### **FAQs**

1. **Who is your customer?**

One common challenge we encounter in analyzing complex samples is the difficulty of separating out the spectra of the sample to identify individual components. By utilizing this product, we will be able to efficiently and accurately decompose complex spectra into individual components’ spectra, thus enabling us to have deeper insight into the composition and properties of martian soil mixtures. This product could be valuable to other customers working in similar fields, especially biologists working with complex biological samples.

1. **How will this make your customer’s life significantly better? Solve their pain point?**

The product will improve the customer’s life by automating their task while increasing accuracy and reliability. The customer needs to identify the individual components of Martian soil from Raman spectra. Traditional methods for spectral decomposition require manual examination, so the product will reduce the amount of labor needed for research.

1. **Why is this a problem that needs to be solved right now?**

Oftentimes, vast amounts of complex data are usually generated in chemistry and biological labs. Traditional methods for the analysis of these data can be labor-intensive and time-consuming and may not provide accurate and reliable results . Our goal is to have models that can overcome these deficiencies. This product will meet those needs and make scientific research more efficient.

1. **What might disappoint the customer?**

The resulting product might not be as robust as requested. The model may have lower accuracy and reliability than the state-of-the-art model.

1. **How will the customer discover or find our product?**

Our overarching aim is to publish our work in a peer-reviewed chemistry journal. The build ups of our model and the methodologies used in data collections will be made available for public access. We will also publish our code in Github so others can download and run locally on their machines.

1. **How will you measure success?**

We will measure success through various stages. First, we will compare our model’s training accuracy to our validation/test set accuracies to ensure that we are not overfitting our training data. With the baseline model created that has an accuracy better than guessing, we will have achieved our first measure of success.

Next, we will compare our model to the industry standard, which is available through the most current research papers and techniques; falling within a reasonable range of an industry standard model’s accuracy would be our second measure of success.

If we can achieve an accuracy comparable to the industry standard, we will then compare our model to a human experimenter experienced with manually labeling Raman Spectra. By matching or bettering a person’s labeling accuracy, we will have achieved our third measure of success and successfully created a viable product that can be employed in research labs.

1. **What is the most contentious aspect of your product? (What sparks the most intense debate?)**

Using a deep neural network on Raman spectroscopy will likely spark concern in one area: the network’s performance at generalization. When working with these types of networks, there is always concern about a model’s output when it is given something it has yet to see. If the model is well trained, the output should be correct and useful. However, if the model is over-, or under-, trained, the model’s performance will suffer, making it a less desirable product. In our case, there are numerous types of samples and materials, and ensuring that the model maintains its accuracy and provides reliable results will be one of the more important aspects to attain.

1. **What is the dataset for your project?**

The dataset will be collected using confocal Raman instruments. This instrument utilizes a 532nm diode laser. The green laser passes through optics and opto-mechanics (Thorlabs) and enters an inverted Nikon Ti-u microscope with a 20x objective (Nikon, NA 0.60). The sample mixture to be measured contains three Martian soil analogs , Gypsum, Olivine and Soapstone. Mixture dataset of about 3000 raman spectra will be collected. Raman spectra of individual components will also be collected. This will serve as our ground truths.

1. **What are the computational needs for your idea?**

The computation needs for our idea can be met by a Google Collab Notebook using the freely available CPU and RAM resources. Our dataset can fit into RAM and does not require expensive hardware resources. For these reasons, we should be able to create a viable product using the free resources available through a Google Collab Notebook.

1. **What are the key milestones in your roadmap:**

|  | **Milestone/Feature Description** | **Priority** | **Size Guess**  **(S, M, L, XL)** |
| --- | --- | --- | --- |
| **1** | **PRFAQ (March 9th)** | **High** | **M** |
| **2** | **Data Collection (March 1st)** | **High** | **L** |
| **3** | **Baseline Model (March 29th)** | **High** | **L** |
| **4** | **Midterm Report(April 14th)** | **High** | **XL** |
| **5** | **Refined Model (April 22nd)** | **High** | **M** |
| **6** | **Project Video(April 28th)** | **High** | **L** |