Garrick Morley

SENG 309 / 2 PM Wednesdays

Deliverable 1

For this project I'll be deducing whether or not water is potable (drinkable) using machine learning and a preset dataset. Ideally this dataset will train the model and determine whether or not a set of water descriptors will likely yield water that is safe for human consumption. I plan on using linear regression for this project but this may change as I progess through it.

The features for this dateset / project include the following:

ph, Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic Carbon, Trihalomethanes, and Turbidity.

These features will be used to calculate the odds of whether or not the resulting water is potable. While these features are slightly limited I believe that they give an accurate portrayal in regards to water sanitation. The only way I could see these features ase being better would be if they had laymans terms for each category, as this would help newcomers such as myself.

From what I've seen all of the features are relatively equal as far as usefulness goes, and extrapolating from our previous work and some speculation I believe that the ph value and hardness scale will be the most impactful on the result.

The dataset that I will be using for this project is called "water\_potability.cvs" and contains 3,276 values that can be used to train the model. The problem at hand is to determine whether or not the water is likely to be potable based on the numeric inputs for each feature. The specific value that I'll be trying to predict is the probability of potability based on the inputs. Some more potential features that could be utilized here would be more human factors such as the relative water pollution around each area and the average income of the surrounding population.

Regarding the last portion of the instruction, I feel that calculating water potability is definitely essential in the modern world, especially in light of the recent shortages of countless critical supplies during the epidemic. Hopefully we can eventually use machine learning to streamline the process of aquiring drinking water, both in our country and around the world.

NOTES FOR FINAL WRITEUP:

1. Need to add a gui to be used after the model is trained so we can input a set of features as an example and determine the odds of the water being potable.

2. See if we can just cancel some stuff out, like if the ph value is too low or too high and definitely makes the water non-potable we can rule all those ones out right off the bat. This is an example of boundaries within the features.

3. ieee- (something) .com is another good website like kaggle, include this as someone's work if possible.

4. He posted something in the week 10 section just now about tensor flow and GPU on a laptop.