

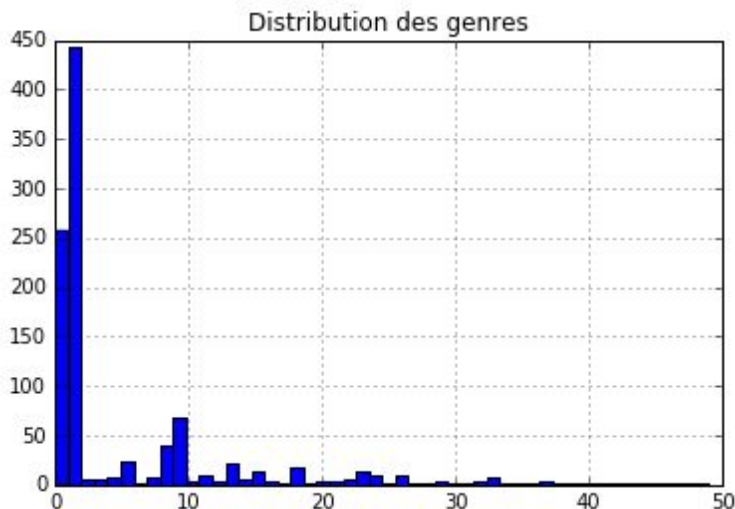
Projet : Classification de textes de théâtre par genre

Eléonore Bartenlian, Adrien Pavao



Corpus : 1030 textes, 310 auteurs, 50 genres

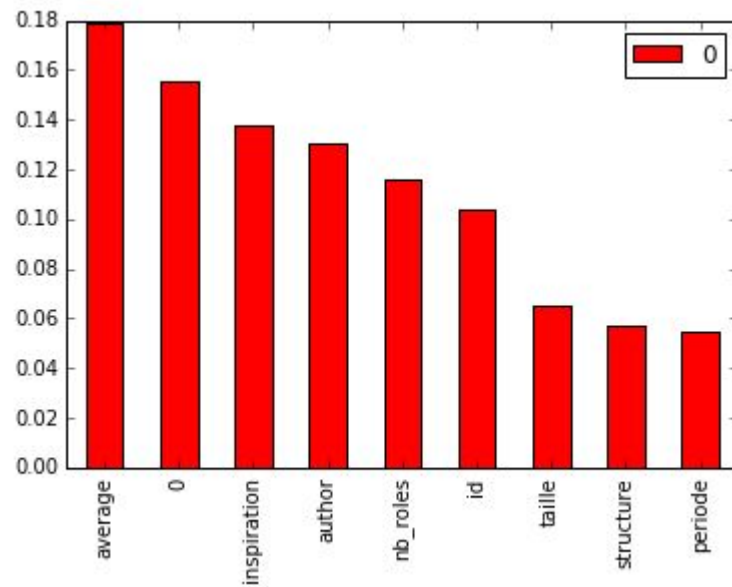
Ballet
Mélodrame
Farce
Ambigu poissard
Proverbe
Bergerie
opuscule dramatique
Pastorale héroïque
Mascarade
Pantomime
Comédie
Dialogue
Pièce
Pièce dramatique
Divertissement
Pièce épisodique
Tragédie
Tragi-comédie
Comédie héroïque
Opéra Bouffe
Saynette
Opéra
Comédie-ballet



Attributs :

- L'auteur,
- La taille moyenne des phrases,
- La longueur du texte,
- La date,
- L'inspiration,
- La structure,
- La période,
- Le nombre de personnages,
- Le genre pour lequel les mots du texte ont le plus de correspondance.

	author	average	id	inspiration	nb_roles	periode	structure	taille	0
741	105	8.143103	741	2	11	16	5	1500	0.0
638	236	6.232143	638	8	3	19	1	0	1.0
993	305	3.445714	993	1	2	18	1	0	47.0
279	140	7.885151	279	6	14	16	5	2000	1.0
796	60	8.856734	796	2	8	16	5	1750	1.0
240	127	6.056782	240	8	7	18	1	500	1.0
204	112	4.803235	204	1	6	17	1	500	1.0
941	68	5.484517	941	1	10	17	5	0	1.0
5	1	5.335731	5	1	15	16	3	0	1.0
706	66	5.657895	706	1	6	17	1	0	1.0
703	25	6.183099	703	1	9	16	5	0	1.0
746	7	4.525000	746	1	1	18	1	0	1.0
821	142	4.962366	821	1	14	17	1	500	1.0



Avec Random Forest Classifier, 90% entraînement, 10% test :
Taux de réussite : **79%**