Language Classification

DSIR-113020
Capstone Presentation
Luke Heeringa

Objective

Problem Statement

Can machine learning be used to identify the language being spoken in an audio recording?

Parameters

5 second audio clips

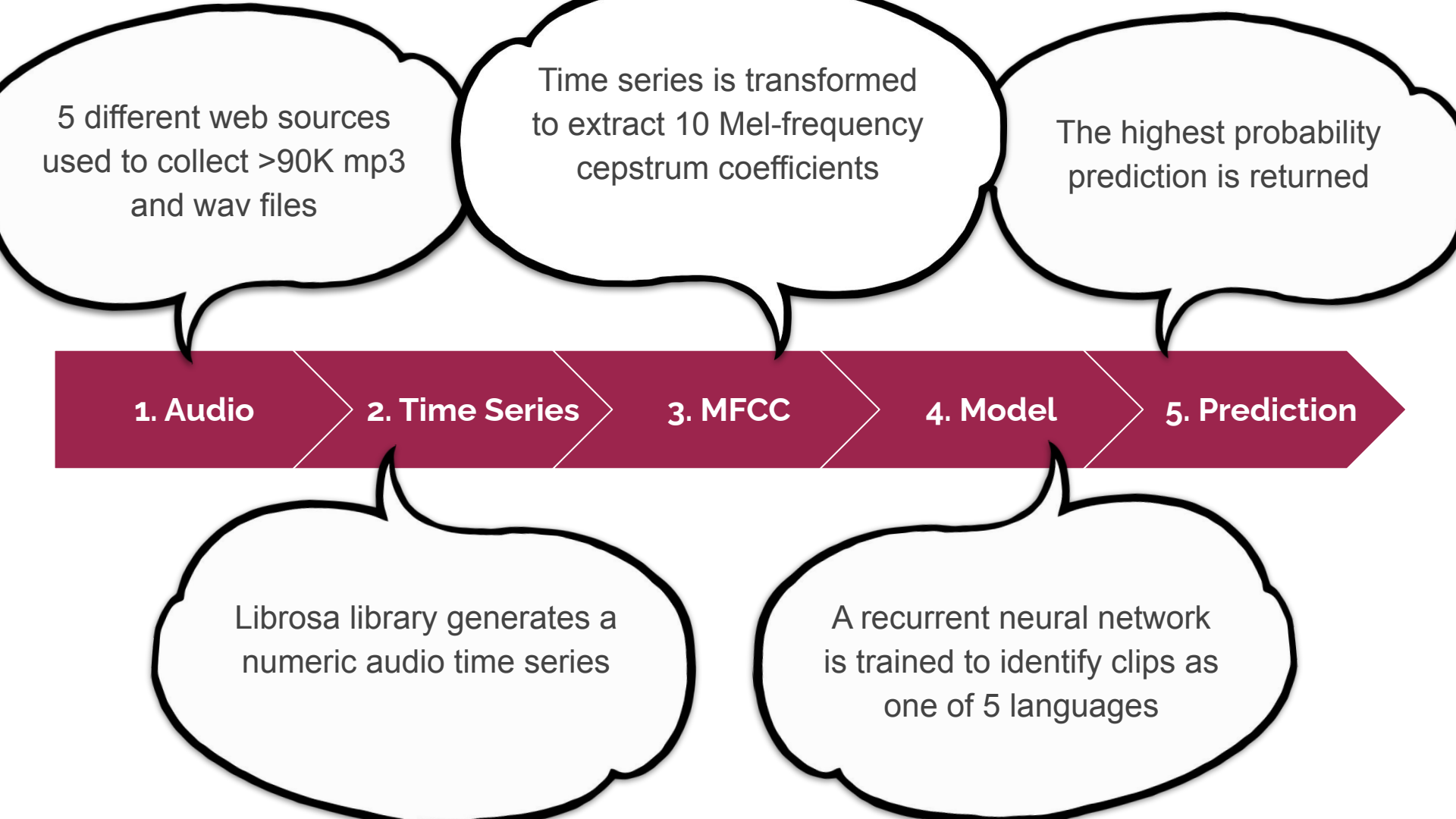
5 major world languages

10 Mel-frequency cepstrum coefficients

Scenarios

Connect clients with appropriate interpretation services

Identify context for auto-transcription and auto-translation software


5 different web sources used to collect >90K mp3 and wav files
Time series is transformed to extract 10 Mel-frequency cepstrum coefficients
The highest probability prediction is returned
1. Audio
2. Time Series
3. MFCC
4. Model
5. Prediction
Librosa library generates a numeric audio time series
A recurrent neural network is trained to identify clips as one of 5 languages

Why these 5?

	Millions of Speakers	
	Native	Total
English	379	1,348
Spanish	480	543
French	77	267
Russian	154	258
Mandarin	918	1,120

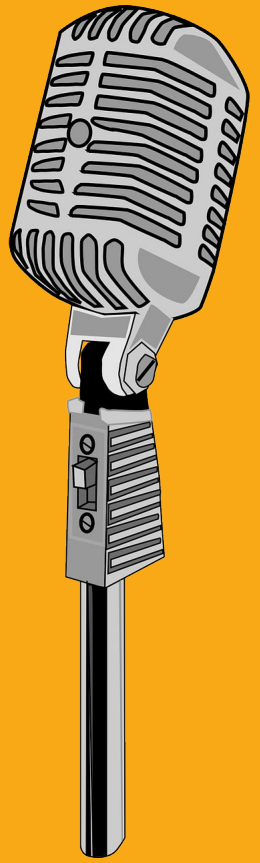
Data Collection

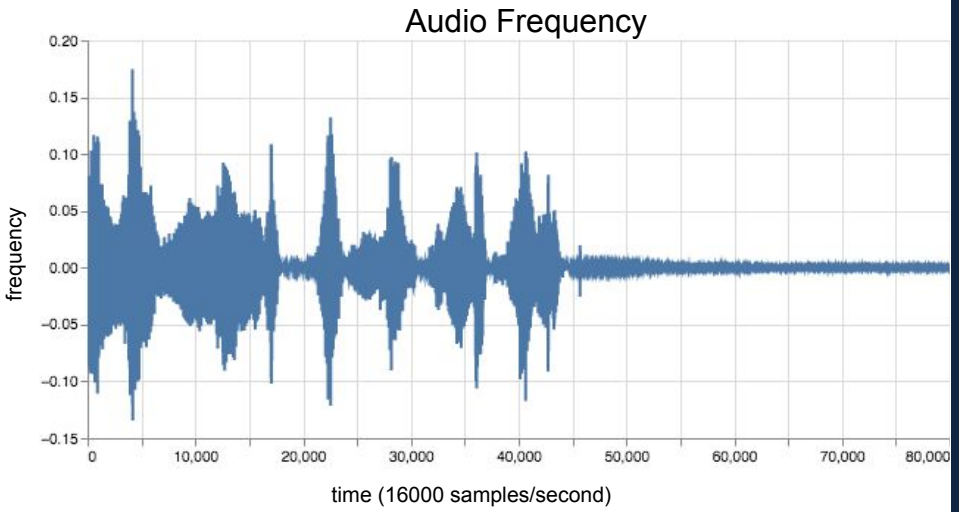
BY SOURCE

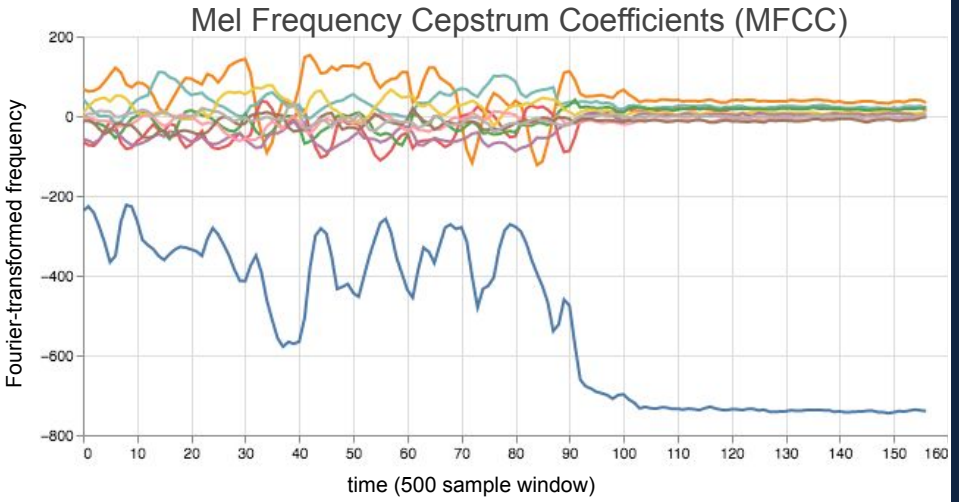
- Audio-Lingua.eu
- EveryTongue.com
- Omniglot.com
- VoxForge.org
- CommonVoice.mozilla.org

BY LANGUAGE

- English - 19,280
- Spanish - 18,885
- French - 18,021
- Russian - 18,919
- Mandarin - 18,163







What is a Mel Frequency Cepstrum Coefficient (MFCC)?

Neural Network Modeling

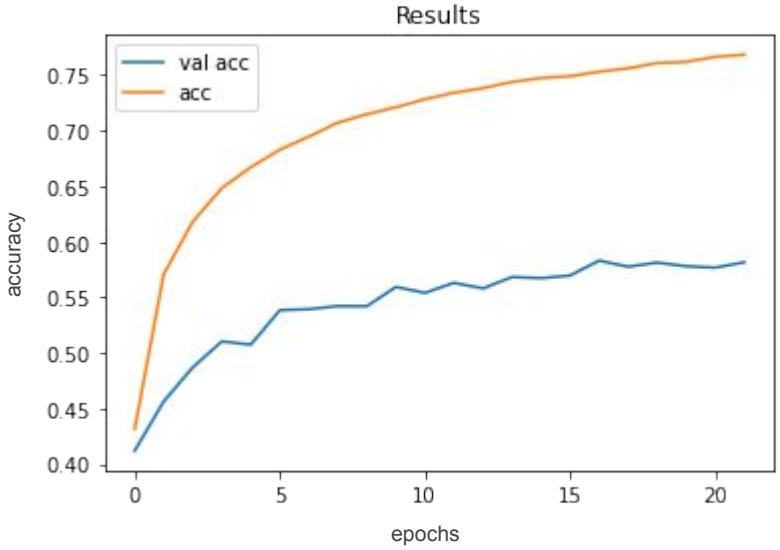
Validation Accuracy

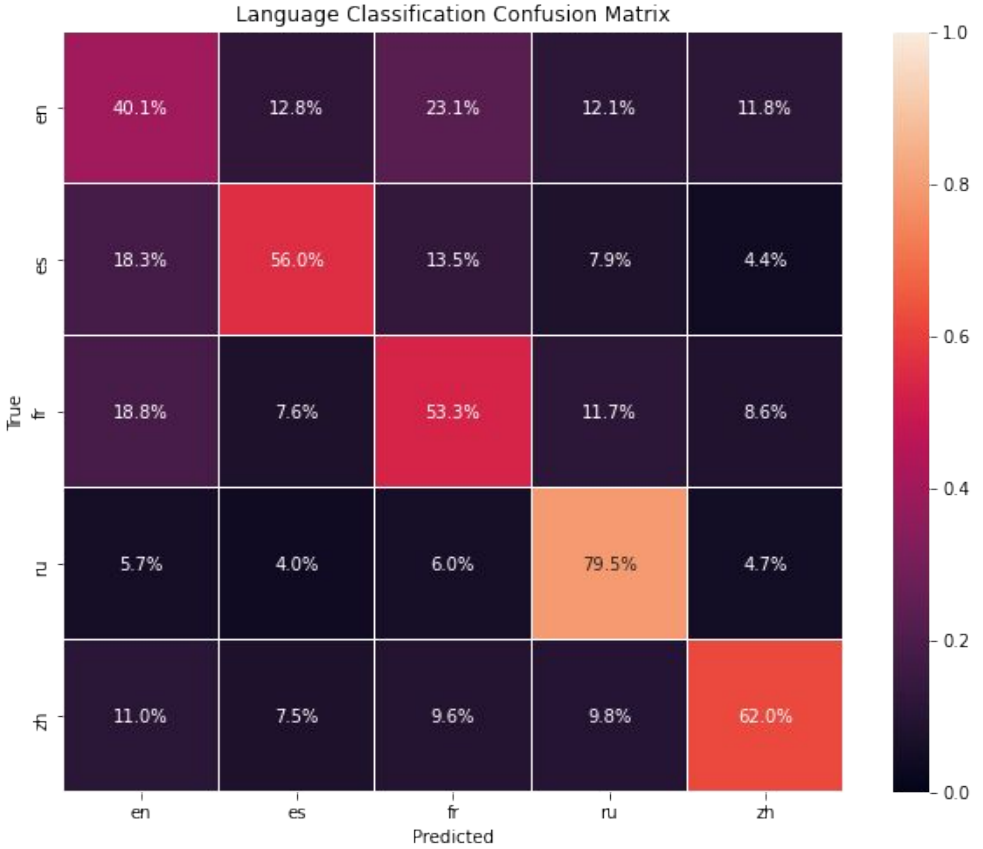
Null model: **19.8%**

Dense layer model: **37.5%**

Gated Recurrent Units: **58.2%**

Gated Recurrent Unit Network: Results





Demonstration