in each food group and much more. but we cant explore every detail in this project because this is a much larger subject, and i don't have domain knowledge to do so.

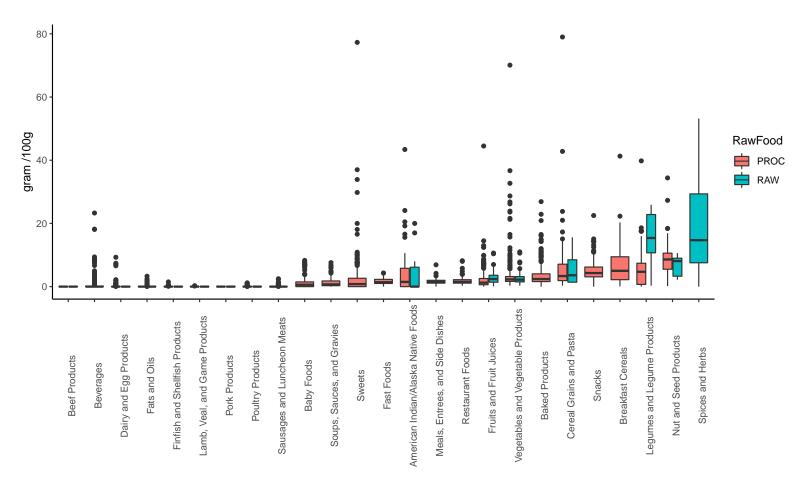


Figure 3: Fiber In Food

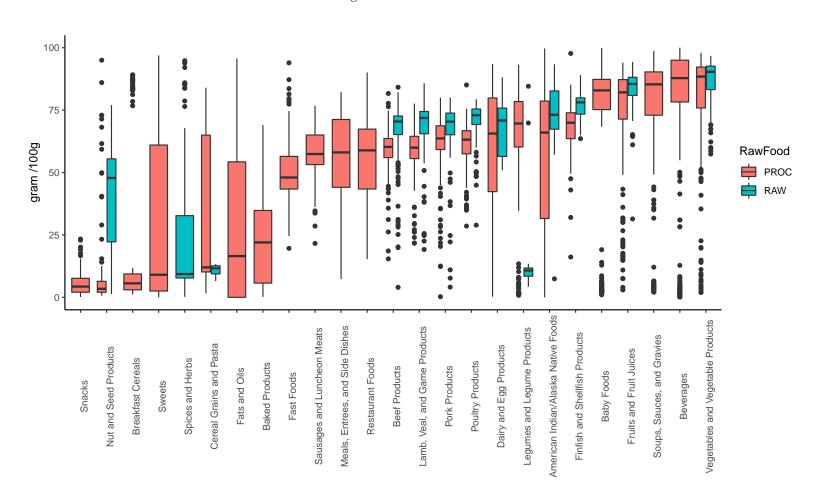


Figure 4: Water In Food

	merican Indian/Alaska Native Foods	Baby Foods	Baked Products	Beef Products	Beverages	Breakfast Cereals	Cereal Grains and Pasta	Dairy and Egg Products	Fast Foods	Fats and Oils	Finfish and Shellfish Products	Fruits and Fruit Juices	Lamb, Veal, and Game Products	Legumes and Legume Products	Meals, Entrees, and Side Dishes	Nut and Seed Products	Pork Products	Poultry Products	Restaurant Foods	Sausages and Luncheon Meats	Snacks	Soups, Sauces, and Gravies	Spices and Herbs	Sweets	Vegetables and Vegetable Products	
n, Ca	53.75 22 7	125 34 11	102.5 52 26	16 12 7	40.25 5 2	242.5 55 19	34 17 8	574.75 141.5 94	154 105 51	13.5 1 0	73.5 37 14	21 11 8	22 14 8	123 54.5 35	108.75 48.5 29	141 71 46	19 10 7	19 14 12	89 34 17	42 15 10	135 58 21	33 17 10	0 0 0	117 54.5 8.75	48.75 26 14	PROC
Calcium,	24.5 13 6.25	0 0 0	0 0 0	16 12 5	0 0 0	0 0 0	28.5 21 9	113 64 56	0 0 0	0 0 0	47 25 12	24 12 8	15 10 6	174.5 130.5 89.75	0 0 0	54.25 33.5 17.5	16 12 6.5	13 11 7.5	0 0 0	0 0 0	0 0 0	0 0 0	1139.5 383 112.5	0 0 0	67 35.5 17	RAW
um, Mg	62 27 19	21 10 6	40 25 16	24 22 20	11 3 1	106 61 27	93.75 40.5 21	28 16 11	25 22 19	2 0 0	42.5 36 30	15 9 5	26 24 22	74 46 29	30.25 20 15.75	304.25 217.5 109.75	24 21 18	26 23 20	28 22 17	19 16 13	109.25 75.5 50.75	17 8 4	0 0 0	44.25 14 3	30 20 13	PROC
P Magnesium, Mg	27.5 25 21.5	0 0 0	0 0 0	23 20 17	0 0 0	0 0 0	129.5 79 30	13 11 9	0 0 0	0 0 0	34 30 25	17 11 8	24 22 19	196.5 183 157	0 0 0	87 55 32	23 21 16.5	24 22 19	0 0 0	0 0 0	0 0 0	0 0 0	285 174.5 73.75	0 0 0	38 23 14	RAW
orus, P N	303.5 170 59.5	109 41 20	202.75 122 89	233 210 183	31.5 5 1	346.5 222 81	274.25 119.5 76	477.25 136.5 98.5	225 184 131	16 0 0	304.5 257 215.5	25 12 9	237 210 187.25	231 142 99	165.75 117.5 72	666 456 202	267 237 216.25	236.75 203 169.5	235 130 89	192 157 129.5	292 207 145	65 35 20	0 0 0	142 82 8	70 42.5 27	PROC
Phosphorus,	216.5 150 40	0 0 0	0 0 0	213 198 175	0 0 0	0 0 0	298 221 111.5	390 208 186	0 0 0	0 0 0	245 210 187	26 17 12	208 186 163	439.5 396.5 367	0 0 0	178.25 117.5 89.5	227.5 207 186	216.5 184 160	0 0 0	0 0 0	0 0 0	0 0 0	316 173.5 69	0 0 0	75.25 49.5 30.75	RAW
	458 309 114	200 125 77	201.25 136.5 103	365 316.5 272.5	116.5 25 8	358 226 120.5	287 140 56.5	219 152 115	250 202 177	36.75 10 0	469 353 251.5	200 128 89	339 306 248	438.75 311 155.5	242.75 191.5 152.25	733 582 424	388 344 275	266 242.5 199	305 187 136	323.25 252.5 192.75	412.75 286 216.75	255 140 71	0 0 0	284 143 36.5	378.75 249.5 164	PROC
Potassium, K	370 317 148	0 0 0	0 0 0	343.75 317 283	0 0 0	0 0 0	274 195 100.5	163 135 121	0 0 0	0 0 0	394 350 261	266 178.5 126.5	323.25 288.5 237.75	1405.25 1250 978.5	0 0 0	569.25 465.5 364.25	376.5 341 280	256 228 204	0 0 0	0 0 0	0 0 0	0 0 0	1441 1000 458	0 0 0	418.5 317 228.5	RAW
, Na	220.5 71 13.75	49 25.5 6.75	580 438 327	69 63 54.75	48.25 13 4	537.5 344 13.5	12 5 2	628 112 53	671.25 566 432.75	779 28 0	399 124 73.5	7 4 2	84 72 61	281.5 65 10	591 423 349	143 28 4	988 100 67	237 99 75	618 430 344	1149.5 960 813.5	538.5 318 204.5	708.25 406.5 306.5	0 0 0	188.75 79 33.25	255 120.5 14	PROC
Sodium, Na	54.5 43 8.5	0 0 0	0 0 0	67 58 53	0 0 0	0 0 0	5 5 2.5	166 142 128	0 0 0	0 0 0	104 70 52	4 1.5 1	83 70.5 59	17.75 13 7.25	0 0 0	16.25 6 2.75	87 63 52.5	113.5 77 63	0 0 0	0 0 0	0 0 0	0 0 0	118 50 19	0 0 0	27 13 5	RAW
де, Т	0 0 0	21.125 8.15 1.025	49.15 25.05 15.25	22.4 22.4 22.4	78.65 60.6 34.75	71.6 53.35 21	18 7 6	12 8.7 3.05	29.65	27.525 22.55 13.675	140.25 31 31	62.95 19.8 7.15	26.5 21 13	9.7 3.1 2.4	56.5 56.5 56.5	7.5 5 2.5	0 0 0	14.7 14.7 14.7	0 0 0	41.2 36 20.3	2.5 0.1 0	65.5 42.7 35.95	0 0 0	25.7 7 1.6	37 25.6 6.9	PROC
Fluoride,	0 0 0	0 0 0	0 0 0	22.4 22.4 22.4	0 0 0	0 0 0	0 0 0	1.1 1.1 1.1	0 0 0	0 0 0	33.025 25.55 18.075	4 2.2 1.5	0 0 0	2.2 2.2 2.2	0 0 0	3.4 3.4 3.4	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	18.1 2 1.75	0 0 0	3.6 2 1.2	RAW
m, Se	53.575 16.55 5.25	8.4 1.9 0.6	23.4 13.1 6.1	33.3 28.7 22.875	0.5 0.1 0	21.95 12.8 6.15	31.9 15.5 8.3	14.75 7.35 2.6	24.9 20.1 16.575	1.6 0 0	50.025 45.3 32.75	0.4 0.3 0.1	26.175 13.5 9.3	7.15 2.8 1.3	20.1 13 9.5	34.4 11 4.1	44.7 37.9 27.5	31.4 26.55 21.275	18.85 15.2 6.95	22.825 17.45 12.5	15.7 8.95 5.8	4.85 2.8 1.1	0 0 0	3.2 1.9 0.7	1 0.7 0.4	PROC
Selenium,	30.45 14.7 4.35	0 0 0	0 0 0	25.2 22.2 19.2	0 0 0	0 0 0	22.25 15.1 12.025	53.5 36.4 30.7	0 0 0	0 0 0	36.5 36.5 23.7	0.6 0.3 0.1	19.275 8.9 5.5	10.925 8.2 4.225	0 0 0	10.1 7 6.2	34.6 29.4 22.65	22.9 20.6 15.5	0 0 0	0 0 0	0 0 0	0 0 0	12.45 5.1 2.95	0 0 0	0.9 0.7 0.5	RAW
e B	4.715 2.6 0.988	2.38 0.8 0.3	3.43 2.53 1.62	3.062 2.67 2.277	0.25 0.06 0.02	28.6 9.09 3.48	4.188 2.015 1.05	0.42 0.16 0.05	2.32 1.765 0.998	0.2 0.02 0	1.455 0.85 0.495	0.75 0.35 0.252	2.388 2.03 1.518	2.59 1.7 1.105	2.13 1.34 0.95	5.4 3.53 2.53	1.13 0.94 0.82	1.677 1.305 0.98	1.22 0.84 0.66	1.768 1.24 0.94	3.185 1.89 1.42	0.94 0.58 0.288	0 0 0	1.152 0.405 0.09	1.268 0.73 0.462	PROC
Iron,	4.7 1.8 0.5	0 0 0	0 0 0	2.27 1.97 1.69	0 0 0	0 0 0	3.66 1.96 1.045	3.7 3.4 1.74	0 0 0	0 0 0	1.13 0.59 0.34	0.55 0.25 0.15	1.918 1.55 1.205	8.252 6.6 5.005	0 0 0	2.8 1.865 1.008	0.985 0.84 0.615	2.61 1.09 0.77	0 0 0	0 0 0	0 0 0	0 0 0	32.125 13.97 4.49	0 0 0	1.7 0.85 0.47	RAW
Zu	3.195 1.43 0.41	1.15 0.49 0.14	1.01 0.67 0.46	7.393 5.77 4.79	0.13 0.03 0.01	6.755 2.5 0.795	1.92 1.05 0.642	2.768 0.8 0.4	1.725 1.25 0.71	0.11 0 0	1.295 0.76 0.51	0.18 0.1 0.07	5.27 4.195 3.032	1.8 1.03 0.62	1.058 0.88 0.542	5.29 3.76 2.025	2.94 2.38 1.97	2.982 2.1 1.52	1.16 0.74 0.51	2.545 2.1 1.495	2.64 1.42 0.98	0.64 0.3 0.17	0 0 0	0.892 0.36 0.07	0.49 0.32 0.2	PROC
Zinc,	1.545 0.52 0.24	0 0 0	0 0 0	5.315 4.07 3.54	0 0 0	0 0 0	2.13 1.68 1.13	2.3 1.41 1.3	0 0 0	0 0 0	1 0.62 0.42	0.14 0.1 0.065	3.728 3.07 2.285	3.65 3.205 2.615	0 0 0	1.172 0.93 0.518	2.47 1.93 1.695	2.735 1.84 1.28	0 0 0	0 0 0	0 0 0	0 0 0	4.7 2.84 1.07	0 0 0	0.61 0.37 0.22	RAW
er, Cu	0.22 0.113 0.048	0.115 0.055 0.04	0.205 0.129 0.073	0.112 0.09 0.078	0.037 0.01 0.005	0.34 0.202 0.079	0.284 0.181 0.098	0.05 0.028 0.016	0.109 0.085 0.074	0.009 0 0	0.228 0.078 0.046	0.107 0.063 0.04	0.155 0.129 0.113	0.44 0.25 0.178	0.141 0.096 0.072	1.6 1.18 0.69	0.119 0.1 0.078	0.114 0.075 0.06	0.097 0.069 0.06	0.11 0.08 0.06	0.406 0.264 0.148	0.118 0.081 0.04	0 0 0	0.235 0.056 0.018	0.149 0.088 0.052	PROC
Copper, (0.147 0.094 0.051	0 0 0	0 0 0	0.091 0.073 0.063	0 0 0	0 0 0	0.412 0.277 0.215	0.064 0.062 0.032	0 0 0	0 0 0	0.188 0.054 0.034	0.092 0.068 0.042	0.138 0.11 0.091	1.085 0.897 0.764	0 0 0	1.186 0.505 0.374	0.09 0.072 0.056	0.142 0.077 0.054	0 0 0	0 0 0	0 0 0	0 0 0	0.926 0.572 0.336	0 0 0	0.189 0.109 0.062	RAW
ese, Mn	0.409 0.11 0.038	0.144 0.07 0.016	0.439	0.016 0.012 0.01	0.122 0.026 0.007	2.745 0.852 0.231	1.283 0.679 0.376	0.038 0.015 0.006	0.311 0.241 0.171	0.03 0 0	0.074 0.027 0.019	0.235 0.088 0.039	0.03 0.022 0.011	1.174 0.51 0.382	0.374 0.224 0.186	2.603 2.077 1.421	0.022 0.017 0.011	0.028 0.019 0.015	0.268 0.212 0.135	0.059 0.038 0.021	1.458 0.976 0.44	0.25 0.146 0.068	0 0 0	0.358 0.057 0.008	0.371 0.22 0.14	PROC
Manganese,	0.32 0.038 0.024	0 0 0	0 0 0	0.014 0.012 0.008	0 0 0	0 0 0	2.243 1.322 1.094	0.055 0.038 0.032	0 0 0	0 0 0	0.064 0.019 0.015	0.143 0.055 0.032	0.024 0.015 0.009	1.795 1.536 1.183	0 0 0	1.523 1.252 0.55	0.012 0.011 0.009	0.02 0.018 0.011	0 0 0	0 0 0	0 0 0	0 0 0	7.642 2.374 1.169	0 0 0	0.404 0.229 0.153	RAW

	1310 Merican Indian/Alaska Native Foods	Baby Foods	Baked Products	Beef Products	Beverages	Breakfast Cereals	Cereal Grains and Pasta	Dairy and Egg Products	Fast Foods	Fats and Oils	Finfish and Shellfish Products	Fruits and Fruit Juices	Lamb, Veal, and Game Products	Legumes and Legume Products	Meals, Entrees, and Side Dishes	Nut and Seed Products	Pork Products	Poultry Products	Restaurant Foods	Sausages and Luncheon Meats	Snacks	Soups, Sauces, and Gravies	Spices and Herbs	Sweets	Vegetables and Vegetable Products	
Vitamin A, IU	1310 261 40 599 210 84.5	1040 197 40 0 0	106.5 6 0 0 0	14 5 0 12 6 0	49 0 0 0 0	1681.25 148 0 0 0	10 0 0 0 0	753.5 255 72 1043 570 540	292 164.5 58.5 0 0	273.5 0 0 0 0	180 88.5 50.25 148.25 91.5 46.25	376 73 19.25 326 66 33	6 0 0 11 0	24 0.5 0 39.75 8 0	336 206 101 0 0	66 14.5 0 213.5 32.5 0	32 7 0 8 6	105.25 55.5 21.75 141.5 59 21	340 189 25.5 0 0	74.5 0 0 0 0	226.75 41.5 0 0 0	514 196 12 0 0	0 0 0 4127 363	157.5 25 0 0 0	1852 244 12 2167 333 13	PROC RAW
Vitamin A, RAE	95.5 19 0 67 19 1.5	113.25 34 3 0 0	23 1 0 0 0 0	4 2 0 4 2 0	2 0 0 0 0	490.75 7 0 0 0 0	0 0 0 0	197.25 62.5 14 313 171 156	63 27 6 0 0	18 0 0 0 0	54.5 27 15 44.5 27.5	19 4 1 16 3 2	2 0 0 3 0	1 0 0 2 0	39 23 10 0 0	3 1 0 10.5 1.5	10 2 0 2 1	31 16 6 42.5 18 6.5	55.5 13.5 4 0 0	13 0 0 0	15.5 2 0 0 0	37.75 17 2 0 0	0 0 0 210.5 17 0	33.5 1 0 0 0	93.5 15 1 108 17 1	PROC RAW
Vitamin D	71.5 15.5 0 0 0	41.5 1 0 0 0	0 0 0 0 0	6 5 3 4 3	0 0 0 0	0 0 0 0 0	0 0 0 0 0	40 15 0 123.75 75.5 59.5	2 0 0 0	0 0 0 0 0	512.75 99.5 3.5 408 99 8	0 0 0 0 0	1 0 4 1 0	0 0 0 0 0	1 0 0 0 0	0 0 0 0 0	38 31 23 25.25 20.5 14.75	10 5 3 17 12.5 5	7 4 0 0 0 0	40 27 12 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0	0 0 0 0 0 0	PROC RAW
Vitamin D2 (ergocalciferol)	0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0 0	1 1 0 0 0	0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0 0	1.2 1.05 1 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0 0	0.2 0.2 0.2 0 0	0 0 0 0 0	0 0 0 0 0	0.7 0.3 0.2 26.675 2.9 0.275	PROC RAW
Vitamin D3 (cholecalciferol)	0 0 0 0 0	0.4 0 0 0 0	0 0 0 0 0	0.2 0.1 0.1 0.1 0.1 0.1	9.1 2.4 1.075 0 0	3.55 3.25 1.8 0 0	0 0 0 0	1.2 0.5 0.1 3.1 2 1.6	0.2 0.2 0.15 0 0	4.8 1.6 0 0 0	12.1 2.2 0.1 9.05 1.7 0.2	0.25 0 0 0 0	0.1 0 0 0.1 0	0 0 0 0 0	0.15 0.1 0.05 0 0	0 0 0 0 0	1 0.8 0.6 0.625 0.5 0.4	0.3 0.1 0.1 0.4 0.3 0.1	0.125 0.05 0 0 0	0.9 0.6 0.2 0 0	0.125 0 0 0 0	0 0 0 0 0	0 0 0 0 0	0.2 0.1 0 0 0	0.2 0 0 0 0	PROC RAW
Vitamin D (D2 + D3)	1.8 0.35 0 0 0	1.05 0 0 0 0	0 0 0 0 0	0.2 0.1 0.1 0.1 0.1 0.1	0 0 0 0	3.2 0 0 0 0	0 0 0 0 0	1 0.4 0 3.1 1.85 1.475	0.2 0.1 0 0 0	0 0 0 0 0	12.825 2.5 0.1 10.2 2.5 0.2	0 0 0 0 0	0.1 0 0 0.1 0	0 0 0 0 0	0.1 0 0 0 0	0 0 0 0 0	1 0.8 0.6 0.625 0.5 0.4	0.3 0.1 0.1 0.4 0.3 0.1	0.2 0.1 0 0 0	1 0.7 0.3 0 0	0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0	PROC RAW
Vitamin C, total ascorbic acid	1.925 0 0 10.1 4 0	28.9 8.4 1.7 0 0	0.2 0 0 0 0	0 0 0 0 0	24.7 0.4 0 0 0 0	2.7 0 0 0 0	0 0 0 0 0	0.8 0 0 0 0	1.225 0.5 0 0 0	0.1 0 0 0 0	1.525 0 0 1.6 0	21.1 3.8 1.7 39.275 26.25 10	0 0 0 0 0	1.2 0 0 4.5 0.1	2.4 0.7 0 0 0	2.2 0.7 0 26.575 4.45 1.025	0 0 0 0.35 0	0 0 0 2.35 0	2.3 0.95 0.275 0 0	1.175 0 0 0 0	8.325 0.3 0 0 0	1.9 0.6 0 0 0	0 0 0 32.1 13.3 0.7	0.8 0.2 0 0 0	18.2 9.7 4.6 38.2 17.1 5.6	PROC RAW
Vitamin K (phylloquinone)	1.2 0.2 0 14.8 1.4 0.2	9.8 4.7 0.6 0 0	9.5 5 1.8 0 0	1.6 1.6 1.5 1.5 1.5 1.3	0.1 0 0 0 0	1.9 0.65 0 0 0	1.25 0.3 0 2.05 0.9 0.35	2.325 1 0.2 0.7 0.35 0.3	9.5 6.5 4.5 0 0	72.95 39.5 10.8 0 0	0.4 0.1 0.1 0.1 0.1 0.1	4.3 1.5 0.3 4.5 2.2 0.4	5.2 3.9 1.45 3.4 1.25 1.1	3.5 2.3 0.375 9 5.6 5	10 3.8 1.7 0 0	7.55 2.7 0 17.15 0.2 0.15	0 0 0 0	2.4 0 0 2.4 0	32.95 22.2 4.75 0 0	1.6 0.3 0 0 0	13.3 5.2 1.6 0 0	5.2 1.6 0.3 0 0	0 0 0 289.25 24.6 0.2	2.975 0.4 0 0 0	47.025 8.7 2 74.2 10.95 1.325	PROC RAW
Vitamin E (alpha- tocopherol)	0.865 0.44 0.11 1.37 0.75 0.395	1.32 0.55 0.145 0 0	1.34 0.65 0.3 0 0	0.39 0.19 0.12 0.31 0.23 0.17	0.112 0 0 0 0	0.86 0.44 0.11 0 0	0.508 0.17 0.05 0.655 0.11 0.095	0.402 0.18 0.04 2.243 1.185 0.78	0.97 0.71 0.38 0 0	11.495 5 1.988 0 0	1.742 1.065 0.638 1.2 0.745 0.498	0.535 0.2 0.088 0.32 0.18 0.1	0.485 0.385 0.2 0.348 0.26 0.21	1.01 0.35 0.07 0.715 0.22 0.21	0.8 0.53 0.27 0 0	22.89 3.805 0.725 0.9 0.54 0.24	0.27 0.23 0.19 0.235 0.19 0.13	0.33 0.26 0.15 0.245 0.2 0.11	1.21 0.82 0.53 0 0	0.45 0.25 0.19 0 0	6.06 2.735 0.8 0 0	0.72 0.28 0.08 0 0	0 0 0 7.48 1.69 0.315	0.6 0.18 0.01 0 0	0.96 0.29 0.08 0.847 0.355 0.103	PROC RAW
Vitamin B–12	6.987 2.57 0.055 7.298 2.615 0.288	0.4 0.12 0 0 0	0.14 0.025 0 0 0	3.325 2.56 1.94 2.957 2.075 1.632	0.028 0 0 0 0	2.925 0 0 0 0	0 0 0 0 0	1.185 0.5 0.32 1.9 1.61	0.74 0.51 0.31 0 0	0.062 0 0 0 0	4.97 2.69 1.6 3.71 2.03 1.3	0 0 0 0	2.97 2.52 1.62 2.73 2.34 1.68	0 0 0 0	0.448 0.28 0.14 0 0	0 0 0 0	0.79 0.64 0.502 0.83 0.56 0.51	1.058 0.38 0.28 2.05 0.68 0.395	0.74 0.25 0.125 0 0	1.672 1.07 0.668 0 0	0.155 0 0 0 0	0.2 0.05 0 0 0	0 0 0 0 0	0.283 0.01 0 0 0	0 0 0 0 0	PROC RAW
Vitamin E, added	0 0 0 0 0 0 0	1.2 0 0 0 0	0 0 0 0 0	0 0 0 0 0 0 0	0 0 0 0 0	0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0	0 0 0 0 0	0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	PROC RAW
Vitamin B-12, added	0 0 0 0 0 0 0	0.215 0 0 0	0 0 0 0	0 0 0 0 0 0 0	0 0 0 0 0 0 0	4.84 0 0 0	0 0 0 0 0 0 0	0 0 0 0 0 0 0	0 0 0 0 0 0 0	0 0 0 0	0 0 0 0 0 0 0	0 0 0 0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0 0 0 0	0 0 0 0 0 0 0	0 0 0 0 0 0 0	0 0 0 0 0 0 0	0 0 0 0 0 0 0	0 0 0 0 0 0 0	0 0 0 0 0 0 0	0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	0 0 0 0 0 0 0	0 0 0	PROC RAW
Vitamin B–6	0.377 0.143 0.06 0.361 0.196 0.079	0.13 0.06 0.039 0 0	0.105 0.06 0.038 0 0	0.563 0.41 0.322 0.612 0.498 0.361	0.057 0.011 0 0 0	1.66 0.8 0.13 0 0	0.318 0.114 0.046 0.388 0.171 0.164	0.08 0.05 0.037 0.35 0.226 0.15	0.204 0.132 0.09 0 0	0.01 0 0 0 0	0.389 0.264 0.133 0.4 0.273 0.146	0.075 0.044 0.027 0.1 0.055 0.042	0.387 0.208 0.13 0.395 0.19 0.132	0.231 0.106 0.074 0.422 0.366 0.32	0.145 0.117 0.087 0 0	0.641 0.323 0.146 0.405 0.324 0.163	0.515 0.437 0.346 0.618 0.486 0.37	0.538 0.374 0.233 0.628 0.475 0.35	0.231 0.132 0.084 0 0	0.322 0.218 0.15 0 0	0.532 0.212 0.118 0 0	0.09 0.046 0.02 0 0	0 0 0 1.044 0.397 0.158	0.055 0.032 0.005 0 0	0.18 0.115 0.07 0.209 0.141 0.08	PROC RAW

3 Building Our Model

We want to build a model that can classify if a food product is Raw or Processed using its nutrition values. We will use tree based models, because they are easy to use and they will reveal more differences between raw and processed foods by finding variable importance in our model. we will split our data set based on RawFood field into training and test sets. any nutrients defined as added will be deleted. Below shows the structure of our training set.

we will compare *ctree* from *party* package, *gbm* from *gbm* package and *ranger* from *ranger* package. we will not use FoodGroup field to build our model, but we will keep it to compare the performance of our models through different food groups.

```
## Classes 'tbl_df', 'tbl' and 'data.frame':
                                                 3895 obs. of 97 variables:
                                        : Factor w/ 25 levels "American Indian/Alaska Native Foods",..: 8 8 8 8 8
##
   $ FoodGroup
                                        : Factor w/ 2 levels "PROC", "RAW": 1 1 1 1 1 1 1 1 1 1 ...
##
   $ RawFood
                                               0.029 0.67 0.659 0.409 0.639 0.823 0.958 0.691 0.741 0.757 ...
##
   $ Alanine
                                        : num
                                               0 0 0 0 0 0 0 0 0 0 ...
##
   $ Alcohol, ethyl
                                         : num
##
   $ Arginine
                                        : num
                                               0.031 0.874 0.883 0.467 0.47 ...
##
   $ Ash
                                               2.11 3.18 3.6 1.27 5.2 3.79 4.3 3.55 3.27 3.83 ...
   $ Aspartic acid
##
                                               0.064 1.588 1.502 0.963 0.779 ...
                                        : num
                                                4 0 0 0 0 0 0 0 0 0 ...
##
   $ Beta-sitosterol
##
   $ Betaine
                                               0.3 0 0 0.6 0 0 0 0 0 0.7 ...
                                        : num
##
   $ Caffeine
                                               0 0 0 0 0 0 0 0 0 0 ...
##
   $ Calcium, Ca
                                               24 674 643 111 493 ...
                                        : num
##
   $ Campesterol
                                               0 0 0 0 0 0 0 0 0 0 ...
                                        : num
                                               0.06 2.79 4.78 4.76 3.88 1.55 0.36 0.68 2.77 5.58 ...
##
   $ Carbohydrate, by difference
                                        : num
   $ Carotene, alpha
                                               0 0 0 0 0 0 0 0 0 0 ...
##
                                        : num
                                               158 76 0 6 3 32 33 78 41 51 ...
##
   $ Carotene, beta
                                        : num
   $ Cholesterol
                                               215 94 103 12 89 116 110 89 64 64 ...
##
                                        : niim
##
   $ Choline, total
                                              18.8 15.4 0 16.3 15.4 15.4 15.4 15.4 15.4 14.2 ...
                                        : num
                                               0 0.024 0.042 0.033 0.032 0.025 0.032 0.032 0.025 0.033 ...
##
   $ Copper, Cu
                                        : num
##
   $ Cryptoxanthin, beta
                                               0 0 0 0 0 0 0 0 0 0 ...
                                        : num
                                               0.008 0.131 0.117 0.062 0.083 0.261 0.304 0.123 0.144 0.124 ...
##
   $ Cystine
                                        : num
##
  $ Dihydrophylloquinone
                                              0000000000...
                                        : num
   $ Energy
                                        : num
                                               717 371 387 81 265 389 413 373 254 295 ...
   $ Fatty acids, total monounsaturated: num
                                               21.021 8.598 8.671 0.516 4.623 ...
##
##
   $ Fatty acids, total polyunsaturated: num
                                               3.043 0.784 0.87 0.083 0.591 ...
##
   $ Fatty acids, total saturated
                                        : num
                                               51.37 18.76 19.48 1.24 13.3 ...
                                               3.278 0 0 0.067 0 ...
##
   $ Fatty acids, total trans
                                        : num
##
   $ Fatty acids, total trans-monoenoic: num
                                               2.982 0 0 0.054 0 ...
   $ Fatty acids, total trans-polyenoic: num
                                               0.296 0 0 0.013 0 0 0 0 0 0.138 ...
##
   $ Fiber, total dietary
                                               0 0 0 0 0 0 0 0 0 0 ...
   $ Fluoride, F
                                               2.8 0 0 0 0 0 0 0 0 0 ...
##
                                        : num
##
   $ Folate, DFE
                                        : num
                                               3 20 18 8 32 6 10 18 9 27 ...
##
   $ Folate, food
                                               3 20 18 8 32 6 10 18 9 27 ...
                                        : num
   $ Folate, total
                                               3 20 18 8 32 6 10 18 9 27 ...
                                        : num
##
   $ Folic acid
                                               0 0 0 0 0 0 0 0 0 0 ...
                                        : num
                                               0 0 0 0 0 0 0 0 0 0 ...
##
   $ Fructose
                                        : num
                                               0 0 0 0.12 0 0 0 0 0 0.78 ...
##
   $ Galactose
                                        : num
   $ Glucose (dextrose)
                                        : num
                                               0 0 0 0 0 0 0 0 0 0 ...
   $ Glutamic acid
##
                                               0.178 5.515 5.718 2.446 2.421 ...
                                         : num
                                               0.018 0.437 0.403 0.209 0.097 0.457 0.533 0.422 0.464 0.551 ...
##
   $ Glycine
                                        : num
##
   $ Histidine
                                               0.023 0.823 0.821 0.306 0.397 ...
                                         : num
                                               0 0 0 0 0 0 0 0 0 0 ...
##
   $ Hydroxyproline
                                         : num
##
   $ Iron, Fe
                                               0.02\ 0.43\ 0.21\ 0.13\ 0.65\ 0.23\ 0.17\ 0.72\ 0.22\ 0.22\ \dots
##
   $ Isoleucine
                                               0.051 1.137 1.451 0.556 0.803 ...
                                        : num
##
   $ Lactose
                                               0 0 0 3.87 0 0 0 0 0 1.12 ...
                                               0.083 2.244 2.238 1.049 1.395 ...
##
   $ Leucine
                                        : num
##
   $ Lutein + zeaxanthin
                                               0 0 0 0 0 0 0 0 0 0 ...
##
                                               0 0 0 0 0 0 0 0 0 0 ...
   $ Lycopene
                                        : num
##
   $ Lysine
                                               0.067 2.124 1.945 0.878 1.219 ...
##
   $ Magnesium, Mg
                                               2 24 21 9 19 14 36 27 23 27 ...
                                        : num
   $ Maltose
                                               0 0 0 0 0 0 0 0 0 0 ...
##
                                        : num
  $ Manganese, Mn
                                               0 0.012 0.012 0.015 0.028 0.014 0.017 0.011 0.01 0.039 ...
##
                                        : num
   $ Menaquinone-4
                                               0 0 0 0 0 0 0 0 0 4.1 ...
##
   $ Methionine
                                         : num 0.021 0.565 0.612 0.253 0.368 0.706 0.822 0.641 0.677 0.551 ...
```

```
## $ Niacin
                                              : num 0.042 0.118 0.08 0.103 0.991 0.15 0.106 0.093 0.105 0.111 ...
## $ Pantothenic acid
                                              : num 0.11 0.288 0.413 0.524 0.967 0.429 0.562 0.21 0.079 0.429 ...
## $ Phenylalanine
                                             : num 0.041 1.231 1.231 0.543 0.675 ...
## $ Phosphorus, P
## $ Potassium, K
## $ Proline
## $ Protein
## $ Retinol
## $ Riboflavin
## $ Selenium, Se
## $ Sodium, Na
## $ Starch
## $ Starch
## $ Sugars, total
## $ Theobromine
## $ Tocopherol, beta
## $ Tocopherol, delta
## $ Tocotrienol, alpha
## $ Tocotrienol, delta
## $ Tocotrienol, gamma
## $ Tocotrienol, gamma
## $ Tocotrienol, gamma
## $ Phosphorus, P
                                             : num 24 451 464 150 337 346 605 444 463 548 ...
                                             : num 0000000000...
                                             : num 24 136 95 125 62 64 81 81 84 188 ...
                                             : num 0.082 2.575 2.634 1.155 1.378 ...
                                             : num 0.85 23.24 23.37 10.45 14.21 ...
                                              : num 671 286 220 68 125 258 268 192 124 223 ...
                                             : num 0.034 0.351 0.293 0.251 0.844 0.204 0.279 0.39 0.303 0.353 ...
                                             : num 1 14.5 14.5 11.9 15 14.5 14.5 14.5 14.4 27.6 ...
                                             : num 0.046 1.289 1.366 0.601 1.169 ...
                                             : num 643 560 700 308 1139 ...
                                             : num 0000000000...
                                             : num 0000000000...
                                             : num 0000000000...
                                             : num 0.06 0.51 0 4 0 1.55 0.36 0.5 1.13 1.9 ...
                                            : num 0000000000...
                                            : num 0.005 0.014 0.046 0.02 0.154 0.021 0.06 0.015 0.018 0.024 ...
                                            : num 0.038 0.882 0.832 0.47 0.637 ...
                                             : num 0000000000...
                                            : num 0000000000...
                                            : num 0 0 0 0 0 0 0 0 0 0.03 ...
                                             : num 0000000000...
                                             : num 0000000000...
                                            : num 0000000000...
                                            : num 0000000000...
## $ Tocotrienol, gamma
## $ Total lipid (fat)
                                             : num 81.11 29.68 30.6 2.27 21.49 ...
## $ Tryptophan
                                             : num 0.012 0.324 0.3 0.138 0.2 0.361 0.421 0.315 0.339 0.551 ...
## $ Tyrosine
                                            : num 0.041 1.115 1.128 0.568 0.668 ...
## $ Valine
                                             : num 0.057 1.472 1.56 0.703 1.065 ...
## $ Vitamin A, IU
                                                     2499 1080 985 236 422 ...
                                             : num
## $ Vitamin A, RAE
                                             : num 684 292 233 68 125 261 271 198 127 227 ...
## $ Vitamin B-12
                                             : num 0.17 1.26 0.83 0.47 1.69 1.68 1.6 0.83 0.82 1.68 ...
## $ Vitamin B-6
                                              : num 0.003 0.065 0.074 0.057 0.424 0.083 0.081 0.079 0.07 0.1 ...
## $ Vitamin C, total ascorbic acid
                                             : num 0000000000...
## $ Vitamin D
                                             : num 0 22 0 0 16 23 24 22 12 15 ...
## $ Vitamin D (D2 + D3)
                                             : num 0 0.5 0 0 0.4 0.6 0.6 0.6 0.3 0.4 ...
## $ Vitamin D2 (ergocalciferol)
                                            : num 0000000000...
## $ Vitamin D3 (cholecalciferol)
                                             : num 0 0.5 0 0 0.4 0.6 0.6 0.6 0.3 0.4 ...
## $ Vitamin E (alpha-tocopherol)
                                            : num 2.32 0.26 0 0.08 0.18 0.27 0.28 0.26 0.14 0.5 ...
                                             : num 7 2.5 0 0 1.8 2.6 2.7 2.5 1.6 1.3 ...
## $ Vitamin K (phylloquinone)
## $ Water
                                              : num 16.2 41.1 37.6 81.2 55.2 ...
## $ Zinc, Zn
                                              : num 0.09 2.6 2.79 0.51 2.88 3.5 3.9 3 2.76 3.62 ...
```

Table 6: Model Performance, Training Set

	$ctree_Training$	gbm_Training	ranger_Training	$ctree_Test$	${\rm gbm_Test}$	${\rm ranger_Test}$
Sensitivity	0.811	0.842	1.000	0.742	0.758	0.770
Specificity	0.968	0.978	1.000	0.947	0.970	0.975
Pos Pred Value	0.848	0.896	1.000	0.758	0.850	0.872
Neg Pred Value	0.958	0.965	1.000	0.943	0.948	0.950
Precision	0.848	0.896	1.000	0.758	0.850	0.872
Recall	0.811	0.842	1.000	0.742	0.758	0.770
F1	0.829	0.868	1.000	0.749	0.801	0.818
Prevalence	0.182	0.182	0.182	0.182	0.182	0.182
Detection Rate	0.147	0.153	0.182	0.135	0.138	0.140
Detection Prevalence	0.174	0.171	0.182	0.178	0.162	0.160
Balanced Accuracy	0.889	0.910	1.000	0.844	0.864	0.872

Table.6 shows the performance of each model, we can see that ranger has the highest accuracy of 1, followed by gbm. a perfect accuracy may be a result of over fitting. applying our models to the test set we can see that accuracy of ranger falls to .87 and gbm falls to .86. which indicates that gbm model is more consistent and not over fitted.

One more thing to notice that if we set RAW as our positive test result we get a Prevalence of 0.182, thus the high Specificity relative to Sensitivity. trying to balance our data set by sampling up and down we get the results in *table.7*.

Table 7: Model Performance, Balanced Data Set

	up	$_{ m gbm}$	down
Sensitivity	0.968	0.842	0.980
Specificity	0.930	0.978	0.901
Pos Pred Value	0.754	0.896	0.687
Neg Pred Value	0.992	0.965	0.995
Precision	0.754	0.896	0.687
Recall	0.968	0.842	0.980
F1	0.847	0.868	0.808
Prevalence	0.182	0.182	0.182
Detection Rate	0.176	0.153	0.178
Detection Prevalence	0.233	0.171	0.259
Balanced Accuracy	0.949	0.910	0.941

as we can see in *Table.8*, sampling up/down increased Balanced Accuracy and increased Sensitivity on the expense of Specificity, F1 and Precision. this means that our model started to predict more raw foods, enough to correctly cover 0.97-0.98 of the RAW population but on the expense of predicting more processed foods as raw foods.

Below shows a breakdown for the performance of gbm, and indeed we see that sampling enhanced Sensitivity, but decreased Specificity. so in conclusion gbm is the more stable model and to chose to sample up, down or not to sample at all depends on the importance of predicting positive(raw) test results more than negative results. but in our case i would prefer a balanced model and will chose gbm without sampling.

Table 8: GBM, Model Performance Per Group

FoodGroup	Sensitivity	Specificity	Precision	Recall	F1	Prevalence	Balanced Accuracy
Dairy and Egg Products	0.71	1.00	1.00	0.71	0.83	0.05	0.86
Spices and Herbs	0.77	NA	1.00	0.77	0.87	1.00	NA
Baby Foods	NA	0.96	0.00	NA	NA	0.00	NA
Fats and Oils	NA	1.00	NA	NA	NA	0.00	NA
Poultry Products	0.92	0.98	0.96	0.92	0.94	0.32	0.95
Soups, Sauces, and Gravies	NA	1.00	NA	NA	NA	0.00	NA
Sausages and Luncheon Meats	NA	0.99	0.00	NA	NA	0.00	NA
Breakfast Cereals	NA	1.00	NA	NA	NA	0.00	NA
Snacks	NA	1.00	NA	NA	NA	0.00	NA
Fruits and Fruit Juices	0.67	0.99	0.98	0.67	0.80	0.34	0.83
Pork Products	0.81	0.99	0.97	0.81	0.89	0.26	0.90
Vegetables and Vegetable Products	0.67	0.93	0.77	0.67	0.72	0.25	0.80
Nut and Seed Products	0.33	1.00	1.00	0.33	0.50	0.14	0.67
Beef Products	0.99	0.94	0.92	0.99	0.95	0.41	0.96
Beverages	NA	1.00	NA	NA	NA	0.00	NA
Finfish and Shellfish Products	0.88	0.88	0.86	0.88	0.87	0.45	0.88
Legumes and Legume Products	0.92	0.99	0.96	0.92	0.94	0.16	0.95
Lamb, Veal, and Game Products	0.98	0.94	0.93	0.98	0.95	0.43	0.96
Baked Products	NA	1.00	NA	NA	NA	0.00	NA
Sweets	NA	1.00	NA	NA	NA	0.00	NA
Cereal Grains and Pasta	0.00	1.00	NA	0.00	NA	0.04	0.50
Fast Foods	NA	1.00	NA	NA	NA	0.00	NA
Meals, Entrees, and Side Dishes	NA	1.00	NA	NA	NA	0.00	NA
American Indian/Alaska Native Foods	0.38	0.98	0.89	0.38	0.53	0.24	0.68
Restaurant Foods	NA	1.00	NA	NA	NA	0.00	NA

Table 9: GBM, Model Performance Per Group, Sampling UP

FoodGroup	Sensitivity	Specificity	Precision	Recall	F1	Prevalence	Balanced Accuracy
Dairy and Egg Products	0.86	1.00	1.00	0.86	0.92	0.05	0.93
Spices and Herbs	0.91	NA	1.00	0.91	0.95	1.00	NA
Baby Foods	NA	0.94	0.00	NA	NA	0.00	NA
Fats and Oils	NA	0.99	0.00	NA	NA	0.00	NA
Poultry Products	1.00	0.91	0.85	1.00	0.92	0.32	0.96
Soups, Sauces, and Gravies	NA	0.99	0.00	NA	NA	0.00	NA
Sausages and Luncheon Meats	NA	0.99	0.00	NA	NA	0.00	NA
Breakfast Cereals	NA	0.99	0.00	NA	NA	0.00	NA
Snacks	NA	1.00	NA	NA	NA	0.00	NA
Fruits and Fruit Juices	0.95	0.91	0.85	0.95	0.90	0.34	0.93
Pork Products	1.00	0.94	0.84	1.00	0.91	0.26	0.97
Vegetables and Vegetable Products	0.97	0.76	0.58	0.97	0.72	0.25	0.87
Nut and Seed Products	1.00	0.91	0.64	1.00	0.78	0.14	0.95
Beef Products	0.99	0.90	0.87	0.99	0.93	0.41	0.94
Beverages	NA	0.99	0.00	NA	NA	0.00	NA
Finfish and Shellfish Products	0.94	0.75	0.76	0.94	0.84	0.45	0.85
Legumes and Legume Products	1.00	0.90	0.67	1.00	0.80	0.16	0.95
Lamb, Veal, and Game Products	0.99	0.85	0.83	0.99	0.90	0.43	0.92
Baked Products	NA	1.00	0.00	NA	NA	0.00	NA
Sweets	NA	1.00	NA	NA	NA	0.00	NA
Cereal Grains and Pasta	0.00	0.96	0.00	0.00	NaN	0.04	0.48
Fast Foods	NA	1.00	NA	NA	NA	0.00	NA
Meals, Entrees, and Side Dishes	NA	1.00	NA	NA	NA	0.00	NA
American Indian/Alaska Native Foods	0.81	0.80	0.57	0.81	0.67	0.24	0.80
Restaurant Foods	NA	1.00	NA	NA	NA	0.00	NA

4 Conclusion

Data exploration shows that there is a clear difference between raw and processed food if we consider food as groups, such as the differences in Fluoride, F, Sodium, Na and Vitamin C values between raw and processed food. Our model shows that we can predict raw/processed food with a Balanced Accuracy of 0.91, Fig. 7 shows the most important differences between raw and processed foods, again we can see that Sodium, Na and Vitamin C are important differences between raw and processed foods.

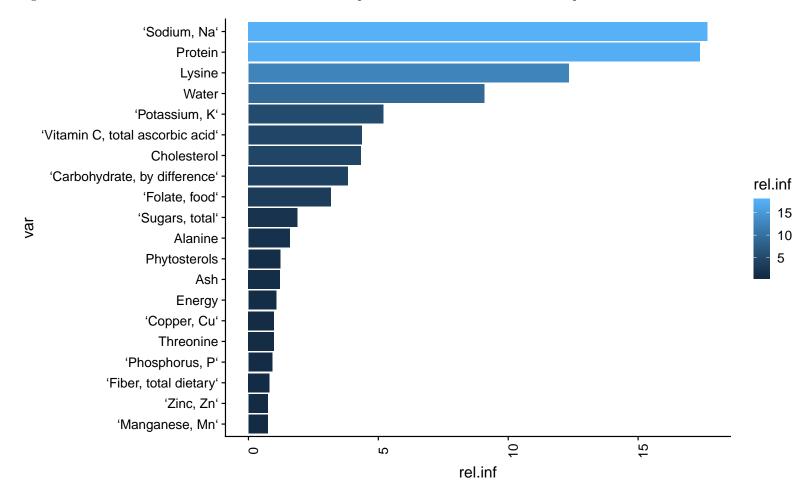


Figure 7: Var Importance