

Obtenir un crédit sans historique bancaire : la data comme garantie

Machine learning, scoring alternatif et inclusion financière par la donnée

Pinel Baudelaire T. DAHOUI

Data Scientist & IA | Ingénieur en Travaux Statistiques
tpineldahoui@gmail.com | github.com/Baudelaire12

02 Avril 2026

Série LinkedIn STEM au Quotidien — Jour 15/40

Plan de la présentation

- 1 Mise en contexte
- 2 Concepts clés
- 3 Développement mathématique
- 4 Cas pratique
- 5 Conclusion
- 6 Références

1,4 milliard d'adultes sans compte bancaire

Chiffre clé

Selon la Banque Mondiale (Global Findex, 2022), **1,4 milliard d'adultes** dans le monde restent non-bancarisés. En Afrique subsaharienne, ce chiffre atteint **57%** de la population adulte — soit plus de 400 millions de personnes exclues du crédit formel.

- Au Bénin, le taux de bancarisation est d'environ **38%** (BCEAO, 2023) — mais le taux de pénétration du mobile money dépasse **65%**.
- Le score FICO, standard aux États-Unis, repose sur 5 variables classiques. Il est inutilisable pour les personnes sans historique bancaire (credit invisible).
- Des startups comme **Branch** (Kenya/Nigeria), **Jumo** (Afrique du Sud) et **Experian AI** utilisent des données alternatives pour scorer des millions de non-bancarisés.

Question centrale

Comment construire un score de crédit fiable pour quelqu'un qui n'a jamais eu de compte bancaire, en utilisant uniquement des données alternatives ?

Score FICO vs. Scoring Alternatif

Score FICO (traditionnel) :

Basé sur 5 variables pondérées :

- Historique de paiement : 35%
- Montant dû : 30%
- Durée de l'historique : 15%
- Nouveaux crédits : 10%
- Types de crédit : 10%

Limites : exclut 45 millions d'Américains (CFPB, 2015) et la quasi-totalité des Africains.

Données alternatives utilisées :

Source	Signal
Mobile Money	Régularité des flux
Téléphonie	Recharges, roaming
E-commerce	Fréquence d'achat
Réseaux sociaux	Profil, connexions
GPS/Mobilité	Stabilité du lieu de vie
Comportement app	Heure d'utilisation

Régression logistique — Modèle de base du scoring

Modèle logistique binaire (défaut = 1, remboursement = 0) :

$$P(\text{défaut} = 1 \mid \mathbf{x}) = \sigma(\mathbf{w}^T \mathbf{x} + b) = \frac{1}{1 + e^{-(\mathbf{w}^T \mathbf{x} + b)}}$$

Score de crédit transformé :

$$\text{Score} = \text{Score_base} + \text{PDO} \cdot \log_2 \left(\frac{1 - p}{p} \right)$$

où PDO (Points to Double the Odds) est typiquement fixé à 20 points.

Fonction de perte — Entropie croisée avec régularisation L2 :

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{p}_i) + (1 - y_i) \log(1 - \hat{p}_i)] + \frac{\lambda}{2} \|\mathbf{w}\|^2$$

Importance des variables et modèles avancés

Weight of Evidence (WOE) — Encodage des variables catégorielles :

$$WOE_i = \ln \left(\frac{\text{Distribution des bons}_i}{\text{Distribution des mauvais}_i} \right) = \ln \left(\frac{p_i^G / P^G}{p_i^B / P^B} \right)$$

Information Value (IV) — Pouvoir prédictif d'une variable :

$$IV = \sum_i (p_i^G - p_i^B) \times WOE_i$$

Règle d'interprétation : $IV < 0.02$ (inutile), $0.02-0.1$ (faible), $0.1-0.3$ (moyen), > 0.3 (fort).

Modèles avancés en production

XGBoost + SHAP : performance supérieure + explicabilité pour les régulateurs (Exigence Bâle III, article 22 RGPD).

LightGBM : vitesse d'entraînement sur des millions de clients.

Cas pratique : Branch International en Afrique

L'entreprise

Branch International (fondée 2015) opère au Kenya, Nigeria, Tanzanie et Inde. Elle accorde des micro-prêts via smartphone en analysant **plus de 10 000 points de données** comportementales par client — sans historique bancaire requis.

Données mobilisées (non-exhaustif) :

- Métadonnées de contacts téléphoniques (fréquence, diversité du réseau)
- Régularité de recharge de crédit téléphonique
- Historique des transactions M-Pesa
- Comportement in-app (heure d'ouverture, durée, navigation)

Résultats (Branch, 2023) :

- Plus de **5 millions** de clients servis en Afrique
- Taux de remboursement : **> 90%**
- Montants accordés : 2 \$ à 700 \$, avec décision en **< 24 heures**

Synthèse

Le scoring de crédit alternatif basé sur les données comportementales constitue une révolution pour l'inclusion financière en Afrique. Il permet d'évaluer la solvabilité d'individus invisibles au système bancaire traditionnel, en utilisant uniquement les traces numériques de leur vie quotidienne.

Perspectives :

- Réglementation du scoring algorithmique : droit à l'explication (RGPD Art. 22, directives BCEAO)
- Risque de biais discriminatoires : les données alternatives peuvent encoder des inégalités sociales
- Open Banking en Afrique : partage sécurisé des données financières entre institutions
- Scoring coopératif : utiliser les réseaux de tontines comme signal de crédit communautaire

« Dans un monde où tout laisse une trace numérique, l'exclusion financière n'a plus d'excuse mathématique. »

Références bibliographiques

-  World Bank. (2022). *The Global Findex Database 2021 : Financial Inclusion, Digital Payments, and Resilience in the Age of COVID-19*. World Bank Group.
-  Banque Centrale des États de l'Afrique de l'Ouest. (2023). *Rapport annuel sur les systèmes de paiement*. BCEAO.
-  Consumer Financial Protection Bureau. (2015). *Data Point : Credit Invisibles*. CFPB Office of Research.
-  Siddiqi, N. (2012). *Credit Risk Scorecards : Developing and Implementing Intelligent Credit Scoring*. Wiley.
-  Chen, T., & Guestrin, C. (2016). XGBoost : A scalable tree boosting system. *ACM SIGKDD*, 785–794.
-  Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions (SHAP). *NeurIPS*, 30.

Merci pour votre attention

Pinel Baudelaire T. DAHOUI

Data Scientist & IA | Ingénieur en Travaux Statistiques

`tpineldahoui@gmail.com | github.com/Baudelaire12`

LinkedIn : @Pinel Baudelaire DAHOUI

Série STEM au Quotidien — 40 jour