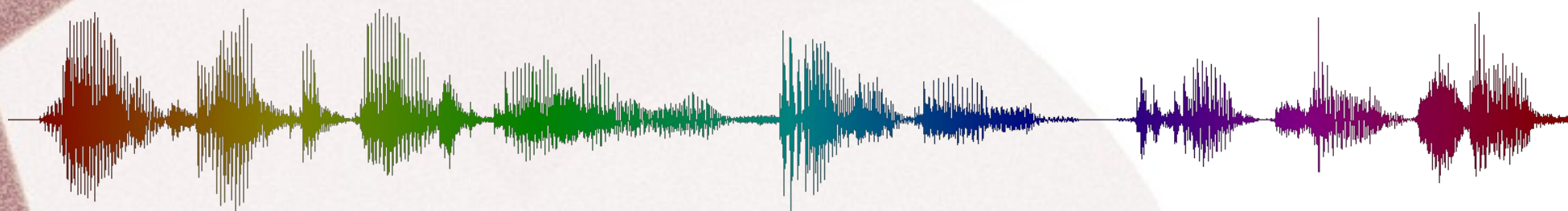


Speech Recognition



By Alex Billinger


Purpose

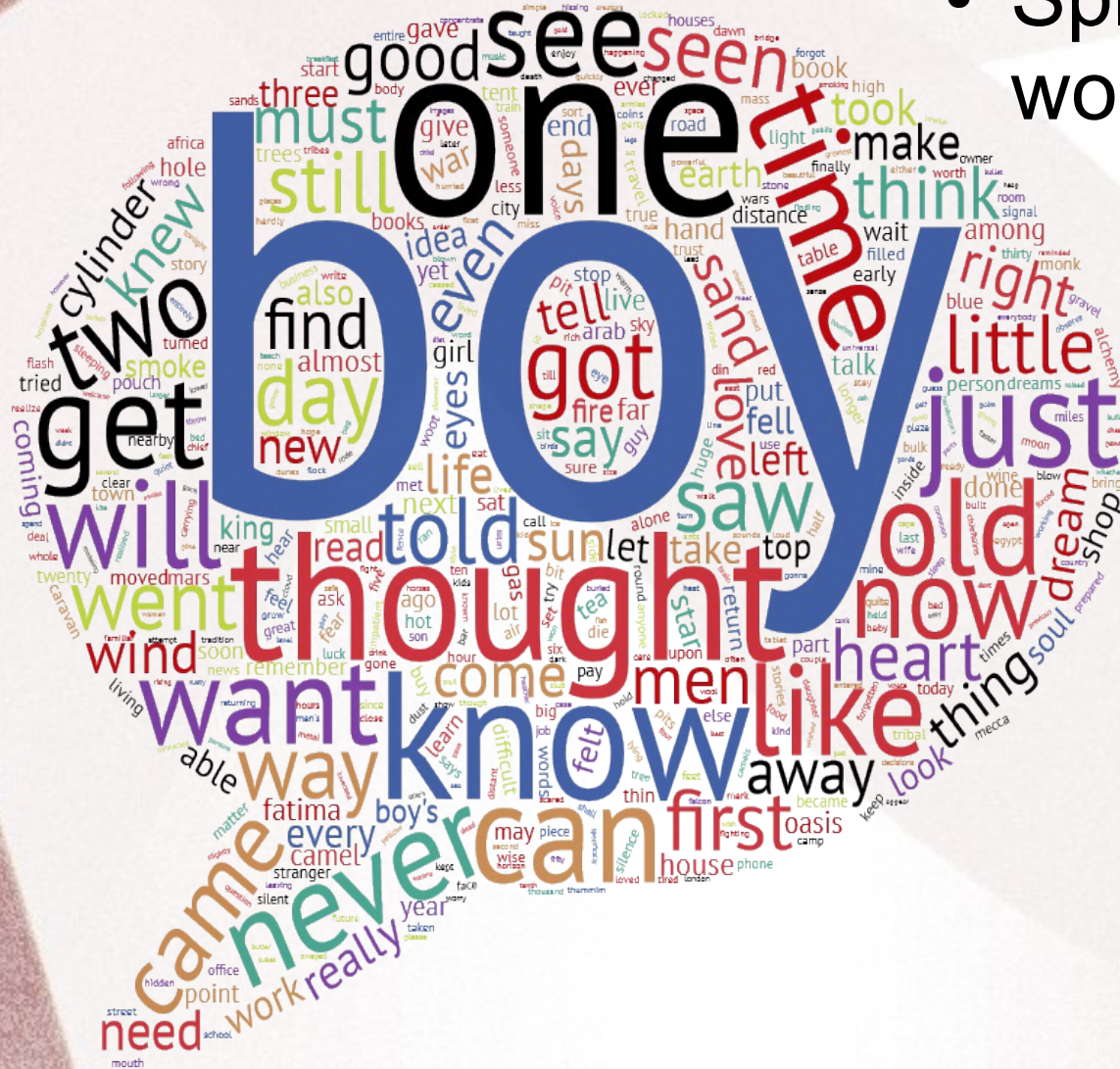
- Build algorithm to predict word based on audio data
- Can be used for subtitle generation or speech-to-text features
- Important as accessibility feature

Data

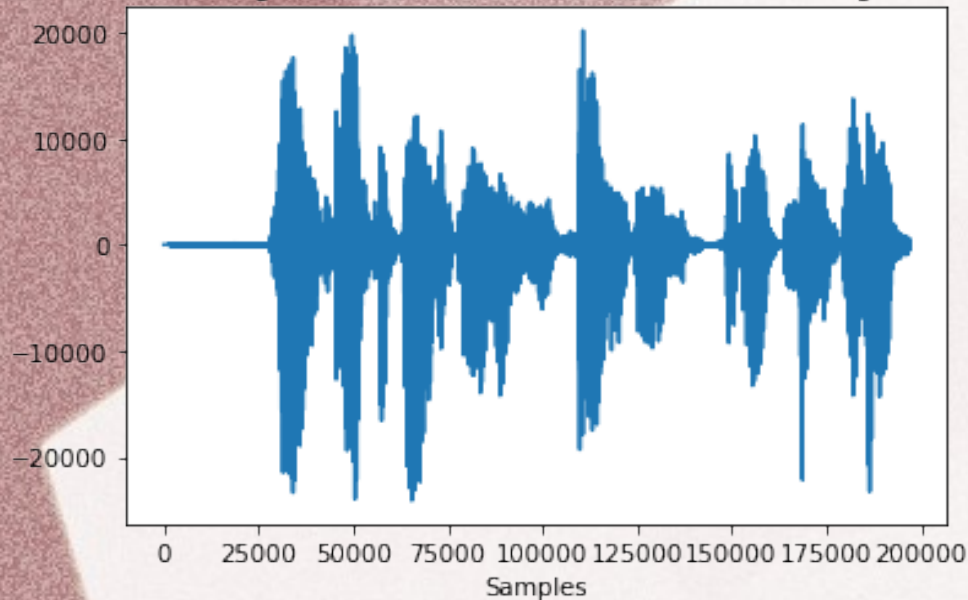
- <https://www.kaggle.com/mozillaorg/common-voice>
- Audio files of short phrases, spoken by many different people
- This data set contains 8000 unique words, the model is only trained with the 1000 most common words

Processing the Data

- Split phrases into words
- 
- Normalize volume
 - Cut silence
 - Get syllable count

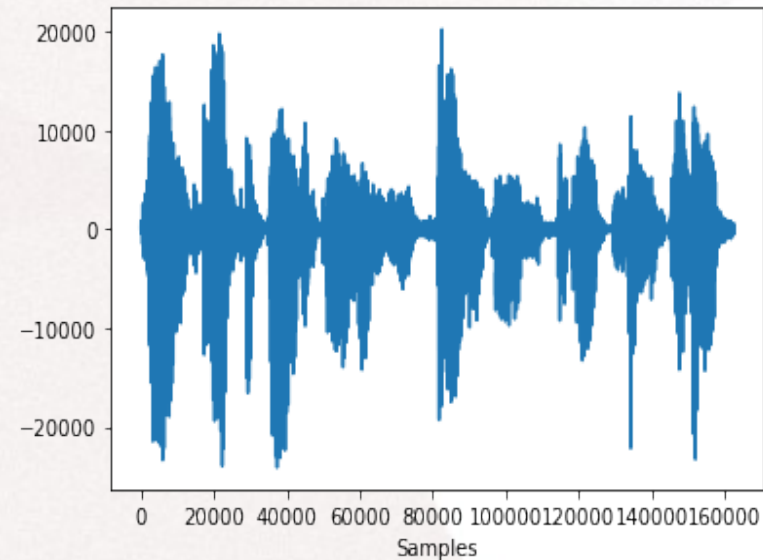


Full Audio Sample
learn to recognize omens and follow them the old king had said

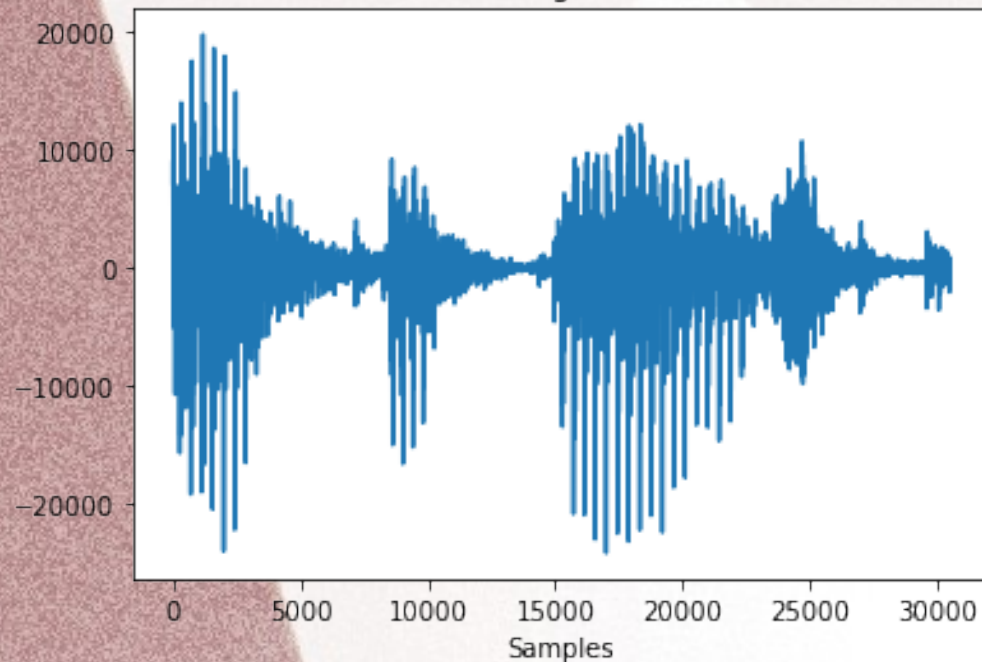


Audio Data

Nonsilent Audio

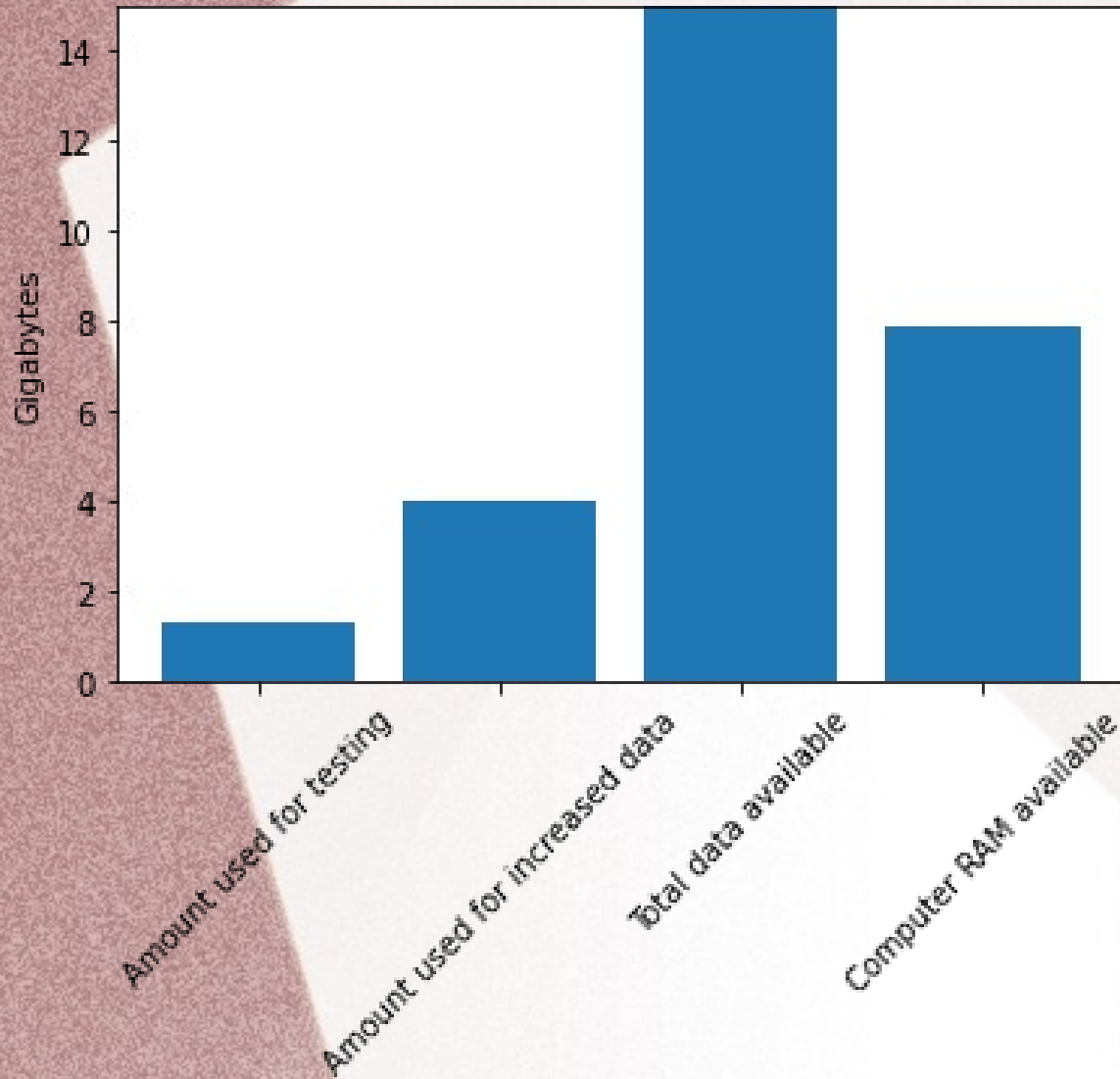


recognize



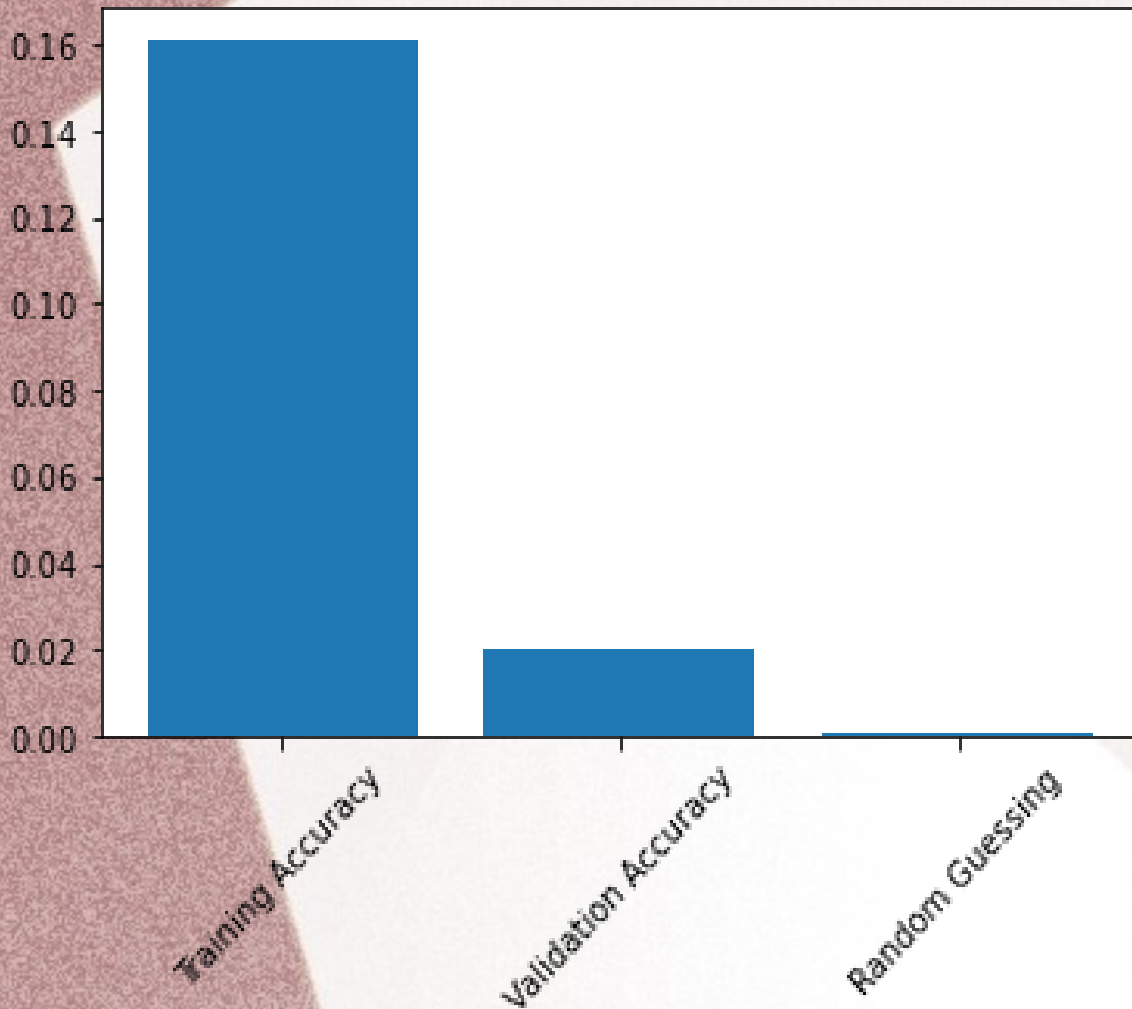
- Phrase example
- “learn to recognize omens and follow them the old king had said”
- Original, silence removed, single word (recognize)

Computing Difficulties



- Even 2.5% of data set is large amount of available RAM
- Could not use all data
- Could not run fast enough to test parameters

Accuracy of final model



- Trying to predict the correct word out of 1000 words it's trained with
- Often overfit
- Best model-training data 16% accuracy; validation accuracy only 2%
- Oddly, increasing amount of data processed caused lower accuracy

Future Work

- Need more computing power, to utilize all data
- Determine how to split an unlabeled audio file, to predict multiple words in a phrase



Thank you!