



Università
di Genova

SCUOLA DI
SCIENZE MATEMATICHE,
FISICHE E NATURALI

Implementazione di un riconoscitore vocale di emozioni per rilevare rabbia e stress

Corso di laurea triennale in Informatica
2023/2024

Relatrici:

Francesca Odone

Nicoletta Noceti

Candidato:

Sara Rosselli

Obiettivo

Implementare un **riconoscitore delle emozioni** in grado di **rilevare il sentimento di rabbia/stress** attraverso l'utilizzo di un **metodo predittivo**.

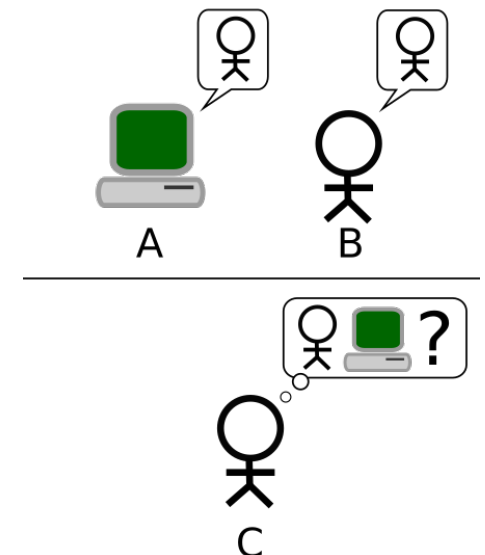
Cosa si intende?

Il riconoscitore vocale delle emozioni è un sistema che analizza il tono e il timbro della voce per identificare le emozioni espresse da una persona.

Questo campo di ricerca combina la prosodia con il Machine Learning.

Applicazioni nel mondo dell'interazione uomo-macchina

- Miglioramento nel campo dell'intelligenza assistiva
- Realizzazione di una macchina che possa superare il Test di Turing, imitando il comportamento umano non solo nel ragionamento logico, ma anche nell'espressione e nella comprensione delle emozioni



Come funziona?

Il riconoscitore vocale delle emozioni utilizza algoritmi di Machine Learning per analizzare le caratteristiche acustiche del suono vocale, come:

- Frequenza e ampiezza del suono
- Tempo e ritmo della parola
- Intensità e cadenza del tono di voce

Queste caratteristiche sono associate a diverse emozioni, come felicità, tristezza, paura, rabbia e disgusto.

Passi per l'implementazione del mio modello

*1. Data Loading
and
Preprocessing*

*2. Data
Visualization*

*3. Data
Augmentation*

*4. Features
Extraction*

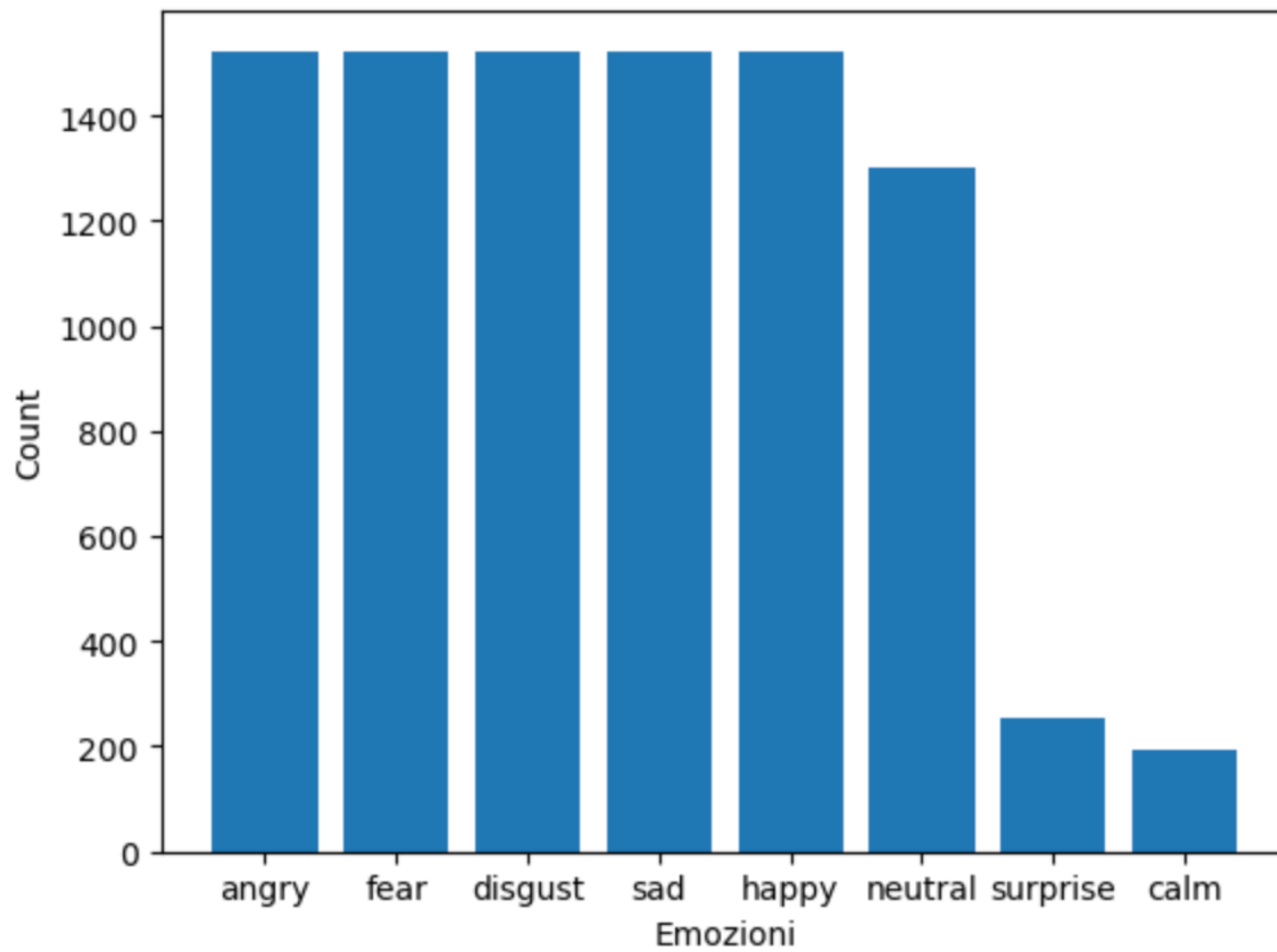
*5. Model Training
and Evaluation*

Data Loading and Preprocessing

Ho utilizzato 3 dataset di kaggle.com ([RAVDESS](#), [CREMA-D](#), [SAVEE](#)), ognuno dei quali presenta tracce audio di attori (maschi e femmine, di varie nazionalità) che pronunciano frasi con diverse emozioni, individuabili nel nome del file.

Ho elaborato i file audio, estraendo le tracce degli audio e le relative etichette delle emozioni.

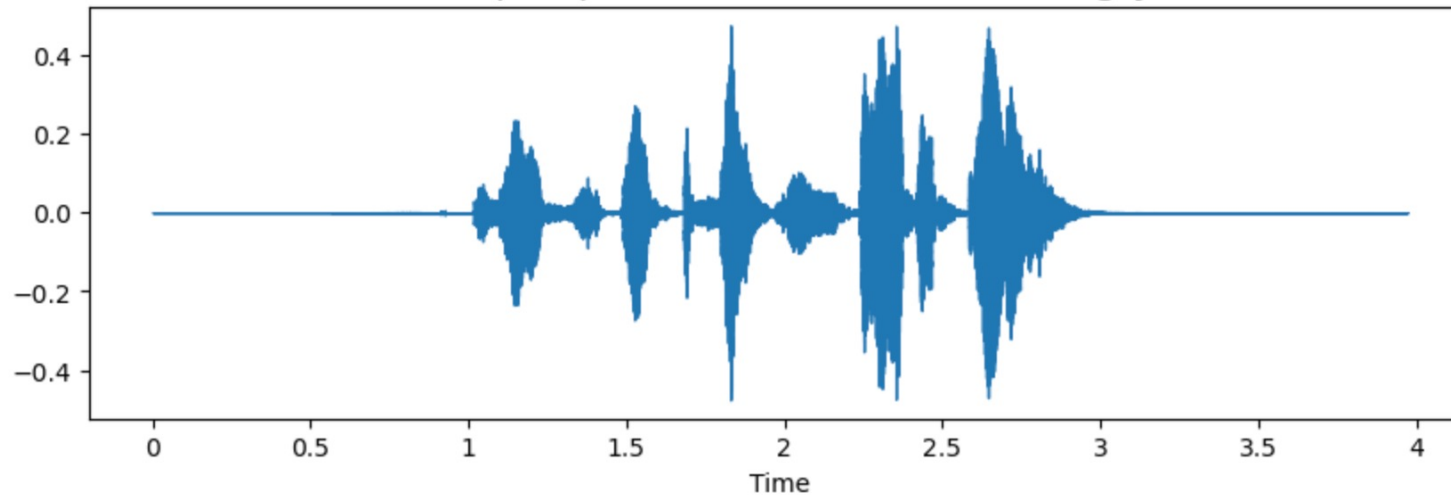
Ho poi combinato i 3 dataset in un unico DataFrame di 9362 istanze.



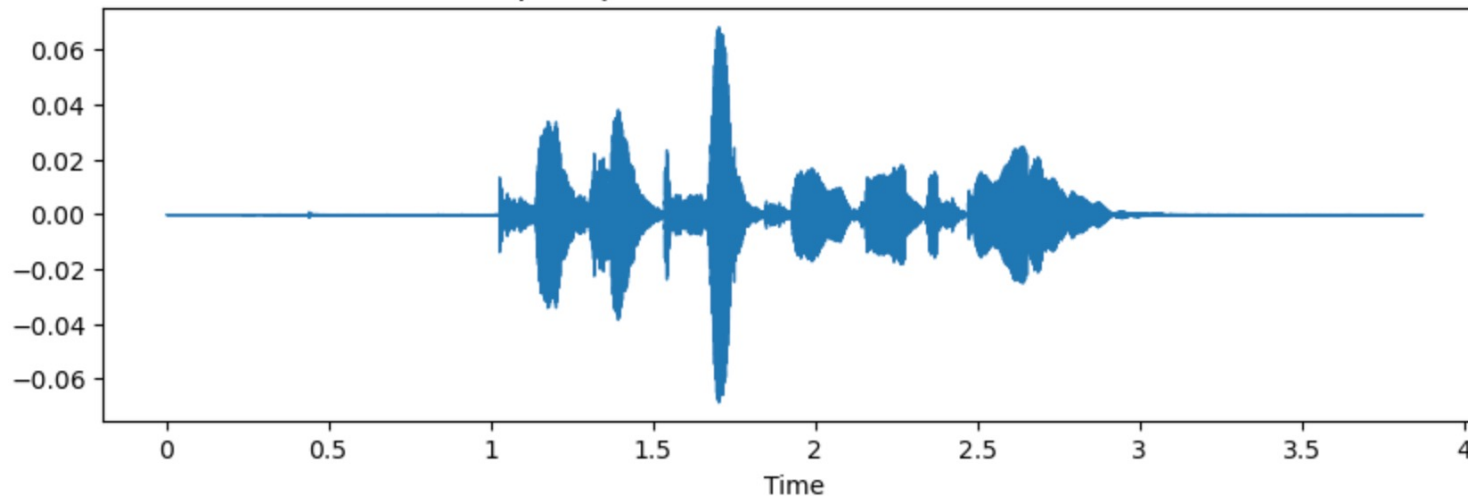
Data Visualization

Dato che stiamo manipolando segnali audio, è utile visualizzare dei **waveplots** per individuarne l'ampiezza nel dominio del tempo e osservarne le variazioni

Waveplot per audio con emozione: angry

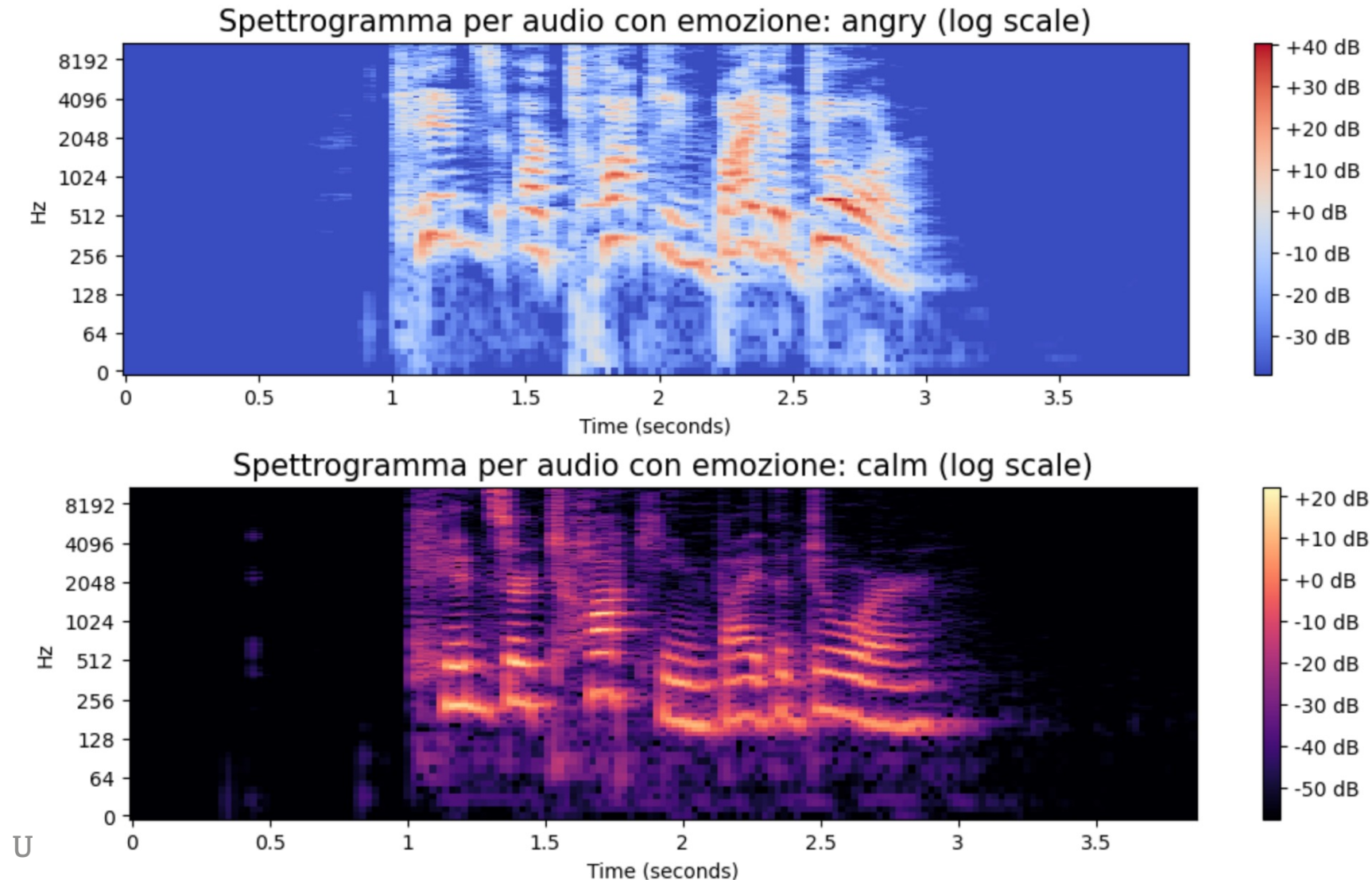


Waveplot per audio con emozione: calm



Data Visualization

Dato che stiamo manipolando segnali audio, è utile visualizzare degli **spettrogrammi** per visualizzare la distribuzione delle frequenze del segnale nel tempo.

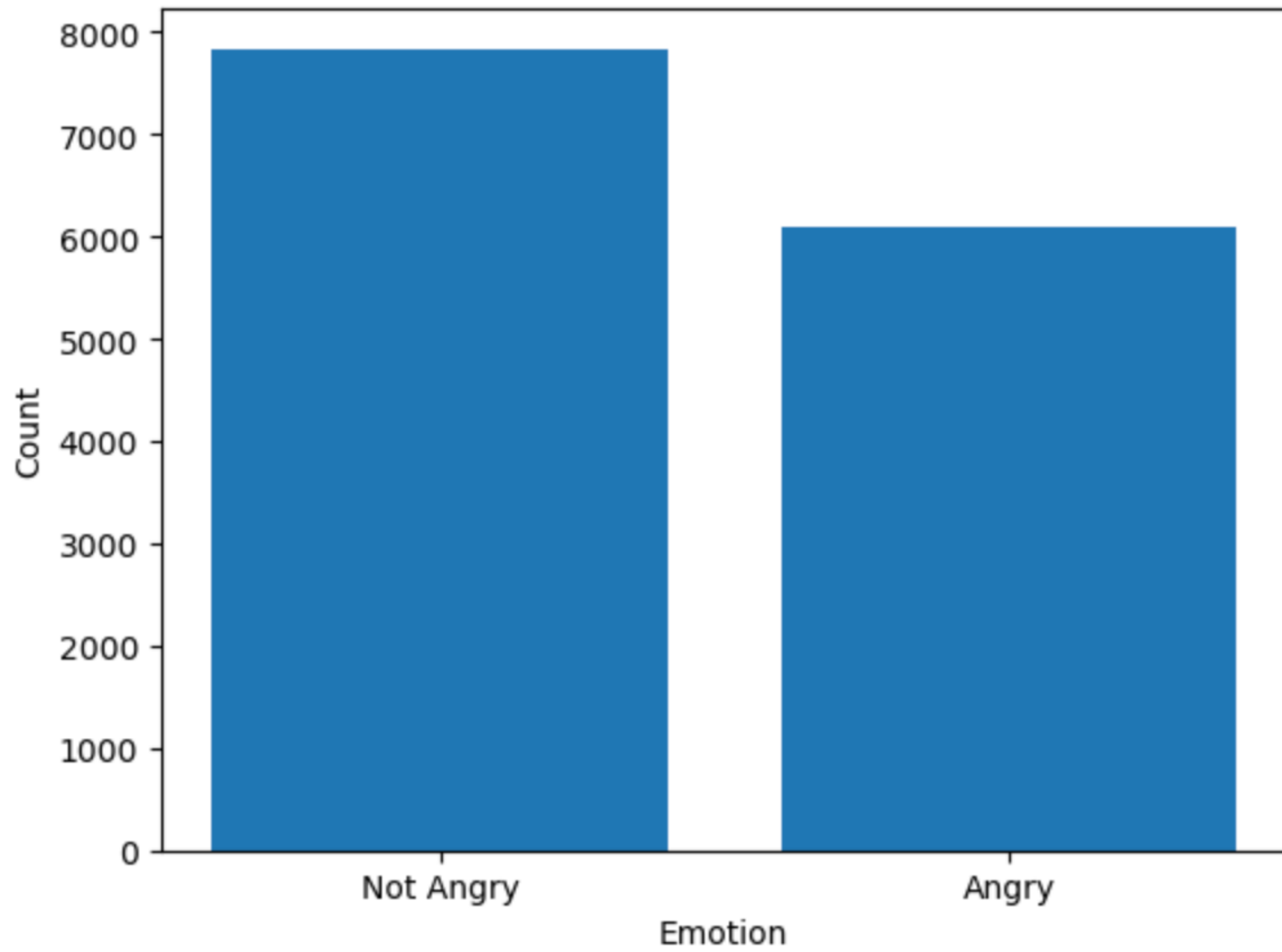


Data Augmentation

Considerato il rapporto tra il numero di istanze la cui categoria è "angry" e tutte le altre, possiamo effettuare un'operazione di Data Augmentation perturbando leggermente alcune tuple del dataset con emozione = "angry", così da **creare artificialmente altre istanze appartenenti a questa categoria** e **diminuire così la discrepanza nel numero di tuple**, che potrebbe influenzare negativamente il processo di predizione; inoltre quest'operazione aiuta a rendere il nostro modello più "generalizzante" e robusto a variazioni nei dati.

PERTURBAZIONI EFFETTUATE:

- aggiunta di rumore
- pitch shift

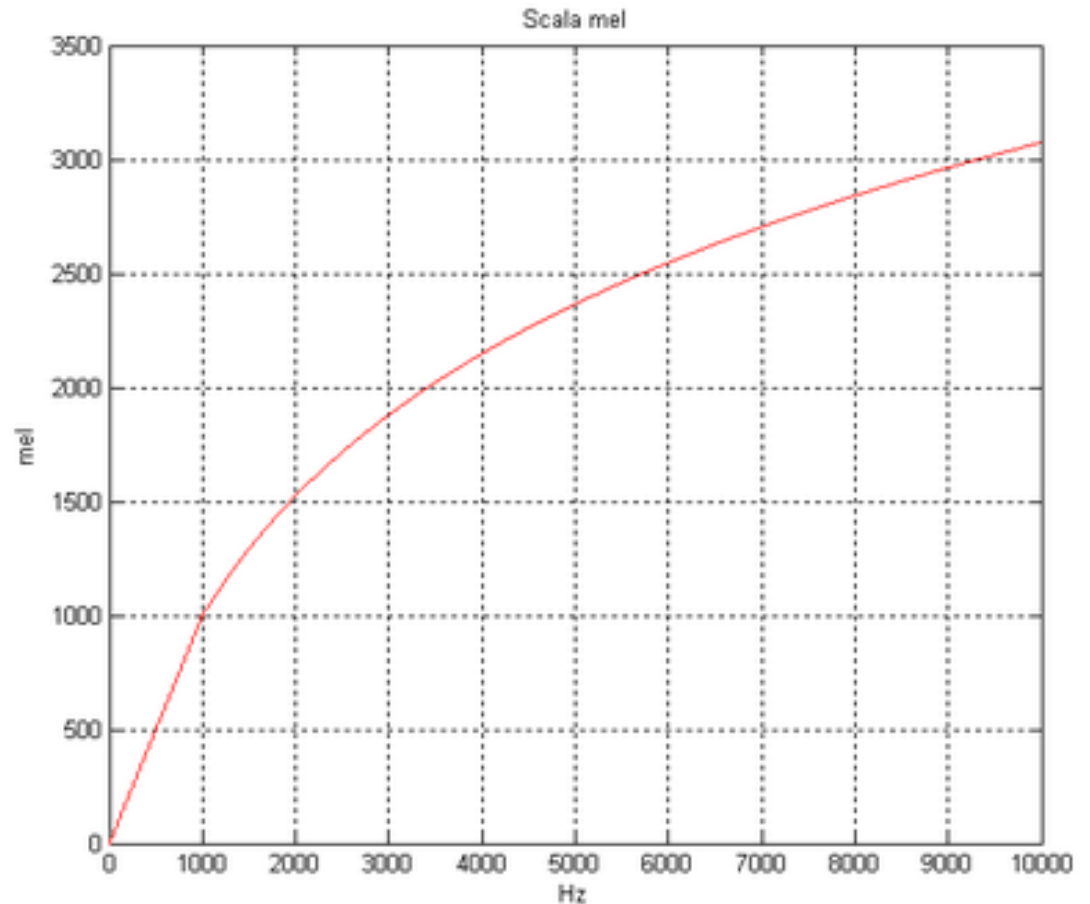


Features Extraction con il metodo MFCC

Mel-frequency cepstral coefficients (MFCC)

E' una rappresentazione del segnale tramite **coefficienti** che catturano le **caratteristiche spettrali di un segnale audio** in modo da **modellare la percezione umana del suono**, tenendo conto di come l'orecchio umano percepisce le diverse frequenze.

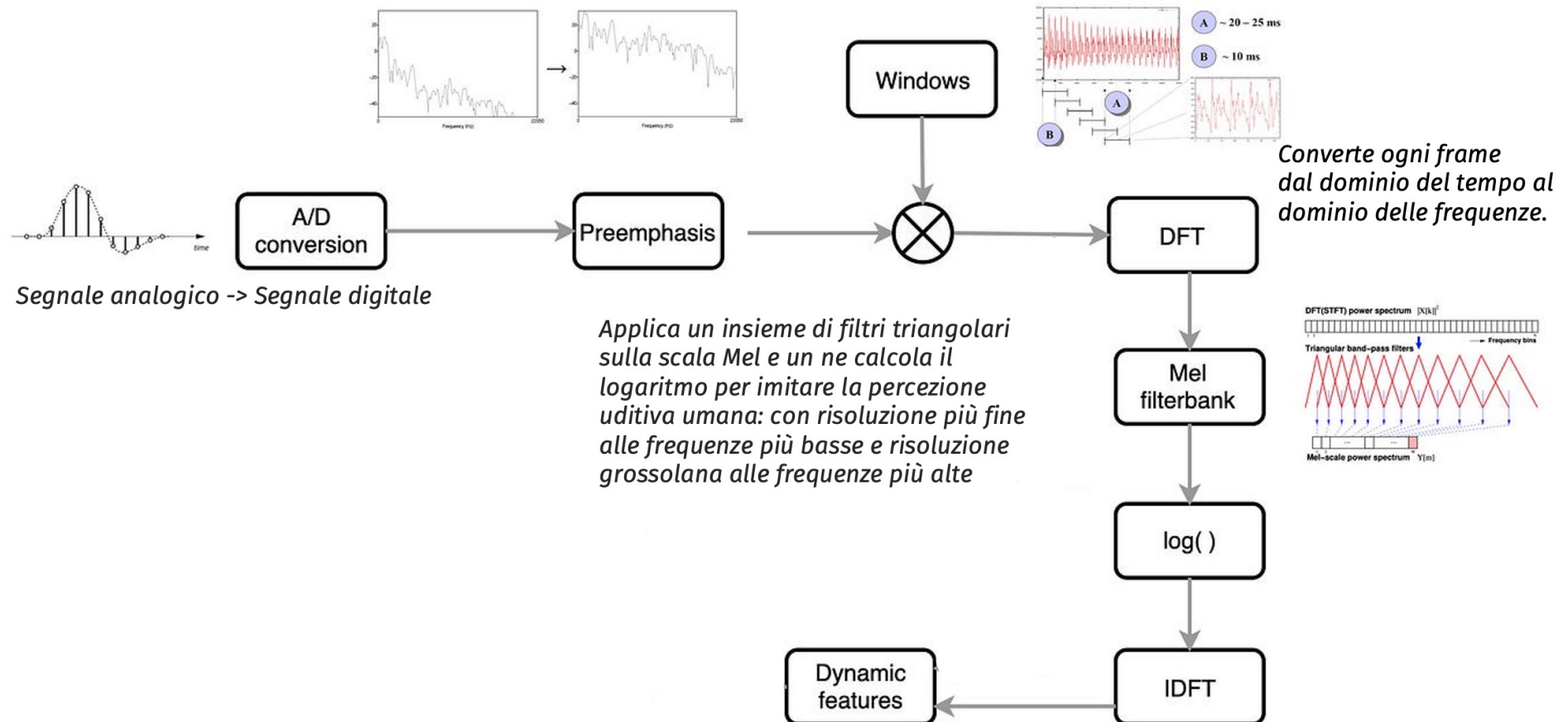
Utilizza la **scala Mel**, una scala di frequenze percettive progettata per approssimare la risposta della coclea umana alle frequenze percepite. Infatti **la percezione umana delle frequenze non è lineare, ma logaritmica**: siamo più sensibili ai cambiamenti nelle frequenze più basse rispetto a quelle più alte.



Features Extraction con il metodo MFCC

Aumenta la quantità di energia nelle alte frequenze per compensare la perdita di recezione dell'orecchio umano

Divide il segnale in piccoli frames e ne minimizza le discontinuità ai lati dei frames



Model Training and Evaluation (Regressione Logistica)

Perché la Regressione Logistica?

E' un metodo predittivo per i problemi di classificazione binaria, dove l'obiettivo è prevedere uno tra due possibili risultati. In questo contesto, la regressione logistica ci può aiutare a determinare se chi sta parlando è arrabbiato o meno sulla base di caratteristiche estratte dai MFCC.

PASSI:

1. Divido il dataset in training-set (80%) e test-set (20%)
2. Viene istanziato un oggetto della classe *LogisticRegression*
3. Il modello viene allenato (sulla base delle caratteristiche dei MFCC)
4. Viene utilizzato il modello imparato per effettuare predizioni sul test-set

Model Training and Evaluation (Regressione Logistica)

Efficienza del modello

```
Accuratezza del mio modello: 0.860  
MAE: 0.140  
MSE: 0.140  
RMSE: 0.374
```

```
Modello nullo:  
MAE: 0.492  
MSE: 0.246  
RMSE: 0.496
```

Report di classificazione:

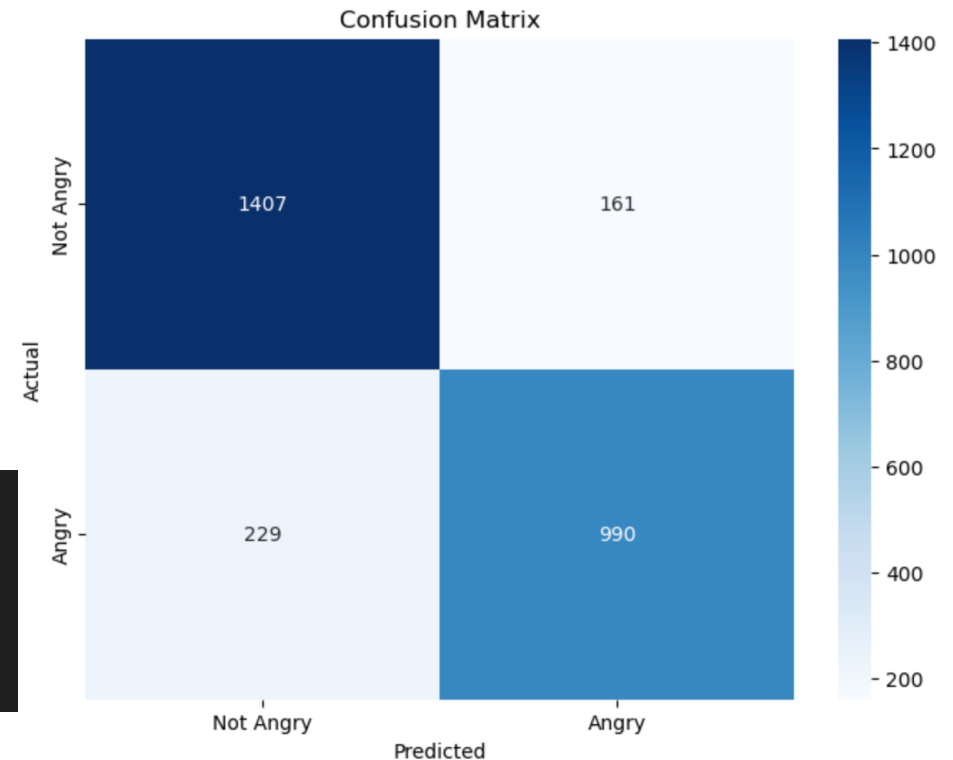
	precision	recall	f1-score	support
0	0.86	0.90	0.88	1568
1	0.86	0.81	0.84	1219

Precision: veri positivi + falsi positivi

Recall: veri positivi + falsi negativi

F1-score: media tra precision e recall

Support: numero di campioni di quella classe presenti nel test-set



Conclusioni

Il **modello** di Regressione Logistica si è rivelato **abbastanza efficace** come metodo predittivo per il nostro scopo, soprattutto **grazie** all'integrazione di nuove istanze di tipo "angry" durante l'operazione di **Data Augmentation**.

Sarebbe interessante esplorare la **Feature Extraction** combinando l'MFCC **con altri metodi** di estrazione per vedere se la precisione del nostro modello migliora ulteriormente. Potremmo anche considerare di implementare un modello capace di **classificare ogni emozione**, utilizzando ad esempio una Rete Neurale Convoluzionale.



Università
di Genova

SCUOLA DI
SCIENZE MATEMATICHE,
FISICHE E NATURALI

Fonti

Speech Emotion Recognition Using Machine Learning:

<https://www.sciencedirect.com>

Speech Recognition – Feature Extraction, Jonathan Hui (2019)

Audio Signal Feature extraction

<https://medium.com>

MFCC Technique for Speech Recognition

<https://www.analyticsvidhya.com>

UniGe

SCIENZE
MFN