



Anacostia Riverkeeper Uses DataRobot to Predict Water Quality in the Anacostia River



The Anacostia River runs through the heart of Washington, DC and parts of Maryland. Like many urban waterways, it is heavily polluted. Swimming has been illegal since the 1970s due to health concerns about pollution. Current methods for testing water quality take days to return results. This creates a delay between when the water is tested and when the results are shared with the public. And because water quality can rapidly change with weather conditions (for example, if it rains), test results can become outdated before they're even returned.

Anacostia Riverkeeper is charged with protecting and restoring the Anacostia River. Through DataRobot's AI for Good program, the nonprofit partnered with DataRobot to develop a system to predict if *E. coli* levels are above safe levels. This system helps shorten the delay between sample taking and results by making water quality predictions multiple times per day. Although it's not a replacement for physical samples, it helps add another layer of information. Frequent predictions provide better oversight and monitoring of water quality in the Anacostia River than manual efforts alone.



Company Info:

Name: Anacostia Riverkeeper
Location: Washington, DC
Industry: Nonprofit

Anacostia Riverkeeper is a nonprofit whose goal is to protect and restore the Anacostia River for all who live, work, and play in its watershed, and to advocate for a clean river for all its communities.

“Having a model that can predict water quality fluctuations across the day could help human health in other cities across the world.”

Robbie O'Donnell
Watershed Program Manager,
Anacostia Riverkeeper



The first step was to find a data source. The team needed data about the conditions of the river that they could use to train the models and make predictions. Fortunately, the United States Geological Survey (USGS) has sensors in rivers across the United States that stream data about their conditions. The team collected data about the discharge, gauge height, temperature, and more from 28 sensors at different locations in the Anacostia River. In addition to the immediate sensor readings, these features were aggregated over the preceding 12 and 24 hours to capture a historical sense of the river's conditions. The target was if the *E. coli* measurements were above a safe level.

The team used DataRobot to build and train dozens of binary classification models, and then selected and deployed the best model. The solution engineers then built a script that pulls the USGS sensor data, aggregates it, and sends it to DataRobot for scoring. The results are stored in a database and visualized using Tableau.

The team has since shared their results with other stakeholders in the water quality field. Judging from conversations with those individuals, this is one of the first times AI has been used to predict water quality across the whole Chesapeake Bay region. "We haven't seen anyone doing this before and we wanted to take a modern approach instead of doing simple models," said Robbie O'Donnell, the Watershed Program Manager at Anacostia Riverkeeper. The team hopes the results of this project can create other opportunities to replicate the results of this project in other waterways.



An Anacostia Riverkeeper volunteer collects water quality samples along the Anacostia River in Washington DC

"The impact of this project could be really huge and it's really meaningful to us."

— Olivia Anderson
Former Project Coordinator,
Anacostia Riverkeeper

Contact Us

DataRobot
225 Franklin Street, 13th Floor
Boston, MA 02110, USA

www.datarobot.com

info@datarobot.com