Angela Zhu | MAIS 202: Deliverable 3 | March 22nd 2024

Final Training Results

My previous model was a binary classification one, where it predicted whether a bacterial strain had resistance against azithromycin or not. For this iteration, I made the model multi-class, so it can now predict whether there is resistance against azithromycin, ciprofloxacin or cefixime, all of them, none of them or any combination. To do this, samples which didn't have available resistance information for all three drugs were removed.

The model yields the following confusion matrix:

Predicted resistance Actual resistance	Azithromycin Resistant	Azithromycin and Ciprofloxacin Resistant	Ciprofloxacin Resistant	No Resistance
Azithromycin Resistant	26	0	1	53
Azithromycin and Ciprofloxacin Resistant	0	4	27	2
Ciprofloxacin Resistant	0	0	767	45
Ciprofloxacin and Cefixime Resistant	0	0	3	0
No Resistance	5	0	39	914

Interestingly, none of the test strains had resistance only against cefixime. While this says nothing about the accuracy of the model, it may suggest some overlap between having resistance against cefixime and other antibiotics.

As is consistent with the results obtained previously, for single antibiotic resistance, the model is good at labeling a strain as resistant only if it truly has resistance, but is less good at finding all resistant samples. Strangely enough, it also has a tendency to label samples that are resistant to two antibiotics as being resistant to ciprofloxacin. In both cases, resistant strains were labeled as ciprofloxacin resistant only, it may be because there are simply more unitigs associated with ciprofloxacin than cefixime. Another very likely cause is that, looking at the metadata sheet, there are 447 samples that are azithromycin resistant, 5 that are cefixime resistant and a whooping 223909 that are ciprofloxacin resistant, thus most likely biasing the model towards ciprofloxacin and also explaining why there are no cefixime. Thus, this method might not have been the greatest. It is perhaps better to train a separate model for each antibiotic and then combine the results. If there is indeed correlation between resistance against different antibiotics, then it could also be relevant to add the resistance to other drugs as a feature, rather than having it as a label.

Final Demonstration Proposal

The final product will be a landing page where the sequence, in the form of a string can be input and the webpage will indicate if the strain has antibiotic resistance against azithromycin, ciprofloxacin, and or cefixime. First, the sequence will be searched to find what unitigs are present, which will then be fed into the model. To connect the frontend with the backend, Flask will be used. Since I have no experience with Flask, I will watch online tutorials such as the "Learn Flask for Python - Full Tutorial" video by freeCodeCamp.org. The frontend will be made with basic HTML, CSS and Javascript without using a framework since that is what I have experience with.

The webpage will look something like this:

Strain Analyzer

Background
Some description about unitigs.

Analyze Your Strain

	Enter

Results

Drug Name	Resistant

Metrics