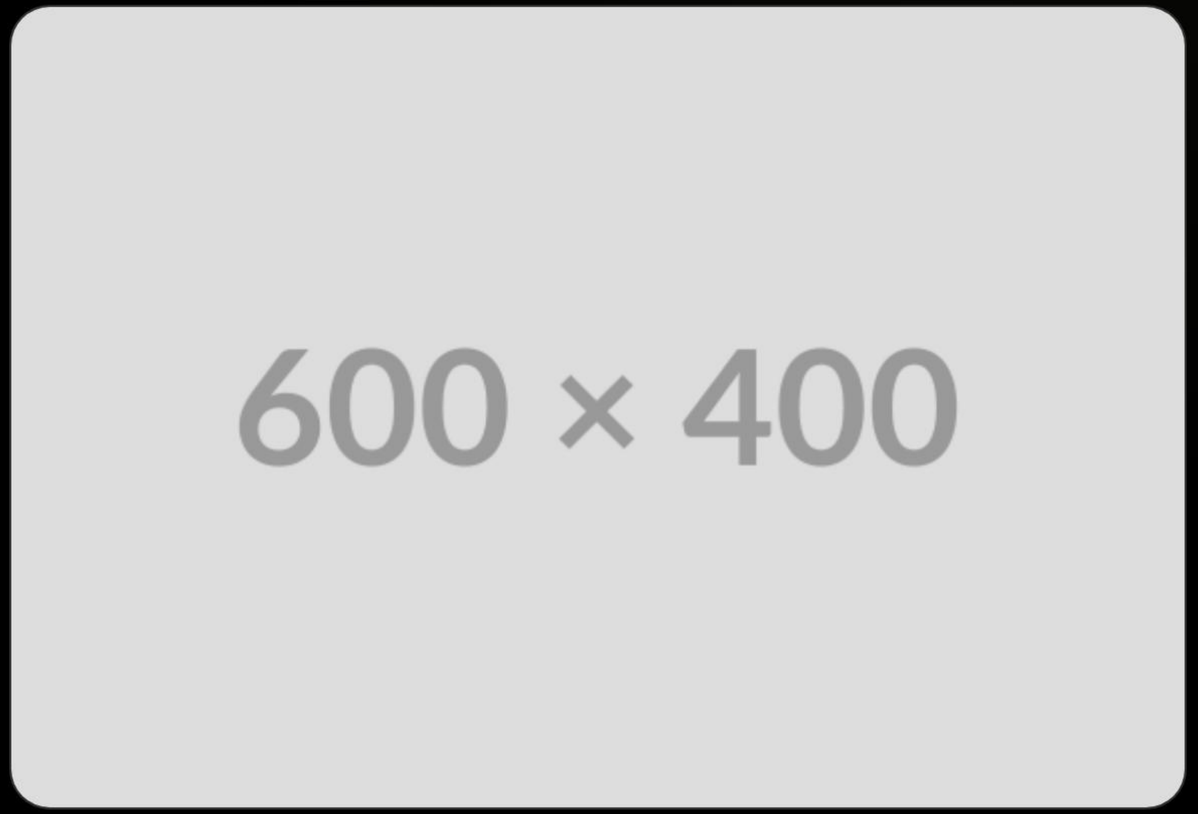
Détection de Fraude par Machine Learning

Conception d'un système intelligent basé sur l'algorithme Random Forest pour sécuriser les paiements mobiles.

Le Défi : Sécuriser les Transactions

- **Objectif :** Identifier automatiquement les transactions frauduleuses parmi des milliers de paiements légitimes.
- **Enjeu :** Minimiser les pertes financières (Faux Négatifs) sans bloquer les clients honnêtes (Faux Positifs).
- **Approche :** Apprentissage supervisé sur un historique de 2000 transactions.



Analyse des Données

Amount

Le montant de la transaction. Variable clé soupçonnée d'être discriminante.

Hour & Freq.

L'heure du paiement et le nombre de transactions effectuées le même jour.

Context

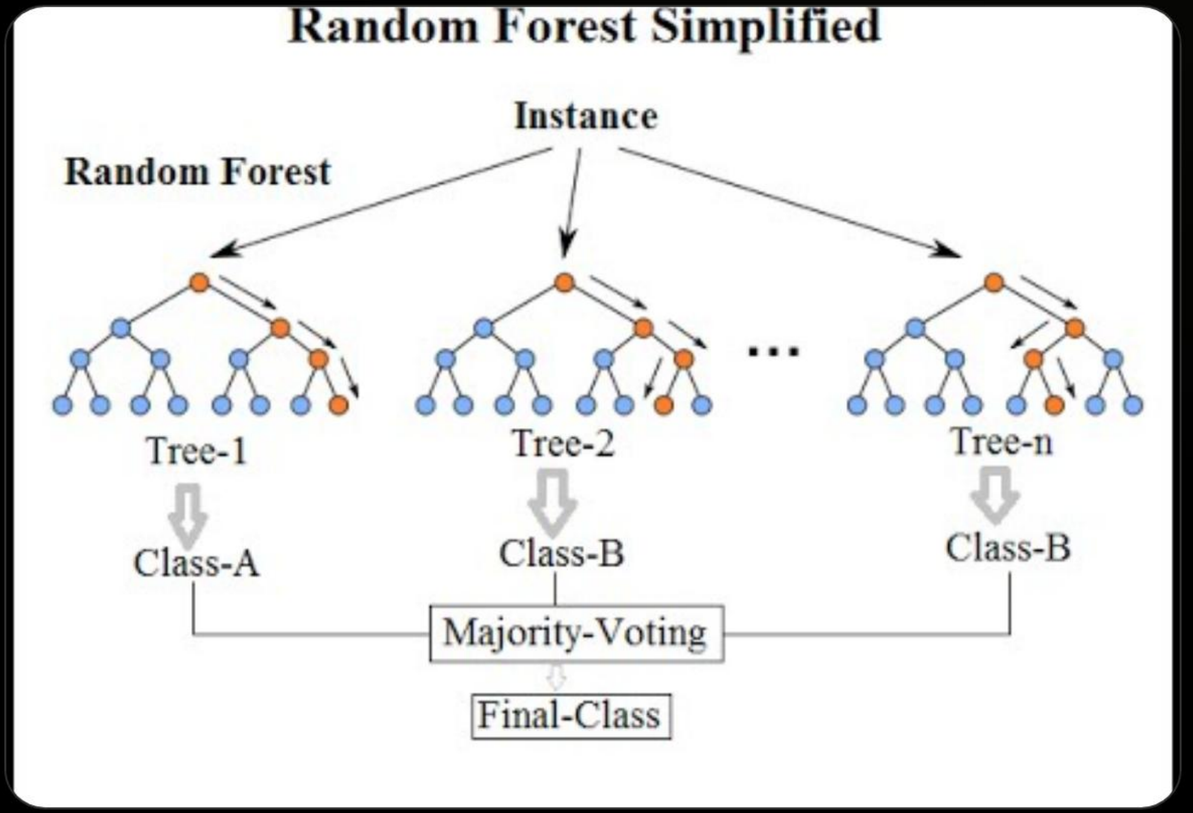
Changement soudain de lieu (Location Change) ou d'appareil (Device Change).

Méthodologie : Random Forest

Pourquoi ce choix ?

Nous avons opté pour une **Forêt Aléatoire** composée de 100 arbres de décision.

- **Robustesse :** Réduit le risque de surapprentissage (Overfitting) par rapport à un arbre unique.
- **Polyvalence :** Gère parfaitement les données mixtes (montants numériques et indicateurs binaires 0/1).
- **Interprétabilité :** Permet de mesurer l'importance de chaque variable.



Performance du Modèle

91%

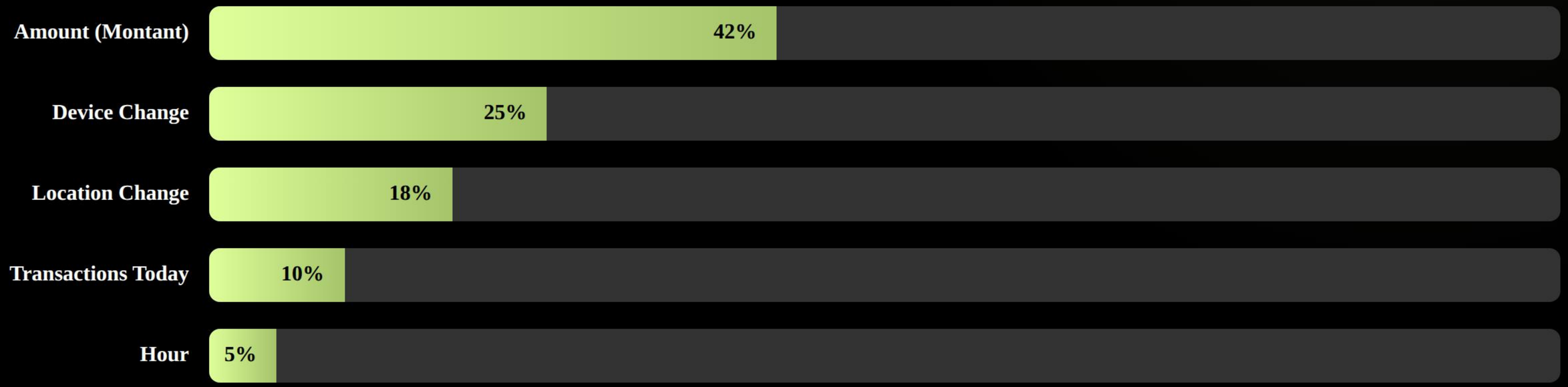
Rappel (Recall)

100

Arbres de Décision

Le modèle détecte plus de 9 fraudes sur 10.
La priorité a été donnée à la détection des pertes (minimisation des Faux Négatifs).

Facteurs Clés de Détection



Le montant est le facteur déterminant majeur, suivi par les changements techniques.

Analyse : Succès

Le Cas "Flagrant Délit" (Index 1860)

Le modèle a correctement intercepté cette fraude grâce à des signaux forts :

- **Montant :** 12 450 € (Anormalement élevé).
- **Contexte :** Changement simultané d'appareil ET de lieu.
- **Verdict :** Probabilité de fraude > 90%.

Le modèle fonctionne parfaitement sur les fraudes "classiques" (gros montants + comportement suspect).

Analyse : Échec (Faux Négatif)

Le Cas "Attaque Furtive" (Index 584)

Le modèle a laissé passer cette fraude. Pourquoi ?

- **Montant :** 22.15 € (Jugé inoffensif par le modèle).
- **Technique :** Pas de changement d'appareil.
- **L'indice raté :** 48 transactions en 1 jour !

Leçon : Le modèle sous-estime la fréquence (`transactions_today`) par rapport au montant.

Biais & Éthique

Le problème du "Voyageur"

Notre analyse d'importance des variables (Feature Importance) révèle un risque éthique majeur.

Les variables **Location Change** et **Device Change** pèsent lourd (43% combinés). Un utilisateur légitime qui voyage souvent (digital nomad) ou change de téléphone risque d'être bloqué à tort (Faux Positif).

Pistes d'Amélioration

1. Règles Métier (Hard Rules)

Pour contrer les failles du modèle (comme le cas Index 584), nous devons imposer des limites strictes :

- Bloquer automatiquement > 20 transactions/heure.
- Double authentification pour tout changement d'appareil.

2. Enrichissement des Données

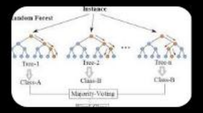
Pour affiner la précision :

- Calculer la vitesse des transactions.
- Intégrer le type de marchand (ex: Casino vs Supermarché).
- Réentraîner le modèle sur les schémas de "petites fraudes".

Questions ?

Merci de votre attention.

Image Sources



https://www.nvidia.com/content/dam/en-zz/Solutions/glossary/data-science/random-forest/img-3.png

Source: www.nvidia.com

https://cdn.dribbble.com/userupload/43050518/file/original-169564b152e3d6fee6b358d3434a8a03.jpg?resize=400x0

Source: dribbble.com
