

Projeto Final: Emoções através de musicas

Alex Junio Ribeiro Campos

2017004555

TinyML - Aprendizado de Máquina aplicado para dispositivos IOT embarcados.

1 Resumo

O Projeto final da disciplina de TinyML tem como objetivo de colocar em prática e testar todos os recursos aprendidos ao longo do semestre. Neste projeto foi abordado o tema de detecção de emoções através das melodias das músicas, em que, através do trecho analisado é possível classificar a música entre 3 classes de sentimentos possíveis, sendo eles Passionate (apaixonada), Cheerfull (alegria), e Wistful (melancolia).

2 Introdução

A música está ao nosso redor. Está na rua e no ônibus, no elevador, na academia, no rádio e na TV, na igreja, nos supermercados, nos pubs e, surpreendentemente, na Internet. Onde quer que o homem esteja, há música. A música esteve, está e sempre estará presente na vida das pessoas, tanto a nível individual como social, seja nas dimensões cultural, lúdica, religiosa ou profissional da existência. Como forma de festa, a música sempre acompanhou os momentos festivos do homem. Na expressão da religiosidade humana, a música é sinônimo de oração e aproxima o homem do transcendente e do infinito. Nas atividades esportivas, a música pode ser usada para manter o atleta motivado, relaxá-lo ou ajudá-lo a se concentrar. A música é a forma de comunicação mais eloquente, exprimindo “aquilo que não se pode traduzir em palavras e aquilo que não se cala” (Victor Hugo). Associamos a música aos momentos mais singulares das nossas vidas e a música faz parte do nosso imaginário individual e social [1]. Assim, de modo a preservar uma parte tão importante do nosso cotidiano e do cotidiano dos nossos antepassados, foi criado o Music Information Retrieval (MIR), é uma área de pesquisa fortemente interdisciplinar que evoluiu da necessidade de gerenciar enormes coleções de música digital para “preservação, acesso, pesquisa e outros usos” [2]

3 Dataset

Uma vez que o tema da pesquisa foi definido, a próxima etapa do projeto foi escolher um dataset que comportasse essas expectativas, de se trabalhar com musicas envolvendo as emoções humanas. Desse modo, o dataset escolhido foi: Multi-modal MIREX Emotion Dataset, no qual contém 903 clipes de áudio (30 segundos), rganizados em cinco pastas (clusters) e subpastas (representando seus rótulos / subcategorias), 764 letras e 193 midis. Desse modo, apesar de os banco de dados conter diversos tipos fontes de musicas como audio, letra, video, para esta analise foi utilizado somente o o conjunto com 903 audios clueterizado em suas subcategorias, afim de classificar as emoções através da melodia.

Assim, foi realizada uma pequena amostra dentro do conjunto de dados apresentado, e foram escolhidas somente 3 classes sendo elas:

- Passionate - Paixão
- Wistful - Melancolico
- Cheerfull - Alegre

Com isso, o conjunto de dados apos esta amostra ficou com cerca de 100 musicas, divididos entre estas três categorias. Na figura 1 é possível ver o conteúdo presente no dataset.

```
Dataset following the MIREX Mood Classification dataset.
More details at the website.

Important notes:
- The dataset consists of:
  *) 903 30 second clips. The files are organized in five
     folders (clusters) and subfolders
  (representing their labels/subcategories).
  *) 764 lyric files in txt format
  *) 196 midi files

- The three formats follow the same file naming scheme:
  *) Eg., file 004.mp3, 004.txt and 004.mid correspond to
  the same song.

- Due to size and bandwidth constraints the files are in
  mp3, they should be converted to the same format used in
  MIREX:
  22 KHz
  Sample size: 16 bit
  Number of channels: 1 (mono)
  Encoding: WAV

- 7 additional files are included representing:
  * clusters.txt - the cluster of each clip represented by
  one line in dataset.txt.
  * categories.txt - similar to clusters.txt but using the
  respective subcategories.
```

Figure 1: Conteúdo do Dataset.

4 Tratamento dos dados

Como os dados baixados no banco de dados eram em sua maioria no fomato .mp3, que não é aceito na plataforma do Edge Impulse, foi necessário realizar um pré processamento desses dados. Assim, utilizando o software” Free Áudio Converter”, foi possível transformar o formato dos arquivos em .mp3 em .wav, sendo um formato aceito na plataforma do Edge Impulse. Na figura 2 é possível observar um exemplo de tela do programa.

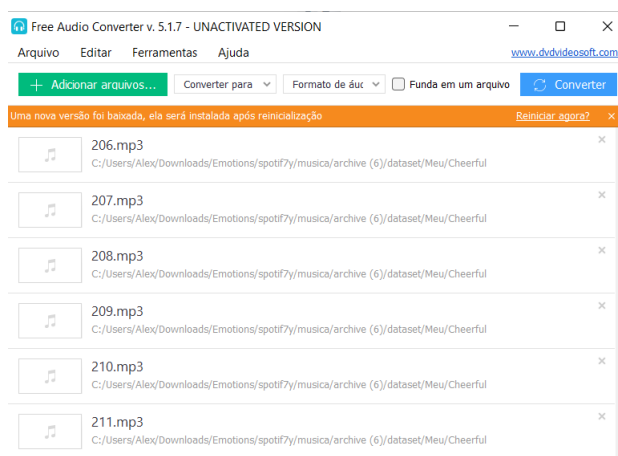


Figure 2: Free Audio Converter.

Entretanto, como se utilizou uma versão gratuita do software, foi adicionado uma marca d'água no início e final de cada audio. Desse modo, outra etapa de pré processamento realizada foi cortar essas marcas d'água com a ferramenta "crop sample" do Edge Impulse. Na figura 3 nota-se como foi realizado este corte dos trechos dos áudios.

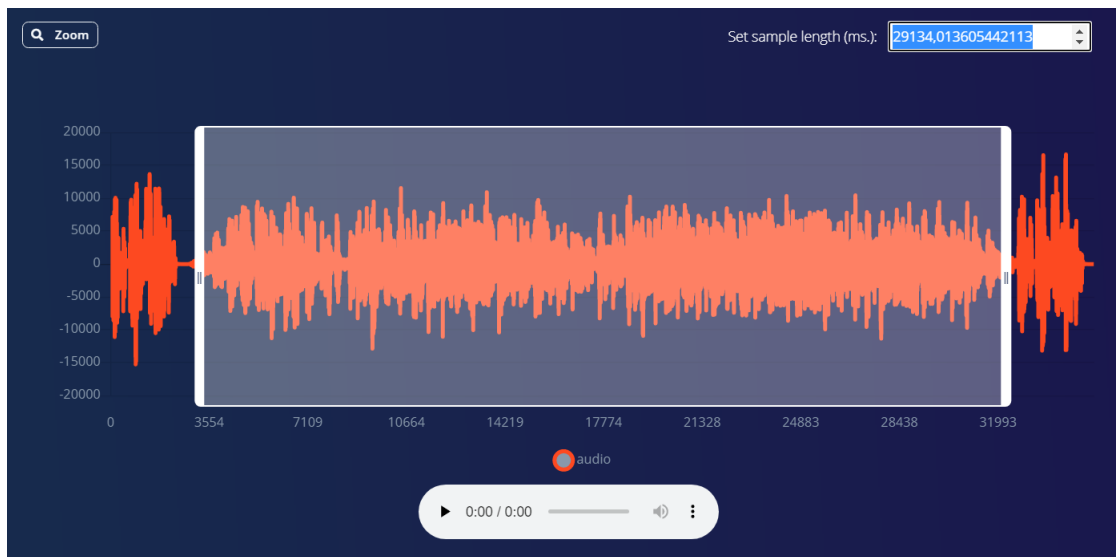


Figure 3: Exemplo Crop Sample Edge Impulse.

Desse modo, após estes procedimentos o modelo estava pronto para iniciar o treinamento.

5 Treinamento da rede

Por fim pode-se dar início ao treinamento do modelo, em que, foi utilizado um bloco de série de tempo no Edge Impulse com uma frequência de 44100 hz, com uma janela de tamanho de 1000 ms. Além disso, foi adicionado também um bloco recomendado para trabalhar com recursos de voz não humana, MFE (Mel-filterbank energy features), em que extrai um espectograma de sinais através do conjunto de áudios utilizado. Por fim, foi adicionado ainda um bloco de classificação que utiliza o Keras, e com isso é possível obter uma saída. Na figura 16 é possível observar como foi construído este modelo.

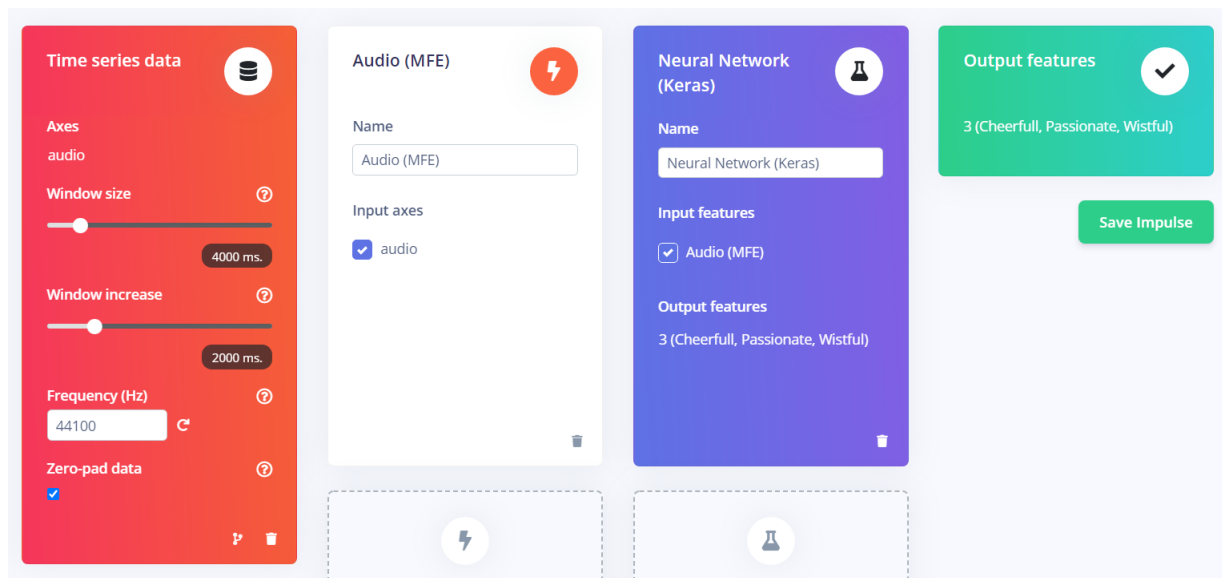


Figure 4: Bloco para o Treinamento da Rede .

Além disso, a arquitetura utilizada para treinar o modelo e na figura 5, pode ser observado como ficou a arquitetura de classificação do modelo.

Architecture presets ? 1D Convolutional (Default) 2D Convolutional



Figure 5: Arquitetura do modelo.

6 Resultados

Após isso, o treinamento do modelo foram gerados as características dos modelos, na guia Spectral features. Em que, dentre os modelos gerados o que melhor separou e distinguiu as classes por suas cores foi o modelo RMS para os 3 eixos. Entretanto, pode-se notar não haver uma separação clara dos modelos, tornando a difícil identificação de cada modelo. Desse modo, na figura 6 é possível constatar a pouca separação dos modelos.

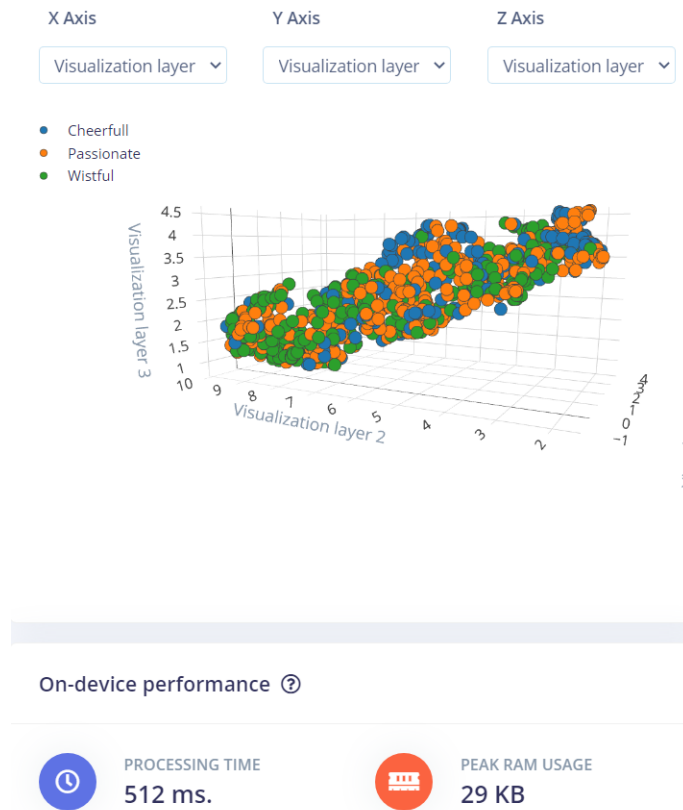


Figure 6: General Features do modelo.

Além disso, pelo NN Classifier é possível notar a matriz de confusão do modelo, no qual possui uma acurácia de 74,4%, e uma função de perda de apenas 0,69. Pode-se notar ainda, que todas as três classes obtiveram valores de Verdadeiros Positivos, maiores de 69%. Na figura 7 é possível analisar a matriz de confusão do modelo.

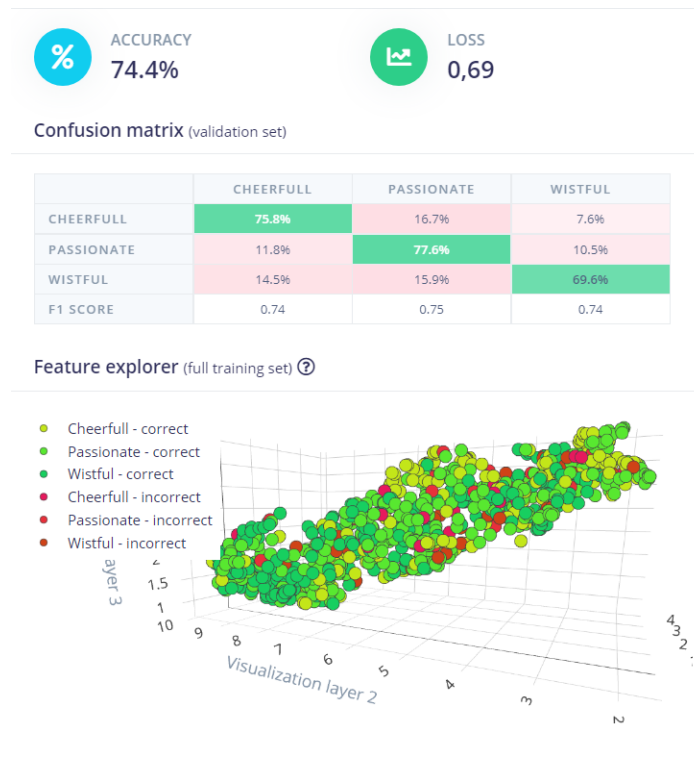


Figure 7: Matriz de confusão do modelo.

7 Teste

Por fim, apos todas as etapas do processo, foi possível testar este modelo dentro da plataforma do Edge Impulse, através da aplicação de Live Classification. Desse modo, é possível adicionar novos conjuntos de dados através do micro-processador e classifica-los da plataforma Web. O modelo de cada musica com a classificação obtida pode ser visto a seguir:

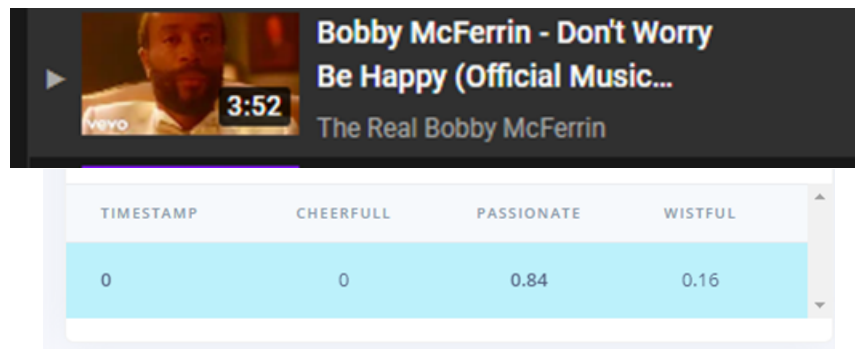


Figure 8: Musica 1 Classificada como Passionate.

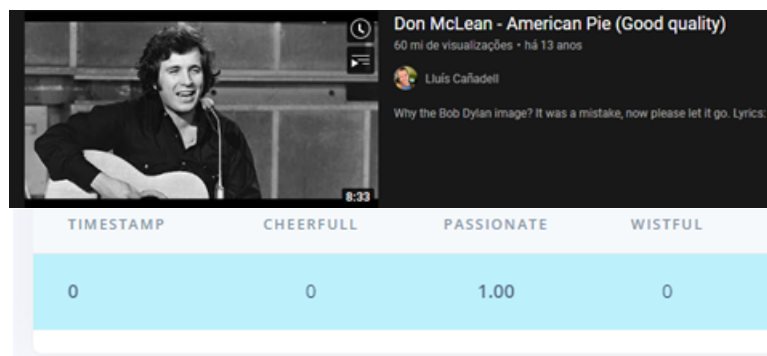


Figure 9: Musica 2 Classificada como Passionate.

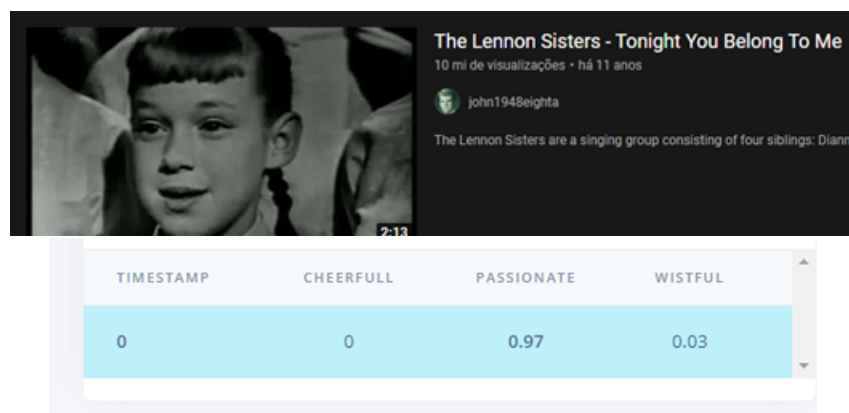


Figure 10: Musica 3 Classificada como Passionate.



Figure 11: Musica 1 Classificada como Wistful.

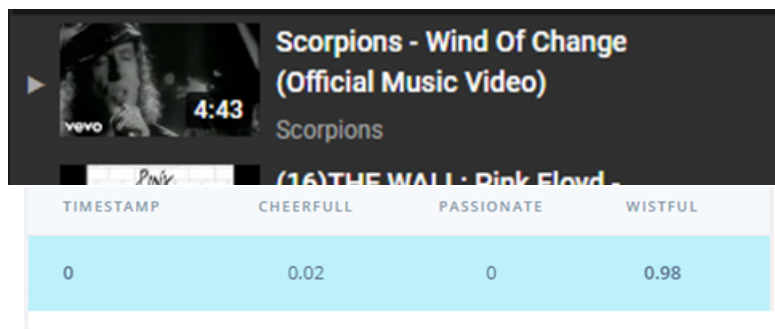


Figure 12: Musica 2 Classificada como Wistful.

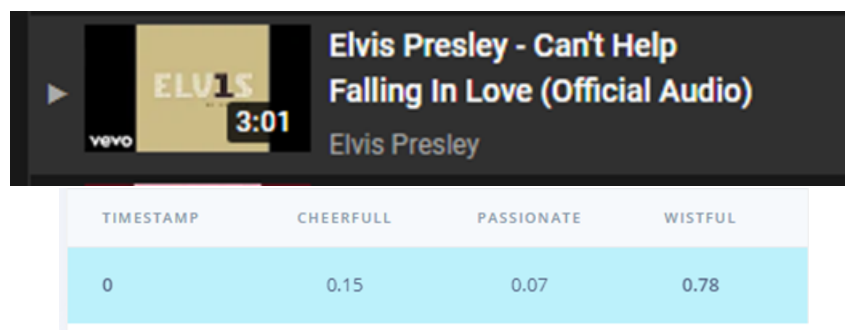


Figure 13: Musica 3 Classificada como Wistful.

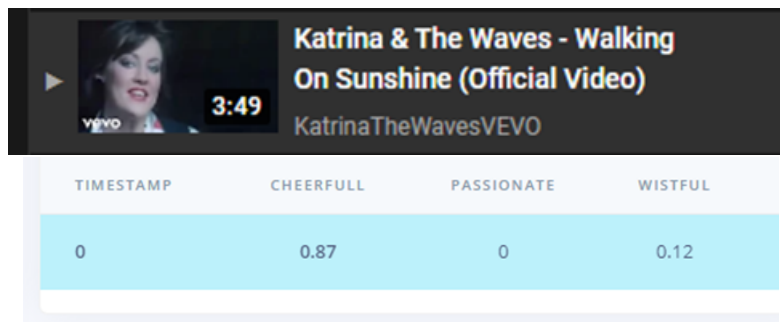


Figure 14: Musica 1 Classificada como Cheerfull.

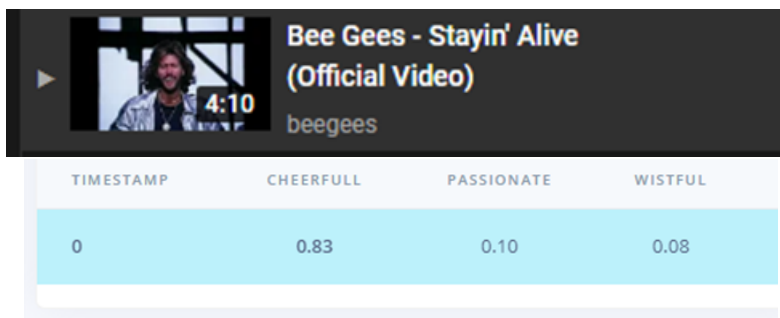


Figure 15: Musica 2 Classificada como Cheerfull.

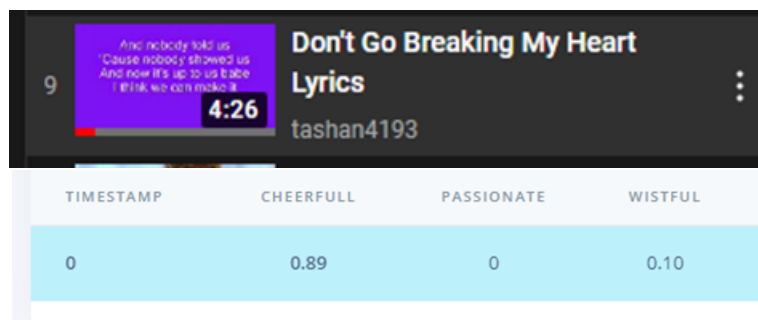


Figure 16: Musica 3 Classificada como Cheerfull.

8 Conclusão

Por fim, apesar de não conseguir testar o modelo no dispositivo físico do Arduino devido alguns problemas de atualização do dispositivo. Entretanto, ainda sim,

foi possível notar que o modelo se comportou bem em sua classificação, através da aplicação do Live Classification, dentro da plataforma do Edge impulse. Assim, dependendo do trecho da melodia identificado do áudio, é possível determinar a qual emoção humana este corresponde. Além disso, o Projeto do Edge impulse e a lista das musicas que continham no dataset podem ser consultados abaixo:

- Projeto no Edge Impulse
- Lista de musicas

Obs: para melhor visualizar a lista de musica, baixe o arquivo .html, e abra no navegador de sua preferencia

References

- [1] R. E. S. Panda, *Emotion-based Analysis and Classification of Audio Music*. PhD thesis, 00500:: Universidade de Coimbra, 2019.
- [2] J. Futrelle and J. S. Downie, “Interdisciplinary research issues in music information retrieval: Ismir 2000–2002,” *Journal of New Music Research*, vol. 32, no. 2, pp. 121–131, 2003.