Male Female Voice Recognition As An Initial Design For Voice Authentication Alternatives

Hermawan Setiawan[1], Tio Hana Lolita Lumban Tobing[2]*, Jonathan Sebastian Marbun[3], Bahteramon Bintang Sanjaya Manurung[4], I Gede Maha Putra[5], M. Rakhmat Dramaga[6]

[1,3,4,5,6]Cryptography Engineering, Politeknik Siber dan Sandi Negara, Indonesia
[2*] Badan Siber dan Sandi Negara, Indonesia
Email: [1]hermawan.setiawan@poltekssn.ac.id, [2,*] tio.hana@bssn.go,id, [3]jonathan.sebastian@student.poltekssn.ac.id, [4]bahteramon.bintang@student.poltekssn.ac.id, [5]igede.maha@student.poltekssn.ac.id, [6]m.rakhmat@student.poltekssn.ac.id

A B S T R A C T

Artificial Intelligence (AI) technology has experienced a very rapid development, where AI has become a part of daily life. One of the applications of AI that has been widely implemented is the Voice Gender Recognition service. With the help of AI, it can be used to find out the gender of the voice sample uploaded to the application. Speech recognition is not something easy to do. Not a few did not manage to get the desired result. This study is the first step that sound becomes a way of authenticating a person. Applications are built using the R language. CART models are used to make decision-making easier. A sample of 20 respondents was taken with various conditions when recorded. The final result of the application gets an accuracy percentage of 82%.

Keywords: Artificial Intelligence, CART, R, Voice Gender

Corresponding Author:

Hermawan Setiawan
Cryptography Engineering, Politeknik Siber dan Sandi Negara
Email: hermawan.setiawan@poltekssn.ac.id

1. INTRODUCTION

The human voice contains a lot of data in it. The data contained in this can be used to identify various things from the owner of the voice such as age, gender, feelings, and even mental condition. Therefore, in this very rapid development era, the development of technology, especially artificial intelligence, is used to process the existing data. From its application, we can process the data obtained from the voice to determine certain things that make our lives easier with the technology that was created [1].
With the development of technology, especially Artificial Intelligence (AI) or Machine Learning, several ways have been found to make machines do what only humans should be able to do[2], [3]. Smart Assistants (such as Siri, Alexa, etc.), Online Advertisements, and ChatBots are some examples of the application of Artificial Intelligence.
Gender prediction is useful in many applications such as interactive systems, targeted advertising, health care systems through mobile phones, recognition systems, crime analysis (gender identification taken part in crimes by voice) and so on[4]. In general, voices and speech are used for gender classification. The natural voice recognition system is the human ear. The human ear has a mechanism for distinguishing the sex by voice and speech based on various factors [5]. physiological signals that represent information at different levels such as linguistic content (such as language, words, accents etc.) and paralinguistic content (such as gender, age, emotions etc)." Along with the improvement of technology and the use of electronic devices (such as Google assistant mobile phones, alexa) has been introduced and in peak demand according to the market value in question.

The application of Voice Gender Recognition is very important for several security systems such as the gender identification of the speaker which is quite crucial[6]. Another example of application is in the field of entertainment, namely its use to identify the duration when men and women appear on film screens. In oral communication, humans can easily distinguish the gender of the voice's owner. The human brain can automatically determine the characteristics of a male or female voice. The characteristics that are easiest to distinguish are pitch, intonation, speed, and so on[7], [8]. These characteristics can be used as a reference in modeling an AI program on sound samples.

Sound features can be considered sound prints capable of identifying the sex, age, origin, and emotional state of the speaker. The gender recognition system automatically extracts features from sound signals and those features will be used to determine the speaker's gender[9], [10]. This gender recognition system can be utilized in several practical applications, such as determining the gender of the user which will be conducive to providing more targeted services based on gender interoperability. Also, speaker gender recognition systems implemented in Human-computer interaction can determine the scope of the user interface and improve the user experience in Most IoT applications. The information can be used to provide gender adjustment in those IoT applications and promote the level of security in those applications, one of which is authentication [11].

Artificial Intelligence in Voice Gender Recognition will be implemented in the form of an application built using the R language. The R language is a programming language devoted to statistical calculations. The CART model will then be used in making decisions.

A. Gender-Based Voice Classification

Voice-based gender classification essentially predicts the gender of a voice by analyzing several different parameters of the voice sample [12]. Some of the distinguishing features of the human voice are intonation, speaking rate, and duration. With the R language, we can exploit these properties and perform calculations to classify whether the voice belongs to a male or a female. The main parameters that can be used when performing the analysis are frequency and tone [13].

B. R Languange

The R language is a language dedicated to performing statistical and graphical computations [14]. The R language is an evolution of the S and Scheme languages, both of which are also languages for statistical data analysis. R is written using the ANSI standard C because it is the most widely used programming environment. In the R language, each function belongs to a different package. So if we want to use a function, we must first install the package that contains the function. Then the package used in the session must be declared. R has its console called R Console. However, there is also software that makes using the R language easier, namely RStudio. RStudio provides services to deploy web applications using R language with the Shiny app[15], [16].

C. CART Model

CART, or Classification and Regression Tree, is a binary recursive training technique that can handle continuous attributes such as goals and predictors [17]. Starting with a root node, the data is split into 2 branches, and each child has then split again. There is no rule about the maximum size of the tree that the process will only stop when the branch can no longer be split because of data shortage. CART does not aim to build just 1 single tree, but a tree that keeps growing. The ideal tree is obtained by evaluating the predictions of each tree. Tree performance is evaluated using independent test data.

2. EASE OF USE

A. Gender Voice Recognition Service

The preference and use of voice biometric systems is evidenced by the improvement of voice search applications [18]. The Voice Gender Recognition program essentially performs an acoustic analysis using the Specan function in the warbleR package on the input voice sample and decomposes the characteristics of the voice into 20 parameters. The 20 parameters are meanfreq (mean frequency), sd (standard deviation), median (median frequency), q25 (first quantile), q75 (third quantile), iqr (distance between quantiles), skew (skewness), kurt (kurtosis), sp.ent (spectral entropy), sfm (spectral flatness), mode (mode frequency), centroid (frequency centroid), meanfun (average fundamental frequency), minfun (fundamental frequency minimum), maxfun (fundamental frequency maximum), meandom (fundamental frequency dominant average), mindom (fundamental frequency dominan minimum), maxdom (fundamental frequency dominan maximum), maxdom (fundamental frequency dominan maximum), dfrange (jarak frekuensi dominan), and modindx (indeks modulation) [19]–[22].

After the sound has been broken down into these 22 parameters, the data is saved in csv format. Data on the results of the vote split are compared with the collected data set. Here are the upper and lower bounds of the male and female voice records.

Table I. comparison of male and female characteristics

	Male (high)	Male (low)	Female (high)	Female (low)
maxfun meandom	0.2758620 0.0078125	0.25 0.4609375	0.2758620 0.2729640 15151515	0.2711864 0.2270220 58823529
mindom maxdom dfrange	0.0078125 0.0078125 0	0.0878906 0.7861328 0.6982421	0.046875 0.7421875 0.6953125	0.0078125 0.5546875 0.546875
modindx	0	0.5521145	0.3398876	0.35
Q25	0.0150714	0.1078787	0.0623500	0.0700715
Q75	0.0901934	0.2167965	0.2245521	0.2508273
IQR	0.0751219	0.1089177	0.1622021	0.1807558
skew	12.863461	3.6485932	2.8013439	1.7050291
kurt	274.40290	20.799924	19.929616	5.7691153
sp.ent	0.8933694	0.8982253	0.9521609	0.9388294
sfm	0.4919177 66397811	0.4011686 33192776	0.6792233 12622536	0.6015288 10198165
mode	0	0.1106493	0.0499260	0.2677017
centroid	0.0597809	0.1587811	0.1581079	0.1655089
meanfun	0.0842791	0.1152992	0.1850417	0.1856069
minfun	0.0157016	0.0793650	0.0230215	0.0622568
x	0.2758620	0.25	0.2758620	0.2711864
meandom	0.0078125	0.4609375	0.2729640	0.2270220
mindom	0.0078125	0.0878906	0.046875	0.0078125
maxdom	0.0078125	0.7861328	0.7421875	0.5546875
dfrange	0	0.6982421	0.6953125	0.546875
modindx	0	0.5521145	0.3398876	0.35

B. CART Model

A CART model can be constructed based on the upper and lower bounds of the parameter data sets for male and female vocal characteristics.

The CART model is built considering the data set to determine the limits of the decision of whether to classify a voice as male or female based on the parameters. The main parameters used as decision branches are 7 parameters that show significant differences between females and males in the data set. The main determinant or root node is taken from the frequency mode, which is the main feature in distinguishing tones. Then branch splitting is done based on other parameters like maxdom, minfun, q25, median, meanfun, and skew[23], [24].

In an analysis process, modeling is performed 5 times with CART to increase accuracy[25], [26]. The CART model that has been compiled and becomes the decision-maker in the system is shown in Figure 1.

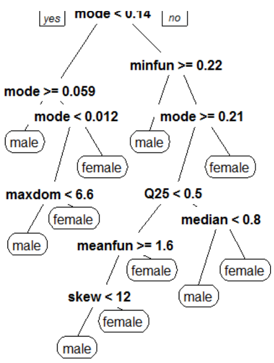

Fig 1. The CART model of the Men and Women voice discriminatory parameters(Becker, n.d.).

C. Application Voice Gender Recognition

The development of the speech recognition application in this study uses RStudio 1.1.463 with the MacOS operating system. The application is built using the Shiny package, so the application becomes web-based. The representation of the created application can be seen in Figure 2. The application has a function to upload the sound file that you want to use. Uploaded sound files must be in waveform (.wav) format. This is because the .wav format audio file does not undergo any changes in either frequency or size, so the results of the analysis are what they are. .mp3 audio files, on the other hand, have undergone resizing and compression, making them easier to distribute [27].

Fig 2. Display when the audio file is undergoing the analysis process.

After the analysis process is completed, the results are displayed in 3 ways namely: frequency graph, spectrogram and calculation details. To make it easier for users to see the results of the analysis, the words "female" are colored when the result is a woman pink, while the words "then" are colored blue[28].

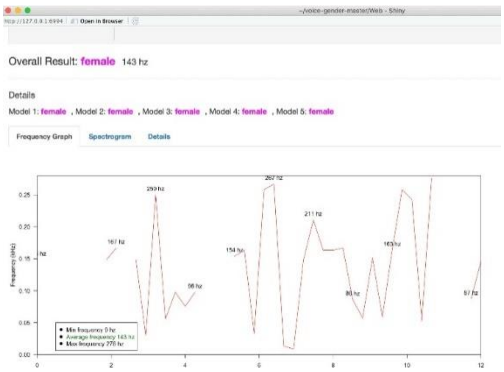

Fig 3. Calculation results by displaying a frequency graph

Fig 4. Calculation results by displaying the spectogram.

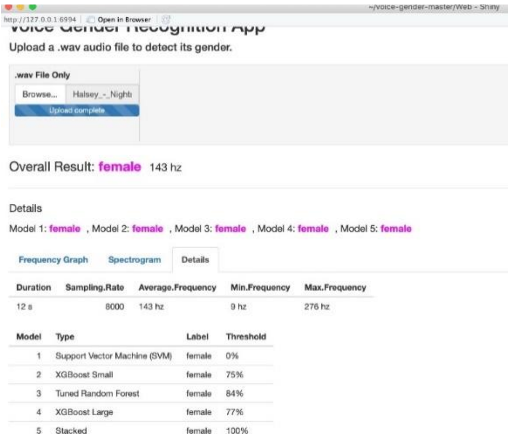

Fig 5. Calculation results by displaying details.

3. RESEARCH RESULT

A. Calculation Result And Accuracy

The application has been tested with 20 samples of audio files with different types of voices, be it the voice of someone singing, the voice of someone speaking, or the voice of someone giving a speech. Based on the results of using the application using a pattern, the results are shown in Table 2. The percentage is calculated by looking at the accuracy 5 times when creating the CART model. If the model gives a correct result, it will be marked with a tick (v), while an incorrect result will be marked with a cross (x).

Tabel II. results of male and female calculations

Sampel	1	2	3	4	5	Persentase
A	v	v	v	v	v	100%
B	v	x	v	v	v	80%
C	v	v	x	x	x	40%
D	v	v	v	v	v	100%
E	v	v	v	v	v	100%
F	v	v	x	v	v	80%
G	v	v	v	v	v	100%
H	x	x	x	x	v	20%
I	v	v	v	v	v	100%
J	v	v	v	v	v	100%
K	v	v	v	v	v	100%
L	v	v	v	x	v	80%
M	v	x	v	v	x	60%
N	v	x	v	v	v	80%
O	x	v	v	v	v	80%
P	v	v	x	x	v	60%
Q	v	v	v	v	v	100%
R	x	v	v	v	v	80%
S	v	v	v	v	v	100%
T	v	x	v	v	v	80%

Based on the results of the analysis, the average accuracy percentage was found to be 82%. This can be caused by audio recording results that do not match the original or inconsistent CART modeling. However, with results close to 100%, it can be said that the application went well.

4. CONCLUSION

Artificial Intelligence Speech Gender Detection Application written in R language is a web application for gender determination of the analyzed sound. The uploaded sound file must be in .wav format as it does not experience frequency changes and size compression like in .mp3 format. The uploaded sound file is subjected to an acoustic analysis using the specan function in the warbleR package and broken down into 20 parameters. The results of the acoustic analysis are then modeled using CART and compared according to the dataset created.The application was tested with 20 different sound samples and based on the calculation of the percentage of accuracy, it was found that the application achieved an accuracy of 82%.There are problems testing the application when there are sound files that cannot be processed. This may be because the sound file is not perfect. There is also a problem when there are sound files which, when analyzed more than once, produce different analysis results with different accuracy percentages.Applications can still be improved and redesigned to achieve greater accuracy.

5. REFERENCES

[1] P. Kumar and S. B. Rana, “Development of modified AES algorithm for data security,” *Optik (Stuttg).*, vol. 127, no. 4, pp. 2341–2345, 2016, doi: 10.1016/j.ijleo.2015.11.188.

[2] H. Wu and Y. Qi, “Application of Computer Software Processing Technology in Performance Information Management System,” 2021. doi: 10.1109/IWCMC51323.2021.9498597.

[3] R. Chavan, [1] H. Harb and L. Chen, “Voice-Based Gender Identification in Multimedia Applications,” J Intell Inf Syst, vol. 24, no. 2, pp. 179–198, 2005, doi: 10.1007/s10844-005-0322-8.

[2] I. H. Sarker, “AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems,” SN Comput Sci, vol. 3, no. 2, Mar. 2022, doi: 10.1007/s42979-022-01043-x.

[3] Y. Xu et al., “Artificial intelligence: A powerful paradigm for scientific research,” The Innovation, vol. 2, no. 4. Cell Press, Nov. 28, 2021. doi: 10.1016/j.xinn.2021.100179.

[4] L. Kuang, S. Pobbathi, Y. Mansury, M. A. Shapiro, and V. K. Gurbani, “Predicting age and gender from network telemetry: Implications for privacy and impact on policy,” PLoS One, vol. 17, no. 7 July, Jul. 2022, doi: 10.1371/journal.pone.0271714.

[5] B. S. Prasad, “GENDER CLASSIFICATION THROUGH VOICE AND PERFORMANCE ANALYSIS BY USING MACHINE LEARNING ALGORITHMS,” INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATIONS AND ROBOTICS www.ijrcar.com, vol. 7, pp. 1–11, 2019, [Online]. Available: www.ijrcar.com

[6] P. Rani and M. B. Yadav, “A Survey on Gender and Emotion Recognition Using Voice,” 2018.

[7] E. Rodero, “Intonation and Emotion: Influence of Pitch Levels and Contour Type on Creating Emotions,” J Voice, vol. 25, pp. e25-34, Dec. 2011, doi: 10.1016/j.jvoice.2010.02.002.

[8] E. Wei, “Intonation Characteristics of Singing Based on Artificial Intelligence Technology and Its Application in Song-on-Demand Scoring System,” Math Probl Eng, vol. 2021, pp. 1–11, Dec. 2021, doi: 10.1155/2021/5510401.

[9] S. Levitan, T. Mishra, and S. Bangalore, “Automatic identification of gender from speech,” Dec. 2016, pp. 84–88. doi: 10.21437/SpeechProsody.2016-18.

[10] M. Ali, M. Islam, and M. A. Hossain, “GENDER RECOGNITION SYSTEM USING SPEECH SIGNAL,” vol. Vol.2, Dec. 2012.

[11] H. Zimeng, M. Ylianttila, M. Liyanage, T. Advisor, and J. Okwuibe, “DEGREE PROGRAMME IN ELECTRICAL ENGINEERING or DEGREE PROGRAMME IN WIRELESS COMMUNICATIONS ENGINEERING SPEAKER GENDER RECOGNITION SYSTEM,” 2017.

[12] D. Kaushik, N. Jain, and A. Majumdar, “Gender Voice Recognition through speech analysis with higher accuracy,” Dec. 2014. doi: 10.13140/2.1.1331.5842.

[13] A. Raahul, R. Sapthagiri, K. Pankaj, and V. Vijayarajan, “Voice based gender classification using machine learning,” in IOP Conference Series: Materials Science and Engineering, Dec. 2017, vol. 263, no. 4. doi: 10.1088/1757-899X/263/4/042083.

[14] R. Ihaka and R. Gentleman, “R: A Language for Data Analysis and Graphics,” Journal of Computational and Graphical Statistics, vol. 5, no. 3, pp. 299–314, 1996, doi: 10.2307/1390807.

[15] J. Doi, “Web Application Teaching Tools for Statistics Using R and Shiny,” Technology Innovations in Statistics Education, vol. 9, pp. 1–32, Dec. 2016.

[16] J. Wojciechowski, A. Hopkins, and R. Upton, “Interactive Pharmacometric Applications Using R and the Shiny Package,” CPT Pharmacometrics Syst Pharmacol, vol. 4, pp. 146–159, Dec. 2015, doi: 10.1002/psp4.21.

[17] X. Wu et al., “Top 10 algorithms in data mining,” Knowl Inf Syst, vol. 14, no. 1, pp. 1–37, 2008, doi: 10.1007/s10115-007-0114-2.

[18] A. K. and S. Aithal, “Voice Biometric Systems for User Identification and Authentication – A Literature Review,” International Journal of Applied Engineering and Management Letters, pp. 198–209, Dec. 2022, doi: 10.47992/IJAEML.2581.7000.0131.

[19] Y. Taspinar, M. SARITAŞ, I. Cinar, and M. Koklu, “Gender Determination Using Voice Data,”

International Journal of Applied Mathematics Electronics and Computers, pp. 232–235, Dec. 2020, doi: 10.18100/ijamec.809476.
[20] T.-H. Li, “Quantile-Frequency Analysis and Deep Learning for Signal Classiication,” 2022, doi: 10.21203/rs.3.rs-1855496/v1.
[21] H. Elo€, C. Frouin-Mouy, and M. O. Hammill, “In-air and underwater sounds of hooded seals during the breeding season in the Gulf of St. Lawrence,” 2021, doi: 10.1121/10.
[22] Z. Y, “Application of Data Mining Techniques in the Analysis of Acoustic Sound Characteristics,” J Inf Technol Softw Eng, vol. 08, no. 03, 2018, doi: 10.4172/2165-7866.1000238.
[23] S. Kushwah, S. Singh, K. Vats, and V. Nemade, “Gender Identification Via Voice Analysis,” International Journal of Scientific Research in Computer Science, Engineering and Information Technology, pp. 746–753, Dec. 2019, doi: 10.32628/CSEIT1952188.
[24] S. Sanjay, T. B. Student, and T. B. Kute Researcher, “PERFORMANCE EVALUATION OF ALGORITHMS FOR GENDER CLASSIFICATION,” 2020. [Online]. Available: http://www.ijeast.com
[25] J. Kaur and J. Singh Gurm, “Optimizing the Accuracy of CART Algorithm by Using Genetic Algorithm,” International Journal of Computer Science Trends and Technology, vol. 3, 2013, [Online]. Available: www.ijcstjournal.org
[26] E. Ersoy, E. Albey, and E. Kayış, “A CART-based genetic algorithm for constructing higher accuracy decision trees,” in DATA 2020 - Proceedings of the 9th International Conference on Data Science, Technology and Applications, 2020, pp. 328–338. doi: 10.5220/0009893903280338.
[27] M. Araya-Salas and G. Smith-Vidaurre, “warbleR: an r package to streamline analysis of animal acoustic signals,” Methods Ecol Evol, vol. 8, no. 2, pp. 184–191, Feb. 2017, doi: 10.1111/2041-210X.12624.
[28] lakhan jasuja and akhtar rasool, “Voice Gender Recognizer RECOGNITION OF GENDER FROM VOICE USING DEEP NEURAL NETWORKS,” Proceedings, International Conference on Smart Electronics and Communication (ICOSEC 2020) : 10-12, September 2020, 2020.
[29] “Analysis of Fashion Industry Business Environment,” *Latest Trends Text. Fash. Des.*, vol. 2, no. 4, 2018, doi: 10.32474/lttfd.2018.02.000144.