Model_Exploration

April 17, 2021

Objectives

Provide an overview of some machine learning models, using the latest iteration of the input data.

Applied Data Filters

The dataframe included in this analysis contains the following modifications of the original data set:

1. Meal level aggregation
2. Meals that are only lunch or dinner
3. Meals that have both seafood and meat, where there is ambiguity in the ratio, are dropped
4. Meals that are more than 0 KCAL
5. Meals of participants older than 18 years of age
6. Meals that are consumed at home
7. Meals that are non-vegeterian

Section 1: Logistic Regression Model

Model Evaluation

This section evaluated the performance of the logistic regression model over several model fittings. This is done to account for the random sampling of the model data. The performance is measured by a split of the data into training and test sets, using an 80% training and 20% test split. The prediction success rate of the model is used as the evaluation criteria.

[90]:
```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from warnings import simplefilter
from sklearn.exceptions import ConvergenceWarning
simplefilter("ignore", category=ConvergenceWarning)

def nhanes_full_log_reg(df, fped_vars, non_sfd_class_n, sfd_class_n,␣
 ↪test_ratio):

    #Sample the seafood and non-seafood classes, create model input df
    df_non_sfd = df[df['seafood_meal']==0].sample(n=non_sfd_class_n)
    df_sfd = df[df['seafood_meal']==1].sample(n=sfd_class_n)
    df_mdl = pd.concat([df_non_sfd, df_sfd])
    #Add the classification target variable to the df input list
    fped_vars.append('seafood_meal')
    #Use variable combination selected by loop
    df_mdl = df_mdl[fped_vars]
    #Split the training and test data
    X_train, X_test, y_train, y_test = train_test_split(df_mdl.
 ↪drop(['seafood_meal'], axis=1), df_mdl['seafood_meal'], test_size=test_ratio)
    #Fit the logistic regression model
    log_reg = LogisticRegression()
    log_reg.fit(X_train, y_train)
    #Obtain predictions on test set and calculate success rate
    y_pred = log_reg.predict(X_test)
    n_correct = sum(y_pred == y_test)
    pred_sr = [str(n_correct/len(y_pred))]
    #Calculate model execution time for combinatorial variable selection
    #non_cmb_time = time.time() - startTime
    #non_cmb_time_df = pd.DataFrame([non_cmb_time ])
    #non_cmb_time_df = non_cmb_time_df.rename({0: 'Runtime(Seconds)'}, axis=1)
    #Create a dataframe with variables used and their success rate
    pred_sr_df = pd.DataFrame(pred_sr)
    pred_sr_df = pred_sr_df.rename({0: 'Success Rate'}, axis=1)
    fped_vars.remove('seafood_meal')
    var_list_df = pd.DataFrame([fped_vars])
    #model_result = pd.concat([var_list_df, pred_sr_df, non_cmb_time_df],␣
 ↪axis=1)
    model_result = pd.concat([var_list_df, pred_sr_df], axis=1)
    return model_result

#Level 5 has all components of level4, but breaks the total fruit into␣
 ↪subcomponents
food_cmp_level5 = ['F_CITMLB', 'F_OTHER', 'F_JUICE',
                   'V_DRKGR', 'V_REDOR_TOMATO', 'V_REDOR_OTHER',␣
 ↪'V_STARCHY_POTATO',
                   'V_STARCHY_OTHER', 'V_OTHER', 'V_LEGUMES',
                   'G_WHOLE','G_REFINED',
                   'PF_EGGS', 'PF_SOY', 'PF_NUTSDS', 'PF_LEGUMES',
                   'D_MILK', 'D_YOGURT', 'D_CHEESE',
                   'OILS', 'SOLID_FATS', 'ADD_SUGARS', 'A_DRINKS']

df = pd.read_csv('../../Data/nhanes_full_pre_proc.csv')
#df = df[df['meal_energy']=='Medium-Low']

sr_tot=[]
for i in range(100):
    model_res_df_x = nhanes_full_log_reg(df = df,
                                   fped_vars = food_cmp_level5,
                                   non_sfd_class_n = 500,
                                   sfd_class_n = 500,
                                             test_ratio = 0.2, )
    sr_tot.append(model_res_df_x['Success Rate'][0])


sr_tot = pd.DataFrame(sr_tot)
sr_tot = sr_tot[0].astype(float)
print(sr_tot.describe())
```

```
count    100.000000
mean       0.668950
std        0.034497
min        0.590000
25%        0.650000
50%        0.670000
75%        0.690000
max        0.755000
Name: 0, dtype: float64
```

The table above is showing the prediction success rate results of the model over 100 runs.

Model Interpretation

This section performs a single fit of the logistic regression model and obtains the regression coefficients. Using the coefficients derived from the fitted model, the probability of a meal containing seafood or not can be obtained by the following equation:

$$p(X) = \frac{e^{\beta_0+\beta_1 X_1+...+\beta_p X_p}}{1+e^{\beta_0+\beta_1 X_1+...+\beta_p X_p}}$$

Where

$$\beta_0$$

is the intercept. When trying to interpret the impact of a single component, we can see that the probability depends on the initial values of the other components. Therefore, the only valuable information from the coefficient values themselves is the sign. If the sign is positive, then we know that increasing the value of the component will increase the probability while the inverse is true of a negative sign.

Below is an example for the model interpretation.

```
[51]: def nhanes_full_log_reg_params(df, fped_vars, non_sfd_class_n, sfd_class_n,␣
       ↪test_ratio):

          #Sample the seafood and non-seafood classes, create model input df
          df_non_sfd = df[df['seafood_meal']==0].sample(n=non_sfd_class_n,␣
       ↪random_state=1)
          df_sfd = df[df['seafood_meal']==1].sample(n=sfd_class_n, random_state=1)
          df_mdl = pd.concat([df_non_sfd, df_sfd])
          #Add the classification target variable to the df input list
          fped_vars.append('seafood_meal')
```

```
    #Use variable combination selected by loop
    df_mdl = df_mdl[fped_vars]
    #Split the training and test data
    X_train, X_test, y_train, y_test = train_test_split(df_mdl.
 ↪drop(['seafood_meal'], axis=1), df_mdl['seafood_meal'], test_size=test_ratio)
    #Fit the logistic regression model
    log_reg = LogisticRegression()
    log_reg.fit(X_train, y_train)
    #Obtain predictions on test set and calculate success rate
    y_pred = log_reg.predict(X_test)
    n_correct = sum(y_pred == y_test)
    pred_sr = [str(n_correct/len(y_pred))]
    pred_sr_df = pd.DataFrame(pred_sr)
    pred_sr_df = pred_sr_df.rename({0: 'Success Rate'}, axis=1)
    fped_vars.remove('seafood_meal')
    var_list_df = pd.DataFrame([fped_vars])
    #model_result = pd.concat([var_list_df, pred_sr_df, non_cmb_time_df],␣
 ↪axis=1)
    model_result = pd.concat([var_list_df, pred_sr_df], axis=1)
    return log_reg.coef_, log_reg.intercept_


food_cmp_level5 = ['F_CITMLB', 'F_OTHER', 'F_JUICE',
                   'V_DRKGR', 'V_REDOR_TOMATO', 'V_REDOR_OTHER',␣
 ↪'V_STARCHY_POTATO',
                   'V_STARCHY_OTHER', 'V_OTHER', 'V_LEGUMES',
                   'G_WHOLE','G_REFINED',
                   'PF_EGGS', 'PF_SOY', 'PF_NUTSDS', 'PF_LEGUMES',
                   'D_MILK', 'D_YOGURT', 'D_CHEESE',
                   'OILS', 'SOLID_FATS', 'ADD_SUGARS', 'A_DRINKS']

df = pd.read_csv('../../Data/nhanes_full_pre_proc.csv')

log_reg_coefficients, log_reg_intercept = nhanes_full_log_reg_params(df = df,
                                    fped_vars = food_cmp_level5,
                                    non_sfd_class_n = 500,
                                    sfd_class_n = 500,
                                    test_ratio = 0.2)


print("The logistic regression coefficients are: \n", log_reg_coefficients)
print("\n")
print("The logistic regression intercept is: \n", log_reg_intercept)
```

```
The logistic regression coefficients are:
 [[ 0.428869    0.05885685  0.45412882  0.44847553 -0.51396135  0.30002303
  -0.50495974  0.60357087  0.49907006 -0.06117684  0.13950824  0.01510887
```

```
   0.62869596  0.20375317 -0.24770104 -0.21335972  0.1286632   0.02856547
  -0.81856701  0.03761363 -0.02244882 -0.03039313 -0.12752813]]


The logistic regression intercept is:
 [-0.11213369]
```

The vectors printed above represent the coefficients from the fitted logistic regression model, using the level 5 FPED components:

```
[52]: print("The logistic regression variables are: \n", food_cmp_level5)
      print("\n")
```

```
The logistic regression variables are:
 ['F_CITMLB', 'F_OTHER', 'F_JUICE', 'V_DRKGR', 'V_REDOR_TOMATO',
'V_REDOR_OTHER', 'V_STARCHY_POTATO', 'V_STARCHY_OTHER', 'V_OTHER', 'V_LEGUMES',
'G_WHOLE', 'G_REFINED', 'PF_EGGS', 'PF_SOY', 'PF_NUTSDS', 'PF_LEGUMES',
'D_MILK', 'D_YOGURT', 'D_CHEESE', 'OILS', 'SOLID_FATS', 'ADD_SUGARS',
'A_DRINKS']
```

However, due to the high number of components/parameters that are used in the model, even the signs of the parameters are random in nature. This is because the model fit convergence can arrive at different solutions. The model fitting algorithms use a random point for initializing the algorithm. And with the model of this size, this will affect the parameter sign, depending on when the algorithm arrives at a solution. For example, running the model again will yield the following solution, where we can see that some of the parameter signs are different:

```
[53]: log_reg_coefficients, log_reg_intercept = nhanes_full_log_reg_params(df = df,
                                   fped_vars = food_cmp_level5,
                                   non_sfd_class_n = 500,
                                   sfd_class_n = 500,
                                   test_ratio = 0.2)



      print("The logistic regression coefficients are: \n", log_reg_coefficients)
      print("\n")
      print("The logistic regression intercept is: \n", log_reg_intercept)
```

```
The logistic regression coefficients are:
 [[-0.06971927 -0.13028231  0.52887761  0.42064084 -0.55416455 -0.0284824
  -0.49418688  0.57011787  0.51343868 -0.08268138  0.21338709  0.02828873
   0.81689587  0.19440342 -0.22312681 -0.30616341  0.38247213  0.06824989
  -1.05100914  0.0363042  -0.02564836 -0.03278337  0.0121854 ]]


The logistic regression intercept is:
 [-0.08784222]
```

Therefore, it is not advisable to draw inferences from the parameter values of the model. This may also be due to the low performance of the model, i.e. the model is not very representative of the real world. This may lead the model to arrive at different solutions, especially for components that do not provide a clear contribution. Below is an example of how to interpret the model derived above. First, we can obtain a meal at random from the dataframe:

[56]:
```
random_meal = df.sample(n=1, random_state=1)
random_meal[food_cmp_level5]
```

[56]:
```
      F_CITMLB  F_OTHER  F_JUICE  V_DRKGR  V_REDOR_TOMATO  V_REDOR_OTHER  \
4939       0.0      0.0      3.6     0.03             0.2            0.0

      V_STARCHY_POTATO  V_STARCHY_OTHER  V_OTHER  V_LEGUMES  …  PF_SOY  \
4939               0.0              0.0     0.49        0.0  …     0.0

      PF_NUTSDS  PF_LEGUMES  D_MILK  D_YOGURT  D_CHEESE  OILS  SOLID_FATS  \
4939        0.0         0.0     0.0       0.0       0.0  1.96        0.09

      ADD_SUGARS  A_DRINKS
4939         0.0       0.0

[1 rows x 23 columns]
```

The component values for the randomly selected meal are displayed above. We can plug these in to the logistic probability equation from above, and can derive the probability of the meal containing seafood using the fitted model parameters.

[83]:
```
from math import exp
X = np.array(random_meal[food_cmp_level5])
Beta = np.array(log_reg_coefficients).T
X_Beta = np.dot(X, Beta)
b0 = log_reg_intercept
p = (exp(b0 + X_Beta))/(1+exp(b0 + X_Beta))
print("The probability of this meal containing seafood is: ", p)
```

```
The probability of this meal containing seafood is:  0.8924882523661772
```

From this initial point, we can then calculate how this probability changes if we change one component value while leaving the others constant. For example, we can try to decrease the F_JUICE component by 2, to 1.6 from 3.6:

[85]:
```
X[0,2] = 1.6
print("The new X vector is: ", X)
X_Beta = np.dot(X, Beta)
b0 = log_reg_intercept
p = (exp(b0 + X_Beta))/(1+exp(b0 + X_Beta))
print("\nThe probability of this meal containing seafood is: ", p)
```

```
The new X vector is:  [[0.   0.   1.6  0.03 0.2  0.   0.   0.   0.49 0.   0.
2.76 0.   0.
  0.   0.   0.   0.   0.   1.96 0.09 0.   0.  ]]

The probability of this meal containing seafood is:  0.7424340046290475
```

As was stated above, we can see in which direction the probability will change based on the sign for the component parameter. However, the actual change in probability depends on the values of the other parameters. This is due to the non-linear nature of the exponential function that the logistic regression uses to derive probabilities. Finally, did the model predict this meal correctly?

[87]:
```
random_meal['species']
```

[87]:
```
4939    shrimp
Name: species, dtype: object
```

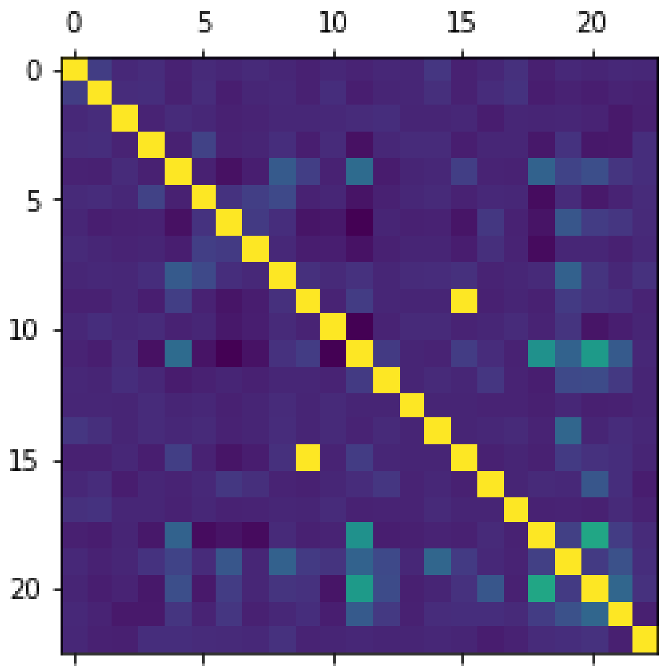
This meal did indeed contain shrimp, so the model was able to classify this meal correctly.

Model Diagnostics

The logistic regression success rate is a bit low, at around a 0.66 prediction success rate. So this model does not seem to be very representative of the real world behavior of food consumers. This may very well be due to the statistical distribution characteristics of the FPED components, which was explored in other reports. But we also know that regression models can be negatively affected by a dataset with a high number of variables, which is the case for this NHANES observations space. The model above was fitted using 23 variables, and the model solution will have a high variance if there is correlation within those components. Below is the correlation matrix for these components:

[96]:
```
import matplotlib.pyplot as plt

df = pd.read_csv('../../Data/nhanes_full_pre_proc.csv')
def histogram_intersection(a, b):
    v = np.minimum(a, b).sum().round(decimals=1)
    return v
df = df[food_cmp_level5]
corr = df.corr()
#plot(corr.style.background_gradient(cmap='coolwarm'))
plt.matshow(corr)
plt.show()
```



The correlation matrix does not show any high correlations among the components. The correlation values are actually quite low. Just as an extra cautionary step, the section below attempts to use a PCA transformation of the observation space and use the principal components to re-fit the model.

[91]:
```
import numpy as np
from sklearn.decomposition import PCA

df = pd.read_csv('../../Data/nhanes_full_pre_proc.csv')

X = df[food_cmp_level5]
pca = PCA(n_components=5)
pca_x = pca.fit(X)
df_pca = pca_x.transform(X)
df_new = pd.DataFrame(df_pca)


df_new['seafood_meal'] = df['seafood_meal']

var_sel = [0,1,2,3,4]

sr_tot=[]
for i in range(100):
    model_res_df_x = nhanes_full_log_reg(df = df_new,
                                fped_vars = var_sel,
                                non_sfd_class_n = 1000,
                                   sfd_class_n = 1000,
                                   test_ratio = 0.2)
    sr_tot.append(model_res_df_x['Success Rate'][0])


sr_tot = pd.DataFrame(sr_tot)
sr_tot = sr_tot[0].astype(float)
print(sr_tot.describe())
print(pca_x.explained_variance_ratio_)
```

```
count    100.000000
mean       0.648350
std        0.024163
min        0.592500
25%        0.634375
50%        0.648750
75%        0.662500
max        0.700000
Name: 0, dtype: float64
[0.57740584 0.32557929 0.0784561  0.01022015 0.00256172]
```

The model above is fitted 100 different times, to account for the variances in the sampling of the data. The results above are showing the success rate distribution of these 100 runs. Five principal components, which explain most of the variation within the data, were used to re-fit the model. This model is actually performing worse than the original, re-iterating that correlation within the components is not an issue for this dataset.

Section 2: Component-Wise Linear Regression

This section explores a linear regression model, targetting each one of the components while using the others. In this manner we can see how changing the amount of seafood on a plate will affect each of the components.

[123]:
```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
from sklearn.metrics import r2_score
import numpy as np

def linear_regressor(df, obs_space, target, n_sfd, n_non_sfd):
    df_non_sfd = df[df['seafood_meal']==0].sample(n=n_non_sfd)
    df_sfd = df[df['seafood_meal']==1].sample(n=n_sfd)
    df = pd.concat([df_non_sfd, df_sfd])
    df_x = df[obs_space]
    df_y = df[target]
    lin_reg = LinearRegression()
    lin_reg.fit(df_x, df_y)
    y_pred = lin_reg.predict(df_x)
    lin_mse = mean_squared_error(df_y, y_pred)
    lin_mse = np.sqrt(lin_mse)
    lin_r2 = r2_score(df_y, y_pred)
    return target, lin_r2, lin_mse, df[target].describe()['mean'], df[target].
 ↪describe()['std']


food_cmp_level6 = ['F_CITMLB', 'F_OTHER', 'F_JUICE',
                   'V_DRKGR', 'V_REDOR_TOMATO', 'V_REDOR_OTHER',␣
 ↪'V_STARCHY_POTATO',
                   'V_STARCHY_OTHER', 'V_OTHER',
                   'G_WHOLE','G_REFINED',
                   'PF_EGGS', 'PF_SOY', 'PF_NUTSDS', 'PF_LEGUMES',
                   'PF_CUREDMEAT', 'PF_ORGAN', 'PF_POULT', 'PF_MEAT',␣
 ↪'PF_SEAFD_TOT',
                   'D_MILK', 'D_YOGURT', 'D_CHEESE',
                   'OILS', 'SOLID_FATS', 'ADD_SUGARS', 'A_DRINKS']

#Read the pre-processed dataframe.
df = pd.read_csv('../../Data/nhanes_full_pre_proc.csv')

result = []
for var in food_cmp_level6:
    obs_space = list(food_cmp_level6)
    obs_space.remove(var)
    #print(variables)
    target = var
    result_temp = linear_regressor(df, obs_space, var, 500, 500)
    result.append(result_temp)

result = pd.DataFrame(result)
result = result.rename(columns={0: "Regression Target", 1: "R^2", 2: "RMSE", 3:␣
 ↪"Target Mean", 4:"Target SD"})
result = result.sort_values(by="R^2", ascending = False)
print(result)
```

```
   Regression Target       R^2       RMSE  Target Mean  Target SD
24        SOLID_FATS  0.488401  10.013990     10.55589  14.007465
22          D_CHEESE  0.468449   0.437460      0.21444   0.600320
23              OILS  0.325445   9.751514      9.85552  11.879022
10         G_REFINED  0.323397   1.823277      2.09176   2.217700
19      PF_SEAFD_TOT  0.288836   2.525525      1.83201   2.996290
18           PF_MEAT  0.208648   1.481895      0.70173   1.666671
6   V_STARCHY_POTATO  0.192106   0.321119      0.14430   0.357443
4     V_REDOR_TOMATO  0.171867   0.256515      0.14051   0.282021
13         PF_NUTSDS  0.154644   0.391540      0.05461   0.426062
17          PF_POULT  0.153676   1.587469      0.68087   1.726451
25        ADD_SUGARS  0.153480   5.298738      3.61063   5.761971
8            V_OTHER  0.141773   0.462492      0.33273   0.499483
```

```
15       PF_CUREDMEAT  0.139455   0.816721      0.28877   0.880855
11            PF_EGGS  0.116634   0.288373      0.10681   0.306974
9             G_WHOLE  0.076042   0.546682      0.18344   0.569018
20             D_MILK  0.071919   0.401900      0.14483   0.417390
5       V_REDOR_OTHER  0.071725   0.134720      0.04519   0.139898
3             V_DRKGR  0.053090   0.285205      0.09484   0.293238
14        PF_LEGUMES  0.048688   0.815307      0.20966   0.836328
26           A_DRINKS  0.038677   0.561471      0.12194   0.572941
7     V_STARCHY_OTHER  0.037122   0.224318      0.06454   0.228715
1             F_OTHER  0.034316   0.316904      0.07205   0.322647
12             PF_SOY  0.019993   0.162955      0.01793   0.164691
0            F_CITMLB  0.018741   0.240987      0.04065   0.243400
2             F_JUICE  0.015150   0.378592      0.07032   0.381684
16           PF_ORGAN  0.013154   0.233679      0.01647   0.235349
21           D_YOGURT  0.010256   0.038004      0.00292   0.038220
```

The table above is showing the results of the linear regression model, targeting each component. The performance metric is the the standard model Root Mean Square Error, or RMSE. Given that each component has different units and distribution characteristics, the regression target component statistics are disaplyed above. The table is sorted by the R-squared result of the model. This provides an indication of how well the model predicts the quantities of the target component. As we can see, the better performing models are when regressing on SOLID_FATS and D_CHEESE. However, even for these models, the R-squared result is not very high.

Model Interpretation

The linear regression model is represented by the following equation:

$$\hat{y} = \beta_0 + \beta_1 X_1 + ... + \beta_p X_p$$

The interpretation here is that changing the value of a component by some unit, while holding all the other parameters constant, will change the prediction value by:

$$\delta\hat{y} = \beta_n \delta X_n$$

Since the linear regression on solid fats performed the best, we can use this as an example.

[124]:
```python
def linear_regressor_params(df, obs_space, target, n_sfd, n_non_sfd):
    df_non_sfd = df[df['seafood_meal']==0].sample(n=n_non_sfd)
    df_sfd = df[df['seafood_meal']==1].sample(n=n_sfd)
    df = pd.concat([df_non_sfd, df_sfd])
    df_x = df[obs_space]
    df_y = df[target]
    lin_reg = LinearRegression()
    lin_reg.fit(df_x, df_y)
    y_pred = lin_reg.predict(df_x)
    lin_mse = mean_squared_error(df_y, y_pred)
    lin_mse = np.sqrt(lin_mse)
    return lin_reg.coef_, lin_reg.intercept_


obs_space = list(food_cmp_level6)
obs_space.remove('SOLID_FATS')
lin_reg_coefficients, lin_reg_intercept = linear_regressor_params(df,␣
 ↪obs_space, 'V_LEGUMES', 500, 500)

print("The linear regression predictor variables are: \n", obs_space)
print("\n")
print("The linear regression coefficients on SOLID_FATS are: \n",␣
 ↪lin_reg_coefficients)
print("\n")
print("The linear regression intercept on SOLID_FATS is: \n", lin_reg_intercept)
print("\n")
print("The linear regression coefficient for PF_SEAFD_TOT is: \n",␣
 ↪lin_reg_coefficients[19])
print("\n")
print("So increasing the PF_SEAFD_TOT value by 1 unit, will change the amount␣
 ↪of solid fats by: ", lin_reg_coefficients[19])
print("\n")
print("And decreasing the PF_SEAFD_TOT value by 1 unit, will change the amount␣
 ↪of solid fats by ", -lin_reg_coefficients[19])
```

```
The linear regression predictor variables are:
 ['F_CITMLB', 'F_OTHER', 'F_JUICE', 'V_DRKGR', 'V_REDOR_TOMATO',
'V_REDOR_OTHER', 'V_STARCHY_POTATO', 'V_STARCHY_OTHER', 'V_OTHER', 'G_WHOLE',
'G_REFINED', 'PF_EGGS', 'PF_SOY', 'PF_NUTSDS', 'PF_LEGUMES', 'PF_CUREDMEAT',
'PF_ORGAN', 'PF_POULT', 'PF_MEAT', 'PF_SEAFD_TOT', 'D_MILK', 'D_YOGURT',
'D_CHEESE', 'OILS', 'ADD_SUGARS', 'A_DRINKS']


The linear regression coefficients on SOLID_FATS are:
 [ 9.36136806e-05 -3.07388895e-04  5.23706205e-04  5.68146634e-05
  1.42595451e-04 -5.59135261e-04  6.08344288e-05 -2.41502561e-04
 -2.14181140e-04  1.63839757e-05  9.71694764e-05 -6.33311412e-04
  3.22683943e-04  1.83353777e-04  2.48870266e-01 -5.36502668e-05
 -1.32913699e-05  5.42045282e-05  9.21734886e-06 -2.70821736e-06
  2.38046579e-04  1.12222444e-03 -1.67284936e-04  1.71555880e-06
 -1.61058470e-05  1.46566527e-04]


The linear regression intercept on SOLID_FATS is:
 -9.028702816244599e-05


The linear regression coefficient for PF_SEAFD_TOT is:
```

```
-2.7082173596260176e-06
```

```
So increasing the PF_SEAFD_TOT value by 1 unit, will change the amount of solid
fats by:  -2.7082173596260176e-06
```

```
And decreasing the PF_SEAFD_TOT value by 1 unit, will change the amount of solid
fats by  2.7082173596260176e-06
```

So even in this model, the coefficient for PF_SEAFD_TOT is too small to be used reliably for making inferences on how solid fats are consumed among seafood and non seafood meals. In addition to this, the poor performance on the linear regression when targeting the other components, indicates that a linear regression does not best represent the behavior in the real world. This model is most likely experiencing similar issues as the logistic regression model, where the distribution characterisitcs of the input data are not suited for regression.

Section 3: Decision Tree Model

Decision trees are an alternative to the regression approach, while still providing a certain level of explainability of the data features. Given the statistical characteristics of the features in this dataset, a decision tree may provide a more suitable model of the real world behavior.

[113]:
```
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score
from sklearn.tree import export_text

def tree_regressor(df, components, n_sfd, n_non_sfd):
    df_non_sfd = df[df['seafood_meal']==0].sample(n=n_non_sfd)
    df_sfd = df[df['seafood_meal']==1].sample(n=n_sfd)
    df = pd.concat([df_non_sfd, df_sfd])
    df_x = df[components]
    df_y = df['seafood_meal']
    X_train, X_test, y_train, y_test = train_test_split(df_x, df_y, test_size=0.
 ↪2)
    decision_tree = DecisionTreeClassifier()
    decision_tree.fit(X_train, y_train)
    y_pred_tree = decision_tree.predict(X_test)
    #score = decision_tree.score(X_test, y_test)
    score = accuracy_score(y_test, y_pred_tree)
    tree_rules = export_text(decision_tree, feature_names=list(X_train.columns))
    return score, tree_rules


df = pd.read_csv('../../Data/nhanes_full_pre_proc.csv')

food_cmp_level5 = ['F_CITMLB', 'F_OTHER', 'F_JUICE',
                   'V_DRKGR', 'V_REDOR_TOMATO', 'V_REDOR_OTHER',␣
 ↪'V_STARCHY_POTATO',
                    'V_STARCHY_OTHER', 'V_OTHER', 'V_LEGUMES',
                    'G_WHOLE','G_REFINED',
                    'PF_EGGS', 'PF_SOY', 'PF_NUTSDS', 'PF_LEGUMES',
                    'D_MILK', 'D_YOGURT', 'D_CHEESE',
                    'OILS', 'SOLID_FATS', 'ADD_SUGARS', 'A_DRINKS']


result = []
for i in range(100):
    model_fit, tree_rules = tree_regressor(df, food_cmp_level5, 1000, 1000)
    result.append(model_fit)

result = pd.DataFrame(result)
print(result.describe())
```

```
                0
count  100.000000
mean     0.645750
std      0.023381
min      0.595000
25%      0.630000
50%      0.643750
75%      0.662500
max      0.705000
```

The prediction success rate is not much higher for this model.

Model Interpretation

The resulting decision tree is depicted below. The tree contains a high amount of branches, due to the high amount of components.

```
[114]: print(tree_rules)
```

```
|--- SOLID_FATS <= 2.86
|   |--- OILS <= 2.75
|   |   |--- G_REFINED <= 0.15
|   |   |   |--- V_REDOR_OTHER <= 0.02
|   |   |   |   |--- SOLID_FATS <= 2.18
|   |   |   |   |   |--- V_DRKGR <= 0.00
|   |   |   |   |   |   |--- V_REDOR_TOMATO <= 0.22
|   |   |   |   |   |   |   |--- V_OTHER <= 1.00
|   |   |   |   |   |   |   |   |--- V_REDOR_TOMATO <= 0.00
|   |   |   |   |   |   |   |   |   |--- D_MILK <= 0.04
|   |   |   |   |   |   |   |   |   |   |--- V_STARCHY_POTATO <= 0.85
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 12
|   |   |   |   |   |   |   |   |   |   |--- V_STARCHY_POTATO >  0.85
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- D_MILK >  0.04
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- V_REDOR_TOMATO >  0.00
|   |   |   |   |   |   |   |   |   |--- V_OTHER <= 0.01
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- V_OTHER >  0.01
|   |   |   |   |   |   |   |   |   |   |--- SOLID_FATS <= 0.82
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |   |--- SOLID_FATS >  0.82
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- V_OTHER >  1.00
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- V_REDOR_TOMATO >  0.22
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- V_DRKGR >  0.00
|   |   |   |   |   |   |--- class: 1
|   |   |   |   |--- SOLID_FATS >  2.18
|   |   |   |   |   |--- OILS <= 1.53
|   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- OILS >  1.53
|   |   |   |   |   |   |--- class: 1
|   |   |   |--- V_REDOR_OTHER >  0.02
|   |   |   |   |--- V_STARCHY_OTHER <= 0.13
|   |   |   |   |   |--- V_REDOR_OTHER <= 0.69
|   |   |   |   |   |   |--- V_OTHER <= 0.05
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- V_OTHER >  0.05
|   |   |   |   |   |   |   |--- V_OTHER <= 0.55
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- V_OTHER >  0.55
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- V_REDOR_OTHER >  0.69
|   |   |   |   |   |   |--- class: 1
|   |   |   |   |--- V_STARCHY_OTHER >  0.13
|   |   |   |   |   |--- F_OTHER <= 0.34
|   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- F_OTHER >  0.34
|   |   |   |   |   |   |--- class: 0
|   |   |--- G_REFINED >  0.15
|   |   |   |--- V_REDOR_TOMATO <= 0.15
|   |   |   |   |--- G_REFINED <= 0.49
|   |   |   |   |   |--- F_OTHER <= 0.45
|   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- F_OTHER >  0.45
|   |   |   |   |   |   |--- class: 1
|   |   |   |   |--- G_REFINED >  0.49
|   |   |   |   |   |--- ADD_SUGARS <= 0.25
|   |   |   |   |   |   |--- V_OTHER <= 0.09
|   |   |   |   |   |   |   |--- V_STARCHY_OTHER <= 0.03
|   |   |   |   |   |   |   |   |--- F_JUICE <= 0.81
|   |   |   |   |   |   |   |   |   |--- OILS <= 2.35
|   |   |   |   |   |   |   |   |   |   |--- V_OTHER <= 0.06
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 4
|   |   |   |   |   |   |   |   |   |   |--- V_OTHER >  0.06
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- OILS >  2.35
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- F_JUICE >  0.81
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- V_STARCHY_OTHER >  0.03
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- V_OTHER >  0.09
|   |   |   |   |   |   |   |--- V_DRKGR <= 0.69
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- V_DRKGR >  0.69
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- ADD_SUGARS >  0.25
|   |   |   |   |   |   |--- G_REFINED <= 0.72
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- G_REFINED >  0.72
|   |   |   |   |   |   |   |--- G_REFINED <= 3.32
|   |   |   |   |   |   |   |   |--- OILS <= 0.49
|   |   |   |   |   |   |   |   |   |--- V_OTHER <= 0.27
|   |   |   |   |   |   |   |   |   |   |--- G_REFINED <= 1.02
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |   |--- G_REFINED >  1.02
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 3
|   |   |   |   |   |   |   |   |   |--- V_OTHER >  0.27
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- OILS >  0.49
|   |   |   |   |   |   |   |   |   |--- V_OTHER <= 0.28
|   |   |   |   |   |   |   |   |   |   |--- F_JUICE <= 0.01
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 3
|   |   |   |   |   |   |   |   |   |   |--- F_JUICE >  0.01
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 3
|   |   |   |   |   |   |   |   |   |--- V_OTHER >  0.28
|   |   |   |   |   |   |   |   |   |   |--- G_REFINED <= 2.12
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 2
|   |   |   |   |   |   |   |   |   |   |--- G_REFINED >  2.12
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- G_REFINED >  3.32
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |--- V_REDOR_TOMATO >  0.15
|   |   |   |   |--- OILS <= 2.34
|   |   |   |   |   |--- G_REFINED <= 0.35
|   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- G_REFINED >  0.35
|   |   |   |   |   |   |--- class: 0
|   |   |   |   |--- OILS >  2.34
|   |   |   |   |   |--- class: 1
|   |--- OILS >  2.75
|   |   |--- PF_LEGUMES <= 1.01
|   |   |   |--- ADD_SUGARS <= 2.97
|   |   |   |   |--- V_STARCHY_POTATO <= 0.84
|   |   |   |   |   |--- D_CHEESE <= 0.05
|   |   |   |   |   |   |--- G_REFINED <= 1.69
|   |   |   |   |   |   |   |--- V_OTHER <= 2.64
|   |   |   |   |   |   |   |   |--- V_DRKGR <= 1.83
|   |   |   |   |   |   |   |   |   |--- V_REDOR_TOMATO <= 0.11
|   |   |   |   |   |   |   |   |   |   |--- OILS <= 33.44
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 6
|   |   |   |   |   |   |   |   |   |   |--- OILS >  33.44
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 2
|   |   |   |   |   |   |   |   |   |--- V_REDOR_TOMATO >  0.11
|   |   |   |   |   |   |   |   |   |   |--- V_REDOR_TOMATO <= 0.14
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |   |--- V_REDOR_TOMATO >  0.14
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 7
|   |   |   |   |   |   |   |   |--- V_DRKGR >  1.83
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- V_OTHER >  2.64
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- G_REFINED >  1.69
|   |   |   |   |   |   |   |--- SOLID_FATS <= 2.73
|   |   |   |   |   |   |   |   |--- D_MILK <= 0.93
|   |   |   |   |   |   |   |   |   |--- ADD_SUGARS <= 1.56
|   |   |   |   |   |   |   |   |   |   |--- ADD_SUGARS <= 1.32
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 12
|   |   |   |   |   |   |   |   |   |   |--- ADD_SUGARS >  1.32
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 3
|   |   |   |   |   |   |   |   |   |--- ADD_SUGARS >  1.56
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |--- D_MILK >  0.93
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- SOLID_FATS >  2.73
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- D_CHEESE >  0.05
|   |   |   |   |   |   |--- V_REDOR_TOMATO <= 0.01
|   |   |   |   |   |   |   |--- OILS <= 6.48
|   |   |   |   |   |   |   |   |--- G_WHOLE <= 0.14
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- G_WHOLE >  0.14
|   |   |   |   |   |   |   |   |   |--- V_OTHER <= 0.06
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- V_OTHER >  0.06
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- OILS >  6.48
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- V_REDOR_TOMATO >  0.01
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |--- V_STARCHY_POTATO >  0.84
|   |   |   |   |   |--- SOLID_FATS <= 0.77
|   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- SOLID_FATS >  0.77
|   |   |   |   |   |   |--- class: 1
|   |   |   |--- ADD_SUGARS >  2.97
|   |   |   |   |--- G_REFINED <= 2.02
|   |   |   |   |   |--- PF_EGGS <= 0.39
|   |   |   |   |   |   |--- V_REDOR_OTHER <= 0.04
|   |   |   |   |   |   |   |--- SOLID_FATS <= 0.35
|   |   |   |   |   |   |   |   |--- OILS <= 13.16
|   |   |   |   |   |   |   |   |   |--- SOLID_FATS <= 0.02
|   |   |   |   |   |   |   |   |   |   |--- V_REDOR_TOMATO <= 0.04
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 2
|   |   |   |   |   |   |   |   |   |   |--- V_REDOR_TOMATO >  0.04
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- SOLID_FATS >  0.02
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- OILS >  13.16
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- SOLID_FATS >  0.35
|   |   |   |   |   |   |   |   |--- V_STARCHY_POTATO <= 0.25
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |--- V_STARCHY_POTATO >  0.25
|   |   |   |   |   |   |   |   |   |--- PF_EGGS <= 0.01
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- PF_EGGS >  0.01
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- V_REDOR_OTHER >  0.04
|   |   |   |   |   |   |   |--- PF_EGGS <= 0.04
|   |   |   |   |   |   |   |   |--- OILS <= 2.87
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |--- OILS >  2.87
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- PF_EGGS >  0.04
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- PF_EGGS >  0.39
|   |   |   |   |   |   |--- class: 0
|   |   |   |   |--- G_REFINED >  2.02
|   |   |   |   |   |--- V_OTHER <= 0.30
|   |   |   |   |   |   |--- F_JUICE <= 0.01
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- F_JUICE >  0.01
|   |   |   |   |   |   |   |--- G_REFINED <= 2.72
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- G_REFINED >  2.72
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- V_OTHER >  0.30
|   |   |   |   |   |   |--- SOLID_FATS <= 2.21
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- SOLID_FATS >  2.21
|   |   |   |   |   |   |   |--- class: 0
|   |   |--- PF_LEGUMES >  1.01
|   |   |   |--- OILS <= 8.66
|   |   |   |   |--- class: 1
|   |   |   |--- OILS >  8.66
|   |   |   |   |--- V_REDOR_TOMATO <= 0.31
|   |   |   |   |   |--- D_YOGURT <= 0.25
|   |   |   |   |   |   |--- PF_LEGUMES <= 4.47
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- PF_LEGUMES >  4.47
|   |   |   |   |   |   |   |--- G_REFINED <= 2.33
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- G_REFINED >  2.33
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- D_YOGURT >  0.25
|   |   |   |   |   |   |--- class: 1
|   |   |   |   |--- V_REDOR_TOMATO >  0.31
|   |   |   |   |   |--- SOLID_FATS <= 1.66
|   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- SOLID_FATS >  1.66
|   |   |   |   |   |   |--- class: 0
|--- SOLID_FATS >  2.86
|   |--- PF_EGGS <= 0.03
|   |   |--- D_CHEESE <= 0.19
|   |   |   |--- OILS <= 1.44
|   |   |   |   |--- V_STARCHY_OTHER <= 0.62
|   |   |   |   |   |--- PF_SOY <= 0.35
|   |   |   |   |   |   |--- V_REDOR_TOMATO <= 0.21
|   |   |   |   |   |   |   |--- PF_EGGS <= 0.00
|   |   |   |   |   |   |   |   |--- V_STARCHY_POTATO <= 1.78
|   |   |   |   |   |   |   |   |   |--- V_REDOR_OTHER <= 0.48
|   |   |   |   |   |   |   |   |   |   |--- D_MILK <= 1.79
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 3
|   |   |   |   |   |   |   |   |   |   |--- D_MILK >  1.79
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 2
|   |   |   |   |   |   |   |   |   |--- V_REDOR_OTHER >  0.48
|   |   |   |   |   |   |   |   |   |   |--- OILS <= 0.37
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |   |--- OILS >  0.37
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |--- V_STARCHY_POTATO >  1.78
|   |   |   |   |   |   |   |   |   |--- V_STARCHY_OTHER <= 0.25
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |--- V_STARCHY_OTHER >  0.25
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- PF_EGGS >  0.00
|   |   |   |   |   |   |   |   |--- OILS <= 0.51
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- OILS >  0.51
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- V_REDOR_TOMATO >  0.21
|   |   |   |   |   |   |   |--- SOLID_FATS <= 3.59
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- SOLID_FATS >  3.59
|   |   |   |   |   |   |   |   |--- G_REFINED <= 5.17
|   |   |   |   |   |   |   |   |   |--- V_LEGUMES <= 0.66
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- V_LEGUMES >  0.66
|   |   |   |   |   |   |   |   |   |   |--- V_REDOR_TOMATO <= 0.39
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |   |--- V_REDOR_TOMATO >  0.39
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- G_REFINED >  5.17
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- PF_SOY >  0.35
|   |   |   |   |   |   |--- class: 1
|   |   |   |   |--- V_STARCHY_OTHER >  0.62
|   |   |   |   |   |--- ADD_SUGARS <= 1.12
|   |   |   |   |   |   |--- A_DRINKS <= 0.27
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- A_DRINKS >  0.27
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- ADD_SUGARS >  1.12
|   |   |   |   |   |   |--- class: 1
|   |   |   |--- OILS >  1.44
|   |   |   |   |--- D_MILK <= 0.01
|   |   |   |   |   |--- F_OTHER <= 0.96
|   |   |   |   |   |   |--- SOLID_FATS <= 13.30
|   |   |   |   |   |   |   |--- A_DRINKS <= 1.22
|   |   |   |   |   |   |   |   |--- OILS <= 1.57
|   |   |   |   |   |   |   |   |   |--- PF_LEGUMES <= 0.89
|   |   |   |   |   |   |   |   |   |   |--- V_DRKGR <= 0.09
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |   |--- V_DRKGR >  0.09
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- PF_LEGUMES >  0.89
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- OILS >  1.57
|   |   |   |   |   |   |   |   |   |--- V_STARCHY_OTHER <= 0.01
|   |   |   |   |   |   |   |   |   |   |--- D_YOGURT <= 0.26
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 6
|   |   |   |   |   |   |   |   |   |   |--- D_YOGURT >  0.26
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |--- V_STARCHY_OTHER >  0.01
|   |   |   |   |   |   |   |   |   |   |--- OILS <= 10.40
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 5
|   |   |   |   |   |   |   |   |   |   |--- OILS >  10.40
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 4
|   |   |   |   |   |   |   |--- A_DRINKS >  1.22
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- SOLID_FATS >  13.30
|   |   |   |   |   |   |   |--- G_REFINED <= 1.20
|   |   |   |   |   |   |   |   |--- ADD_SUGARS <= 0.91
|   |   |   |   |   |   |   |   |   |--- V_STARCHY_OTHER <= 1.17
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |--- V_STARCHY_OTHER >  1.17
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- ADD_SUGARS >  0.91
|   |   |   |   |   |   |   |   |   |--- G_REFINED <= 0.88
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- G_REFINED >  0.88
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- G_REFINED >  1.20
|   |   |   |   |   |   |   |   |--- V_OTHER <= 0.46
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- V_OTHER >  0.46
|   |   |   |   |   |   |   |   |   |--- V_REDOR_OTHER <= 0.35
|   |   |   |   |   |   |   |   |   |   |--- V_OTHER <= 0.60
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |   |--- V_OTHER >  0.60
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- V_REDOR_OTHER >  0.35
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- F_OTHER >  0.96
|   |   |   |   |   |   |--- class: 1
|   |   |   |   |--- D_MILK >  0.01
|   |   |   |   |   |--- SOLID_FATS <= 18.02
|   |   |   |   |   |   |--- SOLID_FATS <= 4.09
|   |   |   |   |   |   |   |--- D_MILK <= 0.47
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- D_MILK >  0.47
|   |   |   |   |   |   |   |   |--- OILS <= 15.29
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |--- OILS >  15.29
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- SOLID_FATS >  4.09
|   |   |   |   |   |   |   |--- V_REDOR_TOMATO <= 0.50
|   |   |   |   |   |   |   |   |--- OILS <= 9.37
|   |   |   |   |   |   |   |   |   |--- OILS <= 8.30
|   |   |   |   |   |   |   |   |   |   |--- ADD_SUGARS <= 0.66
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 2
|   |   |   |   |   |   |   |   |   |   |--- ADD_SUGARS >  0.66
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 3
|   |   |   |   |   |   |   |   |   |--- OILS >  8.30
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- OILS >  9.37
|   |   |   |   |   |   |   |   |   |--- V_STARCHY_POTATO <= 0.05
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |--- V_STARCHY_POTATO >  0.05
|   |   |   |   |   |   |   |   |   |   |--- V_STARCHY_POTATO <= 0.17
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |   |--- V_STARCHY_POTATO >  0.17
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 2
|   |   |   |   |   |   |   |--- V_REDOR_TOMATO >  0.50
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- SOLID_FATS >  18.02
|   |   |   |   |   |   |--- G_REFINED <= 1.78
|   |   |   |   |   |   |   |--- V_OTHER <= 1.20
|   |   |   |   |   |   |   |   |--- G_REFINED <= 0.41
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- G_REFINED >  0.41
|   |   |   |   |   |   |   |   |   |--- V_OTHER <= 0.10
|   |   |   |   |   |   |   |   |   |   |--- OILS <= 17.61
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |   |--- OILS >  17.61
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |--- V_OTHER >  0.10
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- V_OTHER >  1.20
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- G_REFINED >  1.78
|   |   |   |   |   |   |   |--- class: 0
|   |   |--- D_CHEESE >  0.19
|   |   |   |--- OILS <= 34.49
|   |   |   |   |--- ADD_SUGARS <= 0.11
|   |   |   |   |   |--- V_OTHER <= 0.00
|   |   |   |   |   |   |--- SOLID_FATS <= 35.72
|   |   |   |   |   |   |   |--- V_STARCHY_POTATO <= 0.09
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- V_STARCHY_POTATO >  0.09
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- SOLID_FATS >  35.72
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- V_OTHER >  0.00
|   |   |   |   |   |   |--- OILS <= 3.10
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- OILS >  3.10
|   |   |   |   |   |   |   |--- OILS <= 9.22
|   |   |   |   |   |   |   |   |--- G_REFINED <= 0.66
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |--- G_REFINED >  0.66
|   |   |   |   |   |   |   |   |   |--- V_STARCHY_OTHER <= 0.11
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- V_STARCHY_OTHER >  0.11
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- OILS >  9.22
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |--- ADD_SUGARS >  0.11
|   |   |   |   |   |--- D_MILK <= 2.12
|   |   |   |   |   |   |--- F_OTHER <= 2.02
|   |   |   |   |   |   |   |--- PF_EGGS <= 0.02
|   |   |   |   |   |   |   |   |--- V_REDOR_TOMATO <= 0.00
|   |   |   |   |   |   |   |   |   |--- G_REFINED <= 1.75
|   |   |   |   |   |   |   |   |   |   |--- D_MILK <= 0.03
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 4
|   |   |   |   |   |   |   |   |   |   |--- D_MILK >  0.03
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- G_REFINED >  1.75
|   |   |   |   |   |   |   |   |   |   |--- OILS <= 0.05
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |   |--- OILS >  0.05
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- V_REDOR_TOMATO >  0.00
|   |   |   |   |   |   |   |   |   |--- V_OTHER <= 0.47
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- V_OTHER >  0.47
|   |   |   |   |   |   |   |   |   |   |--- V_OTHER <= 0.50
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |   |--- V_OTHER >  0.50
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 3
|   |   |   |   |   |   |   |--- PF_EGGS >  0.02
|   |   |   |   |   |   |   |   |--- D_CHEESE <= 0.71
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- D_CHEESE >  0.71
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- F_OTHER >  2.02
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- D_MILK >  2.12
|   |   |   |   |   |   |--- class: 1
|   |   |   |--- OILS >  34.49
|   |   |   |   |--- D_CHEESE <= 1.06
|   |   |   |   |   |--- class: 0
|   |   |   |   |--- D_CHEESE >  1.06
|   |   |   |   |   |--- class: 1
|   |--- PF_EGGS >  0.03
|   |   |--- D_MILK <= 0.01
|   |   |   |--- D_CHEESE <= 0.24
|   |   |   |   |--- V_REDOR_TOMATO <= 0.49
|   |   |   |   |   |--- ADD_SUGARS <= 0.03
|   |   |   |   |   |   |--- V_STARCHY_OTHER <= 0.01
|   |   |   |   |   |   |   |--- PF_EGGS <= 0.09
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- PF_EGGS >  0.09
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- V_STARCHY_OTHER >  0.01
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- ADD_SUGARS >  0.03
|   |   |   |   |   |   |--- V_OTHER <= 1.67
|   |   |   |   |   |   |   |--- SOLID_FATS <= 50.96
|   |   |   |   |   |   |   |   |--- V_REDOR_OTHER <= 1.23
|   |   |   |   |   |   |   |   |   |--- OILS <= 5.87
|   |   |   |   |   |   |   |   |   |   |--- G_REFINED <= 0.09
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |   |--- G_REFINED >  0.09
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 7
|   |   |   |   |   |   |   |   |   |--- OILS >  5.87
|   |   |   |   |   |   |   |   |   |   |--- V_STARCHY_POTATO <= 0.77
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 3
|   |   |   |   |   |   |   |   |   |   |--- V_STARCHY_POTATO >  0.77
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 2
|   |   |   |   |   |   |   |   |--- V_REDOR_OTHER >  1.23
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- SOLID_FATS >  50.96
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- V_OTHER >  1.67
|   |   |   |   |   |   |   |--- ADD_SUGARS <= 4.38
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- ADD_SUGARS >  4.38
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |--- V_REDOR_TOMATO >  0.49
|   |   |   |   |   |--- V_REDOR_OTHER <= 0.00
|   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- V_REDOR_OTHER >  0.00
|   |   |   |   |   |   |--- class: 1
|   |   |   |--- D_CHEESE >  0.24
|   |   |   |   |--- OILS <= 11.27
|   |   |   |   |   |--- V_REDOR_TOMATO <= 0.31
|   |   |   |   |   |   |--- A_DRINKS <= 0.50
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- A_DRINKS >  0.50
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- V_REDOR_TOMATO >  0.31
|   |   |   |   |   |   |--- class: 1
|   |   |   |   |--- OILS >  11.27
|   |   |   |   |   |--- OILS <= 26.79
|   |   |   |   |   |   |--- PF_LEGUMES <= 1.74
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- PF_LEGUMES >  1.74
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- OILS >  26.79
|   |   |   |   |   |   |--- SOLID_FATS <= 20.41
|   |   |   |   |   |   |   |--- SOLID_FATS <= 7.76
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- SOLID_FATS >  7.76
|   |   |   |   |   |   |   |   |--- PF_EGGS <= 0.05
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- PF_EGGS >  0.05
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- SOLID_FATS >  20.41
|   |   |   |   |   |   |   |--- class: 0
|   |   |--- D_MILK >  0.01
|   |   |   |--- F_CITMLB <= 0.00
|   |   |   |   |--- OILS <= 7.73
|   |   |   |   |   |--- OILS <= 0.55
|   |   |   |   |   |   |--- V_REDOR_TOMATO <= 0.07
|   |   |   |   |   |   |   |--- D_MILK <= 0.56
|   |   |   |   |   |   |   |   |--- ADD_SUGARS <= 3.74
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |--- ADD_SUGARS >  3.74
|   |   |   |   |   |   |   |   |   |--- ADD_SUGARS <= 4.75
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- ADD_SUGARS >  4.75
|   |   |   |   |   |   |   |   |   |   |--- F_OTHER <= 0.07
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |   |--- F_OTHER >  0.07
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- D_MILK >  0.56
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- V_REDOR_TOMATO >  0.07
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- OILS >  0.55
|   |   |   |   |   |   |--- D_MILK <= 0.03
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- D_MILK >  0.03
|   |   |   |   |   |   |   |--- V_REDOR_OTHER <= 0.00
|   |   |   |   |   |   |   |   |--- V_OTHER <= 0.21
|   |   |   |   |   |   |   |   |   |--- D_CHEESE <= 0.03
|   |   |   |   |   |   |   |   |   |   |--- G_REFINED <= 0.77
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |   |--- G_REFINED >  0.77
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 5
|   |   |   |   |   |   |   |   |   |--- D_CHEESE >  0.03
|   |   |   |   |   |   |   |   |   |   |--- ADD_SUGARS <= 0.27
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |   |--- ADD_SUGARS >  0.27
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- V_OTHER >  0.21
|   |   |   |   |   |   |   |   |   |--- V_OTHER <= 1.50
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- V_OTHER >  1.50
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- V_REDOR_OTHER >  0.00
|   |   |   |   |   |   |   |   |--- OILS <= 1.76
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- OILS >  1.76
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |--- OILS >  7.73
|   |   |   |   |   |--- D_CHEESE <= 0.82
|   |   |   |   |   |   |--- G_REFINED <= 0.15
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- G_REFINED >  0.15
|   |   |   |   |   |   |   |--- V_DRKGR <= 0.03
|   |   |   |   |   |   |   |   |--- G_REFINED <= 2.86
|   |   |   |   |   |   |   |   |   |--- SOLID_FATS <= 18.15
|   |   |   |   |   |   |   |   |   |   |--- G_REFINED <= 1.37
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 2
|   |   |   |   |   |   |   |   |   |   |--- G_REFINED >  1.37
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 4
|   |   |   |   |   |   |   |   |   |--- SOLID_FATS >  18.15
|   |   |   |   |   |   |   |   |   |   |--- V_REDOR_OTHER <= 0.00
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |   |--- V_REDOR_OTHER >  0.00
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |--- G_REFINED >  2.86
|   |   |   |   |   |   |   |   |   |--- V_REDOR_OTHER <= 0.28
|   |   |   |   |   |   |   |   |   |   |--- V_REDOR_TOMATO <= 0.01
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |   |--- V_REDOR_TOMATO >  0.01
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 3
|   |   |   |   |   |   |   |   |   |--- V_REDOR_OTHER >  0.28
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- V_DRKGR >  0.03
|   |   |   |   |   |   |   |   |--- G_REFINED <= 7.83
|   |   |   |   |   |   |   |   |   |--- OILS <= 48.51
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |--- OILS >  48.51
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- G_REFINED >  7.83
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |--- D_CHEESE >  0.82
|   |   |   |   |   |--- SOLID_FATS <= 13.56
|   |   |   |   |   |   |--- G_REFINED <= 3.08
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- G_REFINED >  3.08
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- SOLID_FATS >  13.56
|   |   |   |   |   |   |--- class: 0
|   |   |   |--- F_CITMLB >  0.00
|   |   |   |   |--- class: 1
```

Support Vector Machine Model

A Support Vector Machine model does not provide any explainability of the model features, but is explored here to see if it can provide any improvement in performance.

[117]:
```python
import numpy as np
from sklearn import datasets
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import StandardScaler
from sklearn.svm import LinearSVC

df = pd.read_csv('../../Data/nhanes_full_pre_proc.csv')
df = df[df['meal_energy']=='Medium-Low']


def svm_regressor(df, variables, n_sfd, n_non_sfd, param):

    df_non_sfd = df[df['seafood_meal']==0].sample(n=n_non_sfd)
    df_sfd = df[df['seafood_meal']==1].sample(n=n_sfd)
    df = pd.concat([df_non_sfd, df_sfd])

    X = df[variables]
    Y = df['seafood_meal']
    X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.2)

    #svm_nhanes = LinearSVC(C=1, loss="hinge")

    svm_nhanes = Pipeline([
        ("scaler", StandardScaler()),
        ("linear_svc", LinearSVC(C=param, loss="hinge"))
    ])

    #svm_nhanes = LinearSVC(C=1, loss="hinge")
    svm_nhanes.fit(X_train, y_train)
    y_pred_svm = svm_nhanes.predict(X_test)
    score = accuracy_score(y_test, y_pred_svm)
    return score



result = []
for i in range(100):
    model_fit = svm_regressor(df, food_cmp_level5, 500, 500, 1)
    result.append(model_fit)

result = pd.DataFrame(result)
print(result.describe())
```

```
               0
count  100.00000
mean     0.65960
std      0.03379
min      0.59500
25%      0.63500
50%      0.66000
75%      0.68000
max      0.74000
```

This model does not provide any improvement in performance from the previous models.

```
[ ]:
```