GA-DSI-123

Group Project

TIME FREQUENCY ANALYSIS

Chris Landschoot
Elio Marcaccio
Farhad Farahani
Ayako Homma

March 24th, 2023

FCEA ANALYTICS TEAM

CHRIS LANDSCHOOT

ELIO MARCACCIO

FARHAD FARAHANI

AYAKO HOMMA

PROBLEM STATEMENT

Spotify has hired us at FCEA Analytics as they want to develop a music genre classification model that can better organize the works of smaller or independent artists who may not accurately classify their music.

To achieve this, our team will conduct studies using web scraping data from Spotify and audio files from GTZAN dataset transformed into images to build a classification model. We aim to develop the best model to predict the genre of music.

TODAY'S AGENDA

01 Background

02 Study 1 (audio feature data)

SPOTIFY API

RESEARCH PROCESS

- DATA COLLECTION
- DATA PREPROCESS
- DATA MODELING & EVALUATION

03 Study 2 (audio files)

GTZAN DATASET

04 Conclusion & Recommendations

01

BACKGROUND

BACKGROUND

- Spotify face some challenges when it comes to classify music genre:
  - Genre Ambiguity: Many songs contain elements of multiple genres, making it difficult to assign them to a specific category.
  - Genre Evolution: Music genres evolve over time, with new sub-genres emerging, making it challenging to keep up with the latest trends and accurately classify songs.

We at FCEA Analytics is here to build a model that ensures accurate classification and organization of their vast libraries of songs.

02

Study 1 (audio feature data):

SPOTIFY WEB API

DATA COLLECTION

Data source: Spotify Web API (initially)

- Rate limiting impacted the possibility of creating a fresh data set
- Implemented: Cameron Watt's Dataset on the same API

Datasets

- 32,200 rows x 11 columns
- 8,576 rows x 21 columns once cleaned
- Features: danceability, energy, key, loudness, mode, acousticity, instrumentalness, liveness, balanca, tempo

Music Genre:

Dance Pop	Contemporary Country	Alternative Metal	Alternative Hip Hop
Album Rock	Alt Rock	Adult Standards	Alt Dance

DATA PREPROCESS

- Removal of genres with less than 500 appearances
- Removal of unknown columns
- Encoding of key as a categorical variable
- Transforming the list of genres into one genre per entry

CLASSIFICATION METHODS

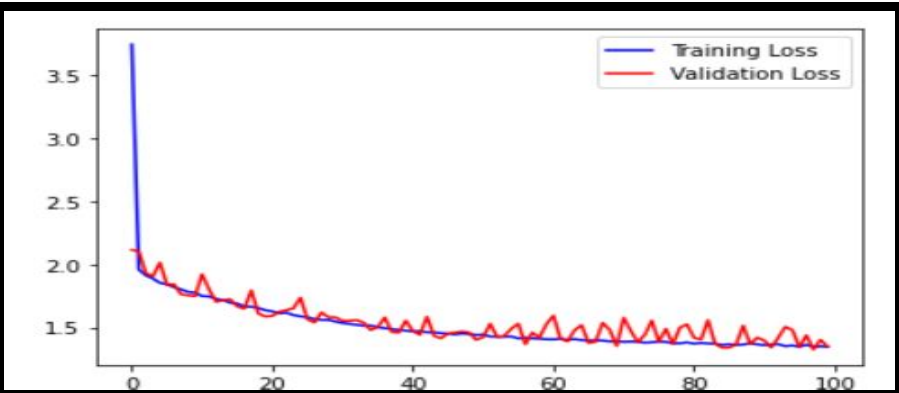
01

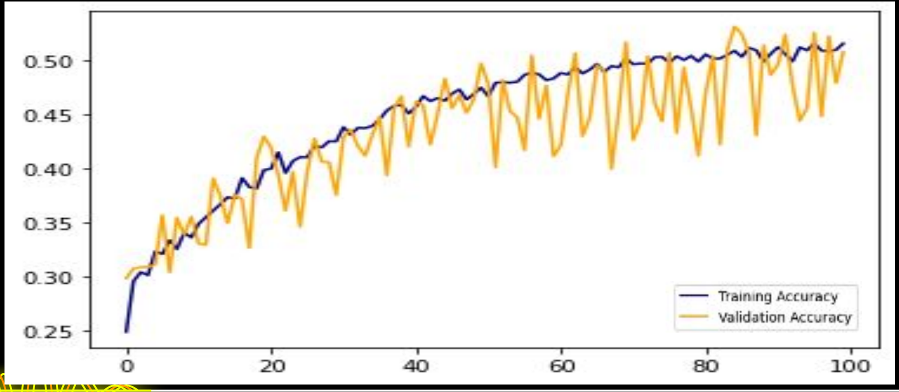
LOGISTIC REGRESSION

02

DENSE NEURAL NETWORK

DENSE NEURAL NETWORK





One input layer, one hidden layer, one output layer:

- 30/ 30/ 8 neurons
- 'ReLU' activation

Softmax output activation

100 epochs

Test Loss: 1.347
Accuracy: 0.507

Study 1 (audio feature data): Spotify Web API

MODEL PERFORMANCE EVALUATION

STUDY 1: SPOTIFY WEB API	
Model	**Test Accuracy**
Logistic Regression	.479
Dense Neural Network	.507

Baseline: .293

03

Study 2 (audio files):

GTZAN DATASET

DATA COLLECTION

Data source: GTZAN Dataset From Kaggle

Genres original

- 10 genres with 100 audio files per each genre
- Each having a length of 30 seconds (wav.file) and classified as one of the following genres:

Music Genre:

Blues	Classical	Country	Disco	Hiphop
Jazz	Metal	Pop	Reggae	Rock

DATA PREPROCESS

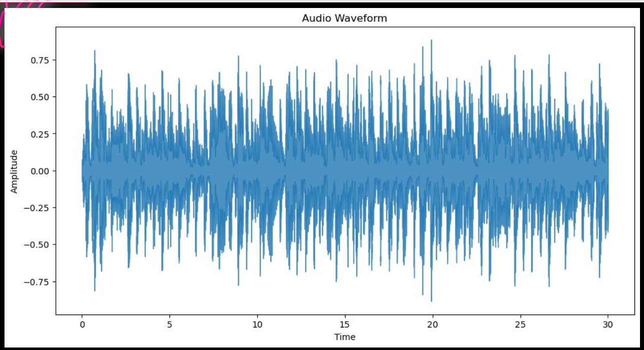
WHAT IS LIBROSA?

Librosa is used for analyzing processing audio signals, particularly for music and speech-related applications. It provides a variety of tools including:

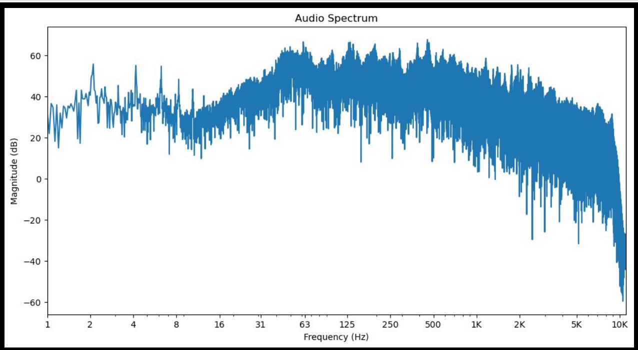
- Loading and playing audio files in various formats (WAV and MP3)
- Extracting various features from audio signals (spectrograms)
- Visualizing audio data and features (waveform plots)
- Transforming audio data for machine learning algorithms for music analysis and classification tasks

Study 2(audio files): GTZAN Dataset

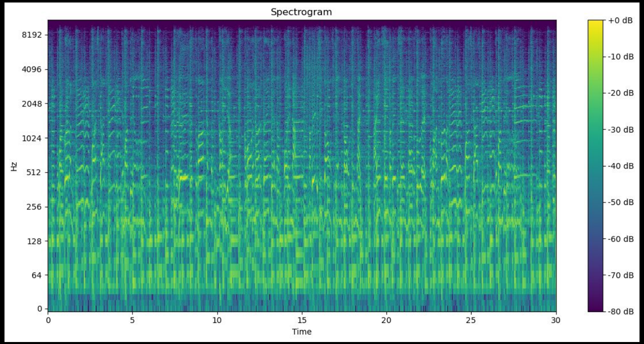
SOUND VISUALIZATION



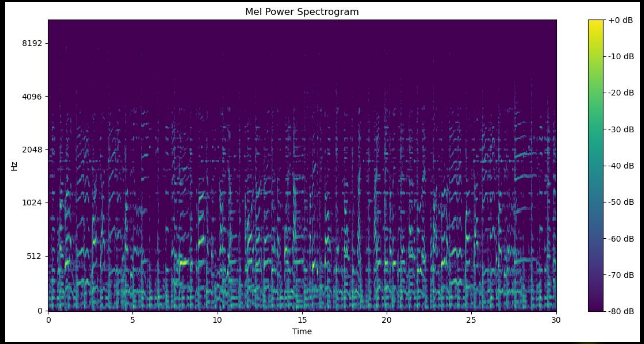
Waveform



Spectrum



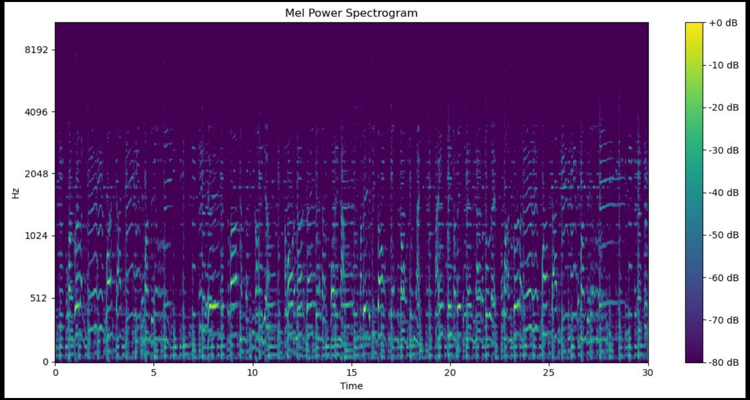
Spectrum



Mel Spectrogram

WHAT IS THE MEL SPECTROGRAMS?

- The mel spectrogram is a visual representation of the frequency spectrum of an audio signal that has been transformed using a mel-scale.
- The mel scale is a non-linear frequency scale that is more perceptually relevant for human hearing than the linear frequency scale used in traditional spectrograms.



Study 2(audio files): GTZAN Dataset

CLASSIFICATION METHODS

01

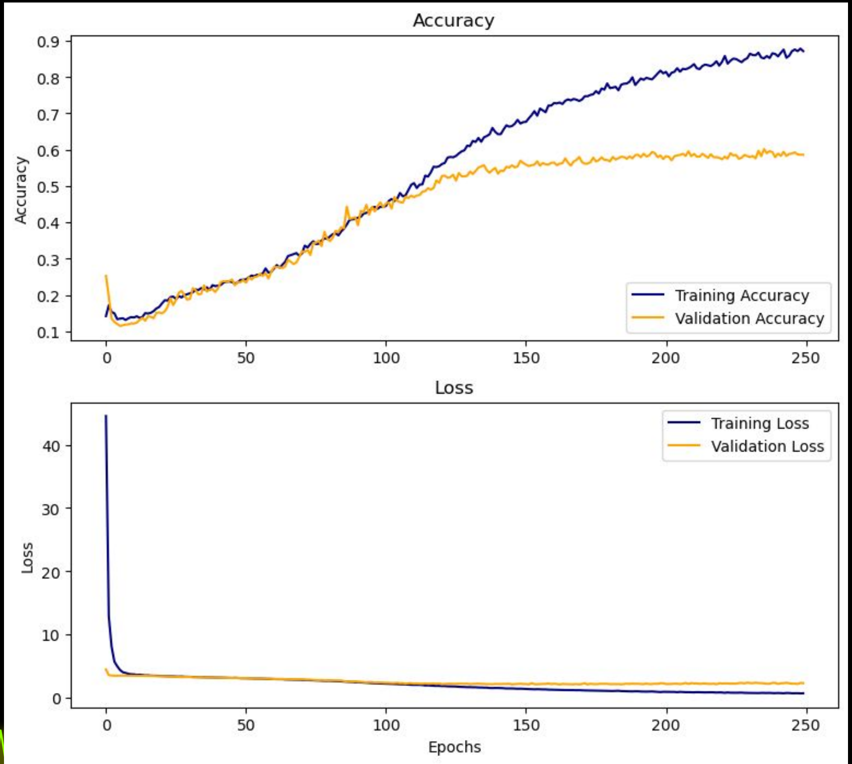
DENSE
NEURAL NETWORK

02

CONVOLUTIONAL
NEURAL NETWORK

Study 2(audio files): GTZAN Dataset

DENSE NEURAL NETWORK



Multiple dense Layers:

- 512 / 256 / 64 neurons
- ‘ReLU’ activation

Dropout regularization (0.3)

L2 regularization

Output layer:

- 10 neurons
- ‘Softmax’ activation

Test Loss: 2.415
Accuracy: 0.578

CNN : KEY FEATURES THAT HELP IMPROVE OUR MODEL PERFORMANCE

BATCH NORMALIZATION

IMPROVED OUR MODEL ACCURACY BY 5%

DROPOUT

IMPROVED OUR MODEL ACCURACY BY 6%

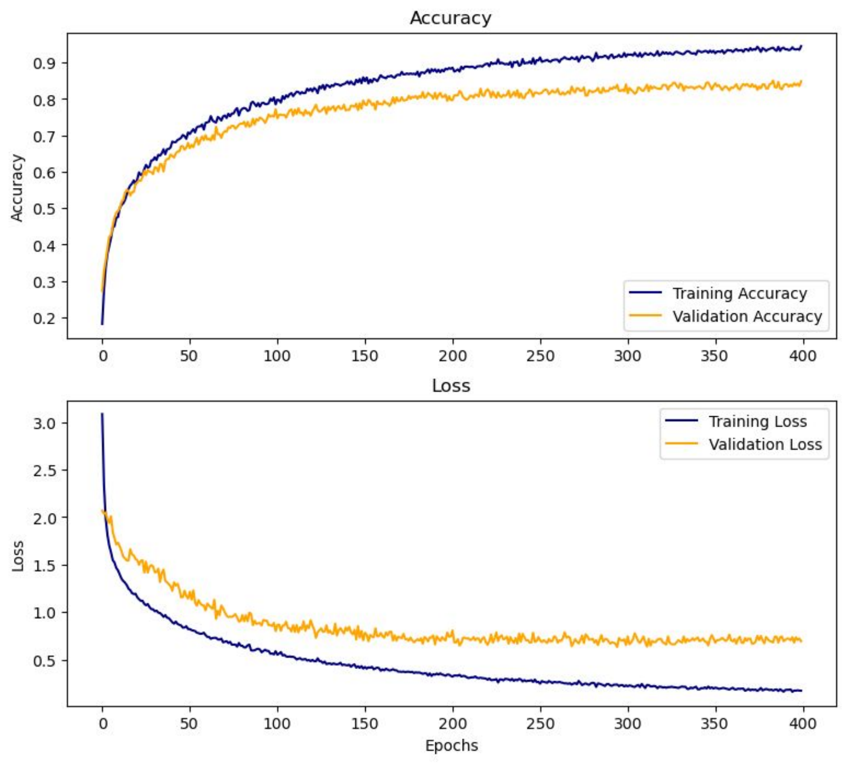
EPOCHS

INCREASED EPOCHS FROM 50 TO 400

"IMAGE" DATA AUGMENTATION

IMPROVED OUR MODEL BY SENDING AUDIO FORWARD AND REVERSED AUDIO THROUGH THE MODEL

CONVOLUTIONAL NEURAL NETWORK



Multiple convolutional blocks:

- Conv2D: 32 / 64 / 64 filters
- Kernel size (3x3 / 3x3 / 2x2)
- ‘ReLU’ activation
- Maxpooling2D layer
- Flatten layer

Dense layer

- 128 neurons
- ‘Relu’ activation
- Dropout regularization (0.5)

Output layer:

- 10 neurons
- ‘Softmax’ activation

Test Loss: 0.634
Accuracy: 0.838

Study 2(audio files): GTZAN Dataset

MODEL PERFORMANCE EVALUATION

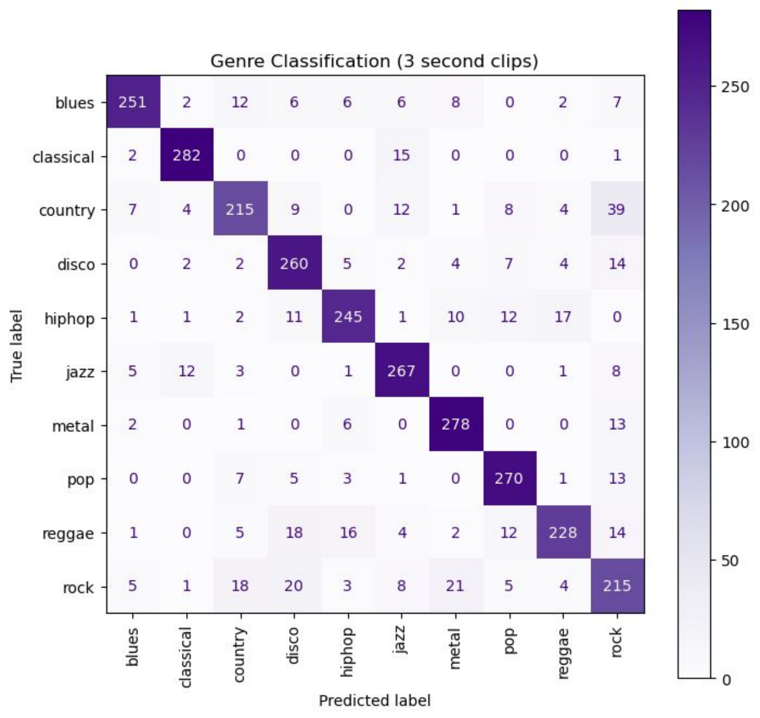
STUDY 2: GTZAN DATASET	
Model	**Test Accuracy**
DNN	0.578
CNN (without regularization)	0.691
CNN (with regularization)	0.801
CNN (with regularization and data manipulation)	**0.838**

Baseline: 0.100

Study 2(audio files): GTZAN Dataset

Music Genre		Blues	Classical
Hiphop	Jazz	Hiphop	Jazz
Metal	Pop	Reggae	Rock

CONFUSION MATRIX



04

CONCLUSION & RECOMMENDATIONS

CONCLUSION

- In this project, we explored different audio data and built music genre classification models.
- The model that used pre-engineered features from Spotify's API achieved an accuracy of 0.507. However, using audio files and converting them into images resulted in better classification models. In fact, the top performing model was a convolutional neural network (CNN), achieving an accuracy of 0.838.
- Also, CNN model performance was significantly improved by using batch normalization, dropout, increasing the number of epochs and flipping the audio 'image'.
- This project demonstrated the ability to learn complex audio patterns in audio features and build a successful model for music genre classification.

RECOMMENDATIONS

- Here are some additional areas that we can explore to optimize the performance of our music genre classification models:
    - Increase the size of the datasets and conduct additional data augmentation techniques such as altering the tempo or adding background noise
    - Experiment with different hyperparameters such as batch size, dropout rate, and number of epochs and layers
    - Test different lengths of audios or collect more diverse datasets
- The best performing model could be leveraged for recommendation systems to provide personalized music recommendations to Spotify users.

THANKS!

Any questions?