

# IA-Lyophilisation

## Romain CAPOCASALE

Travail de Bachelor 2020

Informatique – Développement Logiciel et Multimédia

Professeur: Cédric BILAT

Expert: Diego ANTOGNINI

### Description

Dans le cadre de son processus de lyophilisation, l'industrie Takeda récolte une multitude d'informations. Ces informations correspondent à un lot de travail et se présentent sous forme de variables physiques en fonction du temps. Ses paramètres sont notamment des valeurs de pression et de température récupérées via des capteurs.

Chaque lot de travail possède une valeur de rendement qui permet de quantifier la qualité de production du produit.

La quantité de données et de paramètres physiques induit un problème avec beaucoup de dimensions. Différentes techniques seront explorées dans le but de réduire le nombre de dimensions du problème.

Le but du projet consiste, à l'aide de différents outils informatiques tels que des techniques de visualisation, des statistiques descriptives et différentes techniques de *machine learning*, de constater si les paramètres physiques influencent le rendement ou non.

Dans un second temps, l'objectif est d'utiliser différents modèles de *machine learning* dans le but de prédire le rendement ou la classe du rendement d'un lot de travail.

### Déroulement

Principales phases du projet :

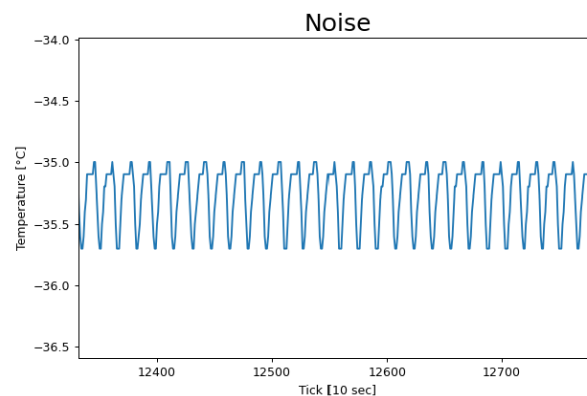
- Acquisition des données.
- Uniformisation de la composante temporelle et recentrage des courbes.
- Visualisation et statistiques descriptives.
- Suppression du bruit.
- Réduction des dimensions.
- Création d'une base de données.
- Classification non supervisée et *clustering*.
- Classification supervisée.
- Régression.
- *Machine learning* automatique.

### Problèmes

Premièrement, le nombre de lot de travail (122 pour la catégorie la plus grande) est insuffisant pour réaliser des statistiques descriptives et utiliser des techniques de *machine learning*.

Deuxièmement, il s'agit d'un problème avec beaucoup de dimensions. En effet, les 6 courbes utilisées pour le projet sont chacune composées d'environ 45'400 points, ce qui produit une matrice de caractéristiques avec beaucoup de colonnes.

La dernière difficulté du projet est le bruit sur les courbes dû à la régulation. Différents algorithmes ont été implémentés dans le but de lisser les différentes courbes.



Bruit sur les courbes

### Résultats

Comme le projet est confidentiel, les résultats obtenus ne peuvent pas être dévoilés. Uniquement les résultats non confidentiels sont indiqués ci-dessous.

Premièrement, une base de données a été mise en place pour faciliter la gestion et le chargement de la grande quantité de données.

Les différents outils et techniques utilisés n'ont pas pu montrer un impact des paramètres physiques à disposition pour le projet sur la variation du rendement.

Les algorithmes de *machine learning* utilisés n'ont pas réussi à classer les lots de travail dans deux groupes selon leur rendement ou encore de prédire cette valeur de rendement.