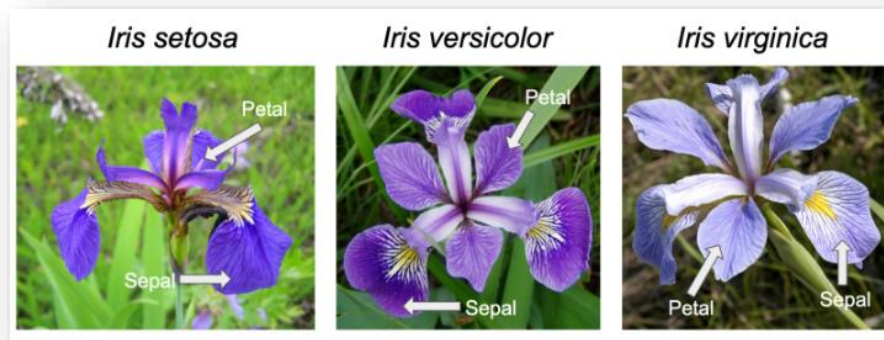




Iris Classification project using R



SOMMAIRE



Mise en place des données



Visualisation



Data Science : SVM



Bilan



I – Mise en place des données

I - Création du Dataframe

Data Exploration.R x df_iris x

Filter

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa
7	4.6	3.4	1.4	0.3	setosa
8	5.0	3.4	1.5	0.2	setosa
9	4.4	2.9	1.4	0.2	setosa
10	4.9	3.1	1.5	0.1	setosa
11	5.4	3.7	1.5	0.2	setosa

Showing 1 to 12 of 150 entries, 5 total columns

features

classes



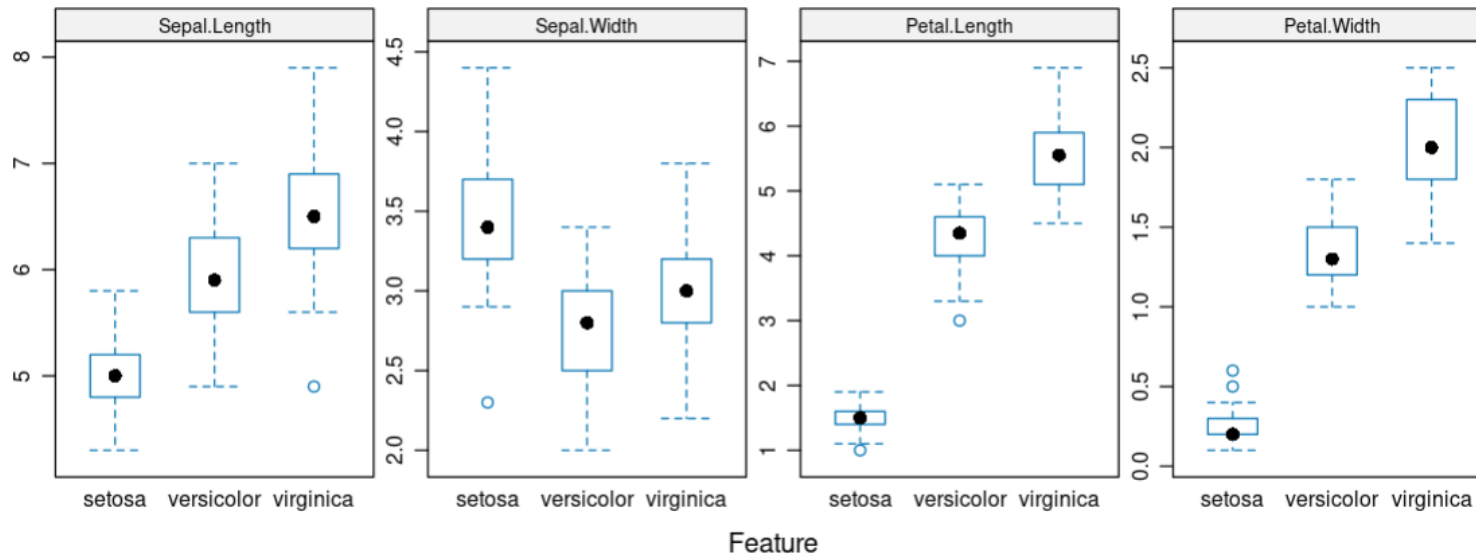
II - Visualisation

II - Visualisation des données

II.1 - Visualisation, statistiques

Data Exploration.R x stats x													
Filter													
	skim_type	skim_variable	Species	n_missing	complete_rate	numeric.mean	numeric.sd	numeric.p0	numeric.p25	numeric.p50	numeric.p75	numeric.p100	numeric.hist
1	numeric	Sepal.Length	setosa	0	1	5.006	0.3524897	4.3	4.800	5.00	5.200	5.8	
2	numeric	Sepal.Length	versicolor	0	1	5.936	0.5161711	4.9	5.600	5.90	6.300	7.0	
3	numeric	Sepal.Length	virginica	0	1	6.588	0.6358796	4.9	6.225	6.50	6.900	7.9	
4	numeric	Sepal.Width	setosa	0	1	3.428	0.3790644	2.3	3.200	3.40	3.675	4.4	
5	numeric	Sepal.Width	versicolor	0	1	2.770	0.3137983	2.0	2.525	2.80	3.000	3.4	
6	numeric	Sepal.Width	virginica	0	1	2.974	0.3224966	2.2	2.800	3.00	3.175	3.8	
7	numeric	Petal.Length	setosa	0	1	1.462	0.1736640	1.0	1.400	1.50	1.575	1.9	
8	numeric	Petal.Length	versicolor	0	1	4.260	0.4699110	3.0	4.000	4.35	4.600	5.1	
9	numeric	Petal.Length	virginica	0	1	5.552	0.5518947	4.5	5.100	5.55	5.875	6.9	
10	numeric	Petal.Width	setosa	0	1	0.246	0.1053856	0.1	0.200	0.20	0.300	0.6	
11	numeric	Petal.Width	versicolor	0	1	1.326	0.1977527	1.0	1.200	1.30	1.500	1.8	
12	numeric	Petal.Width	virginica	0	1	2.026	0.2746501	1.4	1.800	2.00	2.300	2.5	

II.2 - Visualisation, Distribution des valeurs





III -Data Science : SVM pour la Classification

III – Data Science

III.1 – Entrainement du jeu de données

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa
7	4.6	3.4	1.4	0.3	setosa
8	5.0	3.4	1.5	0.2	setosa
9	4.4	2.9	1.4	0.2	setosa
10	4.9	3.1	1.5	0.1	setosa
11	5.4	3.7	1.5	0.2	setosa

Showing 1 to 12 of 150 entries, 5 total columns

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa

Showing 1 to 5 of 120 entries, 5 total columns

80% du jeu de données total

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
6	5.4	3.9	1.7	0.4	setosa
10	4.9	3.1	1.5	0.1	setosa
12	4.8	3.4	1.6	0.2	setosa
30	4.7	3.2	1.6	0.2	setosa
31	4.8	3.1	1.6	0.2	setosa

Showing 1 to 5 of 30 entries, 5 total columns

20% du jeu de données total

III – Data Science

III.2 – Evaluation du modèle

Prediction	Reference		
	setosa	versicolor	virginica
setosa	40	0	0
versicolor	0	36	1
virginica	0	4	39

Matrice de confusion du jeu d'entraînement

Prediction	Reference		
	setosa	versicolor	virginica
setosa	10	0	0
versicolor	0	10	0
virginica	0	0	10

Matrice de confusion du jeu de test

Prediction	Reference		
	setosa	versicolor	virginica
setosa	40	0	0
versicolor	0	36	1
virginica	0	4	39

Matrice de confusion après validation croisée



BILAN

Résultats Globaux Satisfaisants :

Les différentes phases de développement du modèle SVM ont abouti à des résultats globalement satisfaisants.

Succès dans la Préparation des Données :

La phase de prétraitement des données a été efficace, incluant la suppression des valeurs aberrantes, la normalisation des caractéristiques et la gestion des données manquantes.

Légère Difficulté de Prédiction sur "Versicolor" :

Malgré les succès globaux, une légère difficulté de prédiction a été observée, en particulier concernant la classe "Versicolor".

Performances Acceptables sur les Autres Classes :

Le modèle a montré de bonnes performances de classification pour les autres classes, ce qui indique une capacité de généralisation satisfaisante.