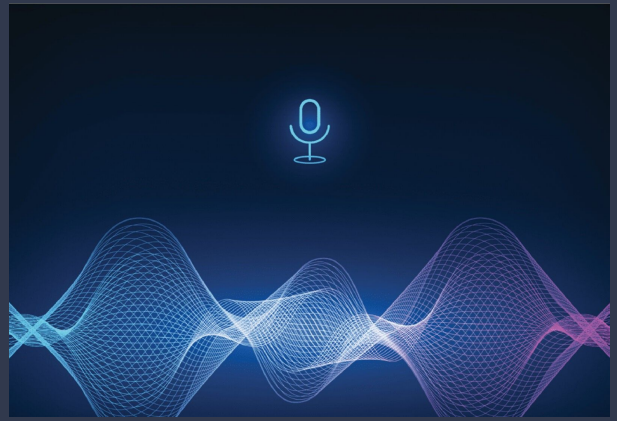
Hey Jarvis

Daniel Sarria, Kyle Clemente, Phi Lu

Objectives



- Utilize machine learning to build a system that can recognize audible events, particularly the human voice through audio classification
- Works similarly to "Hey Siri" or "OK, Google" and is able to recognize keywords or other audible events, even in the presence of other background noise or background chatter
- Activate an LED to indicate voice recognition of keyword

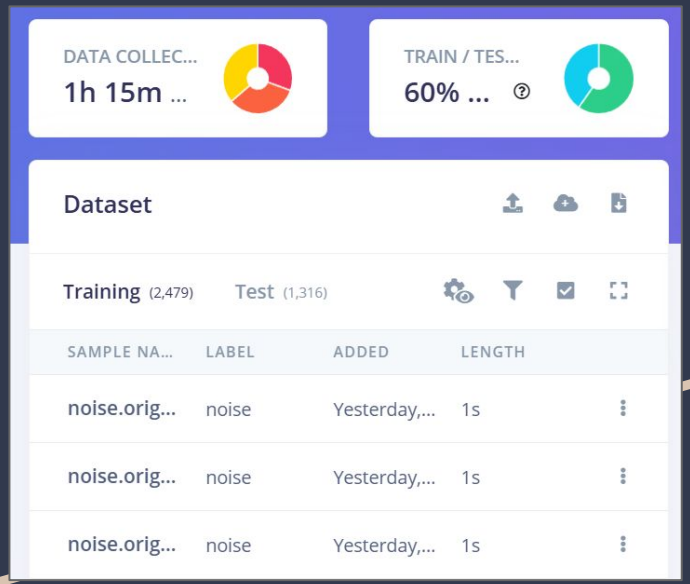
Hey Jarvis

- Audio classification system using Arduino Nano 33 BLE Sense that recognizes the keyword “Hey Jarvis” and works in the presence of other background noise or background chatter
- Potential to add on other commands following the keyword “Hey Jarvis”, such as turning on the lights, opening the garage door, and other smart tasks

Steps Overview

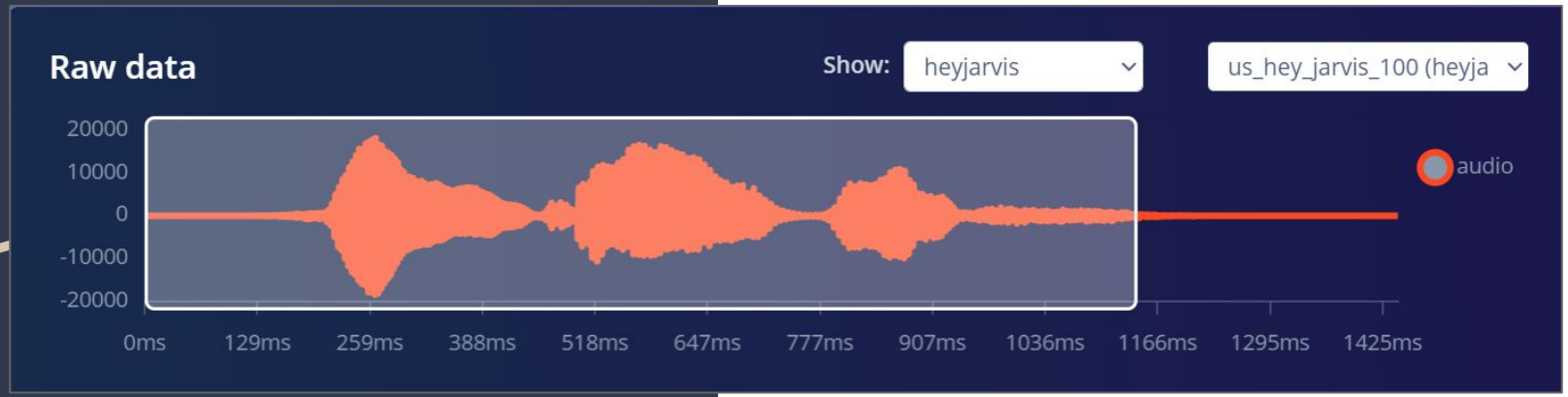
1. Device Setup
2. Data Acquisition
3. MFCC
4. Training the Model
5. Deployment
6. Arduino Development
7. Testing

Data Acquisition



- Data Collection
  - Noise (~15 mins)
  - Unknown (~16 mins)
  - Trigger Word: ‘Hey Jarvis!’ (~14 mins)
- Train/Testing
  - Testing (~30 mins)
  - Training (~45 mins)

Mel Frequency Cepstral Coefficient (MFCC)



Training The Model

Model

Model version: Quantized (int8)

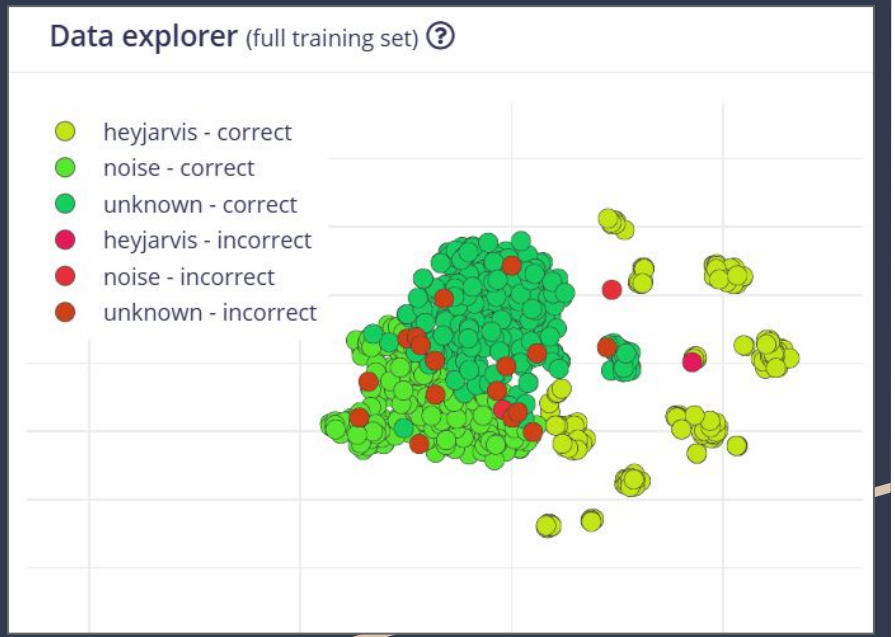
Last training performance (validation set)

ACCURACY
96.8%

LOSS
0.22

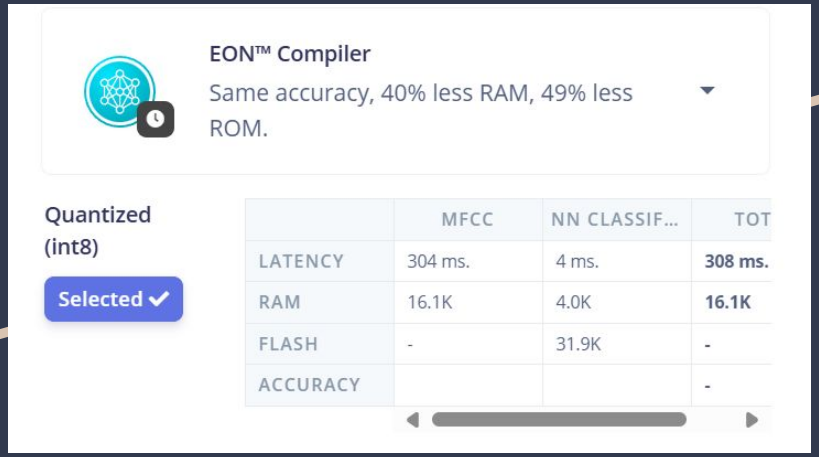
Confusion matrix (validation set)

	HEYJARVIS	NOISE	UNKNOWN
HEYJARVIS	99.2%	0%	0.8%
NOISE	0%	98.4%	1.6%
UNKNOWN	0.6%	6.1%	93.4%
F1 SCORE	0.99	0.96	0.95



- Visualization helps audio preprocessing
- Clusters demonstrate similarities and differences based on features
- Ideal to have tight clusters
- Majority of points are classified correctly

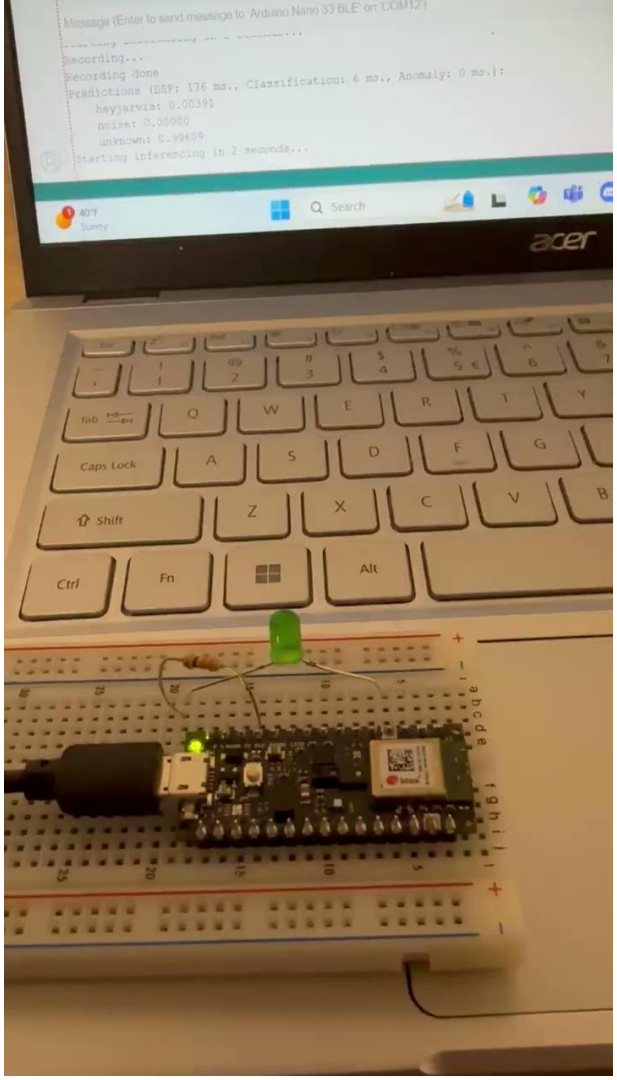
Deployment



	MFCC	NN CLASSIF...	TOT
LATENCY	304 ms.	4 ms.	**308 ms.**
RAM	16.1K	4.0K	**16.1K**
FLASH	-	31.9K	-
ACCURACY			-

- Edge Impulse builds firmware
- Export Arduino library
- Develop C++ code to connect trigger word to LED toggling

Demo (LED)


Message (Enter to send message to 'Arduino Nano 33 BLE' on 'COM12')
Recording...
Recording done
Predictions (DSP: 176 ms., Classification: 6 ms., Anomaly: 0 ms.):
hey jarvis: 0.00391
noise: 0.00000
unknown: 0.99609
Starting inferencing in 2 seconds...
Search
acer

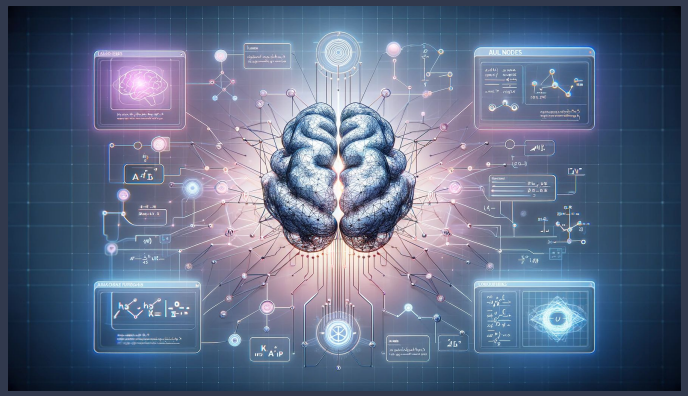
Demo (Output)

Output Serial Monitor ×

Message (Enter to send message to 'Arduino Nano 33 BLE' on 'COM12')

```
Recording...
Recording done
Predictions (DSP: 176 ms., Classification: 6 ms., Anomaly: 0 ms.):
    heyjarvis: 0.00391
    noise: 0.00000
    unknown: 0.99609
Starting inferencing in 2 seconds...
```

Key Takeaways



- For accurate training, the dataset should have relatively equal amounts of keywords, noise, and unknown audio files
- The model size needed to fit the hardware so the dataset had to be adjusted
- The training dataset used in this project was large which increased our model’s accuracy but significantly increased our training time