

Implementasi Metode *K-Nearest Neighbor* Dalam Mengklasifikasikan Jenis Suara Berdasarkan Jangkauan Vokal

Ni Made Putri Wahyuni^{a1}, Luh Arida Ayu Rahning Putri^{a2}, I Gusti Ngurah Anom Cahyadi Putra^{b3}, I Dewa Made Bayu Atmaja Darmawan^{b4}, Made Agung Raharja^{b5}, Agus Muliantara^{b6}

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Bali, Indonesia

¹nimadeputriwahyuni29@email.com (Corresponding author)

²rahningputri@unud.ac.id

³anom.cp@unud.ac.id

⁴dewabayu@unud.ac.id

⁵made.agung@unud.ac.id

⁶muliantara@unud.ac.id

Abstract

Humans have voice characteristics with different vocal ranges, namely the male voice consists of Tenor, Baritone, and Bass, while the female voice consists of Soprano, Mezzosoprano, and Alto. Determining the voice range, especially for a singer, requires a vocal trainer or musical instrument that is quite difficult to access. Therefore, a sound classification system created based on vocal range using the Harmonic Product Spectrum (HPS) feature extraction method and the K-Nearest Neighbors (KNN) classification method uses k parameters from 1 to 40. The test gets the highest accuracy on parameter k=8, which is 88.88%, so that from the resulting accuracy to prove the K-Nearest Neighbor (KNN) method gives good results in classifying the type of voice.

Keywords: Classification, Vocal range, Harmonic Product Spectrum, K-Nearest Neighbors

1. Pendahuluan

Bagi penyanyi pemula dalam meng-cover sebuah lagu, banyaknya pilihan lagu dengan jangkauan nada yang bervariasi seringkali membuat orang yang memiliki hobi bernyanyi salah dalam memilih lagu. Hal ini biasanya yang membuat antara suara dan instrumen lagu tidak sesuai. Untuk melakukan cover terhadap lagu, seseorang harus menyesuaikan suara yang dimiliki dengan nada instrumen pada lagu tersebut. Oleh karena itu untuk mengatasi masalah tersebut diperlukan pengolahan suara yang dapat mendeteksi nada vokal dan mengidentifikasi jenis suara berdasarkan jangkauan vokal, sehingga dapat membantu para penyanyi pemula dalam meng-cover lagu[1].

Penelitian identifikasi jenis suara juga dilakukan oleh Annisa [2] dalam penelitian ini suara yang diteliti berfokus pada suara anak-anak, peneliti menggunakan metode Fast Fourier Transform (FFT). FFT dapat mengubah sinyal dari domain waktu ke domain frekuensi, metode ini digunakan untuk melihat frekuensi yang dihasilkan oleh suara anak tersebut kemudian diklasifikasikan dengan tipe suara. Namun dalam penelitian [2] ini hanya mengetahui frekuensi yang dihasilkan dari suara yang telah direkam menggunakan metode Fast Fourier Transform (FFT). Proses pengolahan data suara akan dihasilkan beberapa macam frekuensi suara, yaitu frekuensi untuk pelafalan suara a untuk mengetahui frekuensi nada dasar, suara do re mi fa sol la si do' digunakan untuk mendeteksi range frekuensi rata-rata suara dan suara do' digunakan untuk mendeteksi suara tertinggi yang dapat dijangkau. Dari hasil data frekuensi tersebut akan dilihat jangkauan frekuensi yang dihasilkan, yang nantinya akan digunakan untuk menentukan tipe suara[3].

Berdasarkan paparan penelitian yang telah dilakukan sebelumnya, penulis tertarik dalam mengklasifikasikan jenis suara berdasarkan jangkauan vokal dengan menggunakan metode klasifikasi KNN. Pemilihan metode K-NN dilakukan karena metode KNN merupakan metode yang paling sederhana, sehingga akan berpengaruh pada efisiensi waktu. Algoritma Harmonic Product Spectrum (HPS) digunakan sebagai metode ekstraksi fitur untuk melihat frekuensi yang terdapat pada sinyal suara, sinyal suara tersebut sebelumnya telah diubah ke domain frekuensi menggunakan Fast Fourier

Transform (FFT). Hasil dari ekstraksi fitur menggunakan metode HPS ini kemudian akan klasifikasi menggunakan klasifikasi *K-Nearest Neighbor* (KNN) untuk menentukan jenis suara, yaitu sopran, mezzo-sopran, alto, tenor, baritone, atau bass.

2. Metode Penelitian

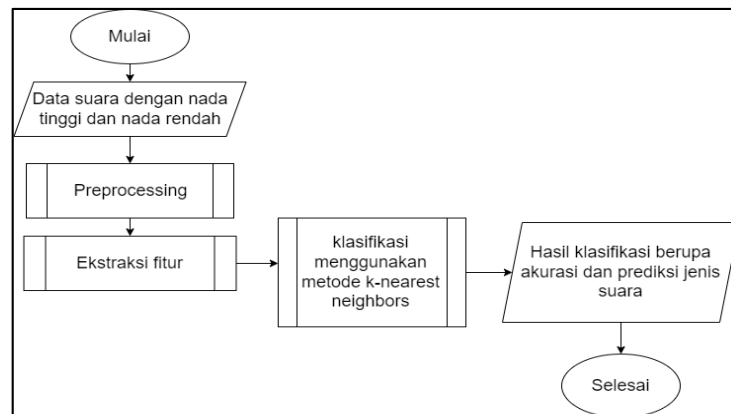
2.1. Pengumpulan Data

Dalam penelitian ini menggunakan data primer suara yang diperoleh langsung dari melakukan perekaman kepada 30 pria dan 30 wanita yang mewakili semua kelas, setiap partisipan memiliki rentang usia antara 15-23 tahun. Terdapat 10 orang dengan masing-masing jenis suara, setiap kelas telah divalidasi oleh guru vokal. Proses perekaman dilakukan dua kali kepada masing-masing partisipan, dengan melafalkan huruf vokal "a" dengan mengambil nada terendah dan nada tertinggi yang bisa dicapai oleh masing-masing partisipan. File suara yang direkam disimpan dalam format (.wav), *sample rate* yang digunakan yaitu dengan frekuensi sebesar 44100 Hz yang terdapat 44100 sampel per detik. Jumlah data yang didapatkan sebanyak 60, yang memiliki fitur dengan frekuensi fundamental nada tinggi dan frekuensi fundamental nada tinggi.

2.2. Perancangan Sistem

Tahapan pertama merupakan *input* data yang diambil dari hasil rekaman suara berupa melafalkan huruf vokal "a" dengan mengambil nada terendah dan nada tertinggi yang bisa dicapai setiap partisipan. Data rekaman menggunakan format "wav" dengan frekuensi *sampling* yang digunakan pada masing-masing suara adalah 44100 Hz. Kemudian amplitudonya dinormalisasikan, setelah itu sinyal dari stereo ke mono. Selanjutnya dilakukan tahap *windowing* dengan *hamming window* untuk mencegah kebocoran frekuensi ketika proses FFT. Keluaran dari proses FFT adalah *spectrum* frekuensi yang terdiri dari fundamental frekuensi dan harmoniknya. Spektrum ini yang dimasukkan ke dalam algoritma *Harmonic Product Spectrum* untuk dikompresi dan kemudian menentukan frekuensi dari nada yang terdeteksi pada sinyal suara.

Proses HPS akan mengalikan nilai spektrum frekuensi untuk menghilangkan sinyal harmonik pada frekuensi dan dapat mengenali frekuensi dari nada yang terdeteksi pada sinyal suara masukan. Klasifikasi KNN untuk melakukan pencocokan objek berdasarkan data yang paling dekat dengan objek tersebut. Secara umum, langkah-langkah perancangan sistem digambarkan dalam diagram alir seperti gambar 1.



Gambar 1. Diagram alir perancangan sistem.

2.3. Preprocessing

Dalam proses ini, masukan mengalami beberapa proses sehingga menghasilkan keluaran yang nantinya menjadi masukan untuk proses selanjutnya. Adapun tahap preprocessing terdiri dari tahap normalisasi yaitu menyamakan rentang amplitudo, selanjutnya mengubah sinyal audio menjadi monostereo.

2.3.1. Normalisasi Sinyal Suara

Proses ini bertujuan untuk menyamakan amplitudo dari setiap suara nyanyian yang direkam oleh sistem sehingga berada dalam rentang -1 dan +1. Proses normalisasi awal dilakukan dengan cara

membagi tiap nilai data masukan yaitu suara terekam dengan nilai absolut maksimal dari data masukan tersebut. Berikut rumus untuk proses normalisasi [4].

$$X_{norm} = \frac{x}{\max(|x|)} \quad (1)$$

Dimana :

X_{norm} = hasil data sinyal normalisasi

x = data input

2.3.2 Mengubah Stereo ke Mono

Pada tahap ini, data suara akan diubah dari stereo ke mono dengan mencari rata-rata dua saluran pada data stereo yang akan dikonversi menjadi hanya satu saluran (mono). Data yang semula adalah kanal stereo dan memiliki frekuensi sampling sebesar 44100 Hz dirubah menjadi kanal mono dengan cara mengubah data yang dua kanal menjadi satu kanal atau satu kolom matriks datanya saja. Tahap ini berguna untuk mempersingkat waktu kerja sistem karena mengolah data yang lebih sedikit. Hal ini tidak menghilangkan informasi karena matriks data kolom satu tidak jauh berbeda dengan matriks kolom dua[5].

2.4. Ekstraksi Fitur

Adapun tahap ekstraksi fitur terdiri dari tahap *windowing* menggunakan *hamming window*, kemudian dilanjutkan dengan tahap mengubah sinyal suara kedalam domain frekuensi dengan metode *Fast Fourier Transform* (FFT). Selanjutnya dilakukan tahap mencari frekuensi dari nada yang terdeteksi pada sinyal suara dengan menggunakan metode *Harmonic Product Spectrum* (HPS)[6].

2.4.1. Hamming Window

Hamming Window mempunyai *side lobe* yang paling kecil dan *main lobe* yang paling besar sehingga hasil *windowing* akan lebih halus dalam menghilangkan efek diskontinuitas. Ketika digunakan FFT untuk mengukur frekuensi dari data, maka harus ditentukan dasar analisa pada sebuah data yang pasti. FFT mengasumsikan bahwa data yang tetap adalah sebuah periode dari period signa [7]!. Secara matematis fungsi *hamming window* dituliskan dengan persamaan :

$$w(n) = \alpha - \beta \left(\frac{2\pi n}{N-1} \right) \quad (2)$$

Dimana:

$$\alpha = 0.54$$

$$\beta = 1 - \alpha$$

$$= 0.46$$

2.4.2. Fast Fourier Transform (FFT)

Proses selanjutnya adalah transformasi sinyal untuk menghasilkan spektrum dari sinyal hasil *windowing*, *spectrum* domain waktu diubah menjadi sinyal dalam domain frekuensi dengan menggunakan algoritma *Fast Fourier Transform* (FFT). Spektrum yang dibangkitkan inilah yang akan dianalisis untuk mengetahui frekuensi dari nada yang terdeteksi oleh sinyal suara. Proses FFT sinyal dibagi menjadi beberapa bagian yang lebih kecil yang bertujuan untuk memperoleh waktu proses yang lebih cepat. DFT adalah prosedur matematis yang digunakan untuk menentukan harmonik, atau frekuensi serta isi dari urutan sinyal diskrit [2]. Berikut rumus dasar *Discrete Fourier Transformation*:

$$F(k) = \sum_{n=0}^N f(n) \cos\left(\frac{j2\pi nk}{N}\right) - j f(n) \sin\left(\frac{j2\pi nk}{N}\right) \quad (3)$$

Untuk mendapatkan nilai j menggunakan persamaan seperti persamaan 3

$$|f(\mu)| = [R^2 + l^2]^{1/2} \quad (4)$$

Dimana :

N = jumlah sampel yang akan diproses.

x(n) = nilai sampel signal

k = variable frekuensi discrete

$\pi = 3,14$

n = indeks data nilai sampling

2.4.3. Harmonic product spectrum (HPS)

Harmonic Product Spectrum merupakan metode *pitch detection algorithm* yang paling simpel dan bekerja dengan baik dalam berbagai kondisi, metode HPS berfungsi untuk melihat frekuensi dasar yang terdapat pada sinyal input. HPS adalah suatu metode yang berfungsi untuk melihat frekuensi dasar yang terdapat pada sinyal input [8]. Secara matematis HPS dapat dirumuskan seperti persamaan 5

$$HPS(k) = (\prod_{n=1}^N Y(n k)) \quad (5)$$

Dimana:

HPS = hasil spektrum harmonik

k = indeks frekuensi spektrum harmonik

Y = besar spektrum pada frekuensi positif

N = jumlah harmonik yang digunakan

2.5. Klasifikasi *K-Nearest Neighbor* (KNN)

K-Nearest Neighbor (KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek yang berdasarkan dari data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Metode ini banyak digunakan dalam bidang pengenalan pola, klasifikasi KNN didasarkan pada membandingkan sebuah data uji dengan sejumlah data latih. Data latih terdiri dari n atribut, tiap data merepresentasikan titik pada sebuah ruang berdimensi n, dengan begitu semua data latih disimpan di dalam ruang pola berdimensi n. Ketika diberikan sebuah data yang tidak diketahui kelasnya, *k-nearest neighbor* akan mencari pola ruang untuk data latih k yang terdekat. Data latih k ini merupakan k "*nearest neighbor*" dari data yang tidak diketahui tersebut [9]. Jarak antara data latih dan data uji dihitung menggunakan persamaan Euclidean Distance [10]. *Euclidean Distance* dapat didefinisikan pada persamaan 6.

$$dist(X_1, X_2) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2} \quad (6)$$

Dimana :

dist = jarak kedekatan

x_1 = data latih

x_2 = data uji

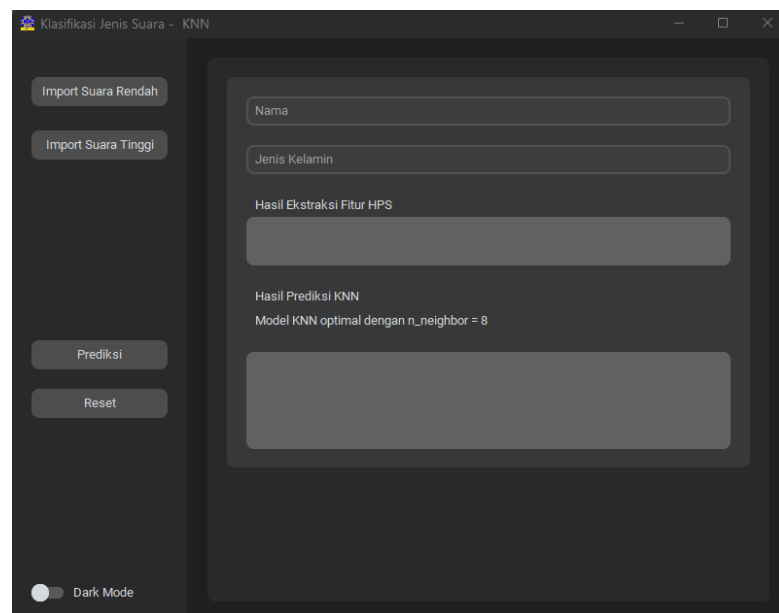
n = jumlah atribut antara 1 sampai n

i = atribut antara 1 sampai n

3. Hasil dan Pembahasan

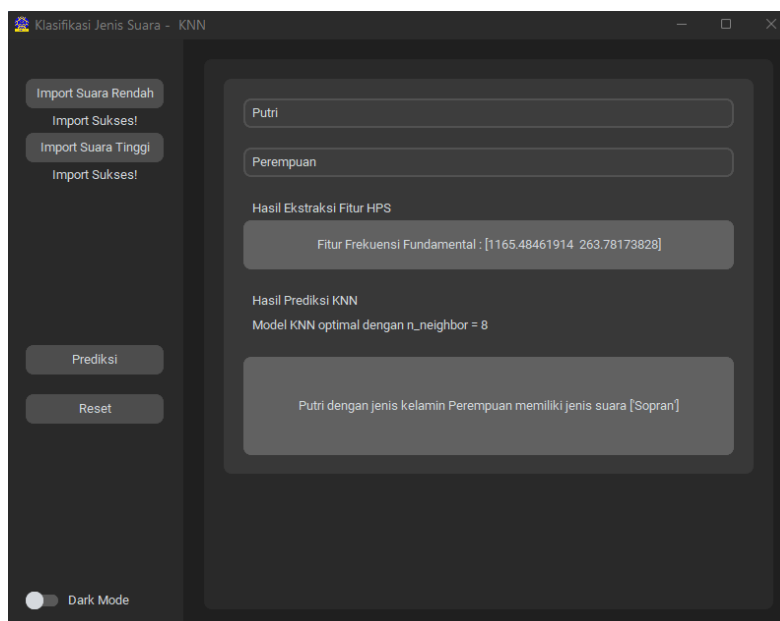
3.1. Tampilan Antar Muka Sistem

Antarmuka sistem terdiri dari satu halaman, Sistem menggunakan Bahasa pemrograman *python* berbasis GUI. Berikut pada gambar 2 merupakan tampilan antarmuka sebelum dijalankan.



Gambar 2. Implementasi antarmuka sistem

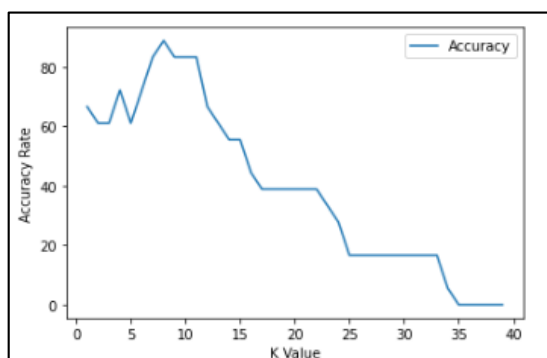
Pengguna dapat melakukan prediksi jenis suara dengan memasukan data suara dengan format .wav. Pengguna memasukan dua file suara yaitu suara terendah dan tertinggi yang bisa dicapai dengan nemakan tombol "Import Suara Rendah" untuk data suara rendah dan tombol "Import Suara Tinggi" untuk data suara tinggi. Setelah data sukses dimasukkan, pengguna mengisikan data diri berupa nama dan jenis kelamin. Selanjutnya pengguna menekan tombol "Prediksi" untuk menjalankan proses ekstraksi fitur dan prediksi. Keluaran dari hasil prediksi akan menampilkan frekuensi fundamental dari hasil ekstraksi fitur dan menampilkan prediksi jenis suara dari pengguna. Tombol "Reset" digunakan untuk mengatur ulang sistem. Implementasi antar muka sistem setelah dijalankan ditunjukkan pada gambar 3.



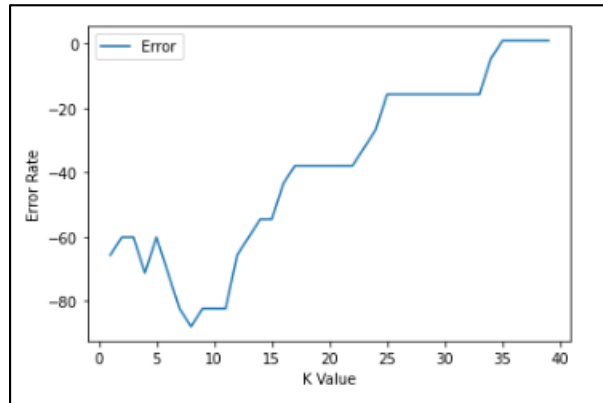
Gambar 2. Implementasi antarmuka sistem hasil

3.2. Pengujian Sistem

Hasil ekstraksi fitur terdiri dari dua fitur yaitu frekuensi tinggi dan frekuensi rendah, jumlah fitur HPS sebanyak 60 data. Pengujian jenis suara berdasarkan jangkauan vokal yaitu alto, mezzo, sopran, tenor, bariton, dan bass menggunakan ekstraksi fitur HPS dan klasifikasi KNN. Pembagian dataset antara data latih dan data uji yaitu 70:30. Jadi 70% merupakan data untuk pelatihan model sedangkan 30% merupakan data untuk pengujian model, dimana 42 data sebagai data uji dan 18 sebagai data latih. Penelitian ini menggunakan nilai k tetangga dari 1 sampai 40, berdasarkan pengujian pada setiap nilai k dilihat gambar 4, akurasi tertinggi diperoleh k=8 dengan akurasi sebesar 88.88%. Sedangkan dari k=14 sampai dengan k=40 memperoleh akurasi dibawah 60%.



Gambar 3. Grafik akurasi setiap parameter k

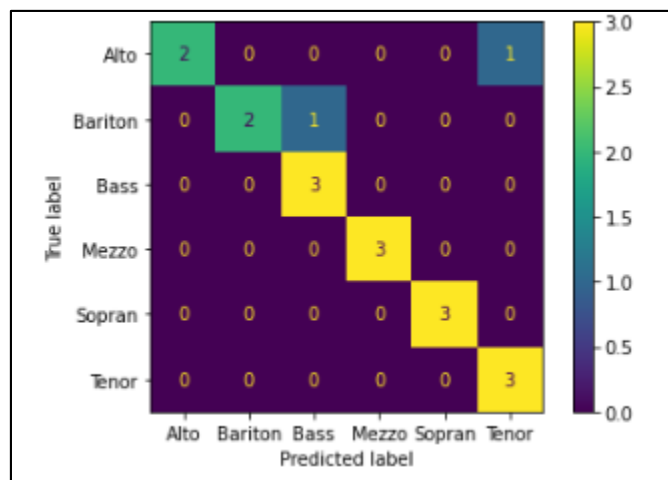


Gambar 4. Grafik Galat KNN

Berdasarkan hasil keseluruhan akurasi terlihat membentuk pola menurun dan terjadi pengulangan nilai akurasi pada beberapa pengujian nilai k, seperti pada pengujian k 17 – 22 dimana mendapatkan akurasi yang sama yaitu 38.8 % dan pengujian k 25 – 33 mendapatkan akurasi yang sama yaitu 16.6 %. Pengujian ini membuktikan bahwa model klasifikasi menggunakan metode KNN pada klasifikasi jenis suara berdasarkan jangkauan vokal memiliki pola menurun terlihat pada gambar 5, dimana semakin besar nilai k maka semakin kecil akurasi yang didapatkan.

Pengujian dilakukan dengan membagi dataset yang berjumlah 60 data menjadi dua bagian yang terdiri atas 48 data sebagai data training dan 18 data sebagai data uji. Proporsi jumlah data untuk masing-masing kelas data pada data training dan data uji dibuat 70% dan 30%. Proses training dilakukan dengan menggunakan metode *K-nearest neighbor* dengan nilai parameter k = 1 sampai dengan k = 40. Berdasar hasil pengujian model, dapat dilihat bahwa model yang terbaik adalah model dengan nilai parameter k=8. Pertimbangannya adalah model ini selain memiliki tingkat akurasi yang baik yaitu sebesar 88.88% .

Sehingga berdasarkan pengujian model dengan akurasi terbaik sebesar 88.88% pada parameter k = 8, hal ini dapat dibuktikan dari confusion matriks ditampilkan pada gambar 6, dapat dilihat gambar 6 pada setiap pengujian terdapat 3 data yang menjadi data uji setiap kelas. Untuk kelas dengan jenis suara alto terdeteksi benar sebanyak 2 dan 1 data terdeteksi dengan jenis suara tenor. Untuk kelas bariton terdeteksi benar sebanyak 2 dan 1 data terdeteksi dengan jenis suara bass. Untuk data dengan kelas bass, mezzo, sopran dan tenor semua terdeteksi dengan benar.



Gambar 5. Hasil prediksi k=8

4. Kesimpulan

Pada penelitian ini mengklafikasikan jenis suara berdasarkan jangkauan vokal menggunakan metode ekstraksi fitur *Harmonic Product Spectrum* (HPS) dan metode klasifikasi *K-Nearest Neighbor*

(KNN) dengan pengujian parameter nilai k, parameter k yang digunakan dari 1 sampai 40. Berdasarkan hasil penelitian dapat disimpulkan hasil sebagai berikut.

1. Akurasi tertinggi yang dihasilkan dari pengujian klasifikasi jenis suara berdasarkan jangkauan vokal menggunakan metode klasifikasi KNN dengan nilai $k = 8$ yaitu sebesar 88.88%, sehingga dari akurasi yang dihasilkan membuktikan metode *K-Nearest Neighbor* (KNN) memberikan hasil yang baik dalam mengklasifikasikan jenis suara.
2. Akurasi terendah pada $k = 14$ sampai $k=40$ dengan akurasi dibawah 60%, sehingga nilai k yang tinggi kurang bagus digunakan dalam mengklasifikasikan jenis suara.
3. Pengujian ini juga membentuk pola menurun antara akurasi dengan parameter k, dimana semakin tinggi parameter k maka akurasi semakin rendah.

Daftar Pustaka

- [1] F. Rifqi, *Aplikasi Rekomendasi Transpose Lagu Berdasarkan Vocal Range Menggunakan Algoritma Genetika Berbasis Android Aplikasi Rekomendasi Transpose Lagu Berdasarkan Vocal Range Menggunakan Algoritma Genetika Berbasis Android Skripsi*. 2019.
- [2] R. Nugra Annisa, D. Suprayogi, and H. Bethaningtyas, "Klasifikasi Suara Anak-Anak Dengan Menggunakan Metode Fast Fourier Transform Classification of Children'S Voices Using Fast Fourier Transform," vol. 6, no. 1, pp. 1141–1148, 2019.
- [3] P. D. Prasetyo, I. G. P. Suta Wijaya, and A. Yudo Husodo, "Klasifikasi Genre Musik Menggunakan Metode Mel-Frequency Cepstrum Coefficients dan K-Nearest Neighbors Classifier," *J. Teknol. Informasi, Komputer, dan Apl. (JTika)*, vol. 1, no. 2, pp. 189–197, 2019, doi: 10.29303/jtika.v1i2.41.
- [4] I. Wijayanto and R. Dwifebrianti, "Jenis Tipe Jangkauan Suara Pada Pria Dan Wanita Menggunakan Metoda Mel-Frequency Cepstral Coefficient," *Konfrensi Nas. Sist. dan Inform.*, no. October 2013, pp. 2–10, 2013.
- [5] G. Harsemadi, M. Sudarma, and N. Pramaita, "Implementasi Algoritma K-Nearest Neighbor pada Perangkat Lunak Pengelompokan Musik untuk Menentukan Suasana Hati," *Maj. Ilm. Teknol. Elektro*, vol. 16, no. 1, pp. 14–20, 2017, doi: 10.24843/mite.1601.03.
- [6] W. S. M. Sanjaya and Z. Salleh, "Implementasi Pengenalan Pola Suara Menggunakan Mel-Frequency Cepstrum Coefficients (MFCC) dan Adaptive Neuro-Fuzzy Inferense System (ANFIS) sebagai Kontrol Lampu Otomatis," *AI-HAZEN J. Phys.*, vol. 1, no. 1, 2014, [Online]. Available: <http://journal.uinsgd.ac.id/index.php/ahjop/article/view/129>.
- [7] J. Adler, M. Azhar, and S. Supatmi, "Identifikasi Suara dengan MATLAB sebagai Aplikasi Jaringan Syaraf Tiruan Speech Recognition in MATLAB as Artificial Neural Network Application," vol. 1, no. 1, pp. 16–23, 2013.
- [8] V. A. Egisthi, D. Andreswari, and Y. Setiawan, "Aplikasi Latih Vokal Dengan Menggunakan Metode Harmonic Product Spectrum (Hps) Dan Boyer Moore Berbasis Android," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 7, no. 2, p. 501, 2016, doi: 10.24176/simet.v7i2.761.
- [9] F. Darmadi, A. Rizal, and U. Sunarya, "Deteksi Sleep Apnea Melalui Analisis Suara Dengkuran Dengan Metode Mel Frekuensi Cepstrum," vol. 2, no. 2, pp. 2681–2686, 2015.
- [10] E. Susanti, S. Sasongko, and I. G. Suta W, "Klasifikasi Suara Berdasarkan Usia Menggunakan Mel Frequency Cepstral Coefficient (Mfcc)," vol. 4, no. 2, pp. 120–126, 2017.

This page is intentionally left blank