

Game Belajar Membaca Kata Bahasa Inggris Dengan Pendekatan Metode MFCC dan KNN

I Wayan Agus Juniarta^{a1}, I Gede Santi Astawa^{a2}, Made Agung Raharja^{a3},
I Gusti Ngurah Anom Cahyadi Putra^{a4}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Udayana

Jalan Raya Kampus Unud, Badung, 08361, Bali, Indonesia

¹agusjuniarta16@gmail.com@gmail.com

²santi.astawa@cs.unud.ac.id

³made.agung@unud.ac.id

⁴anom.cp@unud.ac.id

Abstract

Learning English, especially in terms of speaking, at the elementary school level in rural areas with minimal technology is a complex challenge. In an era of globalization filled with technological advances, it is important for students to learn English from an early age. Internet and smartphone technologies, which are becoming increasingly prevalent in children's lives, can be used as innovative tools for learning English. A multimedia-based active and interactive learning approach can enhance the effectiveness of learning in this context. This research aims to develop innovative learning methods by focusing on creating internet and smartphone-based games to facilitate English speaking skills. These games use voice features extracted with the Mel Frequency Cepstral Coefficients (MFCC) method, which effectively recognizes differences in language and accent in speech recognition. The classification method employed is K-Nearest Neighbors (K-NN), chosen for its simplicity in implementation and suitability for datasets with many features. By integrating these technologies, the proposed approach can provide immediate feedback and engage students more effectively, making the learning process enjoyable and accessible, especially for those in rural areas. This innovation represents a significant step towards improving English education in resource-limited settings. .

Keywords: *Speaking, MFCC method (Mel Frequency Cepstral Coefficients), K-NN (K-Nearest Neighbors) classification method, Learning effectiveness, English learning*

1. Pendahuluan

Para siswa yang hidup di zaman globalisasi dimana teknologi sudah berkembang pesat dan semua sistem sudah menggunakan bahasa Inggris, secara tidak langsung pembelajaran bahasa Inggris akan sangat diperlukan untuk diajarkan sejak bangku sekolah dasar. Pembelajaran multimedia di dalam kelas dapat ditingkatkan didasari oleh pendekatan pembelajaran aktif yang dapat memicu dan memperkuat stimulus dan respon anak dalam pembelajaran, pendekatan interaktif dapat merangsang indra pendengaran dan penglihatan guna untuk memperkuat daya ingat anak dalam mempelajari bahasa Inggris [1]. Pembelajaran multimedia di dalam kelas dapat ditingkatkan didasari oleh pendekatan pembelajaran aktif yang dapat memicu dan memperkuat stimulus dan respon anak dalam pembelajaran, pendekatan

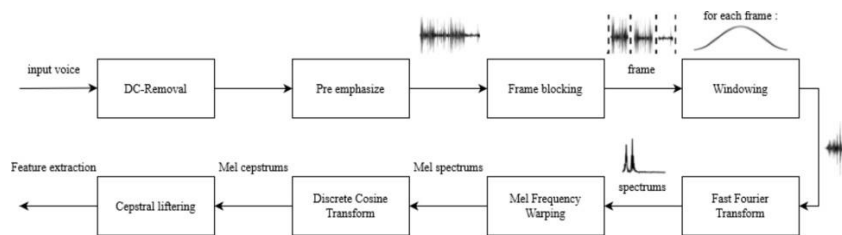
interaktif dapat merangsang indra pendengaran dan penglihatan guna untuk memperkuat daya ingat anak dalam mempelajari bahasa Inggris. Adapun penelitian serupa yang dilakukan oleh (Al Irsyadi, Annas, dan Kurniawan 2019) tentang Game Edukasi Pembelajaran Bahasa Inggris untuk Pengenalan Benda-Benda di Rumah bagi Siswa Kelas 4 Sekolah Dasar di dapatkan bahwa berdasarkan pengujian *User Acceptance Test* dari 5 pertanyaan yang diajukan ke pengguna mendapatkan persentase sebesar 69,19%[5].

Didalam penelitian yang diajukan ini akan difokuskan kedalam pengembangan game pembelajaran speaking dalam bahasa Inggris[5]. Didalam penelitian yang diajukan ini akan difokuskan kedalam pengembangan game pembelajaran speaking dalam bahasa Inggris. Metode yang digunakan untuk mengekstraksi fitur dari suara pengguna game adalah MFCC (Mel-Frequency Cepstral Coefficient) serta metode klasifikasi K-NN digunakan untuk mencari kedekatan antara suara pengguna dengan suara pakar yang sudah disimpan dalam database.

2. Metode Penelitian

2.1 Mel Frequency Cepstral Coefficient

Metode ini biasanya dipilih karena dapat mendefinisikan parameter suara dengan baik dan mampu menangkap karakteristik suara yang penting bagi proses pengenalan suara, dapat menghasilkan data seminimal mungkin tanpa mengurangi informasi-informasi penting yang terkandung dalam sinyal tersebut, dan dapat mereplikasikan organ pendengaran manusia dalam hal mendengar suara. Ada beberapa tahap untuk mendapatkan capstrum mel dapat dilihat pada gambar 1.



Gambar 1 Visualisasi Suara Cowok Soal 1 (Indonesia is my country)

2.1.1. DC-Removal

Dc-Removal merupakan persamaan untuk menghitung mean dari data sampel dan mengurangi nilai setiap sampel suara dengan rata-rata tersebut dengan tujuan untuk mendapat normalisasi dari data suara. Persamaan *DC-Removal* dapat dilihat pada persamaan 2.1

$$y[n] = x[n] - \bar{x}, 0 \leq n \leq N - 1 \dots \dots \dots (2.1)$$

Dimana:

$y[n]$ = sampel sinyal hasil proses DC-Removal

$x[n]$ = sampel sinyal asli

\bar{x} = nilai rata-rata sampel sinyal asli

N = panjang sinyal

2.1.2. Pre-Emphasize

Pre-Emphasize Filtering merupakan salah satu jenis filter yang sering digunakan untuk mempertahankan frekuensi-frekuensi tinggi pada sebuah spektrum yang umumnya tereliminasi pada proses selanjutnya. Contoh Pre Emphasis pada gambar 2. Persamaan yang umum digunakan dalam *pre-emphasize* filter dapat dilihat pada persamaan 2.2

$$y[n] = s[n] - \alpha s[n - 1], 0.9 \leq \alpha \leq 1.0 \dots\dots\dots (2.2)$$

Dimana :

y[n] = sinyal hasil pre-emphasize filter s

[n] = sinyal sebelum pre-emphasize filter

2.1.3. Frame Blocking

Karena signal suara terus mengalami perubahan akibat adanya pergeseran artikulasi dari organ produksi vokal, sinyal harus diproses secara *short segment* Panjang *frame* yang biasanya digunakan untuk pemrosesan sinyal antara 10-30 milidetik[2].

2.1.4. Windowing

Proses *window* berfungsi untuk mengurangi kemungkinan terjadinya kebocoran spektral atau *aliasing*. *Aliasing* adalah signal baru dimana memiliki frekuensi yang berbeda dengan signal aslinya [2]. Representasi fungsi window terhadap sinyal suara dapat dilihat pada persamaan 2.3

$$x(n) = x_t(n)w(n), n = 0, 1, 2, \dots, N - 1 \dots\dots\dots (2.3)$$

Dimana :

x(n) = nilai sampel sinyal dari hasil windowing x_t(n) = nilai sampel dari frame sinyal ke-i

w(n) = fungsi window

N = frame size, merupakan kelipatan

2.1.5. Fast Fourir Transform

Fast Fourier transform (FFT) adalah salah satu metode pengkonversian dari sinyal analog ke sinyal frekuensi [3]. Proses FFT dilakukan pada semua frame dari sinyal yang sudah melalui tahap windowing. FFT merupakan salah satu algoritma yang cepat untuk menerapkan *Discrete Fourier Transform* (DCT) dimana FFT menghilangkan proses perhitungan kembar dalam DCT FFT dapat dihitung menggunakan persamaan 2.4

$$f(n) = \sum_{k=0}^{N-1} W_k e^{-2\pi jkn/N}, 0 \leq n \leq N - 1 \dots\dots\dots (2.4)$$

Dimana :

F(n) = Hasil proses FFT

W_k = Windowing

2.1.6. Discrete Cosine Transform

Mel Frequency wrapping umumnya dilakukan menggunakan *filter bank*. *Filter bank* merupakan salah satu bentuk *filter* yang dilakukan dengan tujuan untuk mengetahui ukuran energi dari *frequency* band tertentu dalam sinyal suara. *Filter bank* dapat dihitung menggunakan persamaan 2.5

$$Y[t] = \sum_{j=1}^N S[j]H_t[j] \dots \dots \dots (2.5)$$

Dimana :

N= Jumlah magnitude spectrum (N E N)

S[j]= magnitude spectrum pada frekuensi j

H_t[j]= koefisien filterbank pada frekuensi (0 < t < M-1)M = jumlah channel dalam filterbank

2.1.7. Cepstral Lifering

Tahap selanjutnya adalah *Discrete Cosine Transform* (DCT), yaitu perubahan konversi dari domain frekuensi ke domain waktu. Namun hasil dari DCT mendekati PCA (*principle component analysis*). PCA merupakan metode *static* klasik yang digunakan secara luas dalam analisa data dan kompresi [2]. DCT dapat dihitung menggunakan persamaan :

$$C = \sum_{k=1}^K (\log S_k) \cos \left[n \left(k - \frac{1}{2K} \right) \right], \quad n = 1, 2, \dots, K$$

Dimana :

S_k= Keluaran dari proses filterbank pada index

K= jumlah koefisien yang diharapkan

2.1.8. K-Nearest Neighbor

Algoritma *K-Nearest Neighbor* (KNN) merupakan suatu metode yang digunakan dalam proses klasifikasi. Klasifikasi *Nearest Neighbor* didasarkan pada pembelajaran dengan analogi, yaitu mencari kelas terdekat dari objek yang dimiliki dengan dataset yang sudah ada dengan mengukur jarak dari setiap objek dengan dataset sebanyak jumlah dataset yang dimiliki. Untuk mengukur jarak dari dua buah data dapat menggunakan berbagai macam metode seperti metode euclidean distance yang dapat dilihat pada persamaan

$$\text{dist}(X_1, X_2) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2}$$

Dimana

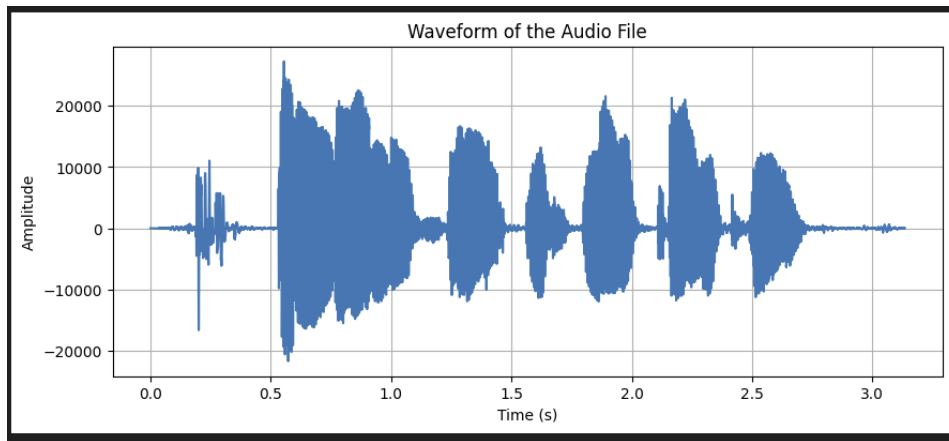
X₁ dan X₂ = Data yang diuji

x_1 dan x_2 = Atribut dari setiap data

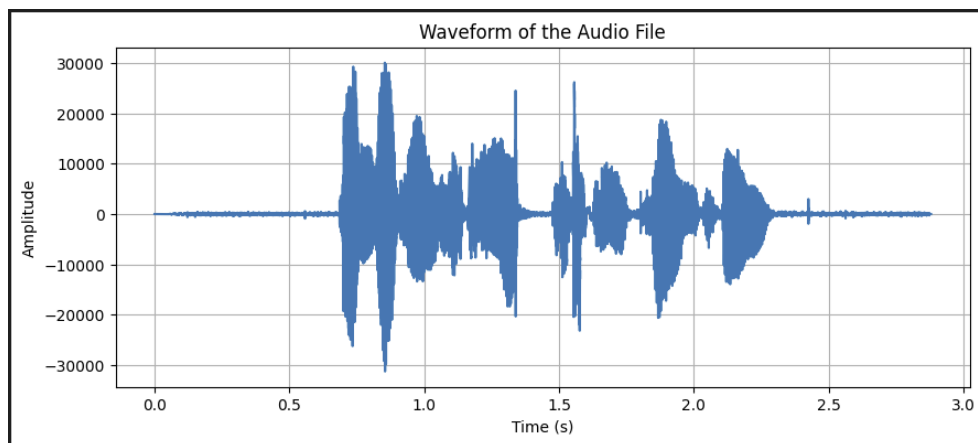
3. Hasil dan Pembahasan

3.1. Pengolahan Data Awal

Proses pengolahan *dataset* merupakan proses untuk mengolah data sampel yang berupa *file .wav* menjadi sekumpulan fitur yang disimpan menjadi dataset. Pada proses ini terdiri dari tahap *preprocessing*, dan ekstraksi fitur menggunakan metode MFCC. Dataset pelatihan yang sudah dikelompokkan dalam variabel word menjadi enam kelas. Tiap dataset nantinya akan diberi label sesuai dengan angka 1-6. Dan untuk contoh visualisasi dataset seperti pada gambar 2 dan gambar 3.



Gambar 2 Visualisasi Suara Cowok Soal 1 (Indonesia is my country)



Gambar 3 Visualisasi Suara Cewek Soal 1 (Indonesia is my country)

3.2. Preprocessing

Preprocessing merupakan proses pertama yang dilakukan oleh sistem dimana pada tahap ini setiap suara yang ada di folder referensi akan melewati 2 proses yaitu *noise reduction* dan *silence removal*. *Noise reduction* merupakan proses untuk menghilangkan suara derau yang ada pada sinyal suara inputan. Setelah itu masuk ke proses *silence removal* dimana dalam proses ini suara yang merupakan *silence* akan dihilangkan. Sehingga, hasil dari proses ini adalah sinyal suara yang tidak memiliki *noise* dan *silence*.

3.3. Ekstraksi Fitur

Proses ekstraksi fitur merupakan proses untuk mendapatkan fitur-fitur dari sinyal yang telah melalui proses *preprocessing* dengan menggunakan algoritma MFCC. Data hasil *preprocessing* berupa vektor sinyal suara akan menjadi input dari proses ekstraksi fitur ini. Hasil dari proses ekstraksi fitur adalah fitur vektor yang berupa matriks dengan ukuran $nframe \times jkoef$, dimana $nframe$ adalah jumlah frame yang dihasilkan sedangkan $jkoef$ adalah koefisien MFCC yang ditentukan. Hasil dari proses ini akan disimpan di database yang telah ditentukan.

3.4. Klasifikasi Data

Klasifikasi data merupakan tahap pencocokan antara input suara dari pengguna dengan dataset yang ada pada sistem. Pada proses klasifikasi menggunakan algoritma yaitu KNN (*K-Nearest Neighbor*). Adapun proses klasifikasi dengan menggunakan metode KNN dengan perhitungan jarak dengan metode *euclidean distance*.

3.5. Skenario Pengujian

Dalam skenario pengujian dibagi menjadi dua yaitu *functional testing* dan *non functional testing*.

3.5.1. Functional testing

Functional testing merupakan pengujian yang berfokus pada hasil eksekusi dari setiap fitur pada aplikasi yang dibangun. Dalam functional testing akan dilakukan pengujian UAT (User Acceptance Testing). Pengujian UAT adalah pengujian yang ditujukan untuk mengetahui persepsi dari pengguna terhadap aplikasi yang akan dibuat dengan cara memberikan kuisioner yang berisi penilaian perspektif kepada pengguna. Pada penelitian ini pengujian dilakukan terhadap fitur yang ada dalam aplikasi. Bobot dari pertanyaan UAT bisa dilihat pada tabel 1 berikut.

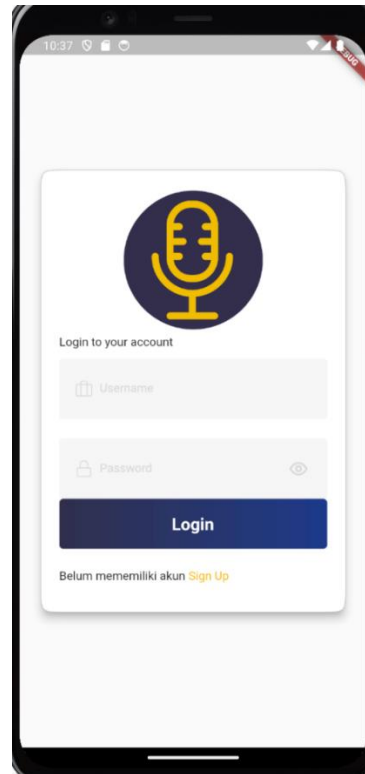
3.5.2. Non-Functional Testing

Non-functional Testing merupakan pengujian terhadap performa dan kegunaan aplikasi yang dibangun. Pada penelitian yang akan dilakukan, *non-functional testing* yang dilakukan adalah menggunakan *confusion matrix*. Pengujian ini digunakan untuk menguji akurasi data yang dihasilkan oleh sistem. *Confusion matrix* merupakan salah satu metode pengujian sistem yang berbentuk tabel. Pengujiannya adalah dengan mengukur kinerja suatu set data. Tujuan dari pengujian ini adalah untuk mengetahui akurasi data yang dihasilkan oleh sistem. *Precision* digunakan untuk menghitung efektivitas dari rekomendasi sedangkan relevansinya dihitung dengan *recall*.

4. Kesimpulan

4.1. Implementasi Aplikasi "E-Voice"

Perancangan Aplikasi menggunakan bahasa pemrograman *Dart* dan menggunakan framework *Flutter* sebagai front-end dari Aplikasi yang dibangun dan untuk registrasi user dan logic algoritma menggunakan bahasa pemrograman *Python* untuk tampilan *Home Page* bisa dilihat pada gambar 4.



Gambar 4 Home Page

4.2. Implementasi Ekstraksi Fitur

4.2.1. Implementasi MFCC (*Mel Frequency Cepstral Coefficients*)

Ekstraksi fitur audio dengan menggunakan metode MFCC dilakukan untuk mendapatkan fitur suara dengan frekuensi rendah. Ekstraksi fitur MFCC dimulai dengan melakukan beberapa preprocessing audio pada gambar 5.

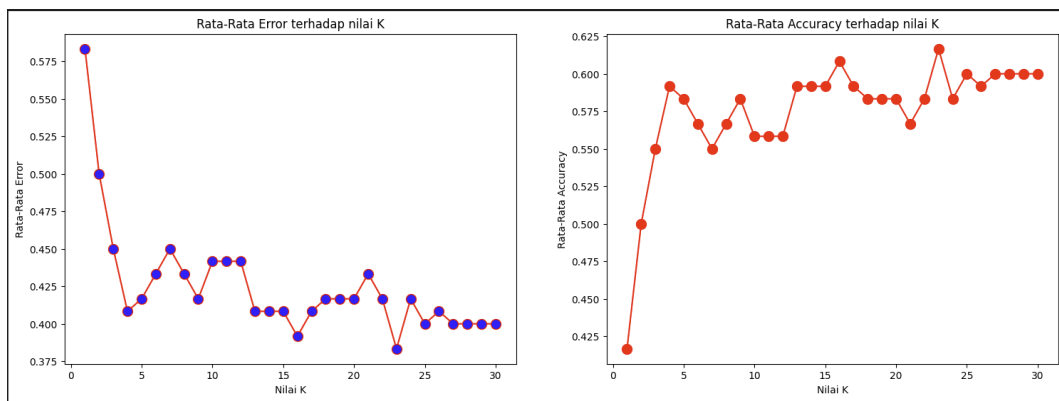
	prefix_0	prefix_1	prefix_2	prefix_3	prefix_4	prefix_5	prefix_6	prefix_7	prefix_8	prefix_9	prefix_10	prefix_11	word
0	10.926098	77.255434	29.190295	-2.808359	-28.318793	-120.671233	-139.781283	-147.947482	-93.860842	-107.040543	-97.056621	-96.418836	1
1	10.926098	77.255434	29.190295	-2.808359	-28.318793	-120.671233	-139.781283	-147.947482	-93.860842	-107.040543	-97.056621	-96.418836	1
2	10.926098	77.255434	29.190295	-2.808359	-28.318793	-120.671233	-139.781283	-147.947482	-93.860842	-107.040543	-97.056621	-96.418836	1
3	-77.386866	23.865492	-162.655166	-172.529460	-73.967493	33.722486	-188.301666	-21.403098	-11.849547	16.603345	-56.774844	-174.451905	1
4	-44.366627	-111.261167	-26.035281	-141.130345	-202.873212	-38.846477	-42.438222	-98.243156	-70.292935	148.700753	-0.785832	32.702702	1
...
594	-109.701990	47.294023	-29.356625	14.456736	-51.483257	-69.374035	-205.149561	19.333543	-86.308915	-115.024882	-84.077266	-131.972032	6
595	-96.657691	39.197153	-118.327525	-32.774413	-95.009707	-146.828364	-307.829753	-150.600359	-133.124126	-43.860060	25.238984	-154.398437	6
596	-123.340253	-6.918728	-75.609002	-27.155422	-2.339270	-130.187311	-332.882578	-103.535060	63.198664	98.552062	130.013041	4.818957	6
597	-99.787915	-33.851389	-179.170548	-6.716594	-48.168468	-113.756870	-330.919031	-76.775333	-5.174227	-31.688621	90.013368	-77.847021	6
598	-123.424273	6.881012	67.821285	-21.364139	-60.936034	-39.290792	-76.624841	-135.441761	50.661665	-105.639182	-52.780499	-128.799425	6

Gambar 5 Hasil Pengolahan MFCC

Hasil dari ekstraksi fitur mengubah gelombang suara menjadi beberapa tipe parameter seperti cepstral coefficient yang mempresentasikan audio file. Selain itu MFCC menghasilkan fitur vektor yang mengonversi sinyal suara menjadi beberapa vektor untuk pengenalan fitur suara. Nilai fitur ini digunakan menjadi inputan model klasifikasi KNN sehingga setiap audio (kelas) dari soal bahasa inggris dapat di bedakan berdasarkan dari nilai fiturnya. Dan menambahkan kelas word untuk semua dataset secara manual. Gambar dibawah merupakan contoh untuk nilai fitur MFCC

4.2.2. Implementasi Metode Klasifikasi dengan Metode KNN

Pada proses ini akan dibuat sebuah model untuk menyimpan data hasil pelatihan dari klasifikasi dengan metode KNN. Pertama, program akan membaca input dataset pelatihan yang sudah dikelompokkan dalam variabel word menjadi enam kelas. Tiap dataset nantinya akan diberi label sesuai dengan angka 1-6. Dimana nantinya dataset ini akan dibagi menjadi data latih dan data testing dengan perbandingan 80% dan 20%. Selanjutnya, seluruh dataset akan melalui proses ekstraksi fitur dengan metode MFCC. Setelah mendapatkan nilai-nilai fitur dari dataset maka data nilai fitur ini akan disimpan ke dalam file ".csv". File ini nantinya yang akan digunakan dalam proses pembuatan model KNN. Hasil rata-rata dan akurasi dari model yang dibuat bisa dilihat pada gambar 6.



Gambar 6 Akurasi dan Error Pada Model

4.2.3. Akurasi Model Klasifikasi KNN

Dalam pengujian model KNN pada K = 23 mendapatkan akurasi, recall dan presisi pada tabel 1 berikut:

Tabel 1. Pengujian Pada Model

Akurasi	Presisi	Recall
61.66%	61.34%	61.92%

4.3. Pengujian UAT (User Acceptance Testing)

User Acceptance Testing adalah suatu proses pengujian yang dilakukan oleh pengguna dengan hasil output sebuah dokumen hasil uji yang dapat dijadikan bukti bahwa software sudah diterima dan sudah memenuhi kebutuhan yang diminta.[4]. Untuk Mengetahui tanggapan responden (user) terhadap aplikasi, maka dilakukan pengujian dengan memberikan 15 pertanyaan kepada 20 responden dimana jawaban dari pertanyaan tersebut terdiri dari tingkatan yang dapat dipilih pada pada tabel 2 berikut:

Tabel 2. Bobot Jawaban Pengujian UAT

Jawaban	Bobot
---------	-------

A. Sangat Baik : Mudah/bagus/sesuai/jelas	5
B. Baik : Bagus/sesuai/jelas	4
C. Cukup : Sulit/bagus/sesuai/jelas	3
D. Kurang : Sulit/jelek/tidak sesuai	2
E. Sangat Kurang : Sulit/jelek/tidak sesuai/tidak jelas	1

Data yang didapat diatas oleh dengan cara mengalikan setiap poin jawaban dengan bobot yang sudah ditentukan sesuai dengan tabel bobot nilai jawaban. Dan hasil perhitungan dengan menghasilkan setiap jawaban bobot yang sudah ditentukan maka didapat hasil sebagai berikut: Dari hasil pengujian User Acceptance Test maka dapat ditarik kesimpulan pada tabel 3 berikut:

Tabel 3. Hasil Pengujian UAT

No	(a)	(b)	(c)
	Jumlah nilai dari responden	Rata – rata (a/jumlah responden)	Hasil presentase (b/bobot tertinggi*100)
1	88	4,4	88%
2	85	4,25	85%
3	85	4,25	85%
4	80	4	80%
5	88	4,4	88%
6	85	4,25	85%
7	89	4,45	89%
8	84	4,2	84%
9	89	4,45	89%
10	89	4,45	89%
11	87	4,35	87%
12	87	4,35	87%
13	88	4,4	88%
14	90	4,5	90%
15	87	4,35	87%
Jumlah			87%

Dari data diatas dapat disimpulkan bahwa aplikasi mempunyai tampilan yang menarik, menu-menu pada aplikasi android mudah di pahami, aplikasi mudah dan nyaman digunakan, mudah dioperasikan, kemampuan dan fungsi sesuai harapan serta sistem sangat di butuhkan. Berdasarkan hasil yang telah didapatkan dari penelitian ini, maka dapat disimpulkan bahwa perancangan aplikasi *E-Voice* yang mengimplementasikan dalam bentuk aplikasi android dan mampu mengimplementasikan algoritma ekstraksi fitur MFCC dan Klasifikasi KNN dalam mengklasifikasikan suara pengguna dengan suara pakar. Dari hasil penelitian yang dilakukan dapat dilihat bahwa sistem dapat melakukan proses klasifikasi suara dengan cukup baik dimulai dari proses *register* akun pengguna di halaman *register* setelah itu pengguna bisa login menggunakan akun yang sudah didaftarkan di halaman *login* lalu pengguna masuk ke halaman utama yang berupa halaman latihan disana pengguna bisa langsung mencoba

beberapa soal berbahasa Inggris yang disediakan dengan cara merekam suara pengguna di halaman *record* dengan menggunakan *microphone* lalu bisa memutar ulang rekaman yang dibuat sekaligus melihat seberapa akurat suara pengguna dengan suara soal berdasarkan intonasi dari suara rekaman pengguna. Performa dan akurasi dari pengujian yang dilakukan didapat akurasi sebesar 61.66% presisi sebesar 61.34% dan *recall* sebesar 61.92% menggunakan dataset suara peneliti. Testing menggunakan UAT (*User Acceptance Testing*) mendapatkan hasil sekitar 87%.

Referensi

- [1] Al Irsyadi, F. Y., Annas, R., & Kurniawan, Y. I. (2019). Game Edukasi Pembelajaran Bahasa Inggris untuk Pengenalan Benda-Benda di Rumah bagi Siswa Kelas 4 Sekolah Dasar. *Jurnal Teknologi Dan Informasi*, 9(2), 78–92. <https://doi.org/10.34010/jati.v9i2.1844>
- [2] Helmiyah, S., Fadlil, A., Yudhana, A., Dahlan, A., & Studi Teknik Elektro, P. (2018). Pengenalan Pola Emosi Manusia Berdasarkan Ucapan Menggunakan Ekstraksi Fitur Mel-Frequency Cepstral Coefficients (MFCC) Speech Based Emotion Pattern Recognition Using Mel-Frequency Cepstral Coefficients (MFCC) Feature Extraction. *Cogito Smart Journal*, 4(2), 372–381.
- [3] Pujiyanto, P., Mujito, M., Prabowo, D., & Prasetyo, B. H. (2020). Pemilihan Warga Penerima Bantuan Program Keluarga Harapan (PKH) Menggunakan Metode Simple Additive Weighting (SAW) dan User Acceptance Testing (UAT). *Jurnal Informatika Universitas Pamulang*, 5(3), 379. <https://doi.org/10.32493/informatika.v5i3.6671>
- [4] Putra, D., & Resmawan, A. (2011). Verifikasi Biometrika Suara Menggunakan Metode Mfcc Dan Dtw. *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, 2(1), 8–21.
- [5] Rahayu, S. L., & Fujiati, F. (2018). Penerapan Game Design Document dalam Perancangan Game Edukasi yang Interaktif untuk Menarik Minat Siswa dalam Belajar Bahasa Inggris. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(3), 341. <https://doi.org/10.25126/jtiik.201853694>