

Compte-rendu

Rencontre Aptar / INSA du 19/04/2018

Synthèse de la seconde réunion du 19 Avril 2018 entre les tuteurs Aptar (Franck Fabre, Stéphane Benard), l'encadrant INSA (Bruno Portier) et l'alternant (Antoine Barbosa). Points évoqués :

- **Description du travail effectué depuis début Mars au Vaudreuil, dans le cadre du déploiement d'un modèle prédictif prévu pour fin Avril sur le Val-de-Reuil** : analyse, compréhension et extraction des composantes d'intérêt d'un projet Machine Learning fait par une équipe de l'ESIGELEC, connecter en live ces éléments (4 modèles de classification) aux sources de données, évaluer la cohérence des résultats.
- **Définition des 2 axes d'étude de la thèse** : construction de deux modèles statistiques (mélange et vulcanisation), structures des matrices de données en entrée des modèles, étape de statistiques descriptive et de nettoyage pour déterminer la nature des données (corrélations, variabilité des données process).
- **Redéfinition du plan de la thèse** : inclure les différentes tâches d'extraction d'informations effectuées sur le projet de l'ESIGELEC, idée de comparer cette approche à celle à deux modèles.

Résumé

Le développement de la thèse professionnelle a été modifié dès le début de cette troisième session en entreprise. Un **projet analogue à celui de la thèse** (développement d'une solution Machine Learning pour la prédiction de conformité/non-conformité en phase de vulcanisation) a été mis en place par une **équipe de l'ESIGELEC**, et livré fin Février à l'équipe Business Intelligence du Vaudreuil sous la forme d'un dossier de travail « gitlab ». Une **première tâche** constituait à **comprendre leur approche de modélisation**, et en extraire les **4 modèles optimaux** (un par critère qualité prédit en vulcanisation) basés sur le **random forest**, et construits sur **deux types de données d'entrée** : le premier incluant **l'ensemble des données mélange/vulcanisation** (capteurs, et qualité mélange), le second associé seulement aux **données qualité mélange et capteurs vulcanisation**.

Parallèlement à cette mission d'extraction d'informations, l'objectif de fond de la thèse a été retravaillé. Cette dernière sera structurée autour de **deux modèles statistiques indépendants**, l'un pour l'étape de mélange, l'autre pour la vulcanisation, avec une approche davantage **exploratoire** :

- **Mélange** : l'idée est **d'étudier les liens fonctionnels entre les données capteurs** (numériques) **et 4 critères qualités mesurés en sortie de mélangeur** (ML, MH (Lbs), ts2, t90 (min)).
Remarque : une observation définit une **charge**, mais on dispose d'une **séquence de données capteurs** par charge. Pour le **nettoyage des données**, construire un premier modèle en décomposant la séquence en quatre statistiques descriptives (**moyenne, écart-type min, max**). Evaluer la qualité du modèle. En fonction des résultats, trouver une alternative à cette approche de décomposition adaptée à des **signaux dispersés propres au mélangeur**. Exemple : obtenir un vecteur de coefficients caractérisant les fluctuations du capteur sur la période de production d'une charge. Tester alors dans l'immédiat un modèle linéaire, random forest, pour premiers résultats.
- **Vulcanisation** : adopter le même objectif **de modélisation que pour le modèle mélange**, en intégrant l'ensemble des données capteurs (numériques), ainsi que les données qualités mélange mesurées sur les boîtes consommées, en entrée du modèle, pour la **prédiction de 4 PMC** (PMC = propriétés mécaniques, mesures numériques qualité en vulcanisation) en sortie. **Remarque** : Etude et prise en compte de la **corrélations entre les 4 PMC** → Recherche et sélection d'un **modèle multidimensionnel**, pour la **prédiction simultanée de plusieurs variables** en fonction d'une matrice de données d'entrée. Garder l'approche (**moyenne, écart-type min, max**) pour des **séquences de données capteurs plus tendancielles, lisses**, en entrées du modèle.

Concernant le **plan de la thèse**, il pourrait être judicieux de détailler dans une partie le travail effectué sur l'analyse du projet ESIGELEC, i.e expliquer leur **méthode de data cleaning**, résumer leur **approche de construction des modèles** en ajoutant des **indicateurs de performances**, etc. pour une comparaison future avec l'approche choisie pour la modèle vulcanisation de la thèse.

Planning et objectifs

- Modèle mélange : construire un premier modèle basé sur la décomposition des données d'entrée en (min, max, mean, stdev). L'évaluer, puis travailler sur une mise en forme différente des données capteurs pour comparer les résultats (ex : transformation de fourier, etc.).
- Modèle vulcanisation : étudier le niveau de corrélation des 4 PMC. Rechercher dans l'état de l'art des modèles multivariés répondant aux objectifs du modèle.
- Présentation à la codirection des 4 modèles de l'ESIGELEC le 25/04.