Team Viva Aqua

1. How does your visual inform a decision or action that furthers one or more of the key competition SDGs (zero hunger, clean water and sanitation, climate action)?

Our visual focuses on providing high-resolution, near real-time groundwater level data, which is a key metric for evaluating and monitoring groundwater sources. Sustainable usage of groundwater sources, in turn, plays an extremely important role in advancing not only SDG 6, but also SDGs 2 and 13.

Naturally, our project is most impactful with respect to SDG 6 (Clean Water), as it helps communities lacking clean drinking water to cost effectively identify locations for borehole wells, which are a common way to access clean groundwater. The ability to precisely monitor groundwater level is also highly beneficial to any groundwater dependent population or ecosystem, as well as for the purpose of informing international cooperation around transboundary aquifers.

Our project also plays a critical role in achieving SDG 2 (Zero Hunger), as access to groundwater is essential for agricultural practices and food production in many regions. After all, water scarcity can lead to decreased agricultural productivity, exacerbating issues of hunger and malnutrition. Finally, in terms of SDG 13 (Climate Action), our project provides the necessary tools to promote sustainable water resource management and create resilient water management systems, thus aligning with the climate adaptation strategies outlined in SDG 13.

2. How did you create your submission? Include the tools you used (e.g., Python, Excel, specific python packages), how you processed the data, and (if applicable) how you managed your codebase. If you have a public repository with code, you can share a link here.

Our project, AquaViva (yes, we just reversed our team name), was primarily developed using Python, mostly with Jupyter Notebooks but also some Python scripts. Key libraries used are: Pandas and NumPy for data manipulation and numerical calculations, Matplotlib and Seaborn for data visualization, Geopandas for handling geospatial data, Scikit-Learn for machine learning, netCDF4 and Xarray for working with .nc files, and ClimateservAccess (a custom library we built for accessing the ClimateSERV API). A summary of our process is as follows:

1. **Data Collection:** From IGRAC's GGIS (Global Groundwater Information System), we downloaded piezometric (groundwater level) data from 2015-2022 for 36 wells distributed across Gambia. Then we gathered corresponding data for 13 input variables: soil moisture, streamflow, evapotranspiration, enhanced vegetation index (EVI), mid-infrared reflectance (MIR), normalized difference vegetation index (NDVI), precipitation, estimated groundwater level range, curvature, drainage density, slope, hydrogeological region, and elevation, sourced from AppEEARS, ClimateSERV, BGS, and GGIS. QGIS was also used to process data.

2. **Data Cleaning and Preprocessing:** We used Jupyter notebooks to manage the various data formats (.nc4, .nc, .csv), visualize/analyze the raw data, and account for missing/erroneous data using nearest neighbor algorithms and linear interpolation.
3. **Data Integration:** Using pandas & geopandas, we merged datasets based on date, latitude, and longitude to form our primary dataset, with ~6300 rows. We then split this dataset into a training set (83%) and a testing set (17%).
4. **Model Selection and Training:** We trained 6 different regression models using scikit-learn: SVR, AdaBoostRegressor, GradientBoostingRegressor, RandomForestRegressor, SGDRegressor, and LinearSVR.
5. **Model Evaluation:** We employed metrics like Mean Squared Error (MSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2) for performance assessment, achieving our best result (MAE = 2.6 m, R^2 = 0.42) with Linear SVR.
6. **Model Optimization:** Applied Cross-Validation and GridSearchCV for hyperparameter tuning, and combined LinearSVR with Nystroem for kernel optimization.
7. **Visualization Data:** We gathered feature data for our area of interest, processed and compiled it as before, and ran it through the Linear SVR model to get predicted groundwater levels. Due to time constraints, we only used data at 500m resolution, but higher resolutions would have otherwise been entirely feasible.
8. **Visualization Creation:** We first used IDW (Inverse Distance Weighting) interpolation in QGIS to increase the resolution to about 177m. Then we used kepler.gl to put together our interactive visualization, exported it to an html file, and customized it to create our Aqua Viva website.

All code, data, and corresponding documentation is on our GitHub repository (with the exception of a few data files that were too large). We have strived to organize it according to standard best practices and to be easily understood by the scientific community or any interested parties who would like to use, modify, or build upon our work.

3. What motivated you to choose this topic?

The idea for this project originated from a volunteering trip that one of our team members made to Honduras to design a pipe system for delivering clean water to a rural community. In this community, there were many hand-dug wells that suffered from contamination and water borne illnesses, so a nonprofit came in, conducted a hydrogeological survey, and drilled a borehole well to provide uncontaminated groundwater. So our idea was to reduce the costs associated with hydrogeological surveying by using machine learning and satellite/climatic data to create an open-source tool that can map out key information such as groundwater level data.

Now, the central question that we grappled with was - what region should we focus on? We spent weeks researching various parts of the world, looking for an availability of open-source data as well as a lack of access to clean water. Our literature review took us from Argentina, to South Africa, to Brazil, Chile, and Mongolia, but eventually we settled on the West African region, which faces some of the worst water scarcity issues on this planet (affecting 1 in 3 people, according to the World Health Organization). In fact, one of our team members was born, raised, and currently lives there (in Senegal), having witnessed and been affected by water scarcity on a daily basis.

Although we knew our project would have limited time and resources, we nevertheless wanted to help as many people as possible. So we utilized various tools from applied social sciences, including the review of demographic surveys and digital ethnography, to determine the areas where our work could have the greatest impact. Using this approach, we identified The Gambia to be an optimal focus region, as according to Our World In Data, over half of the population lacks access to safely managed drinking water. Just in this past year, we found papers published by UNICEF and The Center for Collaborative Investigative Journalism that report a major lack of access to clean water in Gambia, deteriorating quality of life and leaving people highly susceptible to various diseases. Furthermore, we discovered that on crowdfunding platforms such as GoFundMe and JustGiving, Gambia had the single highest number of projects seeking funding for groundwater wells.

Most of these projects were located in the capital city of Banjul & surrounding urban areas of the Western subdivision, which encompass a total population of around 1,078,000. Indeed, according to UN-Habitat, the majority of Gambia's population is concentrated in this western region, leading to challenges in providing basic services such as access to clean water. Thus, due to a combination of all these factors, along with the availability of open-source groundwater level data, we decided to narrow our focus to this western region of Gambia.

4. How did you learn about the broader context of your chosen issue (e.g., historical, social, political)? This could include drawing on the lived experiences of team members, reading articles and literature, conducting interviews with community members, etc. Did what you learned change your approach?

To gain deeper insights on Gambia, we got in contact with The Association of Non-Governmental Organizations in the Gambia (TANGO), an umbrella organization founded in 1983 to coordinate NGOs operating in the country. Graciously, TANGO connected us with various community leaders who were actively involved in similar issues to the ones that we were tackling within Gambia, and we had the opportunity to interview some of them. They quickly confirmed the severity of the clean water scarcity and how it affects people's daily lives, describing it as one of the main reasons people seek to emigrate from Gambia. This bleak reality really resonated with us, as we realized that the water situation was at such a low point that people were willing to leave their homes just to find somewhere where they could have as basic of a human right as clean drinking water. It served as a crucial insight into what it was like to be in such a situation, and only reaffirmed our resolve to keep working towards our goal.

But of course, we first had to figure out if our idea was truly a sound one. Could machine learning really be used for these purposes? How useful would the data produced be to well drillers? To gain a better understanding of these questions, we performed a literature review and found research studies such as this one, which used machine learning to predict groundwater levels of GDEs (groundwater dependent ecosystems) in California. So it turned out that projects like ours had been done before, albeit without our focus on providing open-source tools for regions in dire need. We compiled a list of the various input features that these studies had used, and we also reached out to hydrogeology researchers at Texas A&M University who suggested additional variables related to the water balance equation, and helped us to carefully select the features that we would use in our model. They also brought to our attention other methods such as GPR (ground penetrating radar), which can provide very high accuracy measurements of hydrogeology. However, we realized that this equipment can be expensive for poorer communities, so even though we couldn't match its level of detail, we could focus on ensuring that our tool is open-source and free to use.

5. What are the ethics and/or equity issues you considered? What are some possible strategies or approaches for addressing them?

For our project, ethics plays a pivotal role as we prioritize the protection of human rights in Gambia. After engaging with representatives from the community, we are confident that the implementation of additional groundwater wells would very much have a net positive impact on the daily lives of the people. It would serve as a means to prevent diseases associated with the consumption and daily use of contaminated water. And it would contribute to curbing the ongoing emigration from the country, where individuals are forced to leave their homeland due to the environmental crisis, a situation exacerbated by the lack of access to clean, drinking water.

That is not to say that there are no potential issues. We believe that if our project, which aims mainly to drill more groundwater wells, is implemented, extensive work must be done to prevent inequitable distribution of these resources. We must always prioritize the interests of the inhabitants of Gambia. Under no circumstances should the project jeopardize their integrity or violate the fulfillment of human rights. And in cases where drilling might have negative consequences (whether ecological, cultural, or otherwise), all stakeholders should be consulted and a balanced decision be made. For this purpose, it would be crucial to collaborate with various key actors and institutions within Gambia and external to the country. These may include the United Nations, different NGOs, environmental organizations, etc.

Ultimately, the goal is to ensure access to clean water for the population of Gambia in a democratic manner, initially focusing on the most affected regions. This way, we could contribute to improving the quality of life and well-being for people in an equitable manner.

6. Would your team like to share the URL of an interactive visualization?

franfurey.github.io/aquaViva/