Food symphony

Applied
Data Analysis
2018

Jonathan Chansin
Etienne Dubois
Tom Suter

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

dlab

1 Text processing

Raw:
— #6124: 1 pack (3/2 oz) of (red, yellow or green) cherry tomatoes
— 1 teaspoon chopped fresh dill or 1/2 teaspoon dried dill

Filter 1: Wise handling of special characters and numeric conversion.
— ~~#6124~~: 1 pack (3/2 oz) of (red, yellow or green) cherry tomato
— 1 teaspoon chopped fresh dill or 1/2 teaspoon dried dill

Filter 2: Parenthesis handling
— ~~1 pack~~ (1.5 oz) of ~~(red, yellow or green)~~ cherry tomato
— 1 teaspoon chopped fresh dill or 0.5 teaspoon dried dill

Filter 3: Selecting only one word from each list we create (techniques, units, ingredients)
— 1.5 oz ~~of~~ cherry tomato
— 1 teaspoon chopped ~~fresh~~ dill ~~or~~ ~~0.5 teaspoon dried dill~~

Final:
— 1.5 oz cherry tomato
— 1 teaspoon chopped dill

90 g

15 g

100 g

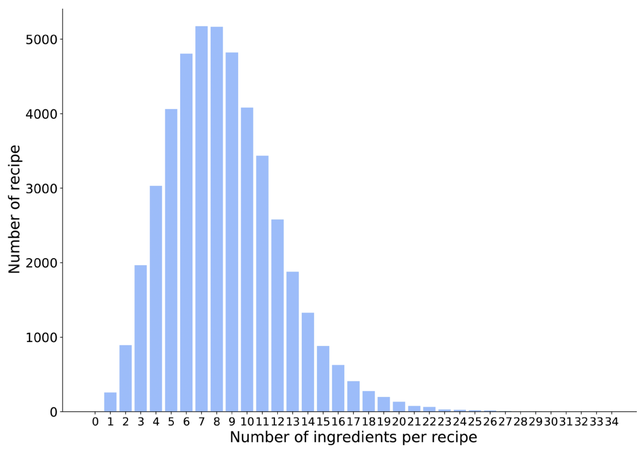
89 g

17 g

71 g

155 g

2 Dataset Comprehension



Positive skewed normal distribution of ingredients number per recipe. Gives the range of the ingredients to use while creating a new recipe.

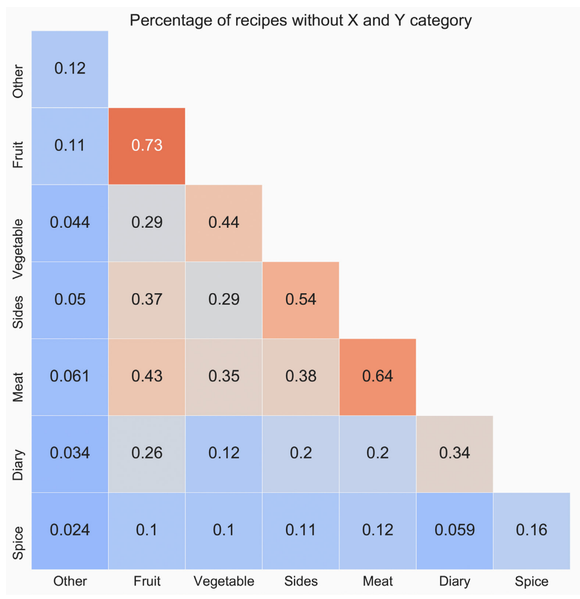


Table expressing the percentage of recipes without the union of the categories.

The following informations can be deduced: 64 % of recipe are vegetarian, 20 % are vegan and 27 % of them contain fruits. The ones that are not desserts in those 27 %, are a potential source of new ingredients association.

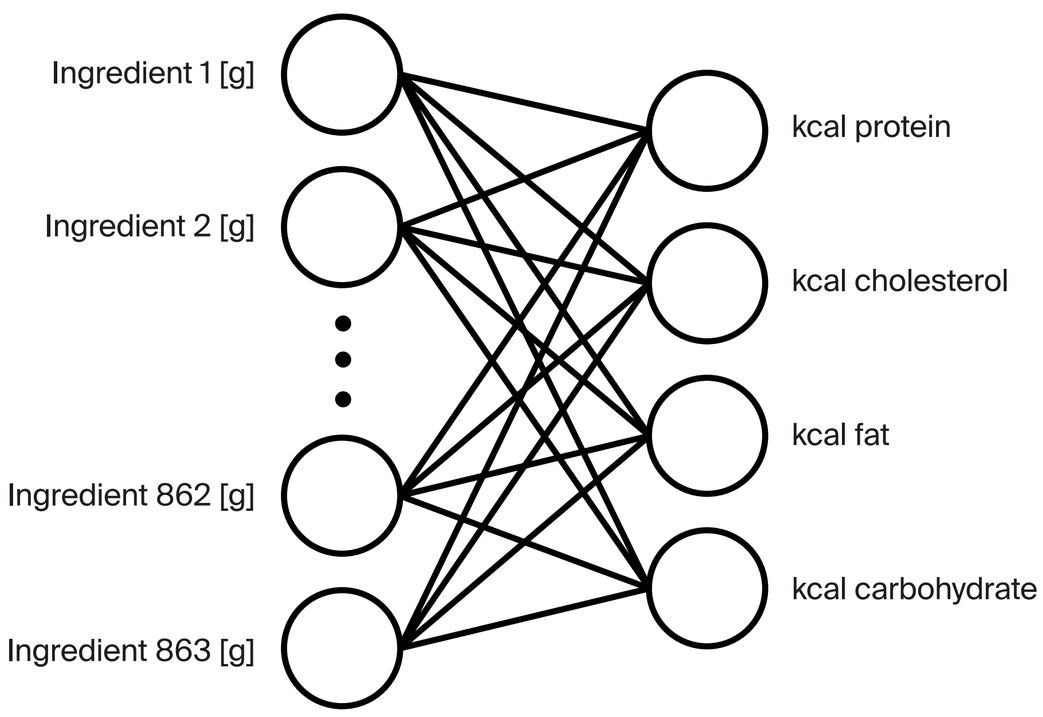
3 How to choose ingredients to form a recipe?

Co-occurence:
Relation intensity between ingredients

K-Nearest Neighbors:
Return values of relative distance between ingredients according to co-occurence matrice. It spots closest ingredients that don't necessarily have direct links or the highest co-occurence values.

4 Perceptron: Prediction of nutrients

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}[\log(\mathbf{wX_i}) - \log(Y)]^2 + \gamma\sum_{i=1}^{N}\mathbf{w}_i^2$$



Nutrition Facts

Serving Size 100g

Calories	334.7
Total Fat	9.8 g
Cholesterol	3.6 mg
Total Carbohydrate	41.0 g
Protein	20.7 g

5 Conclusion:

— Good predictions on nutrients values per gram considering the assumptions
— Recipe generator gives a lot of flexibility
— Creating recipes with precise quantities of total nutrients
— Creating recipes including only low-fat ingredients

Improvements:
— Strong assumptions on ingredients' quantities to relax
— Number of ingredients per recipe to be drawn from distribution
— Use ranking of KNN instead of uniform distribution in top 20 neighbours