



# Data Mining Project

RAVDESS Dataset Analysis

Author: *Biancamaria Bombino, Sara Hakim, Davide Ricci*

Instructor:	<i>Pedreschi, Guidotti, Spinnato</i>
Master's degree:	Data Science & Business Informatics
Date:	31 dicembre 2022

---

# Indice

1. Introduzione . . . . .	2
2. Data Understanding & Data Preparation . . . . .	2
2.1 Data Semantics . . . . .	2
2.2 Distribuzione delle variabili e statistiche . . . . .	3
2.3 Data Quality . . . . .	5
2.4 Trasformazione variabili . . . . .	7
2.5 Eliminazione variabili e correlazione attributi . . . . .	7
3. Clustering . . . . .	8
3.1 K-MEANS - Metodi basati su centriodi . . . . .	8
3.2 DBSCAN - Metodi basati sulla densità . . . . .	11
3.3 Clustering Gerarchico . . . . .	12
3.4 Valutazioni finali di clustering . . . . .	13
4. Classificazione . . . . .	14
4.1 Pre-processing . . . . .	14
4.2 Scelta degli Iperparametri . . . . .	14
4.3 Emotion . . . . .	15
4.4 Emotional Intensity . . . . .	16
4.5 Regressione . . . . .	19
5. Pattern Mining . . . . .	20
5.1 Pre-processing . . . . .	20
5.2 Estrazione e analisi dei frequent patterns al variare di support . . . . .	21
5.3 Estrazione e analisi di regole al variare di confidence . . . . .	22
5.4 Predizione variabile target e valutazione accuratezza . . . . .	24

# Data Understanding and Clustering

## 1. Introduzione

Il Ryerson Audio-Visual Database of Emotional Speech (RAVDESS) è un dataset composto da una serie di audio digitali, registrati da un numero finito di persone. In particolare, a 24 attori con accento americano è stato richiesto di pronunciare o cantare due frasi lessicamente corrispondenti. Tali espressioni sono caratterizzate da determinate emozioni e sono enunciate con una certa intensità emozionale.

In ambito scientifico, i file audio, e in particolar modo il canto, rappresentano un'innovazione nel settore della ricerca medica. Infatti, questi ultimi vengono impiegati per stimolare il sistema nervoso dei pazienti e analizzare i diversi effetti ottenuti su ogni individuo.

Gli elementi presenti nel dataset, oggetto del nostro studio, rappresentano solo una porzione del dataset originale; ai dati analizzati sono state aggiunte nuove features.

## 2. Data Understanding & Data Preparation

Il Dataset contiene 2452 records e 37 features ed è composto da due tipi di file audio digitali: *vocale* e *cantato*. La tipologia dell'audio è visualizzabile attraverso la feature **vocal\_channel**. Nella seguente sezione vengono riportate le principali analisi effettuate durante la fase di Data Understanding, in modo da apprendere il maggior numero di informazioni possibili attraverso i dati oggetto delle nostre analisi.

### 2.1 Data Semantics

Inizialmente sono stati analizzati i singoli attributi contenuti nel dataset per familiarizzare con l'ambito a cui appartengono le nostre variabili. Per questo motivo, abbiamo distinto gli attributi in base a tre tipologie: numerico, categorico e ordinale. Questa distinzione è visualizzabile nelle figure 1, 2 e 3.

Nome	Tipologia	Descrizione	N. valori distinti
Modality	Categorico	indica il tipo di file	1
Vocal channel	Categorico	identifica se il file è parlato o cantato	2
Emotion	Categorico	indica il tipo di emozione provata nel file audio	8
Emotional intensity	Categorico	intensità dell'emozione [strong, normal]	2
Statement	Categorico	affermazione pronunciata dagli attori	2
Repetition	Categorico	affermazione ripetuta per la prima o seconda volta	2
Actor	Categorico Numerale	id rappresentante il singolo attore	24
Sex	Categorico	indica il sesso dell'attore [M,F]	2
Channel	Categorico Numerale	numero di canale [1:mono, 2:stereo audio]	2
Sample width	Numerico	numero di bits per rappresentare l'audio	1
Frame rate	Numerico	frequenza ovvero numero di campioni per secondo	1
Frame width	Numerico	numero di byte usato per ogni sample	2
Lenght ms	Numerico	durata dell'audio in millisecondi	95
Frame count	Numerico	numero di frames contenuti nel sample	158
Intensity	Numerico	volume audio in dbFS	989

Figura 1: Semantica delle Features

Nome	Tipologia	Descrizione	N. valori distinti
Zero crossing sum	Numerico	somma delle velocità con cui cambia il segnale	2176
Mean	Numerico	media statistica del segnale dell'audio originale	2450
Std	Numerico	deviazione standard del segnale dell'audio originale	2451
Min	Numerico	minimo del segnale dell'audio originale	2148
Max	Numerico	massimo del segnale dell'audio originale	2166
Kur	Numerico	allontanamento della normalità distributiva	2451
Skew	Numerico	misura dell'assimetria della distribuzione di probabilità	2451
Mfoc mean	Numerico	media della distribuzione della potenza del segnale	2451
Mfoc std	Numerico	deviazione standard della distribuzione della potenza	2449
Mfoc min	Numerico	minimo della distribuzione della potenza del segnale	2451
Mfoc max	Numerico	max della distribuzione della potenza del segnale	2449
Sc mean	Numerico	valore medio del centro di massa dello spettro	2451
Sc std	Numerico	deviazione std del centro di massa dello spettro	2451
Sc min	Numerico	minimo centro di massa dello spettro	1431
Sc max	Numerico	max centro di massa dello spettro	2423

Figura 2: Semantica delle Features

Nome	Tipologia	Descrizione	N. valori distinti
Sc kur	Numerico	curtosi del centro di massa dello spettro	2451
Sc skew	Numerico	misura dell'assimetria del centro di massa spettro	2451
Stft mean	Numerico	ampiezza media di un segnale	2451
Stft std	Numerico	deviazioni std dell'ampiezza di un segnale	2451
Stft min	Numerico	minimo dell'ampiezza di un segnale	1431
Stft max	Numerico	massimo dell'ampiezza di un segnale	1
Stft kur	Numerico	curtosi dell'ampiezza di un segnale	2451
Stft skew	Numerico	assimetria dell'ampiezza di un segnale	2451

Figura 3: Semantica delle Features

## 2.2 Distribuzione delle variabili e statistiche

In questa sezione sono state esaminate, attraverso diverse tecniche di visualizzazione, tutte le tipologie di features contenute nel dataset; in modo da comprendere la loro distribuzione ed effettuare eventuali analisi statistiche. In figura 4, è possibile visualizzare la ripartizione della variabile categorica **emotion**, utilizzata come variabile target, tra i vari records. Da una prima analisi si osserva che gli oggetti, rispetto a questa variabile, sono distribuiti in modo equilibrato ad eccezione delle emozioni: *disgust*, *surprised* e *neutral*; le quali risultano essere meno frequenti.



Figura 4: Distribuzione delle emozioni

Un caso analogo riguarda la variabile categorica **emotional\_intensity**, in cui si nota una distribuzione pressochè bilanciata. Infatti sono presenti 1230 records *normal* e 1132 records *strong*. Infine, per quanto riguarda le variabili categoriche **statement** e **repetition** risultano anch'esse perfettamente equilibrate. Attraverso l'analisi della variabile **actor**, è stato possibile visualizzare che i 24 attori si suddividono in 12 donne e 12 uomini. Come raffigurato mediante la tabella in figura 5, ai soggetti di sesso femminile corrisponde un identificativo pari, mentre ai soggetti maschili è associato un identificativo dispari. Durante lo studio di questa variabile, è stato notato che i cinque utenti più attivi all'interno del dataset sono rappresentati da un ID pari. Di conseguenza è stato dedotto che, nonostante risultino esserci più records interpretati da uomini, gli attori che interpretano maggiormente le due affermazioni corrispondono a cinque donne.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
F	0	58	0	52	0	55	0	61	0	56	0	63	0	62	0	58	0	35	0	61	0	65	0	58
M	51	0	51	0	58	0	51	0	51	0	55	0	60	0	44	0	55	0	60	0	55	0	51	0

Figura 5: Distribuzione degli attori

Relativamente alle variabili numeriche, è stata osservata la presenza di attributi distribuiti in modo sbilanciato. Per esempio, la variabile **channels** presenta records in prevalenza connotati da un valore pari a 1; mentre **frame\_width** è caratterizzata da records in prevalenza connotati da un valore pari a 2. Le features **sample\_width** e **frame\_rate** sono invece composte solamente da valori costanti, rispettivamente pari a 2 e a 48000. Tramite l'analisi della distribuzione delle emozioni rispetto alla variabile **vocal\_channel**, è stato compreso che le emozioni di disgusto e di sorpresa compaiono solo per gli audio con la modalità *speech*; che è quella maggiormente utilizzata nelle interpretazioni delle frasi messe a disposizione per gli attori. Inoltre, tramite la distribuzione dei valori di **vocal\_channel** rispetto a più variabili; è emerso che, per entrambe le modalità *speech* e *song*, i records con un'intensità *normal* sono in quantità maggiore rispetto a quelli caratterizzati da un'intensità *forte*. Successivamente, è stato osservato che gli audio aventi modalità *song* sono più lunghi di quelli aventi modalità *speech*; e gli audio cantati sono caratterizzati da emozioni quali la calma e la tristezza. Gli audio cantati sono maggiormente vocalizzati da uomini, mentre quelli parlati sono pronunciati principalmente da donne. Di conseguenza, si può affermare che le donne si interfacciano con

tutte le emozioni contenute nel dataset; mentre gli uomini si interfacciano principalmente con emozioni caratterizzate da una minore reazione emotiva. Infine per quanto riguarda la variabile **statement**, si può affermare che la frase *"dogs are sitting by the door"* viene pronunciata più frequentemente in entrambe le modalità *speech* e *song*.

## 2.3 Data Quality

Nella fase di valutazione della qualità dei dati, sono state effettuate delle analisi per comprendere se sono presenti errori, outliers e/o valori mancanti. Di conseguenza, è stata presa in considerazione la possibilità di eliminare determinate features poiché contengono un unico valore oppure valori ricavabili da altre variabili.

### 2.3.1 Missing Values

Analizzando il dataset, sono stati individuati alcuni missing values inerenti ad alcune features che caratterizzano i dati. Dalla distribuzione dei missing values, ottenuta in figura 6, è possibile comprendere che solo 3 features su 38 presentano valori nulli. Le variabili con valori mancanti sono: **Vocal\_channel** con 196 valori nulli, **Actor** con 1126 valori nulli e **Intensity** con 816 valori nulli.

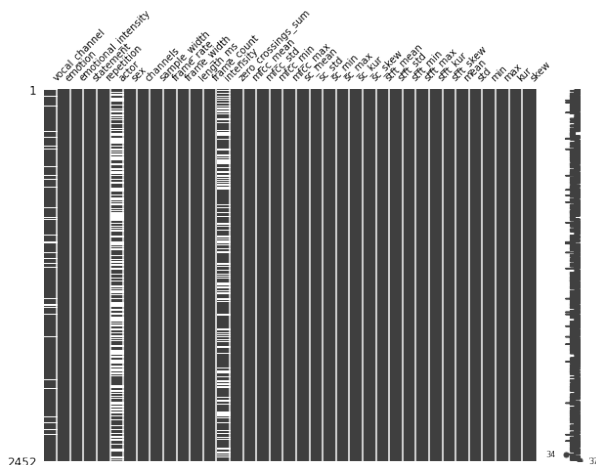


Figura 6: Rappresentazione dei Missing Values

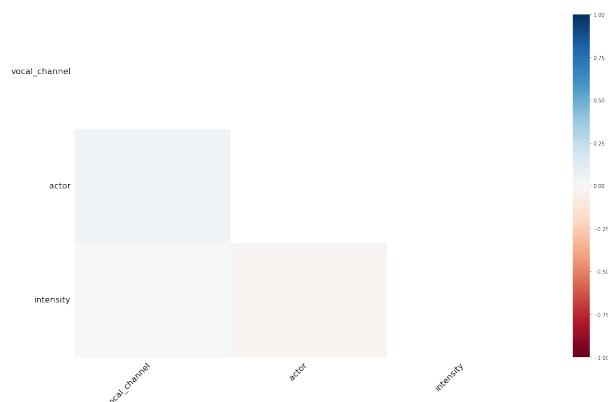


Figura 7: Rappresentazione Correlazione Valori Nulli

Attraverso una analisi quantitativa dei dati, è stato possibile visualizzare che l'attributo riguardante gli attori presenta il 45,92% di valori mancanti. Questa features è quindi quella maggiormente esposta alla presenza di valori nulli. I missing values sono stati trattati come appartenenti alla categoria **MCAR**; ovvero i dati mancanti non seguono uno schema particolare, ma sono casuali. Infatti la mancanza di questi valori non è correlata con le variabili rimanenti, come è possibile visualizzare nella figura 7. Inoltre, attraverso le restanti features, non è possibile prevedere i valori non presenti nelle 3 variabili elencate precedentemente. Per quanto riguarda i valori mancanti della features **actor**, possiamo solo comprendere se il dato mancante può corrispondere ad un numero pari o dispari, attraverso la variabile **sex** che rappresenta il sesso. Ovvero, se il sesso è uguale ad *M* probabilmente l'ID dell'attore corrisponde ad un numero dispari tra 1 e 24. Lo stesso ragionamento viene effettuato se il sesso è pari a *F*. Considerando il tipo dei missing values, per la loro

gestione è stato scelto di adottare un metodo simile all'imputazione singola. In particolare, questo metodo è stato utilizzato per la variabile **vocal\_channel**. Infatti, essendo quest'ultima una variabile categorica e qualitativa, essa viene sostituita con la moda; ovvero il valore più frequentemente ritrovato nei records. Le variabili **actor** e **intensity**, invece, sono state selezionate per essere eliminate nella fase successiva di preparazione dei dati, poiché quasi la metà dei loro valori è assente. Tutte le altre 35 features non presentano missing values, ma alcune di esse sono caratterizzate da un unico valore per tutti i records presenti nel dataset analizzato, come riportato nella Sezione 2.2. Queste ultime, avendo un unico valore distinto, non sono state ritenute utili per le nostre analisi.

### 2.3.2 Gestione Outliers ed Errori

L'individuazione di outliers ed errori è stata effettuata attraverso la rappresentazione di ogni valore di tipo numerico nei boxplot. Dall'analisi è emerso che varie features presentano degli outliers. In particolare **frame\_count** presenta 35 records con valore pari a -1. Essendo inoltre il **frame\_count** il numero di frames del sample, non può essere rappresentato da un numero negativo. Questi outliers risultano quindi essere dei dati non validi e sono stati etichettati come errori. In aggiunta, è emerso che anche le altre features che compongono il dataset, presentano alcuni outliers. Anche in questi casi, molti di loro risultano in records in cui il valore rappresentante l'attore è nullo. Essendo basso il numero di questi outliers, non è stata attribuita loro particolare importanza per le analisi effettuate.

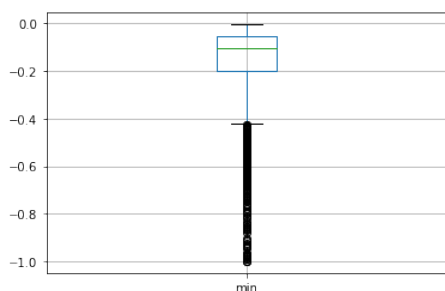


Figura 8: Outliers Minimo

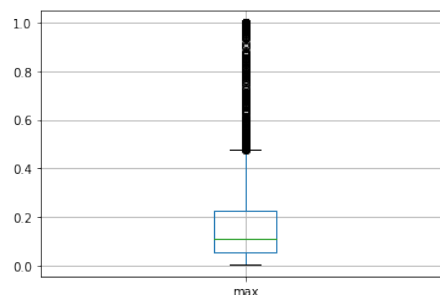


Figura 9: Outliers Massimo

Invece, tra le variabili con tanti outliers, è presente il minimo del segnale dell'audio originale, ovvero **min**. Dal boxplot in figura 8, è stato dedotto che questa feature contiene numerose istanze al di sotto del *whiskers*. In particolare, sono stati rilevati 134 records identificati come outliers. Questi records hanno in comune l'emozione, ovvero la rabbia, e un'intensità emozionale forte, che caratterizzano gli audio corrispondenti a tali casi. Questi outliers inizialmente sono stati ritenuti anomali poiché a forti reazioni emotive, caratterizzate da forti intensità emozionali, dovrebbero corrispondere segnali audio elevati. Di conseguenza è stato notato, mediante il grafico in figura 9, che anche il massimo del segnale audio contiene 123 outliers presenti con le stesse condizioni. Quindi si hanno dei minimi molto bassi che corrispondono a dei massimi alti. Ciò è stato interpretato come un cambiamento drastico del segnale audio originale, ed è stato ritenuto normale in presenza di forti emozioni. Per questo motivo, questi valori sono stati considerati come "casi particolari" piuttosto che errori, ipotizzando che siano degli outliers fittizi intenzionali e inseriti per testare gli audio in termini di validità emotiva, intensità e soprattutto di genuinità.

## 2.4 Trasformazione variabili

Durante questa prima fase di esplorazione dei dati, non è stato ritenuto necessario effettuare trasformazioni preliminari di variabili. L'eventuale processo di trasformazione delle features verrà svolto durante le fasi di applicazione degli algoritmi di classificazione e pattern mining; in particolare nella sezione di pre-processing dei dati.

## 2.5 Eliminazione variabili e correlazione attributi

In prima istanza sono stati eliminati dal dataset gli attributi **modality**, **sample\_width**, **frame\_rate** e **stft\_max**. Questa cancellazione è stata effettuata poiché essi contengono un unico valore per tutti i records. Inoltre, durante la fase di Data Understanding, è stata riscontrata la presenza di valori nulli in circa la metà dei records per gli attributi **actor** e **intensity**, come riportato nella Sezione 2.3.1; di conseguenza queste variabili sono state rimosse. In seguito, è stata effettuata un'accurata ricerca di valori anomali per ogni attributo appartenente al dataset. Come descritto nella Sezione 2.3.2, l'esistenza di errori è stata riscontrata per l'attributo **frame\_count**; per i relativi 35 records contenenti il valore -1 è stata eseguita un'operazione di cancellazione.

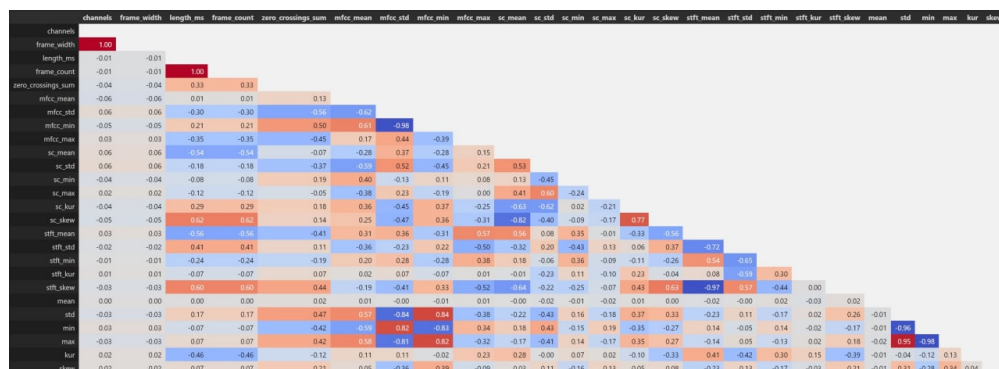


Figura 10: Matrice di correlazione di Pearson

Successivamente è stata analizzata la correlazione tra tutte le coppie di attributi, utilizzando il coefficiente di Pearson, come mostrato in figura 10. Da questa analisi è emersa una correlazione pari ad 1 per le coppie **frame\_width, channels** e **frame\_count, length\_ms**. I gruppi di funzionalità altamente correlate non potrebbero fornire significative informazioni aggiuntive e mantenendoli si avrebbe un aumento della complessità degli algoritmi di apprendimento supervisionato e non. Per tale motivo è stato deciso di non mantenere situazioni di correlazione perfetta; dunque per la coppia **frame\_width, channels** è stata rimossa la prima feature contenente solamente due valori distinti, ovvero 2446 records con un numero pari a 2, e 6 records con un numero pari a 4. Per quanto riguarda la seconda coppia **frame\_count, length\_ms**, è stato tolto il primo attributo poiché la lunghezza del file audio originale è stata ritenuta più informativa e vantaggiosa in previsione dei prossimi task. I 35 records, eliminati dopo aver individuato la presenza di errori nella variabile **frame\_count** sono stati ritenuti comunque vani, nonostante sia stata eliminata completamente questa variabile. La scelta di non reintrodurre tali records nel dataset, è stata presa poiché, come detto in precedenza, contenevano valori nulli in altre features (*actor, modality*) e perciò sono stati considerati come errori effettivi del record in generale, e non solamente relativi all'attributo di conteggio del numero



di frames del sample. Dalla matrice di correlazione sono emerse anche altre coppie di attributi con una correlazione piuttosto alta, sia negativamente che positivamente. Per queste coppie è stato però deciso di non proseguire con ulteriori cancellazioni, in modo da non ridurre ulteriormente la dimensione del dataset.

### 3. Clustering

In questa sezione sono state riportate le analisi effettuate attraverso diversi algoritmi di clustering: **K-Means**, **DBScan** e **Hierarchical**. Questi algoritmi sono stati applicati al dataset modificato e ai risultati ottenuti in seguito agli studi condotti e riportati nella Sezione 2. Inoltre sono state sperimentate anche alcune varianti di questi metodi, ovvero, **Bisecting K-Means**, **X-Means** per l'algoritmo basato sui centroidi, e **Optics** come estensione del DBScan. Per l'applicazione degli algoritmi, sono state utilizzate le features numeriche continue. Quest'ultime sono state normalizzate attraverso il *Robust Scaler*, in modo tale da gestire la presenza di outliers, individuati nella fase di Data Understanding. Inizialmente, è stato compreso che adoperare tutte le features non avrebbe portato a risultati validi. Per tale motivo, è stato deciso di suddividere il dataset nei cinque sottoinsiemi riportati in figura 11 e di sperimentare diversi casi.

```
original_audio=['length_ms','mean','frame_width','zero_crossings_sum','std','min','max','kur','skew']
original_audio2=['length_ms','zero_crossings_sum','std','min','max','kur','skew']
mfcc_audio=['mfcc_mean','mfcc_std','mfcc_min','mfcc_max']
sc_audio=['sc_mean','sc_std','sc_min','sc_max','sc_kur','sc_skew']
stft_audio=['stft_mean','stft_std','stft_min','stft_kur','stft_skew']
```

Figura 11: Suddivisione variabili numeriche per Clustering

#### 3.1 K-MEANS - Metodi basati su centriodi

##### 3.1.1 Scelta degli attributi

In prima istanza, è stato testato l'algoritmo per ognuno dei cinque insiemi precedentemente definiti. In seguito ad un'attenta analisi in termini di "Trade Off" tra SSE, Silhouette, cluster generati e distribuzione dei dati all'interno dei cluster; è stato deciso che il miglior insieme da utilizzare per effettuare clustering è quello definito dalla variabile **original\_audio2**. Questo insieme contiene le statistiche dell'audio originale, la lunghezza e la zero\_crossing\_sum. Nelle successive sezioni, vengono riportati lo studio del parametro  $k$ , le considerazioni sui corrispondenti cluster generati e i risultati relativi alla variabile target.

##### 3.1.2 Identificazione del miglior K

Per identificare il miglior parametro  $k$ , è stata definita una funzione che restituisce il valore presumibilmente ottimale, utilizzando il metodo *Knee Locator*. Quest'ultimo analizza gli SSE calcolati per valori di  $k$  compresi tra [2,30]. Il grafico in figura 12, mostra le variazioni dell'SSE rispetto al  $k$  considerato. La funzione restituisce inoltre SSE e Silhouette relative ad ogni istanza del K-Means.

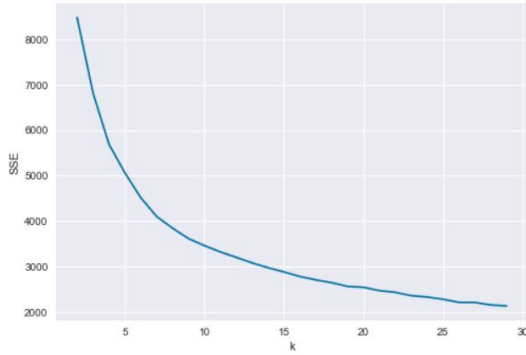


Figura 12: SSE al variare del parametro k



Figura 13: Silhouette per K=9

Confrontando i vari coefficienti di Silhouette risultanti dalle istanze dell'algoritmo ed effettuando un pò di "Trade Off" con il valore restituito dal KneeLocator, è stato scelto come valore ideale  $k = 9$  a cui corrispondono coefficienti di **Silhouette** = **0.220** e di **SSE** = **3610.577**. Nella figura 13 viene mostrato il grafico *Silhouette Plot*, utile per valutare la qualità dei clusters ricavati. Come si può evincere dal grafico, il grado di coesione non è ottimo; ma questo risultato era potenzialmente atteso.

### 3.1.3 Cluster derivati ed esplorazione dei dati

I cluster ottenuti sono stati analizzati in termini di coesione e separazione. Come si evince dal *Parallel Plot* in figura 14, le coordinate dei centroidi che determinano una migliore separazione sono quelle relative agli attributi **std** e **max**; mentre le altre non conducono a esiti positivi.

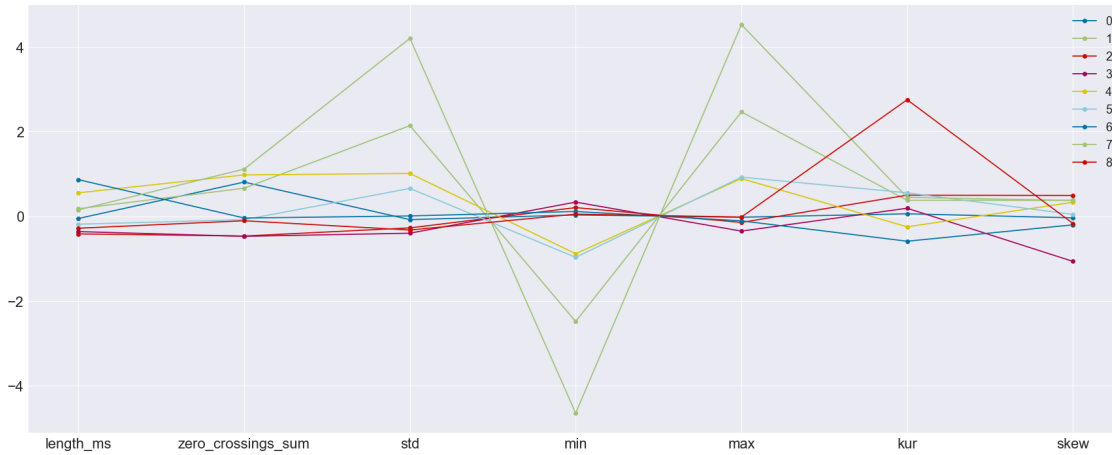


Figura 14: Analisi centroidi tramite Parallel Plot

Per quanto riguarda la coesione e la distribuzione dei dati nei diversi clusters, sono stati ottenuti i risultati visualizzabili nelle figure 15 e 16.

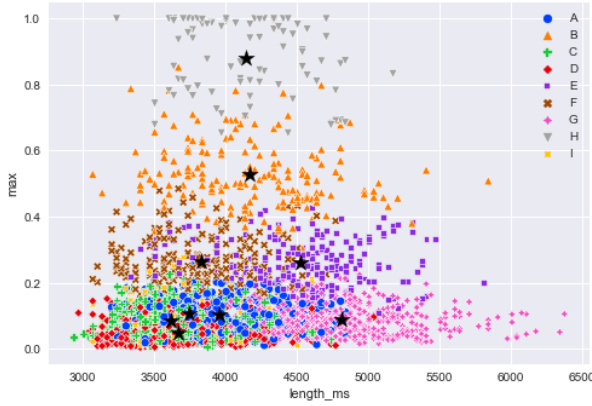


Figura 15: Analisi centroidi tramite Scatter Plot

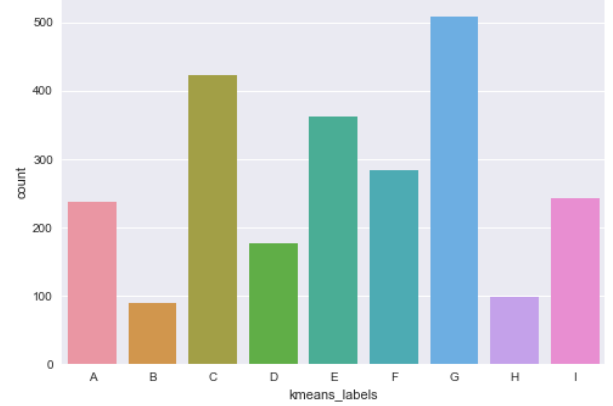


Figura 16: CountPlot dei 9 clusters

Nello *Scatter Plot*, in figura 15, realizzato prendendo in considerazione la lunghezza dell'audio e il massimo del segnale audio; è possibile notare come i centroidi dei clusters **H,B,E,G** siano separati dai centroidi dei restanti clusters. Sono state inoltre definite diverse *Cross Tab* per effettuare un confronto tra le variabili categoriche e i clusters generati in seguito all'esecuzione del K-Means con parametro  $k = 9$ . Mediante questo studio, è stato osservato che la maggior parte dei clusters contiene dati relativi a tutte le emozioni presenti nel dataset. Solamente nel cluster **H** è possibile osservare la presenza di solamente 3 emozioni (*angry*, *fearful* e *happy*), come raffigurato nella tabella in figura 17. In particolare, la maggior parte dei dati presenti sono caratterizzati dalle due emozioni negative (*angry*, *fearful*). Durante questa fase, sono stati testati differenti valori di  $k$  con il fine di riuscire a individuare un unico cluster contenente elementi relativi a un'unica emozione. Questa analisi non ha prodotto esiti positivi, probabilmente a causa della sparsità dei valori relativi alla variabile target **emotion**. Inoltre, dalle successive analisi di altre *Cross Tab*, è stato ottenuto un buon bilanciamento dei dati nei 9 clusters per gli attributi **statement** e **repetition**, come riportato nella tabella in figura 18.

kmeans_labels	A	B	C	D	E	F	G	H	I
emotion									
angry	34	83	18	9	60	55	24	71	14
calm	23	0	50	114	8	0	163	0	10
disgust	66	1	42	36	12	7	0	0	28
fearful	40	50	44	52	60	60	49	12	5
happy	30	38	55	42	63	66	65	3	7
neutral	12	0	48	42	7	1	70	0	6
sad	44	3	44	76	26	17	134	0	28
surprised	34	0	63	47	6	33	2	0	5

Figura 17: Distribuzione delle emozioni nei cluster ottenuti

kmeans_labels	A	B	C	D	E	F	G	H	I
repetition									
1st	250	126	203	127	96	42	179	46	136
2nd	244	112	207	146	96	46	170	48	143

kmeans_labels	A	B	C	D	E	F	G	H	I
statement									
Dogs are sitting by the door	97	43	220	90	161	176	254	23	137
Kids are talking by the door	140	45	203	86	200	108	254	75	105

Figura 18: Distribuzione dei valori di Repetition e Statement per i cluster ottenuti

### 3.1.4 Esperimenti con le varianti del K-Means

E' stata utilizzata la medesima routine anche per la variante Bisecting K-Means ed è stato scelto come parametro migliore  $k=7$ , condizionato da **Silhouette=0.217** e **SSE=4219.790**. Come si può notare dai

numeri, non sono stati ottenuti miglioramenti rispetto alla variante standard del K-Means, anzi in termini di coesione e separazione la situazione è peggiorata; si è dunque deciso di sperimentare anche la variante X-Means che, rispetto al precedente algoritmo, introduce una novità per quanto riguarda il criterio di terminazione anticipata (BIC) cercando di evitare situazioni di *oversplitting*. Anche questo algoritmo non ha restituito risultati sufficienti, ma ha prodotto un numero elevato di cluster, pari a 20, e un indice di coesione e separazione bassissimo. In definitiva, si è potuto affermare che per il dataset in questione il miglior algoritmo di clustering basato sui centroidi è senza ombra di dubbio il K-Means.

## 3.2 DBSCAN - Metodi basati sulla densità

### 3.2.1 Scelta degli attributi

Gli attributi presi in considerazione appartengono sempre all'insieme **originalaudio2**. Queste features sono state scelte nuovamente, con il fine di agevolare il confronto tra i differenti algoritmi. La funzione di distanza scelta è l' *Euclidean*, poiché non sono stati ottenuti risultati migliori con le altre tipologie di distanze.

### 3.2.2 Selezione dei parametri di clustering

L'utilizzo di una metrica come la Silhouette potrebbe risultare inefficiente per la scelta dei due parametri dell'algoritmo: **Epsilon** e **MinPts**; in quanto il DBscan non tiene conto della forma dei clusters ma della densità. Per risolvere questo problema è stata definita una funzione che esegue l'algoritmo DBScan con tutte le combinazioni di **Epsilon**, compreso tra [0.1,2] con uno *step*=0.1, e di **MinPts** compreso tra [5,50] e con uno *step*=5. Tale funzione restituisce 2 dataframe che rappresentano rispettivamente la distanza media tra i Noise Points e i 6 punti più vicini, e il numero di clusters che vengono individuati. In particolare, nella figura 19 vengono mostrate le *Heat Map* relative ai due dataframe generati, con  $x=MinPts$  e  $y=Epsilon$ . Da queste rappresentazioni è possibile notare come, per valori molto bassi di *Epsilon*, una gran parte dei punti sono identificati come **Noise Points** poiché vengono generati 0 clusters. Viceversa, per valori elevati si tende ad avere un solo cluster.



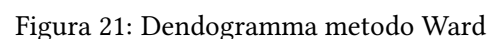
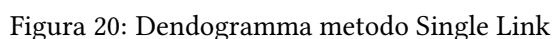
Figura 19: HeatMap dei due Dataframe

### 3.2.3 Esperimenti con variante del DBScan

### 3.3 Clustering Gerarchico

Per il clustering gerarchico sono stati utilizzati gli stessi attributi scelti per i modelli precedenti, ovvero **original\_audio2**, invece come funzione di distanza l'*Euclidea*.

L'algoritmo utilizzato è di tipo agglomerativo ed è stato eseguito variando diverse configurazioni di parametri relativi al criterio di connessione tra i clusters (*single, average, complete, ward*) e all'uso di una matrice di connettività, per il calcolo delle distanze. Tale matrice viene generata dal grafo ottenuto dai primi 100 vicini di ogni nodo. In prima istanza l'algoritmo è stato testato senza il numero di cluster predefinito passato come parametro e senza matrice di connettività; i dendrogrammi relativi ai primi esperimenti sono visibili nelle Figure 20, 21, 22 e 23.



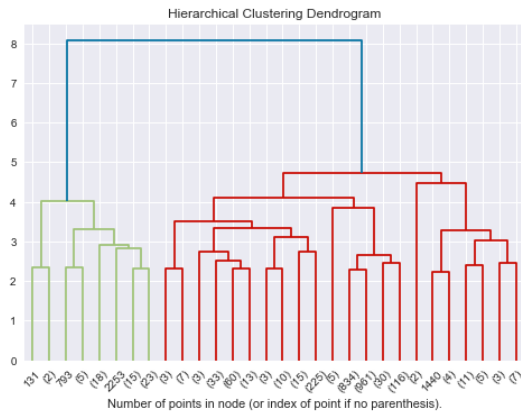


Figura 22: Dendrogramma Metodo Average

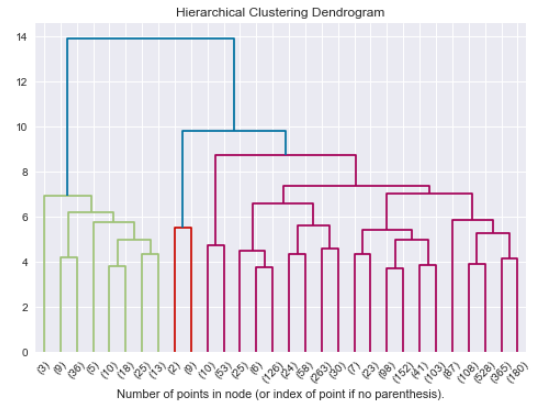


Figura 23: Dendrogramma Metodo Complete

Dopo aver stabilito i tagli più interessanti, ovvero quelli che generano rispettivamente 3, 4 e 5 clusters; è stato rieseguito l’algoritmo agglomerativo variando i criteri di collegamento ed utilizzando la matrice di connettività. Nella tabella in figura 24 vengono sintetizzati i risultati ritenuti più significativi tra quelli conseguiti.

METODO	DIMENSIONE CLUSTERS	COEFF. SILHOUETTE
Ward	318, 1034, 1065	0.201
Connectivity + Ward	318, 1033, 1066	0.201
Average	33, 66, 2318	0.408
Connectivity + Average	2, 2, 2413	0.543
Complete	11, 63, 119, 2224	0.433
Connectivity + Complete	2, 2, 450, 1963	0.071

Figura 24: Confronto metodi clustering gerarchico

Il risultato migliore è stato ottenuto tramite il criterio **Ward**, il quale realizza clusters nettamente più bilanciati rispetto agli esperimenti precedenti. Inoltre il risultato, indipendentemente dall’ausilio della matrice di connettività, non cambia; persiste però un basso coefficiente di separazione causato da un basso valore di Silhouette. E’ possibile notare che per i criteri **Average** e **Complete** si ottengono raggruppamenti totalmente sbilanciati, nonostante la presenza di un modesto coefficiente di Silhouette tendente a 0.400 per entrambi i metodi. Questi due criteri, se eseguiti in combinazione con la matrice di connettività, conducono ad un risultato peggiore del precedente poiché viene creato un cluster contenente il 99% dei dati per il criterio *Average* e l’81% dei dati per *Complete*.

### 3.4 Valutazioni finali di clustering

Nella tabella in figura 25 vengono riportati i risultati migliori ottenuti dagli algoritmi di clustering sperimentati. Gli esiti insoddisfacenti degli algoritmi basati sulla densità, ovvero DBScan e Optics, non sono

stati presi in considerazione per questa valutazione finale in quanto non sono stati ritenuti sufficienti per via dei valori delle metriche affrontate.

ALGORITMO	N. CLUSTERS E DIMENSIONI	SILHOUETTE
K-Means	9 → 283, 177, 354, 88, 235, 416, 108, 497, 259	0.220
Gerarchico(ward)	3 → 318, 1034, 1065	0.201
Gerarchico(ward + connectivity)	3 → 318, 1033, 1066	0.201

Figura 25: Confronto algoritmi di Clustering

Il confronto finale è stato effettuato tra K-Means e Clustering Gerarchico. Per quanto riguarda l'algoritmo basato sui centroidi, sono stati ottenuti buoni risultati in termini di distribuzione e coesione dei dati; nonostante un basso valore di Silhouette. Invece, per l'algoritmo di clustering gerarchico che utilizza la tecnica agglomerativa con una configurazione di tre cluster, è stato ottenuto un risultato sufficiente per la suddivisione degli elementi, ma un basso indice di separazione inter-clusters. I medesimi risultati sono stati ottenuti con e senza matrice di connettività. In conclusione, si può affermare che il **K-Means** è l'algoritmo che meglio ha performato in questo dataset; soprattutto per quanto riguarda l'esplorazione dei dati.

## 4. Classificazione

Nella seguente sezione vengono illustrati gli algoritmi di classificazione utilizzati per prevedere i valori delle variabili categoriche, ponendo particolare attenzione agli attributi *emotion* ed *emotional\_intensity*. Per questo task sono stati utilizzati tutti gli attributi presenti nel dataset, pre-processando opportunamente i dati. La sezione si conclude con due esperimenti di regressione univariata e multivariata con variabile dipendente *length\_ms*.

### 4.1 Pre-processing

Il dataset di riferimento è sempre quello definito in fase di Data Preparation. I dati sono stati preprocessati a seconda della variabile target in questione, in particolare per la classificazione delle emozioni è stata conseguita una trasformazione tramite il metodo "Label Encoding" per gli attributi categorici; per tutti gli altri esperimenti invece è stata definita una codifica "One Hot Encoding" per la variabile *emotion* e nuovamente "Label Encoding" per le altre categoriche. I dati numerici sono stati trattati con la stessa metodologia della sezione di Clustering, quindi normalizzati con il metodo Robust Scaler per poter gestire l'ingente numero di outliers. Infine le informazioni sono state preparate per i task di classificazione e regressione andando a dividere il dataset in 70% training e 30% test, applicando una stratificazione solamente per l'esperimento con la variabile target *emotion*, in quanto sbilanciata.

### 4.2 Scelta degli Iperparametri

In questa fase è stata utilizzata la funzione *GridSearchCV* assistita da una *5-fold Cross Validation* ripetuta 10 volte in modo da identificare la migliore configurazione di iperparametri per allenare i modelli basati



su Decision Tree e KNN. La metrica definita per la resituzione della configurazione migliore è l' F1\_macro in quanto l'Accuracy potrebbe portare a modelli privi di informazioni a causa dello sbilanciamento del dataset per determinate variabili. Per quanto riguarda il modello che sfrutta il Naive Bayes non è stato effettuato alcun tuning di iperparametri. Nella Figura 26 viene mostrata la griglia utilizzata durante la fase di validazione dei classificatori.

Decision Tree	KNN
criterion: Gini, Entropy	n_neighbours: arange (3,30,2)
max_depth: None,2,5,10,15,20	weights: uniform,distance
min_samples_split: 2,5,10,15,20	metric: euclidean,manhattan,minkowski
min_samples_leaf: 1,5,10,15,20	

Figura 26: Griglia tuning iperparametri

## 4.3 Emotion

In questa sezione vengono riportati i risultati ottenuti dalla classificazione della variabile categorica *emotion*.

### 4.3.1 Decision Tree

La migliore combinazione di iperparametri restituita dalla Grid Search è data da criterion=Gini, max\_depth=20, min\_samples\_split=2, min\_samples\_leaf=1 che permette di ottenere i valori esposti in figura 27. E' possibile osservare che le classi *angry*, *calm* e *surprised* hanno riportato i risultati migliori per la metrica f1\_score.

	Precision	Recall	f1-score
angry	0.59	0.61	0.60
calm	0.54	0.49	0.51
disgust	0.35	0.41	0.38
fearful	0.43	0.36	0.39
happy	0.34	0.31	0.32
neutral	0.34	0.34	0.34
sad	0.37	0.41	0.39
surprised	0.40	0.51	0.45
accuracy			0.43
macro avg	0.42	0.43	0.42

Figura 27: Performance classificatore

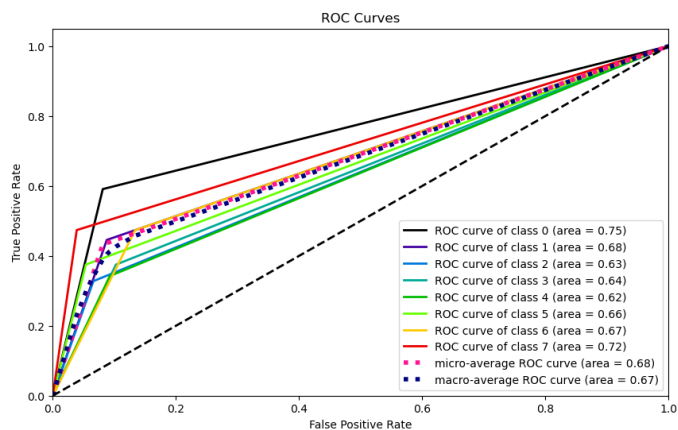


Figura 28: Roc curve delle diverse classi

In Figura 28 è rappresentata la relativa ROC curve. Anche da questo grafico si può notare come le classi sopracitate vengano classificate meglio rispetto alle altre, in quanto la loro curva si avvicina maggiormente al valore 1 di TPR.



#### 4.3.2 KNN e Naive Bayes

Dall'esecuzione della Grid Search è stata ottenuta la seguente combinazione di iperparametri: `n_neighbours=12`, `metric=manhattan`, `weights=distance` per l'allenamento del modello. Anche per il KNN sono stati ottenuti risultati migliori per le solite tre classi, ma valori comunque inferiori rispetto a quelli ottenuti con il Decision Tree. Per il Naive Bayes non è stata condotta alcuna ricerca esaustiva degli iperparametri e sono state testate la tipologia Gaussiana e Multinomiale. Nella Figura 29 sono sintetizzati i risultati del miglior classificatore bayesiano e del KNN.

	KNN	Naive Bayes
F1_macro avg	0.41	0.35
Accuracy	0.43	0.35

Figura 29: Performance classificatori

#### 4.3.4 Confronto tra classificatori

Tra i tre modelli utilizzati per classificare la variabile target *emotion*, l'algoritmo che ha prodotto risultati migliori è stato il Decision Tree il quale, come mostrato in figura 27, ha riportato performance preferibili rispetto agli altri due algoritmi esaminati; in particolare si hanno valori più elevati anche per le classi intermedie. L'elevato sbilanciamento di valori per questa variabile target porta a risultati di performance dei classificatori che non superano il 50%, si è proceduto infatti a classificare i dati utilizzando altri attributi categorici.

### 4.4 Emotional Intensity

In questa sezione vengono riportati i risultati ottenuti dalla classificazione della variabile target *emotional\_intensity*.

#### 4.4.1 Decision Tree

Attraverso l'utilizzo del modello allenato con la configurazione di parametri restituita dalla ricerca esaustiva, sono stati ottenuti i risultati presenti in figura 30. Tramite questi valori è possibile notare come per entrambe le classi *strong* e *normal* si ottiene un valore abbastanza alto di F1\_score, pari rispettivamente a 0.72 e 0.75. Conseguentemente sono stati ottenuti buoni valori anche per le metriche precision e recall per entrambe le classi. Questi ultimi indicano che il classificatore commette un numero basso di errori falsi positivi e di errori falsi negativi. Per quanto riguarda l'accuracy, essendo anche le due classi abbastanza bilanciate, è stato ottenuto un valore pari a 0.74.

	Precision	Recall	f1- score
<b>normal</b>	0.75	0.75	0.75
<b>strong</b>	0.72	0.72	0.72
<b>accuracy</b>			0.74
<b>macro avg</b>	0.73	0.73	0.73

Figura 30: Performance classificatore

	Precision	Recall	f1- score
<b>normal</b>	0.73	0.73	0.73
<b>strong</b>	0.70	0.71	0.70
<b>accuracy</b>			0.72
<b>macro avg</b>	0.72	0.72	0.72

Figura 31: Performance classificatore (10 Features)

Successivamente è stato conseguito lo stesso task utilizzando il sottoinsieme di variabili che sono risultate più rilevanti, le quali sono visibili in Figura 33, per cercare di realizzare un modello più performante. Con l'utilizzo di questo insieme non sono stati ottenuti i risultati sperati in quanto il modello realizzato non risulta migliore sul medesimo test set, come si può notare dalla Figura 31. Dalla heatmap relativa al primo esperimento di classificazione, in Figura 32, si può notare come entrambe le classi presentino il medesimo numero di mismatch.

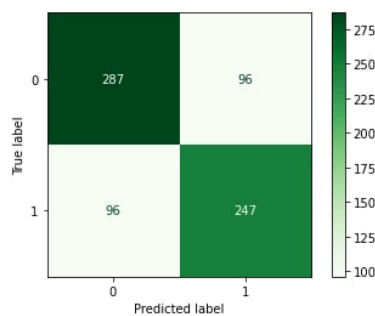


Figura 32: Confusion Matrix

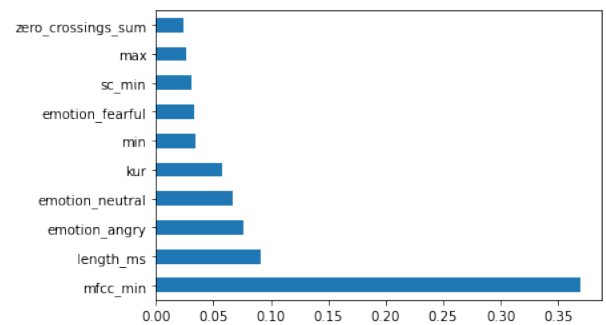


Figura 33: Features Rilevanti

Infine in figura 34, viene visualizzato l'albero ottenuto con il sottoinsieme di variabili selezionate. Da questo grafo possiamo comprendere come lo spettro di potenza del segnale audio fornisca la maggior quantità informativa ai fini della classificazione. In particolare si può già dedurre dal primo split come il sottoalbero destro generi prevalentemente foglie classificate come "strong" mentre la parte duale, a parità di profondità, generi per lo più foglie della restante classe. Da una profondità superiore o pari a 10 le foglie delle relative etichette appaiono in egual misura.

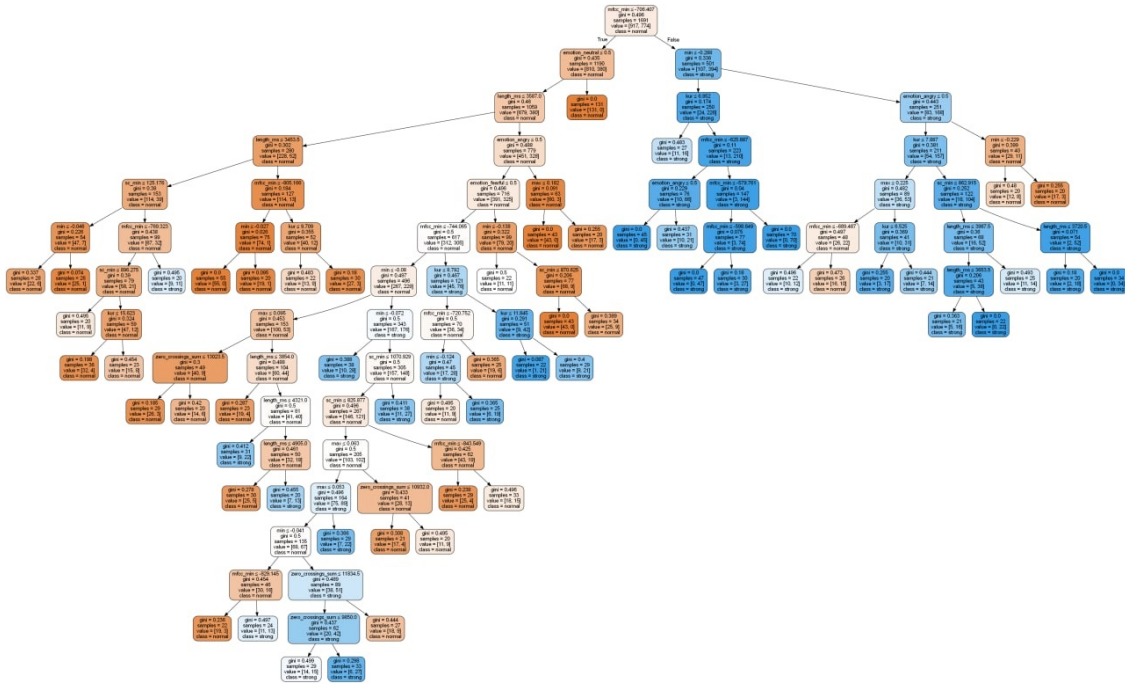


Figura 34: Decision Tree (10 Features)

#### 4.4.2 KNN e Naive Bayes

Dagli esperimenti condotti con questi classificatori, sono stati ottenuti i risultati riportati in tabella 35. Come si può notare dai valori il metodo bayesano, per questa variabile target, ha performato leggermente meglio nonostante la consueta ricerca esaustiva degli iperparametri per il KNN.

	KNN	Naive Bayes
<b>F1_macro avg</b>	0,71	0,73
<b>Accuracy</b>	0,72	0,74

Figura 35: Performance classificatori

#### 4.4.3 Confronto tra classificatori

Confrontando gli esperimenti effettuati su *emotional\_intensity*, attraverso l'utilizzo dei tre classificatori, è possibile concludere che il Decision Tree e il Naive Bayes hanno condotto a risultati migliori ottenendo in linea di massima le stesse performance e di conseguenza riescono a classificare meglio questa variabile target rispetto al KNN.

Infine i dati sono stati classificati anche utilizzando le restanti variabili categoriche ottenendo risultati tendenti anche al 90%, per quanto riguarda ad esempio l'attributo *Sex*. Questi esperimenti non sono stati riportati in quando le variabili in esame risultano perfettamente bilanciate all'interno del dataset e dunque poco interessanti da esaminare per il task corrente.

## 4.5. Regressione

Come variabile target (Y) per la regressione è stata scelta la lunghezza del segnale audio originale; per il primo esperimento di regressione ad una variabile indipendente sono state esaminate differenti X e il risultato presumibilmente migliore è stato ottenuto dall'associazione " $sc\_skew \implies length\_ms$ ". Gli attributi sono stati normalizzati sempre con il RobustScaler. Nella Figura 36 vengono mostrati i risultati dei vari classificatori lineari e non, provati per questo primo esperimento ad una variabile indipendente.

	R2	MSE	MAE
Linear Regression	0.435	0.244	0.388
Ridge	0.435	0.244	0.388
Lasso	-0.001	0.431	0.552
KNN	0.431	0.245	0.378
Decision Tree	0.410	0.254	0.384

Figura 36: Performance regressori (esperimento univariato)

Come si può vedere dalla griglia, il modello con la potenza predittiva migliore (R2) è determinato dal Regressore Lineare che ottiene anche una discreta minimizzazione degli errori quadratici e assoluti. Interessante è il caso del Lasso la cui capacità predittiva è definita da un numero negativo, le cui cause possono essere relative ad una scelta sbagliata del modello o dei parametri. I regressori non lineari, nonostante una ricerca dei migliori iperparametri, non hanno performato come sperato in questo esperimento di regressione univariata. Nelle Figure 37 e 38 viene mostrato il risultato predittivo del primo regressore lineare e del KNN tramite scatter plots.

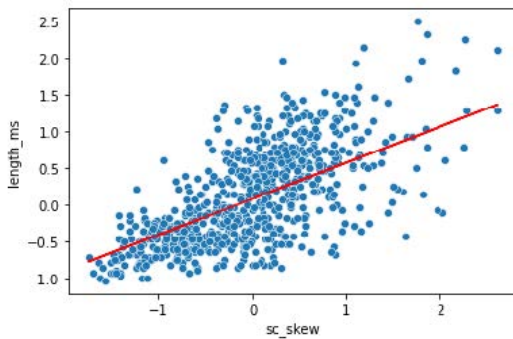


Figura 37: Analisi Regressore Lineare

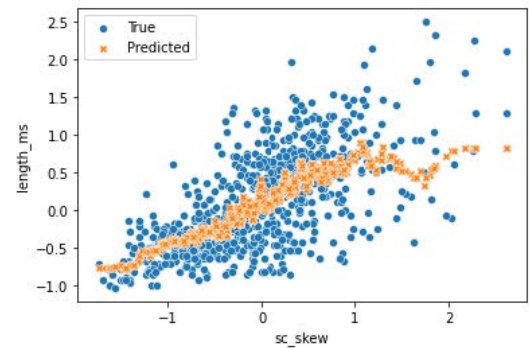


Figura 38: Analisi KNN

I risultati non sono stati considerati soddisfacenti, si è proceduto dunque con il secondo esperimento di regressione multivariata dove sono state utilizzate come variabili indipendenti tutti gli attributi del dataset preparato nella Sezione 2. I dati sono stati pre-processati tramite Label Encoding e RobustScaler. Sono stati utilizzati i medesimi classificatori del primo esperimento, nella Figura 39 i risultati.

	R2	MSE	MAE
Linear Regression	0.779	0.095	0.239
Ridge	0.779	0.095	0.239
Lasso	-0.001	0.431	0.552
KNN	0.772	0.098	0.232
Decision Tree	0.687	0.135	0.278

Figura 39: Performance regressori

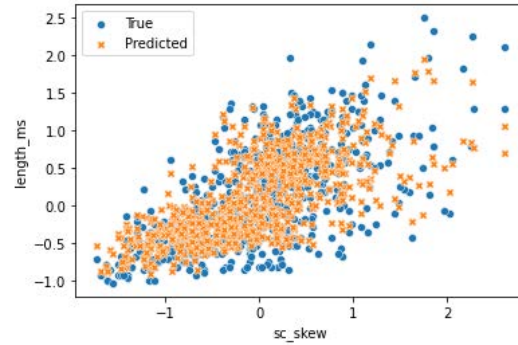


Figura 40: Capacità predittiva KNN

In linea generale i modelli realizzati hanno ottenuto performance decisamente migliori rispetto all'esperimento ad una variabile indipendente. Si può notare come nuovamente i primi due classificatori restituiscano i medesimi risultati, come nel caso precedente; si può dedurre infatti che la regolarizzazione di Tikhonov non intervenga sul calcolo delle relative metriche. Il Lasso soffre nuovamente del solito problema; mentre i regressori non lineari non superano a livello di performance quelli lineari nemmeno in questo secondo esperimento. Nella Figura 40 viene mostrata la capacità predittiva del KNN i cui risultati coincidono quasi con quelli del primo regressore lineare.

## 5. Pattern Mining

In questa sezione finale vengono individuati e analizzati i pattern più frequenti e le regole di associazione più interessanti, ponendo particolare attenzione alla variabile target *emotional\_intensity*. Tutti i risultati sono stati ottenuti mediante l'applicazione dell'algoritmo Apriori.

### 5.1 Pre-processing

Facendo sempre riferimento al dataset realizzato in fase di Data Preparation nella sezione 2, per l'applicazione dell'algoritmo di Pattern Mining, i dati sono stati pre-processati nel seguente modo:

- Eliminazione delle variabili derivate dal file audio originale, ritenute poco interessanti e informative per questo task.
- Discretizzazione in 4 intervalli per gli attributi numerici
- Ridenominazione in *value\_AttributeName* dei valori contenuti in ogni variabile presa in considerazione per questo task, in modo da rendere più leggibili le regole generate dall'algoritmo.
- Trasformazione del dataset in un basket (lista) di transazioni.

Il conseguente dataset transazionale risulta quindi caratterizzato da due gruppi di variabili:

- vocal\_channel, emotion, emotional\_intensity, sex, length\_ms, zero\_crossings\_sum
- statistiche segnale audio originale (mean, min, max, kur, skew)

## 5.2 Estrazione e analisi dei frequent patterns al variare di support

Per individuare i pattern più interessanti sono stati provati diversi valori per i parametri dell'algoritmo. In particolare sono stati estratti i frequent, closed, maximal itemsets al variare del supporto nell'intervallo percentuale [5,50], rispetto a differenti soglie di lunghezza minima: 2,3,4,5. Il risultato dell'esperimento è visibile in Figura 41.

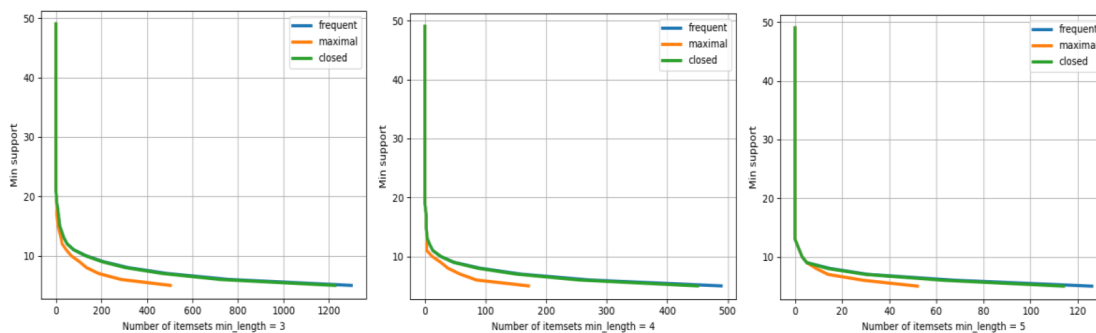


Figura 41: Itemsets al variare della lunghezza minima

In ognuna delle figure sopra riportate si può notare come i frequent e closed itemset coincidano sempre e come per un valore di support superiore al 20% gli insiemi generati siano nulli. Si è deciso di analizzare inizialmente gli insiemi più interessanti, con lunghezza minima 5 e una soglia di support del 10% e si è evinto, come mostrato nello schema 42, una stretta correlazione tra gli intervalli di *std*, *max*, *min* riportati e l'intensità emozionale "strong". Dalla relazione tra i primi due itemset si può dedurre che in linea generale gli oggetti caratterizzati dai precedenti intervalli e da un segnale audio di tipo "speech" sono interpretati da attori femminili, come era stato scoperto anche in sezione 2.

Frequent itemset	Support
{0.026, 0.152]_Std, {0.225, 0.999]_Max, {-1.0, -0.203]_Min, strong_Intensity, F	10.22
{0.026, 0.152]_Std, {0.225, 0.999]_Max, {-1.0, -0.203]_Min, strong_Intensity, speech_Vc	11.17
{-0.00035, 0.00677]_Std, {0.004, 0.055]_Max, {-0.054, -0.006]_Min, normal_Intensity, speech_Vc	12.24

Figura 42: Itemset generati (supp=10% e zmin=5)

Per ottenere associazioni tendenti al 20% di soglia di supporto è necessario considerare gli itemset di lunghezza minima 3; nello schema seguente in figura 43 vengono riportati i primi risultati ottenuti che riprendono in parte le considerazioni relative all'esperimento precedente. Si noti inoltre come nell'ultima transazione venga generato un set che correla la lunghezza del segnale audio ad una intensità questa volta di tipo "normal".

Frequent itemset	Support
{0.0262, 0.152]_Std, {0.225, 0.999]_Max, {-1.0, -0.203]_Min	20.93
{0.225, 0.999]_Max, {-1.0, -0.203]_Min, strong_Intensity	19.61
{0.0262, 0.152]_Std, {-1.0, -0.203]_Min, strong_Intensity	18.61
{0.0262, 0.152]_Std, {0.225, 0.999]_Max, strong_Intensity	18.57
{2935.999, 3604.0]_Length, normal_Intensity, speech_Vc	18.53

Figura 43: Itemsets generati (supp=18%, zmin=3)

Come si evince dalla Figura 44, si è cercato infine di ottenere degli itemset caratterizzati da almeno un oggetto dell'attributo *Emotion* ottenendo però risultati validi con una soglia di support che non supera il 10%, considerata troppo esigua per ottenere conoscenza dai dati.

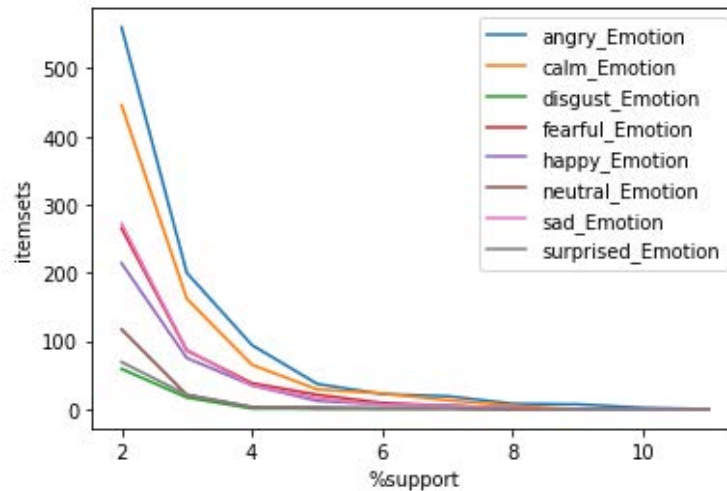


Figura 44: Distribuzione itemsets con emozioni al variare di support

### 5.3 Estrazione e analisi di regole al variare di confidence

E' stata definita una routine per analizzare il numero di regole generate per valori di support e confidence compresi negli intervalli percentuali [5,20] e [40,90] rispettivamente, utilizzando una soglia minima di lunghezza 4. Nella Figura 45 viene mostrato il Lift delle regole generate al variare dei due parametri, in particolare dal grafico di destra si può notare come un lift maggiore di 2.5 venga ottenuto per le regole con indice di confidenza tendenzialmente maggiore del 60%.



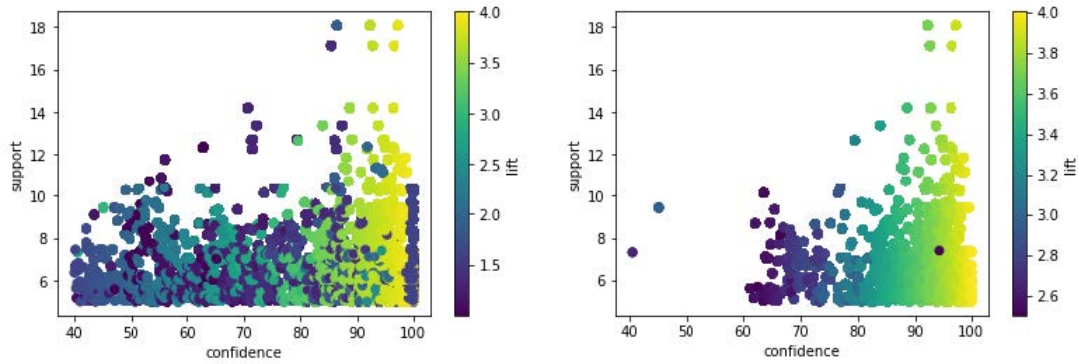


Figura 45: Analisi Lift AR

Considerando i valori di support uguale al 10% e confidence uguale a 60% si ottengono 93 regole di cui solo 3 con lift minore a 1 alle quali corrisponde lo stesso conseguente, come visibile nella Figura 46.

Rule	Support	Confidence	Lift
{0.0262, 0.152]_Std, {0.225, 0.999]_Max, {-1.0, -0.203]_Min, strong_Intensity => speech_Vc	11.17	61.78	0.988
{0.0262, 0.152]_Std, {0.225, 0.999]_Max, strong_Intensity => speech_Vc	11.29	60.80	0.973
{0.0262, 0.152]_Std, {-1.0, -0.203]_Min, strong_Intensity => speech_Vc	11.37	61.11	0.978

Figura 46: AR con Lift minore di 1 (support=10% confidence=60%)

Se il valore di Lift non fosse stato così basso tali regole sarebbero state delle buone candidate per effettuare un'operazione di sostituzione dei valori mancanti o predizione della variabile *Vocal\_channel* che assume valori "song, speech". Nelle successive Figure 47 e 48 viene mostrata, al variare di confidence e lift, la distribuzione delle regole ottenute:

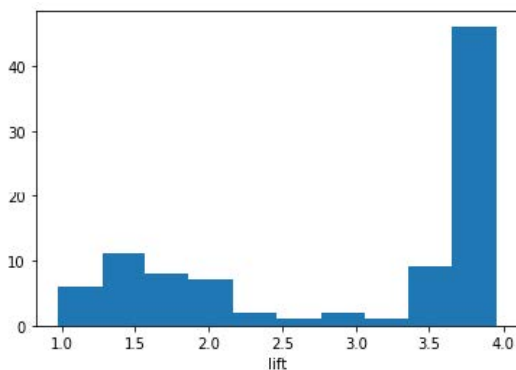


Figura 47: Distribuzione AR al variare di Lift

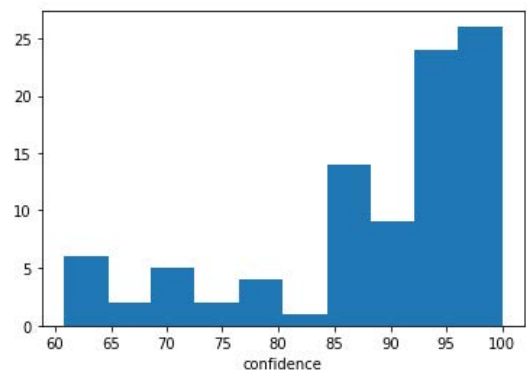


Figura 48: Distribuzione AR al variare di Confidence

Data l'ingente distribuzione di regole con un lift molto alto, si è deciso di analizzare i risultati con un lift maggiore di 3.5 scoprendo che tutti i conseguenti riguardano le statistiche del segnale audio originale divise per intervalli. Nella figura 49 vengono mostrate le cinque con il valore di Lift più alto che denotano una stretta dipendenza tra antecedente e conseguente.



Rule	Support	Confidence	Lift
{0.0262, 0.152]_Std, {0.225, 0.999]_Max, strong_Intensity, speech_Vc => {-1.0, -0.203]_Min	11.17	98.90	3.95
{0.0262, 0.152]_Std, {0.225, 0.999]_Max, speech_Vc=> {-1.0, -0.203]_Min	11.83	98.62	3.93
{0.0262, 0.152]_Std, {-1.0, -0.203]_Min, strong_Intensity, speech_Vc => {0.225, 0.999]_Max	11.17	98.11	3.92
{0.0262, 0.152]_Std, {-1.0, -0.203]_Min, strong_Intensity, F => {0.225, 0.999]_Max	10.21	98.01	3.92
{0.0262, 0.152]_Std, {0.225, 0.999]_Max, {14959.0, 30153.0]_Zcs => {-1.0, -0.203]_Min	10.26	98.02	3.91

Figura 49: AR con Lift maggiore

## 5.4 Predizione variabile target e valutazione accuratezza

Considerando le regole estratte dal precedente esperimento, quindi con supporto e confidence uguali a 10% e 60% rispettivamente, si è deciso di focalizzare l'attenzione sui risultati con Lift maggiore di 1 e in particolare con il conseguente relativo ad una delle due classi della variabile *emotional.intensity*. Nelle Figure 50 e 51 vengono mostrate alcune delle 8 regole con consequent uguale a "normal" e alcune delle 9 regole con consequent uguale a "strong", utilizzate per predire i valori della variabile target.

Antecedent	Support	Confidence	Lift
{-0.000348, 0.00677]_Std, {2935.999, 3604.0]_Length, speech_Vc	10.34	86.20	1.60
{4720.999, 10359.0]_Zcs, {2935.999, 3604.0]_Length, speech_Vc	10.38	75.37	1.40
{-0.0544, -0.00613]_Min, {-0.000348, 0.00677]_Std, speech_Vc	13.36	72.25	1.34

Figura 50: AR consequent=normal

Antecedent	Support	Confidence	Lift
{0.0262, 0.152]_Std, {0.225, 0.999]_Max, speech_Vc	11.29	94.13	2.03
{0.0262, 0.152]_Std, {0.225, 0.999]_Max, {-1.0, -0.203]_Min, F	10.21	86.97	1.88
{0.0262, 0.152]_Std, {0.225, 0.999]_Max, F	10.42	86.89	1.88

Figura 51: AR consequent=strong

Nella Figura 52 viene mostrato il risultato della predizione tramite un report di classificazione e la heatmap relativa alla matrice di confusione. Come si può notare, non sono stati ottenuti risultati migliori rispetto ai classificatori di *emotional.intensity* della sezione 4 per nessuna metrica. Dato anche l'elevato errore nella predizione della classe negativa "strong", questa metodologia non è consigliata per la classificazione della variabile presa in esame.

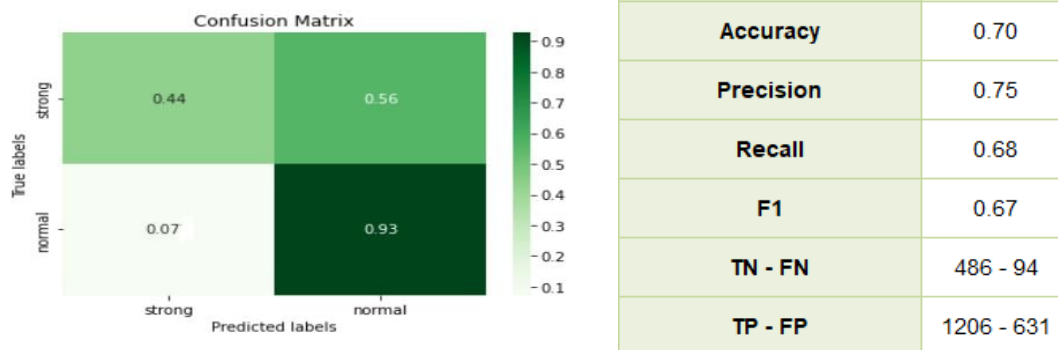


Figura 52: Performance della previsione tramite AR