



**Propositions de sujets pour l'application de l'IA  
au contrôle fiscal  
Domaine impôt**

Janvier 2025

## **1. Détection des anomalies fiscales à l'aide de l'intelligence artificielle**

Développer un modèle d'apprentissage automatique capable de détecter des anomalies dans les déclarations fiscales, telles que des incohérences ou des comportements suspects indiquant une fraude fiscale.

### **Objectifs :**

- Analyser les données fiscales historiques pour identifier des modèles de fraude.
- Mettre en place un système d'alerte pour signaler les anomalies en temps réel.

### **Livrables :**

- Base de données d'entraînement contenant des données anonymisées.
- Modèle d'IA de détection d'anomalies.
- Tableau de bord présentant les alertes et leurs niveaux de criticité.

### **Technologies suggérées :**

- Langages : Python (TensorFlow, PyTorch).
- Bibliothèques : Scikit-learn, Pandas.
- Visualisation : Tableau, Power BI.

### **Contraintes :**

- Garantir la protection des données sensibles.
- Prévoir un taux d'erreur (faux positifs/faux négatifs) acceptable.

## 2. Prédiction des retards de paiement des contribuables

Créer un modèle prédictif pour anticiper les retards ou défauts de paiement des contribuables, basé sur l'historique des transactions et des comportements.

### Objectifs :

- Développer un modèle prédictif basé sur des données passées.
- Établir un scoring de risque pour chaque contribuable.

### Livrables :

- Modèle de prédiction des retards.
- Rapport d'analyse des facteurs influençant les retards.
- Interface de visualisation des prédictions.

### Technologies suggérées :

- Langages : Python ou R.
- Outils : XGBoost, LightGBM.
- Infrastructure : Cloud (AWS, Azure) si besoin d'entraînement lourd.

### Contraintes :

- Valider le modèle avec des données tests réalistes.
- Minimiser les biais discriminatoires liés au modèle.

### 3. Classification automatique des dossiers fiscaux pour un contrôle ciblé

Développer un outil de classification des dossiers fiscaux, permettant de prioriser les contrôles en fonction du risque identifié.

#### **Objectifs :**

- Utiliser des algorithmes de classification supervisée pour classer les dossiers en différentes catégories de risque.
- Proposer une stratégie d'audit basée sur les priorités générées.

#### **Livrables :**

- Modèle de classification avec un algorithme supervisé.
- Interface d'intégration avec le système SIGTAS pour une exploitation directe.

#### **Technologies suggérées :**

- Modèles : Random Forest, SVM, Réseaux neuronaux.
- Intégration : API REST.

#### **Contraintes :**

- Privilégier des modèles interprétables pour justifier les décisions de classification.
- Garantir la compatibilité avec le système actuel.

#### **4. Chatbot intelligent pour l'assistance en matière de fiscalité**

Concevoir un chatbot basé sur le traitement du langage naturel (NLP) pour répondre aux questions fréquentes des contribuables et assister les agents fiscaux.

##### **Objectifs :**

- Développer un chatbot capable de traiter des requêtes fiscales courantes.
- Permettre l'escalade des cas complexes vers un agent fiscal.

##### **Livrables :**

- Modèle NLP entraîné (GPT, BERT).
- Interface utilisateur (chat en ligne ou application mobile).
- Documentation pour l'entraînement futur du modèle.

##### **Technologies suggérées :**

- Modèles : Hugging Face Transformers, OpenAI API.
- Plateformes : Dialogflow, Rasa.

##### **Contraintes :**

- S'assurer que le chatbot respecte les lois fiscales locales.
- Prévoir des mises à jour régulières en fonction des changements législatifs.

## 5. Analyse des schémas de fraude fiscale avec l'analyse de graphes

Utiliser des algorithmes d'analyse de graphes pour identifier les relations suspectes entre contribuables et entreprises.

Objectifs :

- Représenter les interactions fiscales sous forme de graphe.
- Identifier des réseaux suspects de fraude ou de blanchiment.

**Livrables :**

- Graphe interactif des relations fiscales.
- Rapport sur les clusters identifiés comme à risque.

**Technologies suggérées :**

- Outils : Neo4j, NetworkX.
- Modèles : Graph Neural Networks (GNN).

**Contraintes :**

- Garantir une bonne visualisation pour interpréter les résultats.
- Définir clairement les règles d'identification des anomalies.