Identification Of Speaker Recognition For Audio Forensic Using K-Nearest Neighbor

Rusydi Umar, Imam Riadi, Abdullah Hanif, Siti Helmiyah

Abstract: Voice is a part of human that has unique characteristics and can be distinguished from one person to another so that it can be used for loudspeakers as needed to use biometric sounds in the process of logging into applications to be developed. The process of appreciating in this study uses the K-Nearest Neighbor (K-NN) method, where the collection process is carried out using feature extraction data that has been previously obtained using MATLAB R2013a software. The voice data used for speaking speakers consists of 5 speakers with 3 men and 2 women, each speaker says the same word, namely login. The amount of data is 75 voice data taken from 5 speakers, where 50 data are training data, and 25 data are test data. This research has the result of completion of 4 speakers by 40%, and 1 speaker by 20%.

Index Terms: Biometric, Feature Extraction, Identification, K-Nearest Neighbor, Speaker, Speaker Recognition, Voice.

1. INTRODUCTION

VOICE is one way to know and recognize one's character. Human can recognize someone through his voice, for example the speaker's identity, speech style, accent, emotions and health conditions of the speaker. The rapid development of technology has led to the existence of new technologies, namely the processing of human voice signals that will be developed into the speaker recognition. Speaker recognition is the process of identifying a person based on the characteristics of the sound spoken [1]. Speech processing is a diverse field with many applications, such as the introduction of speakers including verification of speakers and identification of speakers [2]. The speech processing architecture can be seen in Fig. 1.

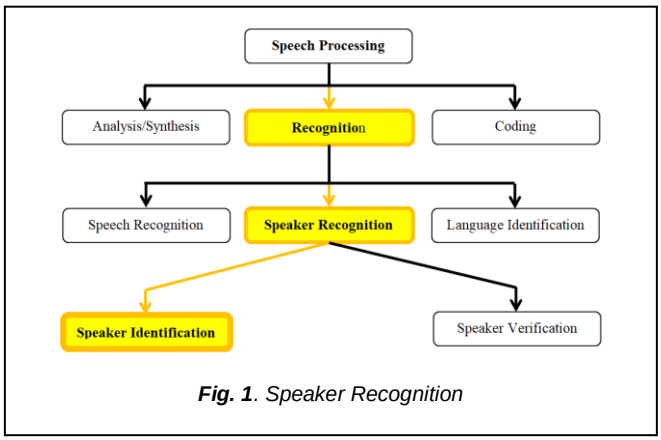

Fig. 1. *Speaker Recognition*

Fig. 1 shows the area of speech processing and the introduction of speakers that are interconnected with other fields; this research focuses on the identification of speaker recognition.

The application of speaker recognition can be used for the biometric field, several biometric categories using characteristics that can be used as indicators to recognize someone like: face, fingerprint, iris, sound etc. Sounds that are used as biometrics have several advantages over other biometrics, namely: login remote is possible, easy to implement and does not require special hardware [3]. Apart from these advantages, biometric sounds still can't be widely implemented because of the many problems, such as: security, intonation, surrounding environmental noise, changes in human voice itself, etc [4].

2 RELATED WORK

2.1 Speaker Recognition

Many methods are used to identify speakers, such as the Mel Frequency Cepstral Coefficient method, MFCC is a method used in the introduction of speakers to describe the characteristics of a signal, which has a relationship to the characteristics of the speaker vocal channel so that it can distinguish between one speaker and speaker others [5]. Praveen and Thomas in [6], obtain results with the same method, using FAR, FRR and EER parameters with attainment rates of 96.18% for cluster 5 size in each coefficientThe research in [7] discussed about speaker recognition using MFCC-VQ and MFCC-GMM for hindi speech samples, the accuracy of text independent recognition is 77.64% and 86.27% respectively. However, for text dependent recognition , the accuracy has increased significantly. In addition of the MFCC method, other methods used in text independent speaker recognition are i-vector methods [8-13]. The research in [14] discussing speaker recognition with hybrid features, the experimental results show that using hybrid features can improve performance in the speaker classification process. In this research, the classification method used is K-Nearest Neighbor (KNN) to identify speakers by using feature extraction data that has been obtained previously.

2.2 Methodology

The method used in this study is the prototype method. Prototype method is the initial form of system design that is used to describe concepts, design experiments, find more problems and provide solutions [15]. With listen to customer, build/revise mock-up, and customer test drives mock-up, as described in Figure 2[16].

- *Rusydi Umar. Department of Informatics. Universitas Ahmad Dahlan, Indonesia, E-mail: rusydi_umar@rocketmail.com*
- *Imam Riadi. Department of Information System. Universitas Ahmad Dahlan, Indonesia, E-mail: imam.riadi@is.uad.ac.id*
- *Abdullah Hanif. Department of Informatics. Universitas Ahmad Dahlan, Indonesia, E-mail: abdullah1708048026@webmail.uad.ac.id*
- *Siti Helmiyah Department of Informatics. Universitas Ahmad Dahlan, Indonesia, E-mail: siti1708048022@webmail.uad.ac.id*

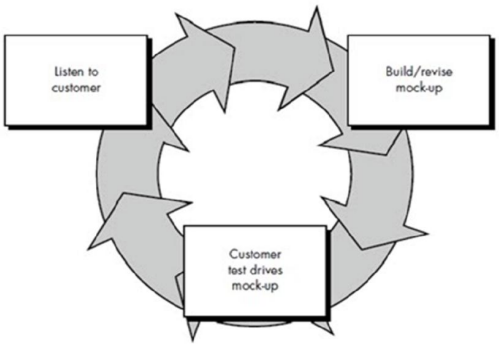

***Fig. 2**. Prototype Method*

The following is an explanation of the prototype method in Figure 2 is: Listen to customer: This first step in this research process is **doing a needs analysis and define needs like as voice sample** data used to identify of speaker recognition. Build or revise mockup: After the data is obtained, the next step is to compile the sound sample data so that it can be analyzed and tested using the K-NN method. Customer test drives mockup: After the voice sample data is compiled, the last step in this prototype method is testing new sample data in order to succeed in identifying who is speaking.

3 K-NEAREST NEIGHBOR

K-Nearest Neighbor (KNN) is a method used to classify objects based on learning data that has the closest distance to the object, this method aims to classify recently entered objects based on training data samples [17]. This method is like a clustering technique, which is a technique for grouping new data that is input based on the distance of new data to some of the closest data [18]. In this research, the object to be classified used K-NN method is the voice sample data obtained from each speaker.

The classification process of the K-NN method can be seen in Figure 3.

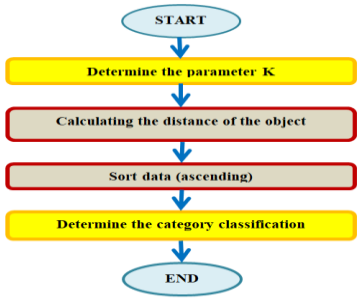

***Fig. 3**. K-NN Method*

The following is an explanation of the classification process: Determine the parameter K (number of closest neighbors). Calculating the distance of the object to the training data provided. Sort data that has the closest distanceDetermine the category of trial data obtained from the number of closest neighbors. The equation for calculating the object's distance from the training of the data provided, which is:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (1)$$

(x1, y1) is the variable value of the new object that is initiated, while (x2, y2) is the variable value of each neighbor in 2 variable.

In this paper there are more than 2 variables to calculate the distance using the Euclidean Distance formula [19], shown in equation (3):

$$d = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_n - y_n)^2}$$

(2)

$$d = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

Equation (2) shows, d = Euclidean Distance, x = The variable value of each neighbor, y = The variable value of new object, i = Variables to, n = Number of variables

4 RESULT AND ANALYSIS

Voice samples used in this study were recorded and extracted features using MATLAB R2013a software which had been obtained in previous studies. Table 1 shows a description of the sample data carried out by 5 speakers. Ten per-speaker voice sample data was taken.

TABLE 1
SAMPLE DATA DESCRIPTION

The Name of The Speaker	Speech	Gender	Repeated Time
Speaker 1	login	man	10 times
Speaker 2	login	woman	10 times
Speaker 3	login	man	10 times
Speaker 4	login	woman	10 times
Speaker 5	login	man	10 times

The feature extraction process using MATLAB R2013a succeeded in getting the characteristics of each voice sample spoken by each speaker, but because the results of the characteristics of each speaker still have a value that is not in the same scope, therefore data normalization is done using Z-ScoreWhere each voice sample has 13 features represented by C1-C13, feature extraction of speaker 1 can be seen in Figure 4.

No.	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
1	3.280	-0.462	-0.285	-0.424	-0.392	-0.079	-0.587	-0.180	-0.079	-0.282	-0.004	-0.157	-0.349
2	3.235	-0.247	-0.627	-0.325	-0.288	-0.078	0.144	0.009	-0.553	-0.542	-0.212	-0.467	-0.050
3	3.288	-0.536	-0.196	-0.228	-0.436	-0.439	-0.245	-0.278	-0.125	-0.158	-0.419	-0.266	0.038
4	3.262	-0.297	-0.219	-0.215	-0.478	-0.386	-0.341	0.279	-0.237	-0.404	-0.262	-0.159	-0.542
5	3.259	-0.522	-0.418	-0.074	-0.224	-0.505	-0.266	-0.366	-0.395	-0.353	0.127	-0.340	0.075
6	3.249	-0.510	-0.084	-0.330	-0.410	-0.215	-0.695	-0.311	0.180	-0.092	-0.225	-0.389	-0.167
7	3.265	-0.600	-0.466	-0.498	-0.418	-0.150	-0.048	-0.187	-0.437	0.001	-0.251	-0.147	-0.063
8	3.278	-0.397	-0.144	-0.521	-0.226	-0.237	-0.599	-0.148	-0.343	-0.064	-0.342	0.016	-0.274
9	3.290	-0.360	-0.185	-0.003	-0.415	-0.368	-0.309	-0.213	-0.150	-0.498	-0.484	-0.082	-0.223
10	3.277	-0.611	-0.133	-0.368	-0.562	-0.150	-0.129	-0.339	-0.231	-0.289	0.037	-0.209	-0.294

***Fig. 4** Speaker 1.*

Feature extraction of speaker 2 can be seen in Figure 5.

No.	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
1	3.280	-0.462	-0.285	-0.424	-0.392	-0.079	-0.587	-0.180	-0.079	-0.282	-0.004	-0.157	-0.349
2	3.235	-0.247	-0.627	-0.325	-0.288	-0.078	0.144	0.009	-0.553	-0.542	-0.212	-0.467	-0.050
3	3.288	-0.536	-0.196	-0.228	-0.436	-0.439	-0.245	-0.278	-0.125	-0.158	-0.419	-0.266	0.038
4	3.262	-0.297	-0.219	-0.215	-0.478	-0.386	-0.341	0.279	-0.237	-0.404	-0.262	-0.159	-0.542
5	3.259	-0.522	-0.418	-0.074	-0.224	-0.505	-0.266	-0.366	-0.395	-0.353	0.127	-0.340	0.075
6	3.249	-0.510	-0.084	-0.330	-0.410	-0.215	-0.695	-0.311	0.180	-0.092	-0.225	-0.389	-0.167
7	3.265	-0.600	-0.466	-0.498	-0.418	-0.150	-0.048	-0.187	-0.437	0.001	-0.251	-0.147	-0.063
8	3.278	-0.397	-0.144	-0.521	-0.226	-0.237	-0.599	-0.148	-0.343	-0.064	-0.342	0.016	-0.274
9	3.290	-0.360	-0.185	-0.003	-0.415	-0.368	-0.309	-0.213	-0.150	-0.498	-0.484	-0.082	-0.223
10	3.277	-0.611	-0.133	-0.368	-0.562	-0.150	-0.129	-0.339	-0.231	-0.289	0.037	-0.209	-0.294

***Fig. 5**. Speaker 2.*

Feature extraction of speaker 3 can be seen in Figure 6.

No.	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
1	3.280	-0.462	-0.285	-0.424	-0.392	-0.079	-0.587	-0.180	-0.079	-0.282	-0.004	-0.157	-0.349
2	3.235	-0.247	-0.627	-0.325	-0.288	-0.078	0.144	0.009	-0.553	-0.542	-0.212	-0.467	-0.050
3	3.288	-0.536	-0.196	-0.228	-0.436	-0.439	-0.245	-0.278	-0.125	-0.158	-0.419	-0.266	0.038
4	3.262	-0.297	-0.219	-0.215	-0.478	-0.386	-0.341	0.279	-0.237	-0.404	-0.262	-0.159	-0.542
5	3.259	-0.522	-0.418	-0.074	-0.224	-0.505	-0.266	-0.366	-0.395	-0.353	0.127	-0.340	0.075
6	3.249	-0.510	-0.084	-0.330	-0.410	-0.215	-0.695	-0.311	0.180	-0.092	-0.225	-0.389	-0.167
7	3.265	-0.600	-0.466	-0.498	-0.418	-0.150	-0.048	-0.187	-0.437	0.001	-0.251	-0.147	-0.063
8	3.278	-0.397	-0.144	-0.521	-0.226	-0.237	-0.599	-0.148	-0.343	-0.064	-0.342	0.016	-0.274
9	3.290	-0.360	-0.185	-0.003	-0.415	-0.368	-0.309	-0.213	-0.150	-0.498	-0.484	-0.082	-0.223
10	3.277	-0.611	-0.133	-0.368	-0.562	-0.150	-0.129	-0.339	-0.231	-0.289	0.037	-0.209	-0.294

***Fig. 6**. Speaker 3.*

TABLE 2
ALL RESULT ECLUDIAN OF 5 SPEAKER

Voice Sample	Speaker 1	Speaker 2	Speaker 3	Speaker 4	Speaker 5
1	0.816	0.862	0.904	0.961	1.135
2	0.835	1.137	1.199	0.685	0.773
3	1.013	1.203	0.965	1.172	1.076
4	0.686	1.043	1.234	1.261	0.982
5	0.924	0.898	1.240	0.718	1.026
6	1.155	0.910	0.827	0.995	0.998
7	0.863	1.090	0.935	1.241	0.707
8	1.076	0.815	0.831	1.147	1.079
9	1.000	1.076	1.188	1.299	1.040
10	0.733	0.765	0.946	1.034	1.402

Feature extraction of speaker 4 can be seen in Figure 7.

No.	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
1	3.270	-0.414	-0.455	-0.576	-0.304	0.101	-0.050	-0.318	-0.221	-0.113	-0.375	-0.416	-0.130
2	3.284	-0.582	-0.329	-0.338	-0.318	-0.511	-0.047	-0.327	-0.080	-0.153	-0.059	-0.275	-0.266
3	3.231	-0.418	-0.199	-0.604	-0.137	-0.436	-0.432	-0.130	-0.738	0.075	-0.130	-0.063	-0.020
4	3.220	-0.492	-0.291	-0.372	-0.062	-0.273	-0.363	0.092	-0.175	-0.521	-0.794	0.105	-0.074
5	3.276	-0.177	-0.514	-0.484	-0.038	-0.374	-0.244	-0.127	-0.121	-0.196	-0.151	-0.225	-0.626
6	3.209	-0.544	-0.193	-0.481	-0.610	-0.188	-0.291	-0.276	0.208	0.200	-0.054	-0.547	-0.433
7	3.264	-0.181	-0.232	-0.483	-0.305	-0.043	-0.440	-0.314	0.081	-0.095	-0.678	-0.322	-0.251
8	3.245	-0.370	-0.492	-0.171	-0.019	-0.539	-0.365	-0.448	-0.434	0.260	-0.156	-0.371	-0.140
9	3.190	-0.377	-0.183	0.006	0.033	0.283	-0.040	-0.663	-0.592	-0.613	-0.201	-0.447	-0.396
10	3.248	-0.586	-0.236	-0.361	-0.397	-0.321	-0.120	0.008	-0.448	0.112	-0.607	-0.038	-0.254

***Fig. 7** Speaker 4.*

No.	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
1	3.270	-0.414	-0.455	-0.576	-0.304	0.101	-0.050	-0.318	-0.221	-0.113	-0.375	-0.416	-0.130
2	3.284	-0.582	-0.329	-0.338	-0.318	-0.511	-0.047	-0.327	-0.080	-0.153	-0.059	-0.275	-0.266
3	3.231	-0.418	-0.199	-0.604	-0.137	-0.436	-0.432	-0.130	-0.738	0.075	-0.130	-0.063	-0.020
4	3.220	-0.492	-0.291	-0.372	-0.062	-0.273	-0.363	0.092	-0.175	-0.521	-0.794	0.105	-0.074
5	3.276	-0.177	-0.514	-0.484	-0.038	-0.374	-0.244	-0.127	-0.121	-0.196	-0.151	-0.225	-0.626
6	3.209	-0.544	-0.193	-0.481	-0.610	-0.188	-0.291	-0.276	0.208	0.200	-0.054	-0.547	-0.433
7	3.264	-0.181	-0.232	-0.483	-0.305	-0.043	-0.440	-0.314	0.081	-0.095	-0.678	-0.322	-0.251
8	3.245	-0.370	-0.492	-0.171	-0.019	-0.539	-0.365	-0.448	-0.434	0.260	-0.156	-0.371	-0.140
9	3.190	-0.377	-0.183	0.006	0.033	0.283	-0.040	-0.663	-0.592	-0.613	-0.201	-0.447	-0.396
10	3.248	-0.586	-0.236	-0.361	-0.397	-0.321	-0.120	0.008	-0.448	0.112	-0.607	-0.038	-0.254

***Fig. 8** Speaker 5.*

Next, feature extraction data will be classified using the K-NN method, following the steps of the classification process: Determine the Parameter KIn this research, the parameter K used is 4, K = Number of closest neighbors. Calculated the Distance of the Object to the Training Data ProvidedBefore performing calculations, new test data to be classified can be seen in Figure 9.

3.245	-0.512	-0.504	-0.297	-0.381	0.299	-0.011	-0.126	-0.231	-0.333	-0.113	-0.364	-0.553

***Fig. 9**. New sample of Speaker 4.*

After the new test data is entered, then it is calculating the closest distance from the new data with the training data obtained from each speaker. Calculation of the closest distance from the new test data to the training data is carried out up to 5 speakers, all the closest distance data from 5 speakers can be seen in Table 2. Sort Data that has the Closest Distance (Ascending) The next step is to sort the data calculated by ascending distance from all the speakers, the results of sorting can be seen in Table 3.

Determine the Category Classification.

The final step of the K-NN classification process is to categorize the ordered data based on the number of closest neighbors that have been determined in the first step, where K = 4, then select 4 data that has the smallest distance as shown in Table 4.

TABLE 4
DETERMINE 4 DATA WITH SMALLEST DISTANCE

The Name of The Speaker	Gender	Ecludian Distance	Smallest Distance Ranking
Speaker 4	Woman	0.685	1
Speaker 1	Man	0.686	2
Speaker 5	Man	0.707	3
Speaker 4	Woman	0.718	4

By sorting the smallest distance from the new test data on the training data, because the priority comparison of the speaker samples is 2 data (Speaker 4) > 1 data (Speaker 1 and Speaker 5), it can be concluded that the speaker test data can be identified as Speaker 4. The K-NN concept based on the results of this research can be seen in Figure 10.

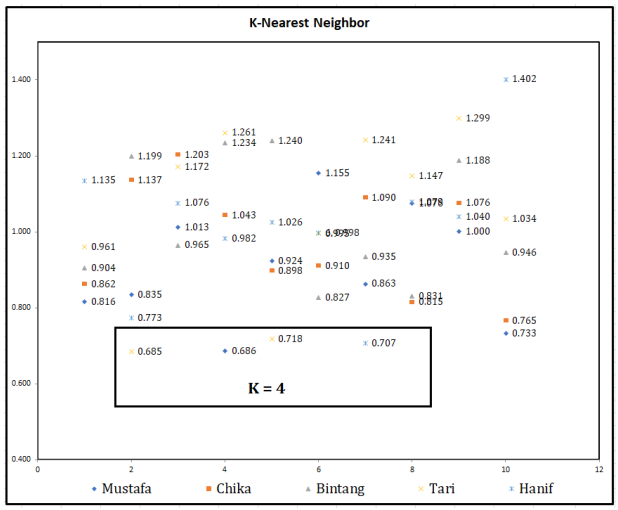

***Fig. 10**. The Concept of Result K-NN*

TABLE 3
20 DATA RESULT ECLUDIAN OF 5 SPEAKER AFTER SORTING

The Name of The Speaker	Gender	Ecludian Distance	Smallest Distance Ranking
Speaker 4	Woman	0.685	1
Speaker 1	Man	0.686	2
Speaker 5	Man	0.707	3
Speaker 4	Woman	0.718	4
Speaker 1	Man	0.733	5
Speaker 2	Woman	0.765	6
Speaker 5	Man	0.773	7
Speaker 2	Woman	0.815	8
Speaker 1	Man	0.816	9
Speaker 3	Man	0.827	10
Speaker 3	Man	0.831	11
Speaker 1	Man	0.835	12
Speaker 2	Woman	0.862	13
Speaker 1	Man	0.863	14
Speaker 2	Woman	0.898	15
Speaker 3	Man	0.904	16
Speaker 2	Woman	0.910	17
Speaker 1	Man	0.924	18
Speaker 3	Man	0.935	19
Speaker 3	Man	0.946	20

The following are the results of testing of all testing data on training data, as can be seen in Figure 11.

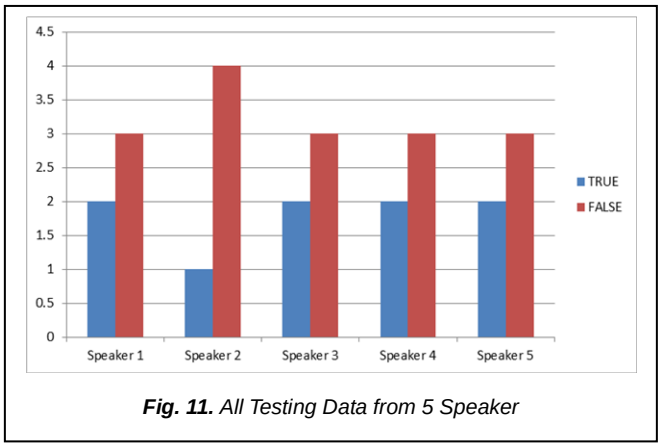

***Fig. 11.** All Testing Data from 5 Speaker*

The analysis results based on Figure 11, obtained the accuracy of the results of the data testing of each speaker are as follows:

$$\text{Speaker 1,3,4,5} = \frac{2}{5} x100\ \% = 40\ \%$$

$$\text{Speaker 2} = \frac{1}{5} x100\ \% = 20\ \%$$

The calculation results of the accuracy of Speaker 1, Speaker 3, Speaker 4 and Speaker 5 are 40% and Speaker 2 is 20%. This happens because of the sound recording process, the speed of sound spoken, the surrounding noise level, and the clarity of the speaker in pronouncing a predetermined word.

5 CONCLUSION AND FUTURE WORKS

Based on the research that has been done, the K-NN classification method successfully identifies the speaker test data on the training data based on the spoken word. Subsequent research is expected to use a far greater amount of data testing, and in the process of sound recording or data retrieval taking into account the conditions of sound collection, such as noise level, spoken sound speed and spoken sound clarity, so that research results can be found better. In the future, the purpose of this research is to create an android-based voice assistant system, so that it is not only to identify who is speaking, but what application users are talking about and can be analyzed for audio forensics.

ACKNOWLEDGMENT

This research is supported by Direktorat Riset dan Pengabdian Masyarakat, Direktorat Jenderal Penguatan Riset dan Pengembangan Kementerian Riset, Teknologi, dan Pendidikan Tinggi Republik Indonesia.

REFERENCES

[1] J.S. Bridle, “Probabilistic Interpretation of Feedforward Classification Network A. Poddar, M. Sahidullah, and G. Saha, “Speaker Verification with short utterances: a review of challenges, trend and opportunities”, IET Biom, Vol. 7, Iss. 2, pp. 91-101, November 2017.

[2] Joseph P. Campbell, Jr., Senior Member, IEEE, “Speaker Recognition: A Tutorial”, Proceedings of the IEEE, vol. 85, no. 9, pp. 1437-1462, September 1997.

[3] Lisa Myers, An Exploration of Voice Biometrics, GSEC Practical Assignment version 1.4b Option 1, 2004
[4] Dr. H B Kekre, Vaishali Kulkarni, “Speaker Identification using Frequency Distribution in the Transform Domain”, (IJACSA) International Journal of Advanced Computer Sciense and Applications, Vol. 3, No.2, 2012.
[5] K. Dash, D. Padhi, B. Panda, and S. Mohanty, “Speaker identification using Melf Frequency Cepstral Coefficient and BPNN”, International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 2, Iss. 4, April 2012.
[6] N. Praveen amd T. Thomas, “Text Dependent Speaker Recognition using MFCC features and BPANN”, International Journal of Computer Applications, Vol. 74, No. 5, July 2013.
[7] Speaker recognition for hindi speech signal using MFCC-GMM approach -Ankur Maurya, Divya Kumar, R.K. Agarwal, 6th International Conference on Smart Computing and Communications (ICSCC),(2017).
[8] D. Snyder, D. Garcia-Romero, G. Sell, D. Povey, and S. Khudanpur, “X-vectors: Robust DNN Embeddings for Speaker Recognition,” in Proceedings of ICASSP, 2018.
[9] E. Variani, X. Lei, E. McDermott, I. L. Moreno, and J. GonzalezDominguez, “Deep neural networks for small footprint textdependent speaker verification,” in 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2014, pp. 4052–4056.
[10] G. Bhattacharya, J. Alam, and P. Kenny, “Deep Speaker Embeddings for Short-Duration Speaker Verification,” in Interspeech 2017, 08 2017, pp. 1517–1521.
[11] G. Heigold, I. Moreno, S. Bengio, and N. Shazeer, “End-toend text-dependent speaker verification,” in 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), March 2016, pp. 5115–5119.
[12] D. Snyder, D. Garcia-Romero, D. Povey, and S. Khudanpur, “Deep Neural Network Embeddings for Text-Independent Speaker Verification,” in Interspeech 2017, Aug 2017.
[13] S. X. Zhang, Z. Chen, Y. Zhao, J. Li, and Y. Gong, “End-toEnd attention based text-dependent speaker verification,” in 2016 IEEE Spoken Language Technology Workshop (SLT), Dec 2016, pp. 171–178.
[14] Ali, H., Tran, S.N., Benetos, E. et al. Neural Comput & Applic (2018) 29: 13.
[15] I. Sommervile, Software Engineering. Ninth Edition, Massachusetts: Addison-Wesley, 2011.
[16] Khosrow-Pour, Encyclopedia of Information Science and Technology (5 Volumes), Idea Group Reference, 2005.
[17] F. Gorunescu, Data Mining: Concepts , Models and Techniques, Springer, Berlin Heidelberg, 2011.
[18] Kursini, Luthfi, E. T, Algoritma Data Mining, Andi Offset, Yogyakarta, 2009.
[19] D. Sinwar and R. Kaushik, “Study of Euclidean and Manhattan Distance Distance Metrics Using Simple K-Means Clustering, (IJRASET) International Journal for Research in Applied Science and Engineering Technology, vol. 2, no.9, 2013.