Klasifikasi Instrumen Musik Menggunakan Metode Machine Learning

Theresia Margaretha Purba[a1], I Gusti Agung Gede Arya Kadyanan[a2]

[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia
[1]purba.2308561076@student.unud.ac.id
[2]gungde@unud.ac.id *(Corresponding Author)*

Abstract

Music plays an important role in human life, and automatic identification of musical instruments is becoming an increasingly relevant field in the digital era. This study aims to classify musical instrument types based on acoustic features using machine learning methods, specifically Support Vector Machine (SVM). The dataset used contains audio recordings of four instruments, namely guitar, piano, drum, and violin. Each audio file goes through a preprocessing process such as sample rate standardization, duration trimming, and framing. Furthermore, feature extraction is carried out from the time domain (Zero Crossing Rate and RMS), frequency domain (Spectral Centroid, Spread, and Roll-off), and cepstral domain (MFCC). The SVM model is trained with a combination of various features and evaluated using accuracy, precision, recall, and F1-score metrics. The experimental results show that the combination of all features produces the best accuracy of 68.33%. Although its performance is not optimal, these results show the potential of a feature-based approach for musical instrument classification and become the basis for further development using more complex methods such as deep learning.

Keywords: *Music Classification, Instrument Recognition, Audio Processing, Feature Extraction, Support Vector Machine, MFCC*

1. Pendahuluan

Music merupakan salah satu jenis seni yang memainkan peran penting dalam kehidupan manusia, keberagaman musik mencerminkan keberagaman budaya dan selera dari setiap orang[1]. Setiap alat musik memiliki karakteristik suara yang membedakannya dari alat musik lainnya. Dalam era digital saat ini, teknologi pemrosesan sinyal audio dan kecerdasan buatan telah berkembang pesat, membuka peluang untuk mengotomatisasi proses pengenalan dan klasifikasi suara alat musik.

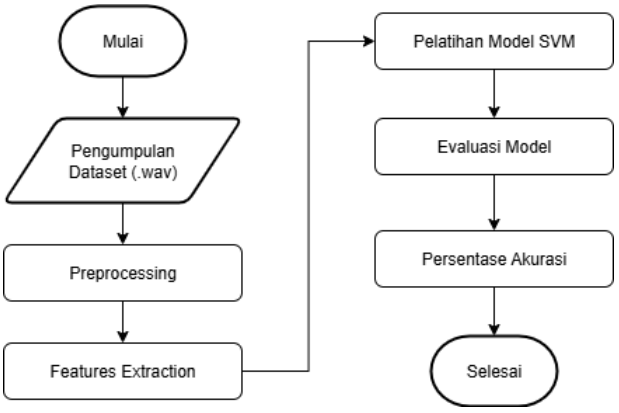
Pengenalan suara alat musik secara otomatis memiliki berbagai aplikasi praktis, mulai dari sistem rekomendasi musik, analisis komposisi musik, hingga bantuan untuk musisi dalam proses pembelajaran dan identifikasi instrumen [2]. Namun, tantangan utama dalam klasifikasi audio alat music terletak pada kompleksitas karakteristik akustik yang dimiliki setiap instrument, variasi dalam teknik bermain, serta pengaruh lingkungan perekaman.

Untuk mengatasi permasalahan tersebut, makalah ini mengusulkan pendekatan *machine learning*, khususnya *Support Vector Machine* (SVM) berbasis ekstraksi fitur. SVM telah terbukti efektif dalam tugas klasifikasi, kemudia dikombinasikan dengan ekstraksi fitur domain waktu, domain frekuensi, dan *Mel-Frequency Cepstral Coefficients* (MFCC) dapat memberikan representasi yang lebih lengkap terhadap karakteristik akustik suara alat music. Dalam jurnal penelitian oleh Chaudhary *et al*, (2021) menunjukkan hasil nilai akurasi mencapai 99%[3].

p-ISSN: 2986-3929
e-ISSN: 3032-1948

2. Metode Penelitian

2.1. Desain Penelitian



Gambar 1. Bagan Alur Penelitian

Penelitian diawali dengan pengumpulan dataset audio dengan format .wav, yang kemudian diproses melalui tahapan *Preprocessing* untuk standarisasi data. Setelah itu, dilakukan ekstraksi fitur guna memperoleh representasi numerik dari audio. Fitur yang dihasilkan digunakan untuk pelatihan model menggunakan algoritma SVM. Model yang telah dilatih kemudian dievaluasi, dan hasil evaluasi berupa persentasi akurasi yang menjadi tolak ukur performa model.

2.2. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini adalah *Musical Instrument Sound Dataset* yang diperoleh dari platform Kaggle . Di dalam dataset terdapat *file* untuk data *train* dan juga data *test.* Dataset ini terdiri dari *file* audio berformat .wav yang merekam suara berbagai instrumen musik, yaitu gitar, piano, drum, dan biola. Total terdapat 2628 *file* audio yang sudah dikategorikan ke dalam masing-masing kelas instrumen. Setiap kelas memiliki sejumlah audio dengan durasi yang bervariasi, yang selanjutnya akan dipotong menjadi segmen pendek (misalnya 3 detik) untuk diproses lebih lanjut. Dataset ini digunakan karena sudah memiliki label yang jelas dan format yang sesuai untuk keperluan klasifikasi berbasis audio.

2.3. Tahap Preprocessing

Tahap *preprocessing* dilakukan untuk menyiapkan data audio agar memiliki format seragam dan siap untuk dianalisis [4]. Setiap *file* audio dalam dataset dibaca menggunakan *library* librosa dengan *sampling rate* sebesar 22.050 Hz, lalu dikonversi dalam bentuk mono untuk menyederhanakan struktur sinyal suara. Mengingat panjang *file* audio yang bervariasi, maka dilakukan *trimming* terhadap audio, sehingga setiap *file* hanya berdurasi tiga detik. Jika audio berdurasi lebih dari tiga detik, dilakukan *padding* untuk menyesuaikan panjangnya. Langkah-langkah ini memastikan bahwa seluruh data audio memiliki panjang dan format yang konsisten sebelum masuk ke dalam tahap ekstraksi fitur.

2.4. Ekstraksi Fitur

Dalam penelitian ini dilakukan ekstraksi fitur audio dengan tiga jenis fitur, yaitu fitur domain waktu menggunakan *Zero Crossing Rate* (ZCR), dan RMS, kedua fitur domain frekuensi dengan *spectral centroind, spectral roll-off, spectral spread*, dan terakhir fitur MFCC mengambil 13 koefisen pertama dari transformasi *log power spectrum* menggunakan DCT.

a. *Zero Crossing Rate* (ZCR)

Pada bagian ZCR, setiap frame sinyal audio diukur frekuensi sinyal yang melintasi sumbu nol (*zero axis*).

b. *Root Mean Square* (RMS)

RMS digunakan untuk mengukur energi atau kenyaringan (*loudness*) sinyal dalam domain waktu.

c. Spectral Centroid

Pada bagian ini, “pusat massa” dari spektrum frekuensi sinyal diukur. Fitur ini sering dianggap sebagai indicator kecerahan (*brightness*) atau kualitas (timbre) suara.

d. Spread dan Roll-Off

Spectral spread mengukur seberapa tersebar energi frekuensi di sekitar spectral centroid. Nilai ini menggambarkan keragaman frekuensi, yang berguna dalam pengenalan suara, analisis *timbre*, dan klasifikasi sinyal audio. *Spectral roll-off* adalah frekuensi batas di mana sejumlah tertentu dari total energi spektrum tercakup. Ini berguna untuk membedakan suara dengan energi rendah (*bass*) dan energi tinggi (*treble/noise*).

e. MFCC

Ekstraksi fitur MFCC dari file audio dalam bentuk frame (framed audio), lalu menyimpannya dalam file CSV, mencakup rata-rata dan standar deviasi MFCC per file audio. Proses ekstraksi fitur ini membagi sinyal audio menjadi beberapa frame kecil, kemudian mengubah menggunakan FFT. Hasil FFT kemudian dihitung nilai *power spectrum*-nya dan dikonversi skala logaritmik. Selanjutnya *power spectrum* ini menggunakan *filterbbank Mel*, setelah itu DCT diterapkan untuk memperoleh koefisien MFCC.

2.5. Arsitektur Model

Model klasifikasi yang digunakan dalam penelitian ini adalah *Support Vector Machine Learning* (SVM). SVM merupakan algoritma *supervised learning* yang efektif untuk pemisahan kelas dalam ruang berdimensi tinggi, termasuk dalam kasus klasifikasi berbasis fitur audio seperti MFCC[5]. Data yang sudah diproses kemudian digunakan untuk melatih model SVM, pada kasus ini model yang digunakan adalah SVM. Pada bagian ini berbagai kombinasi fitur audio seperti MFCC, ZCR, RMS, dan spectral features di eksperimen untuk mencari kombinasi terbaik. Ada 3 parameter grid yang digunakan model ini C digunakan untuk mengontrol regularisasi diatur dengan [0.1, 1, 10], gamma untuk kernel RBF, jenis kernel (di sini hanya "rbf"). Dilakukan juga *K-Fold* dengan nilai *n_splits=5.*
Dalam pelatihan model SVM terdapat 3 tahap utama yaitu: pemilihan fitur, normalisasi (*scaling*), dan pelatihan dengan pencarian parameter terbaik. Pertama memilih kolom fitur tertentu dari list yang sudah ditentukan sebelumnya (misalnya hanya MFCC, atau kombinasi ZCR+RMS, dll), kemudian dilakukan normalisasi untuk mengubah setiap fitur menjadi distribusi dengan mean 0 dan standar deviasi 1, setelah itu SVM ditrain dengan *grid search* untuk mencari kombinasi parameter terbaik.

2.6. Evaluasi

Setelah model SVM dilatih menggunakan data latih, tahap selanjutnya adalah melakukan evaluasi terhadap performa model dengan menggunakan data uji. Evaluasi dilakukan untuk mengukur seberapa baik model mampu mengklasifikasikan jenis instrumen musik berdasarkan fitur yang telah diekstrak sebelumnya. Beberapa metrik evaluasi yang digunakan dalam penelitian ini meliputi, akurasi, precision, *recall*, dan F1-*score*, yang semuanya dihitung berdasarkan nilai-

nilai pada *confusion matrix*. *Confusion matrix* menunjukkan jumlah prediksi benar dan salah dari masing-masing kelas, sehingga dapat digunakan untuk mengukur efektivitas model dalam membedakan instrumen satu dengan yang lainnya.

Evaluasi dilakukan menggunakan fungsi *classification_report* dan *confusion_matrix* dari pustaka *scikit-learn*. Hasil evaluasi ini akan menjadi dasar dalam menganalisis performa model dan menentukan apakah metode klasifikasi yang digunakan sudah cukup baik atau masih memerlukana peningkatan.

3. Hasil dan Diskusi

3.1. Hasil *Preprocessing*

Pada tahap *preprocessing*, dilakukan serangkaian proses untuk menyiapkan data audio agar siap diproses lebih lanjut dalam tahap ekstraksi fitur. Dataset audio yang digunakan memiliki variasi panjang durasi dan format beragama, sehingga perlu dilakukan standarisasi agar proses klasifikasi dapat berjalan konsisten.

```
# target sample rate
target_sr = 44100
# target durasi
target_duration = 3.0  # detik
num_target_samples = int(target_sr * target_duration)

processed_count = 0

for fname in os.listdir(balanced_folder):
    if fname.endswith('.wav'):
        src = os.path.join(balanced_folder, fname)
        dst = os.path.join(standard_folder, fname)

        try:
            y, sr = librosa.load(src, sr=target_sr)

            if len(y) > num_target_samples:
                # potong bagian tengah
                center = len(y) // 2
                start = max(0, center - num_target_samples // 2)
                end = start + num_target_samples
                y = y[start:end]
            elif len(y) < num_target_samples:
                # pad dengan 0
                y = librosa.util.fix_length(y, size=num_target_samples)
```

Gambar 2. Tahap *resampling* data audio

Langkah pertama adalah normalisasi data audio untuk memastikan nilai amplitude berada dalam rentang standar, mencegah perbedaan skala yang dapat memengaruhi ekstraksi fitur. Selain itu, dilakukan proses *resampling* audio ke *sampling rate* yang seragam (44100 Hz) dan standarisasi durasi menjadi 3 detik, agar seluruh data berada pada resolusi frekuensi yang sama.

Setelah normalisasi, diterapkan proses *framing* dan *windowing*. Proses *framing* memecah sinyal audio menjadi potongan-potongan berdurasi pendek agar karakteristik sinyal dapat ditangkap dalam domain waktu yang lebih lokal. Selanjutnya setiap *frame* diberikan *window* fungsi (*Hamming window*) untuk mengurangi efek *spectral leakage.* Selanjutnya dilakukan pembersihan metadata, terutama untuk memastikan label kelas instrument sudah benar dan seragam, misalnya mengoreksi kesalahan penulisan label. Pembersihan metadata ini penting agar proses pelabelan pada model klasifikasi tidak mengalami error atau kebingungan saat proses *training* dan evaluasi.

3.2. Hasil Ekstraksi Fitur

Setelah proses *preprocessing*, dilakukan ekstraksi fitur audio dari seluruh data yang telah distandarisasi. Fitur yang diekstrak mencakup domain waktu, domain frekuensi, dan domain *cepstral* untuk mendapatkan representasi karakteristik sinyal suara secara lebih komprehensif.

p-ISSN: 2986-3929
e-ISSN: 3032-1948

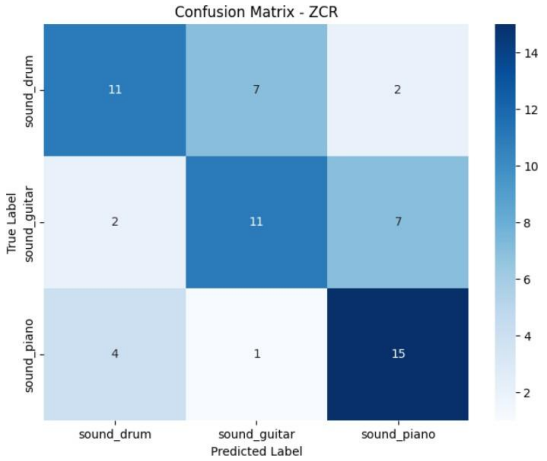
Fitur domain yang diambil antara lain ZCR dan RMS sebagai gambaran energi dan pola tanda sinyal. Dari domain frekuensi diekstraksi *Spectral Centroid, Spectral Spread*, serta *Spectral Roll-off*, yang menggambarkan distribusi energi spektral. Sementara dari domain *cepstral* diambil MFCC, yang banyak digunakan pada pengenalan pola suara karena meniru persepsi manusia terhadap suara.

Tabel 1. Tabel Hasil Eksperimen

Eksperimen	***Best Params***	***Best CV Accuracy***
ZCR	C=10, gamma=scale, kernel=rbf	0.7566
RMS	C=10, gamma=scale, kernel=rbf	0.6475
Spectral	C=10, gamma=scale, kernel=rbf	0.8416
MFCC_13	C=10, gamma=scale, kernel=rbf	0.9208
BestCombo	C=10, gamma=scale, kernel=rbf	0.9216
AllFeatures	C=10, gamma=scale, kernel=rbf	0.9616

3.3. Evaluasi Model

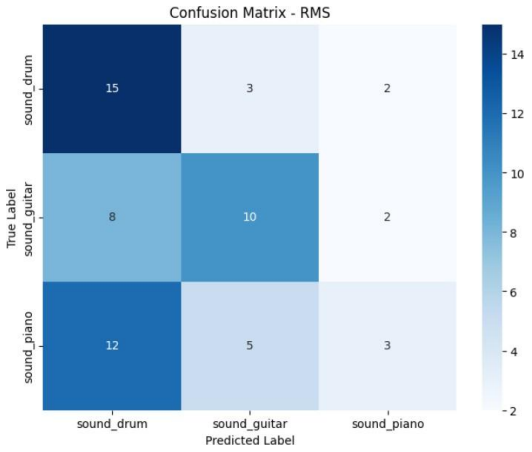
Evaluasi dilakukan menggunakan data uji terpisah yang telah diekstraksi fitur-fiturnya dengan metode yang sama seperti data latih. Proses evaluasi memanfaatkan model SVM terbaik dari masing-masing eksperimen, dengan parameter yang sudah diperoleh melalui *GridSearchCV.*



Gambar 3. *Confusion Matrix* Fitur ZCR

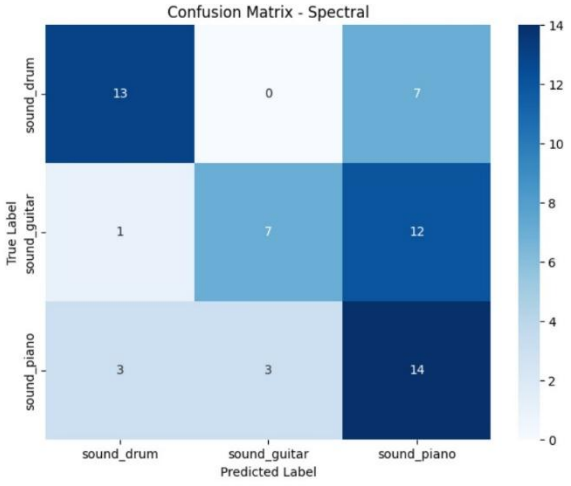
Confusion matrix menunjukkan prediksi banyak salah, terutama pada kelas *sound_gitar* dan *sound_piano* yang sering tertukar.

p-ISSN: 2986-3929
e-ISSN: 3032-1948



Gambar 4. *Confusion Matrix* Fitur RMS

Banyak prediksi benar pada *sound_drum*, namun sering keliru pada *sound_guitar* dan *sound_piano*.



Gambar 5. *Confusion Matrix Spectral*

Confusion matrix memperlihatkan prediksi yang lebih merata, terutama untuk *sound_piano* dan *sound_drum*.

Confusion matrix semakin membaik dengan kombinasi fitur. Pada MFCC_13, sebagian besar prediksi benar di semua kelas. Kemudian kombinasi dari ZCR, RMS, *spectral*, dan MFCC pada eksperimen *BestCombo* meningkatkan ketepatan prediksi di hampir semua kelas. Dan yang terakhir adalah eksperimen *AllFeatures* yang merupakan gabungan seluruh fitur, memiliki tingkat prediksi paling tinggi di semua kelas.

Tabel 2. Tabel Perbandingan Performa Tiap Eksperimen

Eksperimen	**Akurasi (%)**	***Precision* (%)**	***Recall* (%)**	**F1-*Score* (%)**
ZCR	61.67	61.70	61.67	61.35
RMS	46.67	47.09	46.67	43.13
Spectral	56.67	62.96	56.67	56.59
MFCC_13	60.00	59.01	60.00	59.16

p-ISSN: 2986-3929
e-ISSN: 3032-1948

Eksperimen	Akurasi (%)	*Precision* (%)	*Recall* (%)	F1-*Score* (%)
BestCombo	56.67	55.18	56.67	55.09
AllFeatures	68.33	67.97	68.33	67.55

Tabel di atas menunjukkan hasil evaluasi kinerja model SVM pada masing-masing eksperimen fitur. Terlihat bahwa penggunaan semua fitur gabungan (*AllFeatures*) menghasilkan performa terbaik dengan akurasi sebesar 68,33%, *precision* 67,97%, *recall* 68,33%, dan F1-*score* 67,55%. Sementara, eksperimen lain seperti MFCC_13 dan *Spectral* juga menunjukkan performa cukup baik dengan akurasi sekitar 60%.

Eksperimen berbasis fitur tunggal seperti ZCR dan RMS tampak memiliki performa lebih rendah, terutama RMS yang hanya mencapai akurasi 46,67%. Hal ini menunjukkan bahwa kombinasi fitur yang lebih kaya informasi mampu meningkatkan akurasi dalam klasifikasi instrumen musik dibandingkan fitur-fitur sederhana secara terpisah.

4. Kesimpulan

Penelitian ini berhasil mengimplementasikan metode *Support Vector Machine* (SVM) untuk melakukan klasifikasi instrument musik berbasis ciri-ciri audio yang diekstrak melalui domain waktu dan frekuensi. Hasil evaluasi menunjukkan bahwa penggunaan seluruh fitur gabungan (*AllFeatures*) memberikan performa terbaik dengan akurasi 68,33%. Meskipun demikian, tingkat akurasi tersebut masih tergolong sedang dan belum cukup optimal untuk diaplikasikan pada sistem nyata yang membutuhkan keandalan tinggi.

Hasil ini mengindikasikan bahwa metode *machine learning* klasik seperti SVM mampu mengenali pola audio instrumen, tetapi memerlukan pengembangan lebih lanjut agar kinerjanya meningkat. Penelitian di masa mendatang dapat mempertimbangkan penggunaan algoritma *deep learning* (misalnya CNN atau LSTM) yang lebih adaptif terhadap variasi sinyal audio, atau menambahkan teknik augmentasi data serta fitur akustik lain untuk memperkaya representasi data. Dengan demkian, akurasi klasifikasi instrumen musik diharapkan dapat ditingkatkan dan diaplikasikan secara lebih luas pada sistem pengenalan musik otomatis di masa depan.

Daftar Pustaka

[1] I Gusti Agung Istri Agrivina Shyta Devia and I Made Widiartha, “Klasifikasi Genre Musik Menggunakan Metode Support Vector Machine Dengan Multi-Kernel,” JNATIA, vol. 3, no. 1, pp. 127–132, Nov. 2024, [Online]. Available: https://www.kaggle.com

[2] S. K. Mahanta, A. F. U. Rahman Khilji, and P. Pakray, “Deep neural network for musical instrument recognition using MFCCs,” Computacion y Sistemas, vol. 25, no. 2, pp. 351–360, 2021, doi: 10.13053/CyS-25-2-3946.

[3] S. R. Chaudhary, S. N. Kakarwal, and R. R. Deshmukh, “Musical instrument recognition using audio features with integrated entropy method,” Article. [Online]. Available: http://pubs.iscience.in/jist

[4] S. P. Mohanty,” Musical Instrument's Sound Dataset” [Online]. Available: https://www.kaggle.com/datasets/soumendraprasad/musical-instruments-sound-dataset/data?select=Metadata_Train.csv. [Accessed: Jun. 20, 2025].

[5] R. Maulana and S. Redjeki, “Analisis Sentimen Pengguna Twitter Menggunakan Metode Support Vector Machine Berbasis Cloud Computing,” 2016. [Online]. Available: www.akakom.ac.id

p-ISSN: 2986-3929
e-ISSN: 3032-1948

Halaman ini sengaja dibiarkan kosong