**In 400 words, please describe your project:**

We are making a drinkable water identifier that has 2 components; a physical component and a software component. For our software component, we made a machine learning model that takes the information from a variety of sensors and finds out if the water is potable, or safe for people to drink. For this machine learning model we implemented a neural network and random forest regression. For our physical component, we devised a sensor stick that includes many different sensors. These sensors include pH, hardness, Total Dissolved Solids, Chloramines, Sulfate, Conductivity, Total Organic Carbon, Trihalomethanes, and Turbidity sensors. The physical and software components would work together when the user dips the sensor stick into a body of water. Our machine learning model would, then, based on the recorded data values from all of the sensors, accurately inform the user of whether or not the water sample is safe to drink or not.

**How is your idea creative? What sets it apart?**

**We believe our idea is creative because we plan on incorporating a hardware component to work with our software component. We plan on making a sensor stick that will gather all the data values necessary for the machine learning mode. Our hack is generally used for a more individual use.**

**What is the potential impact of this idea? What are all the implications?**

* To people who live in much more underdeveloped areas and those with the lack of good water infrastructure, the local water supply may or may not be dangerous to human health, especially because of the lead pipes which the water passes through. This is especially devastating when people drink this water without knowing if it includes toxic chemicals in it or not. Our identifier can change that; identifying whether or not water is safe to drink is key to potentially saving people’s lives.

**How is this idea relevant to Bioinformatics? What is the connection?**

* Water is essential to our survival. However, when bodies of water are contaminated, the consumption of such water can be detrimental to our health. Our model allows us to ensure that the water which we are testing with our model is absolutely safe to drink.

**How far are you with this idea? Is it complete? If not, how far are you from completing it? What are the next steps?**

* We completed the software part of this project, but have only prototyped the physical component.
* In terms of completing this idea, we are sure that when the model sees tested water as safe, it is close to pure water in terms of water quality and is safe to drink. However, sometimes our models may label some of our water samples as unsafe, even though in reality they may be safe, so our next step is to address that issue.
* We created a machine learning model CreateML app to test how high the accuracy of most machine learning models would get on this data and we got around 68% accuracy. We developed a Random Forest and a Neural Network and both returned around a 67-68% validation accuracy. Once we realized the limits on this, we decided to say the water is drinkable only if we are 95% certain.

**Relevant Links (e.g. to dataset, code, relevant documents, etc.)**

* **https://www.kaggle.com/datasets/adityakadiwal/water-potability**