Music Genre *Classification*

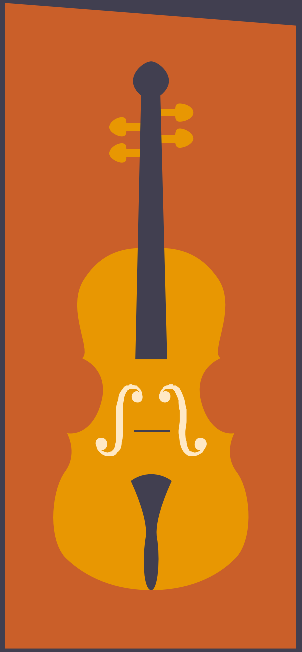
Young Park
General Assembly, DSIR-1019

Agenda

- Problem Statement
- Data
- Feature Extraction
- Modeling
  - Classifiers (Logistic, KNN, Random Forests, SVM, Gradient Boost, XG)
  - * Mel-Spectrogram
  - Convolutional Neural Networks (CNNs)
  - Convolutional Recurrent Neural Networks (CRNNs)
- Model Performance/Evaluation
- Misclassifications
- Conclusions / Future Considerations

Music

Music plays a critical role in our everyday lives. Companies like Spotify, Pandora, Apple, Google, and other major music platforms are constantly looking for new and fresh ways to categorize and classify their music inventory based on users' taste and preference in order to create a more personal experience on their platforms. They want to know what kind of music are you into?



Problem Statement

Can machine learning, with a high level of accuracy, classify/predict various music genres?

GTZAN Genre Collection

- 1,000 audio tracks
- 10 genres, each represented by 100 tracks
- 30 seconds long
- 2000-2001
- Variety of sources (personal CDs, radio, live recordings, and other recording conditions)

Blues	Jazz
Classical	Metal
Country	Pop
Disco	Reggae
Hiphop	Rock

Feature Extraction

Process of computing a compact numerical representation that can be used to characterize a segment of audio

- Timbral texture
- Rhythmic features
- Pitch content

Tempo
Chroma Energy Normalized Statistics (Mean, Std.)
Mel-Frequency Cepstral Coefficients (Mean, Std.)
Spectral Centroid (Mean, Std., Skew)
Spectral Contrast (Mean, Std.)
Spectral Rolloff (Mean, Std., Skew)
Zero-Crossing Rate (Mean, Std., Skew)

Modeling: Part 1

Classifiers

Logistic Regression
(Bagging)

KNN
(Bagging)

Random Forests

SVM

Gradient Boost

XGBoost

SVM Accuracy

Baseline: 10%
State-of-the art: 91%

100%

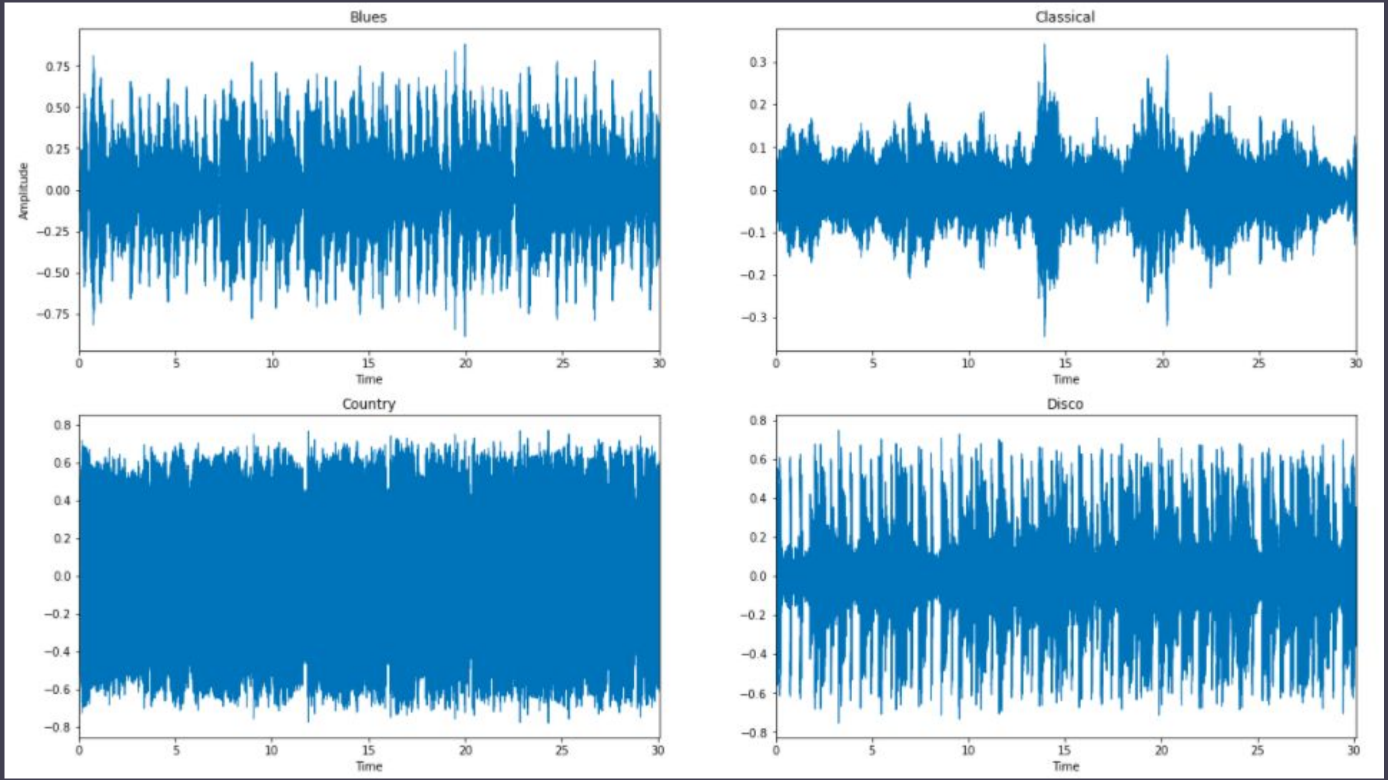
Training

72%

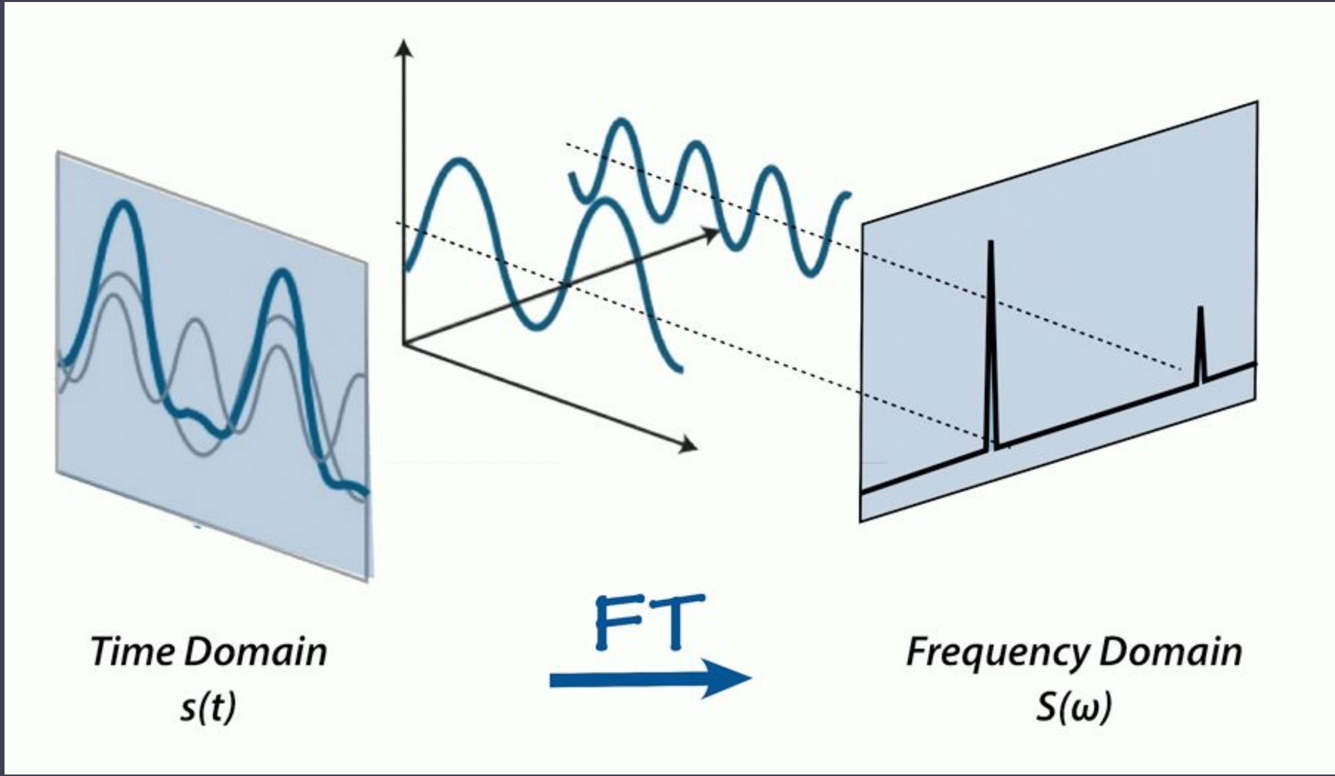
Testing

Modeling: Part 2

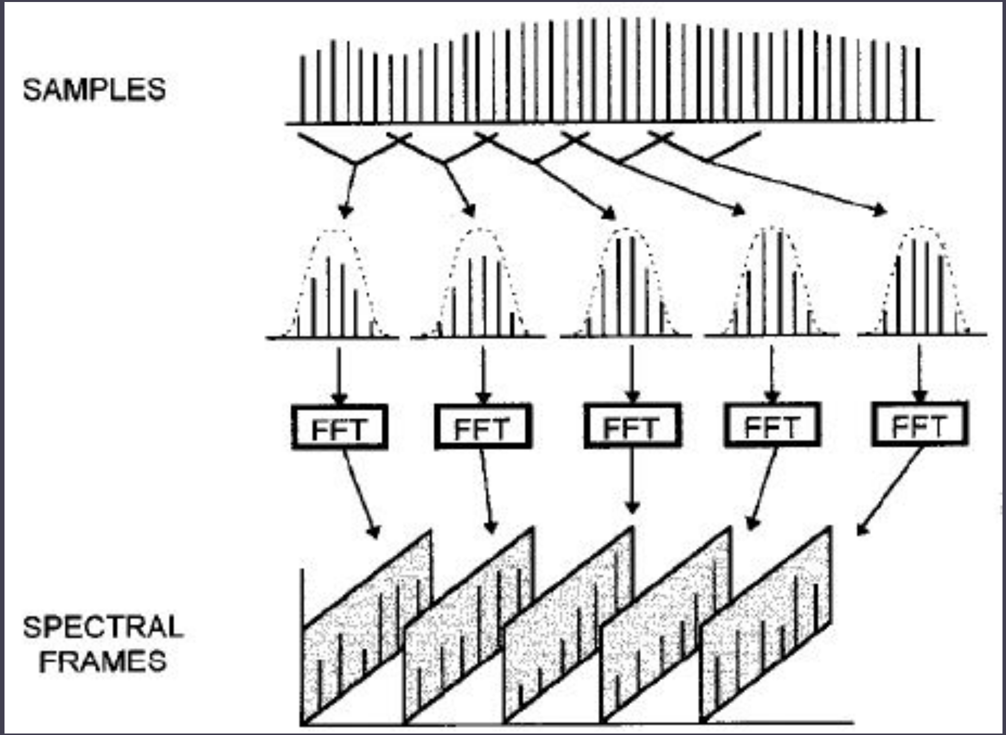
Waveforms


Blues
Classical
Country
Disco
Amplitude
Time
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
0.8
0.6
0.4
0.2
−0.2
−0.4
−0.6
−0.8
0
5
10
15
20
25
30

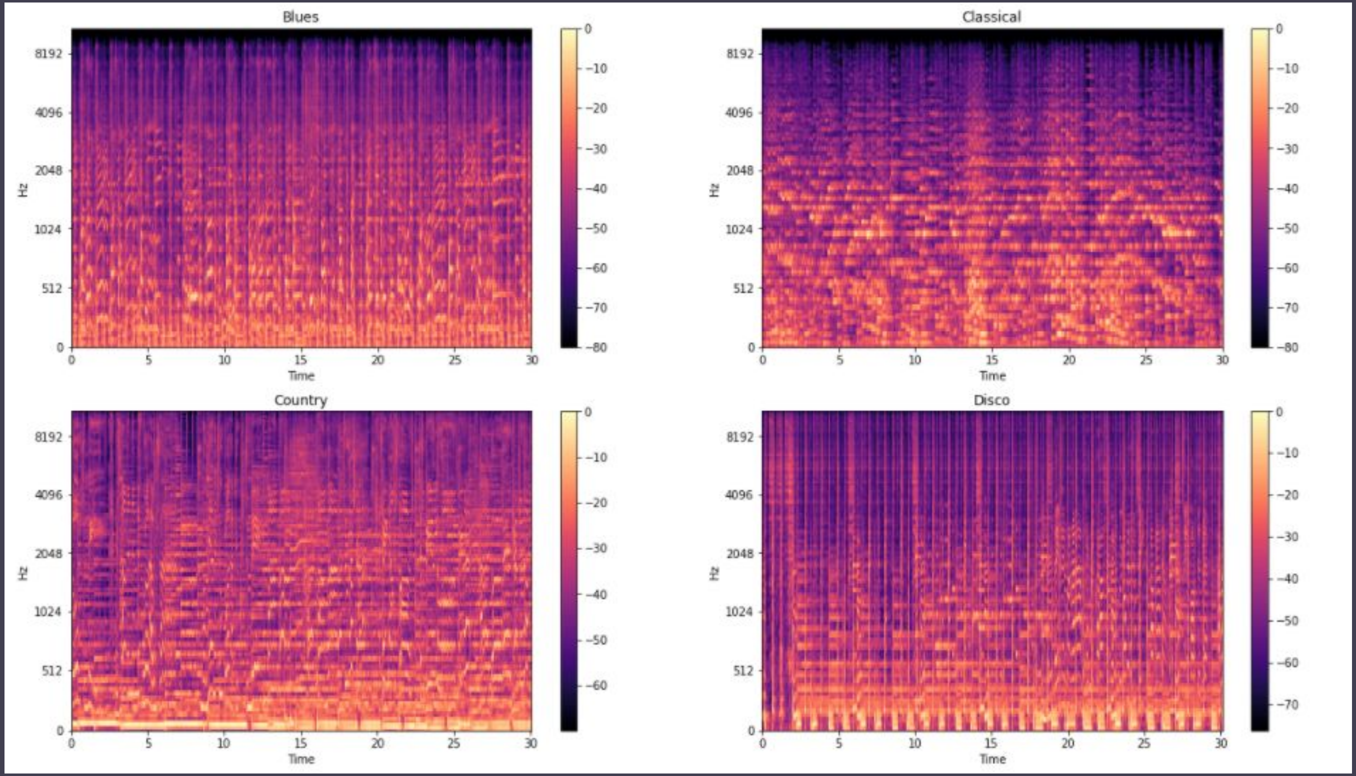
Fourier transform (FFT)



Short-Time Fourier Transform (STFT)



Mel-Spectrogram



Convolutional Neural Network

Baseline: 10%
State-of-the-art: 91%

96.1%

Training

86.1%

Testing

Conv2D

MaxPooling2D

Dense

Dropouts

Softmax

Modeling: Part 3

Convolutional Recurrent Neural Network

Baseline: 10%
State-of-the-art: 91%

98.6%

Training

86.7%

Testing

Conv2D

MaxPooling2D

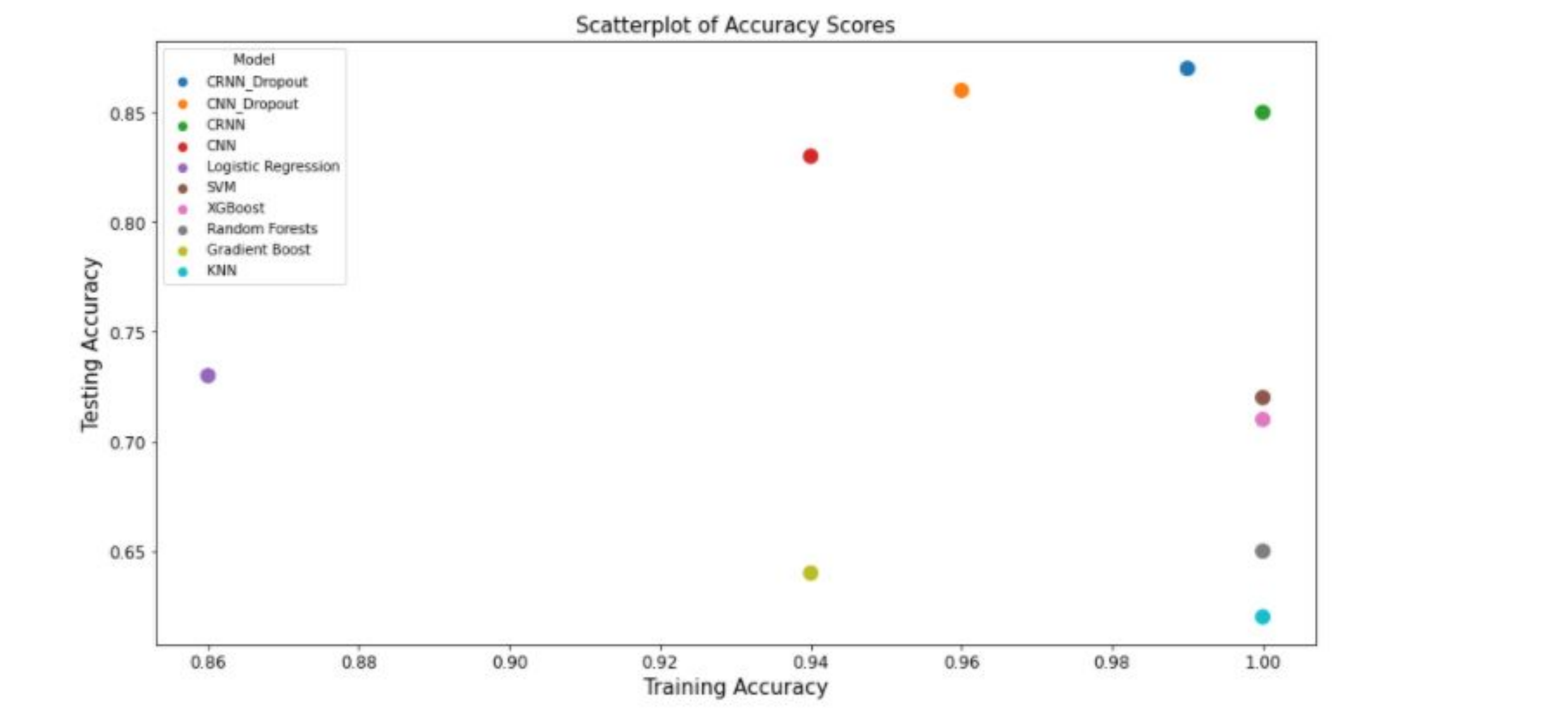
GRU

Dense

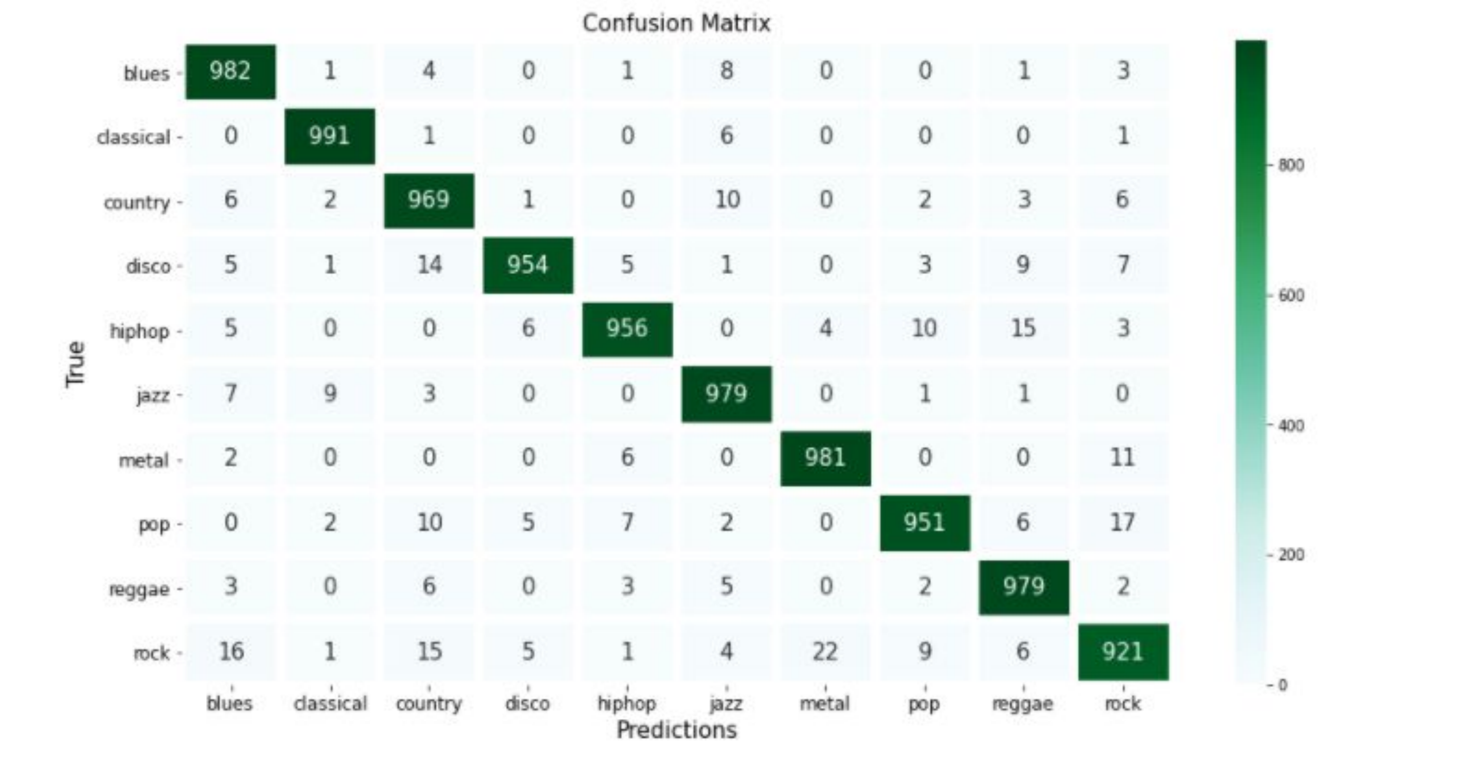
Dropouts

Softmax

Accuracy Score Summary



Confusion Matrix



Misclassifications

Case #1: Jazz misclassified as Classical

Case #2: Rock misclassified as Metal

Conclusions/Future Considerations

- Tendency to overfit across all models. This can be addressed with more data and additional effort on regularization.
- Convolutional Recurrent Neural Network topology, which is gaining popularity, was the best performing model based on accuracy and bias/variance tradeoff.
- GTZAN Genre Collection Dataset is 20 years old. It'd be interesting to deploy this model into production and test it with recent music samples to measure impact on performance.
- Certain songs don't fall "cleanly" into a genre. Instead of assigning to a specific genre, it may be more appropriate to classify them as clusters of genres.
- Group users based on their taste/preference in certain genres and deploy targeted advertisements, study their behaviors for deeper insight, and customize recommendations/layouts for more personal experience.

Publications/Sources:

http://www.cs.cmu.edu/~gtzan/work/pubs/tsap02gtzan.pdf

https://arxiv.org/pdf/1712.08370v1.pdf

https://arxiv.org/pdf/1901.04555.pdf

https://rramnauth2220.github.io/blog/posts/code/200525-feature-extraction.html

https://github.com/subho406/Audio-Feature-Extraction-using-Librosa/blob/master/Song%20Analysis.ipynb

https://www.youtube.com/channel/UCZPFjMe1uRSirmSpznqvJfQ

Thanks

Do you have any questions?

youngathpark1@gmail.com
https://github.com/youngathpark1
www.linkedin.com/in/youngathpark1