CLASSIFYING TEXT BY RHYTHM AND SOUND

Nina Wang

MOTIVATION & GOAL

- Shakespeare's iambic pentameter
- Homer's use of hexameter
- Lyrical nature of poetry
- Academic writing,

- Do different genres of writing "sound" different? Can we quantify that?
- Use rhythm as a metric to classify a piece of writing to its genre

APPROACH

- Transform written text (novels, poetry, and academic papers) into an audio stream
- Develop algorithms to define and analyze the rhythmic quality of every audio file.
- Use established machine learning techniques to classify writing based on its rhythmic patterns.

IMPLEMENTATION

- Use Amazon Polly to accomplish text-to-speech conversion, using a custom script
- Process resulting audio files with Praat, an audio analysis software. Use forced aligners to match audio with the original text
- Complete data analysis using Python scripts
- Use support vector machines as my machine learning model to perform classifications

RESULTS

- N/A

CONCLUSIONS

- N/A