

KLASIFIKASI EMOSI BERDASARKAN SINYAL SUARA MANUSIA MENGUNAKAN METODE KEY-NEAREST NEIGHBOR (K-NN)

CALSSIFICATION OF EMOTIONAL BASED ON HUMAN VOICE SIGNAL USING KEY-NEAREST NEIGHBOR (K-NN)

Fajar Dwi Septria¹, Dr. Ir. Jangkung Raharjo, M.T, Nur Ibrahim, S.T.,M.T.³

^{1,2,3} Prodi S1 Teknik Telekomunikasi, Fakultas Teknik Elektro, Universitas Telkom

¹fajardwiseptria.fds@gmail.com, ²nuribrahim@telkomuniversity.ac.id,

³jangkung.raharjo@gmail.com

Abstrak

Salah satu bentuk interaksi antar manusia adalah dengan berdialog atau berbicara. Biasanya interaksi antar manusia tidak selalu baik dikarenakan pengaruh beberapa faktor seperti berbeda pendapat, harapan tidak sesuai kenyataan atau tidak selalu buruk dikarenakan faktor suasana hati yang sedang baik. Hal-hal tersebut pasti akan sangat mempengaruhi kepada emosi seseorang, emosi tersebut dapat ditentukan melalui sinyal suara. Emosi yang akan diidentifikasi pada penelitian ini adalah bahagia, marah, sedih, dan kaget. Sinyal suara direpresentasikan dengan fitur *Linear Predictive Coding* (LPC). Performa dari sistem akan diukur berdasarkan akurasi dalam ketepatan deteksi emosi. penggunaan metode klasifikasi yaitu *Key-Nearest Neighbor* (K-NN) Pemilihan metode ini akan diujikan dengan objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut, beberapa aturan jarak pada K-NN juga mempengaruhi terhadap akurasi sistem pengujian. Hasil pengujian menunjukkan bahwa akurasi tertinggi yang didapatkan adalah 92.5% dengan menggunakan 6 *feature* statistik LPC yaitu *mean*, *variance*, standar deviasi, *skewness*, *kurtois*, *entropy* dan menggunakan *Distance Cityblock* pada K-NN dari 100 data latih dan 40 data uji. Parameter terbaik yang didapatkan adalah 14 panjang matriks maksimum, 22 *window cepstral*, dan parameter *k* adalah 1 (satu).

Kata kunci : *K-Nearest Neighbor* (K-NN), *Linear Predictive Coding* (LPC), Emosi.

Abstract

One form of interaction between humans is by talking or dialogue. The interaction between humans is not always going well because it influenced by several factors such as differing opinions, expectations not in reality or not always bad due to good mood factors. These things will definitely affect a person's emotions, these emotions can be determined through sound signals. The emotions that will be identified in this study are happy, angry, sad, and shocked. Sound signals are represented by the *Linear Predictive Coding* (LPC) feature. The performance of the system will be measured based on accuracy in the accuracy of emotional detection. This research is proposed using the *Key-Nearest Neighbor* (K-NN) classification method. The selection of this method will be tested with objects based on learning data that is the closest distance to the object, some rules of distance on K-NN also affect the accuracy of the testing system. The result of the test showed that the highest accuracy obtained was 92.5% using 6 LPC statistical features, which contain mean, variance, standard deviation, skewness, kurtois, entropy and *Distance Cityblock* on K-NN from 100 training data and 40 test data. The best parameters obtained are 14 maximum matrix lengths, 22 *cepstral* windows, and the *k* parameter is 1 (one).

Keywords : *K-Nearest Neighbor* (K-NN), *Linear Predictive Coding* (LPC), Emotion.

1. Pendahuluan

Salah satu bentuk interaksi antar manusia adalah dengan berdialog atau berbicara. Biasanya interaksi antar manusia tidak selalu baik dikarenakan pengaruh beberapa faktor seperti berbeda pendapat, harapan tidak sesuai kenyataan, atau tidak selalu buruk dikarenakan faktor suasana hati yang sedang baik. Hal-hal tersebut pasti akan sangat mempengaruhi kepada emosi seseorang. Deteksi emosi dari sinyal suara merupakan bidang dari *speech recognition*, beberapa penelitian menunjukkan adanya hubungan yang erat antara beberapa satuan statistik ucapan dan emosi dari pembicara. Diantara satuan tersebut yang paling populer adalah nada, energi, artikulasi, dan bentuk

spektral [1]. Satuan tersebut dapat membentuk suatu ciri yang melambangkan suatu emosi apakah dia sedang merasakan bahagia, marah, sedih, kaget. Ciri tersebut dapat diklasifikasi dengan beberapa metode yang sudah diujikan sebelumnya.

Pada tugas akhir ini akan dilakukan deteksi dan penentuan keadaan emosi seseorang melalui sinyal bicara manusia menggunakan ekstraksi ciri *Linear Predictive Coding* (LPC) kemudian akan diklasifikasikan menggunakan *K-Nearest Neighbor* (K-NN). LPC sudah lama digunakan pada teori komunikasi, saat ini LPC ditemukan pada aplikasi pengenalan analisis dan perpaduan, pengenalan pembicara, dan pengenalan kata. *Linear Predictive* adalah teknik untuk menganalisis data persamaan waktu yang didapat dari analisis sistem linear [2]. Dengan menggunakan *linear predictive coding*, parameter dari sistem tersebut dapat ditentukan dengan menganalisis masukan dan keluaran sistem.

K-Nearest Neighbor (K-NN) merupakan pengelompokan suatu data baru berdasarkan jarak data ke beberapa data. Metode ini dapat menyederhanakan algoritma perhitungan sehingga berpengaruh pada efisiensi waktu. Penulis akan mencoba melakukan penelitian menggunakan metode *K-Nearest Neighbor* guna untuk menentukan emosional melalui sinyal suara. Selanjutnya, penulis akan menyimpulkan hasil dari penelitian apakah bisa metode tersebut di implementasikan dan mendapatkan hasil akurasi yang lebih baik dari penelitian yang sudah ada.

2. Konsep Dasar

2.1 Sinyal Suara Manusia

Suara manusia merupakan sinyal yang dihasilkan dari bergetarnya pita suara. Suara tersebut merupakan representasi dari pesan yang ingin disampaikan oleh otak kita. Pita suara manusia bergetar akibat adanya aliran udara dari paru-paru, dan getaran itu akan menghasilkan gelombang bunyi. Suara yang kita bunyikan akan bergantung dengan bagaimana kita meletakkan posisi lidah, gigi, dan rahang atau yang sering disebut dengan articulator, sehingga akan menghasilkan bunyi-bunyi vokal tertentu [6].

2.2 Teori Emosi Manusia

Emosi yang seseorang rasakan bukan hanya berupa satu bentuk saja, melainkan bisa dibedakan menjadi beragam istilah yang lebih cocok untuk menggambarkan apa yang dirasakan saat itu [7]. Macam-macam jenis emosi yang ada yaitu:

1. Bahagia

Emosi bahagia akan dirasakan apabila seseorang merasa senang, dan itu berarti ada suatu hal yang menyenangkan hatinya.

2. Marah

Ketika sesuatu kehendak atau harapan seseorang terhadap suatu hal tidak terpenuhi karena adanya hambatan tertentu, maka bisa saja emosi alami yang muncul adalah rasa marah.

3. Sedih

Sedih atau kesedihan adalah perasaan manusia yang menyatakan kecewa atau frustrasi terhadap seseorang atau sesuatu.

4. Kaget

Emosi berupa perasaan kaget akan dirasakan apabila seseorang tidak mempunyai persiapan atau tidak mengetahui apa yang akan terjadi.

2.4 Key-Nearest Neighbor

K-Nearest Neighbor (K-NN) adalah sebuah metode yang digunakan untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut [3]. Algoritma ini merupakan jenis *instance based algorithm* sehingga pada proses klasifikasi tidak menyimpan model dan hanya berdasarkan memori. Metode ini menyederhanakan algoritma perhitungan sehingga sudah berpengaruh pada efisiensi waktu. Untuk menentukan label data *testing* maka dilakukan pengambilan sebanyak k data latih terdekat. Parameter k sendiri adalah jumlah tetangga terdekat yang dilibatkan dan memiliki pengaruh dalam menentukan hasil prediksi [4].

Pada K -NN terdapat beberapa aturan jarak yang dapat digunakan, yaitu [4]:

1. *Eucliden Distance*, dengan rumus:

$$J(v_1, v_2) = \sqrt{\sum_{k=1}^N (v_1(k) - v_2(k))^2} \quad (1)$$

2. *Cityblock* atau *manhattan distance*, dengan rumus:

$$J(v_1, v_2) = \sum_{k=1}^N |v_1(k) - v_2(k)| \quad (2)$$

3. *Cosine*, dengan rumus:

$$\text{Cos}(N_i, N_j) = \frac{\sum_k a_{i,k} \cdot a_{j,k}}{\sqrt{\sum_k a_{i,k}^2} \sqrt{\sum_k a_{j,k}^2}} \quad (3)$$

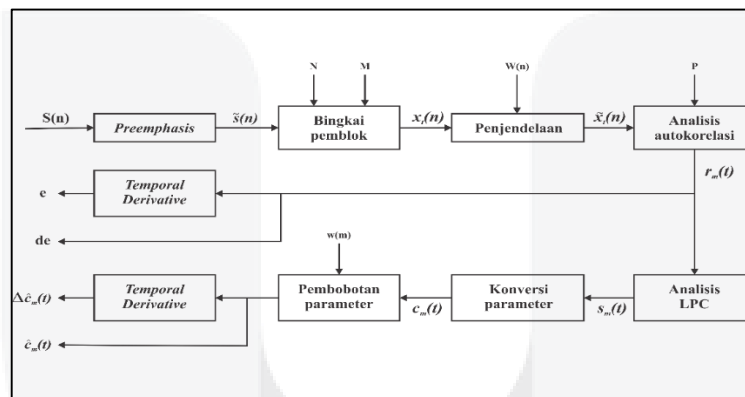
4. *Minkowski*, dengan rumus:

$$J(v_1, v_2) = \sqrt[\lambda]{\sum_{k=1}^N |v_1(k) - v_2(k)|^\lambda} \quad (4)$$

Dimana, v_1, v_2 merupakan data latih dan data uji, $J(v_1, v_2)$ adalah jarak J dari v_1, v_2 , N adalah nilai dimensi data, dan k adalah jumlah data dari tetangga terdekat.

2.4 Linear Predictive Coding

Sinyal suara yang diambil langsung dari rekaman diproses untuk mengekstraksi fitur. Metode yang digunakan untuk proses ekstraksi ciri ini adalah *Linear Prediction Coding* (LPC). Diagram blok ekstraksi ciri dapat dilihat pada gambar 2.5 dengan langkah-langkah analisa dari LPC [5].



Gambar 2.4 Diagram Blok Ekstraksi Ciri linear predictive Coding

Terdapat 6 langkah-langkah analisis LPC untuk pengenalan ucapan, yaitu [5]:

1. *Preemphasis*

Filter ini mempertahankan frekuensi-frekuensi tinggi pada sebuah *spectrum*, yang umumnya tereliminasi pada saat produksi suara. Persamaan *preemphasizer* yang umum digunakan adalah:

$$y(n) = s(n) - \alpha s(n-1) \quad (5)$$

Dimana, $y(n)$ adalah sinyal *preemphasis* filter, $s(n)$ adalah sinyal sebelum *preemphasis* filter dan α merupakan konstanta filter *preemphasis*, biasanya bernilai 0,97. Dalam bentuk dasar operator s sebagai unit filter, persamaan di atas akan memberikan sebuah transfer *function* filter *preemphasis* seperti berikut pada langkah ini sinyal yang telah di *preemphasis* $s(n)$ akan diblok dalam *frame-frame* pada tahap *frame blocking*.

2. *Frame Blocking*

Frame blocking digunakan untuk memotong-motong sinyal suara menjadi beberapa *frame* agar dapat diproses secara *short-time* untuk memperoleh karakter frekuensi yang relatif stabil.

sinyal bicara yang telah di *preemphasis*, diblok dalam *frame-frame* dengan N sample dan digeser sebesar M sample di mana : $N = 2 \times M$, sehingga didapat nilai dari sinyal yang baru adalah:

$$xl(n) = y(M.l + n) \quad (6)$$

Dimana, M adalah *overlapping frame*, n dan l adalah angka yang dimulai dari 0,1,2,3...dan x(n) adalah nilai dari *frame blocking*.

3. Windowing

Pada langkah ini dilakukan fungsi *weighting* pada setiap frame yang telah dibentuk pada langkah sebelumnya dengan tujuan untuk meminimalkan *discontinuities* pada ujung awal dan ujung akhir setiap *frame* yaitu dengan taper sinyal menuju nol pada ujung-ujungnya. Tipikal *window* yang digunakan pada metode auto korelasi LPC adalah *hamming window* yang memiliki bentuk,

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) \quad (7)$$

Dimana, n adalah angka yang dimulai dari 0,1,2,3..., w(n) adalah fungsi *window*, N adalah jumlah sampel dalam masing-masing frame yang merupakan kelipatan 2 dan π adalah bernilai 3,14.

4. Analisa auto korelasi

Pada tahap ini masing-masing frame yang telah di windowing di auto korelasikan untuk mendapatkan:

$$rl(m) = \sum_{n=0}^{N-1-m} \tilde{x}1(n) \cdot \tilde{x}1(n+m) \quad (8)$$

Dimana, rl(m) adalah hasil analisa auto korelasi, N adalah jumlah sampel, m adalah orde dari analisa LPC dan $\tilde{x}1$ adalah nilai dari *frame blocking*.

5. Analisa LPC

Langkah selanjutnya adalah analisa LPC, di mana pada tahap ini p+1 auto korelasi pada setiap *frame* diubah menjadi satu set LPC parameter,

$$E^0 = r(0) \quad (9)$$

$$k = \frac{r(i - \sum_{j=1}^{i-1} a_j^{i-1} r(i-j))}{E^{i-j}} \quad (10)$$

$$a_i^{(i)} = k_i \quad (11)$$

$$a_j^{(i)} = a_j^{i-1} - k_i a_{i-j}^{i-1} \quad (12)$$

$$E^i = (1 - k_i^2) E^{i-1} \quad (13)$$

Dimana, a_m adalah koefisien LPC, k_m adalah koefisien PARCOR (koefisien pantulan), g_m adalah koefisien perbandingan daerah *logaritmis* dan m adalah orde dari analisa LPC *frame blocking*.

6. Mengubah LPC Parameter ke Koefisien *Cepstral*

Sekelompok LPC parameter yang sangat penting yang dapat diperoleh dari penurunan koefisien LPC adalah koefisien cepstral $c(m)$. Persamaan yang digunakan adalah

$$C_m = a_m + \sum_{k=1}^{m-1} \left(\frac{k}{m}\right) \cdot C_m a_m - k \quad (14)$$

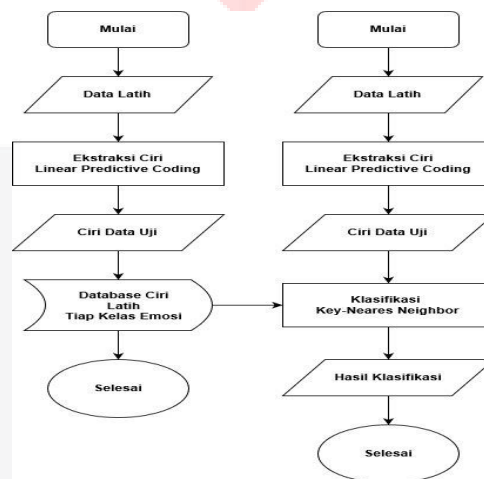
Dimana, C_m adalah koefisien *cepstral*, a_m adalah koefisien LPC, m adalah orde dari analisa LPC dan k_m adalah koefisien pantulan

3 Perancangan3

A. Desain Sistem

Metodologi penelitian yang dilakukan pada proposal ini menggunakan pendekatan eksperimental. Secara garis besar, sistem deteksi ini dibagi menjadi 4 tahap utama, yaitu akuisisi atau pengambilan data suara menggunakan *voice recorder*, ekstraksi ciri dan pembuatan database menggunakan data latih, deteksi emosi melalui sinyal bicara menggunakan data uji, dan menganalisis performansi sistem melalui akurasi sistem.

Diagram blok data uji dan tahap data latih dapat dilihat pada gambar dibawah ini.



Gambar 5 Diagram Alir Sistem

Proses latih merupakan proses pembentukan sistem klasifikasi berdasarkan data latih sebagai acuan. Sedangkan proses uji merupakan proses sesungguhnya sistem yang telah dirancang pada proses latih untuk mengklasifikasi jenis emosi dari data uji yang dipilih pada proses uji. Proses latih dan data uji secara garis besar sama hanya saja pada proses latih berarti membangun sistem klasifikasi menggunakan *K-Nearest Neighbor* (K-NN) yang menghasilkan model klasifikasi K-NN yaitu parameter K-NN yang digunakan pada proses uji klasifikasi.

B. Performansi Sistem

Akurasi adalah ukuran ketepatan sistem dalam mengenali input yang akan diberikan sehingga menghasilkan keluaran yang benar. Secara sistematis dapat dituliskan sebagai berikut:

$$A = \frac{N_B}{N_T} \times 100\% \quad (15)$$

Dimana A adalah akurasi, N_B adalah jumlah data benar, N_T adalah jumlah data keseluruhan.

Waktu komputasi adalah waktu yang dibutuhkan sistem melakukan suatu proses. Secara sistematis dapat dituliskan sebagai berikut:

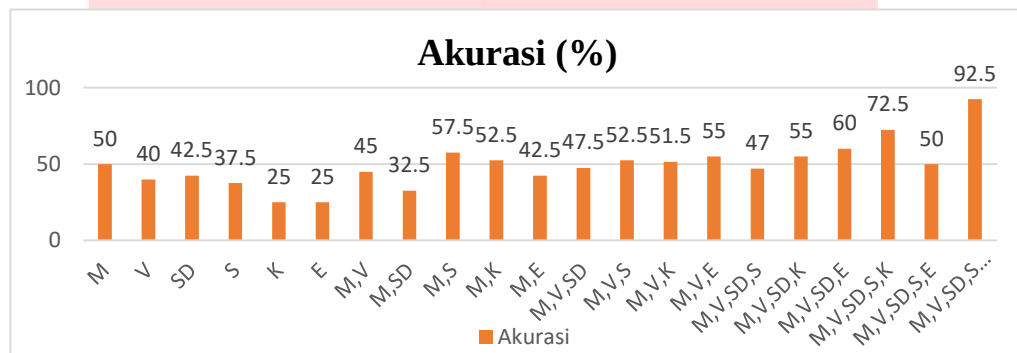
$$W_k = W_s - W_m \quad (16)$$

Dimana w_k adalah waktu komputasi, w_s adalah waktu selesai, w_m adalah waktu mulai.

4. Hasil Pengukuran

A. Pengujian Pengaruh Akurasi *Feature* Statistik Terhadap Ekstraksi Ciri LPC

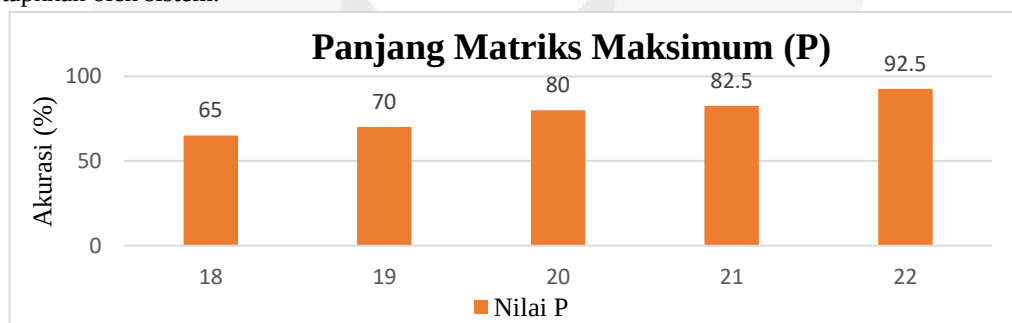
Pada skenario ini pengujian data latih dilakukan terhadap data yang telah dilakukan *training* atau pembelajaran terlebih dahulu. Pengujian ini bertujuan untuk mengetahui ciri yang paling tepat digunakan pada proses ekstraksi ciri berdasarkan nilai pengujian. Data latih yang digunakan pada sistem berjumlah 100 dan data uji berjumlah 40 pada masing-masing fungsi objektif yang ditetapkan. Hasil pengujian akurasi dan waktu komputasi dapat dilihat Gambar 4.



Gambar 4 Grafik Performansi Akurasi Pengujian

B. Pengujian Pengaruh Panjang Matriks Maksimum Terhadap Akurasi dan Parameter Pada Klasifikasi K-NN

Pengujian pengaruh dan Panjang Matriks Maksimum (P) terhadap akurasi pada klasifikasi K-NN merupakan pengujian yang dilakukan terhadap ekstraksi ciri yang telah dilakukan *training* atau pembelajaran terlebih dahulu. Pengujian ini dilakukan untuk mengetahui tingkat akurasi yang terbaik. Parameter yang digunakan pada pengujian ini adalah 1 s/d 20 batasan *range* atau fungsi kendala dan telah ditetapkan jumlah individu yang akan muncul setiap perulangan sebanyak 5 individu. Dimana pada sistem telah ditetapkan bahwa dapat disebut titik konvergen pada fungsi kendala, ketika terjadi pengulangan sebanyak $N+1$. Dimana N merupakan jumlah perulangan yang ditetapkan oleh sistem.

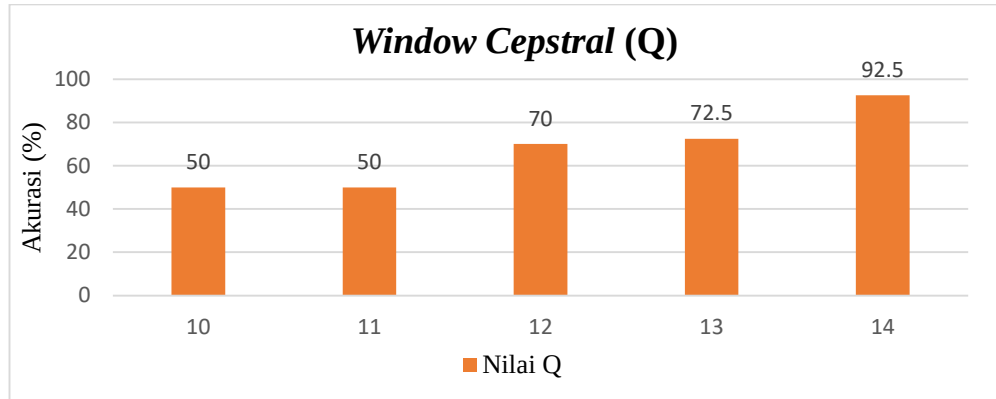


Gambar 5 Grafik Performansi Akurasi Pengujian.

C. Pengujian Pengaruh Panjang *Window Cepstral* Terhadap Akurasi dan Parameter Pada Klasifikasi K-NN

Pengujian pengaruh *window cepstral* (Q) terhadap akurasi pada klasifikasi K-NN merupakan pengujian yang dilakukan terhadap ekstraksi ciri yang telah dilakukan *training* atau pembelajaran terlebih dahulu. Pengujian ini dilakukan untuk mengetahui tingkat akurasi yang terbaik. Parameter yang digunakan pada pengujian ini adalah 1 s/d 20 batasan *range* atau fungsi kendala dan telah ditetapkan jumlah individu yang akan muncul setiap perulangan sebanyak 5 individu. Dimana pada

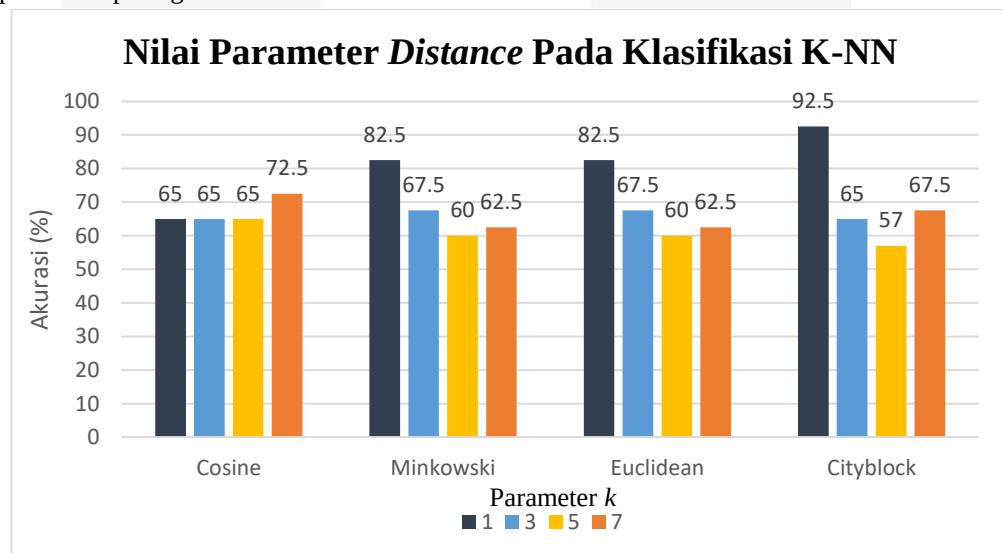
sistem telah ditetapkan bahwa dapat disebut titik konvergen pada fungsi kendala, ketika terjadi pengulangan sebanyak $N+1$. Dimana N merupakan jumlah perulangan yang ditetapkan oleh sistem. Pada Gambar 6 menunjukan hasil akurasi terbaik terdapat pada *window cepstral* (Q) bernilai 14 yang memiliki dataset benar berjumlah 37 dan dataset salah berjumlah 3 dengan akurasi 92.5%. *Window Cepstral* bernilai 14 ini mengalami perulangan sebanyak $N+1$.



Gambar 6 Hasil Pengujian akurasi *window cepstral*.

D. Pengujian Pengaruh Parameter *Distance* pada K-NN Terhadap Akurasi Sistem

Pada skenario ini, akan dilakukan pengujian pengaruh nilai parameter k dan aturan *distance* pada K-NN. Nilai k yang akan diujikan adalah 1, 3, 5, 7 dengan parameter *distance euclidean*, *cityblock*, *cosine*, dan *minkowski*. Dalam pengujian ini parameter statistik yang digunakan adalah *mean*, *variance*, standar variansi, *skewness*, *kurtois*, *entropy* dan juga menggunakan *overlapping*. Hasil akurasi dan waktu komputasi parameter *distance* dan nilai parameter k pada pengujian tahapan ini dapat dilihat pada gambar 7.



Gambar 7 Diagram Performansi Akurasi

Pada skenario ini didapatkan nilai akurasi tertinggi yaitu 92.5% dengan waktu komputasi 0,1194 detik terdapat pada *cityblock* dan parameter k adalah 1. Akurasi terendah bernilai 65% dengan waktu komputasi 0,1408 terdapat pada *cosine*. Nilai *cityblock* mendapat akurasi terbaik karena *cityblock* memiliki perhitungan yang lebih rumit sehingga lebih akurat dalam menentukan jarak tetangga terdekat setiap data.

5. Kesimpulan

Dari hasil pengujian dan analisis yang telah dilakukan pada sistem deteksi pulpitis reversibel melalui sinyal wicara, dapat diambil kesimpulan sebagai berikut:

- Berdasarkan hasil pengujian sistem yang dilakukan serta analisis sistem dalam penelitian klasifikasi emosi berdasarkan sinyal suara, dapat diambil kesimpulan bahwa sistem yang telah dirancang mampu mengenali 4 jenis suara emosi yaitu; bahagia, marah, sedih, kaget dengan menggunakan metode ekstraksi ciri *Linear Predictive Co-ding* (LPC) dan metode klasifikasi *Key-Nearest neighbor* (K-NN).
- Pada metode ekstraksi ciri LPC menggunakan *feature* statistik untuk masing-masing data suara emosi yang telah diakuisisi. Hasil pengujian akurasi dan waktu komputasi bahwa *feature* statistik yang sangat baik digunakan adalah 6 *feature* statistik yaitu: *mean*, *variance*, standar deviasi, *skewness*, *kurtosis*, *entropy*. Ada 2 fungsi objek yang telah ditetapkan adalah *window cepstral* dan panjang matriks maksimum. Akurasi terbaik dengan *window cepstral* 14 dan panjang matriks maksimum 22 dan menggunakan 6 *feature* statistik tersebut serta 0.1105 detik untuk waktu komputasi.
- Pada metode klasifikasi K-NN hasil ekstraksi ciri yang telah didapatkan dengan melihat 6 *feature* statistik setiap suara emosi dapat melakukan klasifikasi untuk mendapatkan akurasi dan parameter terbaik. Pada klasifikasi ini, melakukan pengujian terhadap objek berdasarkan data ekstraksi ciri yang jaraknya paling dekat dengan objek tersebut. Akurasi terbaik dicapai dengan menggunakan aturan jarak K-NN *Cityblock*.
- Jadi performansi terbaik yang telah diperoleh dari semua pengujian adalah akurasi tertinggi 92.5% dengan waktu komputasi 0.1105 detik, pada kondisi 37 suara emosi teridentifikasi sesuai dengan kondisi emosi yang sebenarnya.

Daftar Pustaka

- [1] Nogueiras, A., Marino, J.B., Moreno, A., Bonafonte, A, Speech Emotion Recognition using Hidden Markov Model, In: Proc. European Conf. Speech Communication and Technology (Euro Speech 01). Denmark, 2001.
- [2] A.Yahya and Suryanto, Deteksi Emosi Melalui Pengenalan Suara Menggunakan *Linear Predictive Coding* (LPC) Dan *Hidden Markov Model* (HMM), Tugas Akhir, 2012.
- [3] M. Walid dan A. K. Darmawan, "ISNN: 2088-4591," *Pengenalan Ucapan Menggunakan Metode Linear Predictive Coding (LPC) Dan K-Nearest Neighbor (K-NN)*, vol. 7, no. 1, PP. 13-22-2017.
- [4] R. Nurpalida, "Analisis dan Implementasi Prediksi Churn pada Perusahaan Telekomunikasi Menggunakan Metode NCL DAN K-Nearest Neighbor". Bandung 2017.
- [5] H. Saputra, Sistem Pengenalan Kata dengan Menggunakan, vol. 5, no. September, pp. 1924, 2005.
- [6] McIoughlin, Ian, *Applied Speech and Audio Processing : With Matlab Example* New York: Cambridge University Press, 2009.
- [7] 10 Jenis Emosi Pada Manusia dalam Psikologi – DosenPsikologi., <https://dosenpsikologi.com/jenis-emosi>.

