



**POLYTECHNIQUE
MONTREAL**

**LE GÉNIE
EN PREMIÈRE CLASSE**

INF8215
Groupe 01

TP3

Classifications multiclass : légumes secs

Par

Brando, Tovar **1932052**

Vega, Estefan **1934346**

Équipe : **BrandiniStifini**

Le 14 avril 2022

Table des matières

1	Contexte	2
2	Prétraitement	2
3	Méthodologie	3
4	Résultats	3
5	Discussion	3
6	stuff	3
7	Section 2	4
	Références	6

1 Contexte

Dans ce travail pratique, il nous était demandé de classer des légumes secs dans leur catégorie respectives. Il y en avait 7 en tout; Sira, Horoz, Dermason, Barbunya, Cali, Bombay, Seker et nous devons déterminer la catégorie à l'aide de 16 *features*. Nous avons donc à résoudre un problème de classification multiclass. Nous avons décidé d'utiliser la librairie *scikit-learn* et le modèle que nous avons utilisé se base sur les machines à vecteurs de support (*SVC OneVsOneClassifier*). Nous avons aussi exploré d'autres modèle telles que celui basée sur la descente de gradient stochastique (*SGDClassifier*) et celle basée sur les forêts aléatoires (*RandomForestClassifier*)

2 Prétraitement

Avant de commencé à résoudre le problème, il est utile de se familiariser avec les données. Nos données étaient constitué de 16 *features*. Il y avait 6000 données de test. Nous avons commencé par voir s'il manquait des valeurs dans certaines de no données test ce qui n'était pas le cas. Nous avons ensuite regardé si nos données étaient équilibrées.

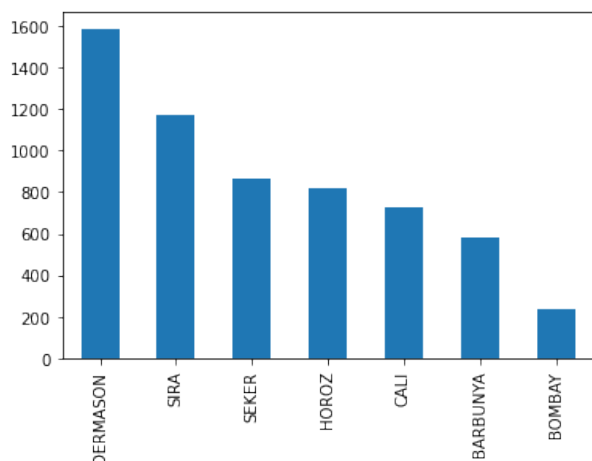


Figure 1 - Diagramme à bandes des catégories

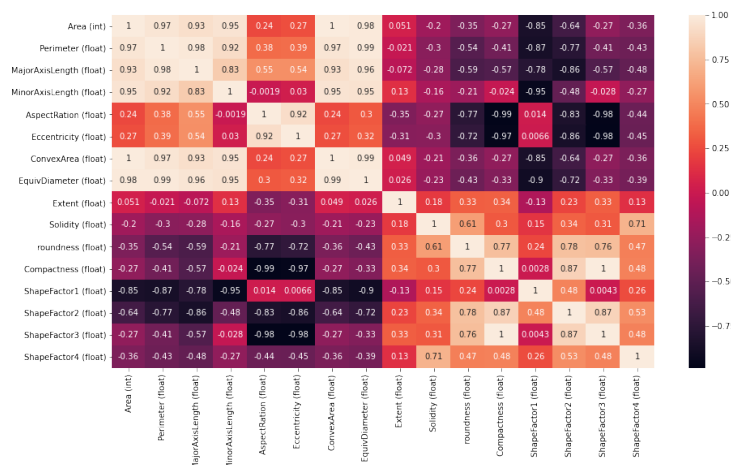


Figure 2 - Matrice de corrélation

Il a été intéressant de voir qu'il y a présence de déséquilibre et que certains attributs étaient très corrélés entre eux soit > 0.9 ou < -0.9 . Nous avons essayé de retirer les attributs corrélés, mais il n'y a pas vraiment eu de gain et nous avons donc décidé de garder tous les attributs.

Une fois l'analyse de données terminée, nous avons dû faire quelques changements dans les données afin de pouvoir utiliser notre modèle. Nous avons en premier lieu, dû transformer les valeurs de X_{train} (attributs) en float. Nous avons ensuite retiré les valeurs de ID dans X_{train} et y_{train} . Nous pouvions ainsi entraîner notre modèle. Étant donné les mauvais résultats initiaux, nous avons dû faire appel à la normalisation. Nous sommes donc passés d'une précision de 0.268 à une de 0.933 sur le classificateur SVM.

3 Méthodologie

4 Résultats

5 Discussion

6 stuff

N	N ²	N ³	N ⁴	sqrt(n)	sqrt[4](N)
1	1	1	1	1	1
2	4	8	16	1.4142	1.1892
3	9	27	81	1.7321	1.3161

Table 1 – A table

N	N ²	N ³	N ⁴	sqrt(n)	sqrt[4](N)
1	1	1	1	1	1
2	4	8	16	1.4142	1.1892
3	9	27	81	1.7321	1.3161

Table 2 – Another table

Table 3 – Regular table

N	N ²	N ³	N ⁴	sqrt(n)	sqrt[4](N)
1	1	1	1	1	1
2	4	8	16	1.4142	1.1892
3	9	27	81	1.7321	1.3161

```
1 ls -l
```

Listing 1 – Source code

```
total          256
-rw-rw-r--    1 bndo  bndo    1406 Apr 13 21:12 config.tex
drwxrwxr-x    2 bndo  bndo    4096 Apr 13 23:18 img
drwxrwxr-x    2 bndo  bndo    4096 Apr 13 23:34 _minted-report
-rw-rw-r--    1 bndo  bndo     459 Apr 13 21:12 packages.tex
-rw-rw-r--    1 bndo  bndo    1378 Apr 13 21:12 README.org
-rw-rw-r--    1 bndo  bndo    2978 Apr 14 01:24 report.bbl
-rw-rw-r--    1 bndo  bndo     7162 Apr 14 01:44 report.org
-rw-rw-r--    1 bndo  bndo   214812 Apr 14 01:24 report.pdf
-rw-rw-r--    1 bndo  bndo     5780 Apr 14 01:24 report.tex
-rw-rw-r--    1 bndo  bndo     738 Apr 13 21:12 template.bib
```

7 Section 2

selon une étude (Lemieux et al., [2021](#))

Références

Lemieux, V. L., Mashatan, A., Safavi-Naini, R. & Clark, J. (2021). A Cross-Pollination of ideas about Distributed Ledger Technological Innovation through a Multidisciplinary and Multisectoral lens : Insights from the Blockchain Technology Symposium '21. *Technology Innovation Management Review*, 58-66. <https://doi.org/10.22215/timreview/1445>