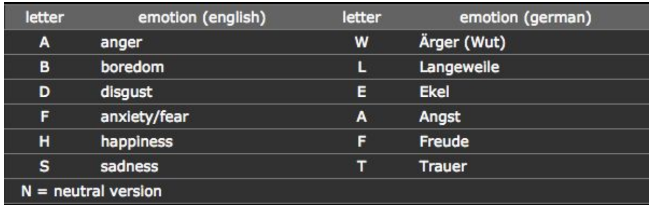
Emotional Speech recognition

Ali Fessi

Introduction.

Using a dataSet containing 535 audio files. The task will be to predict which emotion has been expressed in the audio file. The audio files are in german and they were expressed by men and women from different ages.

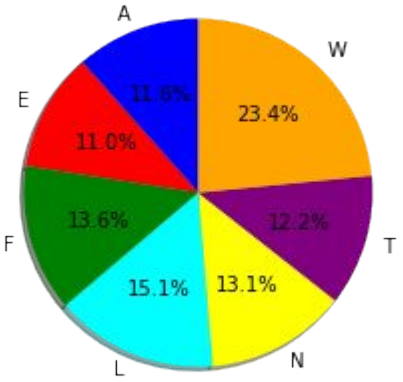
The labels are the following.

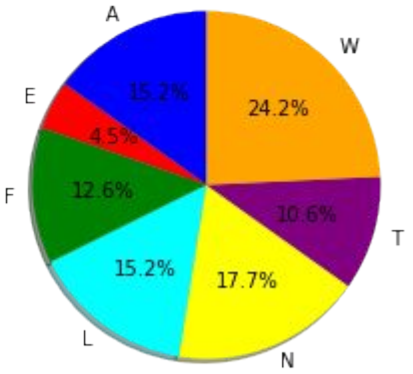


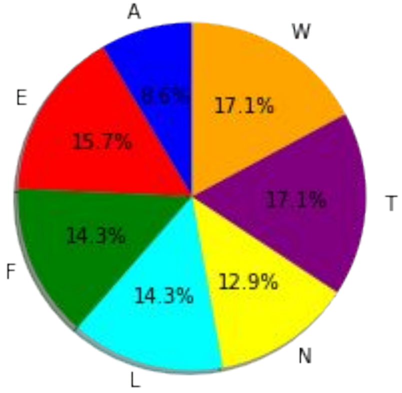
letter	emotion (english)	letter	emotion (german)
A	anger	W	Ärger (Wut)
B	boredom	L	Langeweile
D	disgust	E	Ekel
F	anxiety/fear	A	Angst
H	happiness	F	Freude
S	sadness	T	Trauer
N = neutral version			

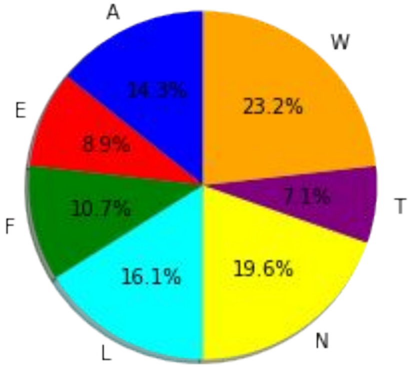
Despite the small size of the data set we will try to implement several models that will give us great insight about the sentiment we want to predict.

Some Insights about the data.


Man sentiment partition


Women sentiment partition


35 years old sentiment partition


25 years old sentiment partition

-Very unbiased data from all sides. It would have been interesting to take these ‘meta data’ informations(Gender, Age) as input of our future mode but would probably lead us to make assumptions that may be true in our case but completely wrong if we extended our dataset.

Libraries used

-Some Classical ones as pandas, numpy, matplotlib, sklearn,keras.

- Specific libraries have been used for audio processing. The library pyAudioAnalysis was really helpful as it did enable us to construct our feature vector by taking into account important metrics about the audio file(Energy, MFCC’s, Spectral flux)

Data processing pipeline

Our data is presented in a folder where each file contains a WAV file. The name of the file contains all the informations related to each sample(Which person said it, Sentiment expressed,etc...)

After several approaches we decided to represent our data into a dataframe wich 34 input features, each feature is the standard deviation of the array of one of the 34 metrics presented in the slides before.

We took the standard deviation as it gives us the better correlation between it and the sentiment expressed.

Feature Engineering.

-We Standardized the input features(Subtracted the mean of each feature and divided by its variance). It significantly improved our model results.

-We also tried to do a principal component analysis and extract the best features. However it did not lead us to a great result improvement.

Modeling and results

As it is a classification task with 7 possible classes. We implemented several different models and compared their results.

We made the assumption that in this classification task, “false positive” and “false negatives” should be treated equally, therefore the models should be evaluated by their predictive accuracy.

Modeling and results

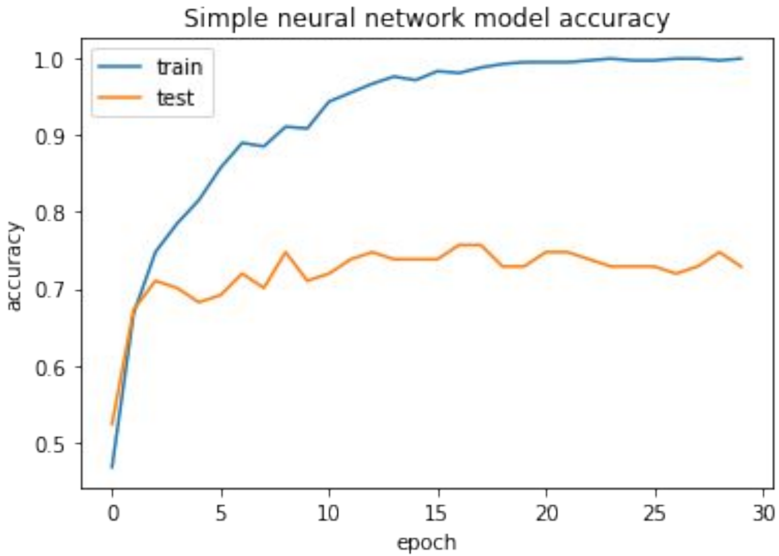
.Logistic regression: Accuracy 70.09% :)

.K-nearest Neighbors: We used K=12 Neighbors. Accuracy 60.75% :(

.SVM: Degree=1 : Accuracy 70.1% :)

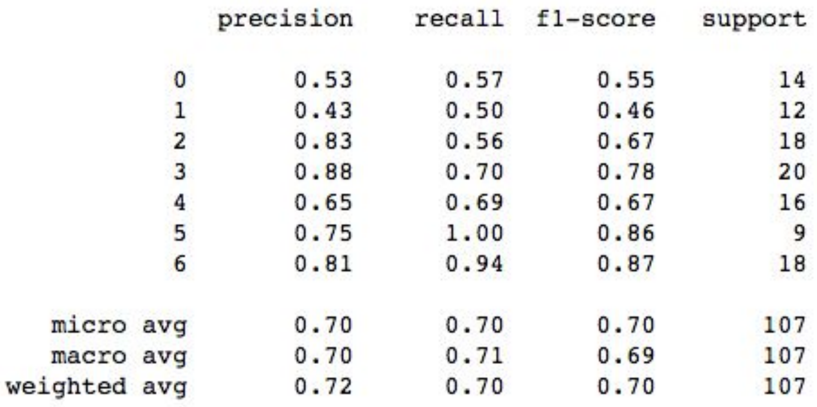
.Simple Neural Network: No hidden layer : Test accuracy of 76.64% :D

Results and comments



Results and comments

SVM Confusion Matrix.



```
              precision    recall  f1-score   support

           0       0.53      0.57      0.55        14
           1       0.43      0.50      0.46        12
           2       0.83      0.56      0.67        18
           3       0.88      0.70      0.78        20
           4       0.65      0.69      0.67        16
           5       0.75      1.00      0.86         9
           6       0.81      0.94      0.87        18

   micro avg       0.70      0.70      0.70       107
   macro avg       0.70      0.71      0.69       107
weighted avg       0.72      0.70      0.70       107
```

We can see the very bad effects of unbalanced data. Some sentiments and much easily predicted because of their predominance in the dataset..

Because of the small size of the dataset we did not want to balance it. As we would find ourselves with a very small dataset.

Conclusion

. In conclusion , we were able with the neural network model to predict the sentiment expressed by almost 8/10 people audios samples.

. More data would help us improve the predictive power of the model.

. More balanced data would help us make the model more robust to changes.

. We could try to implement unsupervised models as K-means with K=7 and see if we could plot interesting results...