



JESCE

(Journal of Electrical and System Control Engineering)

Available online <http://ojs.uma.ac.id/index.php/jesce>

Deteksi Nada Dasar Alat Musik Panting Menggunakan Compressive Sensing dengan MFCC dan SVM

Basic Tone Detection of Important Musical Instruments Using Compressive Sensing with MFCC and SVM

Shinta Kirana Salsabilla Ramadhani¹⁾*, Gelar Budiman²⁾, Bambang Hidayat³⁾

1),2),3) Teknik Telekomunikasi, Telkom University Indonesia

Diterima: 22 Mei 2022; Direview: 24 Mei 2022; Disetujui: 02 Oktober 2022

*Corresponding Email: shintakiranasr@gmail.com

Abstrak

Selama ini penyajian alat musik Panting membutuhkan proses kalibrasi yang hanya mengandalkan pendengaran manusia. Penelitian ini bertujuan untuk merancang sistem pengenalan nada dasar Panting untuk mempermudah proses kalibrasi. Metode yang digunakan adalah Compressive Sensing, MFCC dan SVM. Penelitian ini menggunakan 6 jenis audio, yaitu audio terkompresi dengan lima rasio yang berbeda dan audio asli. Berdasarkan penelitian yang telah dilakukan, sistem deteksi nada Panting dalam keadaan maksimal dengan menggunakan data audio terkompresi 2,5%. Menggunakan skenario ini memperoleh nilai akurasi 97,96% dengan proses klasifikasi selama 0,06274 detik dengan menggunakan kernel Linear.

Kata Kunci: MFCC; SVM; Compressive Sensing; Panting.

Abstract

During this time, Panting's calibration process has been done with mere reliance on human hearing. Thus, this study is intended to assist the calibration process of Panting. The experiment has been done by using Compressive Sensing, MFCC and SVM. Moreover, the six scenarios are applied to the classification process are classification using five audios with different compression ratios and the original audio. This study results the best state of the Panting's note recognition system is given by using the compressed audio with ratio of 2,5%. It is proven by its 97,96% of accuracy that is computed in 0,06274 second of duration.

Keywords: MFCC; SVM; Compressive Sensing; Panting.



This work is licensed under a Creative Commons Attribution 4.0

PENDAHULUAN

Sistem yang dirancang menggunakan metode *Mel Frequency Cepstral Coefficient* (MFCC) sebagai metode ekstraksi ciri, adapun untuk proses klasifikasi menggunakan *Support Vector Machine* (SVM). Selain itu, proses kompresi juga dilakukan dengan menggunakan *Compressive Sensing* (CS) dan algoritma OMP untuk proses rekonstruksi.

Penelitian yang berjudul *Music Classification using MFCC and SVM* (Thiruvengatanadhan, 2018) menghasilkan nilai akurasi 93% dalam proses klasifikasi sinyal nada musik menggunakan MFCC dan SVM. Hal ini menunjukkan bahwa penggunaan SVM dan MFCC dapat memberikan nilai akurasi yang baik. Namun, pada penelitian ini tidak menerapkan tahap kompresi.

Penelitian tentang pengenalan suara masih sedikit yang melibatkan proses metode kompresi. Adanya proses kompresi dapat mengatasi permasalahan efisiensi pada proses pengujian. Penelitian mengenai deteksi sinyal suara menggunakan CS dan MFCC (Gumelar et al., 2020) menghasilkan hasil klasifikasi dari penggunaan SVM dan KNN dengan besar akurasi 91,7% dan 94%. Penelitian ini membuktikan bahwa penerapan CS pada sistem deteksi suara dapat memberikan nilai akurasi yang baik.

Namun, pada penelitian ini tidak ada pengukuran waktu komputasi untuk mengetahui efisiensi kerja sistem secara keseluruhan.

METODE PENELITIAN

Informasikan secara ringkas mengenai materi dan metode yang digunakan dalam penelitian, meliputi subjek/bahan yang diteliti, alat yang digunakan, rancangan percobaan atau desain yang digunakan, teknik pengambilan sampel, variabel yang akan diukur, teknik pengambilan data, analisis dan model statistik yang digunakan.

A. *Compressive Sensing* (CS)

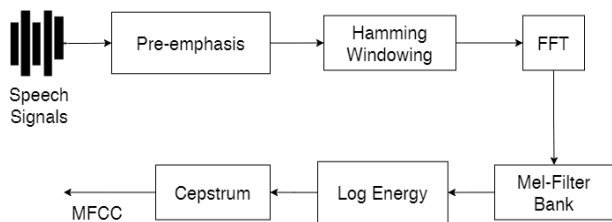
Compressive Sensing (CS) adalah suatu teknik pemrosesan sinyal untuk mengkompresi dan merekonstruksi sinyal. Berdasarkan pada prinsip teorema Shannon/Nyquist, bahwa hanya dibutuhkan frekuensi sampling minimal 2 kali dari besar *bandwidth* untuk merekonstruksi suatu sinyal. Hal ini menyebabkan tidak efisien saat rekonstruksi dikarenakan memerlukan penyimpanan data yang cukup besar. Sehingga dengan adanya CS dapat mengatasi masalah ini dengan melakukan pengukuran terkompresi dari sinyal (Nguyen et al., 2016).

Setelah terkompresi, sinyal melalui proses rekonstruksi untuk memulihkan

sinyal terkompresi menggunakan algoritma *Orthogonal Matching Pursuit* (OMP). OMP bekerja dengan menemukan basis vektor secara berulang, sehingga perulangan kesalahan dalam representasi sinyal akan semakin berkurang (Fatahilah & Winardhi, 2017).

B. Mel-Frequency Cepstral Coefficient (MFCC)

Pada proses ekstraksi ciri, sinyal audio diproses dan dianalisa dengan mengubah ke dalam bentuk parametrik. Penelitian ini menerapkan metode MFCC untuk proses ekstraksi ciri. MFCC menggunakan mel scale untuk proses ekstraksi ciri karena pada saat frekuensi meningkat melebihi 1 kHz, maka frekuensi pada pendengaran manusia akan menunjukkan penurunan pada karakteristik logaritmik secara bertahap. Adapun arsitektur proses ekstraksi ciri dengan metode MFCC dijelaskan pada Gambar 1 (Sen et al., 2019).



Gambar 1. Arsitektur proses MFCC

C. Support Vector Machine (SVM)

SVM merupakan merupakan salah satu algoritma klasifikasi *supervised learning* yang membedakan kelas dengan

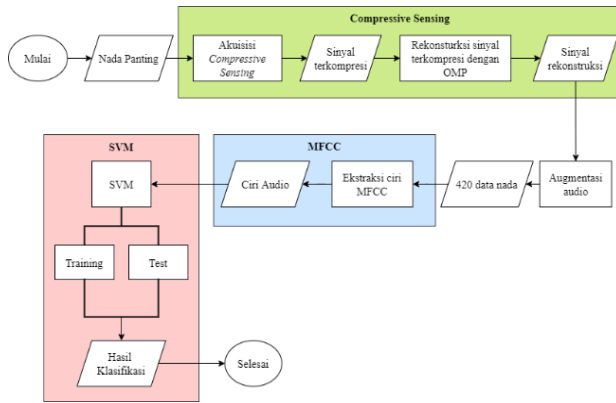
vektor sebanyak n dimensi (Shah & Suthar, 2016). SVM bekerja dengan membangun *hyperplane* atau garis pembatas pada ruang multi dimensi yang bertujuan untuk memisahkan masing – masing kelas berdasarkan pendekatan statistika dengan cara memaksimalkan lebar margin (Patil & Nemade, 2017).

Sebagian besar data tidak bisa dipisahkan menggunakan pembatas linier sehingga membuat proses klasifikasi semakin rumit. SVM menggunakan fungsi Kernel dengan tujuan untuk memetakan persebaran data ke dalam ruang dimensi yang lebih tinggi (Muhdina, 2021). Kernel memiliki berbagai macam jenis, salah satunya adalah *Linear*, *Polynomial* dan RBF.

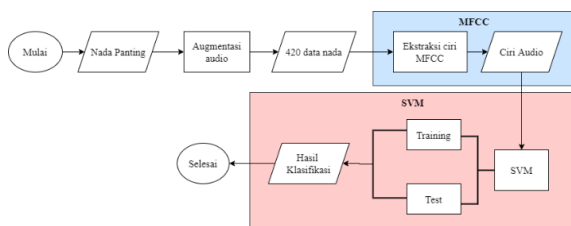
D. MODEL SISTEM PERANCANGAN

1. Desain Sistem

Penelitian ini membandingkan hasil proses deteksi nada musik Panting dengan dan tanpa melalui proses kompresi. Adapun menerapkan kernel *Linear*, *Polynomial* dan RBF di penelitian ini dan hasil dari setiap penggunaan kernel dibandingkan. Berikut adalah gambar 2 dan 3 yang merupakan skema sistem deteksi nada Panting dengan kompresi CS dan tanpa CS.



Gambar 2 Skema sistem deteksi nada panting dengan proses kompresi CS



Gambar 3 Skema sistem deteksi nada Panting tanpa proses kompresi CS

2. AKUISISI DATA

Berikut tabel 1 yang merangkum informasi mengenai proses akuisisi data di penelitian ini.

Tabel 1 Akuisisi data audio

PROSES	KETERANGAN
Sampel audio	Nada dasar do, re, mi, fa, so, la, si dari alat musik Panting
Format audio	WAV
Frekuensi Sampling	44100 Hz
Banyak sampel audio	20 sampel per nada
Durasi audio	1 detik
Total data audio	140 data audio

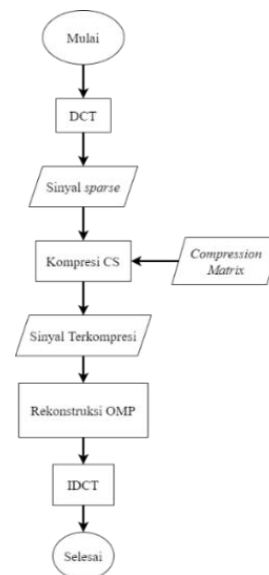
3. Augmentasi Data

Menerapkan proses Augmentasi bertujuan agar data audio menjadi beragam sehingga sistem dapat melakukan proses klasifikasi menjadi lebih optimal.

Penelitian ini melibatkan 3 jenis augmentasi, yaitu *noise injection*, *stretch* dan *pitch*. Setiap 140 sampel data melewati sebanyak 2 kali augmentasi, yaitu penambahan *noise* dan perubahan *pitch* dari audio yang telah dilakukan stretching. Sebanyak 280 data audio baru diperoleh dari proses augmentasi sehingga terdapat 420 total data audio

4. PROSES KOMPRESI

Pada Gambar 4, diagram blok yang menjelaskan setiap tahap pada proses kompresi menggunakan CS ditampilkan.



Gambar 4 Diagram proses kompresi CS

5. EKSTRAKSI CIRI MFCC

Pada proses ekstraksi ciri mfcc terdapat beberapa parameter ciri yang dapat digunakan untuk proses klasifikasi. Berikut adalah parameter ciri yang digunakan dalam penelitian ini.

- Zero Crossing Rate (ZCR)
- Spectral Rolloff
- Ciri MFCC sebanyak 20
- Mel-Spectrogram

6. KLASIFIKASI DENGAN SVM

Klasifikasi merupakan proses pengelompokkan dan prediksi dari audio yang telah melewati proses ekstraksi ciri sesuai dengan labelnya. Dilakukan proses klasifikasi svm dengan tiga jenis kernel yang berbeda, yaitu *linear*, *polynomial* dan *rbf* untuk membandingkan pengaruh jenis kernel dalam proses klasifikasi nada.

HASIL DAN PEMBAHASAN

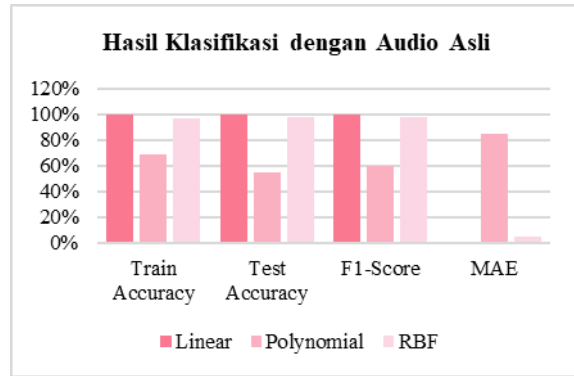
A. KOMPRESI CS DAN REKONSTRUKSI OMP

Pada proses ini, setiap sinyal masukan dibaca dalam bentuk *array* atau matriks berukuran 63000×1 . Matriks sinyal dikalikan dengan matriks A yang berukuran $M \times 63000$. Pada penelitian ini menggunakan lima nilai M yang berbeda sehingga menghasilkan rasio kompresi yang bervariasi pula. Berikut tabel 2 merangkuman hasil rasio yang diperoleh.

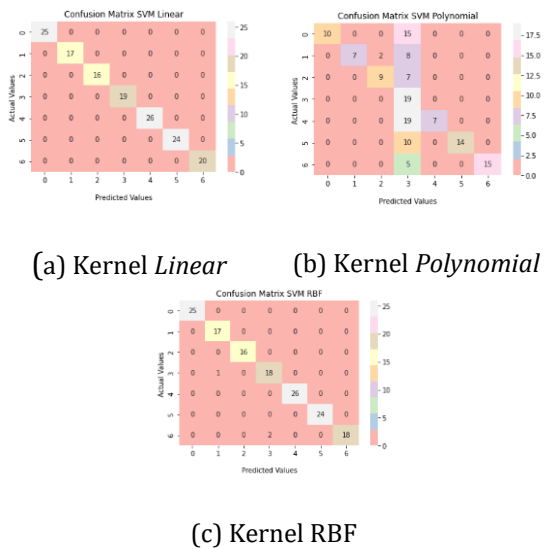
Tabel 2 Rangkuman hasil kompresi CS

Matriks A ()	
M	Rasio Kompresi
394	0,625%
788	1,25%
1575	2,5%
3150	5 %
6300	10%

B. KLASIFIKASI NADA TANPA PROSES KOMPRESI CS



Gambar 5 Hasil klasifikasi nada Panting (audio asli)

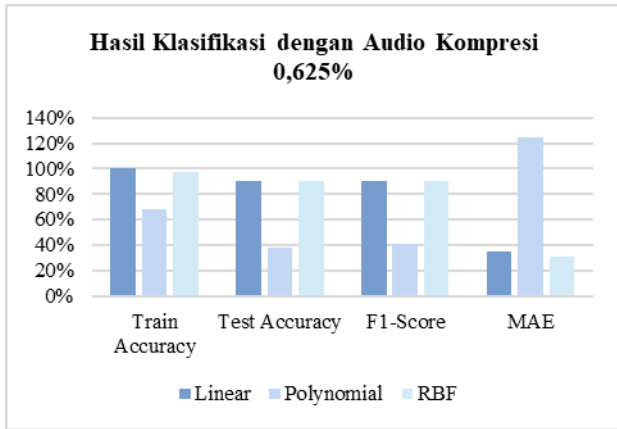


Gambar 6 Confusion Matrix hasil klasifikasi dengan audio asli

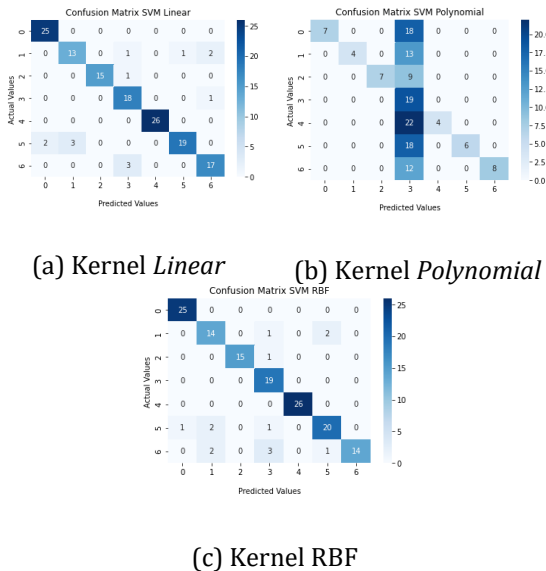
Hasil klasifikasi menggunakan audio tanpa melalui proses kompresi telah terangkum pada gambar 5 dan *confusion matrix* untuk ketiga penggunaan kernel terangkum pada gambar 6. Ketiga parameter kernel SVM menunjukkan hasil yang beragam. Berdasarkan hasil yang telah diperoleh, kernel *Linear* dan kernel RBF merupakan parameter kernel yang bagus untuk model SVM deteksi nada Panting tanpa melalui proses kompresi.

C. KLASIFIKASI NADA DENGAN RASIO KOMPRESI CS 0,625%

terkompresi dengan rasio 0,625%. Kernel *Linear* dan kernel RBF menjadi parameter yang dapat memberikan hasil yang baik di skenario ini. Namun, masih banyak kesalahan prediksi yang ditemukan yang ditunjukkan di *confusion matrix* pada ketiga penerapan kernel.



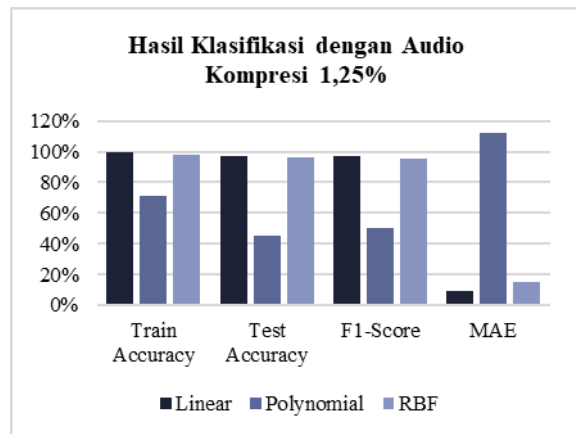
Gambar 7 Hasil klasifikasi nada Panting dengan rasio kompresi CS 0,625%



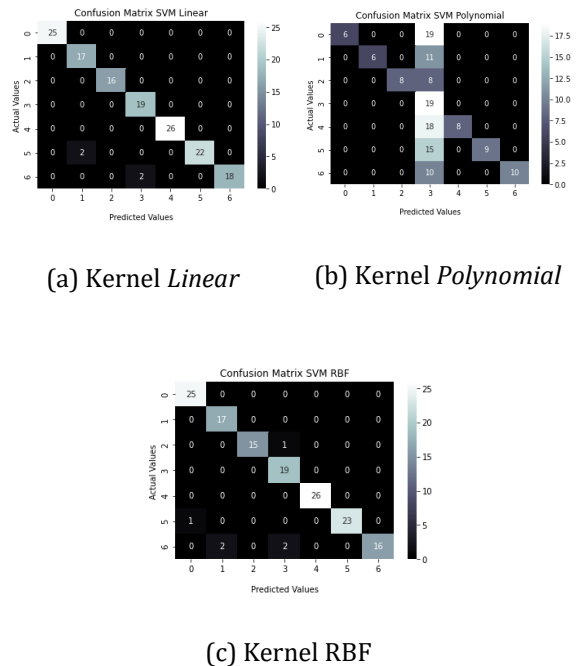
Gambar 8 *Confusion Matrix* hasil klasifikasi dengan audio terkompresi rasio 0,625%

Di skenario ini menggunakan audio yang telah terkompresi dengan rasio 0,625% dan hasilnya terangkum pada gambar 7 dan *confusion matrix* di gambar 8. Berdasarkan hasil yang telah diperoleh, dapat ditarik kesimpulan bahwa secara keseluruhan sistem dengan menggunakan MFCC dan SVM mampu memperoleh hasil klasifikasi yang baik meskipun data telah

D. KLASIFIKASI NADA DENGAN RASIO KOMPRESI CS 1,25%



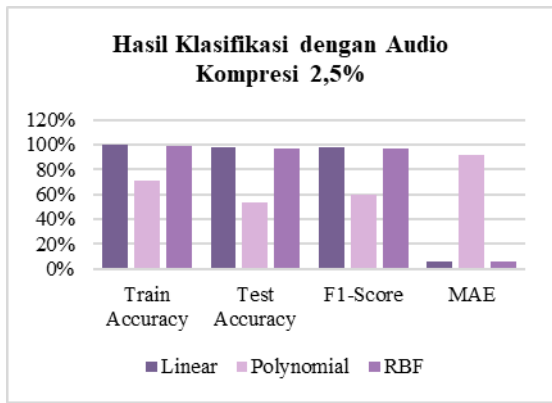
Gambar 9 Hasil klasifikasi nada Panting dengan rasio kompresi CS 1,25%



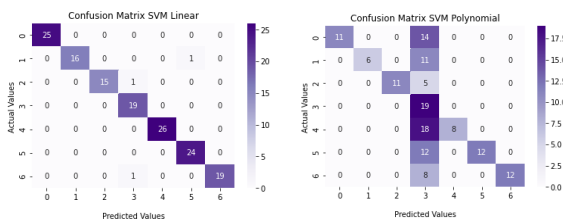
Gambar 10 *Confusion Matrix* hasil klasifikasi dengan audio terkompresi rasio 1,25%

Skenario ini menghasilkan hasil klasifikasi nada yang telah terangkum pada gambar 9 dan *confusion matrix* pada gambar 10. Dari ketiga kernel, penerapan kernel *Linear* dan RBF memberikan hasil terbaik. Sebaliknya, penerapan kernel *Polynomial* tidak memberikan hasil yang memuaskan. Hal ini telah dibuktikan dengan keakuratan prediksi yang dipetakan di *confusion matrix*.

E. KLASIFIKASI NADA DENGAN RASIO KOMPRESI CS 2,5%

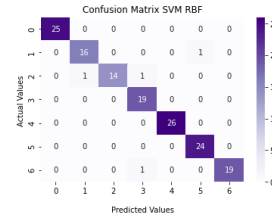


Gambar 11 Hasil klasifikasi nada Panting dengan rasio kompresi CS 2,5%



(a) Kernel *Linear*

(b) Kernel *Polynomial*

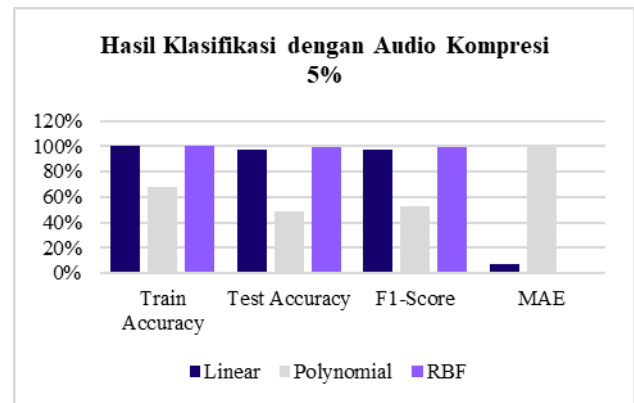


(c) Kernel RBF

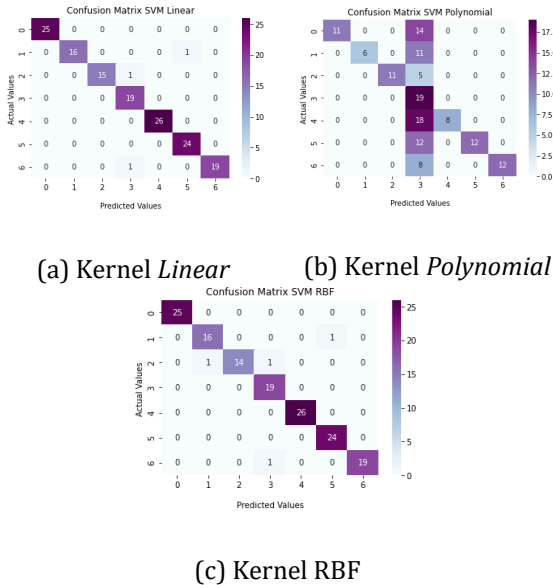
Gambar 12 Confusion Matrix hasil klasifikasi dengan audio terkompresi rasio 2,5%

Hasil klasifikasi pada skenario menggunakan audio terkompresi dengan rasio 2,5% dirangkum pada gambar 11 dan *confusion matrix* pada gambar 12. Berdasarkan hasil dari ketiga parameter SVM yang telah diimplementasikan, kernel *Linear* dan RBF menghasilkan hasil klasifikasi yang lebih unggul. Sebaliknya, kernel *Polynomial* memberikan hasil yang kurang ideal yang ditandai dengan banyaknya kesalahan prediksi.

F. KLASIFIKASI NADA DENGAN RASIO KOMPRESI CS 5%



Gambar 13 Hasil klasifikasi nada Panting dengan rasio kompresi CS 5%



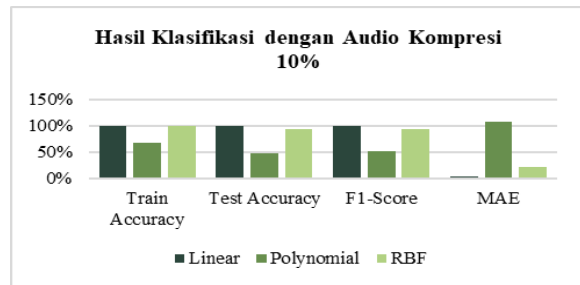
Gambar 14 Confusion Matrix hasil klasifikasi dengan audio terkompresi rasio 5%

Hasil klasifikasi nada menggunakan data audio yang terkompresi CS dengan rasio 5% dirangkum pada gambar 13 dan *confusion matrix* di gambar 14. Pada percobaan dengan tiga kernel berbeda, kernel *Linear* memberikan hasil klasifikasi yang baik dan kernel RBF memberikan hasil klasifikasi yang nyaris sempurna. Sebaliknya, penerapan kernel *Polynomial* pada model tidak memberikan hasil klasifikasi sebaik dua jenis kernel lainnya. Hal ini juga dibuktikan oleh pemetaan hasil prediksi di *confusion matrix*.

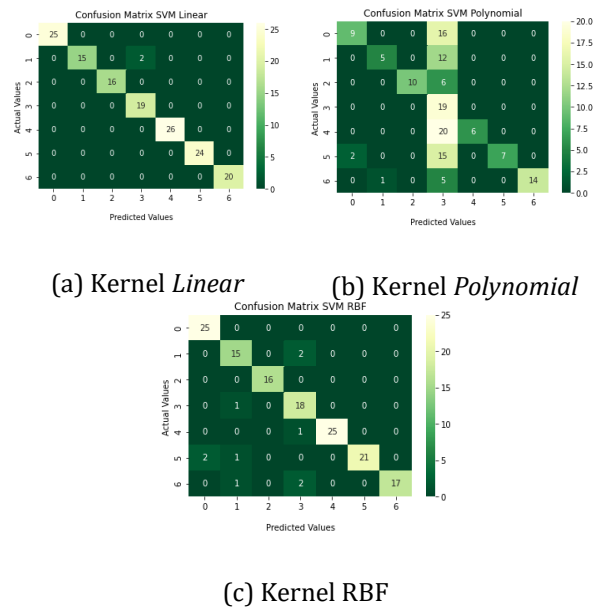
G. Klasifikasi Nada dengan Rasio Kompresi CS 10%

Penerapan tiga jenis kernel pada skenario ini memberikan hasil yang telah terangkum pada gambar 15 dan gambar 16 menampilkan *confusion matrix* dari hasil prediksi di ketiga penerapan kernel.

Dari ketiga jenis kernel yang telah diimplementasikan pada skenario ini, dapat ditarik kesimpulan bahwa penerapan kernel *Linear* dan RBF memberikan hasil klasifikasi yang baik. Sebaliknya, penggunaan kernel *Polynomial* menghasilkan nilai yang kurang baik sehingga berkemungkinan penggunaan jenis kernel ini kurang cocok untuk skenario ini.

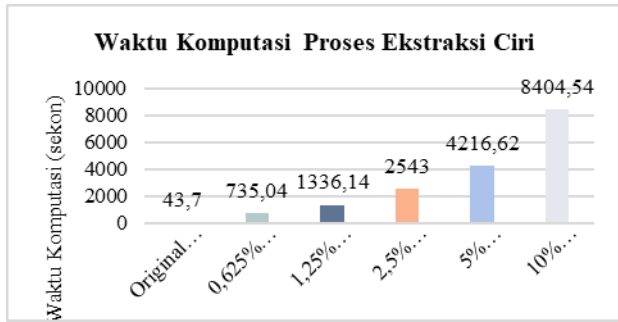


Gambar 15 Hasil klasifikasi nada Panting dengan rasio kompresi CS 10%



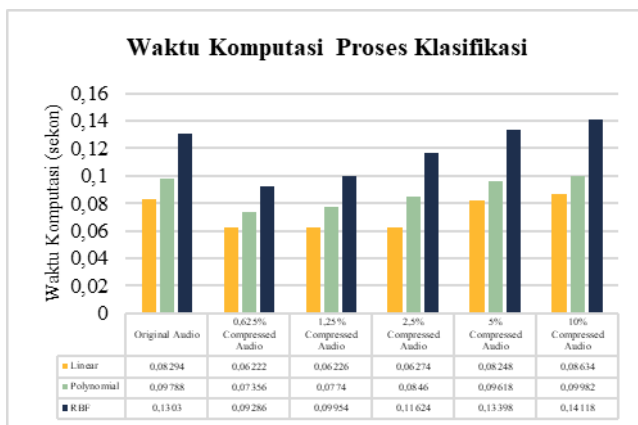
Gambar 16 Confusion Matrix hasil klasifikasi dengan audio terkompresi rasio 10%

H. Waktu Komputasi Sistem



Gambar 17 Waktu komputasi proses ekstraksi ciri

Sistem terbaik adalah sistem yang memiliki hasil klasifikasi yang baik dengan waktu komputasi yang cepat. Pada penelitian ini telah mengukur dua waktu komputasi, yaitu waktu komputasi saat ekstraksi ciri. Berikut merupakan waktu komputasi pada proses ekstraksi ciri dari 420 audio termasuk yang telah diaugmentasi pada setiap skenario yang terangkum pada grafik di gambar 17. Data pada grafik menunjukkan bahwa besarnya rasio kompresi, memiliki pengaruh besar dalam waktu komputasi pada proses ekstraksi ciri.



Gambar 18 Waktu Komputasi saat proses klasifikasi

Setelah waktu komputasi saat ekstraksi ciri MFCC, telah diukur waktu komputasi pada saat klasifikasi menggunakan tiga kernel yang berbeda pada model SVM. Adapun hasil waktu komputasi yang telah terukur dirangkum pada gambar 4.14. Pada grafik, terlihat bahwa jenis kernel dan jenis data audio yang digunakan dalam proses klasifikasi mempunyai pengaruh yang signifikan terhadap waktu komputasi. Pada perhitungan yang dilakukan, dapat ditarik kesimpulan bahwa penggunaan data audio terkompresi 0,625% dan 1,25% memiliki durasi tercepat pada proses klasifikasi nada Panting.

SIMPULAN

Penelitian ini telah menjawab tujuan dari penelitian ini dan telah terangkum pada poin sebagai berikut:

1. Penerapan kernel memiliki pengaruh besar dalam perancangan sistem pendeteksi nada panting. Kernel *Linear* merupakan jenis kernel yang memberikan hasil paling baik untuk sistem ini.
2. Penggunaan audio asli memakan waktu yang lebih lama dibandingkan menggunakan audio terkompresi. Kondisi terbaik sistem adalah saat menggunakan audio terkompresi 2,5%. Pada skenario ini memberikan nilai

akurasi dan $f1$ -score kurang lebih 98% dengan MAE 0,05 dalam waktu klasifikasi selama 0,062 detik.

3. Semakin besar rasio kompresi, maka semakin lama waktu yang dibutuhkan untuk proses kompresi dan ekstraksi ciri dan proses klasifikasi. Kemampuan proses klasifikasi pada sistem menunjukkan bahwa meskipun audio telah terkompresi sangat rendah, sistem dapat memberikan hasil klasifikasi yang baik.

DAFTAR PUSTAKA

- Fatahilah, A. D., & Winardhi, S. (2017). Perbandingan Inversi Reflektifitas Menggunakan Basis Pursuit Inversion dan Orthogonal Matching Pursuit pada Lapisan Tipis. *Jurnal Geofisika*, 15(2), 1-10.
- Gumelar, R. N., Raharjo, J., & Safitri, I. (2020). *Analisis Kompresi Nada Seruling Menggunakan Compressive Sensing Dengan Metode Discrete Fourier Transform Dan Stationary Wavelet Transform*. 9(2).
- Muhdina, A. (2021). *Klasifikasi Tuberkulosis dengan Compressive Sensing dan Support Vector Machine*. Telkom University.
- Nguyen, T. N., Tran, P. T., & Voznak, M. (2016). A Novel Compressed Sensing Approach to Speech Signal Compression. *Springer International Publishing Switzerland*, 75-85.
- Patil, N., & Nemade, Dr. M. U. (2017). Music Genre Classification Using Related papers Music Genre Classification Using MFCC , K-NN. *International Journal of Computer Engineering in Research Trends*, 4(2), 43-47.
- Sen, S., Dutta, A., & Dey, N. (2019). *Audio Processing and Speech Recognition: Concepts, Techniques and Research Overviews* (J. Kacprzyk, Ed.). Springer. <http://www.springer.com/series/10618>
- Shah, R. D., & Suthar, Dr. Anil. C. (2016). Speech Emotion Recognition Based on SVM Using MATLAB. *International Journal of Innovative Research in Computer and Communication Engineering (An ISO Certified Organization)*, 4(3), 2916-2921.
- Thiruvengatanadhan, R. (2018). Music Classification using MFCC and SVM. *International Research Journal of Engineering and Technology*, 5(9), 922-924.