



# **CAHIER DES CHARGES**

## **FONCTIONNEL ET TECHNIQUE**

---

**Sujet de PFE :**

**Système Intelligent de Détection de Fraude  
et de Valorisation Douanière (Auto & Tech)**

**Réalisé par :** Mohammed Amine HAMOUTTI

**Encadré par :** Yassine AMMAMI

# Table des Matières

## Table des matières

<b>1 Contexte et Problématique</b>	<b>2</b>
1.1 Contexte : L'Enjeu du Contrôle Douanier	2
1.2 Problématique	2
1.3 Solution Proposée	2
<b>2 Objectifs du Projet</b>	<b>2</b>
<b>3 Périmètre Fonctionnel (Scope)</b>	<b>2</b>
3.1 Module 1 : Véhicules (Cars)	2
3.2 Module 2 : Produits High-Tech (Phones & Laptops)	3
<b>4 Spécifications Techniques et Méthodologie</b>	<b>3</b>
4.1 1. Architecture des Données (Data Engineering)	3
4.2 2. Moteur d'Estimation (Machine Learning)	3
4.3 3. Visualisation et Interface (Power BI)	3
<b>5 Stack Technologique</b>	<b>4</b>
<b>6 Livrables Attendus</b>	<b>4</b>

# 1 Contexte et Problématique

## 1.1 Contexte : L'Enjeu du Contrôle Douanier

Le contrôle de la valeur transactionnelle des marchandises importées (véhicules d'occasion, matériel informatique) est un défi majeur pour l'administration douanière. La sous-facturation (déclaration d'une valeur inférieure à la réalité) entraîne une perte fiscale significative pour l'État.

## 1.2 Problématique

Les inspecteurs manquent d'outils automatisés pour vérifier instantanément si la valeur déclarée par un importateur correspond à la réalité du marché marocain. Les bases de données statiques (type Argus) sont souvent obsolètes face à la volatilité des prix (ex : inflation des véhicules d'occasion, décote rapide des smartphones).

## 1.3 Solution Proposée

Développement d'une plateforme d'**Audit Automatisé** basée sur l'Intelligence Artificielle. Le système compare la *Valeur Déclarée* par l'importateur à une *Valeur de Référence* prédite par un modèle de Machine Learning, afin de lever des alertes de fraude en temps réel.

# 2 Objectifs du Projet

- **Détection d'Anomalies** : Identifier automatiquement les dossiers suspects présentant un écart injustifié (>20%) entre le prix déclaré et le prix marché.
- **Valorisation Dynamique** : Fournir une estimation "Juste Prix" (Fair Value) basée sur les données réelles du marché marocain (Scraping 2025 + Historique).
- **Aide à la Décision** : Visualiser le risque fiscal via un tableau de bord décisionnel (Power BI).

# 3 Périmètre Fonctionnel (Scope)

Le système traitera trois catégories de produits à fort risque de fraude :

## 3.1 Module 1 : Véhicules (Cars)

- **Variables d'Entrée** : Marque, Modèle, Année de Mise en Circulation, Kilométrage, Motorisation (Diesel/Essence).
- **Spécificité** : Prise en compte de la dépréciation réelle et de la surcote des modèles populaires (ex : Dacia, Renault).

### 3.2 Module 2 : Produits High-Tech (Phones & Laptops)

- **Smartphones** : Analyse par Marque, Modèle et Capacité (Stockage).
- **Laptops** : Analyse par Processeur, RAM et Type de Disque (SSD).
- **Spécificité** : Gestion des configurations multiples pour un même modèle.

## 4 Spécifications Techniques et Méthodologie

### 4.1 1. Architecture des Données (Data Engineering)

Le projet repose sur une approche hybride garantissant la représentativité du marché :

- **Données Réelles (Ground Truth)** :
  - *Véhicules* : Consolidation du dataset MUCars (+100k transactions historiques) et de données web scrapées en 2025 (Moteur.ma).
  - *Tech* : Utilisation de datasets standards nettoyés et convertis en MAD.
- **Simulation de Contrôle (Data Augmentation)** : En l'absence d'accès aux fichiers confidentiels de la Douane, nous générons un jeu de données de test par **Simulation de Monte Carlo** :
  - 85% de déclarations conformes (Prix  $\approx$  Marché).
  - 15% de tentatives de sous-facturation (Prix  $\ll$  Marché) pour entraîner / tester la détection.

### 4.2 2. Moteur d'Estimation (Machine Learning)

- **Algorithme** : *Random Forest Regressor*. Ce choix se justifie par sa capacité à gérer des données hétérogènes (texte + chiffres) et sa robustesse face aux valeurs aberrantes.
- **Traitement du Langage (NLP)** : Utilisation de *TF-IDF* pour transformer les descriptions textuelles (Marque/Modèle) en vecteurs mathématiques exploitables.

### 4.3 3. Visualisation et Interface (Power BI)

L'interface finale sera un Dashboard Power BI intégrant nativement le script Python de prédiction :

- **Indicateur de Risque (Gauge Chart)** : Visualisation immédiate du "Tax Gap" (Manque à gagner fiscal).
- **Cartographie des Importations** : Analyse des flux par origine et par port d'entrée.
- **Statut des Dossiers** : Répartition (Conforme / Suspect / Bloqué).

## 5 Stack Technologique

- **Langage** : Python 3.12 / 3.14
- **Bibliothèques Clés** : Pandas (ETL), Scikit-learn (IA), Joblib (Sérialisation), BeautifulSoup (Scraping).
- **Business Intelligence** : Microsoft Power BI (pour la restitution graphique et l'exécution du modèle).

## 6 Livrables Attendus

1. **Modèles Prédicatifs (.pkl)** : Trois moteurs d'IA entraînés et validés (Auto, Phone, PC).
2. **Datasets Consolidés** : Fichiers CSV propres (Marché Réel + Simulation Fraude).
3. **Dashboard Décisionnel** : Rapport Power BI interactif (.pbix).
4. **Rapport Technique** : Documentant la méthodologie de simulation et les performances du modèle.