LARGE-SCALE MEDIA ANALYTICS

Master of Science in Signal Theory and Communications
TRACK: Signal Processing and Machine Learning for Big Data
Large-Scale Media Analytics

Fernando Marcos Macías

Víctor Gutiérrez García

Contents

Dataset preprocessing

Before

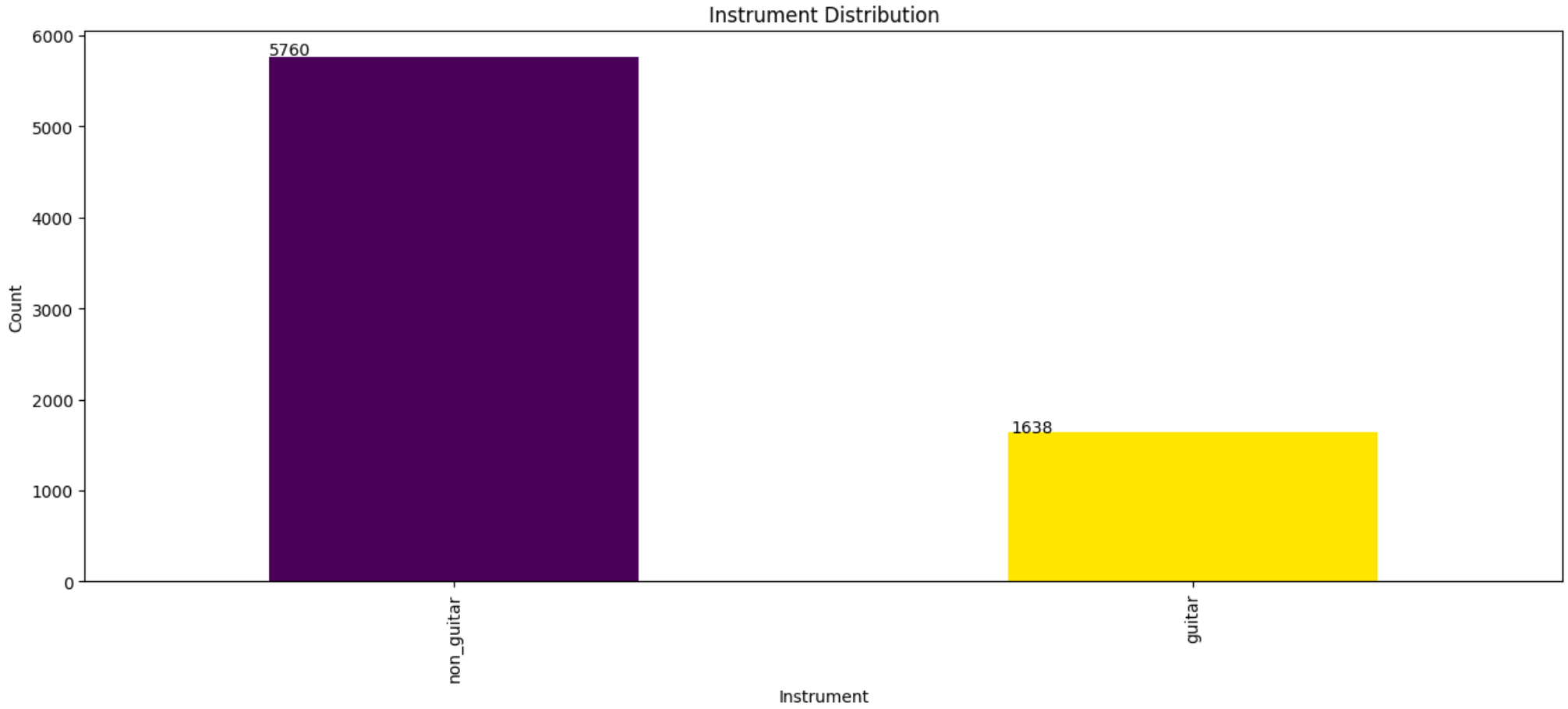
```
,file_name,label,instrument,type_of_sound
0,0000.wav,C:maj,non_guitar,Jazz_Organ
1,0001.wav,C:maj/3,non_guitar,Jazz_Organ
2,0002.wav,C:maj/5,non_guitar,Jazz_Organ
3,0003.wav,C#:maj,non_guitar,Jazz_Organ
4,0004.wav,C#:maj/3,non_guitar,Jazz_Organ
5,0005.wav,C#:maj/5,non_guitar,Jazz_Organ
6,0006.wav,D:maj,non_guitar,Jazz_Organ
7,0007.wav,D:maj/3,non_guitar,Jazz_Organ
8,0008.wav,D:maj/5,non_guitar,Jazz_Organ
9,0009.wav,D#:maj,non_guitar,Jazz_Organ
10,0010.wav,D#:maj/3,non_guitar,Jazz_Organ
11,0011.wav,D#:maj/5,non_guitar,Jazz_Organ
```

After

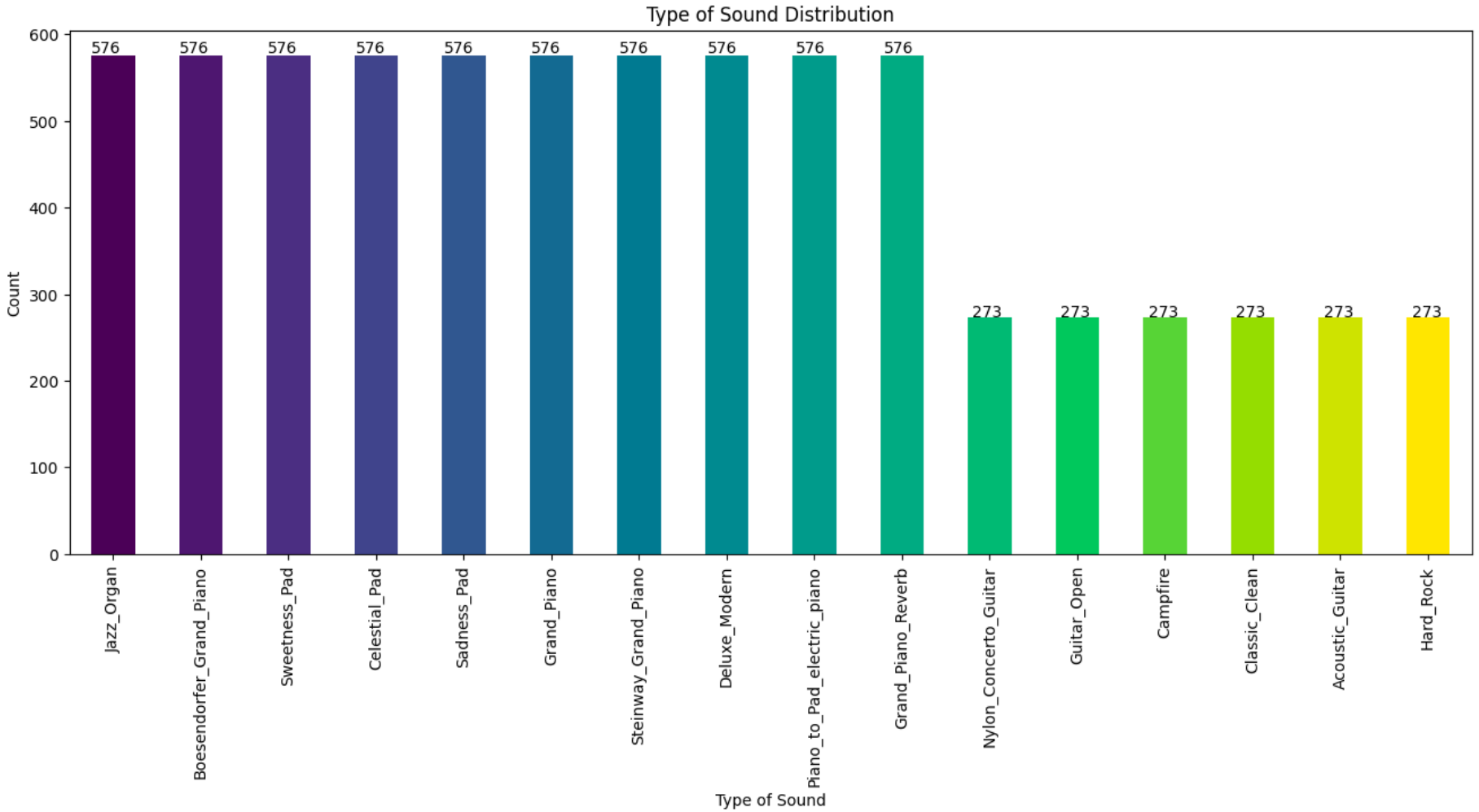
```
file_name,label,instrument,type_of_sound,root_note,chord_type
0000.wav,C:maj,non_guitar,Jazz_Organ,C,maj
0001.wav,C:maj/3,non_guitar,Jazz_Organ,C,maj/3
0002.wav,C:maj/5,non_guitar,Jazz_Organ,C,maj/5
0003.wav,C#:maj,non_guitar,Jazz_Organ,C#,maj
0004.wav,C#:maj/3,non_guitar,Jazz_Organ,C#,maj/3
0005.wav,C#:maj/5,non_guitar,Jazz_Organ,C#,maj/5
0006.wav,D:maj,non_guitar,Jazz_Organ,D,maj
0007.wav,D:maj/3,non_guitar,Jazz_Organ,D,maj/3
0008.wav,D:maj/5,non_guitar,Jazz_Organ,D,maj/5
0009.wav,D#:maj,non_guitar,Jazz_Organ,D#,maj
0010.wav,D#:maj/3,non_guitar,Jazz_Organ,D#,maj/3
0011.wav,D#:maj/5,non_guitar,Jazz_Organ,D#,maj/5
0012.wav,E:maj,non_guitar,Jazz_Organ,E,maj
```

- Deleted index of the dataframe.
- Added two new columns, root note and chord type
- First approach: classify root note (12 classes)

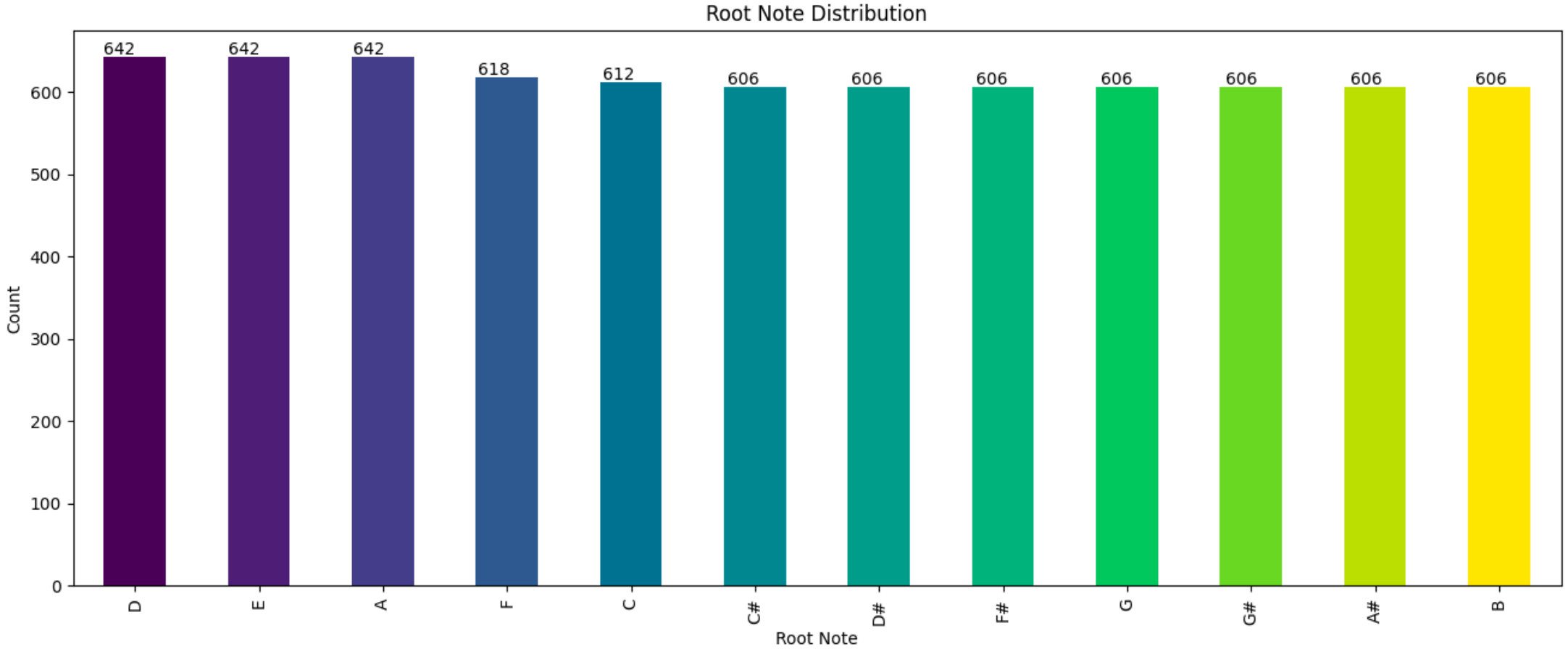
Exploratory Data Analysis



Exploratory Data Analysis

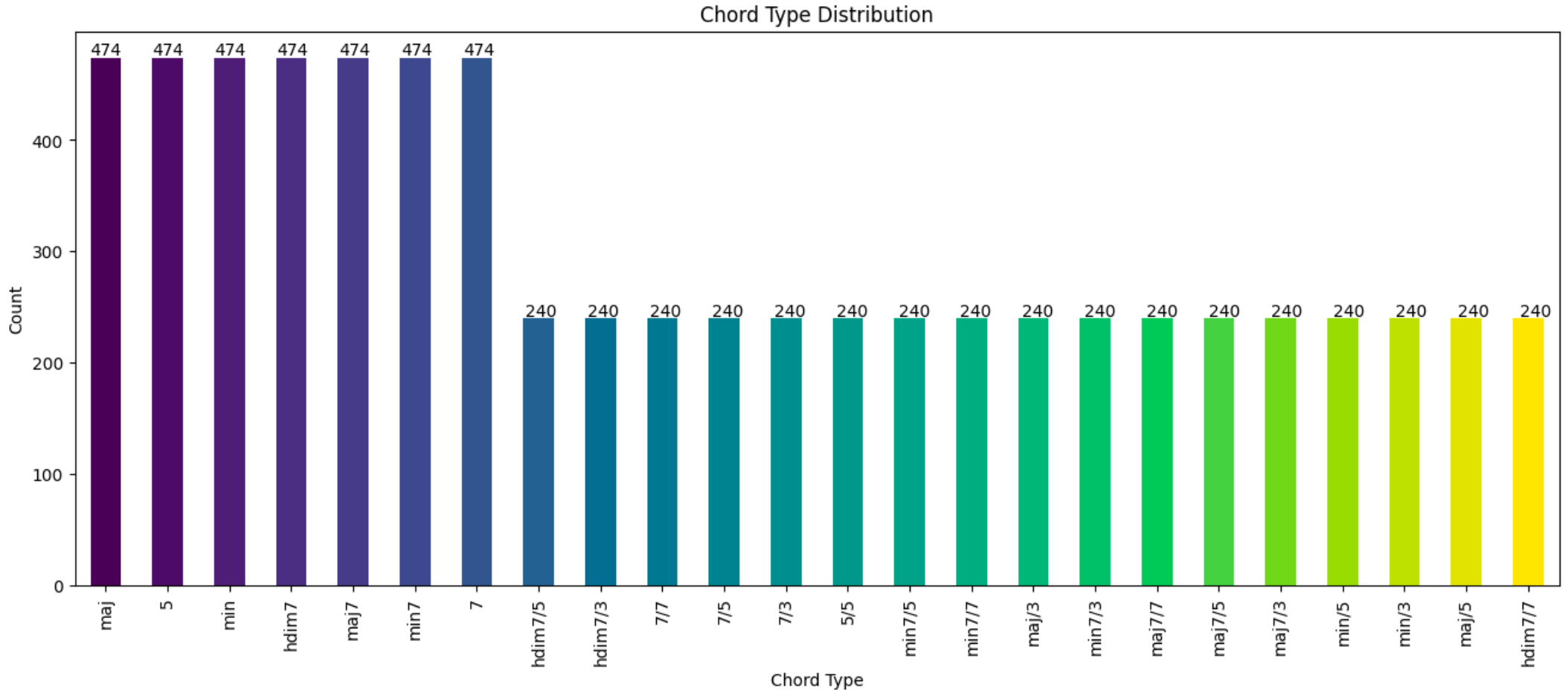


Exploratory Data Analysis

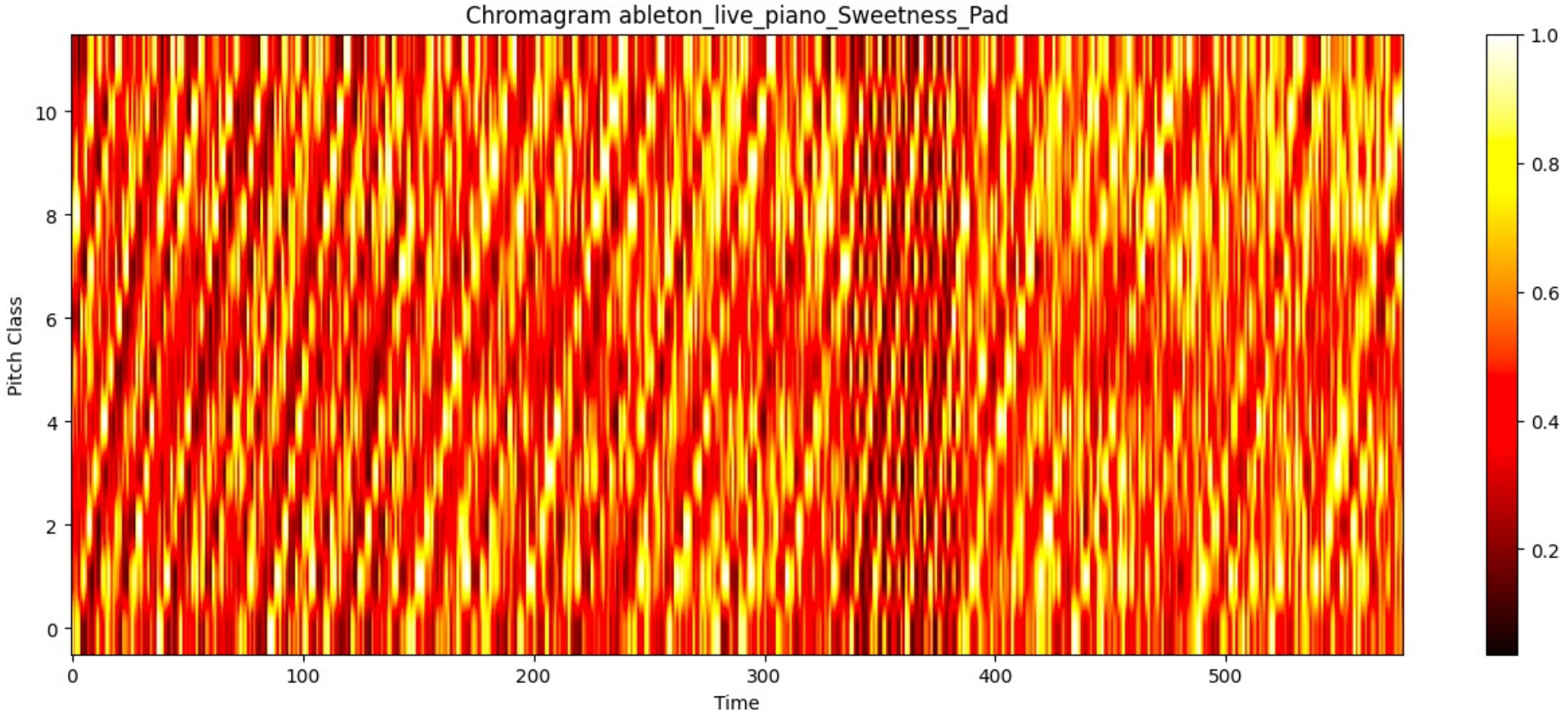


Something strange happens to the dataset, there are different distributions of root notes for the different chords. The imbalance is not too significant, but it is present in the dataset.

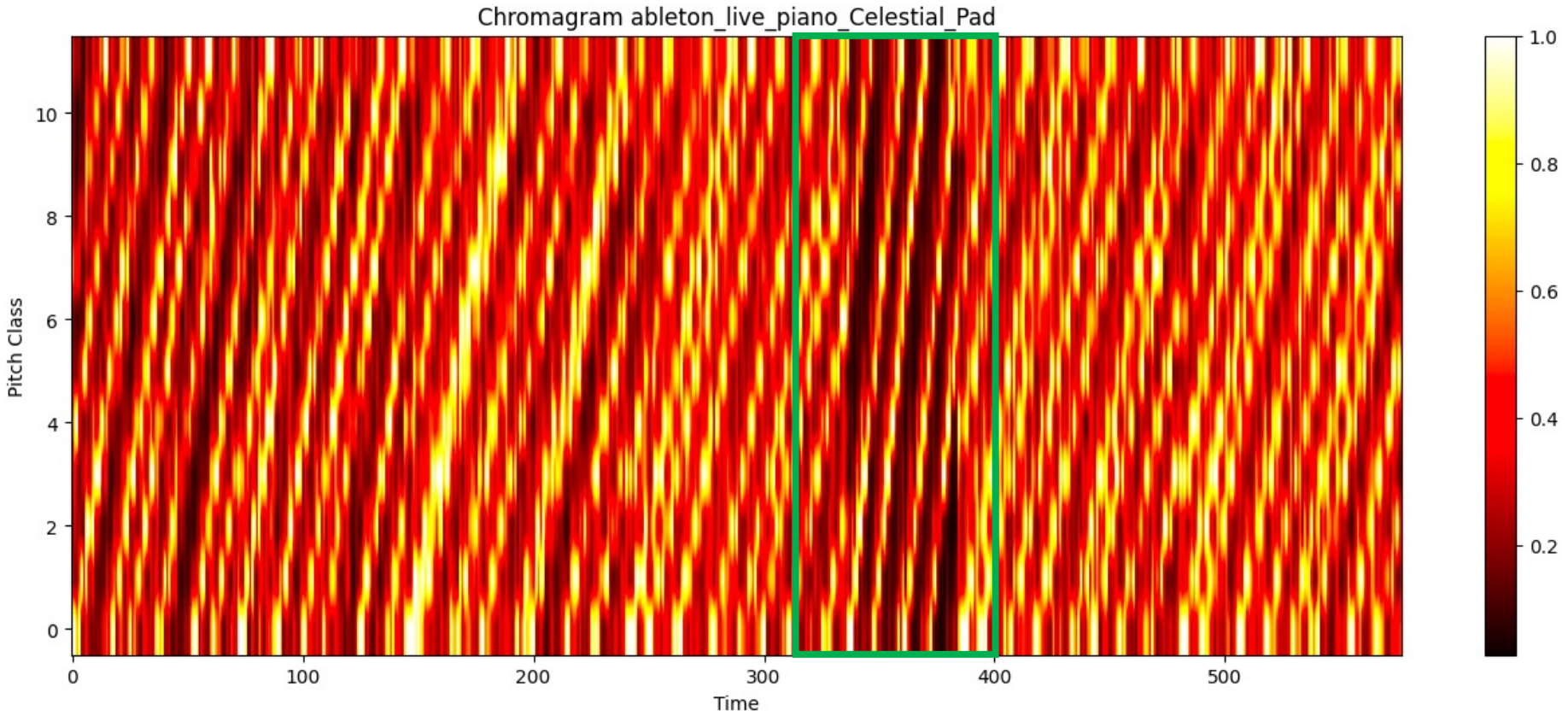
Exploratory Data Analysis



First features – Chromagram

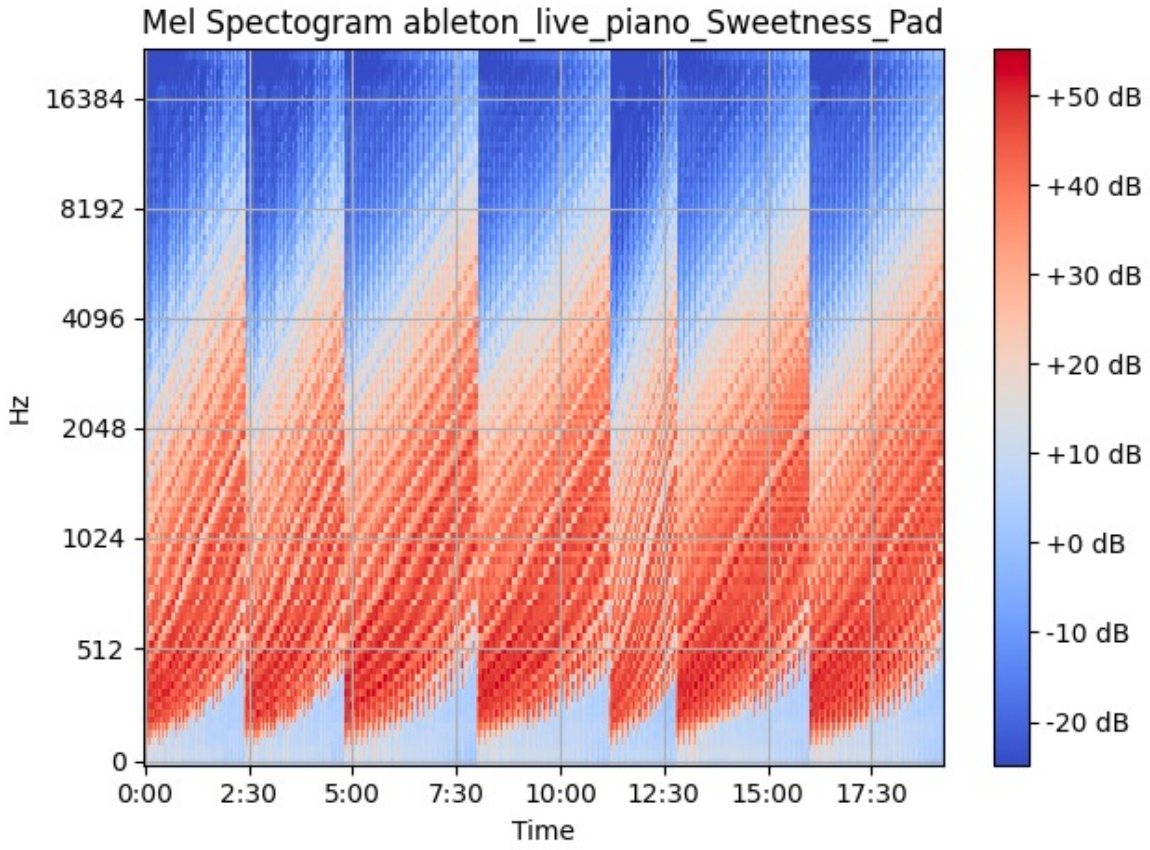


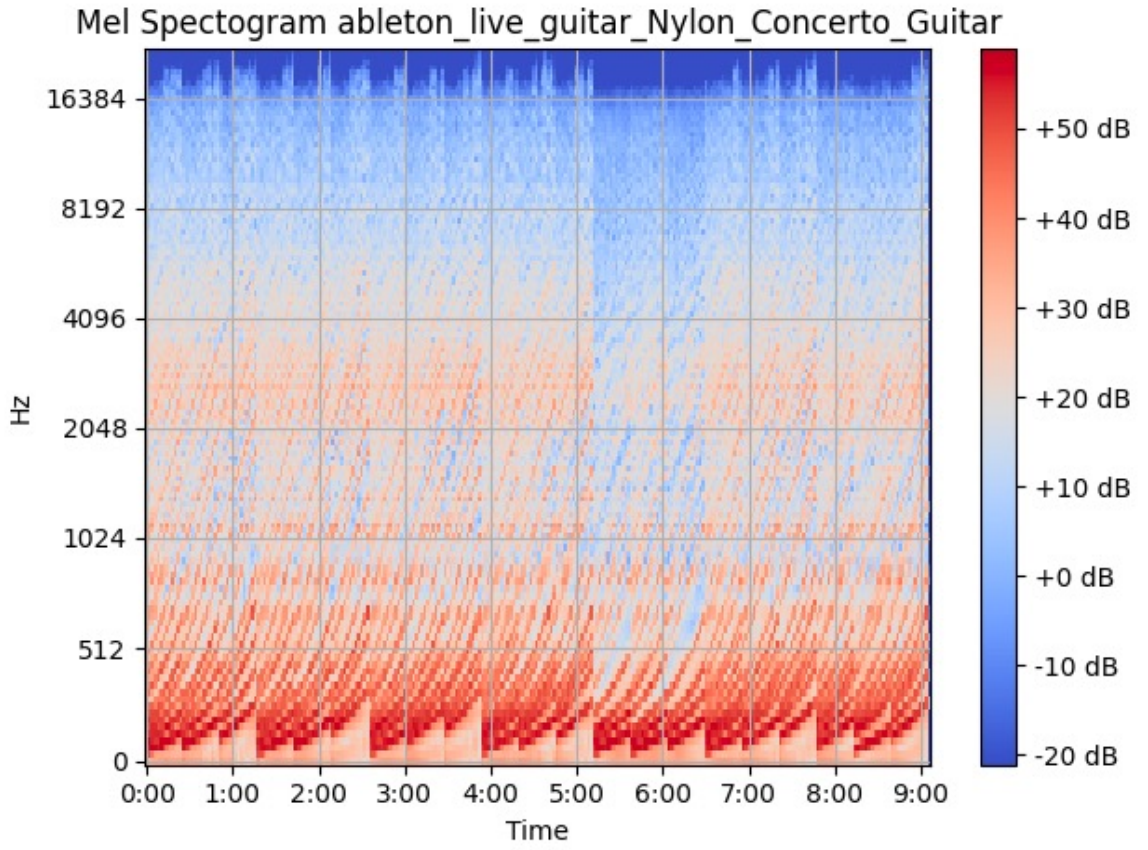
First features – Chromagram



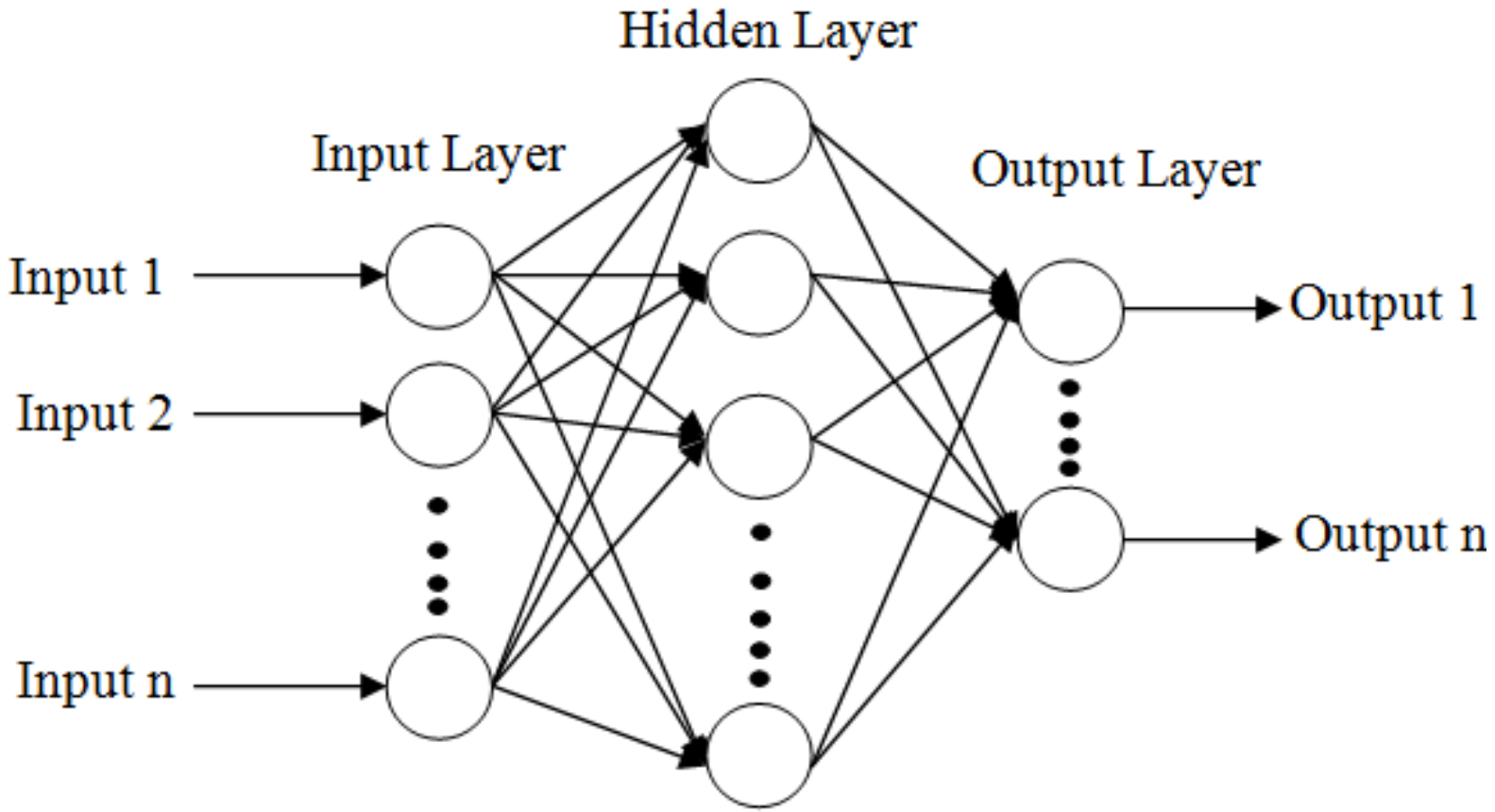
We can distinguish by eye the 5ths!!! Because they have lower energy (2 notes vs. traids and quatriads)

First features – Spectrogram



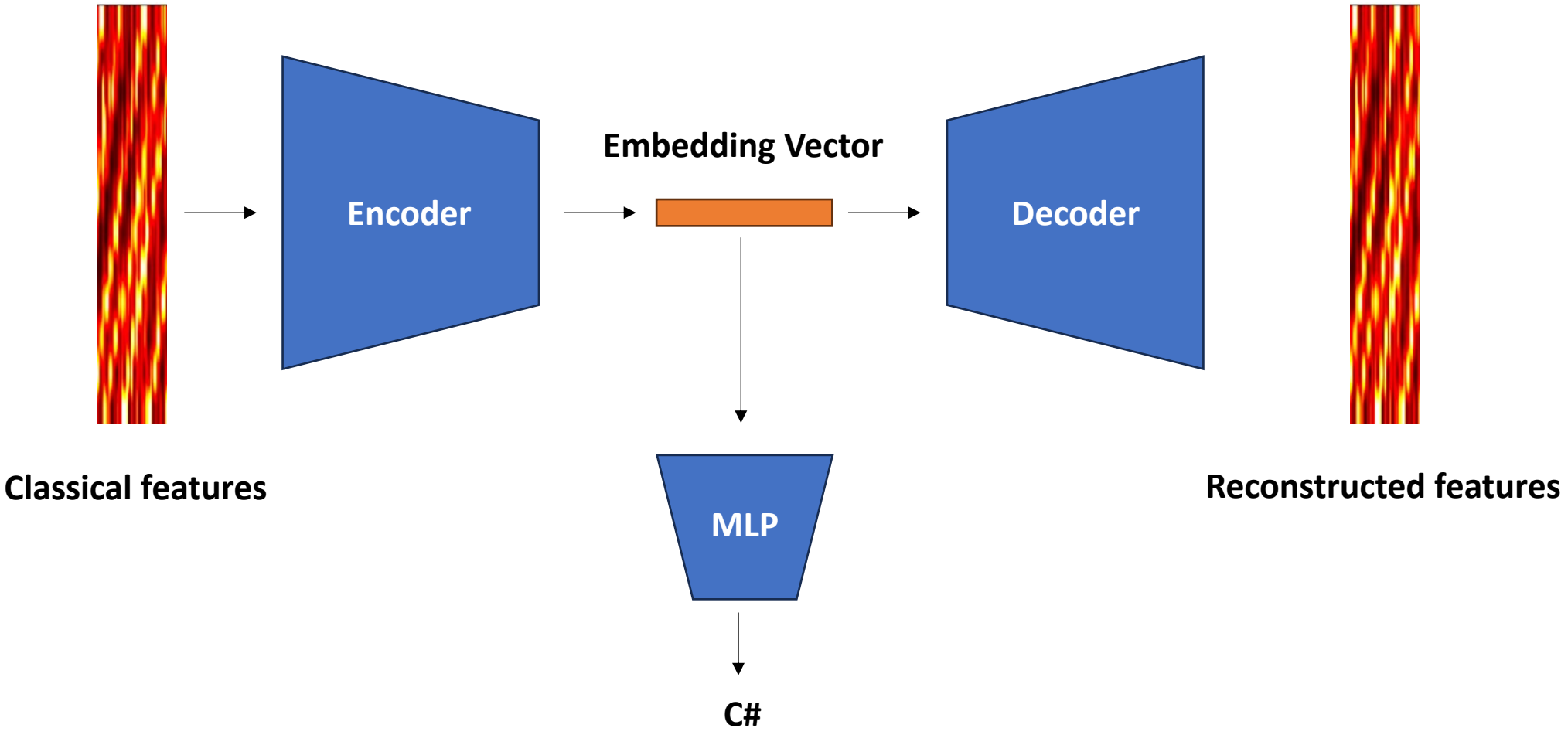


Chordifier MLP v1.0



Chordifier MLP v1.0

Architecture proposal:



Chordifier MLP v1.0

PyTorch Lightning

- High-level management of **pytorch** functionalities
- **Dataset class** incorporates feature extraction in a memory-efficient way

```python
class CustomDataset(Dataset):
    def __init__(self, dataframe, root_dir, feature = 'Chromagram',transform=None):
        self.dataframe = dataframe['file_name']
        self.root_dir = root_dir
        self.feature = feature
        self.transform = transform
        self.label = dataframe['root_note']

    def __len__(self):
        return len(self.dataframe)

    def __getitem__(self, idx):
        audio_file = os.path.join(self.root_dir, self.dataframe.iloc[idx, 0])

        if self.transform:
            audio_data = self.transform(audio_data)
        x, sr = librosa.load(audio_file, sr=None)

        return self.label[idx], self.get_features(x, sr, self.feature)

    def get_features(self, x, sr, feature='Chromagram'):

        rerturned_feature = np.empty((0, 0))
        hop_length = int(44.1e3*2)
        if feature == 'Chromagram':
            n_chroma = 12
            n_octaves =7
            rerturned_feature = librosa.feature.chroma_cqt(y=x, sr=sr, n_chroma=n_chroma, n_octaves=n_octaves, hop_length=hop_length)

        elif feature == 'Mel Spectrogram':
            n_mels = 128
            n_fft = hop_length
            rerturned_feature = librosa.feature.melspectrogram(y=x, sr=sr, n_mels=n_mels, n_fft=n_fft, hop_length=hop_length)

        else: pass # Implement other features

        return rerturned_feature
```

SSR

Questions?