What is the problem you're trying to solve, why is it important to solve it?
I'm trying to make machine understand animal's speech.
Human's language ability is distinctive, while animals do show some primitive speaking abilities.
The first clear evidence was found in vervet monkey, which has been found to use three types of alarms in wild, i.e., snake alarms, eagle alarms, and leopard alarms. Each alarm evokes contrasting responses. Monkeys on the ground respond to leopard alarms by running into trees, to eagle alarms by looking up, and to snake alarms by looking down. Playback experiments confirmed that the responses were triggered by sound.

These alarms sound quite different even to human ears. However, not all animal speech are that easily distinguishable. What if zoologist don't agree to each other? Is there a more objective way to decide the meaning of anmial speech? The answer is yes. I propose a methodology here, using vervet monkey's alarm as an example. First, we use clustering to categorize audios to three groups. If they match the classification by zoologist, then we're confident about our data and method. This method can helps to solves controversial cases where zoologists don't get consensus.

Furthermore, once the classification is confirmed by clustering, we can train a supervised model, which in the future, can be used to interprete new audio records. We may implement it in laptops, or even smartphones. We can easily see how it, if succeeds, will help zoologist in the wild.

It's important because first it can helps us to understand animals' speaking ability. It's fasinating. Second, as mentioned above, it can help zoologists working in the wild.

How does it affect people? How does it affect you?
It can help zoologists to understand animal language in the wild. The results can be interesting to the public as well, me included. As Darwin said, "He who understands baboon would do more towards metaphysics than Locke." Such understanding helps us to think and understand better of ourselves.

Describe a solution In order to solve your problem, you should know what a solution to it looks like. Try to describe such a solution as precisely as you can. For instance, if it is a prediction problem, how accurate do your results need to be? How quickly does your algorithm need to produce an answer?
I'll find vervet mokey's speech records, extract features from it (applydiscrete Fourier transform), use PCA to reduce feature dimension, then send it to clustering algorithms. As there're 3 types of alarms, I should get 3 clusters, and different type of

alarms fall in different clusters. Accuracy and confusion matrix can be used as metrics.
In supervised learning part, conventional metrics, like accuracy, F1 score, can be used. The choice may depend on how imbalance the data are.

In your report, explain briefly why you think your chosen metric(s) is/are applicable.
As I'm using clustering to verify scientists' observation, accuracy can measure how good the correspondance is. Confusion matrix can give detailed results of model

What is the size of your dataset?
I manually trucate the audio records and get 1 record for snake alarm (~26s), 6 for eagle alarms (~5s), and 2 for leopard alarms (6-15s). Eventually I extract 196, 241 and 141 data points for each alarm.

How many features are present?
Initially, after fast Fourier transformed, as I use a 1024 points long sliding window, there are 1024 frequency components as features. After PCA, 97 features are enough to keep 99% variance.

Which features seem most promising?
As the features are magnitudes in discrete Fourier transform, it's not easy to understand them intuitively.

Are there any categorical variables that may need to be converted?
No.

Which ones would you expect to perform well?
As there are large number of features and small number of clusters, I expect k-means to work well, as well as its variants, like partitioning around medoids (PAM).

How easy is it to convert the available data into a suitable form?
The largest difficulty is noise reduction. As audio is recorded in wild, the background noise may be serious, and noise reduction is a hard problem. I use a rather brutal method to handle it (discarding weak frames). This part can be improved but expertise is required.

What pre-processing operations do you have to carry out on the features? (e.g. scaling, normalization, selection, transformation)
Here are some pre-processing.
First I manually cut the original record to get relevant parts;
then records' magnitudes are scaled to [-1,1], and a smoothing window (Hamming window is applied);

then frames too weak (mean absolute magnitude in frequency domain < 0.5) are discarded;
finally PCA is applied to get reduced features.

Are there any incomplete data points or outliers that you have to work around?
Yes, if we take background noise as outliers. As mentioned above, noise reduction is not easy. I discard weak frames to partially solve it.

Using the metric(s) you defined earlier, measure your current performance. Is it close to what you expected?
In the clustering part, I get >75% accuracy in classification. In supervised learning, I get >95% accuracy. It's better than my expectation.

Are there any better metrics you can come up with?
No in my opinion. We can use other usual metrics like F1 score, but as the data are quite balanced, accuracy, together with confusion matrix, are good enough.

For each version of your solution, track what changes you make and how they affect performance.
As stated above, noise reduction is the hardest part, and I discard frames to overcome it. At first, I use 0.2 as the threshold, and don't get good classification. Then I try 0.3, 0.4 and finally 0.5. As the threshold increases, the data size decreases (since I'm discarding more frames), but I get better classification. From trial and error, 0.5 is about the smallest number I can use to get satisfactory classification.

Does it ever become worse? If so, note down and figure out why.
When I make the theshold larger than 0.5, the result doesn't improve too much but I lose too many data. I consider it as becoming worse.

Report how your project evolved, and what changes you made to your specifications (if any).
As discussed above, I play with the threshold to get better noise reduction. That's the main change in the evolution of the project.

What was your experience like working on this project? Do you feel more confident taking on open-ended projects like these in the future?
It's very interesting and I think it's promising. I search the lieterature and only find one paper does something similar. As there're so many animals showing potential language ability, this is can be a fruitful research field in the future.
Yes, I'm more confident!