Exploring Students Eating Habits through Individual Profiling and Clustering Analysis

Michela Natilli

michela.natilli@isti.cnr.it

Introduction

- Individual well-being strongly depends on food habits
- Understanding the real eating habits of people becomes fundamental for a better and more effective intervention to improve the students' diet
- We present two exploratory analyses based on centroid-based clustering to understand the food habits of university students.
  1. It exploits the information on the students' food consumption of specific food categories
  2. It includes the temporal dimension to capture the information on when the students consume specific foods.

Michela Natilli
michela.natilli@isti.cnr.it

Food Dataset

Dataset provided by the Tuscan Institute of Right to Study (DSU)

Meals	10,034,413
Students	82,871
with grant	*19,141*
free meals	*4,730,658*
Dishes	950
Food categories	41
Period	2,551 days
from	*01/01/2010*
to	*12/26/2016*

Michela Natilli
michela.natilli@isti.cnr.it

Food Dataset (2)

dish	attribute	description	long description
Bolognese pasta	a_{11}	meat flours	flours (pasta, couscous, dumplings) with meat/cheese/eggs
Pasta with pesto	a_{13}	veg flours	flours (pasta, couscous, dumplings) with vegetables
Pasta with zucchini	a_{13}	veg flours	flours (pasta, couscous, dumplings) with vegetables
...	...	...	...
Saffron and potato soup	a_{34}	legumes soup	potato and legumes soups
Hamburger with mushrooms	a_{51}	read meat	red meat / salami
...	...	...	...
Green salad	a_{81}	raw veg	raw vegetables
...	...	...	...
...	...	...	...
Fruit	a_{415}	fruit	fruit
Cheesecake	a_{416}	dessert	dessert

Michela Natilli
michela.natilli@isti.cnr.it

Food Dataset

Each meal is described by a record indicating the student who consumed the meal, the type of the meal, and the list of dishes

Student_id	timestamp	Dishes
A4578A	18/04/2015 12:42:00	pasta with tomato sauce, chicken breast, fruit
G23T20	18/04/2015 12:43:00	mushroom risotto, salad, fruit
GE54Y7	18/04/2015 12:44:01	pasta with tomato, fruit

Food Dataset

The total number of meals consumed at the canteen does not vary significantly over the years and it is about 1,400K each year. On the other hand, there is a slightly decrease of the number of students going to the university canteen, passing from the 30k of 2010 to the 27k of 2016.

All students:	
Students with at least 10 meals:	60,112
Students with at least 100 meals:	22,647
Students with at least 1000 meals:	1,448

Michela Natilli
michela.natilli@isti.cnr.it

Methods

Two exploratory analyses based on centroid-based clustering

1. The first clustering analysis exploits the information on the students' food consumption of specific food categories
2. The second exploratory analysis includes the temporal dimension in order to capture the information about when the students consume specific foods.

Pre-processing: select only students who had at least 100 meals and having a distance between two meals of a maximum of 100 days

Michela Natilli
michela.natilli@isti.cnr.it

Method 1: Food Consumption Analytics

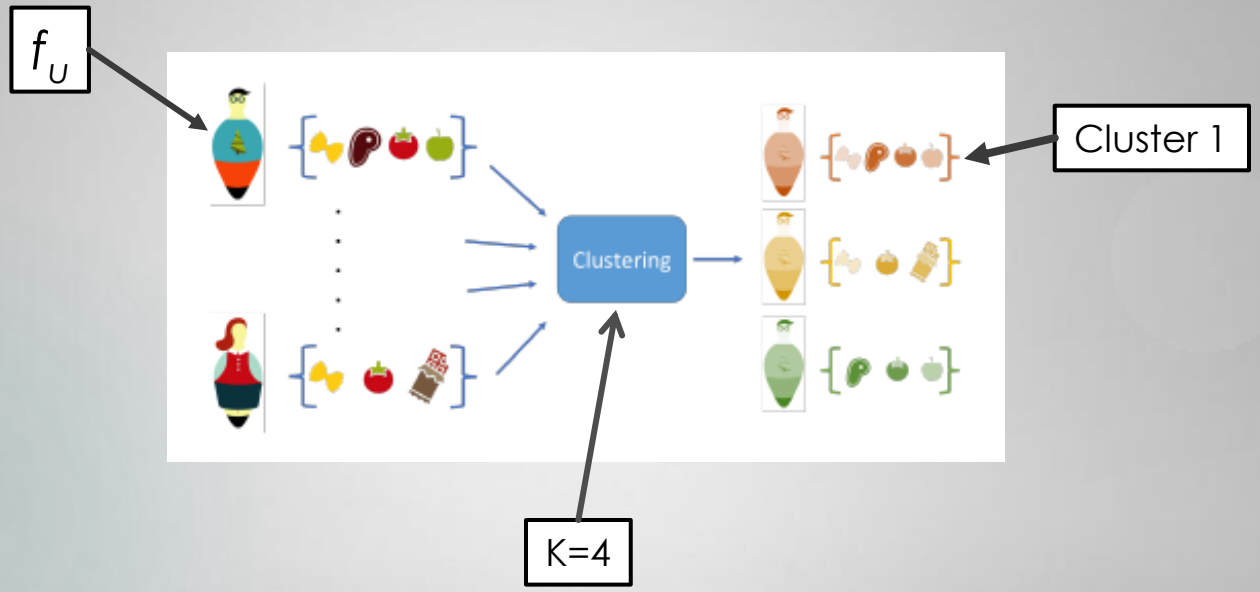
Student food consumption derived by using an individual model *foodprint*.

- We represent a *foodprint* f_u of a student *u* using a vector of 41 attributes a_i with i=1; : : : ; 41 (an attribute for each food category).
- Let $D_i^{(u)}$ be the number of dishes of the student *u* in the food category *i* and $M^{(u)}$ be the total number of meals of the student *u*, the value of the i-th attribute is

$$a_i^{(u)} = \frac{D_i^{(u)}}{M^{(u)}}$$

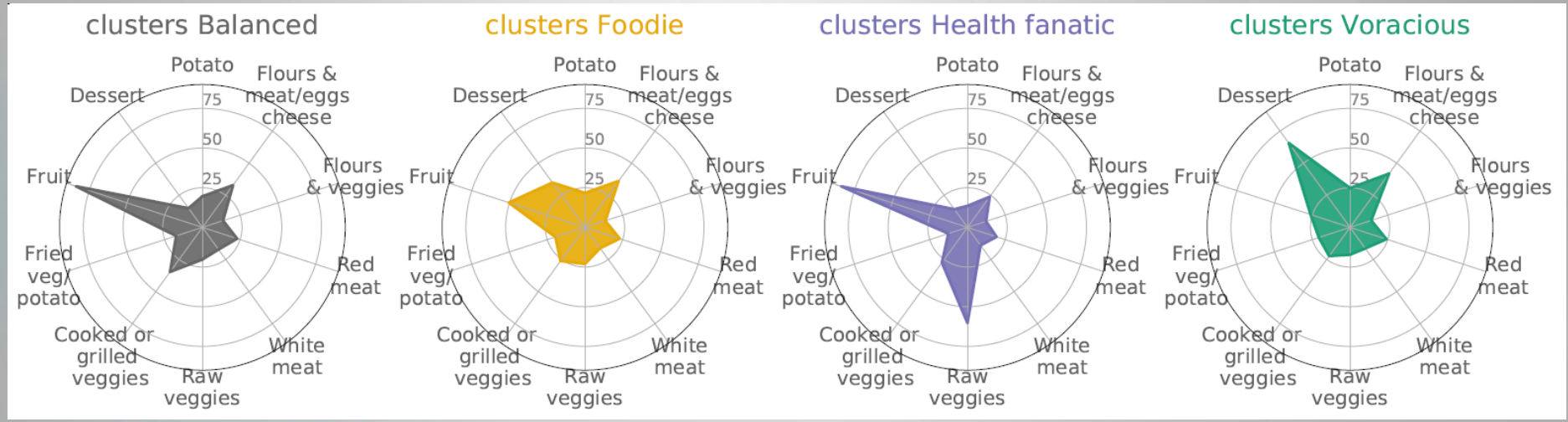
Pre-processing: select only students who had at least 100 meals and having a distance between two meals of a maximum of 100 days

Clustering: k-means (k=4)



Results method 1

- **Cluster 0**: Balanced (30.95%)
- **Cluster 1**: Foodie (17.43%)
- **Cluster 2**: Health fanatic (33.64%)
- **Cluster 3**: Voracious (17.98%)



Michela Natilli
michela.natilli@isti.cnr.it

Methods

Two exploratory analyses based on centroid-based clustering

1. The first clustering analysis exploits the information on the students' food consumption of specific food categories
2. The second exploratory analysis includes the temporal dimension in order to capture the information about when the students consume specific foods.

This clustering process is based on the definition of the **temporal *foodprints*** of a student, and on two tasks called ***food habits discovery*** and ***student grouping***

Michela Natilli
michela.natilli@isti.cnr.it

Method 2: Temporal Food Habits Analytics

The student *temporal foodprint* include a temporal dimension on the year, the season and whether the meal was consumed for lung or dinner

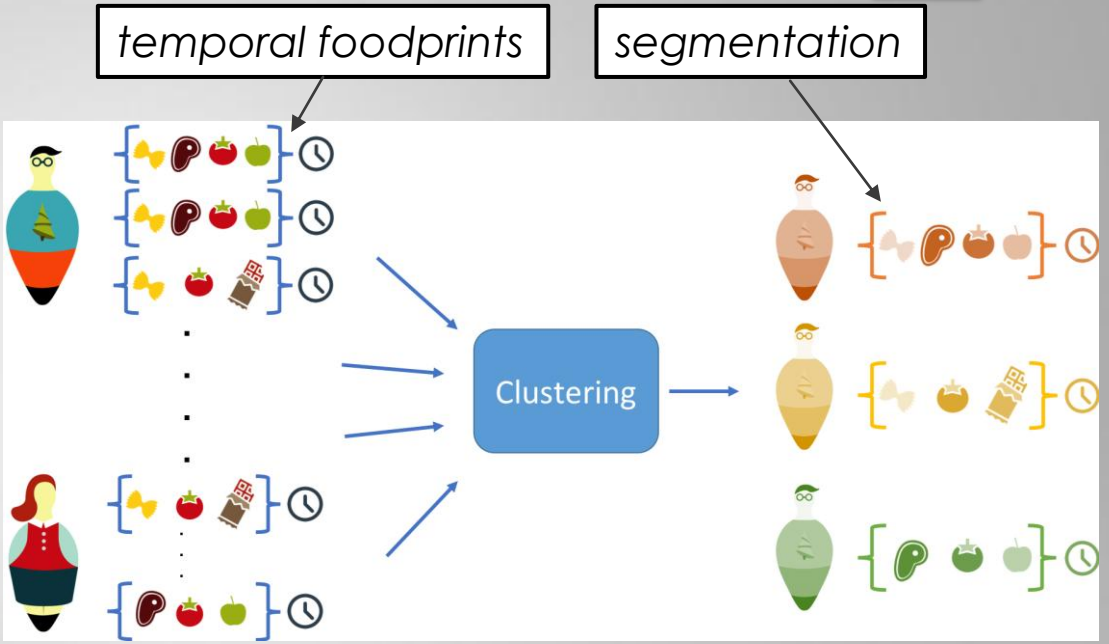
Pre-processing: select only students who had at least 10 meals: 45,952 students

Michela Natilli
michela.natilli@isti.cnr.it

Method 2: food habits discovery

Given the *temporal foodprints*, we apply a clustering algorithm (k-means with k=50) to extract typical food habits
➔ segmentation of the temporal foodprints

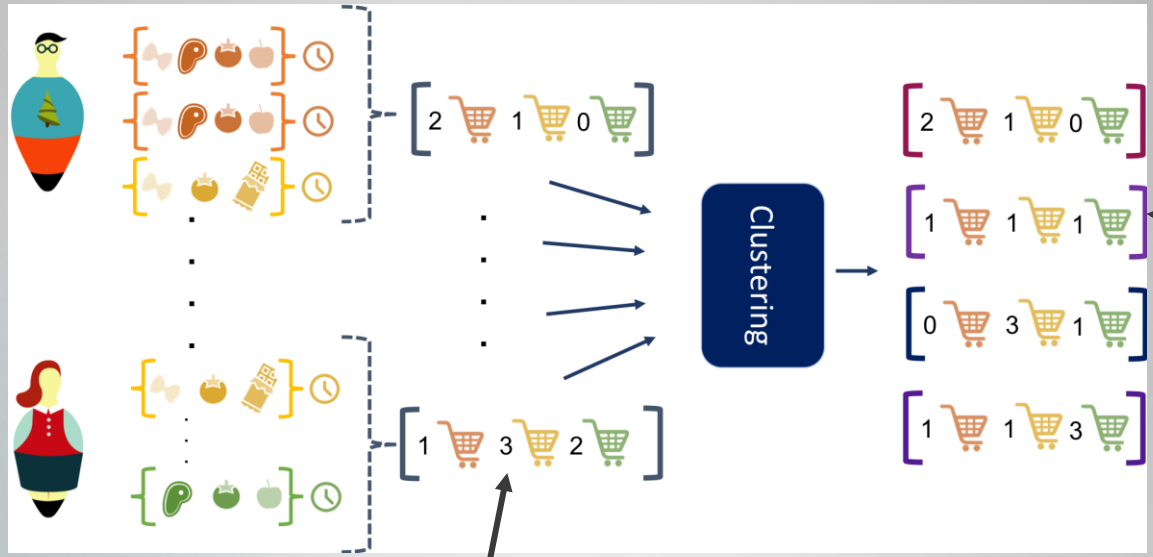
NB: A student can have his *temporal foodprints* distributed over different clusters



Method 2: student grouping

- Construction of a food habit profile: the intensity of the student presence in each cluster (food habit).
- The student profile is a vector of 50 attributes where each attribute is the intensity of a certain temporal food habit.
- Group the students according to their temporal food habits profiles

Method 2: student grouping



Michela Natilli
michela.natilli@isti.cnr.it

Method 2: clustering analysis

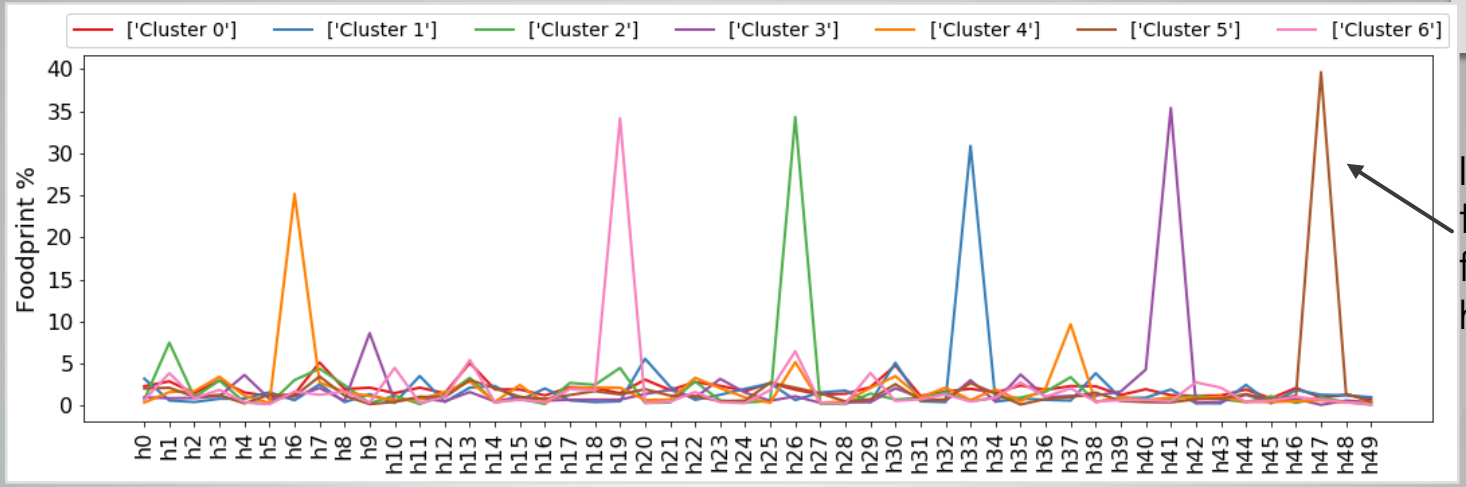
We performed this analysis using different temporal aggregations *t*:

1. year and lunch/dinner information,
2. month and lunch/dinner information,
3. season and lunch/dinner information
4. year, season and lunch/dinner information

Using the elbow method we selected k = 7 as number of different groups for the food habits profiles.

Michela Natilli
michela.natilli@isti.cnr.it

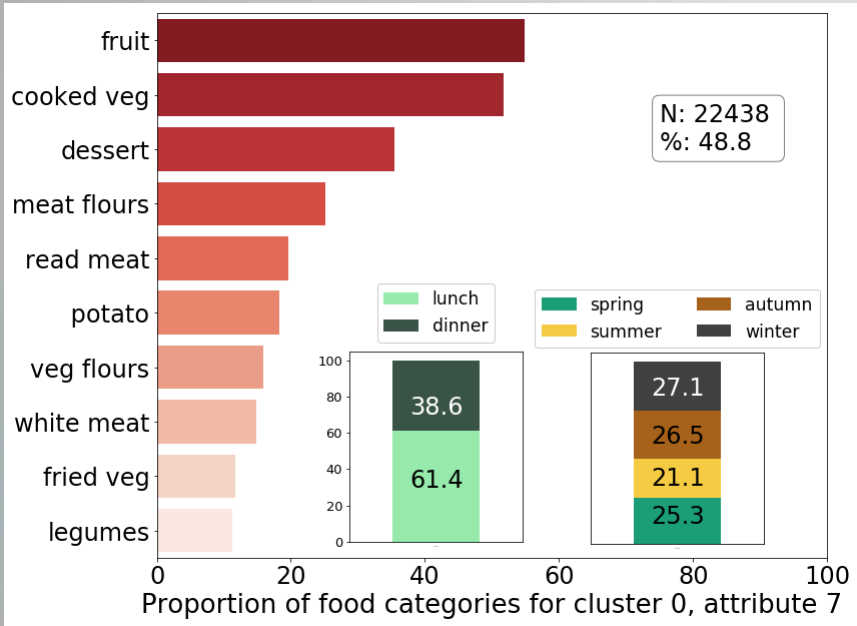
Results method 2

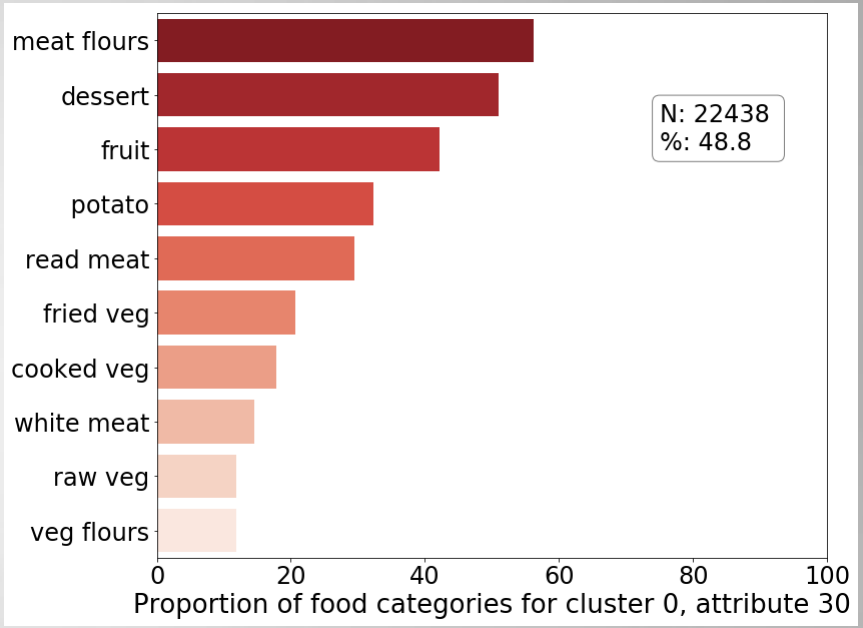


Intensity of temporal food habit

With the exception of cluster 0 (three food habits higher than the rest) for all the other centroids it is possible to isolate a unique ***dominant food habit****.*
We can characterize the seven groups by means of the dominant food habits of their centroids

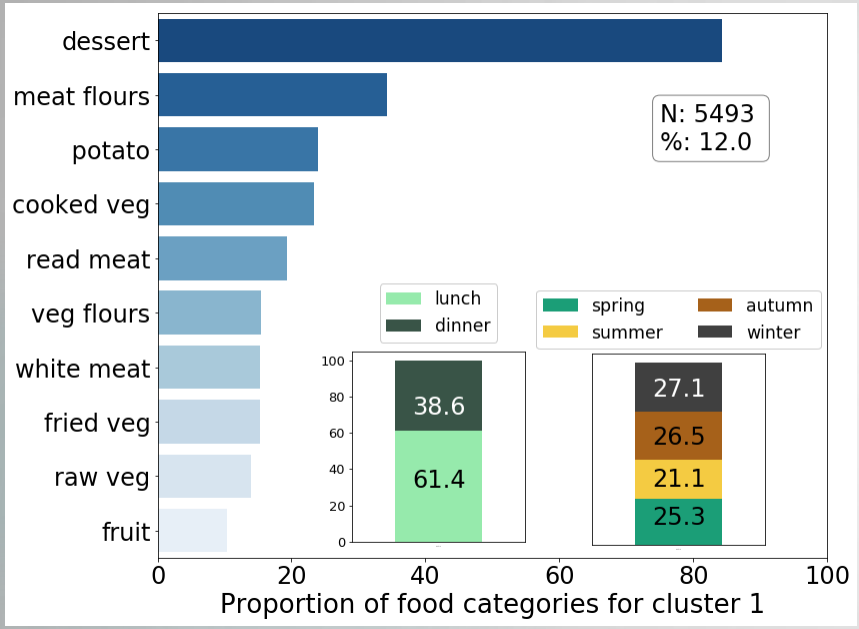
Results method 2: cluster 0

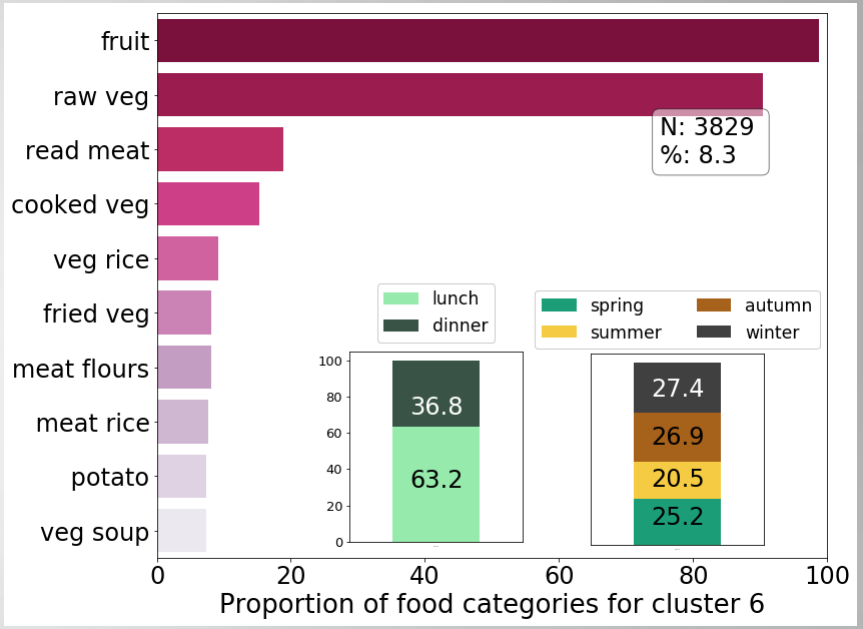




Michela Natilli
michela.natilli@isti.cnr.it

Results method 2: clusters 1 & 6





Michela Natilli
michela.natilli@isti.cnr.it

Conclusions

Two exploratory analyses based on centroid-based clustering aiming at understating the food habits of university students.

The results of our analyses could be useful for suggest improvements to the students diet (privacy)

Future improvements:

- Lower generalization of food categories
- The nutrient values of the meals

Michela Natilli
michela.natilli@isti.cnr.it

Thank you

Michela Natilli

michela.natilli@isti.cnr.it