

Documentation ML Features

provider_ml_features - Detection d'Anomalies pour la Fraude Sante

25+

Features

6

Flags Anomalie

4

Dimensions

8.9M

Providers

Modele de feature engineering pour detection de fraude sans labels. Compare chaque provider a ses pairs (meme specialite + etat) pour identifier les outliers statistiques. Les providers avec plusieurs flags d'anomalie meritent une investigation approfondie.

1. Z-Scores (Comparaison aux Pairs)

Mesure combien d'ecarts-types un provider est eloigne de la moyenne de son groupe de pairs.

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

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

Feature	Description
PAYOUT_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 % medicaments 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 (Medicaments)

L'indice Herfindahl-Hirschman mesure la concentration des prescriptions sur quelques medicaments.

$$HHI = \sum (part_marche)^2$$

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 medicament #1
TOP_3_DRUGS_PCT	% des 3 top medicaments
IS_CONCENTRATED_PRESCRIBER	HHI > 2500

3. Diversité Pharma

Analyse les relations financières entre le provider et les compagnies pharmaceutiques.

Feature	Description
UNIQUE_PHARMA_COMPANIES	Nombre de compagnies payeuses
TOP_PAYER_PCT	% du top payeur
PCT_GENERAL_PAYMENTS	% paiements généraux (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

Position relative du provider par rapport à ses pairs (0-100). Permet d'identifier rapidement les extrêmes.

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

Interpretation :

0 - 75 Normal - Dans la majorité des pairs

75 - 90 Elevé - A surveiller

> 90 Extrême - Investigation recommandée

5. Flags d'Anomalie (0-6)

Compteur aggrege de tous les signaux d'alerte. Un provider avec plusieurs flags merite une investigation prioritaire.

PAYMENT_ZSCORE > 2

RX_COST_ZSCORE > 2

BRAND_PCT_ZSCORE > 2

DRUG_HHI > 2500

Pharma Unique (10k\$+)

Est Exclu (LEIE)

0 Flags

Provider normal

1 Flag

A surveiller

2-3 Flags

Investigation recommandee

4+ Flags

Investigation prioritaire

6. Features Cles

Feature	Type	Description
ANOMALY_FLAG_COUNT	Int (0-6)	Nombre total de flags
IS_MULTI_DIMENSION_OUTLIER	Bool	Outlier dans 2+ z-scores
PAYOUT_TO_RX_RATIO	Float	Paiements / Cout Rx
PAYOUT_VS_PEER_RATIO	Float	Paiements / Moyenne pairs
AVG_COST_PER_CLAIM	Float	Cout moyen par claim

7. Groupement de Pairs

Groupes : SPECIALTY + STATE

Minimum : 5 pairs requis pour statistiques significatives

NULL : Specialites NULL groupées comme 'UNKNOWN'

8. Integration Dashboard

Onglet	Contenu	Utilite
Z-Scores	Radar chart	Comparaison visuelle vs pairs
Concentration	HHI, diversité	Détection pill mills
Percentiles	Bar chart	Classement rapide

Bandeaux d'Alerte

Vert : 0 flags - Aucune anomalie

Orange : 1-2 flags - A surveiller

Rouge : 3+ flags ou multi-outlier