Acta Scientific COMPUTER SCIENCES

Volume 6 Issue 3 March 2024

Research Article

Gender Recognition by Voice Using Machine Learning

Moneerh Aleedy[1], Riham AlSmariy[2], Wejdan Alsurayyi[1]* and Suad Almutairi[1]

[1]*Information Technology Department, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia*

[2]*Computer Sciences Department, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia*

***Corresponding Author:** Wejdan Alsurayyi, Information Technology Department, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Received: February 07, 2024

Published: February 21, 2024

© All rights are reserved by **Wejdan Alsurayyi., *et al.***

Abstract

The gender recognition system by a sample of voice has an excellent mechanism based on many factors and features of sound signals like frequency, pitch, and loudness. This paper implements a machine to distinguish between the male's and female's voice. The experiments are implemented by two techniques to split speech dataset: train_test_split function and 10-fold Cross-Validation on the training set (after splitting the data into train and test set), to get the highest result and improve the model and make a comparison between them. Then, the K-best feature selection is used as a filtering method and Recursive Feature Elimination (RFE) as a wrapper method. This paper investigates various machine learning algorithms, such as logistic regression, random forest, SVM, gradient boosting, and decision tree. The performance was evaluated using different metrics: accuracy, recall, precision, F-score, and confusion-matrix. The results showed that splitting the dataset using 10-fold CV with RFE provides the best result for all ML algorithms and the best model is the random forest, it gives the highest percentage in all evaluation methods.

Keywords: Machine Learning; Voice; RFE; k-Best; Random Forest

Introduction

Determining the gender of a person as male or female, based on a sample of their voice, is an easy task at the beginning. Often, the human ear can detect the difference between male or female voice easily through the first words in speaking. It has an excellent mechanism for recognition and distinguishes the speaker's gender based on many factors and features of the sound signal like frequency, pitch, and loudness [1].

This gender recognition system is used in many practical applications, like identifying the user's gender helps to provide additional targeted services based on gender interoperability. Moreover, this system applied in Human-Computer Interaction (HCI) to allocate the user interface and improve the experience of the user in most IoT applications and provide gender customizations in these applications [2].

This project aims to teach the machine to distinguish between the male's and female's voice as easily as the human ear by determining the characteristics needed by the machine to classify the voice of male and female. The voice of males can range between 85 to 180 Hz while from 165 to 255 Hz in females' voices. It is clear that there is a common domain between them, so the classification based on the frequency is not enough as shown in Figure 1 [1].

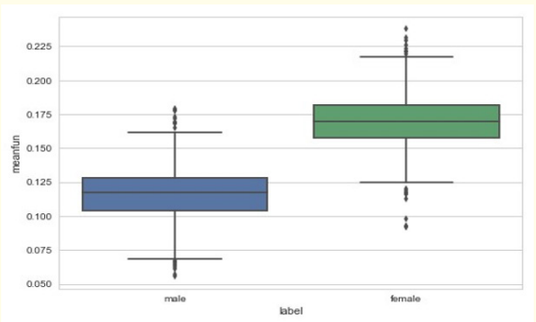

Figure 1: Interference between male and female voice frequency.

This project will distinguish between the gender of the human - male or female – based on characteristics of their voice. Then it analyzes and builds the model that distinguishes the sound based on the training models [1].

There are plenty of methods applied to gain good accuracy in the field of voice recognition. Becker used many models to calculate the highest testing accuracy [3]. He utilized the frequency of a voice with a baseline model to determine the gender, after that he tried using logistic regression which is a powerful algorithm for classifying two categories; it is a binary classifier. It gives a function that is a boundary between two different categories, and it can be extended to work with more than two classes [4].

Digging deeper, he employed a classification and regression tree model to the dataset to determine how the voice features might correspond to a gender classification of male or female. The same procedure followed with a random forest model which is an ensemble method for classification that is containing many decision trees to determine the outcome. This method combines the bagging idea and random selection [5].

Moving to a step further, he applied a generalized boosted regression model. After that, he used Support Vector Machines (SVM) which is a binary classification algorithm that used to classify between two classes by finding the linear hyperplane to separate the two classes with the maximal margin [5]. Moreover, he applied XGBoost algorithm for regression and classification problems, it makes a prediction model in the form of an ensemble of previous models to minimize the overall prediction error [4]. Finally, he used the best 3 models from his experience and combine them in the stack: SVM, random forest, and XGBoost. The results are summarized in Table 1.

Accuracy (%)		
Model	**Train**	**Test**
Logistic regression	72	71
Random forest	100	87
Support Vector Machine	96	85
Frequency-based baseline	61	59
Classification and regression tree	81	78
Boosted tree	91	84
XGBoost	100	87
Stacked	100	89

Table 1: Becker's models for voice recognition [3].

The highest accuracy percentage gained by using a stacked algorithm, and it got 89% in testing accuracy, which is a very high ratio.

The rest of this paper is organized as follows. Section 2 provides research method, dataset collection, and analysis in detail. The results and discussions of this project are discussed in section 3. Finally, the conclusion of the project is provided in Sections 4.

Research method

This section explains the methodology followed for this project. At first, prepare the dataset for modeling. The preparing process includes preprocessing step and features extraction then train the models using a training set and evaluate them with a test set.

The dataset split using two techniques, to get the highest result and improve the model and make a comparison between them. The first splitting way is by using train_test_split function under scikit-learn library, which is a straightforward and helpful function for partitioning data. The second one is using 10-fold CrossValidation on the training set (after splitting the data into train and test set), so the model will be trained on the training set and tested using unseen data which is the test set. The training set partitioned into 10 subsamples, one of them kept for validation, the Cross-Validation repeated 10 times, and the average score will be obtained.

Besides, the most effective features are extracted to reduce complexity and overfitting. Filtering and wrapper methods are more suitable to provide the best subset of features by training a model on it. K-best feature selection is used as a filtering method and Recursive Feature Elimination (RFE) as a wrapper method [6].

After that, four of the most effective machine learning algorithms that have been used in the related work are chosen to be applied to the dataset [7,8]:

- **Logistic Regression:** A basis binary classification algorithm.
- **Random Forest:** A supervised classification algorithm that creates a forest with many trees.
- **SVM:** A popular supervised algorithm that provides good performance when dealing with low-dimensional data
- **Gradient boosting:** A sequential learning technique in which model performance improves over repetition.

Moreover, a new method which is called a decision tree has been selected where the features are tested at each node in the tree, and the correct class is predicted at the leaf node in the tree [9].

Dataset collection and analysis

The speech dataset that has been used in this work is available on the Kaggle website [10]. Original voice samples are the in.WAV format, techniques have been applied to those samples to extract 22 acoustic parameters, peak and mode frequency are ignored since they have the same value for all samples. Moreover, the result saved in CSV file which contained 3168 and 21 columns [3]. The parameters are shown in Table 2.

	Acoustic Properties
Properties	**Description**
Duration	Length of signal
Meanfreq	Mean frequency (in kHz)
Sd	The standard deviation of the frequency
Median	Median frequency (in kHz)
Q25	First quantile (in kHz)
Q75	Third quantile (in kHz)
IQR	Interquantile range (in kHz)
Skew	Skewness
Kurt	Kurtosis
sp.ent	Spectral entropy
sfm	Spectral flatness
Mode	Mode frequency

Centroid	Frequency centroid
Peakf	Peak frequency
Meanfun	Average of fundamental frequency measured across the
	acoustic signal
Minfun	Minimum fundamental frequency measured across the acoustic signal
Maxfun	Maximum fundamental frequency measured across the acoustic signal
Meandom	Average of dominant frequency measured across the acoustic signal
Mindom	Minimum of dominant frequency measured across the acoustic signal
Maxdom	Maximum of dominant frequency measured across the acoustic signal
Dfrange	The range of dominant frequency measured across the acoustic signal
Modindx	Modulation index

Table 2: Acoustic properties of each voice are measured [3].

Description of the testbed

Using Python libraries, the codes have been written. Python Notebook environment (Jupyter version 3.6) combines remarkable power with clear syntax. Moreover, some additional Python libraries should be installed using Anaconda prompt to run this project correctly, which is missingno. Missingno used to visualize and check for missing values. The command used to install it is 'pip install missingno'.

Data preprocessing

In order to prepare the dataset for modeling. First, check the missing values, and observe that all the records do not have missing values especially the label, so the dataset can be considered as supervised. Then encode the Label feature to be 0 and 1 instead of female/male.

For visualization, quick statistical data description, and individual distribution for each feature are provided (see Figure 2) to identify where are their values clustered.

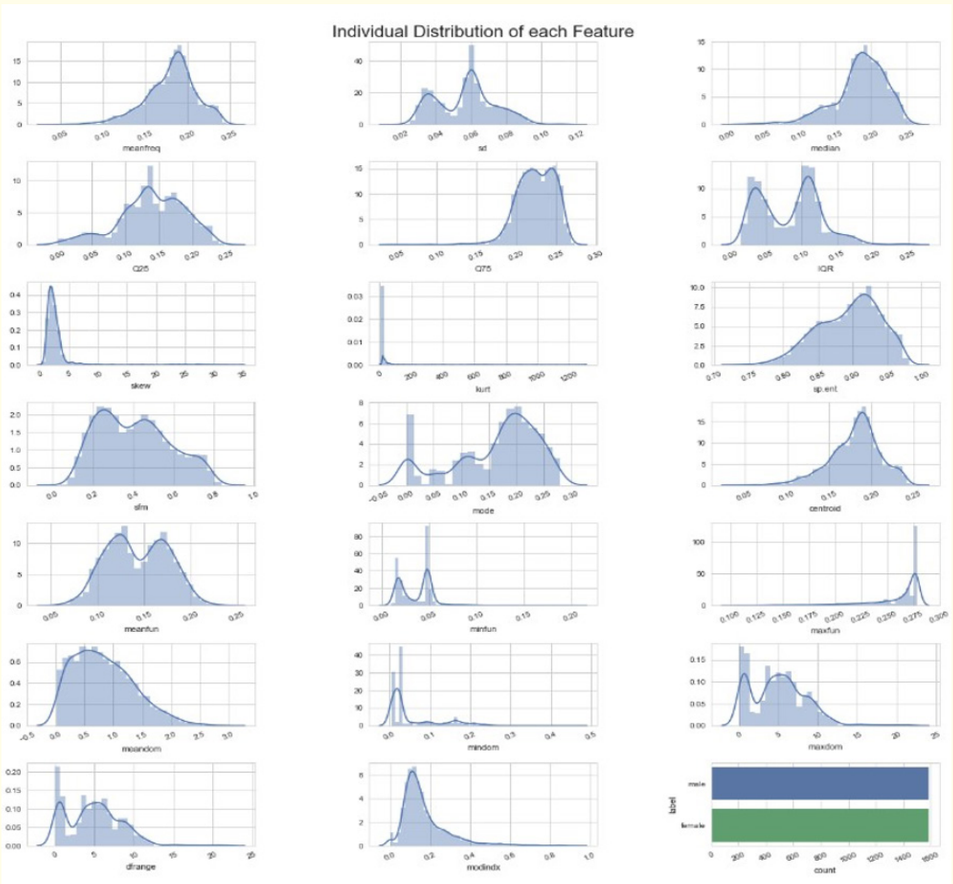

Figure 2: Individual distribution for each feature.

The differences between values of features are very high to observe, to ensure accurate data visualization, doing the normalization is needed. The Violin plot shows that Kurt value is highly scaled, illustrated in Figure 3. The data are normalized as shown in Figure 4.

Feature extraction

In order to analyze the data, first, load the CSV file into the Pandas data frame. After that prepare data for modeling, the label column was changed from string to a binary, with 0 signifying female

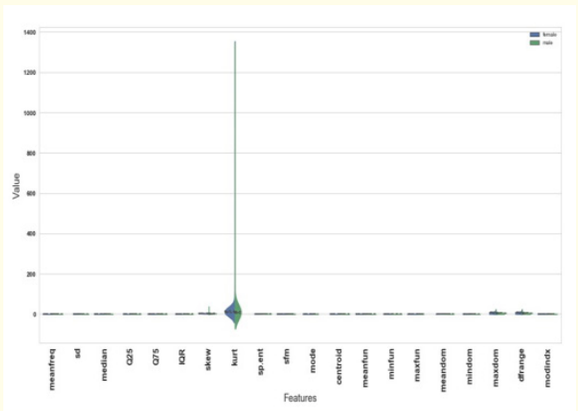

Figure 3: Individual distribution for each feature.

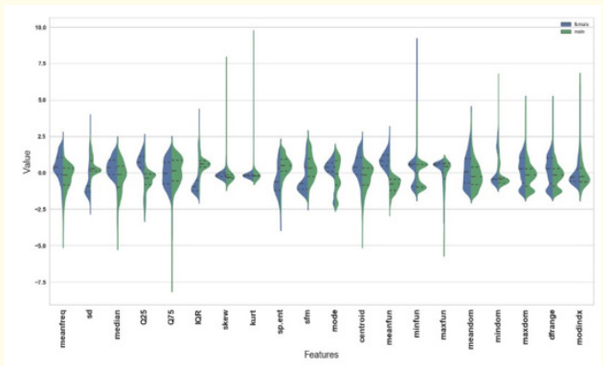

Figure 4: The features values after the normalization.

and all data are normalized. Then, draw the heatmap which shows the correlation between all features. Figure 5, shows that meanfreq, centroid, Q25, and median are highly correlated, so in feature selection, it is enough to choose one of them to reduce the complexity.

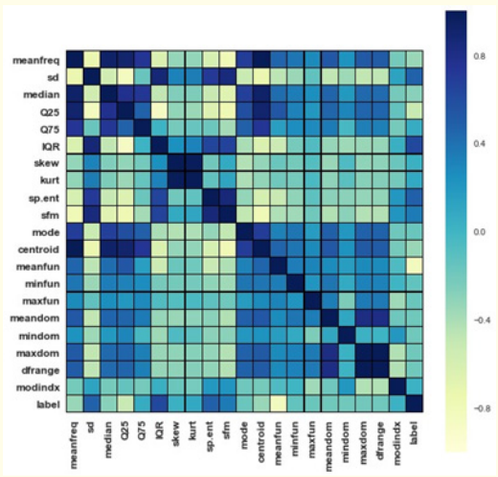

Figure 5: Features correlation using the heatmap.

The most effective features among 20 features have been selected using two feature selection algorithms. By using the k-Best features selection algorithm, the following features considered as the most effective ones (sd, Q25, IQR, sp.ent, and meanfun). Then, those features used with the two splitting algorithms and the chosen five ML algorithms. The second way to select features is Recursive Feature Elimination (RFE). It is used to select the 5 best features and get different features with each ML algorithms.

Modeling

To build the model, first split the dataset by using two methods train_test_split and 10-fold Cross-Validation with the test size percentage that held over for testing is 20%. Then, train the model with five ML algorithms (logistic regression, random forest, SVM, gradient boosting, and decision tree).

In order to improve the model, the 5 best features have been selected, and different features for each ML algorithms are chosen. Again, those features have been used with the two splitting algorithms and ML algorithms applied before.

The model tested by the test set and compare the different ML algorithms by the evaluation methods such as accuracy, recall, precision, F-score, and confusion-matrix.

Evaluation

There are several performance metrics that used to evaluate Machine Learning Algorithms. The results have been evaluated by comparing accuracy, recall, precision, Fscore, and confusion-matrix. Accuracy is the most common metrics used to evaluate the generalization ability of classifiers, it is the number of true predictions made by the trained classifier over all types of predictions made.

Precision and recall are both widely used in information extraction and binary classification. Precision is the number of positive predictions divided by the total of positive predictions of class values, while recall is the number of positive predictions divided by the total of positive class values into the test data.

F-score is the balance between precision and recall. It is called the F-measure. On the other hand, confusionmatrix is one of the easiest performance metrics that used to find the accuracy and correctness of the model, it is used for the classification problem [11].

Results and Discussions

In the end, we studied and compared the obtained results and compared the evaluation methods to get the best model. The results of using the train_test_split function for each ML algorithm are shown in Table 3. On the other hand, the results of using 10-fold Cross-Validation, are shown in Table 4.

	Score using split_train_test (%)			
	Model	**Whole Features**	**5-Best**	**RFE**
1	Random Forest	97.48	97.48	97.00
2	Gradient Boosting Trees	97.48	97.32	97.48
3	Logistic Regression	97.32	97.16	97.79
4	SVM	97.00	97.16	97.63
5	Decision Tree	95.90	95.90	96.37

Table 3: Accuracy for each ML algorithm using SPLIT_TRAIN_TEST.

	Score using 10-fold (%)			
	Model	**Whole Features**	**5-Best**	**RFE**
1	Random Forest	97.63	97.32	97.95
2	Gradient Boosting Trees	97.32	97.79	97.79
3	Logistic Regression	97.00	97.00	97.48
4	SVM	97.00	97.00	97.48
5	Decision Tree	95.74	95.58	96.69

Table 4: Accuracy for each ML algorithm using 10-FOLD.

From the previous results, we can see that splitting the dataset using 10-fold Cross-Validation is doing well and provides higher frequency. Random forest, logistic regression, and gradient boosting tree algorithms are having the highest accuracy among other ML algorithms. By using feature selection, we gain higher ratios, and it naturally decreased the complexity. RFE provides the best result with both data splitting ways. Comparing the results obtained in Table 3 and Table 4, allows us to conclude that splitting the dataset using 10-fold CV with RFE provides the best result for all ML algorithms.

Finally, a comparison of the methods is applied by calculating accuracy, precision, recall, and f-score see Table 5. In conclusion, the best model is the random forest, it gives a higher percentage in all evaluation methods, and 97.95 % inaccuracy, which is higher than what Backers found see Figure 6 for the confusion matrix.

	10-fold with RFE (%)				
	Model	**Accuracy**	**Precision**	**Recall**	**f-score**
1	Random Forest	97.95	98.0	98.0	98.0
2	Gradient Boosting Trees	97.79	98.0	98.0	98.0
3	Logistic Regression	97.48	98.0	97.0	97.0
4	SVM	97.48	98.0	97.0	97.0
5	Decision Tree	96.69	97.0	97.0	97.0

Table 5: Evaluation methods for each ML algorithm using 10-fold CV and RFE.

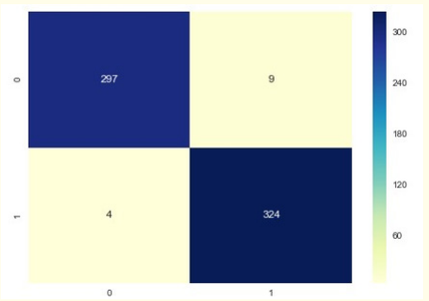

Figure 6: Confusion Matrix for Random Forest.

Conclusion

The results obtained in this paper shows that the acoustic properties of the voices and speech can be used to detect the gender. Two ways of data splitting and five best machine learning algorithms have been used, with and without two kinds of feature selection algorithms, to acquire the best results. Random forest algorithm performs better in classifying gender among other algorithms; the accuracy is higher. Using a larger data set can be maximizing the accuracy of the results and trying other algorithms with a different kind of feature selection could bring higher accuracy in gender recognition.

Acknowledgment

This research was funded by the Deanship of Scientific Research at Princess Nourah bint Abdulrahman University through the Fast-track Research Funding Program.

Bibliography

1. V Ahirkar and N Bansal. “Gender Recognition Using Voice”. 2.3 (2017): 65-69.
2. Z Hong. “Speaker gender recognition system”. (2017): 54.
3. M Buyukyilmaz and A O Cibikdiken. “Voice Gender Recognition Using Deep Learning”. Proc. 2016 Int. Conf. Model. Simul. Optim. Technol. Appl., no. December, (2016).
4. P Harrington. Machine Learning in Action. (2011).
5. A Raahul., *et al*. “Voice based gender classification using machine learning”. IOP Conf. Ser. Mater. Sci. Eng., 263.4 (2017): 42083.
6. S Kaushik. “Feature Selection methods with an example (or how to select the right variables?)”. (2016).
7. D Zhang., *et al*. “A Data-Driven Design for Fault Detection of Wind Turbines radiation forecasting: A review”. Renew. Energy, vol. 105, Using Random Forests and XGboost”. in IEEE Access 6 (2017): 569-582.

8. Nair RR and Vijayan B. “Voice based Gender Recognition”.

9. C Voyant., *et al.* “Machine learning methods for solar radiation forecasting: A review.” *Renewable Energy* 105 (2017): 569-582.

10. A Lindner. “Gender Recognition by Voice”.

11. M Hossin and M N Sulaiman. “Review on Evaluation Metrics for Data Classification Evaluations”. *International Journal of Data Mining and Knowledge Management Process* 5.2 (2015): 1-11.