



PROJET SPAM CLASSIFIER



Analyse du besoin

1/ Contexte

2/ Objectif

3/Caractéristiques des Spams

4/Défis

5/Approche

1/Contexte

Le projet vise à créer un modèle capable de distinguer les SMS légitimes (ham) des SMS indésirables (spam). La base de données fournie contient 5 574 messages SMS en anglais, étiquetés comme ham ou spam.

2/Objectif

Développer un outil automatique qui identifie les spams dans les messages SMS en utilisant des techniques d'intelligence artificielle.

3/Caractéristiques des Spams

- > Utilisation de mots-clés comme "WIN", "FREE", "URGENT"**
- > Contiennent souvent des liens ou des numéros**
- > Souvent plus courts et rédigés pour attirer l'attention**

4/Défis

- > Déséquilibre des classes : Il y a plus de messages ham que spam**
- > Bruit dans les données : Les SMS peuvent contenir des abbréviations, emojis, ou fautes de frappe**
- > Représentation des textes : Transformer les messages en données compréhensibles pour un modèle**

S/Approche

Analyse des données :

- > Identifier la proportion de ham et de spam**
- > Comprendre les caractéristiques fréquentes des spams**

Préparation des données :

- > Nettoyer les messages (mettre en minuscules, supprimer les caractères inutiles...)**
- > Représenter les textes numériquement**

5/Approche(suivie)

Modélisation :

- >Tester plusieurs modèles (comme Régression Logistique, Random Forest, SVM)**
- >Optimiser les modèles pour de meilleures précisions**

Validation :

- >Utiliser des métriques comme le rappel, la précision et le F1-score**
- >Valider avec des techniques comme la validation croisée**

CLOSED