### Compte rendu de nos Analyses Agribalyse

### Mise en Forme et traitement

- Numérisation des données
- Drop DQR > 3
- Matrice de corrélation
- Drop des attributs non pertinents
- Normalisation

## Traitement du problème

Supervisé
Problème du score EF:
non concluant au vu de
la répartition des
classes (<0.1)

Problème du **goupe d'aliments** (prédire le groupe d'aliment à partir des attribbuts

# Classification sur les groupes d'aliments

Accuracy sur les données test

• KNN(3 voisins): 0.66

• Perceptron: 0.42

• Bagging Tree: 0.65

### Conclusion

Le KNN et le Bagging Tree sont assez similaire, on préfère le bagging Tree pour son explicabilité sur la classification.

. . . . . . .

# Silo Non

## Traitement du problème

Application d'un clustering K-means avec 2 clusters

### Analyse des résultats

Moyenne de chaque attribut pour chacun des clusters comparaison de ces moyennes pour avoir les attributs les plus discriminants.

On a vu que les clusters étaient séparés sur le taux d'impact sur l'environnement pour la consommation et l'agriculture

Filtre des ingrédients qui appartiennent au cluster 'polluant" et récupération de leur code CIQUAL

### Association des codes CIQUAL

On fait un "jointure" sur le code CIQUAL entre ingrédients et etape.

Récupération des sous groupe d'aliments auxquels appartiennent les ingrédients.

### Conclusion

On remarque que les ingrédients qui polluent tout le temps appartiennent aux sous groupes herbes, huiles et graisses végétales et pommes de terre et autres tubercules