AI for Tax Fraud Detection

Overview

AI can detect tax fraud even without large labeled datasets by leveraging **unsupervised learning**, **rule-based methods**, and **natural language processing (NLP)**. These approaches help surface suspicious behavior and prioritize cases for further investigation.

1. Unsupervised Learning

Unsupervised methods are effective where confirmed fraud labels are unavailable.

Techniques

- **Clustering (K-Means, DBSCAN)**: Group taxpayers and detect outliers.
- **Isolation Forest, One-Class SVM**: Model normal behavior to flag anomalies.
- **Autoencoders**: Identify records with high reconstruction error.

Use Cases

- Identify inconsistent income versus deduction trends.
- Detect complex returns from first-time filers.
- Find seasonal deviations from industry norms.

2. Rule-Based Detection

Incorporating domain knowledge ensures transparency and robustness.

Example Rules

- **Benford's Law**: Identify unnatural digit distributions.
- **Threshold Checks**: Flag extreme values compared to industry benchmarks.
- **Co-occurrence Rules**: Detect missing logical claim combinations.

Benefits

- Easy to interpret and implement.
- Effective for pre-filtering and hybrid systems.

3. NLP for Textual Evidence

NLP enables the analysis of unstructured text from emails, documents, and social media.

Techniques

- **TF-IDF + Clustering**: Group unusual phrasing patterns.
- **Transformer Models (e.g., BERT)**: Classify risk in communications.
- **Named Entity Recognition (NER)**: Extract and validate entities in receipts or filings.

Use Cases

- Detect evasive or manipulative language in correspondence.
- Flag suspicious or fabricated documents.
- Monitor online platforms for illicit tax avoidance schemes.

4. Feature Engineering: The Key to Success

Strong features are critical to model performance and fraud detection accuracy.

Why It Matters

- Fraudulent behavior mimics legitimate activity.
- Effective features enhance both ML and rule-based detection systems.

Suggested Feature Types and Examples

Feature Type	Example Features
Behavioral	Filing frequency, YoY deduction change, amendment timing
Relational	Similarity to peer filings, shared IPs or addresses
Numerical Patterns	Benford's distribution, excessive rounding (.00 values)
Textual/NLP	Keyword flags, sentiment shifts, entity extraction
Derived Ratios	Expense-to-income, donations per income, tax paid/income