I. Individual information for winner announcement

Team Member 1

- Name: Adam Zheng
- Hometown: East Amherst, New York, USA
- Who you are (mini-bio) and what you do professionally:

  Adam Zheng is a sophomore aerospace engineering student studying at Texas A&M University, with minors in mathematics and philosophy. Last summer, he worked as a systems engineering intern at the Marshall Space Flight Center. He also has experience leading student teams & organizations, conducting research involving space robotics and optimizing fork biodegradability, as well as a variety of other projects spanning machine learning, web development, CAD, Arduino, filmmaking, and more. He loves learning new skills and engaging in innovative, multidisciplinary work to enhance global well-being and promote a sustainable future for humanity. He is particularly passionate about space exploration, and aspires to one day work at a place like NASA, where he can contribute to advancing our understanding of the universe.

Team Member 2

- Name: El Hadji Malick DIEYE
- Hometown: Dakar, SENEGAL
- Who you are (mini-bio) and what you do professionally:

  El Hadji Malick DIEYE, also known as "Jay," is a passionate individual with a keen interest in geospatial sciences. He obtained his associate equivalent degree in Geomatics and a bachelor's degree equivalent in Land Surveying and Planning from the University of Thies/Senegal. He has accumulated a significant amount of experience through his engagements in both public and private services including the National Sanitation Office of Senegal.

  Currently, he is undergoing an internship at the Ecological Monitoring Center, specifically contributing to the Digital Earth Africa project in the role of a GIS assistant. In addition to his professional pursuits, he actively participates as a member of the Senegalese Association of Geomaticians and the Eco-Citizen Association. Notably, he extends his support to geomatics students at his former school by assisting them in crafting their Integrative Projects, facilitating their journey to validate their degrees.
  DIEYE aspires to become an expert in GIS and applied remote sensing, with the ultimate goal of pursuing further studies abroad and later contributing to the development of geospatial sciences worldwide.

Team Member 3

- Name: Malena García Vildoza
- Hometown: Buenos Aires, Argentina
- Who you are (mini-bio) and what you do professionally:

  Malena García Vildoza is an advanced student in Social and Cultural Anthropology at the Universidad Nacional de San Martín - EIDAES. For her undergraduate thesis, she conducted research on the activism of an organization of young immigrants in the city of Buenos Aires. Additionally, she has participated in various interdisciplinary citizen science projects, employing tools from applied social sciences and STEM for research and development purposes. She is interested in engaging in research projects that integrate tools from different sciences to contribute to social and environmental justice, as well as the well-being of marginalized communities, both academically and in an applied manner.

Team Member 4

- Name: Francisco Furey
- Hometown: Cordoba, Argentina.
- Who you are (mini-bio) and what you do professionally:

  I'm Francisco Furey, a lover of programming, python, nature, satellite imagery and drone piloting. I'm a YLAI and Global Shapers fellow. I'm from Argentina but I work for a startup in the US related to the NFL as a data scientist, although my dream is to work in or found a climatech startup. When I was 19 I started a social project that impacted the lives of more than 50 families (installing solar energy). I co-founded a social e-commerce site here in Argentina.

II. Team submission write-up

1. **What motivated you to compete in this challenge?**

   Probably the single most astonishing fact we learned over the course of this challenge was from The Association of Non-Governmental Organizations in the Gambia (TANGO), who told us about how lack of access to clean water was one of the main reasons why people leave their homes and emigrate from Gambia. To think that our work could help prevent this, to help preserve the livelihood of so many people - that really motivated us to do our best. And of course, we had many other reasons to be a part of this challenge, whether it was the unique experience of working on a diverse international team, contributing to the SDGs, gaining more data science experience, or advancing our careers.

2. **High level summary of your approach: what did you do and why?**

   We started out with a very rough premise of using satellite data to obtain information about groundwater. Once the team was formed, we sketched out a plan, made a Gantt chart to lay out our timeline, and scheduled meeting times so as to accommodate our differing time zones. From there, the critical initial step was to conduct a literature review, with the main two objectives being to define the problem (lack of access to

clean water) and to define the project. We researched previous studies on machine learning for groundwater prediction, compiled potential features and outputs for our model, searched for open-source datasets, and identified regions with the most severe clean water scarcity - all in order to identify a region with both a pressing need and the right data for us to work with. After about a week of intensive research, we settled on Gambia as our region of interest.

Over the following 3 weeks, Francisco, Adam, and Jay worked on gathering, pre-processing, and compiling satellite, climatic, and topographic data while Malena continued to research the situation in Gambia using digital ethnography, reviewing demographic surveys, and conducting interviews with members of the communities involved. The pre-processing took longer than expected, but with just under a week remaining, we were able to use our merged dataset to train 6 different ML models. Due to time constraints, we decided to narrow our area of interest down to just West Gambia, because of its significant reliance on groundwater and greater concentration of data points. We obtained feature data for this region, processed and compiled it, and ran it through our best ML model to produce predicted groundwater level data. Finally, we interpolated the data in QGIS to increase the resolution, and visualized it on our website.

3. Did you use any tools for visualization, data preparation, or exploratory data analysis that aren't listed in your submission?

The only tool that we forgot to mention, if it counts, is OpenAI's GPT (free version) - mainly just to assist with code writing.

4. **What are some other things you tried that didn't necessarily make it into the final workflow (quick overview)?**

In terms of the machine learning model, we were limited mainly by the time and resources available to us. For instance, we had planned to include input features such as land use/land cover, biome, and vegetation macrogroup, but didn't end up having time to incorporate them. We also attempted to train models with neural networks in Tensorflow, but lacked the computational resources to do so in a timely manner. These are all elements that may or may not have improved our model.

5. **If you were to continue working on this problem for the next year, what methods or techniques might you try in order to build on your work so far?**

If we had more time and resources to work on this, we would firstly want to improve the machine learning model, make it more robust, take more variables into account, and try other models such as deep learning. We would also like to further verify our model's accuracy using a larger dataset of piezometric groundwater level measurements. Once we are fully confident in our model, we could streamline it for easy usage & customization, enable it to visualize time series data, automate it to retrieve near real-time data, and even try expanding it to other regions around the world.