DOI: 10.14421/ijid.2021.2354 IJID (International Journal on Informatics for Development), *e-ISSN: 2549-7448*
Vol. 10, No. 1, 2021, Pp.15-22

Song Recommendation Application Using Speech Emotion Recognition

Eko Budi Setiawan[1], Al Ghani Iqbal Dzulfiqar[2]
[1,2] Informatics Engineering, Universitas Komputer Indonesia, Bandung, Indonesia
[1]eko@email.unikom.ac.id, [2]alganiiqbal@gmail.com

Article History
Received Feb 9th, 2021
Revised Mar 7th, 2021
Accepted Mar 13th, 2021
Published June, 2021

***Abstract*— This research was conducted to facilitate the interaction between radio broadcasters and radio listeners during the song request process. This research was triggered by the difficulty of the broadcasters in monitoring song requests from listeners. The system is made to accommodate all song requests by listeners. The application produced in this study uses speech emotion recognition technology based on a person's mood obtained from the spoken words. This technology can change the voice into one of the mood categories: neutral, angry, sad, and afraid. The *k*-Nearest Neighbor method is used to get recommendations for recommended song titles by looking for the closeness of the value between the listener's mood and the availability of song playlists. *k*NN is used because this method is suitable for user-based collaborative problems. *k*NN will recommend three songs which then be offered to listeners by broadcasters. Based on tests conducted to the broadcasters and radio listeners, this study has produced a song request application by recommending song titles according to the listener's mood, the text message, the searching songs, and the song requests and the song details that have been requested. Functional test that has been carried out has received 100 because all test components have succeeded as expected.**

Keywords*— *k-Nearest Neighbor; Radio Broadcasters; Radio Listeners; Request; Song

1 INTRODUCTION

Radio is one of the mass media closely related to society's needs, which can provide various kinds of information, entertainment, and education. Radio as an effective mass media in disseminating audio-based information, various kinds of information can be conveyed in clear audio and in an easily understood language by the public [1].

During a radio broadcast, there is an interaction between broadcasters and radio listeners. Radio listeners can request song and give greetings to the broadcaster. Then, broadcasters will later play the requested songs.

When this research was conducted, there was a problem in the song request process. The first is that there is no place to request a song. The second problem is from the listeners' side that the listeners do not know how to request the appropriate song. The absence of song list makes listeners not know what songs are suitable for them.

The problem is the difficulty of broadcasters in monitoring song requests from listeners in order to make a system can accommodate all song requests by listeners. The way that can be done to request songs is by online telephone [2]. Song request using the telephone is done by utilizing the Voice over Internet Protocol (VoIP) network technology [3]. Technically, VoIP technology converts analog voice into digital data packets, which are later transferred over the Internet Protocol (IP). In this study, the VoIP client used is Sinch [4].

Many greetings from listeners that come through several social media accounts make broadcasters have to work twice. Thus, a song request system is designed by saving messages to the database, which later will be retrieved and displayed to the broadcaster. The concept of song request will be the same as the queuing process to be more in order [5].

To find out the listener's mood when making a phone call to request a song, speech emotion recognition technology is used. The use of speech emotion recognition technology in this study can achieve an accuracy up to 66% by using the Vokaturi API service [6]. Vokatori API can work using the PRAAT application to make extracted voice and then to measure based on the frequency and dB [7]. After that, the existing values will be equated with the results from the Vokatori Database's training data.

Four moods, neutral, angry, sad, and afraid, can be identified. Later, each mood will have its value based on the results of the analysis that has been carried out. Each song has been analyzed using the Vokatori API to determine the value of each mood. After that, for song recommendations, the *k*-Nearest Neighbor (*k*NN) recommended to find the closeness of the value between the listener's mood and the existing song. *k*NN is one of the best neighboring algorithms based on collaborative filtering [8]. The three songs with the closest value recommended by *k*NN will be the recommended song that the broadcaster will offer to listeners during the song request session via telephone. The *k*NN method is used in this study because it is in accordance with the user-based collaborative filtering model which focuses on user assessments.

Previous research has been carried out on developing the application of Spotify fire use in recommending songs based on mood. This research is only using the accelerometer sensor found on Android-based smart phones. The using of the accelerometer sensor obtained 84% of 30 users can get song recommendations according to the current mood [9]. On the other hand, there is also research on designing a song recommendation application based on mood using James Russel's Circumplex Model mood. The input is in the form of mood coordinates obtained from listener input. It can be concluded that the use of the Euclidean Distance algorithm results the closest distance between one point and another [10].

Based on some problem descriptions above, this study aims to build an application that can accommodate song requests and song recommendation that are suitable to the listener's emotions.

Researchers have conducted several previous works that have a relationship with the features used, starting from the song's basic concept, mood, recommendation methods, and the technology used. Discussions related to previous work are explained in points 1.1 to 1.7.

1.1 k-Nearest Neighbor

k-Nearest Neighbor is a method that uses a supervised learning algorithm where the results of new instances are classified based on the majority of the k-nearest neighbor category. *k*-nearest neighbor can also be explained as an algorithm that classifies objects based on learning data closest to the object to be tested [8].

*k*NN is the most straightforward formula that is often used in implementing distance search. The formula used in the calculation of *k*NN is the Euclidean Distance formula. The Euclidean distance is used [11] to find the distance between points in class *k*.

Here is the Euclidean formula:

$$Dist(p,q) \quad = \sqrt{\sum_{i=1}^{n}(qi - pi)^2} \qquad (1)$$

Some steps must be taken when using the *k*NN method. The steps can be seen at the following points:

1. Determine the parameter *k* (number of classes)
2. Calculate the distance between the test data and the existing training data using the Euclidean formula
3. Sort the calculation results according to the smallest one, then determine the environment according to the value of *k*
4. Use the class with the highest number and then specify the test data class

1.2 Song Concept

In a song, it can be processed again to find out the melody of the music. The melody can be seen from the way a person mutters or whistles. From the melody, we know the genre of a song [12].

Here are the characteristics of the song:

1. Sound

The voice in the song describes how the sound is notated or written. Sound waves are usually discussed in terms of frequency. The sound fundamental aspects are described in terms of pitch, duration, intensity, and timbre.

2. Tone

The sound is divided into tones that have a specific pitch or tuning. The difference in tuning between two notes is called an interval. The tone can arrange on different scales. The root tone of a song determines the frequency of each letter in the song.

3. Melody

In its most literal sense, the melody is a sequence of notes and duration of notes. At the same time, in another reason, the term includes a succession of other musical elements such as tonal colour, a series of letters in time. The circuit can sound alone.

4. Notation

Musical notation is a written depiction of music. In block notation, the pitch is depicted vertically while the time (rhythm) is depicted horizontally. These two elements make up the tone stick.

5. Rhythm

The time signature indicates the number of beats in the measurement and notes counting and it counts as one beat. Specific notes can be accentuated by applying pressure.

1.3 Mood

The mood is a long-lasting form of emotional state. Mood is different from simple emotions that has less specific and less intense feeling and tend not to be triggered by certain stimuli or events [13]. A person's mood can last only a few hours. Moods can be affected by many unexpected events.

Moods are very different from emotions because emotions do not have to be caused by anything. A person feels happy after getting married or receiving a gift, while a happy mood tends to react to external stimuli. A mood tends to last longer than feelings. Personal characteristics such as optimism and neuroticism influence certain types of moods. Long-term mood disorders, such as clinical depression and bipolar disorder, are considered mood disorders. Mood occurs in both internal and subjective circumstances but can often be inferred from other postures and behaviours.

The mood is a temporary state in a pleasant or unpleasant position. In other words, the mood is a feeling that occurs at a specific time and in a specific situation [14]. The mood can significantly influence someone's behaviour who is experiencing it. Vice versa, mood can be seen from a person's behaviour.

1.4 Speech Emotion Recognition

Speech emotion recognition, one of the latest speech processing challenges, is a technology for identifying emotions from a human speech. Apart from facial expressions, speech has also proven to be one of the most promising modalities for the automatic recognition of human emotions. In the field of security systems, particularly, the development can be observed over the past year. In realistic view, the classification is considered as the technical approaches that only rely on pragmatic decisions about the type, the extent, and the number of emotions according to the situation [15].

There is fact that sound often reflects the underlying emotion through tone and pitch. To successfully implement a speech emotion recognition system, we need to define and model emotions carefully. However, there is no consensus on the definition of emotion, and it is still an open problem in psychology. The discrete emotional and dimensional emotional models are two models have become common in introducing speech emotions.

Microphone dramatically affects the quality of the analyzed sound. Using multiple microphones is more practical because information about the speaker's possible location from a multi-microphone signal usually helps separate speech. One of the crucial points in emotion recognition is what features are used. In recent research, many standard features were extracted, such as energy, pitch, formant, and some spectral features such as linear prediction coefficient (LPC), Mel-frequency cepstrum coefficient (MFCC), and spectral modulation features [16].

1.5 Application Programming Interface

When two systems connect via API in integration, it has two sides, with a unique name each. One side is talking about servers. This is the side that involves the API. It helps to remember that this API is not the only program running on the server. A computer protocol is a set of rules that govern how two computers can communicate with each other. Compared to standard human language, however, computer protocols are very rigid. Think for a moment about the two sentences "My favourite colour is blue" and "Blue is my favourite colour." People can break down each sentence and see that they have the same meaning, even though the words are in a different order. Unfortunately, computers are not that smart [17].

REST is an architecture in designing web services where the REST design has resources that are accessed through a different URL address. URL address refers to a collection of functions or programs that will be executed and provide feedback in the form of a message to the sender [17]. Commands that are sent using the HTTP request method.

1.6 Vokatori

Vokatori were created in 2016 in Amsterdam. Vokaturi API developers design libraries in C and Python form. API vocabulary can reflect state of the art in recognizing emotions from human voices. The algorithm has been designed and continues to be improved, by Paul Boersma, professor of Phonetic Science at the University of Amsterdam, the world's leading principal investigator of pre-speech software analysis. Vokatori can directly measure a person's voice, whether he/she is happy, sad, afraid, angry or has a neutral mood [6].

For prosodic detection, the OpenVokaturi version 2.2b was used to extract the main dimensions of pitch, intensity and spectral slope when detecting emotions. The nine extracted features were compared to two audio recording databases to predict five emotions: happiness, sadness, anger, fear, and neutral. The predictions are returned as a value that represents the probability of each emotion. The highest score

among all predictions represents the dominant mood of the audio recording. Emotion detection using this technology has been validated up to 66.5% [18].

Currently, the open-source version of this software features five emotions with high accuracy, even if you hear the speaker for the first time. Vokatori emotion recognition can be easily integrated into existing software applications. The software has been validated against an existing emotion database and is independent language functioning. The Vokaturi software is available as a drop-in library for most modern platforms, including iOS, Android, Windows, macOS, and Linux.

Vokaturi provides three choices of libraries that developers can use. The libraries provided can be seen in Table 1.

Table 1. Vokatori Libraries

Name	*Accuracy*	*Pricing*
OpenVokaturi	66.5	Free
VokaturiPlus	76.1	Paid
VokaturiPro	76.1	Paid

1.7 Voice Over Internet Protocol

Voice over Internet Protocol (VoIP) is a technology that makes the internet media capable of direct long-distance voice communications. Voice packets are sent over the internet protocol network. VoIP is often referred as IP Telephony, Internet Telephony or Digital Phone. The working principle of VoIP's is to convert the analogue sound obtained from the speakers on the computer into a digital data packet. The PC is forwarded through the Hub / Router / ADSL Modem sent over the internet network and received by the place goals through the same medium. Alternatively, through the media, the telephone is forwarded to a phone adapter that is connected to the internet and can be received by the destination telephone.

Voice over IP (VoIP) is implemented in various ways using proprietary open standards and protocols. The following are the examples of a network protocol for implementing VoIP: Real-time Transport Protocol (RTP), Session Initiation Protocol (SIP), Inter-Asterisk eXchange (IAX), H.323, Media Gateway Control Protocol (MGCP) and Session Description Protocol (SDP)

2 METHOD

This research is used descriptive method. It is a method that aims to explain an ongoing event. Problem-solving describes the state of the subject or object in a research as a person or an institution. At this stage, the researchers took steps to collect data. The data collection methods used in this study were unstructured interviews, questionnaires and literature studies.

The software development method is built by using the waterfall method. This method is systematic, sequential in building an application. The stages consist of requirements definition, system and software design, implementation and unit testing, integration, system testing, and operation and maintenance.

2.1 Mood Value

To find the mood value, an android library called Vokaturi is used. Vokatori can reflect the state of the art in recognizing emotions from human voices. Algorithms have been designed, and are continually being improved. Flowchart of the analysis of mood values can be seen in Figure 1.

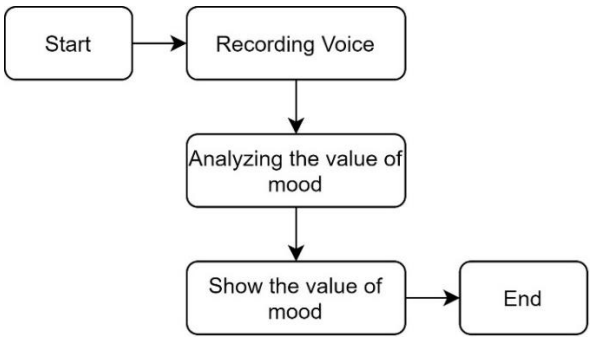

Figure 1. Mood value flowchart

2.2 Song Recommendations

The song recommendations will use the *k*-Nearest Neighbor algorithm. K-Nearest Neighbor is a method that uses a supervised learning algorithm where the results of new instances are classified based on the majority of the *k*-nearest neighbour category. To find the distance between points in class K, the Euclidean distance is used.

In this research, the song will be matched with the user's mood. In this study, the moods used were neutral, happy, sad, and angry. Previously, the song list had been analyzed using the Vocatury API, which produced a value for each mood; this data would become training data. Each song's mood value is stored in the song bank database. The architecture of the song recommendations can be seen in Figure 2.

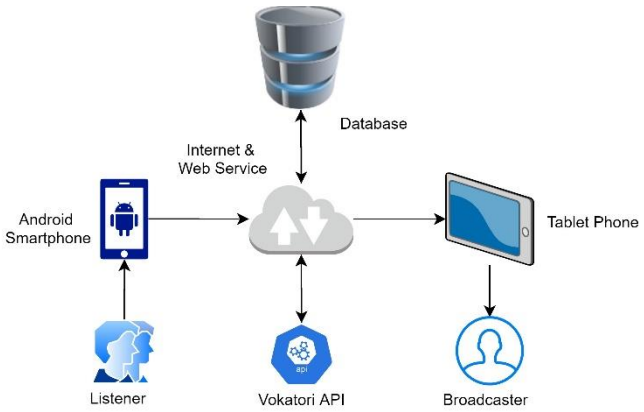

Figure 2. Song title recommendation architecture

3 RESULT AND DISCUSSION

3.1 Song recommendations result

Here are the steps to find the distance from each song which later becomes song recommendations.

1. Prepare a song that has a mood rating. The song table can be seen in Table 2.

Table 2. Value of The Song Mood

No	Title	Neutral	Sad	Happy	Angry
1	Nyaman	0.649	0.348	0	0.003
2	Senorita	0.225	0.184	0.634	0.057
3	Dance Monkey	0.329	0.221	0.471	0.079
4	Memories	0.284	0.471	0.267	0.078
5	Pura-Pura Lupa	0.355	0.381	0.264	0
6	Langit Abu Abu	0.183	0.520	0.087	0.210
7	I Love You but I'm Letting Go	0.91	0.734	0.210	0.065
8	Drowning Pool	0.089	0.164	0.208	0.539
9	Disturbed	0.084	0.365	0.473	0.078
10	Limp Bizkit	0.207	0.143	0.418	0.232

2. Prepare the value of the sound that has been analyzed by vokaturi. The table of voting values can be seen in Table3.

Table 3. Mood Value of Sound

Title	Neutral	Sad	Happy	Angry
Suara_1	0.761	0.238	0	0.001

3. At this stage, the calculation using the *k*-Nearest Neighbor algorithm begins. The *k* is three, so there are three song that become recommendation later. Here is the calculation:

a. Song rating no 1:

$$d1(p+q) = d(q,p)$$
$$= \sqrt{(q1-p1)^2 + \cdots + (qn-pn)^2}$$
$$= \sqrt{(0.649-0.761)^2 + (0.348-0.238)^2} \sqrt{+(0-0)^2 + (0.003-0.001)^2}$$
$$= 0.157$$

b. Song rating no 2:

$$d2(p+q) = d(q,p)$$
$$= \sqrt{(q1-p1)^2 + \cdots + (qn-pn)^2}$$
$$= \sqrt{(0.225-0.761)^2 + (0.184-0.238)^2} \sqrt{+(0.534-0)^2 + (0.057-0.001)^2}$$
$$= 0.761$$

c. Song rating no 3:

$$d3(p+q) = d(q,p)$$
$$= \sqrt{(q1-p1)^2 + \cdots + (qn-pn)^2}$$
$$= \sqrt{(0.329-0.761)^2 + (0.221-0.238)^2} \sqrt{+(0.371-0)^2 + (0.079-0.001)^2}$$
$$= 0.575$$

d. Song rating no 4:

$$d4(p+q) = d(q,p)$$
$$= \sqrt{(q1-p1)^2 + \cdots + (qn-pn)^2}$$
$$= \sqrt{(0.284-0.761)^2 + (0.471-0.238)^2} \sqrt{+(0.167-0)^2 + (0.078-0.001)^2}$$
$$= 0.562$$

e. Song rating no 5:

$$d5(p+q) = d(q,p)$$
$$= \sqrt{(q1-p1)^2 + \cdots + (qn-pn)^2}$$
$$= \sqrt{(0.355-0.761)^2 + (0.381-0.238)^2} \sqrt{+(0.264-0)^2 + (0.0-0.001)^2}$$
$$= 0.505$$

f. Song rating no 6:

$$d6(p+q) = d(q,p)$$
$$= \sqrt{(q1-p1)^2 + \cdots + (qn-pn)^2}$$
$$= \sqrt{(0.183-0.761)^2 + (0.520-0.238)^2} \sqrt{+(0.087-0)^2 + (0.210-0.001)^2}$$
$$= 0.682$$

g. Song rating no 7:

$$d7(p+q) = d(q,p)$$
$$= \sqrt{(q1-p1)^2 + \cdots + (qn-pn)^2}$$
$$= \sqrt{(0.091-0.761)^2 + (0.634-0.238)^2} \sqrt{+(0.210-0)^2 + (0.065-0.001)^2}$$
$$= 0.809$$

h. Song rating no 8:

$$d8(p+q) = d(q,p)$$
$$= \sqrt{(q1-p1)^2 + \cdots + (qn-pn)^2}$$
$$= \sqrt{(0.89-0.761)^2 + (0.164-0.238)^2} \sqrt{+(0.208-0)^2 + (0.539-0.001)^2}$$
$$= 0.889$$

i. Song rating no 9:

$$d9(p+q) = d(q,p)$$
$$= \sqrt{(q1-p1)^2 + \cdots + (qn-pn)^2}$$
$$= \sqrt{(0.084-0.761)^2 + (0.365-0.238)^2} \sqrt{+(0.473-0)^2 + (0.078-0.001)^2}$$
$$= 0.839$$

j. Song rating no 10:

$$d10(p+q) = d(q,p)$$
$$= \sqrt{(q1-p1)^2 + \cdots + (qn-pn)^2}$$
$$= \sqrt{(0.207-0.761)^2 + (0.143-0.238)^2} \sqrt{+(0.418-0)^2 + (0.232-0.001)^2}$$
$$= 0.738$$

The distance from the data that have been processed using *k*-Nearest Neighbor can be seen in Table 4.

Table 4. Song Distance

No	Song Title	Distance
1	Nyaman	0.157
2	Senorita	0.761
3	Dance Monkey	0.575
4	Memories	0.562
5	Pura-Pura Lupa	0.505
6	Langit Abu	0.682
7	I Love You but I'm Letting Go	0.809
8	Drowning Pool	0.889
9	Disturbed	0.839
10	Limp Bizkit	0.738

4. Sort song titles from the smallest value. The value of the sorted songs can be seen in Table 5.

Table 5. Song Distance After Sorted

No	Song Title	Distance
1	Nyaman	0.157
2	Pura-Pura Lupa	0.505
3	Memories	0.562
4	Dance Monkey	0.575
5	Langit Abu	0.682
6	Limp Bizkit	0.738
7	I Love You but I'm Letting Go	0.809
8	Disturbed	0.839
9	Drowning Pool	0.889
10	Senorita	0/761

5. Take the top three song that have the smallest distance. The song is the recommendation song for listeners based on the mood analysis of the listener's voice in the previous table. Here are the recommended song titles: "*Nyaman"* – Andmesh, "*Pura-Pura Lupa"* – Mahen and "Memories" – Maroon 5

3.2 System Implementation

Implementation is the stage of implementing and testing the system based on the analysis and design results that have been done. This application uses the Java programming language with MySQL database. Implementation and testing are carried out to determine whether the system being built meets the objectives or not.

In implementing software, support is needed either from software or suitable hardware. The following is Table 6. It is the hardware used during implementation from both the broadcaster and the listener side.

Table 6. Hardware Specification

Hardware	Broadcaster Specifications	Listener Specifications
Processor	*Octa-core*	*Octa-core*
Ram	2 GB	12 GB
Screen Size	10.5 "	6.5 "
Internal Memory	64 GB	256 GB

The software implementation is software used on each device to support testing in this study. The software here focuses more on using the operating system. Android version 5.0 is used by broadcasters, while the Android operating system version 10 is used by listeners.

Interface implementation is a design that has been previously created and then implemented into the application. The user interface functions to connect the user with the existing backend to be used. The following is an interface that has been implemented into the application:

1. Login

The login interface serves to display fields that are useful for validating accounts. The login interface can be seen in Figure 3.

Figure 3. Login interface

2. Request Song Via Phone

The song request interface via phone displays the song request page via the phone. The song request interface via telephone can be seen in Figure 4.

Figure 4. Interface of request song via phone

3. Request a song via text message

The song request interface via text message displays the field to request the song. The song request interface via text can be seen in Figure 5.

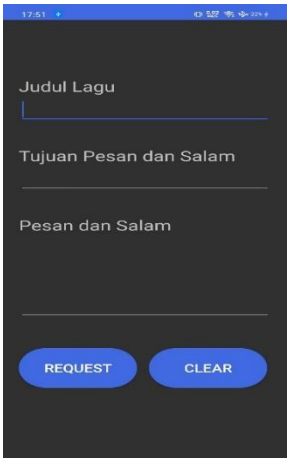

Figure 5. Interface of request song via text message

4. Song Request

The song recommendation interface is used to pick up calls from listeners and display recommended songs. The song recommendation interface can be seen in Figure 6.

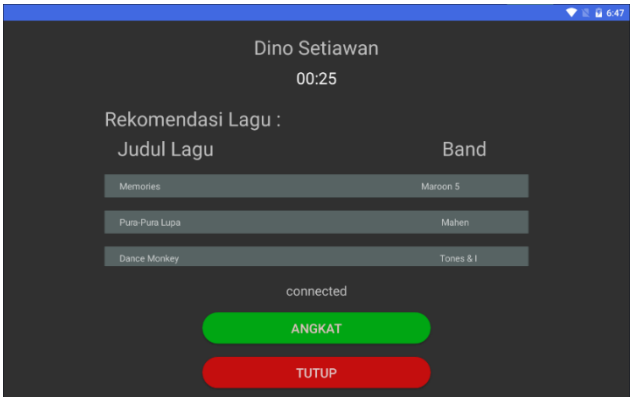

Figure 6. Song Recommendation Interface

5. View Song Request

The interface for viewing song requests serves to display song requests from the listeners. The interface for viewing song requests can be seen in Figure 7.

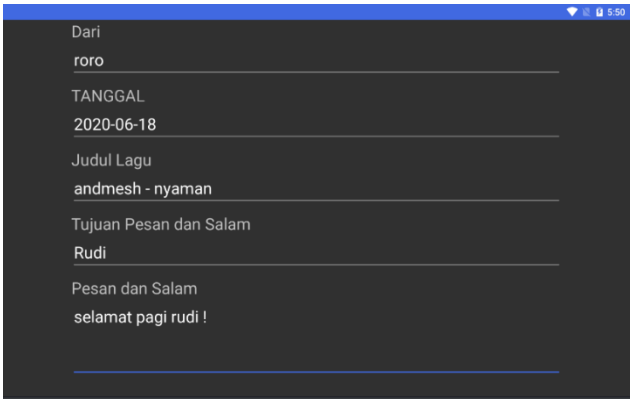

Figure 7. See Request Song

6. Results of song recommendations

The display of Figure 8 is the result of the song recommendations obtained based on the listener's mood.

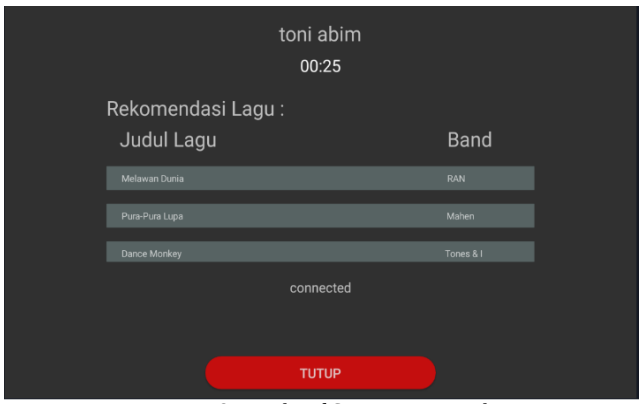

Figure 8. Results of Song Recommendations

3.3 System Testing

System testing is done using the Black Box method. The tested items are the login process, song requests via telephone, song requests via text messages and song recommendations. The results of system testing can be seen in Table 7.

Table 7 System Testing Result

Testing Items	Data Entry	Result
Enter the listener's main menu page	Listener's email address	[√] Success
Enter the broadcaster's main menu page	Broadcaster's email address	[√] Success
Make a phone call to the RRI broadcaster	Id_telp: RRI	[√] Success
Conduct a mood analysis based on the listener's voice	Analog sound	[√] Success
Request a song	Song title, message destination, message content	[√] Success
Displays message missing parameter	song title: (blank) message and greeting destination: (empty) message and greeting: (empty)	[√] Success
Make a phone call to the RRI broadcaster	id_User: 6 neutral: 0.334 happy: 0.112 sad: 0.422 angry: 0243	[√] Success
Move to the song request detail page based on the selected song request form	id_Request Song: 15	[√] Success
Displays song recommendations based on the listener's mood rating	id_User: 6 neutral: 0.334 happy :0.112 sad :0.422 angry :0.243	[√] Success

Based on the results of Black Box testing that has been carried out with several examples of test items, it is found that all functions in the application can be ran and functioned as expected properly. The completion of this research is constrained by the Covid-19 pandemic. Because of that reason, the user acceptance testing (UAT) test is still not finished and will be used as further research. However, researchers have conducted an online demo for the representative employee at the research site. The results showed that the resulting application used in the study has been corresponded to the purpose.

This research has succeeded in combining the KNN method with the Vocatori API. This means that it can be used and ran well on mobile applications using the voice input feature and internet connection. This result is different from research to provide song recommendations only with collaborative filtering methods [19], and Cosine Similarity Algorithms [20].

4 CONCLUSION

As the conclusion, the application built can make the song request process easier for broadcasters and listeners in the song request process. This application runs well and it is stable on the Android operating system version 5.0 to the latest version, Android 10. Currently, the application is still

in the trial phase to improve the content and additional features. It is not yet to be implemented massively on the radio as a research place. Further research is still needed to complete some additional features so that tests can be real-time carried out directly to the radio.

ACKNOWLEDGMENT

The authors acknowledge the support from Kemenrikstek/BRIN using Penelitian Riset Terapan grant in 2019-2021 with multi-year contract number, the Multiple Years Research Contract number 019/SP2H/LT-AMAND/LL4/2020. The funding is used for procuring servers and API services from the Google Cloud Platform.

REFERENCES

[1] A. Murtadha, "PERAN LEMBAGA PENYIARAN RADIO REPUBLIK INDONESIA SEBAGAI SABUK PENGAMANAN INFORMASI DI WILAYAH PERBATASAN KABUPATEN NATUNA PROVINSI KEPULAUAN RIAU," *J. Penelit. Komun. dan Pembang.*, vol. 15, no. 2, 2018.

[2] A. S. B. Nugroho, I. Wardiah, and M. K. Ilmi, "Aplikasi Radio Daring Pemerintah Kabupaten Hulu Sungai Tengah Berbasis Web," *Poros Tek.*, vol. 10, no. 1, 2018.

[3] M. Exsan and U. Fadlilah, "Pembangunan Infrastruktur Voice Over Internet Protocol di Organisasi Perangkat Daerah Boyolali menggunakan Server Elastix," *Emit. J. Tek. Elektro*, vol. 17, no. 2, 2017.

[4] N. C. R., R. R. R., and V. K. R. K., "Voice over IP Via IEEE 802.11 Wireless LAN," *N., Ananthi*, vol. 2, no. 2, pp. 218–222, 2016.

[5] S. B. Aziz, T. A. Riza, and R. Tulloh, "PERANCANGAN DAN IMPLEMENTASI APLIKASI SISTEM ANTRIAN UNTUK PASIEN PADA DOKTER UMUM BERBASIS ANDROID DAN SMS GATEWAY," *J. Elektro dan Telekomun. Terap.*, vol. 2, no. 1, 2016.

[6] J. M. Garcia-Garcia, V. M. R. Penichet, and M. D. Lozano, "Emotion detection: A technology review," in *ACM International Conference Proceeding Series*, 2017, vol. Part F131194.

[7] M. Magdin, T. Sulka, J. Tomanová, and M. Vozár, "Voice Analysis Using PRAAT Software and Classification of User Emotional State," *Int. J. Interact. Multimed. Artif. Intell.*, vol. 5, no. 6, 2019.

[8] G. Li and J. Zhang, "Music personalized recommendation system based on improved KNN algorithm," in *Proceedings of 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference, IAEAC 2018*, 2018.

[9] J. Sarah, "Pemanfaatan Api Spotify Dan Twitter Dalam Merekomendasikan Lagu Berdasarkan Mood Pada Tweet Dan Aktivitas Pengguna Melalui Sensor Accelerometer Berbasis Android," Universitas Komputer Indonesia, 2018.

[10] M. Rahardian, "PERANCANGAN APLIKASI REKOMENDASI PILIHAN LAGU BERDASARKAN MOOD," *Simki-Techsain*, 2015.

[11] N. L. G. P. Suwirmayanti, "Penerapan Metode K-Nearest Neighbor Untuk Sistem Rekomendasi Pemilihan Mobil," *Techno.Com*, vol. 16, no. 2, 2017.

[12] N. Büdenbender and G. Kreutz, "Familiarity of Western melodies: An exploratory approach to influences of national culture, genre and musical expertise," *Music. Sci.*, vol. 20, no. 2, 2016.

[13] I. G. Harsemadi, M. Sudarma, and N. Pramaita, "Implementasi Algoritma K-Nearest Neighbor pada Perangkat Lunak Pengelompokan Musik untuk Menentukan Suasana Hati," *Teknol. Elektro*, vol. 16, no. 1, pp. 14–20, 2017.

[14] R. Sutiama and P. Yandri, "Suasana toko, suasana hati, kualitas pelayanan, dan keputusan pembelian antara konsumen restoran modern dan restoran tradisional," *Indones. J. Econ. Appl.*, vol. 1, no. 2, 2019.

[15] S. Mao, D. Tao, G. Zhang, P. C. Ching, and T. Lee, "Revisiting Hidden Markov Models for Speech Emotion Recognition," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2019, vol. 2019-May.

[16] T. Yoshioka, H. Erdogan, Z. Chen, and F. Alleva, "Multi-Microphone Neural Speech Separation for Far-Field Multi-Talker Speech Recognition," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2018, vol. 2018-April.

[17] B. Cooksey, "An Introduction to APIs," *An Introd. to EDCI 554*, 2015.

[18] C. LYTRIDIS, E. VROCHIDOU, and V. KABURLASOS, "Emotional Speech Recognition toward Modulating the Behavior of a Social Robot," *Proc. JSME Annu. Conf. Robot. Mechatronics*, vol. 2018, no. 0, pp. 1A1-B14, 2018.

[19] D. Sánchez-Moreno, A. B. Gil González, M. D. Muñoz Vicente, V. F. López Batista, and M. N. Moreno García, "A collaborative filtering method for music recommendation using playing coefficients for artists and users," *Expert Syst. Appl.*, vol. 66, 2016.

[20] M. A. Budiman and G. A. V. Mastrika Giri, "Song Recommendations Based on Artists with Cosine Similarity Algorithms and K-Nearest Neighbor," *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 8, no. 4, 2020.