

MODÈLE DE DÉTECTION DE SPAM

1. Introduction

On a choisi un exemple réel de classification de texte à l'aide de l'apprentissage automatique. **L'objectif de ce projet est de former un modèle d'apprentissage automatique de classification de texte en python capable de prédire si un message texte est un spam ou non.**

2 . Dataset

Ensemble de données de collecte de spam SMS » contient 5572 instances et deux attributs v1 et v2. Le v2 correspond aux messages d'entrée qui sont soit du spam, soit du non-spam. L'étiquette prédite v1 a deux classes : 0 = non-spam et 1 spam. Dans les données, 4 900 sont des échantillons de non-spam et 672 sont des échantillons de spam. L' ensemble de données est présenté dans le tableau .

Sample no number	V1	V2
1	Ham	Go until jurong point, crazy.. Available only in bugis <i>n</i> great world la e buffet... Cine there got amore wat...
2	Ham	Ok lar... Joking wif u oni...
3	Spam	Free entry in 2 a wkly comp to win FA cup final tks 21st May 2005. Text FA to 87121 to receive entry question (std txt rate) T&C's apply 08452810075over18's.
.	.	.
.	.	.
.	.	.
.	.	.
.	.	.
.	.	.
.	.	Rofl. Its true to its name
.	.	.
5572	Ham	.

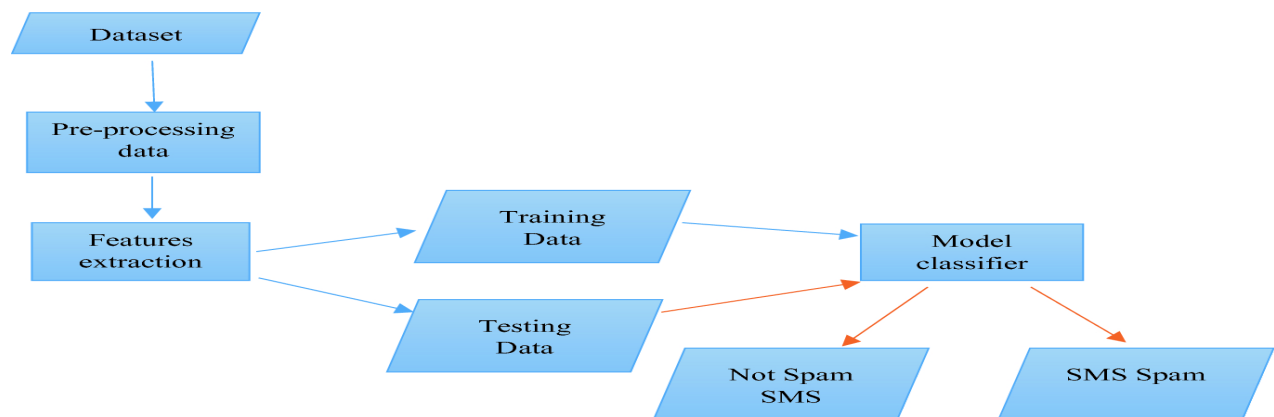
3. Résumé

La détection de spam est un gros problème dans la communication de messages mobiles en raison de laquelle la communication de messages mobiles n'est pas sécurisée. Afin de résoudre ce problème, une méthode précise et précise est nécessaire pour détecter le spam dans la communication par message mobile.

3.1 Problème d'apprentissage

L'objectif de ce projet est de former un modèle d'apprentissage automatique de classification de texte en python capable de prédire si un message texte est un spam ou non.

3.2 Schéma



Détection du spam SMS à l'aide de l'apprentissage automatique

