Slide1: Hello everyone. My name is Steven and this is a tutorial on how you can help humanity find our next home. My teammates are Josephine and Drew, and our sponsors are Dr. Nixon at NASA and Dr. Santerre from SMU.

Slide 2: We are presenting a methodology for automating the classification of spectrally resolved observations of multiple emission lines. We applied a methodology to Titan’s atmosphere and were able to classify 100% of previously discovered molecules.

Slide 3: I know that was a lot of words, but essentially what we were able to do was take NASA’s current manual process for identifying molecules in a planetary atmosphere, and automate it to save time and money.

Slide 4: To help highlight the importance of our work, here is a quote from Stephen Hawking in which he estimates that we only have 100 years to colonize another planet based off of several attributes. Climate change, overdue asteroid strikes, and population growth to name a few.

Slide 5: But have no fear! Utilizing the tools that we have developed, you will be equipped to help humanity search the cosmos for Earth 2.0!

Slide 6: To search for habitable planets you’re going to need some background on spectroscopy, a way to collect data, perform some data preprocessing, and then a method to analyze the data – which we have got you covered on!

Slide 7: There is a lot to know about spectroscopy, some of which I am sure you are already familiar with, however, the main terms that are important for our work are frequency, intensity, and flux.

Slide 8: Essentially, a molecule in an atmosphere reflects a certain amount of light at different frequencies. The way in which it does so, is unique to each molecule, just like a fingerprint is unique to a person.

Slide 9: For our project, the data was collected via the Atacama Large Millimeter/submillimeter Array telescope, or ALMA for short. It is located in Antofagasta, Chile.

Slide 10: To recap, light from a star reflects off of a planetary atmosphere and is then seen through a telescope. This process produces raw data.

Slide 11: To better visualize this process, here is an example of the data we used from Titan’s atmosphere, which is one of Saturn’s moons and the planetary body we applied our processes to. The tall spikes in the data represent possible reflections from molecules.

Slide 12: Once the data has been collected, we look for the frequencies associated with spikes three standard deviations away from the mean. We pass these frequencies through databases, such as Splatalogue and the Jet Propulsion Laboratory, to retrieve all molecules that could possibly contribute to each spike in our data.

Slide 13: This is an example of the spectral lines retrieved back for Ethyl Cyanide, where the small sliver of red is our window of data from the ALMA telescope, and the rest are the millions of known spectral lines that span all frequencies from the databases.

Slide 14: Updating our roadmap, the raw data is sent to databases and the data is enriched to include all the spectral lines for possible molecules in the planet’s atmosphere. Here I will hand it over to Drew to discuss the method behind analyzing the data.

Slide 15: Thanks Steven. Now that we’ve collected all the data we need, the next step is to figure out which emission lines are strongest for each molecule, and match up their catalogued frequencies with the closest frequency bin from the input data. Once we know the frequency in the input data that each emission line is closest to, we can determine if that molecule may be responsible for a 3 sigma spike.

Slide 16: Now, the data is in a format for our algorithm to do all the heavy lifting. The first thing it does is identify a set of molecules that can reproduce the spike signatures found in the input data by implementing a greedy set cover algorithm.

Slide 17: Once a set of probable molecules has been identified, our algorithm performs some calculations to weight the likelihood of each molecule in the set cover as being present or not present. As long as the weight meets the threshold criteria set by you, the molecule is classified as present.

Slide 18: Our full roadmap is displayed here. We identify a list of possible molecules using a set-cover algorithm, and then implement 2 different selection metrics to classify the molecules in the set cover as present or not present.

Slide 19: The molecules we classified as present using the better of the two selection methods are displayed above. Our results line up perfectly with what has already been identified by NASA in this data set, and even identified one possible new molecule for further research.

Slide 20: The outcome of our project can be seen on the left: a plug n’ play jupyter notebook, an $8,000 cost savings and 80 man-hours savings to NASA, a custom API to retrieve enriched data from the underlying databases, and an accurate list of potential molecules on a planetary atmosphere. So what are you waiting for? Go out there and find New Earth!