Università di Pisa

Department of INFORMATICA

Master of Data Science and Business Informatics

PROJECT REPORT

THE RYERSON AUDIO-VISUAL DATABASE OF EMOTIONAL SPEECH AND SONG (RAVDESS)

Submitted by:

Michele DiCandia (657494)

Enrico Cilia (563434)

Academic Year 2022/2023

TABLE OF CONTENTS

Sommario

CHAPTER 1

DATA UNDERSTANDING AND PREPARATION

This report examines the Ryerson Audio-Visual Database of Emotional Speech and Song. RAVDESS contains audio of 24 professional actors (12 female, 12 male), vocalizing two statements in a neutral North American accent. Speech includes calm, happy, sad, angry, fearful, surprise, and disgust expressions, and song contains calm, happy, sad, angry, and fearful emotions. Each expression is produced at two levels of emotional intensity (normal, strong), with an additional neutral expression.

1.1 DATA SEMANTICS

RAVDESS dataset is a dataset consisting of 2452 different audios recorded by both female and male actors with different vocals, intensities, and emotions. Each record within the dataset is described by 38 attributes both categorical, binary and numerical. In Table 1.1 we list for each attribute the name, description, the type and the domain associated with it.

TABLE 1.1: Description of dataset attributes

NAME	TYPE	DOMAIN	DESCRIPTION
Modality	Categorical	Audio – only	Type of recording
Emotion	Categorical	neutral, calm, happy, sad, angry, fearful, disgust, surprised	Attribute that takes on different values depending on the mood expressed in the audio.
Emotional Intensity	Categorical	normal, strong	Each emotion expressed in an audio has two levels of emotional intensity.
Statement	Categorical	"Kids are talking by the door", "Dogs are sitting by the door"	In each audio, the actors recite one of two statements.
Repetition	Categorical	1st repetition, 2nd repetition	Each statement can be repeated once or twice within the same audio.
Actor	Categorical (float)	01 - 24	There are 24 actors in the dataset who participated in the audio recording.

Sex	Binary	M - F	There are 24 actors in the dataset who participated in the audio recording.
Channels	Binary	mono, stereo audio	Attribute indicating the number of channels used.
Sample width	Binary	8-bit, 16-bit	Number of bytes
Frame Rate	Numeric	48000	Frequency of samples used (in Hertz)
Frame width	Numeric	2 - 4	Number of bytes for each frame. One frame contains a sample for each channel.
Length ms	Numeric	2936 -6373	Audio file length (in milliseconds).
Frame Count	Numeric	-1.0 - 305906.0	The number of frames from the sample.
Intensity	Numeric	-63.86 - 30153	Loudness in dBFS (dB relative to the maximum possible loudness)
Zero crossing sum	Numeric	30153 - 4721	Sum of the zero-crossing rate.
Mean	Numeric	-0.00 >N< 0.00	Statistics of the original audio signal.
Std	Numeric	0.00 > N < 0.15	Statistics of the original audio signal.
Min	Numeric	-0.99> N < 0.00	Statistics of the original audio signal.
Max	Numeric	0.00> N < 0.99	Statistics of the original audio signal.
skew	Numeric	-2.35> N < 1.79	Statistics of the original audio signal.
Mfcc_mean	Numeric	-15.49> N <-43.81	statistics of the MelFrequency Cepstral Coefficients
Mfcc_std	Numeric	83.62> N <195.94	statistics of the MelFrequency Cepstral Coefficients
Mfcc_min	Numeric	-1085.47> N <-461.48	statistics of the MelFrequency Cepstral Coefficients
Mfcc_max	Numeric	126.25> N <280.17	statistics of the MelFrequency Cepstral Coefficients
Sc_mean	Numeric	2360.88>N < 7655.33	statistics of the spectral centroid
Sc_std	Numeric	1491.34 > N < 4819.78	statistics of the spectral centroid
Sc_min	Numeric	0.0> N< 2121.41	statistics of the spectral centroid
Sc_max	Numeric	7657.49> N <17477.54	statistics of the spectral centroid
Sc_kur	Numeric	-1.79> N < 3.65	statistics of the spectral centroid
Stft_mean	Numeric	0.72 > N < 0.21	statistics of the stft chromagram
Stft_std	Numeric	0.21 > N < 0.39	statistics of the stft chromagram
Stft_min	Numeric	0.0> N < 0.03	statistics of the stft chromagram
Stft_max	Numeric	1.0>N<1.0	statistics of the stft chromagram
Stft_kur	Numeric	-1.79 >N< 3.65	statistics of the stft

			chromagram
Stft_skew	Numeric	-0.51> N < 1.82	statistics of the stft chromagram

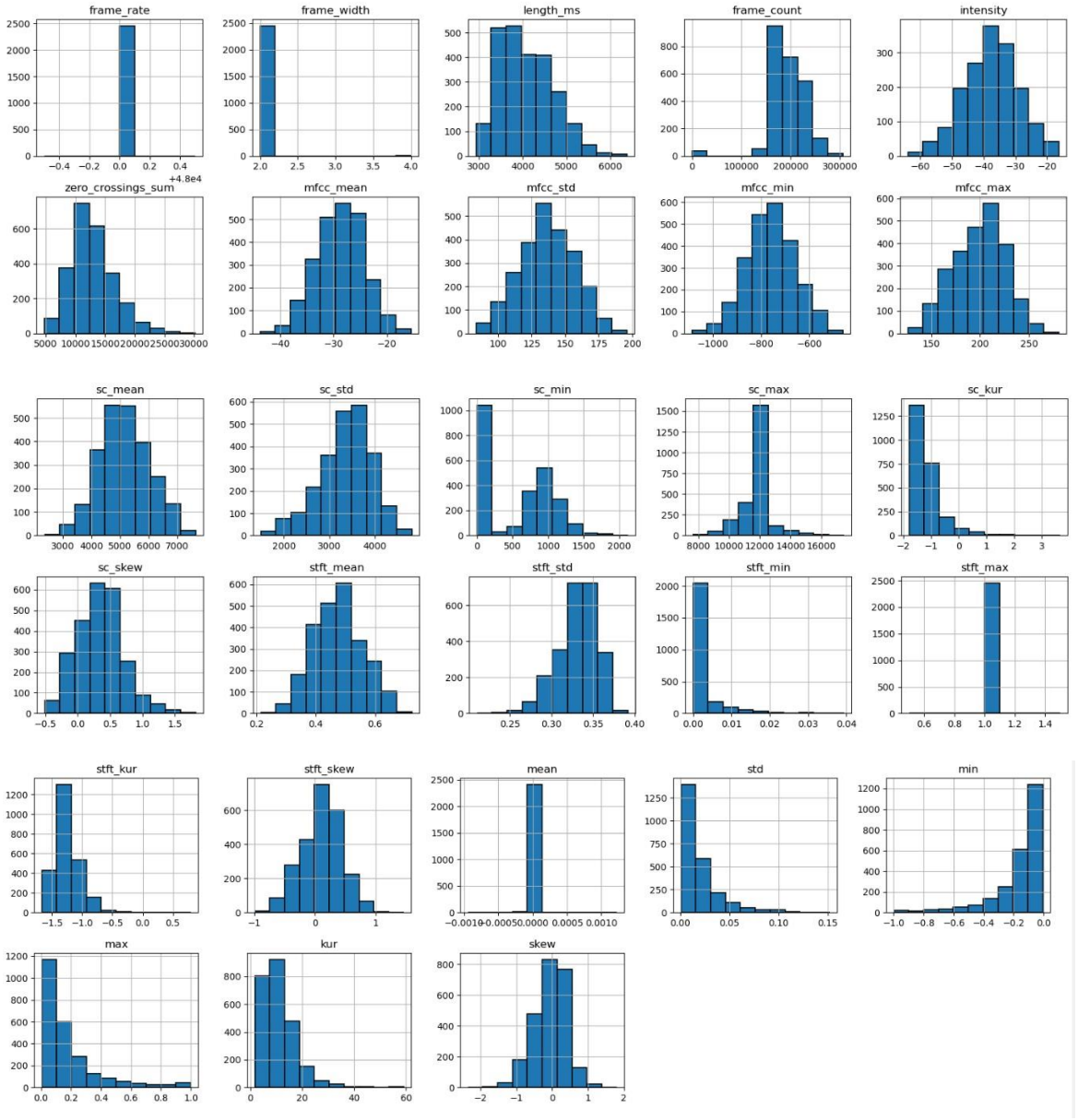
THE MEL FREQUENCY CEPSTRAL COEFFICIENTS (MFCCs) of a signal are a small set of features which concisely describe the overall shape of a spectral envelope.

THE SPECTRAL CENTROID (SC): It is a measure of the amplitude at the center of the spectrum of the signal distribution over a window calculated from the Fourier transform frequency and amplitude information.

STFT CHROMAGRAM: It defines a particularly useful class of time-frequency distributions which specify complex amplitude versus time and frequency for any signal.

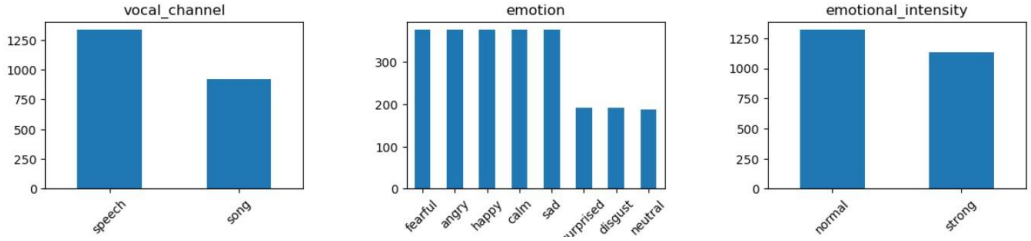
1.2 DISTRIBUTION OF THE VARIABLES AND STATISTICS

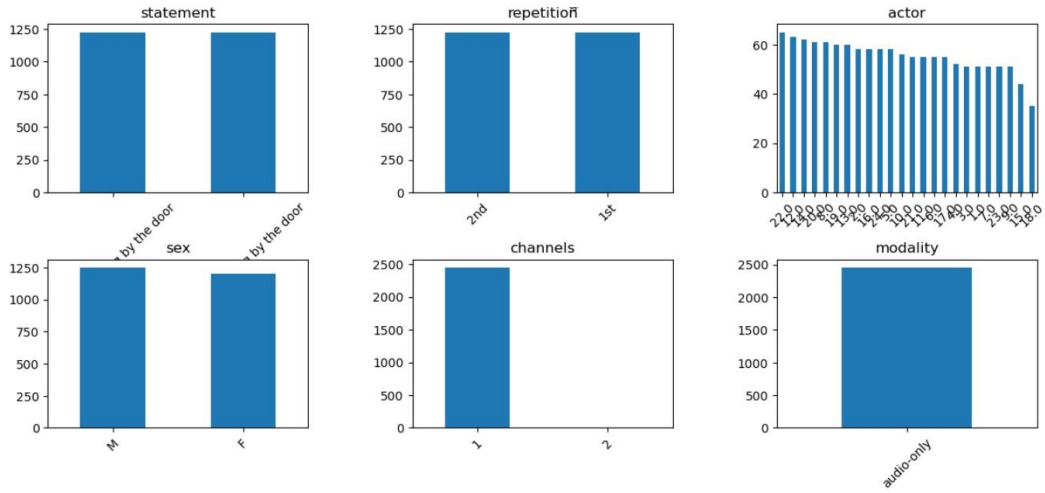
The data we find in the table above can be depicted through the use of certain graphs. Let us start by visualizing the distribution of numerical data using histograms:



These histograms show the frequency distribution of the numerical attributes of the dataset, where the attribute intervals are discretized into a fixed number of intervals (BINS). For each interval, the (absolute) frequency of the values within it is indicated by the height of the single bar.

Next, we plot the categorical variables using Bar Charts:





Bar Charts are better suited to depicting the frequencies of categorical attribute values because they are data measured on a scale with specific potential values.

1.3 EVALUATING DATA QUALITY

This section deals with understanding some general information concerning the data (missing values, outliers, correlation). In the next subsections we are going to implement the Data Preparation phase in which we try to solve problems related to the data we have. The goal is to improve the quality of the data by also going to reduce the size of the dataset, if necessary, and choosing the attributes we are most interested in.

1.3.1 MISSING VALUES

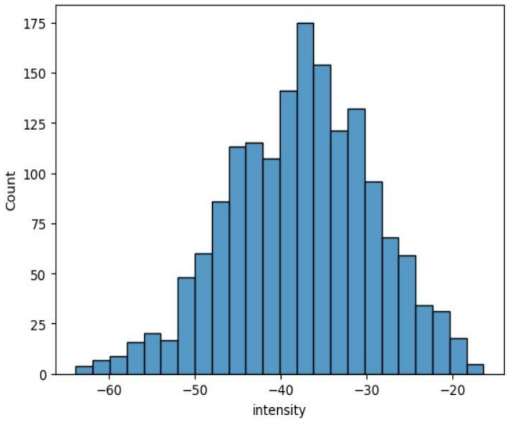
After careful analysis of the dataset, we realized that missing values are present distributed as follows:

- VOCAL CHANNEL: 196
- ACTOR: 1126
- INTENSITY: 816

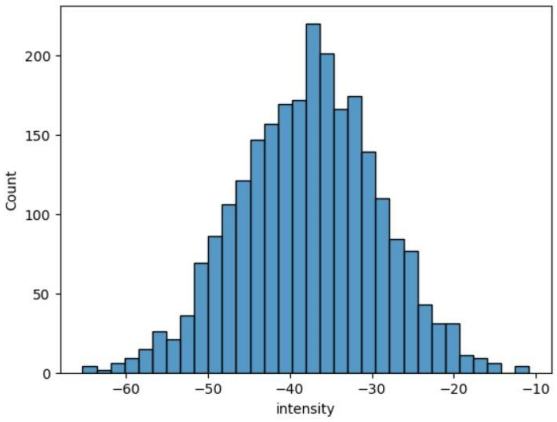
Regarding the VOCAL CHANNEL attribute, we know that it has two modes: speech which has 1335 observations and song which has 921 observations. We replaced 59% of the missing values with speech and the remaining 41% with song.

With reference to the INTENSITY attribute, we kept the shape of the normal distribution of the variable

BEFORE REPLACING MISSING VALUES



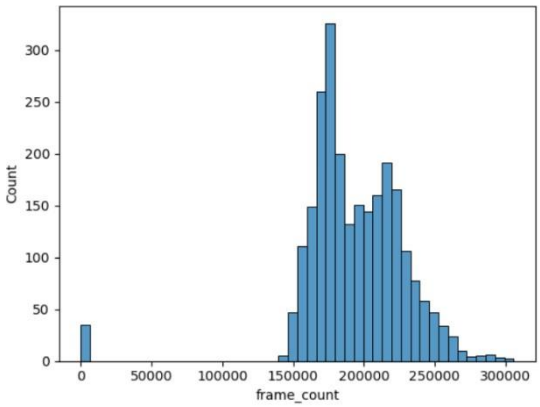
AFTER REPLACING MISSING VALUES



The ACTOR attribute, on the other hand, has very many missing values i.e., 45.92 % of the total values, and this does not allow us to replace the missing values truthfully.

BEFORE REPLACING MISSING VALUES

FIGURE 1



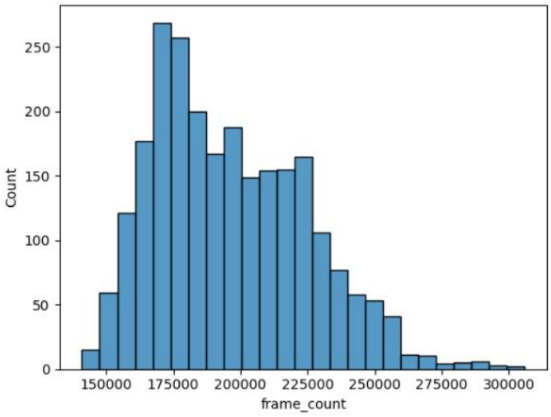
These observations could be classified as missing values, and since we have the need to remove them, we replaced them with the median. The result achieved can be observed in Figure 2

1.3.2 OUTLIERS

A final analysis should be addressed to the FRAME-COUNT attribute, which has 35 observations with the value -1 as can be seen from Figure 1.

AFTER REPLACING MISSING VALUES

FIGURE 2

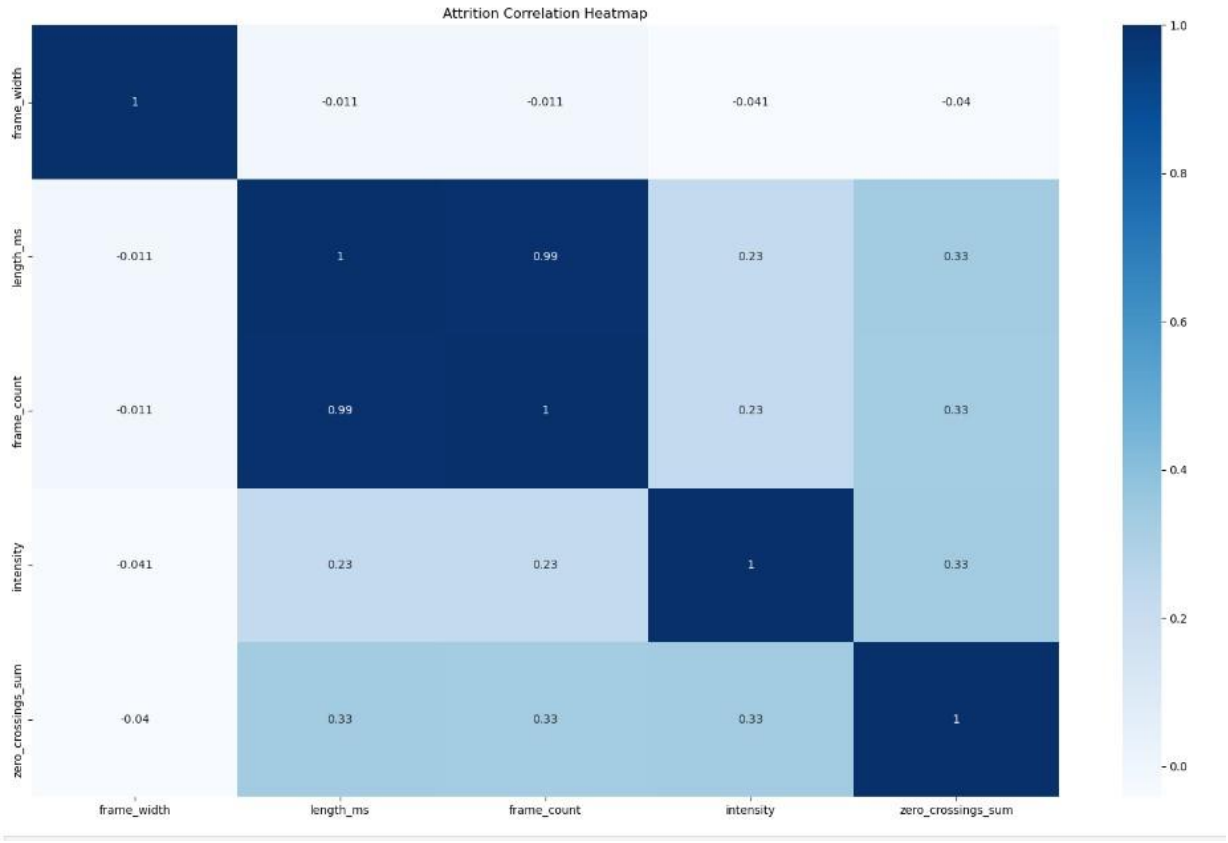


An Outlier is a value that belongs to a certain attribute but is very different from all other values. In most cases, the removal of Outliers is strongly recommended. In fact, ignoring these values could have very serious repercussions on our models and the performance of our results. Identifying Outliers is one of the fundamental pre-processing procedures. The identification of such values is called OUTLIER DETECTION.

There is no precise and rigorous mathematical definition of Outliers but there are empirical ways to be able to understand how to identify them. Usually in statistics the standard deviation is used as a criterion. A value that deviates three times the standard deviation from the mean usually is considered an Outlier. This is the method we used in our analysis: Taking only quantitative variables into account, we translate all observations that have values that are beyond the mean ± 3 times the standard deviation and replace them with the mean ± 3 times the standard deviation.

1.3.3 PAIRWISE CORRELATIONS AND ELIMINATION OF REDUNDANT VARIABLES

FIGURE 3



The heat map (Figure 3) shows the correlation matrix, where the light blue square indicates low correlation, while the dark blue square indicates high correlation. Looking closely at the matrix we notice a higher correlation (0.99) between the variables FRAME_COUNT and LENGHT_MS. Since these attributes are highly correlated, to simplify the model, the variable Frame_count was eliminated. *We choose to keep the variable Lenght_ms because it has 35 observations with the value -1.*

1.3.4 DATA REDUCTION

In this section we are going to define which attributes are unnecessary or superfluous for our data analysis. In section 1.3.1 we had noticed that the **ACTOR** attribute had many missing values and we were unable to replace them; this is a good reason to eliminate that variable from the model. Whereas, in section 1.3.4 we eliminated the variable **FRAME_COUNT** as it was highly correlated with **LENGHT_MS**

Other variables to be eliminated are:

- **MODALITY** as it presents only one mode
- **CHANNEL** was deleted because it has few observations on the second channel (only 6)
- **FRAME_RATE** has the only mode 2 observed
- **SAMPLE_WIDTH** has only one observed mode
- Mfcc (mean, std, min, max); Sc (mean, std, min, max, kur, skew); Stft (mean, std, min, max, kur, skew); mean; std; min; max; kur; skew: have been removed as basic statistics
- FRAME WIDTH presents mode 2 for 2448 times and mode 4 for 4 times

CHAPTER 2

DATA CLUSTERING

For cluster analysis, we considered 3 variables:

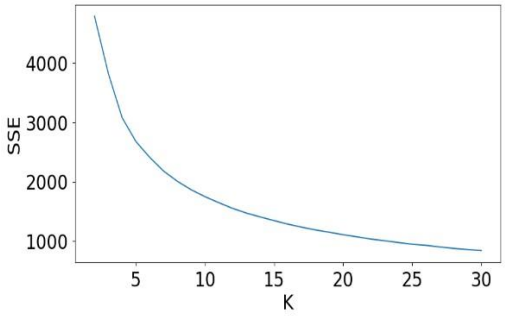
- LENGHT_MS
- INTENSITY
- ZERO_CROSSING_SUM

The first basic step was to standardize numerical variables with zero mean and unit variance. In the following section, several methods are used to perform clustering. The algorithms applied are K-means, density-based (DBSCAN) and hierarchical clustering.

2.1 K-MEANS

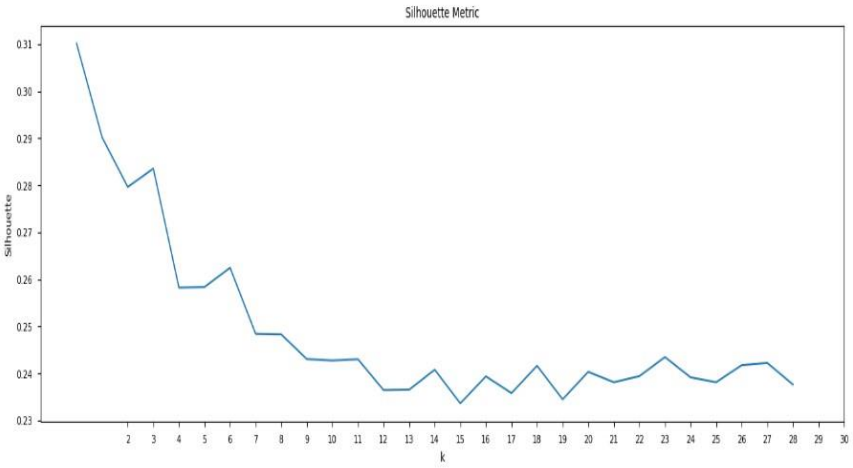
As a first attempt at clustering, we used K-Means, an unsupervised learning method. The goal of this algorithm is to find groups of objects that have common features within them, minimizing their distance (intra-cluster distance), and at the same time, maximizing the distance between different groups (inter-cluster distance). To use the K-Means algorithm, we had to choose the number of K clusters into which to divide the dataset. The study of the value of K to be used was done by taking into consideration the trend of SSE and silhouette as K changes.

FIGURE 4



For the choice of parameter K, we chose a number of clusters ranging from 2 to 30 (Figure 4) Looking at the graph we can see that as the number of k increases, the mean square error (SSE which is the sum of the squared differences between each cluster point and its centroid) falls.

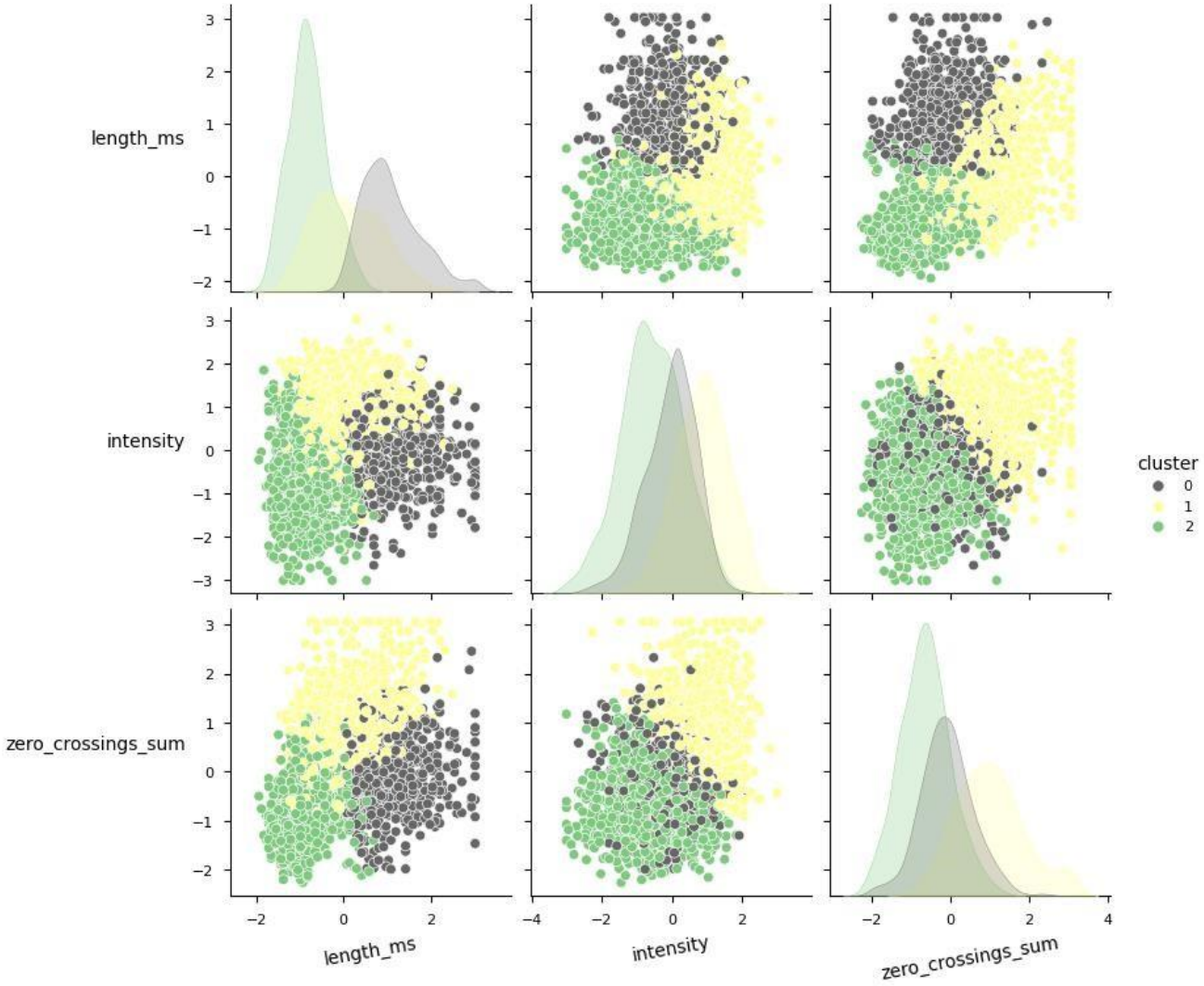
FIGURE 5



As shown in Figure 5, we noticed that the highest value of the silhouette is 3. Comparing the two graphs, we considered the K for which there is a greater decrease in the SSE value and an increase in the silhouette. Therefore, we chose K = 3 as the parameter because it is near the elbow point.

2.1.1 CLUSTER ANALYSIS OBTAINED WITH K-MEANS

To evaluate the goodness of the choice of the number of clusters, we evaluated the "Silhouette Score," which is a metric that assesses intra-cluster and inter-cluster distance in a range from -1 (worst score) to 1 (best score). For k=3 we obtain a Silhouette Score of 0.29



This graph depicts the four quantitative variables we considered grouped into four distinct clusters. The K-MEANS algorithm allowed us to obtain three clusters with the following dimensions:

Cluster 0: 696 Cluster

1: 721

Cluster 2: 1035

2.2 DENSITY-BASED CLUSTERING

For the density-based clustering part, we decided to use DBSCAN. DBSCAN is an unsupervised learning technique used to identify clusters of different sizes and shapes.

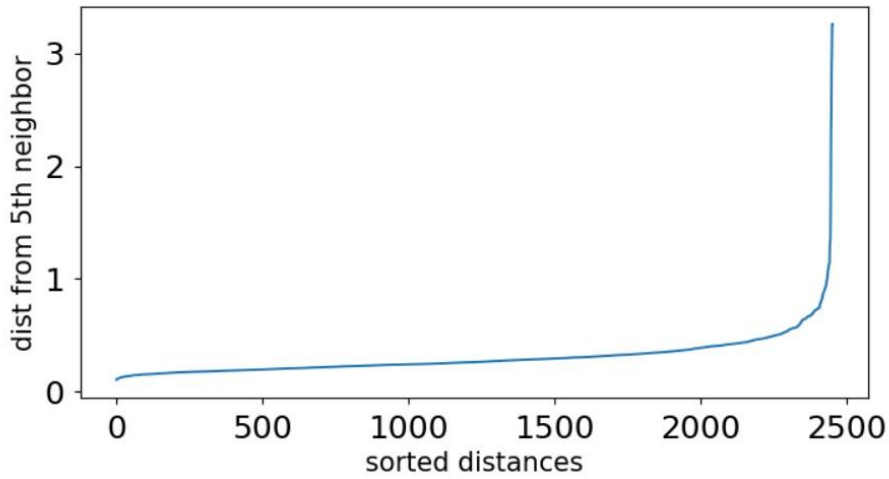
It works with two parameters:

- The radius ε (eps), which specifies the distance within which two points are considered to be neighbors. - MinPoints, which is the minimum number of points needed to form a cluster.

To obtain an interesting result from running DBSCAN it is necessary to estimate the value of the MinPoints and Epsilon parameters. The choice of Epsilon is not trivial, as selecting too small a value would have too many points shown as outliers while too high a

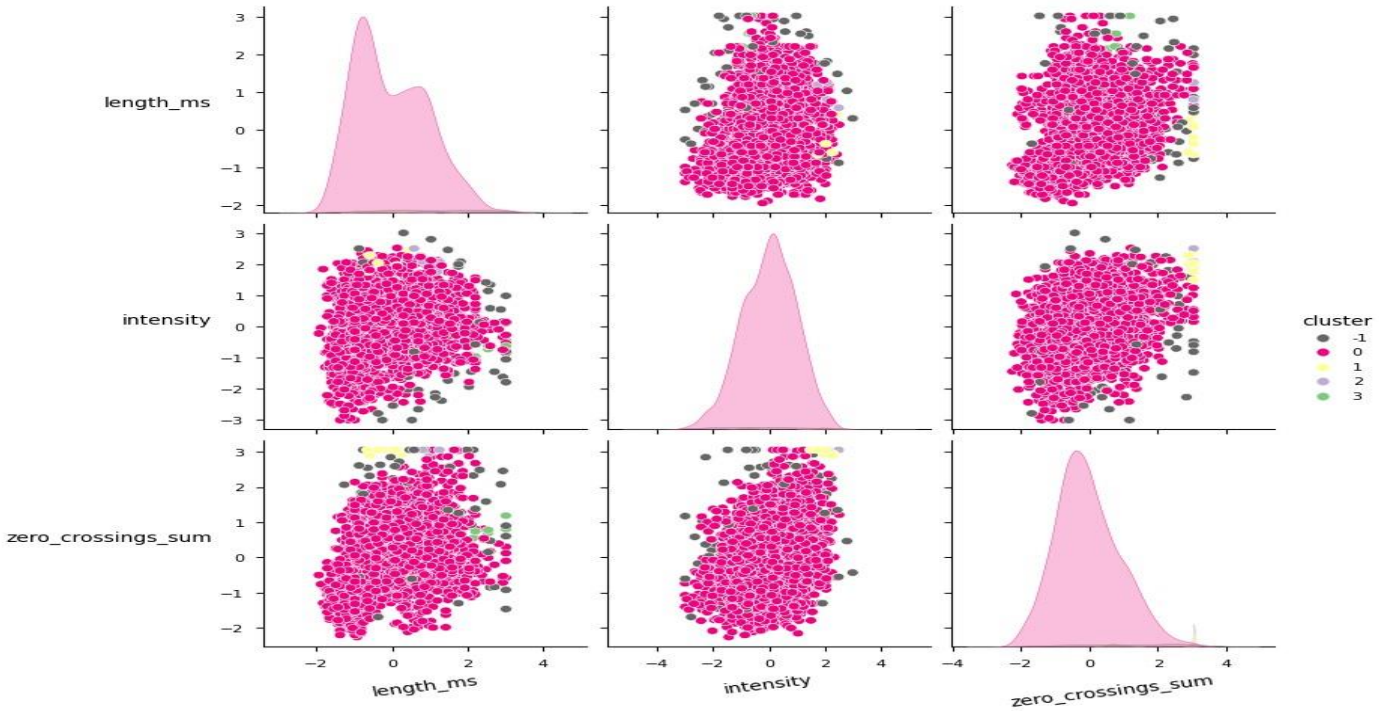
value would result in a single cluster. To estimate the value of Epsilon we used the K-Nearest neighbor method. In the graph in Figure 8 we show the curve obtained by the K-Nearest neighbor method.

FIGURA 8: Average of the distances between each point and the nearest 5 nodes



To estimate the value of epsilon, we chose a point lower than the elbow point so that we would have more meaningful clusters. The value we chose for epsilon is 0.5 and for MinPoints is 5. The result obtained from DBSCAN is the formation of 58 noise points with 4 cluster sizes: **Cluster 0**: 2369; **Cluster 1**: 10; **Cluster 2**: 8; **Cluster 3:** 7

The silhouette score of this DBSCAN model is 0.216



2.3 HIERARCHICAL CLUSTERING

We performed Hierarchical Clustering using three different methods, choosing those that are least susceptible to noise and outliers namely COMPLETE, AVERAGE and WARD.

2.3.1 ANALYSIS OF THE DENDOGRAMS OBTAINED

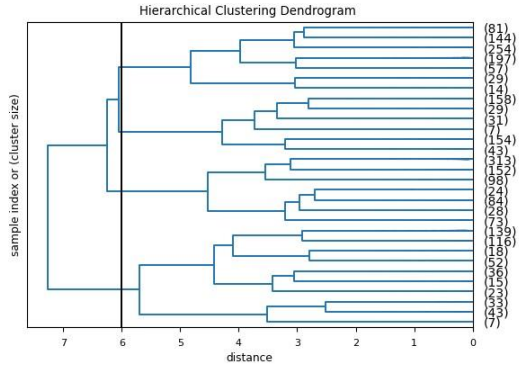


FIGURA 10: DENDROGRAM BY COMPLETE METHOD

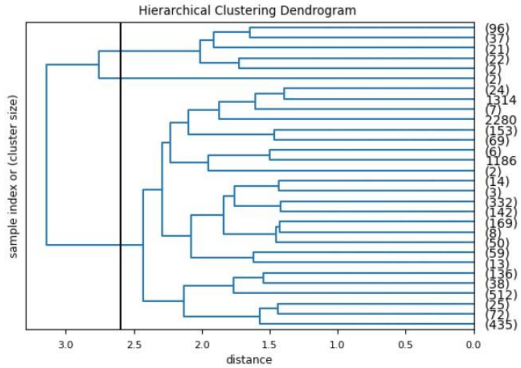


FIGURA 11: DENDROGRAM BY AVERAGE METHOD

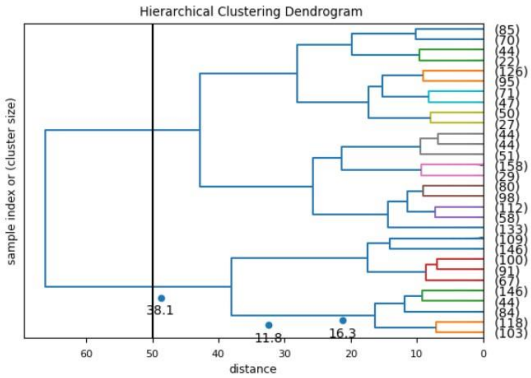


FIGURA 12: DENDROGRAM BY WARD METHOD

For choosing the best number of clusters, we calculated the silhouette score for each of these methods. The results are:

COMPLETE METHOD:

```
Number Clusters:  4
Cluster {0: 482, 1: 776, 2: 772, 3: 422}
Silhouette Score 0.2181664914772812
```

AVERAGE METHOD:

```
Number Clusters:  3
Cluster {0: 325, 1: 225, 2: 1296}
Silhouette Score 0.12748128973774925
```

WARD METHOD:

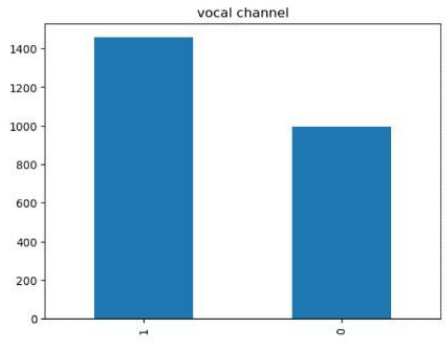
```
Number Clusters:  2
Cluster {0: 1444, 1: 1008}
Silhouette Score 0.26567183730864585
```

The best method obtained is WARD method because it has the highest silhouette score.

CHAPTER 3

DATA CLASSIFICATION

Regarding the classification procedure, we considered as dependent variable the attribute vocal_channel which has 2 modes (speech and song), while as independent variables we chose 4 categorical attributes: emotional_intensity, statement, repetition, sex. The next step was to dichotomize the variables considered and then divide our dataset into training set (70%) and test set (30%).



The graph on the left shows the frequency distribution of the vocal_channel attribute with the two modes speech and song identified as follows:

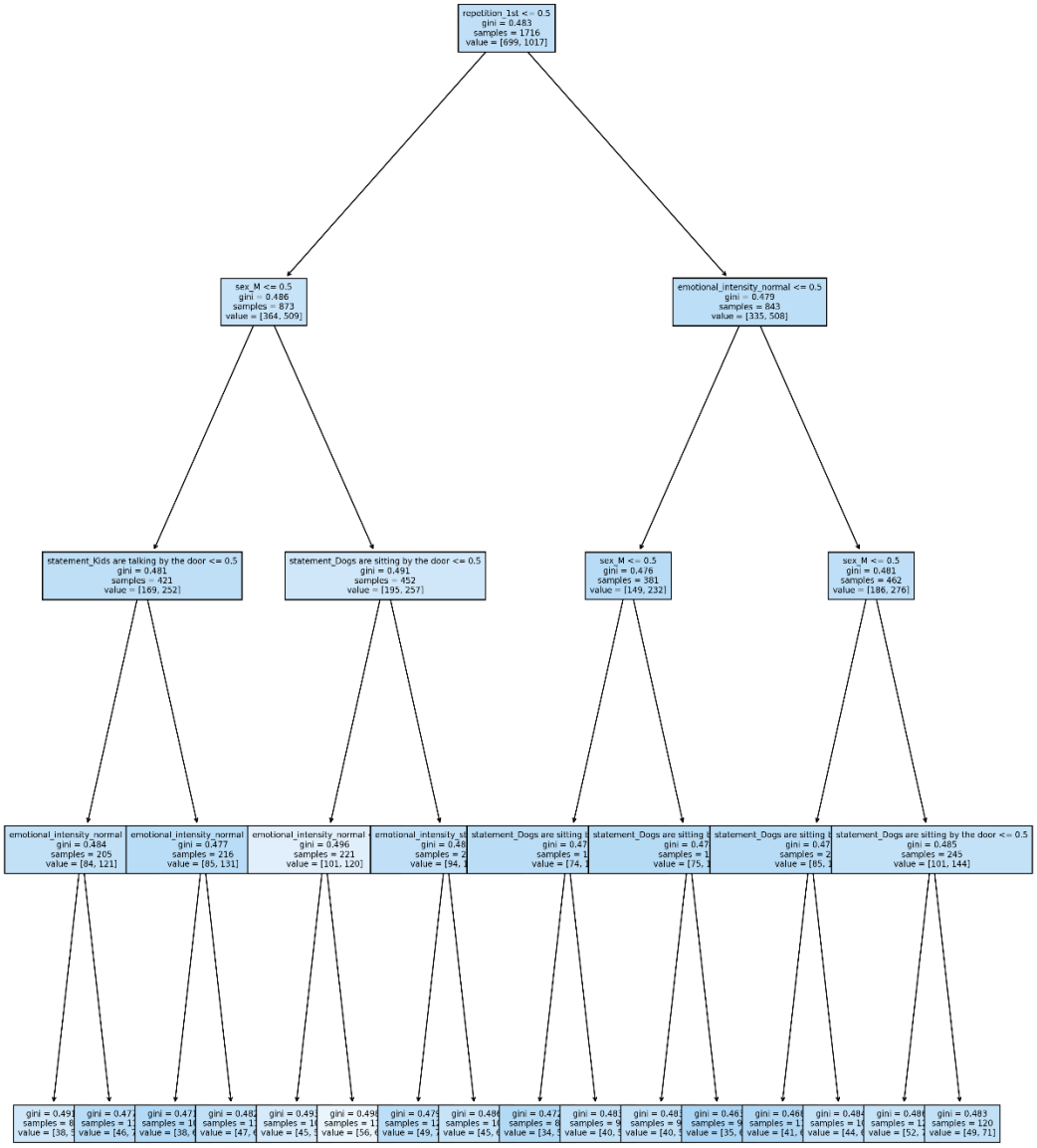
- Speech = 1
- Song = 0

The following paragraphs will explain the three classifiers we used in our analysis, and then conclude the chapter by choosing the classifier that best evaluates our model.

3.1 DECISION TREE

The decision tree model was carried out with the DecisionTreeClassifier function in the sklearn package, we set the number of children =4, obviously putting the transformed vocal channel attribute with speech=1, song=0 as the response variable, and as explanatory variables we put emotion, emotional intensity,statement, repetition and sex. The results are shown here with the following graph:

DECISION TREE



Here instead we have the report with the model performance measures, as we see:

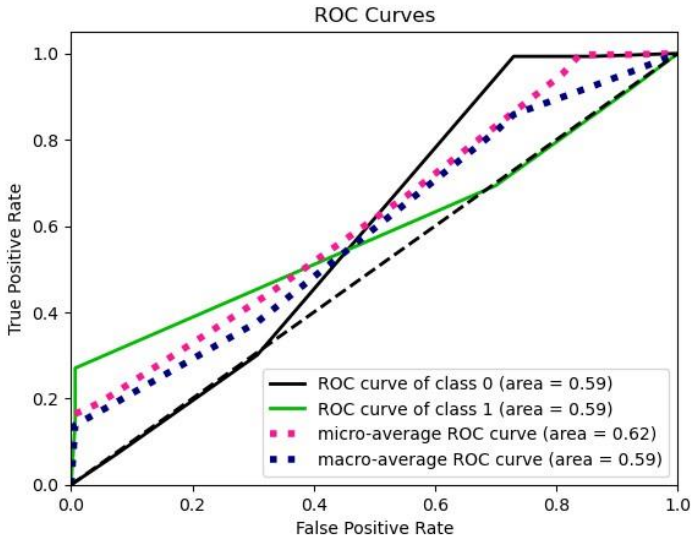
```
precision    recall  f1-score   support

           0       0.40      0.30      0.34       296
1       0.60      0.69      0.64       440

    accuracy                           0.54       736
macro avg       0.50      0.50      0.49       736
weighted avg       0.52      0.54      0.52       736
```

The most important results to display are the accuracy which is equal to 0.54 and the F1-score which is 0.50.

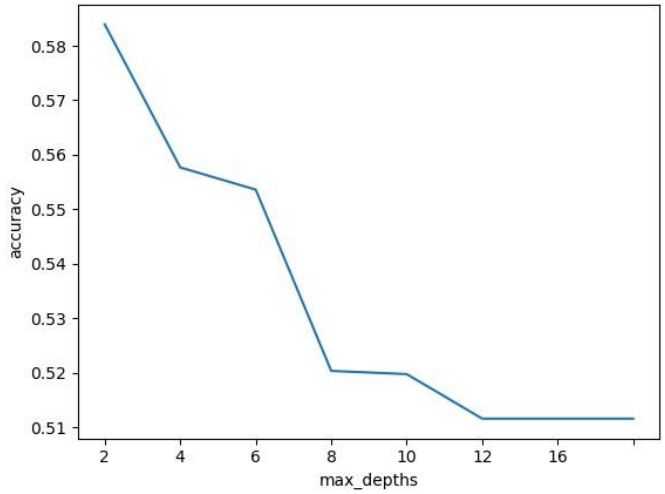
In the table below, the confusion matrix with the actual values in the x-axis and the predicted values in the y-axis is refined.

	True values-song	*True values-speech*
Predicted values - song	89	207
Predicted values -speech	135	305

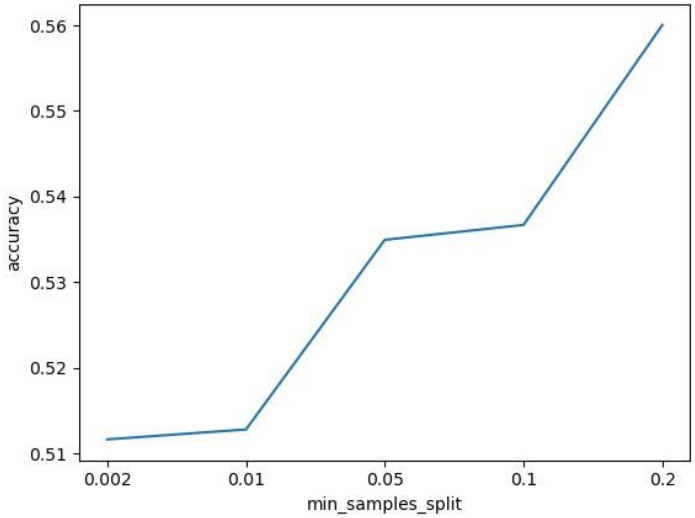
Next, the roc curve of the model, which is a graphical scheme for binary-type Z classification problems, is refined. They are related with the ratio of true positives (TPR) = sensitivity in the y-axis and the ratio of false positives (FPR)=1-specificity in the x-axis.



Next, we tested the decision tree model by changing the number of max_depth (max depth) by trying the model with 2,4,6,8,10,12. We saved the accuracy results and plotted the results and saw that as max depth increases, accuracy decreases never going beyond a certain threshold.

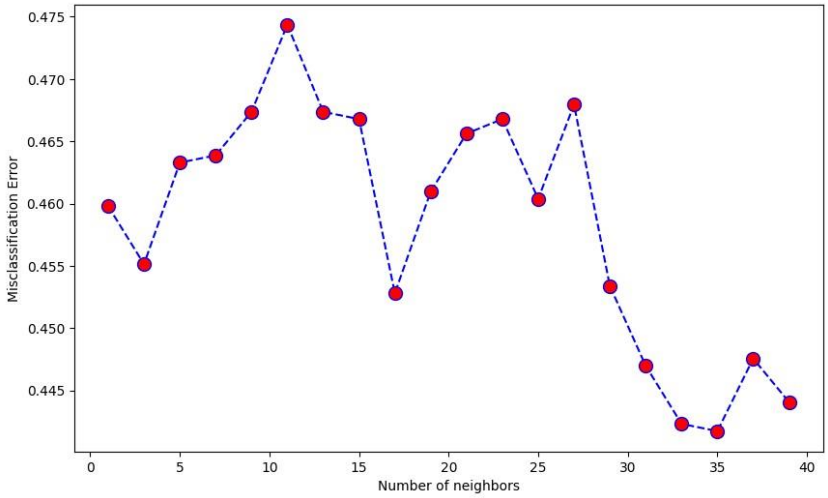


We also tested the decision tree model by gradually changing the number of min_samples_split (minimum number of split samples) by putting them with the values of [0.002, 0.01, 0.05, 0.1, 0.2] and noticed that as the value of min_samples_split increases the accuracy.



3.2 KNN

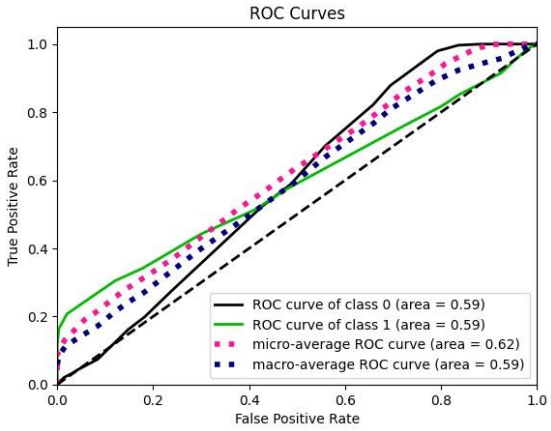
The knn is an algorithm to classify test instances according to its majority class of nearest neighbors. The best number of neighbors was found with an algorithm where for each k (neighbors) the misclassified error is calculated and, the k with the lowest erate classification value is chosen. In this case the best k is 35 as we see from the following graph:



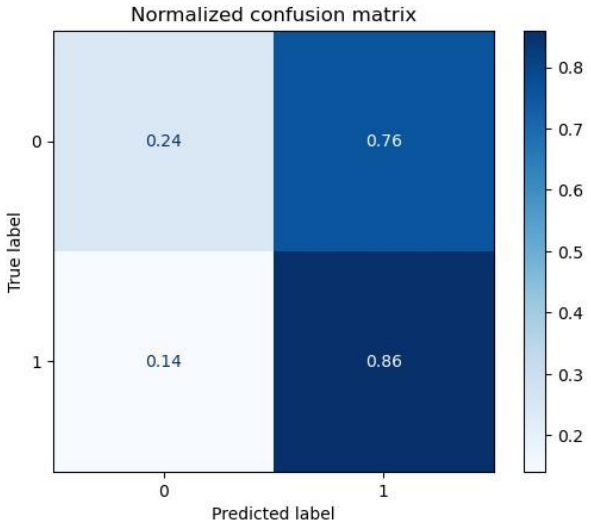
We then fit the knn model by putting in our variables, using uniform weights for all attributes, and derive performance values, which will be as follows:

```
              precision    recall  f1-score   support

           0       0.40      0.22      0.29       297
1       0.60      0.78      0.68       439
     accuracy                           0.55
736    macro avg       0.50      0.50      0.48
736 weighted avg       0.52      0.55      0.52
736
```

Here the roc curve of the KNN model:



Shown next is the confusion matrix with the predicted and real values of speech (=1) and song (=0) the confusion matrix with relative frequency conditional on the real values



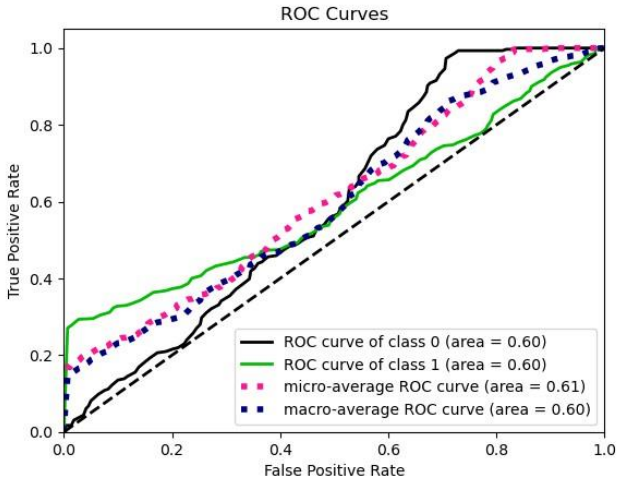
3.3 NAÏVE BAYES

The third proposed model is a naïve bayes classifier model that is a probabilistic classifier model based on the application of Bayes' theorem with assumptions of independence.

The response variables and explanatory variables are the same, and by going to implement the naive Gaussian Bayes model we obt ain the following results:

```
              precision    recall  f1-score   support

           0       0.48      0.99      0.65       296
1       0.98      0.27      0.42       440
     accuracy                           0.56
736    macro avg       0.73      0.63      0.53
736 weighted avg       0.78      0.56      0.51
736
```

As we see, the accuracy is 0.56 and the f1-score for song mode = 0.65 and for speech mode=42 with an average macro of 0.53 Regarding the roc curve:



3.4 FINAL CONSIDERATIONS

Among the three models analyzed, the best one is the naïve bayes because it has higher values for accuracy and f1 score, which means that the values predicted by the model (compared to the other two models) are closer to the actual values.

The worst model among the three is the decision tree because in addition to having lower accuracy and f1 score it also has a nonlinear shape in the roc curve.

CHAPTER 4
PATTERN MINING

In this section we describe the process of association rules analysis. In the first section we discuss the preliminary operations performed on the dataset to prepare the data for subsequent operations. In subsequent sections we extract frequent itemsets and then, from these itemsets, association rules.

To carry out the itemset extractions, we considered only a few variables, such as: vocal channel, emotion, emotional intensity, statement, repetition and sex.

We used as our algorithm for pattern extraction the **A PRIORI ALGORITHM**

4.1 FREQUENT ITEMSET

For Frequent ItemSets we used a MinSupp equal to 20% as a parameter by going to select only sets with 2 items. In the table we give the list of the most frequent ItemSets. The list of ItemSets largely coincides with the list of "Closed" ItemSets, this is because the Frequent ItemSets are an over set of the "closed" ones listed above.

	frequent_itemset	support
37	(normal, speech)	31.606852
31	(Kids are talking by the door, speech)	29.526917
18	(F, speech)	29.486134
27	(2nd, speech)	29.445351
34	(1st, speech)	29.241436
36	(M, speech)	29.200653
23	(Dogs are sitting by the door, speech)	29.159869
35	(M, normal)	27.406199
12	(strong, speech)	27.079935
26	(2nd, normal)	26.916803
22	(Dogs are sitting by the door, normal)	26.916803
30	(Kids are talking by the door, normal)	26.916803
33	(1st, normal)	26.916803
17	(F, normal)	26.427406
21	(Dogs are sitting by the door, M)	25.448613
29	(Kids are talking by the door, M)	25.448613
25	(2nd, M)	25.448613
32	(1st, M)	25.448613
24	(2nd, Kids are talking by the door)	25.000000
28	(Kids are talking by the door, 1st)	25.000000
19	(Dogs are sitting by the door, 2nd)	25.000000
20	(Dogs are sitting by the door, 1st)	25.000000
16	(F, 1st)	24.551387
15	(F, Kids are talking by the door)	24.551387
14	(F, 2nd)	24.551387
13	(F, Dogs are sitting by the door)	24.551387
11	(strong, M)	23.491028
8	(strong, 2nd)	23.083197
7	(strong, Dogs are sitting by the door)	23.083197
9	(strong, Kids are talking by the door)	23.083197
10	(strong, 1st)	23.083197
6	(strong, F)	22.675367
5	(song, normal)	22.226754
4	(song, M)	21.696574
0	(song, Dogs are sitting by the door)	20.840131
3	(song, 1st)	20.758564
1	(song, 2nd)	20.554649
2	(song, Kids are talking by the door)	20.473083

4.2 CLOSED ITEMSET

In the case of closed ItemSets we used as a parameter a MinSupp equal to 20% by going to select only those sets with at least 2 items. In the table we give the list of the closed ItemSets most frequent with the percentage of support next to them.

	closed_itemset	support
37	(normal, speech)	31.606852
31	(Kids are talking by the door, speech)	29.526917
18	(F, speech)	29.486134
27	(2nd, speech)	29.445351
34	(1st, speech)	29.241436
36	(M, speech)	29.200653
23	(Dogs are sitting by the door, speech)	29.159869
35	(M, normal)	27.406199
12	(strong, speech)	27.079935
26	(2nd, normal)	26.916803
22	(Dogs are sitting by the door, normal)	26.916803
30	(Kids are talking by the door, normal)	26.916803
33	(1st, normal)	26.916803
17	(F, normal)	26.427406
21	(Dogs are sitting by the door, M)	25.448613
29	(Kids are talking by the door, M)	25.448613
25	(2nd, M)	25.448613
32	(1st, M)	25.448613
24	(2nd, Kids are talking by the door)	25.000000
28	(Kids are talking by the door, 1st)	25.000000
19	(Dogs are sitting by the door, 2nd)	25.000000
20	(Dogs are sitting by the door, 1st)	25.000000
16	(F, 1st)	24.551387
15	(F, Kids are talking by the door)	24.551387
14	(F, 2nd)	24.551387
13	(F, Dogs are sitting by the door)	24.551387
11	(strong, M)	23.491028
8	(strong, 2nd)	23.083197
7	(strong, Dogs are sitting by the door)	23.083197
9	(strong, Kids are talking by the door)	23.083197
10	(strong, 1st)	23.083197

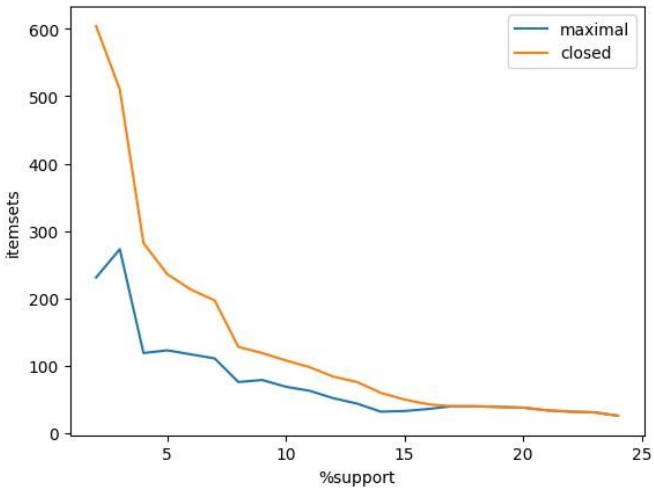
6	(strong, F)	22.675367
5	(song, normal)	22.226754
4	(song, M)	21.696574
0	(song, Dogs are sitting by the door)	20.840131
3	(song, 1st)	20.758564
1	(song, 2nd)	20.554649
2	(song, Kids are talking by the door)	20.473083

4.3 MAXIMAL ITEMSET

The extraction of maximal itemsets was performed using the APriori algorithm using the parameter MinSup = 20% and selecting only sets containing at least 4 itemsets. In Table 3.1 we report frequent itemsets with support between 20% and 31.7%.

	maximal_itemset	support
37	(normal, speech)	31.606852
31	(Kids are talking by the door, speech)	29.526917
18	(F, speech)	29.486134
27	(2nd, speech)	29.445351
34	(1st, speech)	29.241436
36	(M, speech)	29.200653
23	(Dogs are sitting by the door, speech)	29.159869
35	(M, normal)	27.406199
12	(strong, speech)	27.079935
26	(2nd, normal)	26.916806
22	(Dogs are sitting by the door, normal)	26.916803
30	(Kids are talking by the door, normal)	26.916803
33	(1st, normal)	26.916803
17	(F, normal)	26.427406
21	(Dogs are sitting by the door, M)	25.448613
29	(Kids are talking by the door, M)	25.448613
25	(2nd, M)	25.448613
32	(1st, M)	25.448613
24	(2nd, Kids are talking by the door)	25.000000
28	(Kids are talking by the door, 1st)	25.000000
19	(Dogs are sitting by the door, 2nd)	25.000000
20	(Dogs are sitting by the door, 1st)	25.000000
16	(F, 1st)	24.551387
15	(F, Kids are talking by the door)	24.551387
14	(F, 2nd)	24.551387
13	(F, Dogs are sitting by the door)	24.551387
11	(strong, M)	23.491028
8	(strong, 2nd)	23.083197
7	(strong, Dogs are sitting by the door)	23.083197
9	(strong, Kids are talking by the door)	23.083197
10	(strong, 1st)	23.083197
6	(strong, F)	22.675367
5	(song, normal)	22.226754
4	(song, M)	21.696574
0	(song, Dogs are sitting by the door)	20.840131
3	(song, 1st)	20.758564
1	(song, 2nd)	20.554649
2	(song, Kids are talking by the door)	20.473083

The graph below compares the "maximum" feature set with the "closed" feature set considering the percentage of support with the number of features.



It can be seen that with equal support the number of items in the closed itemsets is always greater than the maximum number of itemsets.

We leave the number of items to be selected equal to 2, by increasing the number of items (e.g. equal to 3) there was no itemset with support >=20%, but the max support percentage reached 13%.

Regarding the support percentage we chose the threshold of 20%because raising it further (example 30%) resulted in very few records (at most 2).

4.4 ASSOCIATION RULES

In the table below we have given the list of association rules, deciding to extract these rules by considering only frequent itemsets that had a length greater than or equal to 2.

As a confidence value we set the threshold of 55% as a parameter. In these rules we put the values "speech" and "song" arranged in descending order according to the weight given as postcondition values.

For example, the first record in these association rules stands for the person who is male and the voice intensity is normal then he will simply be speaking (and not singing) [the mode of the vocal channel variable will be speech].

The second rule says that someone who says "kids are talking by the door" and is male will be talking(speech), and so on.

1 to 22 of 22 entries Filter

index	consequent	antecedent	abs_support	%_support	confidence	lift
19	speech	M,normal	387	15.783034257748776	0.5758928571428571	0.963225979341259
16	speech	Kids are talking by the door,M	362	14.763458401305057	0.5801282051282052	0.9703099310875573
9	speech	2nd,M	364	14.845024469820556	0.5833333333333334	0.9756707594361074
20	speech	M	729	29.73083197389886	0.5841346153846154	0.9770109665232448
13	speech	1st,M	365	14.885807504078302	0.5849358974358975	0.9783511736103824
5	speech	Dogs are sitting by the door,M	367	14.9673735725938	0.5881410256410257	0.9837120019589324
6	speech	Dogs are sitting by the door,normal	390	15.905383360522023	0.5909090909090909	0.9883418082599529
3	speech	Dogs are sitting by the door,2nd	363	14.804241435562806	0.5921696574225123	0.990450204638472
14	speech	1st,normal	391	15.94616639477977	0.5924242424242424	0.990876018024722
12	speech	1st,Kids are talking by the door	364	14.845024469820556	0.5938009787928222	0.9931787175989086
21	speech	normal	784	31.97389885807504	0.5939393939393394	0.9934102277894911
10	speech	2nd,normal	393	16.02773246329527	0.5954545454545455	0.9959444375542602
11	speech	2nd	731	29.812398042414358	0.5962479608482871	0.9972714870395635
17	speech	Kids are talking by the door,normal	394	16.06851549755302	0.5969696969696969	0.9984786473190294
18	speech	Kids are talking by the door	732	29.853181076672104	0.5970636215334421	0.9986357435197817
7	speech	Dogs are sitting by the door	734	29.9347471451876	0.598694942903752	1.0013642564802183
15	speech	1st	735	29.975530179445354	0.5995106035889071	1.0027285129604366
8	speech	2nd,Kids are talking by the door	368	15.00815660685155	0.600326264274062	1.004092769440655
0	speech	strong	682	27.814029363784666	0.6024734982332155	1.0076841866765651
4	speech	Dogs are sitting by the door,1st	371	15.130505709624797	0.6052202283849919	1.0122783083219646
2	speech	F	737	30.057096247960846	0.6121262458471761	1.0238291642682644
1	speech	F,normal	397	16.190864600326265	0.6126543209876543	1.0247124113654356