Télécom ParisTech

Projet de filière SR2I

# Détection d'anomalies de classification dans l'IoT via Machine Learning

*Antoine Urban, Yohan Chalier*

encadré par
Jean-Philippe Monteuuis
Houda Labiod

14 juin 2018

**Résumé**

# Chapitre 1

# Démarche et stratégie

## 1.1 Première implémentation

### 1.1.1 Objectif

En premier lieu, nous souhaitions commencer par une vision globale des données et du travail à effectuer. Nous disposions d'une base de données contenant des mesures de voiture, provenant de CarQuery, et contenant 54808 lignes complètes. Dans cette partie, nous allons nous efforcer d'obtenir une première fonction de classification se basant sur des critères très simple : des régions de décision rectangulaires et arbitraires.

### 1.1.2 Mise en œuvre

Puisque l'objectif de cette étude est la détection d'anomalies dans la mesure de longueur et de largeur, nous avons extrait les deux colonnes correspondantes dans une DataFrame du module Pandas, en Python.

Après un premier affichage des données, il est apparu que beaucoup de points apparaissaient en plusieurs fois, aussi la séparation de la base de données en points uniques et points non-uniques se révéla pertinente. Cela permit de réduire le nombre de lignes à 5026.

Manuellement, nous avons alors défini des zones simples (rectangulaires) en tant que régions de décision (Table 1.1.2). Ces zones ont été définies au jugé, afin d'encadrer le plus de points valides sans toutefois englober une zone de l'espace trop large.

| cadre | validité | intervalle de longueur | intervalle de largeur |
|---|---|---|---|
| vert | non-malicieux | 3 à 6,5 mètres | 1,4 à 2,4 mètres |
| gris | malicieux | 3 à 4,1 mètres | 2,05 à 2,4 mètres |
| gris | malicieux | 5,25 à 6,5 mètres | 1,4 à 1,65 mètres |

Table 1.1 – Dimensions des régions de décision arbitraires

Hors de la zone verte, et dans les deux cadres gris, nous avons alors généré aléatoirement 700 points définis comme malicieux. La figure 1.1.2 représente l'affichage de tous les points décrits plus tôt ainsi que des régions de décision.

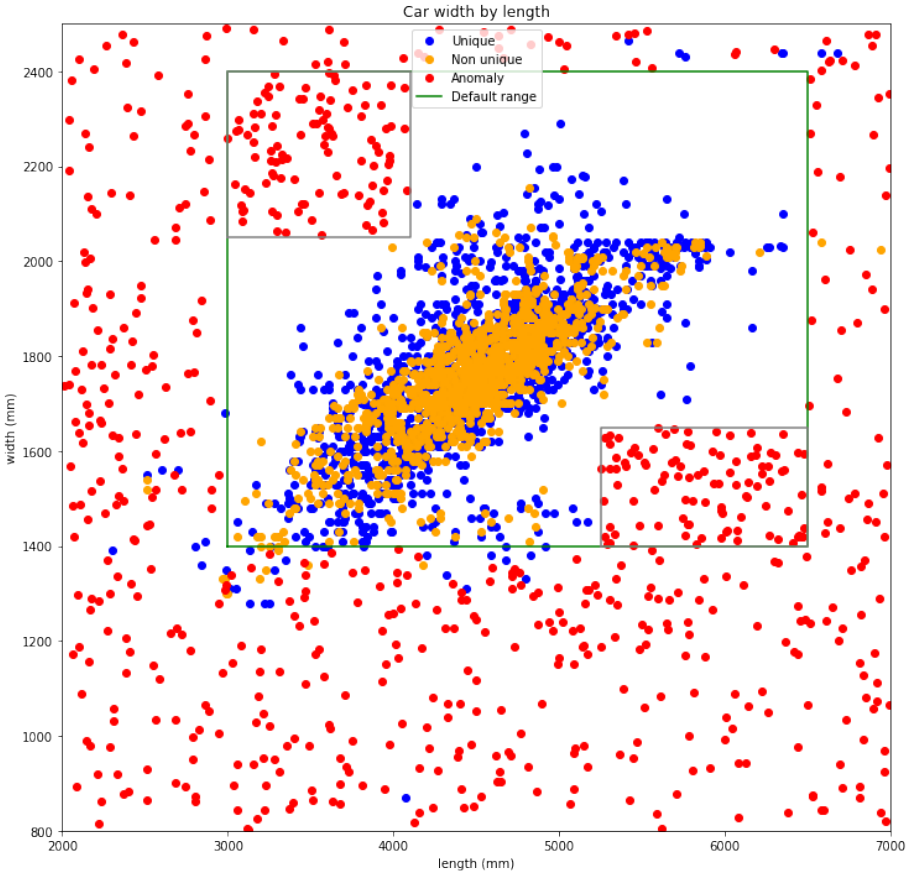


FIGURE 1.1 – Régions de décision manuelles pour des dimensions de voitures

Ainsi faite, notre classification possède, sur le jeu d'entraînement, une précision de 97,57%.

## 1.2 Recherche des bases de données

## 1.3 Environnement de travail

## 1.4 Méthode d'évaluation