Robert Hess

Lisa Lei

1. What is the overall objective of your project?

To see how close the pronunciation is to the “correct” pronunciation.

1. Where/how are you going to obtain the data? For example, do you need to scrape the data from a web source, contact a company or individual for access/permission?

The “correct” pronunciation will come from google translate.

1. Are there any copyright or terms of service limitations that may affect your ability to work with the data?

There is no copy right.

1. What type of machine learning algorithm / process do you anticipate using (e.g., classification, regression, clustering) and if applicable, what variable(s), in particular, will you be trying to classify, etc.

We believe it will be a classification problem. Classes being: You said it correctly/ you said it incorrectly. We think we will be using a Neural Network.

1. How many and what type of attributes do you anticipate? (e.g., less than than 10 vs. 50-100, numeric vs. nominal, time-series data, etc.)

However many attributes audio has. We will be normalizing the audio so that they can be compared.

1. How many instances or records do you anticipate?

Top 500 most used words.

1. What do you see as the biggest risk to the success of your project?

Data wrangling and gathering. And learning to process audio as numbers.