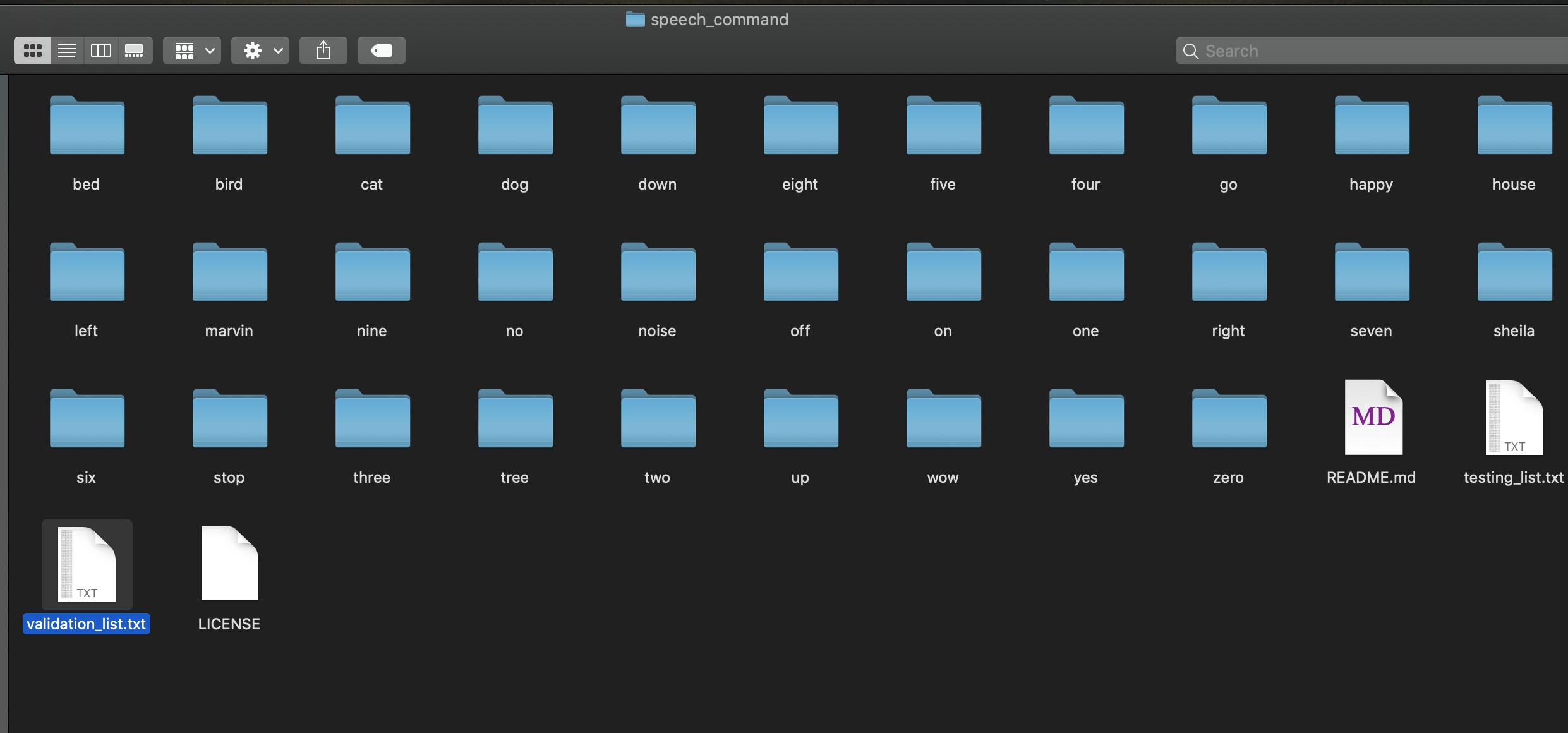
Capstone Project 2 Milestone Report 1

**TensorFlow Speech Recognition Dataset**

About the Data:

The dataset contains 65,000 one-second long utterances of 30 short words, by thousands of different people.

Snapshot of the local repository:



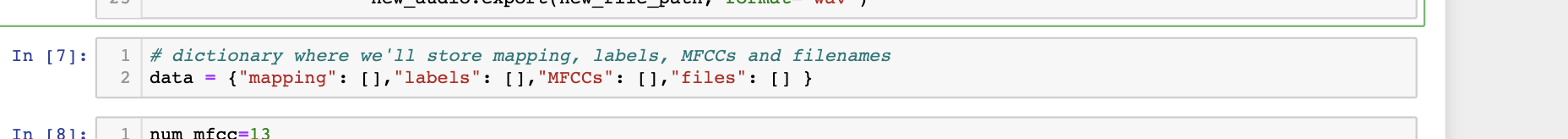
Note: There is also a separate repo which has captured the inputs for some random noises

The below are the words used in the dataset:

'right', 'eight', 'cat', 'tree', 'bed', 'happy', 'go', 'dog', 'no', 'wow', 'nine', 'left', 'stop', 'three', 'sheila', 'one', 'bird', 'zero', 'seven', 'up', 'noise', 'marvin', 'two', 'house', 'down', 'six', 'yes', 'on', 'five', 'off', 'four'

Approach:

1. Create dataset using input audio files
2. Loop through the repository to extract the labels
3. Loop through the sub-directories to extract the frequencies from the audio files
4. Breakdown the files in “Backgroundnoise” folder and extract consecutive 1 second audio of the files in “Backgroundnoise” folder
5. Create json file with the following inputs from the audio files
   1. Mapping
   2. Labels
   3. MFCCs (Frequencies)
   4. Files (path)



1. Store the json file as the input data for building ML models in-order to predict the audio input