Recent Trends in Intensive Computing
M. Rajesh et al. (Eds.)

doi:10.3233/APC210192

Performance Analysis of ML Algorithms to Detect Gender Based on Voice

Raz Mohammad Sahar [a,1], Dr.T.Srivinasa Rao [a], Dr.S.Anuradha [a] and Dr.B.Srinivasa Rao [a]

[a]*Dept of computer science, GITAM. Deemed to be University, Visakhapatnam, India*

Abstract. Gender classification is amongst the significant problems in the area of signal processing; previously, the problem was handled using different image classification methods, which mainly involve data extraction from a collection of images. Nevertheless, researchers over the globe have recently shown interest in gender classification using voiced features. The classification of gender goes beyond just the frequency and pitch of a human voice, according to a critical study of some of the human vocal attributes. Feature selection, which is from a technical point of view termed dimensionality reduction, is amongst the difficult problems encountered in machine learning. A similar obstacle is encountered when choosing gender particular features—which presents an analytical purpose in analyzing a human's gender. This work will examine the effectiveness and importance of classification algorithms to the classification of gender via voice problems. Audial data, for example, pitch, frequency, etc., help in determining gender. Machine learning offers encouraging outcomes for classification problems in all domains. An area's algorithms can be evaluated using performance metrics. This paper evaluates five different classification Algorithms of machine learning based on the classification of gender from audial data. The plan is to recognize gender using five different algorithms: Gradient Boosting, Decision Trees, Random Forest, Neural network, and Support Vector Machine. The major parameter in assessing any algorithm must be performance. Misclassifying rate ratio should not be more in classifying problems. In business markets, the location and gender of people are essentially related to AdSense. This research aims at comparing various machine learning algorithms in order to find the most suitable fitting for gender identification in audial data.

Keywords. Machine Learning; Gradient Boosting; Decision Trees; Neural network; Support Vector Machine, Gender Classification.

1. Introduction

One characteristic of voice which is deeply perceived in humans is dimorphism. Specific features that differentiate human voices are speech rate, Tone, and duration. Specifically in males and females.[1].

[1] Raz Mohammad Sahar, Dept of computer science, GITAM. Deemed to be University, Visakhapatnam Email: raz.sahar2@gmail.com.

The thought of being dimorphic is 98.8% that constitutes the speaker's gender with its frequency. The distinction in gender, nonetheless, cannot make predictions by voice speech. Some voice pitches range between males and females, which makes it tough to identify males and females correctly.

By using Python language, recognition of the gender of a specific speaker is made possible, using methods applied for processing of speech in an asynchronous environment. Vocal cord width is the principal motive with which the difference between genders is calculated. The way a person talks and his present physical states are other motives which contribute to gender difference calculation. In a manner that one can identify a speaker as female or male, these abnormalities are used. Past researches have investigated the distinction between males and females, which includes various variables. Research shows that the major parameter for making speech analysis is derived from frequency, mean frequency, first quantile, and pitch, resulting in identification and classification. Identification of speech helps extract the information regarding gender, age, and accent in which they talk. A tremendous amount of research has been done in this area. Certain speech census is being implemented, which makes use of over period, maximum value, and mean to identify gender.

The audial data is converted into various parameters such as the strength of vocal, first quartile, third quartile, frequency, kurtosis, Spectral Flatness, spectral entropy, and so on. the previously mentioned parameters are trained and tested using various algorithms in ML to identify any genders.

In this paper, a comparative algorithmic approach that identifies gender based on different classification algorithms is proposed. The Predictions are made fundamentally on the dataset where values would be processed coming from audio files. Comparatively, results received are brought into comparison with earlier prediction results, and estimation is done by using classification algorithms to decide which algorithm gives better outcomes in identifying gender-derived specific parameters. Precise prediction of how the comparative algorithmic approach recognizes the gender-derived from the mentioned algorithms is obtained.

2. Literature Survey

Classification of gender, processing, and gender-derived identification is being carried out for a considerable period of time. Some theories used emerged overtime to carry out gender identification. New research based on the identification of gender and identification shows that speech is turned into various parameters. The major parameters are pitch and frequency. Identification is being carried out to distinguish males and females. Firstly, the system is equipped with training data, and then data is presented and assessed for the system's outcome for the data. These results collected may differ for various algorithms and seem to give inconsistent outcomes at different periods. Dimensionality reduction is amongst the major problems encountered in machine learning. An identical obstacle is encountered when selecting gender-specific characteristics—that assist an essential purpose in the classification of the gender of an individual. This work will examine the effectiveness and importance of ML algorithms to the voice-derived gender classification problem. Gender-based identification using F0 frequency [1] This says that Random Forest befits better for speaker identification using pitch and fundamental frequency to distinguish males and females [2]. They are further tuning based on a bucketing method to increase the effectiveness of outcomes achieved.

Voice-derived word extracting laboratory view [3] predicts that the algorithm runs finer for this identification, and it gives promising results to do extraction of vowels in samples of males. Since all samples are being trained then tested, this effectively gives a solution. This is noticed that raising the inexpressible portion in speech related to ‘s’ pitch’s value sound rises, obstructing gender exposure in males. Likewise, booming the sound portion of the speech such as ’a’ reduces the value of pitch. It doesn’t recognize it the instant the speaker delivers 2 varieties of tones. Identification of speech in grown-ups confirms that they can be casual, and vocal length changes and seem girlish since it is not easy to distinguish the males and females.

Certain voices of the female are challenging to investigate on how high or low it is [4]. For instance, measuring a female voice from a single aspect may hardly satisfy all necessities. This paper's (pitch) high or low sound quality in males and females [4] suggests that the voice of females should be recognized with different variables than males like Emotional, Shrill, and Swoopy. This includes parameters in which female persons can vary from each other. Therefore, preprocessing of the dataset has to be done on this before classification of gender. just as the pitch perception. [5-6]. F0 Fundamental frequency contains a compound of dialectal and non-dialectal speakers’ information, and they both correspond to males and females, and it is dependent on the speaker's upper pitch and sound. [7] This guided to put a frequency f0 with not having any range experience and non-syllable outer information. It proves that the speaker’s voice changes between upper and lower pitches between speakers.

Gender identification by support vector machine [8] states that gender’s speech is examined by several speech techniques such as compression of speech, talking on the phone, and distinction in languages, etc. It shows that pitch of male voice, duration, and Mel of frequency approximately 100- 146Hertz and females of around 188-221Hertz. At this point, voice is classified using frequency, and it is obtained and examined. Gender Identification by Voice [9] states that gender recognition using Linear discriminant analysis (LDA) performs well. Nevertheless, even with this model, the test error rate is still larger than 10%. Gender classification by pitch analysis [10] shows that gender identification gives promising results using pitch and signal energy.

3. Algorithms

3.1. Gradient Boosting algorithm

It is ensemble learning. The main concept in gradient boosting is that models are made in series. Gradient boosting uses many weak learners; it converts the week learners into strong learners. Boosting algorithm uses a decision tree as a weak learner then, the errors are obtained at lead nodes it obtains the errors at the leaf nodes and creates a second new tree by using values of error as values of new observation, and this process goes on for a defined number of times. The gradient boosting algorithm is shown in figure 1. Then the evaluation of the trained model is carried out against testing data so the calculation can be done to determine evaluation metrics.

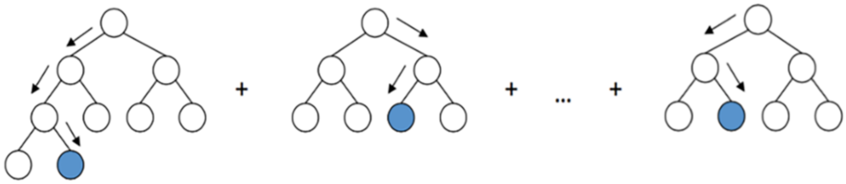


Figure 1.Gradient Boosting

3.2. Decision Tree

Decision Tree A decision tree is a classification algorithm, and it is supervised learning. A Decision tree consists of nodes, leaf nodes, and edges. Every decision tree is a structured tree in which every internal node shows an attribute as well as a feature. Here every decision rule is represented by branch, and every leaf node denotes the output or target variable. the decision tree is one of the most powerful famous tools for classification and prediction, and the decision tree gives good performance in classification related problems.

3.3. Random Forest

The random forest model is an ensemble algorithm. Ensemble algorithm combines algorithms of similar type or different type for the classification of objects or variables. Here classifier makes decision trees from the training dataset. It then checks or calculates the votes from various decision trees to choose the target variable or test object's final class. The features of random forest are more efficient working on large data sets, handles on more variables, estimates which variables are important during classification, estimates missing data, and so on.

3.4. Support Vector Machine

A Support Vector Machine is a classifier that is defined by separating with a hyperplane. It comes under supervised learning, and the hyperplane classifies the target variables or data points. The hyperplane is used for data points categorization. Data points that are on any side of the hyperplane create their own classes.

3.5. Neural Network

Neural network can be described as series of algorithms which aim to recognize underlying relationships in any data set by mimicking how the human brain functions. So, neural networks are related to neuron systems, both organic or artificial naturally. Neural networks can regulate to varying input; therefore, the network produces the greatest feasible result without redesigning the criteria of output. The idea of neural network is that it has its origins in artificial intelligence rapidly gains a reputation in the development of trading systems.

4. Pre-processing

4.1. *Dataset*

The dataset having 62,450 audios samples zipped (tgz) contains ten files can be automatically downloaded from this url= http://www.voxforge.org/ website.

4.2. Feature Extraction

First, all audio file's contents are read; after the essential properties are extracted and stored into a CSV file and, one can parse the README files to extraction metadata: like age, gender, and the pronunciation of the speaker. Reputably, Python comes with the package Scipy wave file is employed to obtain the audio, Scipy stats to do extraction of the prominent features, and NumPy and its FFT (Fast Fourier Transform) and fftfreq to extrapolate the audio data files to frequencies. All wav files' data are registered as amplitude in the domain of time, but the likely exciting features are those which come with a greater discriminative power male/female frequency.

In order to turn the audio to frequencies, DFT should be used, mainly the FFT algorithm. Implementation. Fourier transform receives a signal in the domain of time (set of measurements over a period of time), and it is then converted into a spectrum—a Group of frequencies with equivalent (compound) values. The spectrum never holds any information regarding time! In order to get the frequencies as well as the time at which recording was done for them, a signal of audio is divided in slight, OS (overlapping slices), and FT is applied on all (short-time Fourier transform). np.fft.fft delivers a compound range np. fft. fftfreq delivers the sample frequencies. A sample Directory consists of 10 audio tapes from a specific speaker.

Since Every wave file in the Directory is processed so that the dominating frequencies with 200ms windows sliding (1/5th of one second) can be extracted, when a wave file is 4 Secs long, extraction of a list having 20 frequencies will be done. For a sample folder (user), ten lists corresponding to the ten wav files (having twenty frequencies each) shall be collected in a list of lists. The frequencies are to be filtered to hold values in the speakers' voice within 20 hertz < frequency < 280 hertz. Also, any values in the range 50Hz are possible noise that should be filtered. fig2 shows the flow diagram of the model. as it can be seen in flow diagram, the first the audio files are automatically downloaded, then all the download audio files are unzipped, all the important voice features are extracted, then the extracted features are saved in a CSV file, five different algorithms: Gradient Boosting, Decision Trees, Random Forest, Neural network, and Support Vector Machine are used to classify the gender of a speaker, performance all the five classification algorithms are compared, and the algorithm with the best accuracy is chosen for real-world application.

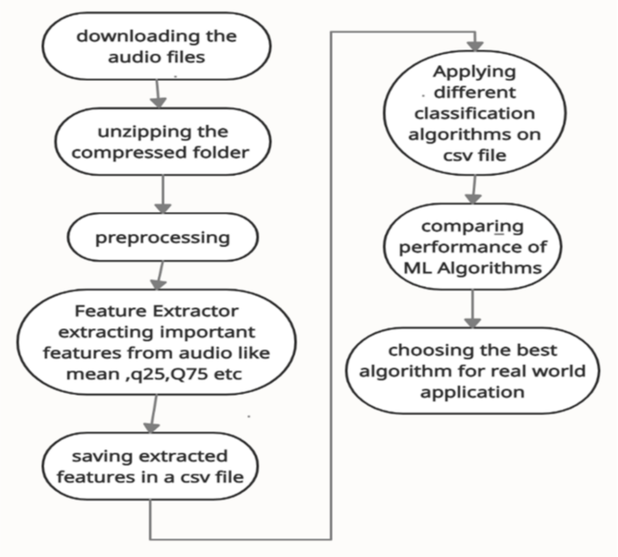

downloading the audio files
unzipping the compressed folder
preprocessing
Feature Extractor extracting important features from audio like mean ,q25,Q75 etc
saving extracted features in a csv file
Applying different classification algorithms on csv file
comparing performance of ML Algorithms
choosing the best algorithm for real world application

Figure 2. flow diagram

4.3. Data Description

After feature extraction, 3000 audio samples of voice are obtained, as every audio sample is an. AIF file. Preprocessing of AIF format files has been carried out for audial analysis using feature extraction. So, by applying the feature extraction technique, around 22 features can be obtained from acoustic signals. The extracted and preprocessed will be stored into a file consisting of 3168 rows and twenty-one 21 columns. Using 20 features, the prediction of the output label is made. The splitting of total data is done into parts, i.e., train data set and test dataset.
acoustics features - 20 Acoustics features are listed below, and Acoustic Properties with Distribution is shown in figure 3.

- Frequency Standard Deviation
- Mode of Frequency
- Median of Frequency
- Q25 Lower Quartile
- Q75 Upper Quartile
- Kurtosis
- Frequency Centroid
- Spectral Flatness
- Spectral Entropy
- Average of Fundamental Frequency Measured Across Acoustic Signal
- Minimum Fundamental Frequency Measured Across Acoustic Signal
- Maximum Fundamental Frequency Measured Across Acoustic Signal
- Minimum of Dominant Frequency Measured Across Acoustic Signal
- Average of Dominant Frequency Measured Across Acoustic Signal
- Maximum of Dominant Frequency Measured Across Acoustic Signal
- Maximum of Dominant Frequency Measured Across Acoustic Signal

- Modulation

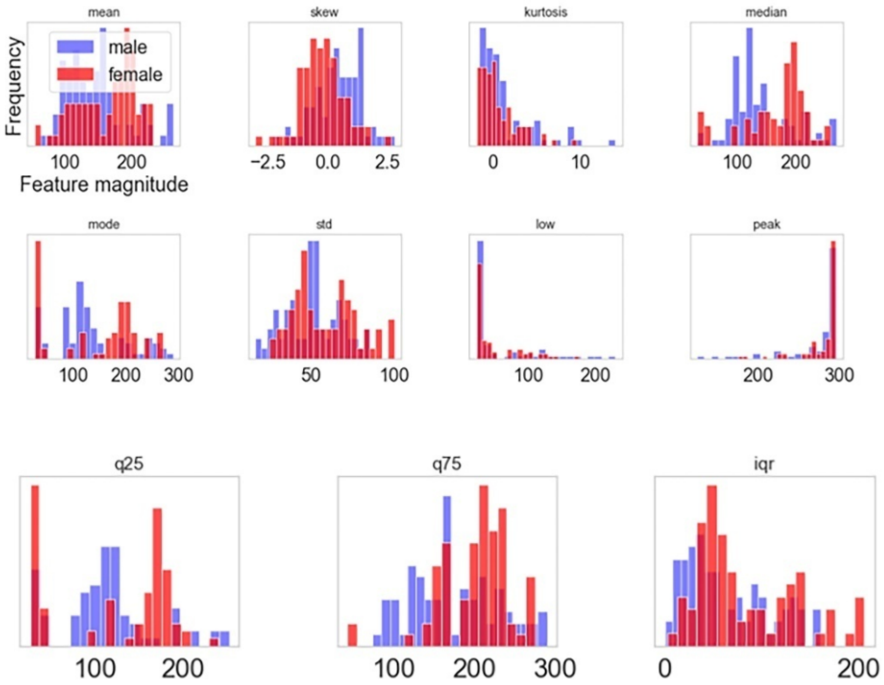


Figure 3. Acoustic Properties with Distribution

5. Results and Discussion

The problem of gender identification based on voice is addressed in this paper. The Mel-frequency Cepstral Coefficients of voice samples are commonly used as gender recognition features. The MFCC-based identification, on the other hand, is highly complicated. This paper proposes a comparative gender classification model that uses frequency pitch, first quartile, etc., to ensure effective gender classification while keeping the system simple.

The audio dataset used for this research can be downloaded from the http://www.voxforge.org/, The models are trained and tested with around 3000 males and female's audio files. The splitting of complete datasets is done into two datasets, testing and training dataset.30% of the entire dataset is considered for the test dataset. As the more influential the training dataset, the greater the model's result. If we train our model with an extensive dataset of training, it is more likely to get vital patterns. This paper evaluates the performance of five ML algorithms to identify gender and the results obtained is given in table 1, the results show that Gradient Boosting gives good Accuracy, the importance of independent and dependent variable for each algorithm is given in figures 4,5,6, it shows that the first quartile is the most important feature in the classification of gender.

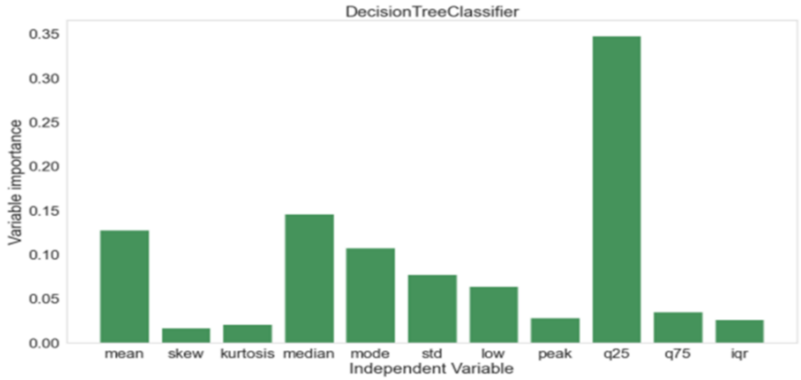
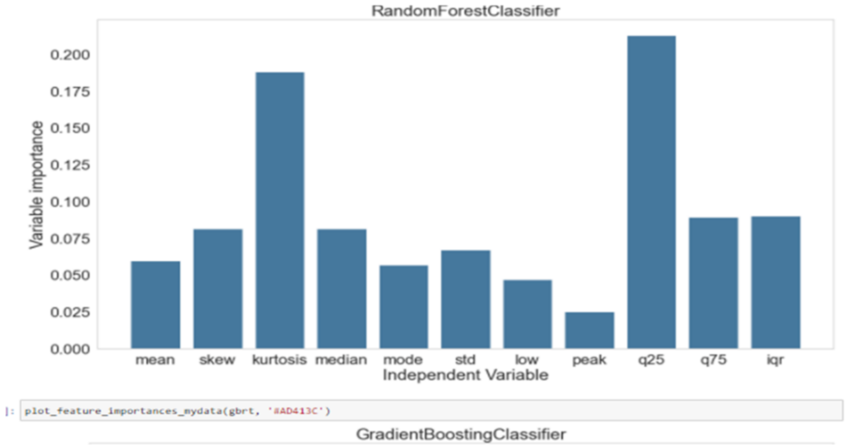


Figure 4. features importance in decision Tree



```
plot_feature_importances_mydata(gbrt, '#AD413C')
```

Figure 5. features importance in Random Forest

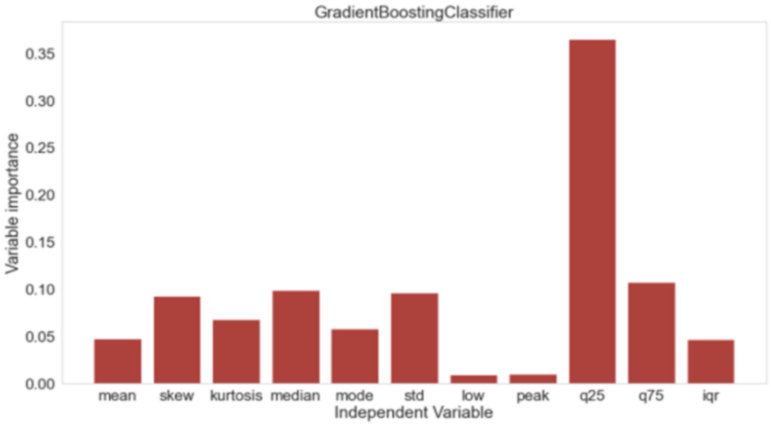


Figure 6. features importance in Gradient Boosting

Table 1. Comparison of results of different algorithms

Algorithms	Accuracy	Precision	Recall	F1 score
Random Forest	89%	81%	95%	88%
Decision Tree	82%	77%	79%	64%
Support Vector	88%	82%	87%	75%
Neural Network	89%	83%	90%	86%
Gradient Boosting	90%	82%	95%	88%

6. Conclusion

The results obtained show that the Gradient Boosting algorithm gives promising results in gender classification. Neural network and SVM also give relatively good results. The obtained results using ML classification algorithms are just for the dataset preprocessed from the 3000 audio files, and the result may change for other datasets. Gradient Boosting has higher accuracy than other algorithms in classifying gender. One of the difficulties in classification is that the collection of samples of the audio is generally done out of loud and noisy environments, limiting classification accuracy. More effective techniques to reduce and eliminate noise can be found out, which can be considered a route for future research.

References

[1] Ericsdotter, C., & Ericsson, A. M. (2001). Gender differences in vowel duration in read Swedish: Preliminary results. Working papers/Lund University, Department of Linguistics and Phonetics, 49, 34-37.

[2] Whiteside, S. P. (1996). Temporal-based acoustic-phonetic patterns in read speech: Some evidence for speaker sex differences. Journal of the International Phonetic Association, 26(1), 23-40.

[3] Byrd, D. (1992). Preliminary results on speaker-dependent variation in the TIMIT database. The Journal of the Acoustical Society of America, 92(1), 593-596.

[4] Henton, C. G. (1989). Fact and fiction in the description of female and male pitch. Language & Communication.

[5] Bishop, J., & Keating, P. (2012). Perception of pitch location within a speaker's range: Fundamental frequency, voice quality and speaker sex. The Journal of the Acoustical Society of America, 132(2), 1100-1112.

[6] Smith, D. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. The Journal of the Acoustical Society of America, 118(5), 3177-3186.

[7] Muhammad, G., Alsulaiman, M., Mahmood, A., & Ali, Z. (2011, July). Automatic voice disorder classification using vowel formants. In 2011 IEEE international conference on multimedia and expo (pp. 1-6). IEEE.

[8] Gaikwad, S., Gawali, B., & Mehrotra, S. C. (2012). Gender Identification Using Svm With Combination Of Mfcc. Advances in Computational Research, 4(1), 69-73.

[9] Jena, B., &Panigrahi, B. P. (2012). Gender classification by pitch analysis. International Journal on Advanced Computer Theory and Engineering (IJACTE), 1.