OPEN ACCESS
Indo-JC
Indonesia Journal of Computing
ISSN 2460-9056
socj.telkomuniversity.ac.id/indojc

Ind. Journal on Computing
Vol. 4, Issue. 1, March 2019. pp. 9-18
doi:10.21108/indojc.2019.4.1.244

Klasifikasi Gender Berdasarkan Suara Menggunakan *Support Vector Machine*

Raditya Budi Handoko [1], Suyanto [2]

School of Computing, Telkom University
Jl. Telekomunikasi Terusan Buah Batu, Bandung, West Java 40257, Indonesia

[1] radityabh@students.telkomuniversity.ac.id, [2]suyanto@telkomuniversity.ac.id

Abstract

A speech gender classification is important in speech recognition and many other applications. The researchers have proposed some methods to develop it, but it is commonly poor for noised speech. In this research, a speech gender classification system is developed and tested for some noise levels. The system contains two steps. First, a feature vector is generated using Mel-Frequency Cepstral Coefficient (MFCC). The feature vector is then classified using Support Vector Machine (SVM). Five-fold cross-validating on 2,264 utterances, which are equally split into two classes: "male" and "female", shows that the highest accuracy of 100% is achieved using SVM with Polynomial kernel and degree 1 for clean-speech. The system is robust, where the accuracy slightly decrease to 95.21%, for low noised speech of -25 dB. But, the accuracy decreases significantly for higher noised speeches.

Keywords: speech gender classification, speech recognition, Mel-Frequency Cepstral Coefficient, Support Vector Machine.

Abstrak

Klasifikasi gender berdasarkan suara sangat penting dalam sistem pengenalan suara dan beragam aplikasi lainnya. Para peneliti telah mengusulkan sejumlah metode untuk membangun sistem tersebut, namun umumnya kurang akurat untuk suara berderau. Pada penelitian ini, sebuah sistem klasifikasi gender berdasarkan suara dibangun dan diuji untuk sejumlah tingkat derau. Sistem ini berisi dua langkah. Pertama, vektor ciri dibangkitkan menggunakan *Mel-Frequency Cepstral Coefficient* (MFCC). Selanjutnya, vektor ciri tersebut diklasifikasi menggunakan *Support Vector Machine* (SVM). Pengujian menggunakan *5-fold cross-validation* terhadap 2,264 suara, yang terbagi secara merata ke dalam dua kelas: "laki-laki" and "perempuan", menunjukkan bahwa akurasi tertinggi 100% dicapai menggunakan SVM dengan kernel *Polynomial* dan *degree* 1 untuk suara bersih. Sistem ini cukup tahan, di mana akurasi sedikit menurun menjadi 95,21%, untuk suara berderau rendah -25 dB. Tetapi, akurasinya menurun drastis untuk suara berderau lebih tinggi.

Kata Kunci: klasifikasi gender, pengenalan suara, *Mel-Frequency Cepstral Coefficient*, *Support Vector Machine*.

I. Introduction

PENGENALAN suara adalah salah satu bidang riset penting yang saat ini banyak digunakan secara luas untuk beragam aplikasi, seperti sistem keamanan, autentifikasi, interaksi manusia-komputer, analisis fisiologi atau psikologi, dan sebagainya [7]. Untuk digunakan secara praktis, pengenalan suara harus memiliki performansi tinggi. Terdapat banyak cara untuk meningkatkan performansi pengenalan suara, salah satunya adalah dengan menambahkan suatu prosedur klasifikai gender. Dengan adanya klasifikasi gender, ruang masalah dalam pengenalan suara dapat dibatasi hanya berdasarkan gender yang telah diklasifikasikan [2]. Klasifikasi gender dilakukan terhadap dua kelas: laki-laki dan perempuan.

Received on Jan 2019. Accepted on March 2019

Suara yang dihasilkan laki-laki dan perempuan memiliki perbedaan yang cukup jelas berdasarkan bentuk tenggorokannya [6]. Perbedaan suara tersebut dapat diketahui menggunakan teknik pengolahan sinyal suara yang tepat. Banyak metode telah diusulkan untuk membangun sistem klasifikasi gender berbasis suara yang berkaurasi tinggi, namun umumnya masih kurang tahan terhadap derau atau *noise*.

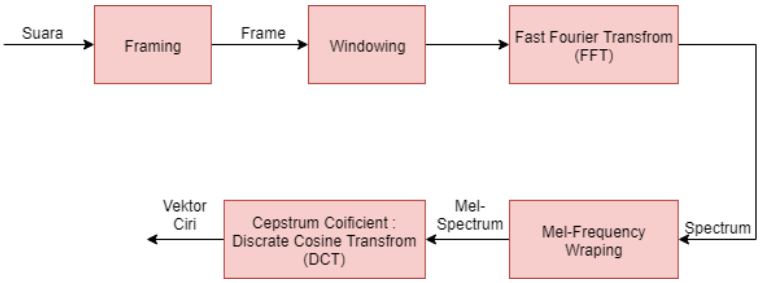
Penelitian ini fokus pada pembangunan sistem klasifikasi gender berdasarkan sinyal suara dan menguji ketahanannya terhadap *noise*. Sistem terdiri atas dua bagian, ekstrasi fitur dan klasifikasi. Ekstraksi fitur dilakukan menggunakan MFCC, yang secara umum memiliki akurasi tinggi dibanding metode lain [11]. Sementara itu, klasifikasi menggunakan SVM yang secara empiris terbukti memiliki akurasi lebih tinggi dibanding berbagai metode lainnya [1].

Selanjutnya, Bab II pada makalah ini akan membahas berbagai studi terkait mengenai klasifikasi gender berdasarkan suara. Bab III menjelaskan desain sistem klasifikasi gender. Pada Bab IV akan dibahas hasil eksperimen dan analisis hasil sistem klasifikasi gender. Terakhir, pada Bab V disampaikan kesimpulan dan saran untuk penelitian selanjutnya.

II. Literature Review

Klasifikasi gender berdasarkan suara merupakan salah satu dari sejumlah teknik yang menggunakan ciri biometrik tidak terlihat [7]. Banyak peneliti telah mengusulkan sejumlah metode untuk ekstraksi fitur suara dan metode klasifikasi. Pada umumnya, metode ekstraksi fitur yang digunakan adalah MFCC [9] sedangkan metode klasifikasinya adalah SVM [3], [6] karena terbukti sangat efektif dalam klasifikasi dan pengenalan suara [5]. Penelitian yang dilakukan oleh Urmila Shrawankar dan Vilas Thakare [11] menunjukkan bahwa MFCC memberikan akurasi paling tinggi untuk ekstrasi ciri suara dibanding metode lainnya. Pada [1] juga dinyatakan bahwa MFCC adalah metode yang simpel, cepat, dan berakurasi tinggi untuk pengenalan suara.

Secara teknis, MFCC didasarkan pada variasi frekuensi *critical bandwidth* dari telinga manusia yang digunakan untuk memperoleh fitur suara [4]. Pada [10] dijelaskan bahwa skala *Mel-frequency* adalah kurang dari 1000 Hz untuk frekuensi linier sedangkan dan lebih dari 1000 Hz untuk frekuensi logaritmik. Penghitungan Mel suatu frekuensi dirumuskan sebagai $mel(f) = 2595 \times \log_{10}(1 + \frac{f}{700})$ [10].



Gambar 1. Blok diagram proses penghitungan MFCC

Gambar 1 menjelaskan diagram blok penghitungan MFCC. Pertama, sinyal suara dipotong untuk menghilangkan keheningan atau gangguan yang mungkin muncul pada awal maupun akhir suara dilakukan *framing*, membagi ke dalam sejumlah *frame* [1]. Selanjutnya, proses *windowing* digunakan untuk meminimalkan diskontinuitas sinyal. *Fast Fourier Transform* (FFT) kemudian diterapkan untuk mengubah setiap *frame* ke domain frekuensi. Dalam *Mel-frequency wrapping block*, sinyal diplot terhadap spektrum Mel untuk meniru pendengaran manusia. Terakhir, *Discrete Cosine Transform* (DCT) dilakukan untuk menghasilkan vektor ciri.

Berdasarkan riset yang dilakukan Jamil Ahmad dkk. [1], SVM menghasilkan akursi tertinggi di antara metode-metode lain: *Naive Bayes* (NB), *Random Forest* (RF), *k-Nearest Neighbor* (kNN), dan *Multilayer Preceptron* (MLP), untuk berbagai aplikasi telepon. Oleh karena itu, pada penelitian ini SVM digunakan untuk melakukan klasifikasi gender berdasarkan fitur suara berbasis MFCC.

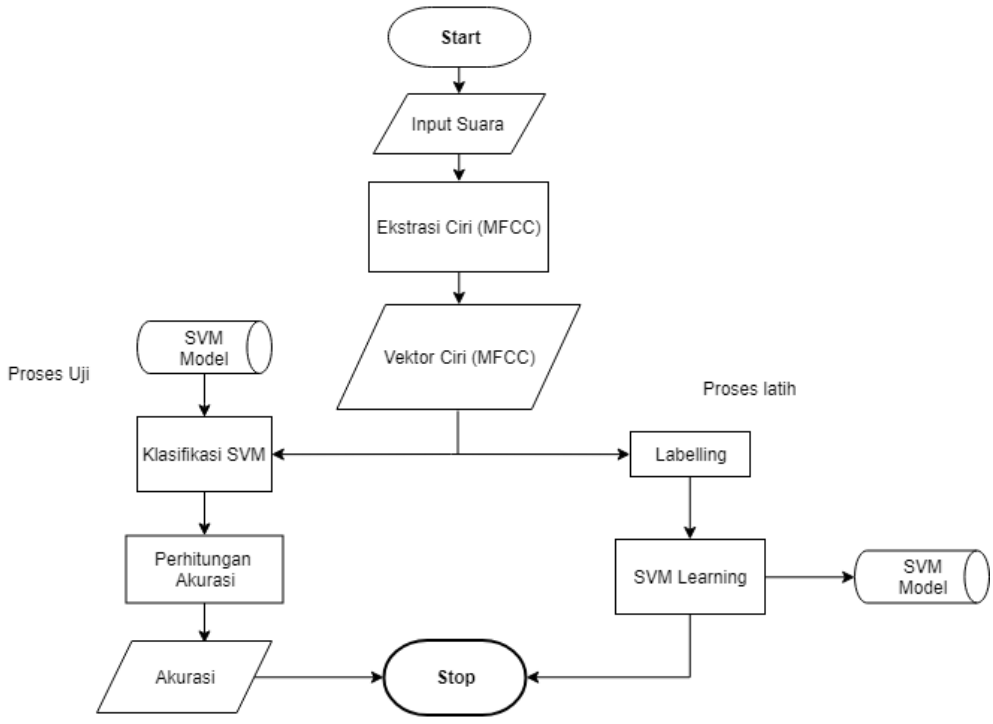
SVM diperkenalkan oleh Vapnik sebagai model klasifikasi dua kelas (*binary classifier*). Metode ini menggunakan proses klasifikasi dua langkah. Pertama, sebuah fungsi kernel melakukan transformasi fitur dimensi rendah ke dimensi tinggi. Transformasi ini mentransformasikan data yang *non-linearly separable* menjadi *linearly separable* pada dimensi yang lebih tinggi. Terdapat berbagai macam kernel yang bisa digunakan, seperti *Polynomial* dan *Radial Basis Function* (RBF). Langkah kedua, dilakukan konstruksi maksimal *margin hyperplane* untuk menentukan batas keputusan setiap kelas. Konsep separasi maksimum mencegah *misclassification of outliers* sehingga menjadikan SVM sebagai metode klasifikasi yang berakurasi tinggi [1]. Pada [5] dijelaskan bahwa untuk sebuah himpunan data latih berlabel $T = \{(x_b, l_i), i = 1, 2, ..L\}$ dengan $x_i \in R^P$ dan $l_i \in \{-1, 1\}$, sebuah data uji diklasifikasikan sebagai

$$f(x) = sign \sum_{i=1}^{L} \alpha_i . l_i . K(x_i, x) + b, \tag{1}$$

di mana α_i adalah *Lagrange Multipliers*, b adalah nilai batas, dan K adalah fungsi kernel. *Support vector* adalah subhimpunan dari data latih dengan $\alpha_i > 0$.

III. RESEARCH METHOD

Sistem klasifikasi suara berdasarkan gender menggunakan SVM dapat diilustrasikan menggunakan *flowchart* pada Gambar 2. Sitem ini terdiri atas proses input suara, kemudian dilakukan ekstrasi ciri untuk menghasilkan vektor ciri MFCC. Selanjutnya, proses terbagi menjadi 2, pelatihan dan pengujian. Pada proses pelatihan, dilakukan pemberian label, yaitu -1 untuk gender pria dan +1 untuk gender wanita, sebagai data latih (*trainset*). Selanjutnya, data ini dilatihkan untuk menghasilkan model SVM. Pada proses uji, vektor ciri akan diklasifikasikan berdasarkan model SVM hasil pelatihan. Terakhir, dilakukan perhitungan akurasi berdasarkan jumlah data yang benar dibagi jumlah seluruh data.

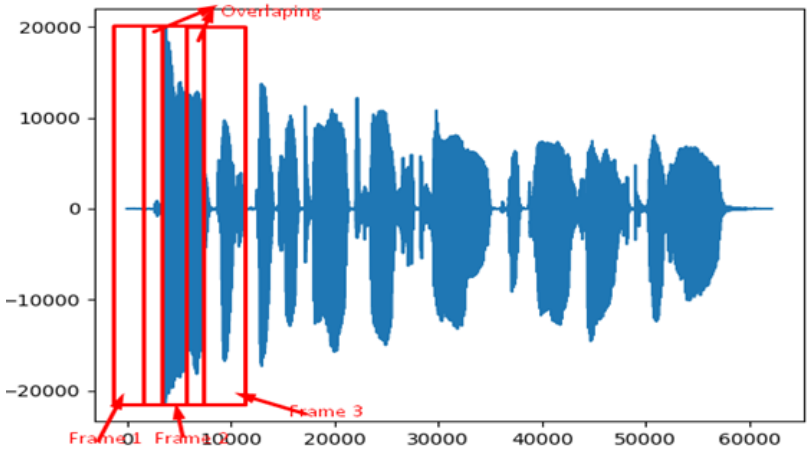


Gambar 2. Flowchat Sistem

Sementara itu, *dataset* yang digunakan dalam penelitian ini merupakan rekaman suara berformat .wav,

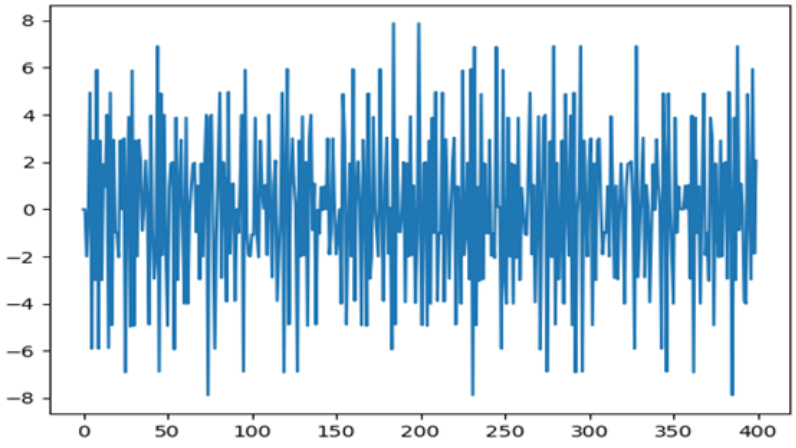
yang didapat dari *Telecommunications and Signal Processing Laboratory* [8]. Penjelasan secara ringkas mengenai *dataset* tersebut adalah: berisi 2.260 rekaman suara (1.130 suara laki-laki dan 1.130 suara perempuan), di mana rekaman suara berisi pengucapan kalimat yang sama dalam bahasa Inggris.

MFCC terdiri atas sejumlah proses. Pada proses *framing*, sinyal suara dibagi menjadi beberapa *frame*, di mana panjang setiap *frame* yang digunakan pada penelitian ini adalah 25 milidetik dengan *overlap* 10 milidetik, seperti diilustrasikan pada Gambar 3.



Gambar 3. Proses *framing*

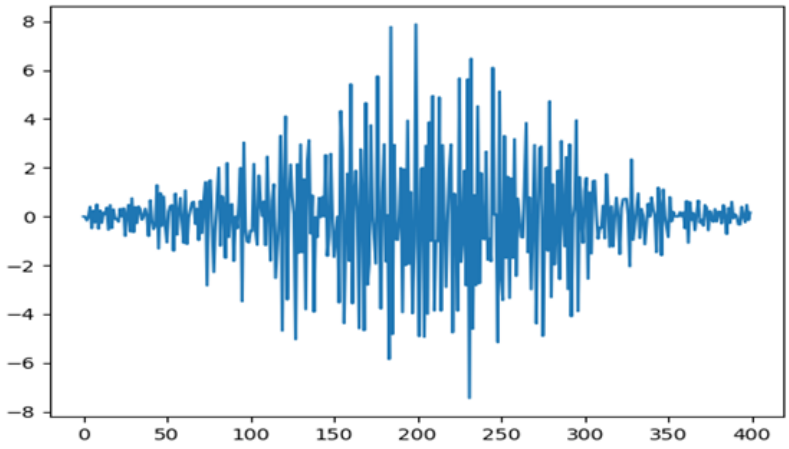
Selanjutnya, proses penjedelaan dilakukan menggunakan *hamming window* untuk meminimasi efek diskontinuitas dari proses *framing*. Gambar 4 dan 5 mengilustrasikan hasil dari proses *framing* dan *windowing*.



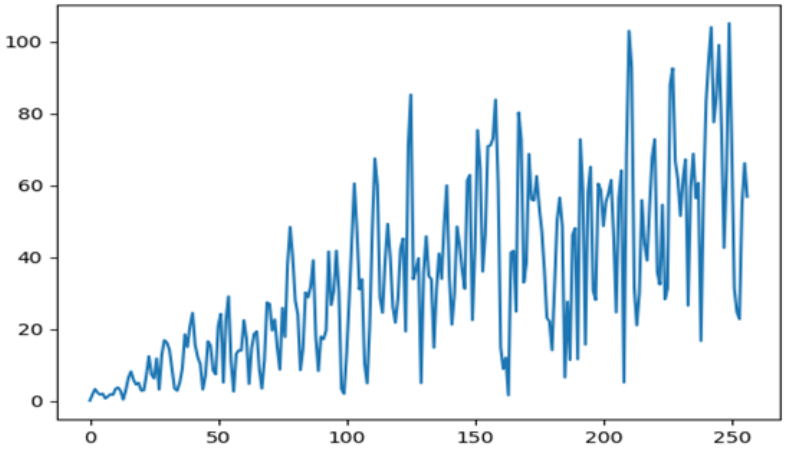
Gambar 4. Hasil proses *Framing*

Kemudian, hasil *dari windowing* tersebut diubah dari domain waktu ke domain frekuensi menggunakan FFT. Pada penelitian ini nilai FFT yang digunakan adalah 256 titik. Keluaran dari proses FFT berupa *spectrum magnitude* seperti diilustrasikan pada Gambar 6.

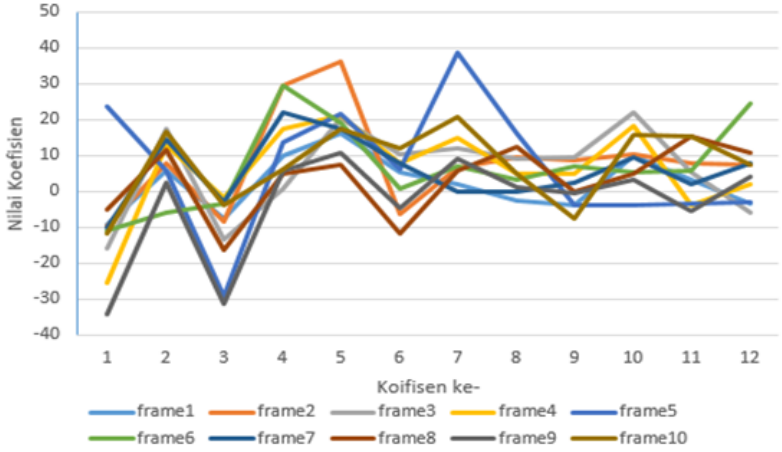
Selanjutnya dilakukan proses *Mel-frequency wrapping* untuk menghasilkan *mel-spectrum*. Penelitian ini menggunakan *filterbank* sebanyak 40 *filter*. Hasil proses tersebut kemudian diubah menjadi *cepstrum* menggunakan DCT dengan 12 koefisien dan diambil 10 *frame* dari setiap data. Hasil ekstrasi ciri diilustrasikan pada Gambar 7.



Gambar 5. Hasil proses *windowing*



Gambar 6. Hasil FFT



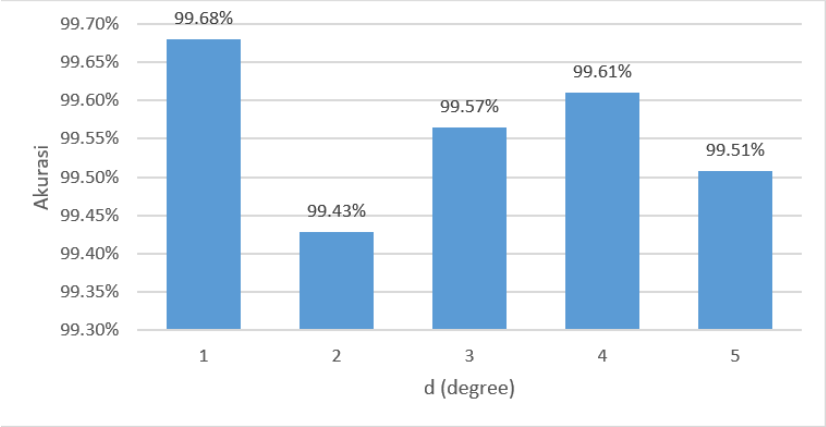
Gambar 7. Hasil ekstrasi ciri

IV. RESULT AND DISCUSSION

Pada bagian ini ditunjukkan hasil pengujian yang dilakukan terkait penggunaan parameter *degree* (d) pada kernel Polynomial, *gamma* (γ) pada kernel RBF, Linier, dan Sigmoid serta penambahan *noise* pada data uji serta ditampilkan pelatihan dengan metode *5-fold cross validation*.

A. Pengaruh Parameter d (degree) pada Kernel Polynomial

Gambar 8 menunjukkan rata-rata akurasi untuk *5-fold cross-validation*. Semakin besar nilai d, akurasi sistem menjadi fluktuatif dan kurang stabil. Hal ini terjadi karena smakin tinggi nilai d maka garis *hypeplane* yang dihasilkan semakin melengkung. Nilai parameter d pada pengujian yang dilakukan sudah optimal pada $d = 1$ dengan akurasi tertinggi sebesar 99.68%. Tabel I mengilustrasikan secara detail nilai akurasi pada setiap *fold*.



Gambar 8. Rata-rata akurasi setiap *fold* pada parameter d kernel Polynomial

Tabel I
AKURASI DETAIL PADA SETIAP FOLD

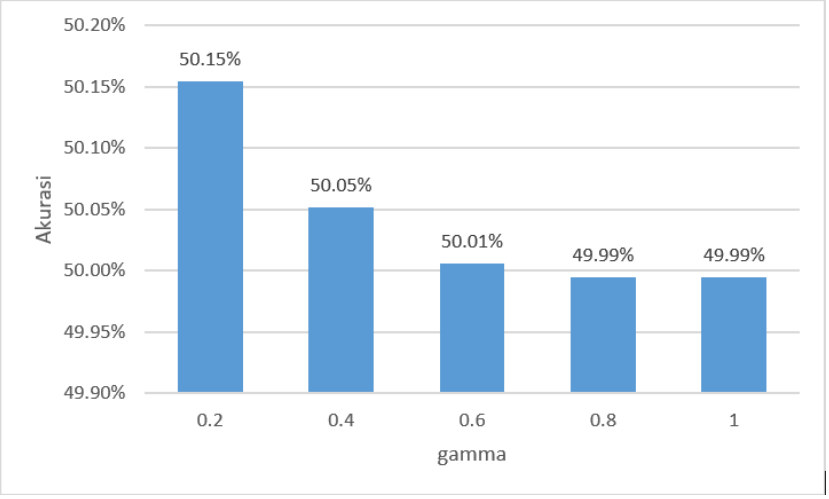
degree	fold1	fold2	fold3	fold4	fold5	average
1	99.8%	99.5%	99.8%	99.7%	99.6%	99.7%
2	99.6%	99.2%	99.4%	99.7%	99.3%	99.4%
3	99.7%	99.5%	99.7%	99.7%	99.3%	99.6%
4	99.7%	99.7%	99.7%	99.5%	99.4%	99.6%
5	99.5%	99.6%	99.7%	99.5%	99.2%	99.5%

B. Pengaruh Parameter gamma pada Kernel RBF

Hasil eksperimen untuk menganalisis parameter *gamma* pada kernel RBF diilustrasikan pada Gambar 9. Hasil ini memiliki kemiripan dengan kernel polynominal. Hal ini disebabkan semakin kecil nilai *gamma*, maka dua titik data yang bejarak jauh dapat dikatakan sama. Pada kernel RBF, nilai parameter optimal dihasilkan pada *gamma* = 0.2 dengan akurasi 50.15 %. Akurasi detail pada setiap *fold* diilustrasikan pada Tabel II.

C. Pengaruh kernel dalam SVM

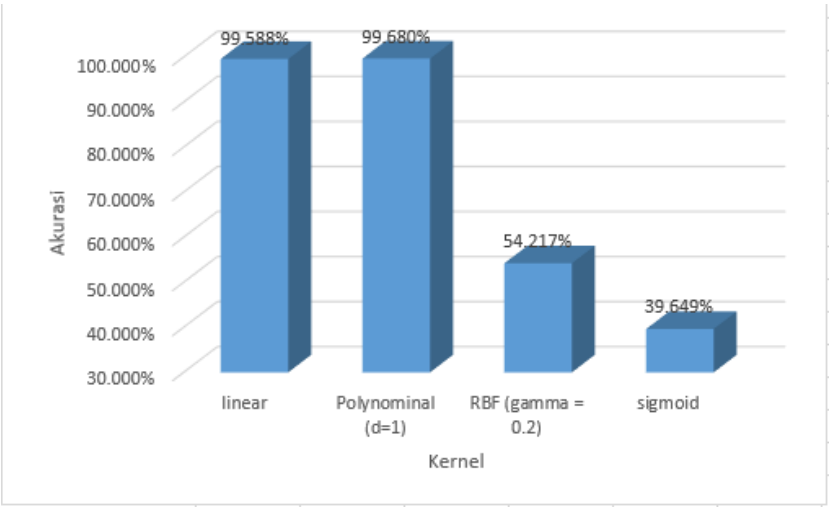
Berikut adalah grafik gabungan akurasi terbaik yang sudah didapat pada subbagian A dan B, ditambah dengan kernel Linier dan kernel Sigmoid. Hasil pada Gambar 10 merupakan akurasi terbaik setiap kernel dari 5-*fold*. Berdasarkan Gambar 10 dapat disimpulkan kernel terbaik adalah kernel Polynomial dengan $d = 1$ yang menghasilkan akurasi tertinggi 99.68%.



Gambar 9. Hasil rata-rata akurasi tiap *k-fold* pada parameter gamma kernel RBF

Tabel II
DETAIL AKURASI TIAP FOLD

gamma	fold1	fold2	fold3	fold4	fold5	average
0.2	50.00%	50.00%	50.00%	50.00%	50.77%	50.15%
0.4	50.00%	50.00%	50.00%	50.00%	50.26%	50.05%
0.6	50.00%	50.00%	50.00%	50.00%	50.03%	50.01%
0.8	50.00%	50.00%	50.00%	50.00%	49.97%	49.99%
1	50.00%	50.00%	50.00%	50.00%	49.97%	49.99%

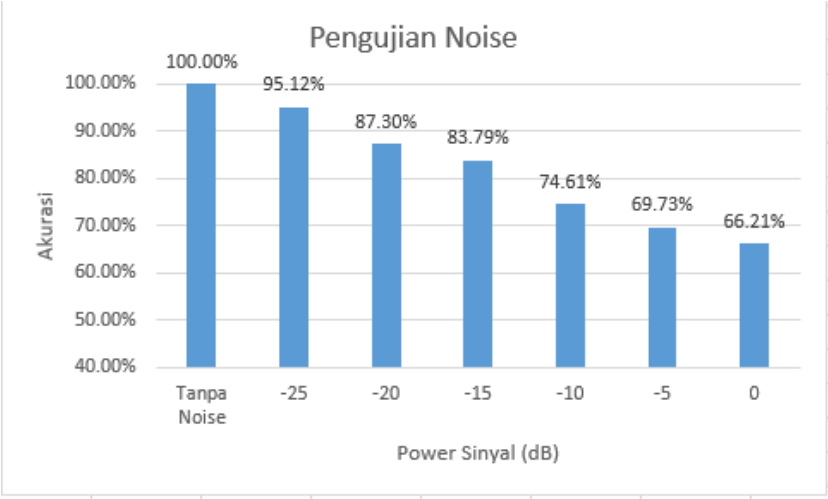


Gambar 10. Hasil akurasi pengaruh kernel pada SVM

D. Pengujian terhadap data ber-noise

Dataset yang berjumlah 2.260 suara dibagi menjadi dua. Pertama, 1.748 suara digunakan untuk pelatihan. Kedua, 512 suara untuk pengujian. Data latih digunakan untuk pembangunan model SVM dengan kernel Polynomial dan $d = 1$. Sementara itu, semua data uji diberikan *noise* yang berisi rekaman suara kemacetan dalam kota dengan sejumlah tingkat derau yang berbeda-beda. Hasil pengujian pada Gambar 11 menunjukkan bahwa sistem yang dibangun mampu menghasilkan akurasi sempurna 100%

pada data tanpa *noise*. Namun, akurasi ini menurun sedikit menjadi 95,12% untuk data uji yang diberi *noise* rendah (-25 dB). Akurasi menurun drastis secara cepat seiring dengan semakin besarnya *noise*. Ketika power sinyal 0 dB, yang berarti suara asli dan *noise* sama besar, akurasi sistem hanya 66,21%.



Gambar 11. Hasil akurasi data tanpa *noise* dan dengan *noise*

V. Conclusion

Sistem klasifikasi gender berdasarkan suara menggunakan SVM yang dibangun mampu memberikan akurasi sebesar 100%. Pemilihan parameter dan kernel SVM sangat mempengaruhi akurasi sistem. Kernel RBF dan *Sigmoid* mempunyai akurasi lebih buruk dibanding *Linear* dan *Polynomial* ($d = 1$). Hal ini disebabkan karena *dataset* yang digunakan bersifat linier. Ketahanan sistem masih relatif rendah. Ketika penambahan *noise* cukup besar, lebih dari -15 dB, akurasi sistem menurun drastis. Hal ini disebabkan sistem sulit membedakan mana suara asli dan mana *noise*. Penelitian ini dapat dilanjutkan dengan fokus pada proses pelatihan menggunakan *trainset* yang sebagian telah diberikan *noise* sintetis yang terukur untuk menghasilkan sistem yang lebih tahan terhadap *noise*.

ACKNOWLEDGMENT

Penulis mengucapkan terimakasih kepada kedua orangtua dan semua kolega di Telkom University yang telah mendukung riset ini.

PUSTAKA

[1] Jamil Ahmad, Mustansar Fiaz, Soon-il Kwon, Maleerat Sodanil, Bay Vo, and Sung Wook Baik. Gender Identification using MFCC for Telephone Applications - a Comparative Study. *arXiv preprint arXiv:1601.01577*, 2016.
[2] Sreedhar Bhukya. Effect of Gender on Improving Speech Recognition System. *International Journal of Computer Applications*, 179:22–30, 2018.
[3] Yashpalsing Chavhan. Speech Emotion Recognition Using Support Vector Machine. *International Journal of Computer Applications*, 1(20):6–9, 2010.
[4] Joyce de Vegte and Yin Xiaoli. *Fundamentals of digital signal processing*. Prentice Hall, 2002.
[5] A Ganapathiraju, J E Hamaker, and J Picone. Applications of support vector machines to speech recognition. *IEEE Transactions on Signal Processing*, 52(8):2348–2355, aug 2004.
[6] Sarah Ita Levitan, Taniya Mishra, and Srinivas Bangalore. Automatic identification of gender from speech. In *Speech Prosody 2016*, pages 84–88, 2016.
[7] Feng Lin, Yan Zhuang, Xi Long, and Wenyao Xu. Human Gender Classification : A Review. *IJBM*, 8(July):275–300, 2015.
[8] University McGill. The Telecommunications & Signal Processing Laboratory, 2019.
[9] Jaume Padrell-sendra and D Fernando. Support Vector Machines for Continuous Speech Recognition. In *the 14th European Signal Processing Conference (EUSIPCO)*, number Eusipco, pages 2–5, 2006.
[10] Kashyap Patel and R K Prasad. Speech recognition and verification using MFCC & VQ. *Int. J. Emerg. Sci. Eng.(IJESE)*, 1(7), 2013.
[11] Urmila Shrawankar and Vilas M Thakare. Techniques for feature extraction in speech recognition system: A comparative study. *arXiv preprint arXiv:1305.1145*, 2013.