

# Méthode de Classification de Mail et de Spam

## APPRENTISSAGE AUTOMATIQUE

Institut Francophone International

Master 2, Promotion 23

Option: Systèmes Intelligents et Multimédia (SIM)

**Présenté par le groupe : 11**

LAMAH Henry Kpakilé

OLEMBO Réel Devin Richmond

**Encadrant : Lê Hong Phuong**

Hanoi, le 12 novembre 2019



- 1 INTRODUCTION
- 2 PRESENTATION DES DIFFÉRENTS RÉSULTATS
- 3 CONCLUSION

## Objectifs visés

Compréhension et implémentation de différentes méthodes d'apprentissage supervisé :

- Machine à vecteurs de support (SVM) ;
- Naïve bayésienne ;
- Forêts d'arbres décisionnels ou Forêts aléatoires (Random forest).

# DataSet

Total = 962

Test = 260 soit 27%

Entrainement = 702 soit 73%

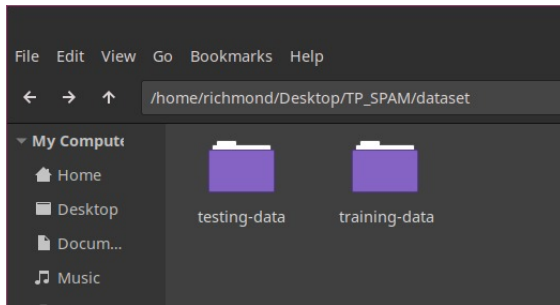


FIGURE – Structure du dataset

## Nomenclature des fichiers

SPAM = spmsgc80

MAIL = 9-612msg2

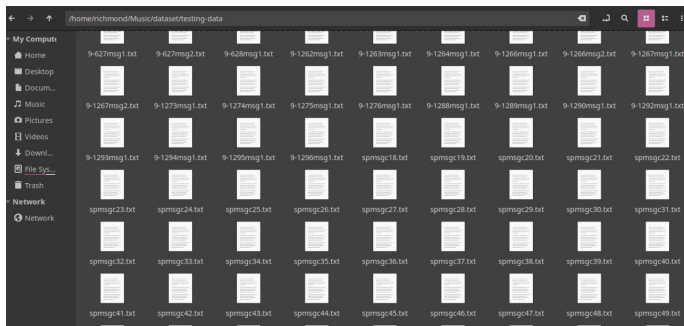


FIGURE – Structure du dataset

- 1 INTRODUCTION
- 2 PRESENTATION DES DIFFÉRENTS RÉSULTATS
- 3 CONCLUSION

# Machine à vecteurs de support (SVM)

## Matrix de confusion

Confusion matrix:

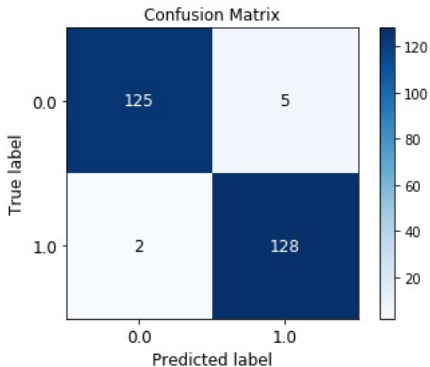


FIGURE – Matrice de confusion

# Machine à vecteurs de support (SVM) 2

Rapport de classification

	precision	recall	f1-score	support
0.0	0.98	0.96	0.97	130
1.0	0.96	0.98	0.97	130
accuracy			0.97	260
macro avg	0.97	0.97	0.97	260
weighted avg	0.97	0.97	0.97	260

FIGURE – Rapport SVM



# Naïve bayésienne

## Matrice de confusion

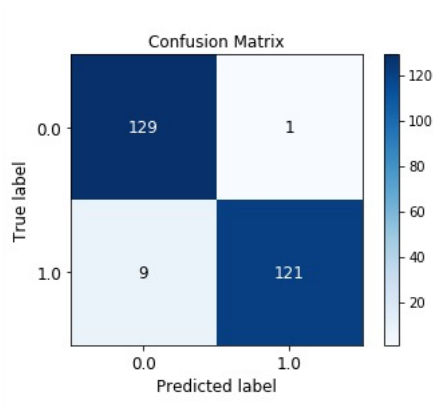


FIGURE – Matrice de confusion

# Naïve bayésienne 2

## Rapport de classification

	precision	recall	f1-score	support
0.0	0.93	0.99	0.96	130
1.0	0.99	0.93	0.96	130
accuracy			0.96	260
macro avg	0.96	0.96	0.96	260
weighted avg	0.96	0.96	0.96	260

FIGURE – Rapport NB

# Forêts d'arbres décisionnels

## Matrice de confusion

Confusion matrix:

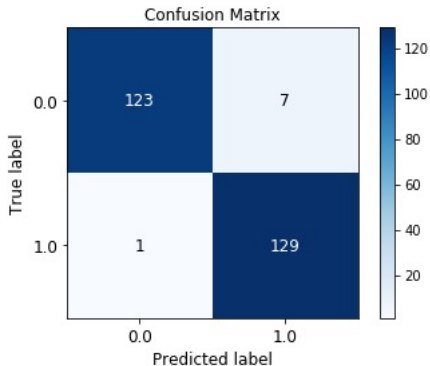


FIGURE – Matrice de confusion

# Forêts d'arbres décisionnels

## Rapport de classification

	precision	recall	f1-score	support
0.0	0.99	0.95	0.97	130
1.0	0.95	0.99	0.97	130
accuracy			0.97	260
macro avg	0.97	0.97	0.97	260
weighted avg	0.97	0.97	0.97	260

FIGURE – Rapport RF

- 1 INTRODUCTION
- 2 PRESENTATION DES DIFFÉRENTS RÉSULTATS
- 3 CONCLUSION

- Il n'existe pas de méthode d'apprentissage parfaite ;
- Les paramètres d'entrée peuvent influencer la qualité des modèles obtenus ;
- Existences de structures de données capable d'améliorer considérablement les résultats.

## Comparaison

Algorithmes	SVM	Random Forest	Naive Bayes
Accuracy	97,30%	95,76%	96,15%

# Référence & Note du groupe

Click on the link below for the dataset

<https://drive.google.com/open?id=1atnQB4whzj8SaUrce7AVXINgjFqahrUo>

Note 1 : 8 / 10 (LAMAH Henry Kpakilé)

Note 2 : 8,5 / 10 (OLEMBO Réel Devin Richmond)