Extimating the age of people from vocal recordings

Bello Renato
Politecnico di Torino
Student id: s341965
s341965@studenti.polito.it

Chiodo Martina
Politecnico di Torino
Student id: s343310
s343310@studenti.polito.it

***Abstract*—The abstract goes here. Keep it short (approx. 3-4 sentences)**

I. Problem overview

The proposed competition is a regression problem about age estimating, in fact based on vocal recordings and some information regarding the person, such as his ethnicity and his gender, we aim to correctly determine the age of the person who is talking. The dataset is divided in two parts:

- a *development* set, containing 2933 elements, each of them labeled
- an *evaluation* set, containing 691 elements

The development set will be used to build a regressor to label the elements of the evaluation set.

We can make some considerations based on the development set. First, the dataset is complete, indeed non of the rows contains missing values. Second, some features requires some preprocessing: *ethnicity* and *gender* are a categorical features and, thus, they need to be encoded in order to be used in the regression; *tempo* is not automatically saved as a float type because the values are enclosed by parenthesis. Third, the dataset also contains the path to the vocal recordings from which we have decided to extract more features from the spectrogram.

We can study the correlation of the feature given by the dataset by analyzing the correlation plot shown in figure (1). From the correlation plot we can notice that some features, such as *num_words* and *num_characters*, are highly correlated and can be redundant in the dataset. This is why we decided to get rid off some features which shared correlation 1 with others.

To better understand the distribution of the features we can plot some histograms. From the histograms shown in figure (2) we can notice that most of the features seems to be distributed as Gaussian distribution with not many outliers. Some exceptions are *max_pitch*, *min_pitch* and *num_characters*: the first two features mentioned have a very wide range of values but almost all their distribution mass is concentrated in a single point, thus there are many outliers; on the contrary, *num_characters* concentrates his mass distribution in two far values.

Another thing worth mentioning is that the values of our response variable, *age*, are not uniformly distributed. As shown in the histogram, most of the recordings are of people in the age range $[15, 35]$. This would probably affect the performance

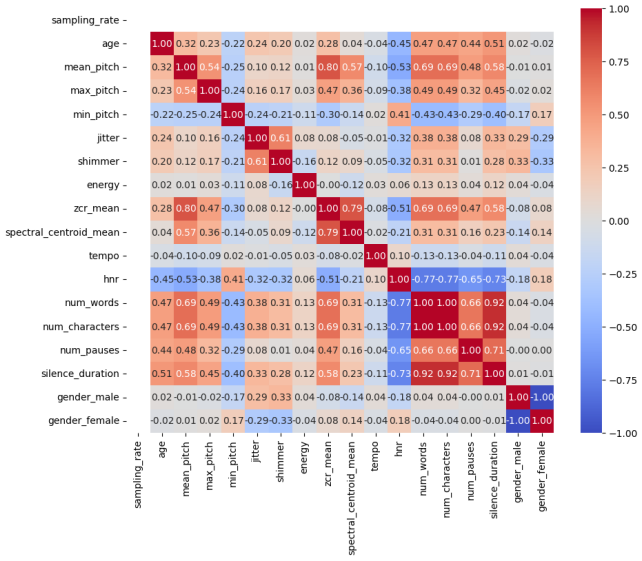

Fig. 1. Correlation among features.

of our regressor because the model might become biased towards predicting ages within this range, potentially leading to less accurate predictions for ages outside this range.

II. Proposed approach

A. Preprocessing

B. Model selection

C. Hyperparameter tuning

You can use citations as follows: [1] (you can add BibTeX citations in the *bibliography.bib* file).

III. Results

IV. Discussion

References

[1] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*, vol. 1. MIT press Cambridge, 2016.

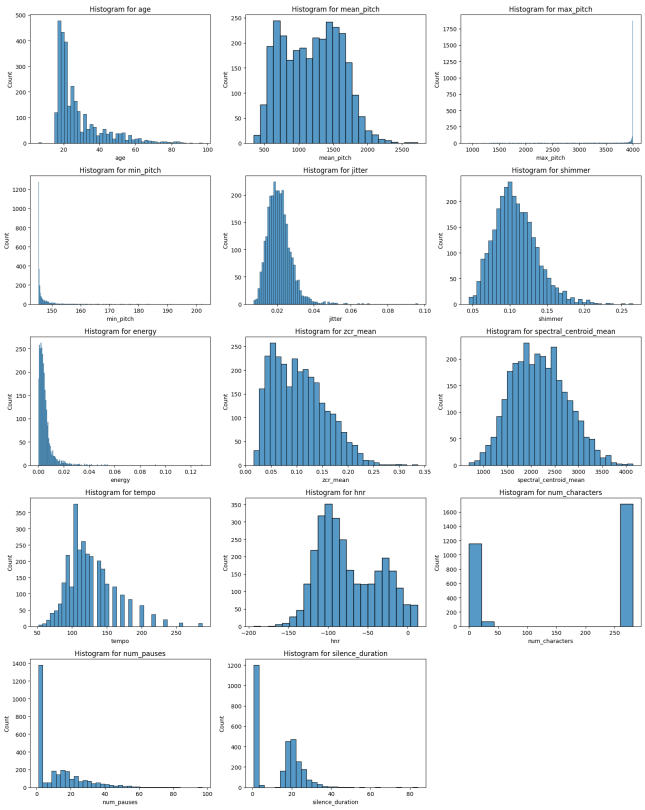

Fig. 2. Histograms of some features.