



VOORTGANGS PRESENTATIE

Anass, Bram, Brian, Ilias, Jurriaan en Milan

AANPASSING VAN ONZE DATASET

In plaats van lists met data, binaire waardes.

Alle bijbehorende tags + ingrediënten.

nutritions		tags	Ingredients			
['energie', '675 kcal', 'koolh		['glutenvrij', 'oven', 'frans', '	['vastkokende aardappel', 'middel			
['energie', '515 kcal', 'koolh		['oven', 'indonesisch', 'aziat	['limoen', 'scharrel kippenpoten na			
['energie', '765 kcal', 'koolh		['oven', 'hollands', 'hoofdge	['knoflook', 'bloemkool', 'ongezou			
['energie', '505 kcal', 'koolh		['vegetarisch', 'oven', 'holla	['milde olijfolie', 'koelverse aardap			
['energie', '635 kcal', 'koolh		['budget', 'oven', 'couscous	['bospeen', 'rode ui', 'kipdijfilet', 'p			
index	name	glutenvrij	oven	frans	hoofdgerecht	indonesisch
0	Tartiflette		1	1	1	0
1	Ajam pangang		0	1	0	1
2	Bloemkoolschotel		0	1	0	0
3	Kapsalon' met vegetarische8		0	1	0	0
4	Couscous met kip, bospeen en cran		0	1	0	0
5	Lekkerbekje met aardappel en groe		0	1	0	0

Ingrediënt pas in list wanneer deze 40 keer voorkomt.

EERSTE MODEL MAKEN/TESTEN

Verschillende modellen getest(KNN, DecisionTree, RandomForest, naive bayes)

Parameters getest doormiddel van GridSearchCV

Ondanks GridSearch nog steeds te hoge Accuracy, Recall en Precision

```
accuracy: 0.9566666666666667 prec: 0.9468085106382979 recall: 0.8090909090909091
```

ONZE VRAAG AAN JULLIE

Hebben jullie ook problemen met overfitting?

Zijn er groepen die hier een oplossing voor hebben gevonden?