

# Documentation ML Features

provider\_ml\_features - Detection d'Anomalies pour la Fraude Sante

## Comment ca marche ? (Version Simple)

Imagine que tu es un inspecteur des fraudes et tu dois surveiller **8.9 millions de medecins**. Impossible de tous les verifier un par un ! L'idee : on compare chaque medecin a ses "collegues similaires" (meme specialite, meme etat). Si un medecin est tres different des autres, c'est suspect.

### 1. Z-Scores (Comparaison aux pairs)

C'est comme les notes a l'ecole : si tout le monde a entre 10 et 15, et qu'un eleve a 25... c'est bizarre. On regarde les paiements recus, cout des prescriptions, % de medicaments de marque.

**Exemple :** Dr. Martin recoit 500 000\$ de labos pharma. Ses collegues cardiologues en Californie recoivent en moyenne 50 000\$. → Z-score = 4.5 → **Alerte !**

### 2. Concentration HHI (Diversite des prescriptions)

Un medecin normal prescrit plein de medicaments differents. Un "pill mill" (usine a pilules) prescrit toujours les memes 2-3 medicaments.

**Exemple :** Dr. Dupont prescrit 80% d'Oxycodone (opioide). Les autres medecins de sa specialite prescrivent 50+ medicaments differents. → HHI = 6400 → **Suspect !**

### 3. Diversite Pharma (Relations avec les labos)

Un medecin qui recoit de l'argent de 10 labos differents → normal. Un medecin qui recoit 50 000\$ d'un SEUL labo → suspect (potentiel pot-de-vin).

**Exemple :** Dr. Bernard recoit 75 000\$ uniquement de Pfizer, aucun autre labo. → **Relation exclusive suspecte**

### 4. Percentiles (Classement)

On classe tous les medecins du moins au plus. Si tu es dans le top 10% pour les paiements recus → a surveiller.

**Exemple :** Dr. Petit est au 98e percentile pour le cout de ses prescriptions. → Il prescrit plus cher que 98% de ses pairs → **A investiguer**

## Le Resultat : Compteur de Flags (0-6)

#### 0 Flags

Medecin normal

#### 1 Flag

A surveiller

#### 2-3 Flags

Investigation

#### 4+ Flags

Prioritaire

**Resultat :** Sur 8.9 millions de medecins analyses, **53 000** ont au moins 2 flags d'anomalie et meritent une investigation approfondie.

# Details Techniques

25+

Features

6

Flags Anomalie

4

Dimensions

8.9M

Providers

## 1. Z-Scores (Comparaison aux Pairs)

$$Z = (X - \mu) / \sigma$$

Seuil d'anomalie :  $z > 2$  = Outlier statistique (top 2.5%)

Feature	Description
PAYMENT_ZSCORE	Z-score des paiements pharma vs pairs
RX_COST_ZSCORE	Z-score du cout des prescriptions vs pairs
BRAND_PCT_ZSCORE	Z-score du % médicaments de marque vs pairs
CLAIMS_ZSCORE	Z-score du nombre de claims vs pairs

### Interpretation :

$z = 0$  : Exactement a la moyenne  
 $z = 1$  : Top 16% des pairs  
 $z = 2$  : Top 2.5% (seuil outlier)  
 $z = 3$  : Top 0.1% (tres suspect)

## 2. Concentration HHI (Médicaments)

$$HHI = \sum (part_{marche})^2$$

Valeur HHI	Interpretation
0 - 1500	Prescripteur diversifie
1500 - 2500	Moderement concentre
> 2500	Tres concentre (potentiel pill mill)

Feature	Description
DRUG_CONCENTRATION_HHI	Indice HHI (0-10000)
TOP_DRUG_PCT	% du médicament #1
TOP_3_DRUGS_PCT	% des 3 top médicaments
IS_CONCENTRATED_PRESCRIBER	HHI > 2500

### 3. Diversite Pharma

Feature	Description
UNIQUE_PHARMA_COMPANIES	Nombre de compagnies payeuses
TOP_PAYER_PCT	% du top payeur
PCT_GENERAL_PAYMENTS	% paiements generaux (repas, voyages)
PCT_RESEARCH_PAYMENTS	% paiements recherche
IS_SINGLE_PAYER_RECIPIENT	1 seule compagnie + >10k\$

**Risque :** Un provider recevant > 10 000\$ d'une seule compagnie pharma est suspect (relation exclusive, potentiel kickback).

**Note :** Les paiements "General" (repas, voyages, honoraires) sont plus suspects que les paiements "Research".

### 4. Classement Percentiles

Feature	Description
PAYMENT_PERCENTILE	Rang des paiements pharma
RX_COST_PERCENTILE	Rang du cout des prescriptions
BRAND_PCT_PERCENTILE	Rang du % de marques
CLAIMS_PERCENTILE	Rang du volume de claims

**Interpretation :**

0 - 75 Normal - Dans la majorite des pairs

75 - 90 Eleve - A surveiller

> 90 Extreme - Investigation recommandee

### 5. Flags d'Anomalie & Features Cles

PAYMENT\_ZSCORE > 2

RX\_COST\_ZSCORE > 2

BRAND\_PCT\_ZSCORE > 2

DRUG\_HHI > 2500

Pharma Unique (10k\$+)

Est Exclu (LEIE)

Feature	Type	Description
ANOMALY_FLAG_COUNT	Int (0-6)	Nombre total de flags
IS_MULTI_DIMENSION_OUTLIER	Bool	Outlier dans 2+ z-scores
PAYMENT_TO_RX_RATIO	Float	Paiements / Cout Rx
PAYMENT_VS_PEER_RATIO	Float	Paiements / Moyenne pairs

**Groupement de Pairs :**

**Groupes :** SPECIALTY + STATE  
**Minimum :** 5 pairs requis  
**NULL :** Groupees comme 'UNKNOWN'

**Dashboard :** Z-Scores (radar), Concentration (HHI), Percentiles (bar chart)