

## Termes de Référence

### **Consultant Data Scientist pour le développement de méthodologies de ciblage basées sur les données téléphoniques et géospatiales pour le Registre Social Dynamique du Togo.**

#### **I. Contexte et justification**

Le Gouvernement du Togo, en collaboration avec des chercheurs de renommée internationale, souhaite améliorer la fourniture de services de protection sociale et de services publics grâce à un nouveau registre social dynamique innovant, le Registre Social des Personnes et des Ménages (RSPM). Géré par l'Agence Nationale d'Identification (ANID), le RSPM est un outil essentiel utilisé par le gouvernement pour identifier les bénéficiaires éligibles aux programmes et services de protection sociale. Il a donc le potentiel d'impacter positivement la vie de millions de citoyens togolais.

Le RSPM en cours de développement prévoit plusieurs innovations. Contrairement aux registres traditionnels qui reposent principalement sur des enquêtes, le RSPM exploitera de multiples sources de données administratives et numériques, ainsi que des méthodes computationnelles de pointe (y compris l'intelligence artificielle et l'apprentissage automatique). Le RSPM sera interopérable avec diverses sources de données, notamment le Réseau Sécurisé Inter-Administration (RSIA), le système d'identification électronique (e-ID) et des sources non traditionnelles (comme les métadonnées des téléphones mobiles, les images satellites, et d'autres données télédéteectées). Grâce à cette combinaison de données et de méthodes innovantes, le RSPM sera plus dynamique et mieux à même de répondre à l'évolution des conditions socio-économiques des ménages, notamment en cas de chocs.

L'ANID recherche actuellement un data scientist possédant des compétences diversifiées et à la pointe de la technologie pour contribuer au développement du RSPM. Le consultant travaillera en étroite collaboration avec des agents gouvernementaux, des chercheurs internationaux en intelligence artificielle, ainsi que divers parties prenantes au Togo. Ce travail s'inscrira dans le prolongement du succès internationalement reconnu du programme Novissi du Togo, qui a utilisé des sources de données innovantes pour cibler les bénéficiaires de transferts monétaires d'urgence durant la crise du COVID-19 (pour plus de détails, voir <https://www.nature.com/articles/s41586-022-04484-9>)

Outre les compétences techniques, il est important que le Data Scientist comprenne les limites et les risques liés à l'utilisation de l'apprentissage automatique sophistiqué et de données non traditionnelles dans la conception des méthodologies de ciblage pour les registres sociaux. Le RSPM nécessitera donc également des mesures de protection des données solides pour garantir un usage des informations sensibles éthique et respectueux de la vie privée.

Les méthodologies développées devront être alignées sur le cadre des critères de vulnérabilité adoptés par le gouvernement togolais, et complémentaires aux méthodes de ciblage existantes. Un cas d'usage particulier concernera le ciblage des interventions en réponse aux chocs.

## II. Objectifs

L'objectif principal de cette mission est de développer une méthodologie innovante et durable pour l'exploitation des données de téléphonie mobile (CDR) et des données géospatiales, afin d'améliorer l'efficacité du RSPM, notamment lors des réponses aux chocs. Cela inclut des analyses statistiques et économétriques classiques, l'application d'algorithmes d'IA et d'apprentissage automatique, l'analyse de données géospatiales, l'ingénierie de pipelines de données avancés, ainsi que l'élaboration de mécanismes de protection des données.

## III. Détails de la mission

- Gestion de projet globale, y compris la coordination régulière avec les partenaires gouvernementaux, opérateurs téléphoniques, fournisseurs de données, chercheurs internationaux, et autres parties prenantes. Cela comprendra aussi une évaluation du cadre réglementaire applicable.
- Construction de pipelines de données durables pour l'intégration continue des données téléphoniques et de télédétection au sein du RSPM via la plateforme RSIA.
- Conception, mise en œuvre, validation et évaluation rigoureuse de modèles d'apprentissage automatique permettant de déterminer l'éligibilité des ménages à partir de données traditionnelles et non traditionnelles. Tests approfondis pour garantir la précision des modèles, identifier les biais, et améliorer les méthodes grâce à des retours d'audiences techniques et non techniques.
- Évaluation comparative de la performance de ces modèles par rapport à d'autres méthodes prédictives basées sur des données auto-déclarées. Analyse de la complémentarité des données non traditionnelles avec des méthodes comme l'évaluation indirecte des ressources .
- Élaboration de protocoles de protection des données, en collaboration avec des équipes juridiques et techniques, pour garantir le respect des standards internationaux et des réglementations locales en matière de confidentialité.
- Réalisation d'analyses géospatiales pour cartographier les zones à forte vulnérabilité à l'aide d'images satellites, de données participatives, ou de cartes de risques naturels, afin de visualiser les disparités géographiques de pauvreté et de vulnérabilité au Togo.
- Élaboration et diffusion de supports de formation pour le transfert de compétences et le renforcement des capacités du personnel technique du RSPM.

## IV. Livrables

- **Évaluation de la disponibilité des données téléphoniques:** note technique sur la disponibilité et la réglementation d'accès aux données de téléphonie mobile au Togo, y compris une analyse de la pénétration mobile et d'Internet, des opportunités de collaboration avec les opérateurs, et des risques associés.
- **Algorithmes entraînés pour prédire la consommation des ménages à partir de données autodéclarées collectées dans le registre social, les données de téléphonie mobile, l'imagerie satellite et d'autres sources de données, y compris le code et la documentation:** modèles d'apprentissage automatique entraînés, calibrés et validés (et les outils de traitement de données nécessaires) pour prédire la consommation des ménages par habitant à partir des données de téléphonie mobile et d'autres sources de données pertinentes. Le code doit être soigneusement documenté.
- **Rapport sur les performances et comparaison avec les référentiels:** évaluation rigoureuse des performances de ces modèles de prédiction, incluant des comparaisons avec les méthodes de ciblage

traditionnelles (par exemple évaluation indirecte des ressources) et les résultats publiés antérieurement sur des modèles similaires dans des contextes similaires.

- **Feuille de route pour un pipeline de données durable** : co-conception de processus automatisés d'injection, de transformation et de stockage des données afin d'intégrer durablement les données de téléphonie mobile au système RSPM via la plateforme d'interopérabilité RSIA.
- **Protocoles de protection des données** : analyse des principaux risques, mesures d'atténuation et garanties nécessaires pour obtenir et traiter en toute sécurité les données de téléphonie mobile, conformément aux exigences réglementaires applicables en matière de protection des données, de confidentialité et de cybersécurité.
- **Cadre d'intégration des données géospatiales** : document technique décrivant le cadre d'intégration des données géospatiales, incluant les méthodologies, les sources de données, les protocoles d'interopérabilité et l'architecture système proposée.
- **Cartes et rapports de vulnérabilité géospatiale** : ensemble de cartes géospatiales et de rapports analytiques mettant en évidence les disparités et zones de forte vulnérabilité à divers risques, notamment aux chocs climatiques, accompagnés d'analyses et de recommandations.

## V. Profil Recherché

- **Formation** : Master (obligatoire) ou doctorat (préféré) en économie, data science, informatique, statistiques, sciences de l'information ou domaine connexe.
- **Expérience** : minimum 3 ans d'expérience pratique dans la modélisation prédictive, l'apprentissage automatique ou l'économétrie avancée.
- **Compétences techniques** : maîtrise d'un langage de programmation orienté objet (Python, C, Java) et d'un langage de statistique (Python, R, STATA, MATLAB). Expérience avec l'IA ou l'analyse géospatiale très appréciée.
- **Compétences analytiques** : excellentes capacités conceptuelles et stratégiques pour réfléchir de manière innovante, au rôle des TIC dans le développement socio-économique et la réduction de la pauvreté.
- **Visualisation de données** : expertise dans la création de visualisations (cartes, tableaux de bord, etc.) accessibles aux non-spécialistes.
- **Compétences en communication** : excellente maîtrise orale et écrite du français et de l'anglais.
- **Gestion de projet** : capacité démontrée à gérer des projets complexes impliquant de nombreux acteurs.

## VI. Durée et lieu

La mission est prévue pour une durée minimale de 12 mois; le poste est basé à Lomé, au Togo.

## VII. Supervision

Le consultant relèvera de l'ANID et travaillera en étroite collaboration avec des chercheurs internationaux et l'équipe de la Banque mondiale appuyant l'ANID.

## VIII. Candidatures

Les candidats intéressés sont invités à postuler en ligne en suivant ce lien. <https://uri.gouv.tg/data-scientist>