

Dataset Optimisé pour le Hackathon Fraude Bancaire

Ce dossier contient un dataset de détection de fraude bancaire optimisé pour le hackathon.

Caractéristiques du dataset

- **Année:** Transactions de 2016 à 2018
- **Taille totale:** 210000 transactions
- **Taux de fraude global:** ~0.15%
- **Défi supplémentaire:** Les clients qui ont commis des fraudes dans l'ensemble d'entraînement sont différents de ceux de l'ensemble d'évaluation

Fichiers pour les participants

- **transactions_train.csv:** Les transactions pour l'entraînement.
- **train_fraud_labels.json:** Les labels pour les transactions d'entraînement.
- **cards_data.csv:** Informations sur les cartes de paiement.
- **users_data.csv:** Informations sur les utilisateurs.
- **mcc_codes.json:** Codes MCC et descriptions.
- **evaluation_features.csv:** Les transactions pour l'évaluation (sans labels). (Attention ce fichier sert seulement pour la phase d'évaluation de vos modèles ça ne sert à rien de l'utiliser pour l'entraînement)

Format des soumissions

Les participants doivent soumettre un fichier CSV contenant les colonnes suivantes:

- `transaction_id`: Identifiant de la transaction (comme dans `evaluation_features.csv`)
- `fraud_prediction`: Prédiction binaire (1 pour fraude, 0 pour non-fraude)

Note importante

Ce dataset présente un défi réaliste de détection de fraude bancaire où les modèles doivent généraliser à de nouveaux clients. Les participants devront développer des algorithmes capables de détecter des fraudes pour des clients jamais rencontrés durant l'entraînement.