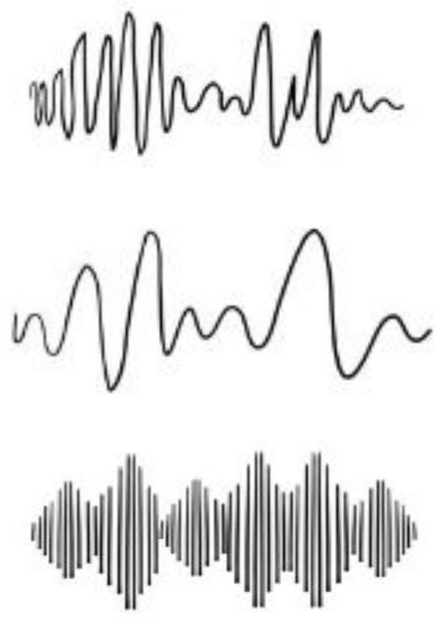
Predicting Gender by Voice and Speech Analysis

Team E6: Kaidi Tootmaa & Hanna Vallner

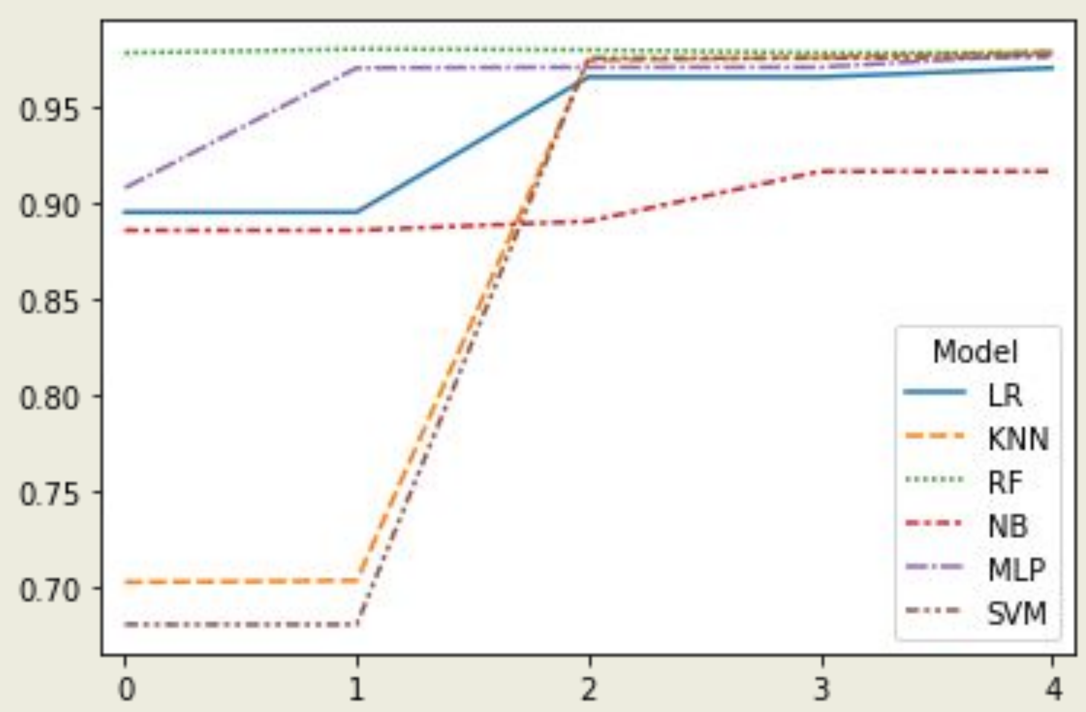
INTRODUCTION



Clearly genderwise defined voices sound more natural to the human ear. The main goal of our project was to deliver a model that predicts the gender of a person based on voice and speech analysis data.
We also wanted to create a thorough analysis on which features have the most impact in defining a person's voice as either feminine or masculine.

Our data is sourced from Kaggle, which contains collective findings from four different databases (e.g The Harvard-Haskins Database of Regularly-Timed Speech). The dataset has 3168 cases, with 21 fields for each of them.

MODELS & DATA PREPARATION



MODEL	ACCURACY
Logistic Regression	97.08%
KNN	97.95%
Random Forest	97.87%
Naive Bayes	91.67%
Multilayer Perceptron	97.83%
Support Vector Machine	97.67%

We trained a machine learning model, which could be used to predict the gender of a person based on voice and speech analysis. Interested parties could use it to verify the gender to match their desired expectations.

For preparation we **filtered** the given features and removed the ones which were highly correlated and thus had no further impact on our models (we used the threshold of 0.95 and removed three features: 'centroid', 'kurt', 'dfrange'). We also considered removing features that consisted of constant values ('minfun', using the threshold of 0.01), but that decreased the overall performance of our models. Therefore we used 17 features (out of 20) for our models. Secondly, we used the **MinMaxScaler**, which gave all the values a new value from 0 to 1. We used MinMaxScaler instead of Standard Scaler because we had many features with bimodal distribution, as opposed to normal distribution.

We trained **six different models** (which are listed in the table on the left). We used **Randomized Search** to find the optimal parameters for the models. We also tried out ROC, but this ended up overfitting. We then chose the 4 models that had an accuracy score that was higher then 97.5% – KNN, Random Forest, Multilayer Perceptron and Support Vector machine – and combined them all together into one model.

In the end, we reached a model with an accuracy score of 98.30% on the training data and 97.79% on the test data. F-measure stayed in a similar range for each of the models used.

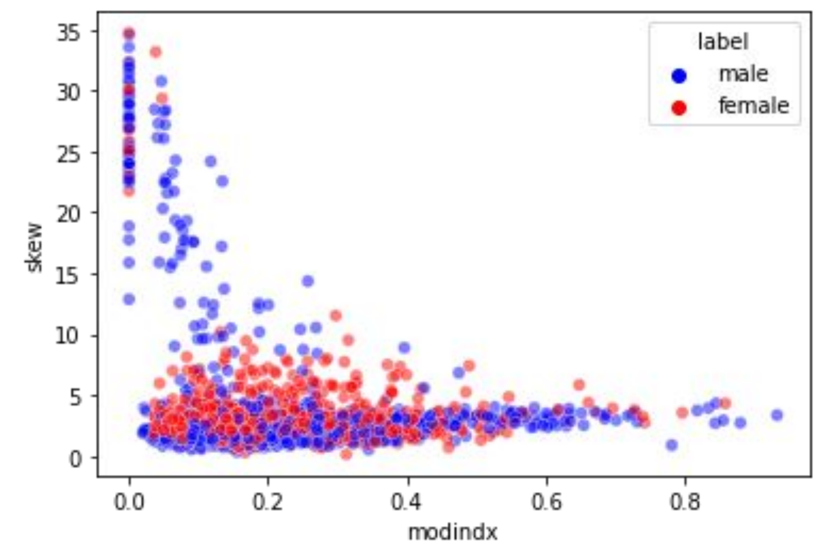
CORRELATION ANALYSIS

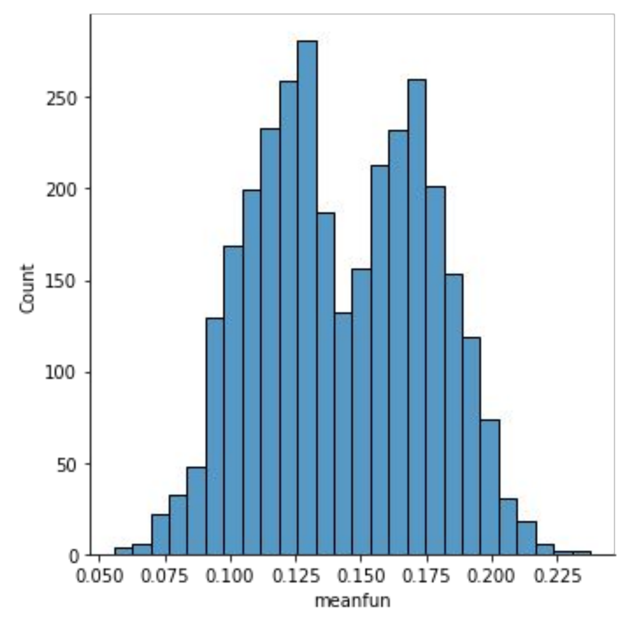
We analysed which speech features play the biggest role in defining a person's gender based on their voice.

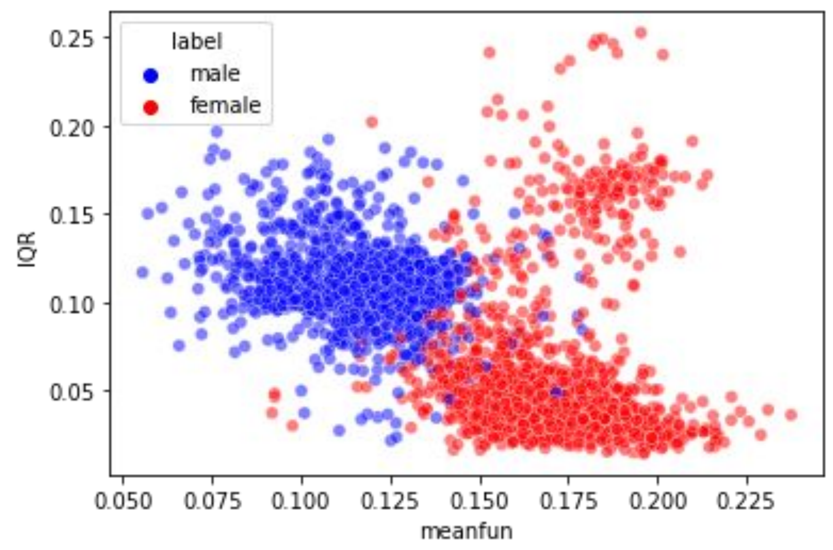
Upon initial analysis, we found some of the measures to be **bimodal**. From there, our hypothesis was that one of the peaks corresponds to the male values and the other to the female values, therefore the features would be highly correlated with the label.

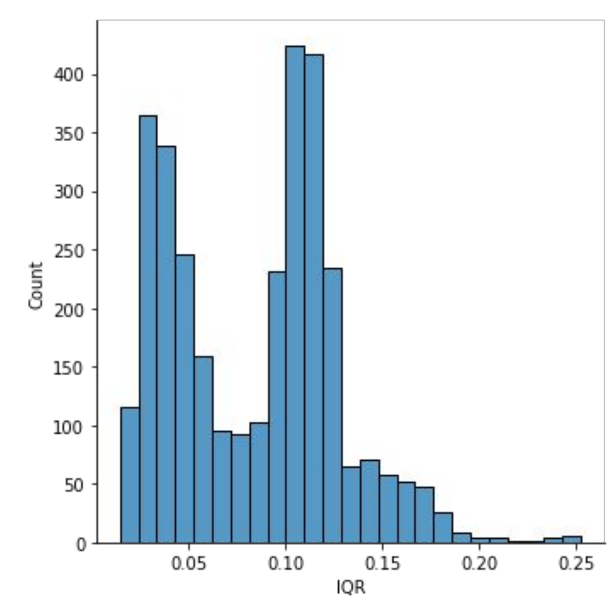
After further analysing all the measures, we found our hypothesis to be correct. The features that had the highest relevance to gender – 'IQR' and 'meanfun' – also had the most recognizably bimodal characteristics. On the right, the distributions of 'meanfun' and 'IQR' can be seen.

Next to these there are two graphs: the lower one shows the relations between 'IQR', 'meanfun' and gender. The upper graph shows the relations between the two least correlated features and gender.









IQR – interquartile range, describes the middle 50% of values when ordered from lowest to highest
Meanfun – average of fundamental frequency measured across acoustic signal

SUMMARY

Our initial goal with the model we created was to reach an accuracy score of at least 95%, which we surpassed, as we reached accuracy scores of 98.30% and 97.79% on training and test data, respectively. Our second goal was to find out which of the speech attributes were the most correlated to gender, and analyse the result. We found that the features with the highest relevancy to a person's gender are 'IQR' and 'meanfun'. We also saw that a lot of the other measures had a correlation that was close to none.