Water Security Project

Background

From the UN-Water, the water security is defined as “The capacity of a population to safeguard sustainable access to adequate quantities of acceptable quality water for sustaining livelihoods, human well-being, and socio-economic development, for ensuring protection against water-borne pollution and water-related disasters, and for preserving ecosystems in a climate of peace and political stability.”[1]. According to the historical and current data of different dimensions, the prediction of changes of water security in the future may be fulfilled, making a guide for the future actions. The Asian Water Development Outlook (AWDO) developed a national Water Security index to assess water security based on five dimensions: rural, economic, urban, environmental, and water-related disasters [2]. The index has five values, ranging from 1 to 5, the higher index represents higher water security. In this project, we develop models based on the data we collected to predict the historical and current National Water Security Index (NWSI) of the countries all over the world.

Data Collection

From Asian Water Development Bank (ADB), we collected the national Water Security Index of countries across Asia and the Pacific of the years 2013, 2016 and 2020. The data used to calculate the NWSI was collected according to the sources provided by Asian Water Development Outlook. However, some sources of their data collection only contain the data of Asian and Pacific countries, the historical data of other countries to calculate the index is not fully available. We selected variables of each dimension which are available for most countries and added other related data collected from FAO AQUASTAT into our whole dataset.

Pre-processing

Some data used by AWDO to calculate the NWSI of 2020 is not the data from 2020, but the data of closest years. When we merged the data, some data may be not from the same year, but is the most related. There are inconsistencies between each year's data and between countries within the same year. For example, few datasets include data in regional scale, such as “EU", “Euro area”, which does not exist in most others, leading to rows with massive missing values. After deleting the variables that have too many missing values, we implemented imputation and normalization on the data.

Model building

The water security index developed by ADB can represent the water security of each country properly. Based on the data of Asian and Pacific countries, we built the model with the selected predictors to predict the water security index of other countries. The six models (Random Forest Classifier, BaggingClassifier, AdaBoostClassifier, Gradient Boosting

Classifier, Lasso and Ridge Regressions) were developed to predict the index of index of other countries. The data was divided into training and testing sets, and the performance of each model was evaluated. The random forest model is selected, and then the features with the importance over 5% were selected to rebuild the random forest model. 8 predictors were used in this rebuilte model. The rebuilt model improved the accuracy of prediction and reduced the number of predictors, finally was decided to be used for prediction.

Predict the situation in the future

After selecting the important features and model, the indexes of different countries in the future could be predicted. Not many years of data could be found so the indexes in the far future could not be predicted. Therefore, the data in 2025 was predicted. Firstly, the important features in 2025 were predicted. Then these predicted data and the obtained random forest model were used to calculate the indexes in 2025. There were two reasons about why the indexes were not predicted directly but the features were predicted and then the indexes were calculated. Firstly, those features had practical meanings so they had tendencies to change over time. If the indexes were predicted directly, the indexes of the past few years should be calculated rather than obtained directly. Therefore, compared with indexes, features which were raw data were more predictable. What's more, only a few years of data could be calculated because only in a few years, all data of every country was available. But for each feature, it could be predicted using the years in which it had data and it was not necessary to consider if all features had data in those years. So more data could be used. 102 countries were predicted. According to the results, in 2025, most of the countries studied will be in the improved conditions. And only a small number of countries have water security issues which is quite optimistic.

Analyse the results with population

To see the relationship of population and prediction of water scarcity, based on the predicted result of 2020, we divided the index (0-4) into two groups: high water scarcity(0) and low water scarcity(1). By using test for homogeneity of variance and feature importance generated by random forest, we find some of the variables related to population show more importance while the others don't. And no strong relationship was shown between population density and the water security index.

Building web app

Web application for the visualizations across the project, including national water indexes and their sub-dimensional parameters, data exploration, and result of several models, was developed using the Plotly framework and deployed on Heroku. The data exploration section includes the following interactive figures:

1. Heatmap with the water security index in the considered country of Asia-Pacific region in a particular year according to the data of Asian Development Bank [2, 3, 4].
2. Two radar charts with insight into individual scores in particular key dimensions. The first one is focusing on showing the difference in key dimensions in the countries

based on their location in the region (Central Asia, East Asia, etc.). The second radar chart is showing the components of the key dimension that contribute to the final score. Both charts include the years 2016 and 2020 to show the progress that countries have made.

3. Bar chart showing how the countries of the particular location score in each of the reports key dimensions. We can observe all available years.
4. Second bar chart showing several columns from the dataset used in the model building by the regions.

The second part of the app shows the results that were obtained during the analysis and model building. It includes the following charts:

1. Bar chart of the model performance.
2. Results of Random Forest model. In particular, scores of the parameters parameters, prediction of indexes for 2020 across all countries, prediction of the indexes in 2025, and overview of the ensembled trees for the two estimators: 0, 5.
3. Results of the linear regressions. More specifically, cross-validation of the alpha parameter, and coefficient behaviour under certain penalties.

We wanted to achieve consistency within the application, however, due to the ill-maintenance of the one particular function in the plotly module tools that allow transferring matplotlib graphics into plotly supported plots, we haven't ended up with the desired results. The last six figures are out of the overall style of the web application due to the issue mentioned earlier. Overall, using the Plotly framework was quite beneficial for the understanding of the data that we were facing and discovering trends in the development of the national water security indexes across the years and regions.

Discussion

According to the national security index of Asian Water Development Outlook (AWDO) we built this model and predicted the NWSI of countries of the future. We gathered the dataset from different sources and selected the important variables for the prediction.Our results about the water security situation in 2025 are quite optimistic, only several countries will suffer from the water security issues,which may be caused by the lack of sufficient datasets. The model of this project is not the most optimal choice, but it is built on the extrapolated data and the logic of pre-existing research.

Project links:

Web app: https://bit.ly/3uyz3py
GitHub: https://bit.ly/3vAJpqb

References:

[1] What is Water Security? Infographic | UN-Water. UN-Water. https://www.unwater.org/publications/water-security-infographic/. Published 2013. Accessed May 29, 2021.

[2] Asian Development Bank. Asian Water Development Outlook 2020: Advancing Water Security across Asia and the Pacific. Asian Development Bank; 2020. https://doi.org/10.22617/sgp200412-2

[3] Asian Development Bank. Asian Water Development Outlook 2016: Advancing Water Security across Asia and the Pacific. Asian Development Bank; 2016. http://dx.doi.org/10.22617/RPT178628-2

[4] Asian Development Bank. Asian Water Development Outlook 2013:: Advancing Water Security across Asia and the Pacific. Asian Development Bank; 2013.