# COMMENT MENER A BIEN MON PROJET

## **Présentation**

Un problème classique dans le domaine de la reconnaissance de formes est celui de la reconnaissance de chiffres manuscrits

Le but de ce projet est de développer un modèle capable d'identifier correctement le chiffre (entre 0 et 9) écrit dans une image. Vous utiliserez pour cela l’algorithme SVM qui devrait correctement classifier les chiffres manuscrits de 0 à 9 en fonction des valeurs de pixels données comme caractéristiques. Chaque image contenant un chiffre est de 28 x 28 pixels et que chaque pixel forme une caractéristique, il y a 784 caractéristiques.

Avec un sous-échantillon de 10 à 20% des données d'entraînement, nous devrions nous attendre à obtenir une précision de plus de 90%.

Vous vous appuierez sur le dataset sur le lien suivant : <https://www.kaggle.com/c/digit-recognizer/data> Il s’appuie sur la base de données **MNIST** qui est une base de données de chiffres écrits à la main : <http://yann.lecun.com/exdb/mnist/>

Pour plus d’informations faites un tour sur wikipédia !

## **Les différentes étapes**

**Etape 1 : Organisez votre projet**

* Commencez par lister les différentes étapes de votre projet et répartissez-vous de façon équitable les tâches
* Organisez des points d’avancement au sein de votre équipe pour partager vos avancées et vos difficultés.

**Etape 2 : Décrivez votre jeu de données**

* Vérifiez les valeurs manquantes, la taille, le type ect…
* Identifiez les valeurs uniques et le nombre de données pour chaque valeur unique

**Etape 3 : Visualisez votre jeu de données**

* Affichez quelques chiffres
* Visualisez la matrice de corrélation de pixels. Est-ce normal ?

**Etape 4 : Préparez vos données**

* D’après vous est-il nécessaire de standardiser vos données ? Pourquoi ?

**Etape 5 : Construisez votre modèle**

* Essayez tout d’abord un modèle lineaire avec l’option linear kernel
* Puis essayez ‘poly’
* Et ‘rbf’

Utilisez dans un premier temps les paramètres par défaut.

**Etape 6 : Evaluez votre modèle**

En affichant la matrice de confusion, le classification report et la courbe ROC pour chacun des modèles que vous avez construit.

**Etape 7 : Choisissez le meilleur et optimisez les paramètres**

En utilisant GridsearchCV, déterminer les paramètres optimaux et réimplémenter le modèle. Quel score obtenez-vous ?

**Etape 8 : Optionnel**

Affichez quelques prédictions

## **CONCLUSION**

Rédigez votre rapport en établissant le contexte, en décrivant vos choix et méthode d’analyse et en explicitant le rôle de chacun dans ce projet. Soyez synthétique ! plus de 2-3 pages écrites et vos lecteurs ne le liront pas en entier