Terms of Reference

Data Scientist Consultant for Developing Targeting Methodologies Based on Mobile Phone and Geospatial Data for Togo's Dynamic Social Registry

I. Context and Background

The Government of Togo, in collaboration with leading researchers, seeks to enhance the delivery of social protection and public services through an innovative new dynamic social registry, the *Registre Social de Personnes et des Ménages* (RSPM). Managed by the *Agence Nationale d'Identification* (ANID), the RSPM is a critical tool used by the government to identify eligible beneficiaries for social protection programmes and services; as such, it has the potential to impact and improve the lives of millions of Togolese citizens.

The RSPM being developed has several planned innovations. In contrast with traditional registries, which rely primarily on survey-based data collection, the RSPM will leverage multiple sources of administrative and digital data, as well as state-of-the-art computational methods (including artificial intelligence and machine learning). The RSPM will be interoperable with a range of data sources, including the *Réseau Securisé Inter-Administration* (RSIA), the e-ID system, and non-traditional sources of administrative data (potentially including mobile phone metadata, satellite imagery, and other remotely sensed data). Through this combination of novel data and methods, the RSPM will be more dynamic and responsive to the changing socioeconomic conditions of households, particularly in case of shocks.

ANID is currently looking for a data scientist with a diverse and state-of-the-art skillset to help develop RSPM. The data scientist will work closely with government officers, leading international artificial intelligence researchers, and various stakeholders in Togo. This work will build on the internationally-recognized success of Togo's Novissi program, which used innovative data sources to target emergency cash transfers during the COVID-19 crisis (for details, see https://www.nature.com/articles/s41586-022-04484-9).

In addition to the core technical competencies, it is important that the data scientist understand the limitations and risks of using sophisticated machine learning, and non-traditional data, in the design of targeting methodologies for social registries. Therefore, the RSPM will also require robust data protection measures to ensure the ethical and privacy-preserving use of sensitive information.

The methodologies developed by the data scientist will need to be aligned with the vulnerability criteria framework adopted by the Togolese government and be complementary to existing targeting methods. A particular use-case for the methodologies developed by the data scientist will be for the targeting of shock-response interventions.

II. Objectives

The primary objective of this consultancy is to develop an innovative and sustainable methodology for utilizing mobile phone Call Detail Records (CDR) and geospatial data to improve the effectiveness of the RSPM in Togo, particularly during shock-response. This involves traditional statistical and econometric analysis, applications of machine learning and artificial intelligence algorithms, geospatial data analysis, sophisticated data pipeline engineering, and the development of data protection safeguards.

III. Scope of Work

- Overarching project management, including regular liaisons with key partners in government, in the private sector (mobile phone operators, data providers), international researchers (who can advise on the data science methods), and other stakeholders. This will also require a comprehensive assessment of applicable regulatory frameworks.
- Build sustainable data pipelines for the continuous integration and processing of mobile phone and remote sensing data within the RSPM through the RSIA interoperability platform.
- Design, implement, validate, and rigorously evaluate machine learning models that determine household eligibility based on traditional and non-traditional data. Conduct rigorous testing to ensure model accuracy, identify potential biases and refine these methodologies through iterative feedback from technical and non-technical audiences.
- Benchmark the performance of state-of-the-art models with other predictive methods involving self-reported data. Explore and assess the complementarity of non-traditional data and models with existing targeting methods utilized by the RSPM such as Proxy Means Testing.
- Develop data protection protocols to ensure the privacy and security of RSPM data, working with legal and technical teams to align data protection protocols and safeguards with international standards and local regulations.
- Conduct comprehensive geospatial analysis to map areas of high vulnerability using satellite imagery, crowd-sourced data, or natural hazards maps (among others), providing visual insights into geographic disparities in poverty and exposure to multiple risks in Togo.
- Develop and deliver training materials for knowledge transfer and capacity building of RSPM technical staff.

IV. Deliverables

- **Mobile phone data availability assessment:** short technical note describing availability and regulations for working with mobile phone data in Togo, including an analysis of mobile phone and internet penetration, legal frameworks regulating access to mobile phone data, opportunities for collaboration with mobile network operators, and key opportunities, risks and challenges.
- **Trained algorithms for predicting household consumption from self-reported data collected in the social registry, mobile phone data, satellite imagery, and other data sources, including code and documentation:** trained, calibrated, and validated machine learning models (and necessary data processing tools) for predicting per capita household consumption from mobile phone data and other relevant data sources. Code must be carefully documented.
- **Report on performance and comparison to benchmarks**. Robust assessment of the performance of these prediction models, including comparisons to traditional targeting methods (such as Proxy Means Testing) and prior published results on similar models in related settings.

- **Sustainable data pipeline roadmap:** co-design of automated processes for data ingestion, transformation and storage to sustainably integrate mobile phone data into the RSPM system through the RSIA interoperability platform.
- **Data protection protocols:** analysis of the key risks, mitigation measures and necessary safeguards to safely obtain and process mobile phone data in compliance with applicable data protection, privacy and cybersecurity regulatory requirements.
- **Geospatial data integration framework:** a technical document outlining the framework for geospatial data integration, including methodologies, data sources, interoperability protocols, and proposed system architecture.
- **Geospatial vulnerability maps and reports:** A set of geospatial maps and analytical reports highlighting disparities, deprivations and areas of high vulnerability to various risks, including to climate shocks, accompanied by insights and recommendations.

V. Qualifications

- **Education:** Master's degree (required) or PhD (preferred) in economics, data science, computer science, statistics, information science, or related field.
- **Experience:** At least 3 years of directly relevant, hands-on experience using predictive modeling, applied machine learning, and/or advanced econometrics.
- **Technical Skills:** Proficiency in at least one object-oriented programming language (e.g., Python, C, Java), and one statistical programming language (e.g., Python, R, STATA, MATLAB). Applied experience with machine learning and/or geospatial data analysis is highly desired.
- **Analytical Skills:** Excellent conceptual and analytical skills, with the ability to think strategically and innovatively about the role of ICT/digital technologies for socioeconomic development and poverty reduction.
- **Data Visualization:** Excellent skills in data visualization, using maps, charts, dashboards and other techniques to communicate complex concepts to non-specialists.
- **Communication Skills:** Excellent oral and written communication skills in English and French.
- **Project Management Skills:** Demonstrated ability to manage large and complex projects with multiple stakeholders

VI. Duration and Location

The consultancy is expected to last for a minimum of 12 months, with the consultant based in Lomé, Togo.

VII. Reporting

The consultant will report to ANID and will work closely with international researchers and the World Bank team supporting ANID.

VIII. Applications

Candidates are invited to apply online: https://uri.gouv.tg/data-scientist