Request for Proposals

Expert in processing machine learning models and data science for automated audio-to-text transcription - DANE
Consultancy

Deadline: 10th September, 2023
Global Partnership for Sustainable Development Data
The Global Partnership is looking for a consultant who is an expert in data science and machine learning model processing; to support the National Institute of Statistics of Colombia (DANE) with the development of the Automated Transcription of Audio to Text pilot for the SIPSA-A Survey of the Supply component. The Global Partnership invites qualified individuals ("Bidders") to submit a Proposal for the requested services. The Contract resulting from this award will be a Service Agreement.

Bidders are encouraged to read this RFP in its entirety, paying specific attention to scope of services, instructions, and requirements. Publication of this solicitation in no way binds the United Nations Foundation to award a contract, nor will it pay for any costs incurred in preparing and submitting a proposal. The agreement resulting from this RFP will be provided to the most responsive Bidder, whose offer will be the most
advantageous to UNF/Global Partnership in terms of cost, functionality, and other factors as specified in this RFP.

Section 1: Background and purpose

Background: United Nations Foundation

The United Nations Foundation links the work of the UN with others around the world, mobilizing the energy and expertise of business and non-governmental organizations to help the UN address issues such as climate change, data, global health, peace and security, women’s empowerment, poverty eradication, energy access, and US-UN relations. The United Nations Foundation hosts and provides administrative, financial and contractual services for the Global Partnership. For more information, visit www.unfoundation.org.

Background: Global Partnership for Sustainable Development Data

The Global Partnership is a fast-growing and dynamic international association that brings together hundreds of different organizations, including governments, United Nations agencies, private companies, civil society organizations and many others. The Global Partnership convenes, connects and catalyzes action to address the problems of poor data use, access, quality and production, and to work with stakeholders to take full advantage of the new opportunities of the revolution of data at the service of sustainable development. The Global Partnership aims to link and align action, capabilities, and resources across geographies, sectors, and data communities. For more information, visit http://www.data4sdgs.org.

Background of the project: Alternative methods for the collection and collection of data for official statistics

The growing demands for information, with greater frequency, granularity, punctuality and precision than current data, are a challenge for statistical production; and responding to these needs using traditional methods is increasingly difficult. For this reason, the official statistical sector has launched various initiatives for standardization, promotion and execution of projects for the use of data from non-traditional sources. Such initiatives include those of the United Nations Global Working Group (GWG) on Big Data for Official Statistics and the Machine Learning Group of the High Level Group for the Modernization of Official Statistics, among others. DANE, in coordination
with the initiative Data For Now, has been developing various projects that incorporate the use of non-traditional methods and sources in statistical production; focused on priority issues for public policy.

Hunger, food security and nutrition are relevant phenomena for the transformation of agri-food systems towards more efficient and sustainable production and consumption and, with it, the achievement of the Sustainable Development Goals (SDGs) at the country level, particularly SDG 2 “Zero Hunger”. Undernourishment is associated with various risk factors, for example, the Food and Agriculture Organization of the United Nations (FAO) relates the increase in market prices of food and the inaffordability of healthy diets with higher levels of severe and moderate or severe food security, quantified by the Food Insecurity Experience Scale - FIES (SDG indicator 2.1.2). For this SDG 2.1.2 indicator, the results for 2022 show that 28.1% of households in Colombia had a prevalence of moderate food insecurity and 4.9% of these severe in 2022; When reviewing these prevalences by departments, it is evident that there are departments in which these prevalences are up to 59.7% (moderate) and 17.5% (severe). The foregoing shows that food insecurity is still a relevant issue in Colombia and having data associated with this phenomenon is key to public policy in the country.

In this sense, the analysis of the behavior of wholesale food prices responds to the information needs related to the identification of trends on the economic access of people to nutritious foods and their correlation with the increase in food prices and data of the indicators of the supply chains of agricultural production. In Colombia, a significant part of the information on food is collected through the statistical operation SIPSA (Information System on Prices and Supply of the Agricultural Sector); and there is a wide list of organizations that are users of the information

In order to guarantee the quality, coverage and timeliness in the collection of SIPSA information, since the end of 2022, DANE together with the Global Partnership for Sustainable Development Data has proposed the inclusion of new data collection and processing methodologies; improving the processes of statistical production, through an innovative development that allows the semi-automation of the collection of information. After the work carried out during the first months of 2023, DANE has a prototype that must be improved to be used as a pilot, before the production test; and for this reason the need for an expert person to support the development of the pilot and improve current developments has been identified; since the pilot seeks to optimize collection and validation times, and minimize errors through digital

---

1 Food and Agriculture Organization of the United Nations - FAO, Rural Development Agency - ADR, Bank of the Republic - BANREP, Agricultural Statistics Board, Colombian Corporation for Agricultural Research - AGROSAVIA, Rural Agricultural Planning Unit - UPRA, Ministry of Agriculture and Rural Development - MADR, Banco de Comercio Exterior - BANCOLED, Banco Agrario de Colombia S.A - Banagrário, Fund for the Financing of the Agricultural Sector - FINAGRO, Society of Farmers of Colombia - SAC, Departmental Secretarats of Agriculture, Members of trade unions , Members of exporters associations, Companies providing agricultural services, Academy (Universities and Research Centers), Citizens (researchers, students or the general public).
developments that assist the collector in capturing and processing information (audios, images, videos, numbers and texts).

In line with what has been described, the Global Partnership is looking for a professional expert in machine learning algorithm coding to support the National Institute of Statistics of Colombia (DANE), with the design and implementation of the Automated Audio Transcription pilot. A Text for the SIPSA-A Survey of the Supply component, whose purpose is to improve the capture, processing and/or quality control processes in real time of the information coming from the SIPSA statistical operation; because after the pilot it is expected to put the tool into production.

Key tasks for the consultant include

- Define best practices and techniques for audio preprocessing, for automated audio-to-text transcription, in the context of natural language analysis and Transformers, a type of natural language processing model. This involves exploring different approaches to handling ambient noise, accents, voice accents, loudness normalization, resampling, and removing silences in audio files.
- Design, implement and optimize audio preprocessing routines using Python adapted to the specific needs of the project. These routines should address the handling of ambient noise, accents, voice accentuation, volume normalization, resampling, and removal of silences in audio files.
- Document in detail the entire process of design, implementation and optimization of preprocessing routines.
- Train and support the DANE project team, by explaining the operation and results of the implemented pre-processing routines, including model performance metrics, and answering possible questions or concerns related to the process.

**Deliverables**

**First month**

- Detailed report documenting research on audio preprocessing best practices and techniques in the context of natural language analysis and Transformers.
- Set of audio preprocessing routines fully implemented in Python, that address each of the mentioned aspects in the context of natural language analysis and Transformers.

**Second month**
• Preprocessed test data set, which will be used to evaluate and validate the effectiveness of the implemented preprocessing routines, by comparing the outcomes of the implemented routine with test outcomes.

Third month
• Final versions of the set of pre-processing routines and final data set with results, incorporating adjustments based on the training and feedback from the Global Partnership.
• Detailed documentation of the implemented code, including explanatory comments and clear descriptions of the functions and procedures used, results and model performance metrics of the final routines. This documentation will be essential to facilitate the understanding and future maintenance of the preprocessing routines.

Ratings and profile
The candidate must have practical experience in the application of Transformers models, such as BERT, GPT-3, or similar, in natural language processing (NLP) tasks and, especially, in the context of audio analysis. The ability to adapt and extend these pre-trained models to obtain high efficiency in audio-to-text transcription will be essential. In addition to in-depth technical knowledge, the candidate is required to possess solid programming skills in Python, in order to carry out the effective implementation of the models and ensure their optimal performance.

Essential
• University degree in statistics, mathematics, systems engineering, computer science, or another field related to data science and machine learning.
• Practical experience in the application of Transformers models, such as BERT, GPT-3, or similar, in natural language processing (NLP) tasks and, especially, in the context of audio analysis.
• Strong programming skills in Python.
• Excellent oral and written communication skills in Spanish.
• Experience working with technical experts from various disciplines, and with people from diverse cultures.

Desirable
• Experience working with the collection, analysis and/or processing of statistical information.
• Experience with projects involving multilateral organizations
• Proven skills in designing, implementing, monitoring and reporting on training initiatives.
• Familiarity with the 2030 Agenda and the SDG indicator framework.
• Experience with projects involving United Nations agencies and other multilateral organizations.

Section 3: RFP Conditions

The United Nations Foundation reserves the right to:

  1. Reject any or all Bids and discontinue this RFP process without obligation or liability to any potential Bidder or other party.
  2. Accept other than the lowest price offered.
  3. Award a contract on the basis of initial offers received, without discussions or requests for best and latest offers.
  4. Award more than one contract.

Nothing in this RFP is, or should be relied upon by Bidder as a promise or representation of the United Nations Foundation. UNF makes no representation or warranty as to the completeness of this RFP and has no liability for any representations (express or implied) contained in, or omissions from, this RFP. This RFP and any response to any written notification are transmitted to the Offeror solely for the purpose of the Offeror preparing and submitting a Proposal. Each Bidder will maintain the confidentiality of the RFP and its content.

Any information or material submitted in response to this RFP and/or as a proposal (whether successful or not) becomes the property of the United Nations Foundation and will not be returned. By submitting a proposal, the Offeror must agree that the offer will remain firm for a period of not less than 120 days from the closing date of the RFP. Failure to comply with the specifications and requirements provided in this RFP may result in disqualification.

Section 4: Instructions for the preparation of the proposal
Proposals are expected to be comprehensive and include the information set out below. Elaborate or unnecessarily bulky proposals are not desired. Proposals must be submitted in Spanish.

1. Proposal narrative, no more than two pages

The proposal narrative will include:

1. A brief description of the Bidder’s experience and knowledge in the field illustrating general qualifications and capabilities to meet the terms of the RFP
2. A brief description of the Bidder’s understanding of the scope of services and the proposed methodology for the engagement
2. Curriculum vitae o CV
3. List of past and current employers/references
4. Cost requirements

The Bidder must include a detailed budget, which at a minimum includes the daily rate and the level of dedication for each product described above. All quotes must be in US dollars. Do not include travel costs in the budget; if any travel is approved, the cost will be refunded.

Proposal Submission: Proposals, including attachments (limited to 6 MB), must be submitted electronically in PDF format to: proposals@data4sdgs.org. Be sure to include in the subject line: Survey Machine Learning Expert - DANE. The United Nations Foundation will not accept proposals received by fax or mail.

All proposals are due on the date and time indicated above. Any proposals received after the required time and date specified for receipt will be considered late and will not respond. Late proposals will not be evaluated.

Section 5: Selection

The Global Partnership will review proposals with the goal of selecting the most advantageous offeror for the DSGP, based on the qualifications listed above, as evidenced by the proposal materials. The Global Partnership will consider the Bidder’s ability to deliver the scope of services and the feasibility of the approach. The Global
Partnership intends to enter into a contract that provides the best value and benefit, not necessarily the lowest price. The Global Partnership may meet with one or more Offerors prior to selection.

Section 6: Terms of payment

The award payment terms will be based on a fixed fee. Payment is contingent upon receipt of a valid invoice and is contingent on the successful completion of the products and related activities, at the sole discretion of the United Nations Foundation. Payment will be made in US dollars by the United Nations Foundation by check or electronic funds transfer/wire transfer. The final payment terms in the contract will prevail, not this RFP. Advance payments will not be provided. The contract will be drawn up in English.