

---

# PREDICTION DE QUALITE DE SOUDURES

---

Le but de ce projet est de prédire la qualité de soudures sur des aciers. Il s'agit d'une problématique d'intérêt pour de nombreux industriels dont le secteur pèse plusieurs milliards d'euros (exemple : soudure de pipe d'éoliennes). De nos jours, la connaissance liée à la qualité des soudures se transmet principalement d'expert à expert soudeur dont les industriels sont dépendants. Il y a un gros enjeu à acquérir de la connaissance via les données pour à la fois extraire et homogénéiser la connaissance experte mais également explorer de nouvelles connaissances via des patterns que l'on pourrait découvrir par l'exploration de la donnée.

---

## OBTENIR LES DONNEES

---

Les données publiques pour réaliser ce projet sont accessibles via le lien suivant : <https://www.phase-trans.msm.cam.ac.uk/map/data/materials/welddb-b.html>

Vous pouvez à présent utiliser ces données grâce à la librairie Python **Pandas**. Vous trouverez à ce [lien](#) un petit tutoriel pour vous aider à manipuler cette librairie. Vous êtes invité à utiliser toutes les librairies proposées dans les labs pour vous faciliter les tâches de pré-processing.

---

## OBJECTIFS DU PROJET

---

Les objectifs du projet sont les suivants :

- Faire une analyse descriptive de la base de données pour bien la maîtriser et identifier les tâches de pré-processing pertinentes à réaliser. Toutes action de pré-processing devra être explicitée, décrite et justifiée. Réfléchir aux unités de mesures des variables, à la nécessité de normaliser ou non les données, appliquer une ACP pour développer de l'intuition sur la problématique, etc.
- Identifier quelles sont les variables représentatives de la qualité de soudure et bien les comprendre. Réfléchir à la façon de prédire la qualité de soudure à partir de ces variables identifiées. Cela implique de réfléchir à une stratégie et il est attendu de bien l'expliciter et la justifier.
- Appliquer différentes approches de ML vues en cours et en TD sur ce dataset pour prédire la qualité de soudure. Pensez à réaliser un protocole de validation croisée rigoureux. Enfin, le dataset n'étant pas complètement labélisé, nous vous invitons à effectuer un travail bibliographique sur des méthodes de ML avancées de type semi-supervisées. Vous êtes invités à en appliquer au minimum une en justifiant votre choix et la pertinence de celle-ci.
- Faire une analyse comparative des performances des différents modèles et méthodes appliqués. Il est donc nécessaire de réfléchir aux choix des métriques pertinentes.
- Conclure sur l'approche la plus appropriée pour prédire la qualité de soudure.
- Après cette étude : quelles sont vos recommandations pour obtenir une bonne qualité de soudure ?
- Écrire un rapport d'environ 5 pages synthétisant le travail effectué et les résultats obtenus. Pensez à indiquer vos noms, le numéro de votre équipe, la date, vos mails et le titre sur la première page. Pensez aussi à mettre quelques figures illustratives. Il sera attendu un plan, une

section rappelant la problématique du projet et votre traduction de celle-ci en problème/tâche de ML. Une section décrivant la base de données est aussi attendue (résultat d'ACP etc.), votre compréhension des variables cibles pour prédire la qualité de soudure et comment est-ce que vous proposez de vous y prendre. Une section sur les méthodes ML que vous choisissiez d'appliquer avec vos justifications (pensez à les présenter synthétiquement, surtout si elles n'ont pas été vues en cours). Une section sur les résultats numériques sera à réaliser avec une étude comparative des performances. Les formules et/ou description des métriques utilisées devront être présentées. Enfin, pensez à ajouter une dernière section de discussion et conclusion avec un diagramme de GANNT contenant le partage en tâches, le temps que chaque tâche a pris et quel membre de l'équipe s'en est chargée. Bien entendu, une section sur les références bibliographiques clôturera le rapport. Pensez à bien citer vos références dans le corps de rapport.

- Remarques générales : un rapport doit être paginé, les figures et tableaux doivent avoir un titre et être cités dans le corps du rapport quand on y fait référence.