



ECOLE MAROCAINE DES
SCIENCES DE L'INGENIEUR

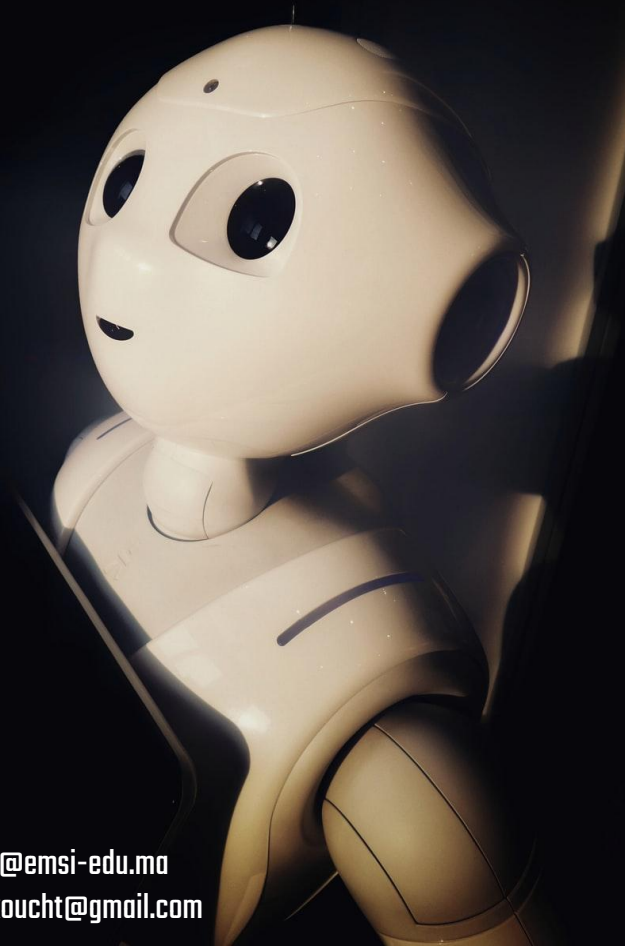
Membre de
HONORIS UNITED UNIVERSITIES

ML / DL

DATA PREPROCESSING

Aissam Outchakoucht

a.outchakoucht@emsi-edu.ma
aissam.outchakoucht@gmail.com



Les modèles de machine learning partent
du principe que les données sont parfaites;
la réalité est souvent bien différente !

DÉFIS

Dans le monde réel, les données sont :

- ❑ Incomplètes : manquent de certaines valeurs ou d'attributs importants, etc.
- ❑ Bruyantes : contiennent des erreurs ou des valeurs aberrantes, souvent dues à des erreurs humaines ou des capteurs défectueux.
- ❑ Incohérentes : manque de compatibilité ou de similarité entre deux faits ou plus, causé par des systèmes de collecte multiples.

ÉTAPES CLÉS DU PRÉTR. DES DONNÉES

- ❑ Intégration des données
- ❑ Nettoyage des données
- ❑ Transformation des données
- ❑ Réduction des données

TECHNIQUES DE TRANSF. DES DONNÉES

- ❑ Scaling (Mise à l'échelle: normalisation, standardisation)
- ❑ Encoding (Encodage - conversion des caractéristiques catégorielles en valeurs numériques)
- ❑ Feature engineering (Ingénierie des caractéristiques)
- ❑ Handling missing data (Gestion des données manquantes)
- ❑ Handling imbalanced data (... déséquilibrées)
- ❑ Dimensionality reduction (Réduction de la dimensionnalité)
- ❑ Embeddings (ex: texte)