Mission: ClearWater Analytica is a public benefit company whose mission is to help achieve universal and equitable access to safe drinking water for all people. Currently, 3 in 10 people lack access to safely managed drinking water services and a large contributing factor are Harmful Algal Blooms (HABs), which are an increasing problem worldwide as climate change alters planet Earth. Toxins produced by HABs are a health hazard to humans and animals and are a large cost to water-quality stakeholders. Up to 48 million Americans get their drinking water from lakes/reservoirs are at risk of algal toxins, and HABs cost the USA approximately $4.6 billion annually. Toxic events can be crushing for local economies, costing them tens of millions of dollars, and municipalities/utilities that manage HABs poorly, risk losing trust with their citizens.

Problem Statement: **Being Caught Unprepared**
ClearWater Analytica started in response to a local need. Our state capital -- the City of Salem in Oregon -- suffered several HABs in 2018. Toxins were detected in the water treatment facility, and the public were advised not to drink tap water for an entire month. This disaster prompted Salem to find solutions so that this would never happen again. Their needs were identified as:

1) Advanced warning of the emergence of HABs in source waters
2) Geospatial intelligence to help direct targeted monitoring of source waters
3) Better understanding of the drivers of HABs.

Since 2018, Clearwater Analytica has worked with the City of Salem to develop a HABs forecasting system that addresses each of these needs[1]. These needs are felt across the country, and Clearwater Analytica is now working with over 30 additional municipalities and water utilities across the USA, as well as clients in Canada and Australia, to deploy the HAB forecasting system.

Value Statement: Forecasting Harmful Algal Blooms with Machine Learning
Just like the weather, value is in knowing ahead of time that a storm is on its way. In this case the storm is a harmful algal bloom. To forecast HABs, we have built a cloud-based software that finds and downloads all relevant data (see Appendix). This makes our technology standout compared to other groups who focus mostly on one form of data. For example, satellite imagery is important, but alone it is not enough to make accurate forecasts. Custom machine learning algorithms use these data to predict with a high level of accuracy (>85%) the likelihood of a HAB occurring in the next week. This system is tailored to each water system, and forecasts are made daily. In addition, long-term forecasts are made, providing information about when the next HAB season will start, and how HAB severity is likely to change in the coming decades. All this information is made available through interactive web-dashboards, and direct access to the data and models.

About Us: Clearwater Analytica was founded by Dr. James R. Watson and Dr. Mathew Titus, to answer the call from the City of Salem to help protect their local source waters. James and Mat met at Oregon State University, where James is also Professor, and they have now deployed their HAB prediction system around the USA and taken ClearWater Analytica through the prestigious Imagine H20 Accelerator[2]. Their team has expertise spanning machine learning, lake biology, data science and GIS. ClearWater Analytica has a strong tradition of research and development, and they continue to publish their results in peer reviewed and industry journals.

[1] https://clrwtr-insights.github.io/detroit-lake/
[2] https://www.imagineh2o.org/accelerator-2021-cohort

Appendix: Data That ClearWater Analytica Obtains and Uses

ClearWater Analytica's HAB forecasting system operates in near real-time by automatically finding and downloading all relevant data required to make accurate predictions. In addition, the custom machine learning algorithms discover patterns in one water system that can be useful for predicting HABs in other water systems. This means that our HAB forecasting system can make predictions at local, state, and national scales. The data we use to train our machine learning algorithms include:

USGS Stream Gage

Data from all stream gages (run by the USGS) that are connected to the water systems that we study via tributaries are automatically downloaded. These data include water temperature, water discharge, the concentration of chlorophyl-a, and other important variables used for measuring water quality.

Weather

Data for several weather variables are automatically collected from multiple sources, including precipitation, relative humidity (min and max), specific humidity, surface downward shortwave radiation (light), wind direction, wind speed, daily reference evapotranspiration, vapor pressure deficit, min/max air temperature. In addition, long-term forecasts of weather variables are obtained from Earth System Models. These are the models that inform the UN Intergovernmental Panel on Climate Change about likely changes to our environment in the coming decades.

Satellite Imagery

Imagery from Landsat 8 and Sentinel 2a are automatically downloaded and stored (~60m resolution, images updated ~weekly). These images contain data on the reflectance of light at different wavelengths or "spectral bands". With band reflectances, various quantities can be calculated, such as chlorophyl-a. In addition, we download and use information from the EPA's CyAN app. This application uses imagery from Sentinel 3 (300m spatial resolution, daily) to compute an index of cyanobacteria abundance.

Water Sample Data

In situ water sample data are obtained through collaborations with local groups, for example water utilities and municipal governments. These data include estimates of the concentration of nutrients (e.g., total phosphates, total nitrates), toxins (e.g., anatoxin-a, cylindrospermopsins, microcystin) and algae (e.g., cyanobacterial biovolume), turbidity (from secchi disk depths). Our machine learning algorithms also make use of information on toxin-expressing genes obtained using qPCR.

Contextual Information

We obtain lake polygons, lake centroid latitude and longitude, watershed type, lake area, shore length, shore development, lake volume, max lake depth, lake discharge flow, lake elevation.