

Javanese Gender Speech Recognition Based on Machine Learning Using Random Forest and Neural Network

Kristiawan Nugroho

AMIK Jakarta Teknologi Cipta

kristiawan1979@gmail.com

Abstract — Speech is a means of communication between people throughout the world. At present research in the field of speech recognition continues to develop in producing a robust method in various research variants. However decreasing the word error rate or reducing noise is still a problem that is still being investigated until now. The purpose of this study is to find the right method with high accuracy to classify the gender voices of Javanese. This research used a human voice dataset of both men and women from the Javanese tribe which was recorded and then processed using a noise reduction preprocessing technique with the MFCC extraction feature method and then classified using 2 machine learning methods, namely Random Forest and Neural Network. Evaluation results indicate that the classification of Javanese accent speech accents results in an accuracy rate of 91.3 % using Random Forest and 92.2% using Neural Network..

Keywords:Speech, Random Forest, Neural Network, Accuration

I. INTRODUCTION

Speech recognition technology is increasingly developing following the development of science and technology. Some technology giant companies have started their research in speech recognition, including Microsoft, Apple and Google by producing various forms of applications such as Google Now, Siri, or Cortana virtual assistants. These various technologies are also still developing as more and more researches are carried out in the field of recognition. sound. The history of research in

the field of speech recognition has begun since 1952 researchers from Bell Labs built a system by the name of Audrey to recognize one-digit speakers. In the 1970s DARPA funded Speech Understanding Research which is speech recognition research that functions to find vocabulary size. A pretty phenomenal speech recognition product was produced by Google in the 2000s in the form of Google Voice Search which is supported by around 30 languages around the world.

Researchers around the world are still trying to build various methods and algorithms that are robust and have high accuracy in speech recognition. Some research on speech recognition includes Speech recognition with artificial neural networks with the method of voice recognition with Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques [1], Voice recognition using Hidden Markov Mode [2] which results in accuracy up to 86.67%, Research speech recognition by combining the Artificial Neural Network method with Hidden Markov Model [3], Hindi voice recognition with Hidden Markov Model [4], Voice recognition for the biometric field with the Vector Quantization method [5]. Research on voice recognition in the groundwater was also carried out using Mel-Frequency Cepstrum Coefficients (MFCC) and Adaptive Neuro-Fuzzy Inferense System (ANFIS) resulting in an accuracy rate of 95.90% [6]. Markov Models with different letter tests produce an accuracy of 54.6% [7].

Various studies on speech recognition as has been done above still do not really produce the best accuracy in speech recognition, researchers are still trying to overcome various problems that often occur in speech recognition including how to

reduce noise and reduce high data dimensions. This paper discusses the recognition of Javanese gender speech using machine learning. Research begins with sound recording and then the sound results are processed using the Adobe Audition application which is then performed feature extraction. The results of feature extraction will be evaluated by measuring the level of accuracy in its classification so that the results of speech recognition accuracy can be compared using 2 methods, namely Random Forest and Neural Network.

II. METHOD

A. Dataset

Research on the javanese speech recognition in this paper uses a private dataset by recording the words "eating", "drinking" and "sleeping", each of which is spoken 10 times by 5 men and 5 women. Here are the sound recording settings in Adobe Audition:

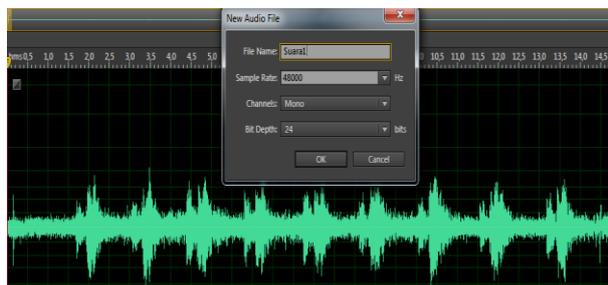


Figure 1. Speech recording settings

Then after recording, each sound is cut in the same duration of time that is 80631 to then be stored in respective folders.

B. Preprocessing

Preprocessing is a stage in processing the raw dataset, in this case the sound dataset to be cleaned of noise disturbances by reducing noise by utilizing the noise reduction feature in Adobe Audition.

C. Feature Extraction

Speech recognition dataset from male and female gender from the Javanese tribe is still in the form of a wav sound file which after being framing will then be processed with a feature extraction deungan to be a form of data that is ready for classification. In

performing the dataset extraction feature, this paper uses Matlab software using the MFCC (Mel Frequency Cepstral Coefficients) method, the MFCC is a method used to contract the unique sounds of humans [8]. Using the MFCC function, each dataset is extracted to produce a unique feature for each sound to produce 150 wav file data record records for later labeling as a requirement for the classification process

D. Classification

1. Random Forest Method

Random forest is one of the methods of machine learning that is widely used by researchers. This method is one way in machine learning to manage large amounts of data. Random forest for the first time was introduced in a paper by Leo Breiman, this paper contains a model for constructing uncorrelated tree forests using procedures such as CART (Classification And Regression Trees), which can be hybridized by optimizing random nodes and bagging [9]. Random Forest is a method that produces a high level of accuracy, but this method requires a high time compared to other methods [10]. However, the Random Forest method also has advantages over other methods, which are suitable for classifying high-dimensional data [11].

2. Neural Network Method

Neural Network is a method implemented in machine learning that is implemented like an imitation of the human brain [12]. Some of the features and advantages of neural networks include [13]:

a. Adaptive learning

Neural Network takes a copy of the human brain along with the ability to learn and adjust the concept of work while learning.

b. Parallel operation

The concept in neural networks works in parallel which can adjust as in the human brain.

c. Classification and recognition

Neural networks can also be used in pattern recognition, data classification and other applications that have unclear data.

d. More fast

Neural networks also have the advantage of faster processing when compared to the human brain.

III. RESULTS AND DISCUSSION

Research on the Javanese speech recognition uses several stages which include:

A. Orange Application

Orange is an opensource application that can be used in machine learning in data mining, analysis and data visualization activities [14]. The Orange application was made by scientists at Ljubljana University using the Python, Cython, C++ and C programming languages [15].

B. Design

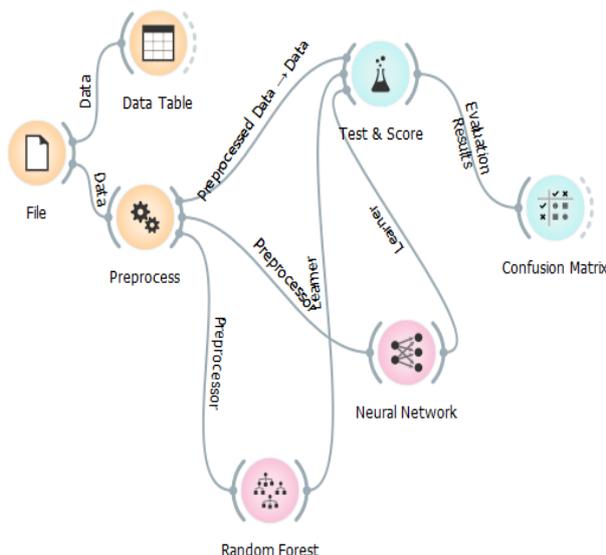


Figure 2. Orange Design

Figure 2 above shows the Javanese gender voice dataset file which is preprocessed later by using the neural newtwork and random forest methods the results of the two methods are evaluated and the results are also displayed in the confusion matrix.

C. Evaluation

Method	AUC	CA	F1	Precision	Recall
Random Forest	0.974	0.913	0.913	0.913	0.913
Neural Network	0.987	0.922	0.922	0.922	0.922

Figure 3. Orange Test Score

The measurement results of 2 methods of neural network and random forest can be seen in Figure 2 above where the random forest method produces an accuracy rate (CA) of 97.4%, F1 measure, precision and recall of 91.3% while the neural network method produces an accuracy level (CA) of 98.7% , F1 measure, precision and recall at 92.2%.

D. Confusion Matrix

1. Neural Network

		Predicted		Σ
		Pria	Wanita	
Actual	Pria	157	23	180
	Wanita	12	258	270
Σ		169	281	450

Figure 4. NN Confusion Matrix

Confusion matrix with the neural network method in Figure 4 above shows the level of accuracy that can be calculated with the formula:

$$\frac{TP+TN}{TP+TN+FP+FN} \times 100\% \dots\dots\dots(1)$$

- TP=True Positive
- TN=True Negative
- FP=False Positive
- FN=False Negative

$$\frac{157 + 258}{157 + 258 + 23 + 12} \times 100\%$$

The result is 92.2%

2. Random Forest

		Predicted		Σ
		Pria	Wanita	
Actual	Pria	161	19	180
	Wanita	20	250	270
Σ		181	269	450

Figure 5. RF Confusion Matrix

Confusion matrix with random forest in Figure 5 above shows the level of accuracy that can be calculated with the formula:

$$\frac{TP+TN}{TP+TN+FP+FN} \times 100\% \dots\dots\dots(2)$$

TP=True Positive
 TN=True Negative
 FP=False Positive
 FN=False Negative

$$\frac{161 + 250}{161 + 250 + 19 + 20} \times 100\%$$

The result is 91.3%

IV. CONCLUSION

Research on the Javanese gender speech recognition has been carried out, the dataset used is a private dataset where the recording process is carried out using a voice recorder that is processed using the Adobe Audition application. The next dataset processing is to do feature extraction using MFCC (Mel-Frequency Cepstrum Coefficients) technique then labeling process. After the dataset is ready to use the dataset is processed and then classified using 2 methods, namely random forest and neural network. The evaluation results show that the neural network method achieved the highest level of accuracy, namely 92.2%, while the random forest method obtained an accuracy of 91.3%.

REFERENCES

- [1] L. Muda, M. Begam, dan I. Elamvazuthi, "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques," vol. 2, no. 3, hlm. 6, 2010.
- [2] Shumaila Iqbal, T. Mehboob, dan M. Sikander Hayat Khiyal, "Voice Recognition using HMM with MFCC for Secure ATM.," 2011.
- [3] M. Frikha dan A. Ben Hamida, "A Comparative Survey of ANN and Hybrid HMM/ANN Architectures for Robust Speech Recognition," *AJIS*, vol. 2, no. 1, hlm. 1–8, Agu 2012.
- [4] V. Mulik, V. Mane, dan I. Jamadar, "Hidden Markov Model Based Robust Speech Recognition," *International Journal of Innovative Research in Advanced Engineering*, vol. 2, no. 2, hlm. 10, 2015.
- [5] Sreelakshmi, "Design of an Intelligent Speaker Recognition System using Mel Frequency Cepstrum Coefficients and Vector Quantization for Biometric Authentication," 2015.
- [6] Z. S. Mada Sanjaya W.S, "Implementasi Pengenalan Pola Suara Menggunakan Mel-Frequency Cepstrum Coefficients (Mfcc) Dan Adaptive Neuro-Fuzzy Inferense System (Anfis) Sebagai Kontrol Lampu Otomatis," 2014.
- [7] Q. Nada, C. Ridhuandi, P. Santoso, dan D. Apriyanto, "Speech Recognition dengan Hidden Markov Model untuk Pengenalan dan Pelafalan Huruf Hijaiyah," vol. 5, no. 1, hlm. 8, 2019.
- [8] A. H.Mansour, G. Zen Alabdeen Salh, dan K. A. Mohammed, "Voice Recognition using Dynamic Time Warping and Mel-Frequency Cepstral Coefficients Algorithms," *IJCA*, vol. 116, no. 2, hlm. 34–41, Apr 2015.
- [9] Leo Breiman, "Random Forests," 2001.

- [10] Computer Science & Engineering &GZSCCET Bhatinda, Punjab, India, E. Goel, Er. Abhilasha, dan Computer Science & Engineering &GZSCCET Bhatinda, Punjab, India, "Random Forest: A Review," *IJARCSSE*, vol. 7, no. 1, hlm. 251–257, Jan 2017.
- [11] B. Xu, J. Z. Huang, G. Williams, Q. Wang, dan Y. Ye, "Classifying Very High-Dimensional Data with Random Forests Built from Small Subspaces:," *International Journal of Data Warehousing and Mining*, vol. 8, no. 2, hlm. 44–63, Apr 2012.
- [12] H. Kukreja, "An Introduction To Artificial Neural Network," vol. 1, no. 5, hlm. 5, 2016.
- [13] O. S. Eluyode dan D. T. Akomolafe, "Comparative study of biological and artificial neural networks," hlm. 11, 2013.
- [14] S. Kodati dan D. R. Vivekanandam, "Analysis of Heart Disease using in Data Mining Tools Orange and Weka," hlm. 7, 2018.
- [15] P. K. Pattnaik, A. Swetapadma, dan J. Sarraf, Ed., *Expert System Techniques in Biomedical Science Practice*: IGI Global, 2018.