

## **IMPLEMENTASI FAST FOURIER TRANSFORM DAN EKSTRAKSI SPEKTRAL UNTUK KLASIFIKASI GENDER SUARA MENGGUNAKAN SUPPORT VECTOR MACHINE**

Raditya Wahyu Pratomo<sup>1\*</sup>, Rakha Alviano Syaputra<sup>1</sup>, Ellena Nayla Firdausi<sup>1</sup>, Muhammad Noor Ridho<sup>1</sup>, Fadilah Raisa Husna<sup>1</sup>, Faisal Himamagung Prabowo<sup>1</sup>, Dino Ifandy<sup>1</sup>, Vera Noviana Sulistyawan<sup>1</sup>

<sup>1</sup>Program Studi Teknik Elektro, Fakultas Teknik, Universitas Negeri Semarang, Semarang, Indonesia

E-mail: [radityawp301105@students.unnes.ac.id](mailto:radityawp301105@students.unnes.ac.id)

### **Abstrak**

Perkembangan teknik pengolahan sinyal membuka peluang luas untuk pengenalan karakteristik suara manusia, termasuk klasifikasi gender. Penelitian ini mengembangkan sistem klasifikasi gender berbasis rekaman suara (.wav) dengan rangkaian pra-pemrosesan yang meliputi normalisasi, penghilangan noise, dan augmentasi data untuk memperkaya variasi sampel. Pada tahap ekstraksi, *Fast Fourier Transform* (FFT) digunakan untuk memperoleh parameter spektral (*mean amplitude*, *spectral centroid*, *bandwidth*, *roll-off*), sementara fitur akustik tambahan seperti MFCC, Chroma, dan *Zero Crossing Rate* (ZCR) turut diekstraksi untuk memperkaya representasi sinyal. Seluruh fitur kemudian distandarisasi menggunakan *StandardScaler* dan direduksi dimensinya melalui *Principal Component Analysis* (PCA) untuk mempertahankan sebagian besar variansi sekaligus mengurangi redundansi. Model klasifikasi dilatih menggunakan *Support Vector Machine* (SVM) dengan kernel *Radial Basis Function* (RBF) yang dioptimalkan melalui *GridSearchCV*. Pengujian menunjukkan model terbaik ( $C = 1$ ,  $\gamma = \text{'auto'}$ ) mencapai akurasi 98,54% dengan nilai precision dan recall di atas 98% pada kedua kelas dengan *Confusion Matrix* yang mendukung performa tinggi tersebut. Hasil ini menegaskan bahwa kombinasi *ekstraksi spektral* dan fitur akustik, dipadukan dengan reduksi dimensi dan penyetelan SVM, efektif untuk membedakan suara Male dan Female, sehingga berpotensi diterapkan pada sistem pengenalan suara dan autentikasi biometrik berbasis audio.

**Kata kunci:** FFT, SVM, MFCC, Klasifikasi Gender, Pengolahan Sinyal Digital

### **Abstract**

Advances in signal processing techniques have opened up vast opportunities for recognizing human voice characteristics, including gender classification. This study developed a gender classification system based on voice recordings (.wav) with a series of pre-processing steps, including normalization, noise removal, and data augmentation to enrich sample variation. At the extraction stage, FFT is used to obtain spectral parameters (*mean amplitude*, *spectral centroid*, *bandwidth*, *roll-off*), while additional acoustic features such as MFCC, Chroma, and ZCR are also extracted to enrich the signal representation. All features are then standardized using *StandardScaler* and reduced in dimension through *Principal Component Analysis* (PCA) to retain most of the variance while reducing redundancy. The classification model was trained using *Support Vector Machine* (SVM) with *Radial Basis Function* (RBF) kernel optimized through *GridSearchCV*. Testing showed that the best model ( $C = 1$ ,  $\gamma = \text{'auto'}$ ) achieved 98.54% accuracy with precision and recall values above 98% in both classes with a *Confusion Matrix* that supported this high performance. These results confirm that the combination of spectral extraction and acoustic features, combined with dimension reduction and SVM tuning, is effective for distinguishing between male and female voices, making it potentially applicable to audio-based voice recognition and biometric authentication systems.

**Keywords:** FFT, SVM, MFCC, Gender Classification, Digital Signal Processing

## 1. Pendahuluan

Perkembangan teknologi digital telah merevolusi cara manusia berkomunikasi, belajar, dan bekerja sehingga analisis sinyal digital menjadi semakin penting dalam berbagai bidang ilmu terapan[1]. Digitalisasi sinyal audio memungkinkan penggunaan teknik pengolahan sinyal digital (PSD) untuk mengekstraksi ciri-ciri yang informatif serta mendukung aplikasi pengenalan otomatis seperti identifikasi pembicara dan pengenalan gender[2]. Transformasi *Fast Fourier Transform* (FFT) digunakan karena mampu secara efisien mengubah sinyal dari domain waktu ke domain frekuensi, sehingga memungkinkan analisis komponen spektral seperti frekuensi dominan dan harmonisa dengan kebutuhan komputasi yang relatif rendah. Dengan representasi frekuensi tersebut, analisis spektrogram dan ekstraksi fitur frekuensi-domain menjadi lebih mudah dan informatif untuk keperluan karakterisasi suara[3].

Akurasi akan meningkat apabila dilakukan penggabungan dengan menggabungkan ekstraksi beberapa jenis fitur: MFCC (*Mel-Frequency Cepstral Coefficients*) yang meniru seperti skala pendengaran manusia dan efektif menangkap timbre suara [4]. Fitur kromatik (chroma) juga dimasukkan untuk merepresentasikan distribusi energi ke dalam kelas pitch/harmonik dan menangkap aspek harmonis sinyal vokal [5]. *Zero-Crossing Rate* (ZCR) dipakai sebagai fitur domain-waktu yang sederhana namun informatif untuk membedakan segmen *voiced* dan *unvoiced* dalam ucapan [6]. Selain itu, fitur-fitur spektral seperti *spectral centroid*, *spectral bandwidth*, dan *spectral roll-off* melengkapi representasi dengan ringkasan bentuk spektrum yang menggambarkan “kecerahan” dan penyebaran energi frekuensi [7].

Penggabungan lintas-domain fitur-fitur tersebut umumnya meningkatkan kemampuan diskriminatif vektor ciri dibandingkan penggunaan satu jenis fitur saja, karena masing-masing fitur menangkap aspek sinyal yang berbeda. Namun, kombinasi fitur sering menghasilkan vektor berdimensi tinggi sehingga menimbulkan masalah *curse of dimensionality*, beban komputasi lebih besar, dan potensi *overfitting* jika tidak ditangani[8]. Mengatasi permasalahan tersebut, maka penelitian ini menerapkan *Principal Component Analysis* (PCA) sebagai teknik reduksi dimensi yang memproyeksikan data ke subruang ortogonal, dengan tujuan mempertahankan variansi terbesar sehingga jumlah fitur dapat dikurangi tanpa kehilangan informasi penting secara berarti. [9].

Algoritma klasifikasi seperti *supervised* difungsikan untuk tugas pengenalan gender (masalah klasifikasi biner), *Support Vector Machine* (SVM) dipilih karena kemampuan margin-maximization-nya yang efektif pada ruang berdimensi tinggi[10]. Untuk memastikan konfigurasi hiperparameter yang optimal, prosedur

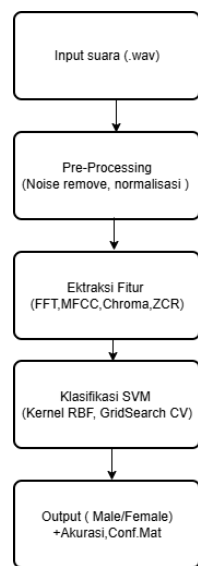
*GridSearchCV* diterapkan sehingga pencarian kombinasi parameter (mis. C, kernel,  $\gamma$ ) dievaluasi melalui validasi silang untuk menghindari pemilihan parameter yang overfit[11]. Evaluasi kinerja model dilakukan dengan menggunakan metrik standar seperti *accuracy*, *precision*, *recall*, dan *F1-score*, sehingga hasil yang diperoleh dapat diinterpretasikan secara menyeluruh serta memungkinkan perbandingan dengan penelitian lain[12]. Selain metrik numerik, matriks kebingungan, kurva ROC, dan visualisasi *Precision-Recall* digunakan untuk memahami tipe kesalahan klasifikasi dan *trade-off* pada berbagai ambang keputusan [13].

Pendekatan gabungan yang meliputi transformasi spektral menggunakan FFT, ekstraksi fitur MFCC, chroma, ZCR, dan fitur spektral, normalisasi serta reduksi dimensi dengan PCA, diikuti klasifikasi menggunakan SVM yang di-tuning melalui *GridSearchCV* dan evaluasi dengan metrik standar, telah banyak diterapkan dan tervalidasi dalam studi pengenalan suara sehingga memberikan dasar empiris yang kuat untuk tugas klasifikasi gender suara [14] [15]. Penerapan implementasi praktis dapat memanfaatkan pustaka open-source seperti *librosa* untuk ekstraksi fitur audio dan *scikit-learn* untuk tahap PCA, *GridSearchCV*, SVM, dan pengukuran metrik sehingga eksperimen menjadi *reproducible* [16]. Permasalahan pemisahan suara manusia inilah yang mendorong penulis mengusulkan fokus penelitian pada Implementasi *Fast Fourier Transform* dan Ekstraksi Fitur Spektral untuk Klasifikasi Gender Suara Menggunakan *Support Vector Machine* dengan tujuan menerapkan transformasi spektral, ekstraksi fitur gabungan, reduksi dimensi, dan klasifikasi tertuning untuk menghasilkan sistem pengenalan gender berbasis suara yang cepat dan akurat.

## 2. Metode

### 2.1 Desain Umum Penelitian

Penelitian ini dilaksanakan dengan pendekatan simulasi eksperimental guna menganalisis performa sistem klasifikasi gender berbasis suara. Metodologi yang diterapkan mengombinasikan *Fast Fourier Transform* (FFT) sebagai metode ekstraksi fitur dan *Support Vector Machine* (SVM) sebagai algoritma klasifikasi berbasis pembelajaran terawasi dalam konteks Machine Learning. Sistem yang dikembangkan terdiri atas tiga komponen utama, yaitu dataset suara, modul ekstraksi fitur, dan modul klasifikasi. Dataset terdiri dari kumpulan file audio berformat .wav yang telah dilabeli sebagai Male dan Female sebagai target klasifikasi. Diagram blok sistem dalam penelitian ini ditunjukkan pada Gambar 1..



Gambar 1. Diagram Blok Sistem

Tahap awal penelitian dimulai dari input suara (.wav) yang diperoleh dari dataset rekaman. Setiap file audio kemudian melewati tahap pre-processing, yang bertujuan meningkatkan kualitas sinyal suara. Proses ini mencakup normalisasi amplitudo agar energi sinyal seragam, penghapusan *noise*, serta pemangkasan bagian senyap (*silence trimming*). Untuk memperluas variasi data tanpa perlu rekaman ulang, dilakukan augmentasi audio menggunakan teknik *pitch shifting* dan *time stretching* melalui fungsi `librosa.effects.pitch_shift()` dan `librosa.effects.time_stretch()`.

Selanjutnya, Setelah sinyal dibersihkan, proses berlanjut ke tahap ekstraksi fitur. Pada tahap ini, sinyal ditransformasikan dari domain waktu ke domain frekuensi menggunakan *Fast Fourier Transform* (FFT) guna memperoleh representasi spektralnya. Berdasarkan hasil FFT tersebut, dihitung berbagai parameter seperti *mean amplitude*, *standard deviation*, *spectral centroid*, *bandwidth*, *roll-off*, serta *zero crossing rate* (ZCR).. Fitur tambahan seperti MFCC (*Mel-Frequency Cepstral Coefficients*) dan *Chroma Features* juga diekstraksi guna merepresentasikan karakteristik persepsi pendengaran manusia. Seluruh fitur ini digabungkan menjadi satu vektor numerik berdimensi tinggi sebagai input bagi model klasifikasi.

Langkah selanjutnya yaitu proses klasifikasi dengan memanfaatkan algoritma *Support Vector Machine* (SVM) yang menggunakan kernel *Radial Basis Function* (RBF). *Pipeline* pelatihan model terdiri dari tiga tahap: *StandardScaler* untuk normalisasi, *PCA* (*Principal Component Analysis*) untuk reduksi dimensi hingga 95% varian, dan SVM sebagai algoritma utama. Optimasi parameter dilakukan menggunakan *GridSearchCV* dengan

validasi silang 5-fold untuk menemukan kombinasi *C*, *gamma*, dan kernel terbaik.

Tahap terakhir merupakan proses keluaran, di mana sistem memberikan hasil prediksi gender suara (Male/Female) serta menampilkan metrik evaluasi seperti akurasi, *precision*, *recall*, *F1-score*, dan *Confusion Matrix* untuk menilai kinerja model. Selain itu, visualisasi plot rata-rata FFT antara suara laki-laki dan perempuan digunakan untuk mengamati perbedaan pola spektrum frekuensi pada masing-masing gender.

## 2.2 Desain Sistem FFT-SVM

Desain sistem dalam penelitian ini difokuskan pada pengembangan model klasifikasi gender berbasis suara dengan memanfaatkan pendekatan *Fast Fourier Transform* (FFT) dan *Support Vector Machine* (SVM). Sistem ini dibuat agar dapat mengidentifikasi perbedaan karakteristik suara antara Male dan Female berdasarkan pola spektrum frekuensi serta fitur akustik yang diperoleh dari tahap ekstraksi fitur.

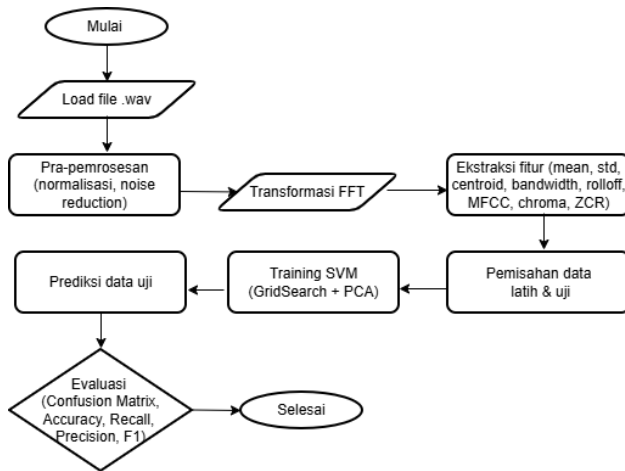
Secara keseluruhan, sistem FFT-SVM terbagi menjadi dua tahap utama, yakni ekstraksi fitur dan klasifikasi. Pada tahap ekstraksi fitur, sinyal suara dikonversi dari domain waktu ke domain frekuensi untuk memperoleh parameter spektral yang menggambarkan karakteristik khas suara manusia. Sementara itu, tahap klasifikasi bertujuan mempelajari pola dari data fitur tersebut menggunakan algoritma Machine Learning, sehingga sistem dapat secara otomatis memprediksi label gender pada suara yang baru..

Setiap tahapan pada sistem saling berhubungan secara berurutan, mulai dari proses akuisisi data suara, pra-pemrosesan, transformasi sinyal dengan FFT, pengambilan fitur tambahan seperti MFCC, Chroma, dan ZCR, hingga pelatihan model SVM dengan optimasi parameter menggunakan *GridSearchCV*. Dalam desain sistem ini diharapkan dapat menghasilkan model klasifikasi dengan akurasi tinggi dan mampu mengenali karakteristik suara manusia secara efisien.

### 2.2.1 Alur Kerja Sistem FFT-SVM

Alur kerja sistem FFT-SVM pada penelitian ini menggambarkan tahapan pemrosesan suara dari awal hingga akhir. Proses dimulai dari akuisisi data suara dalam format .wav, kemudian dilakukan pra-pemrosesan untuk meningkatkan kualitas sinyal. Setelah itu, dilakukan transformasi *Fast Fourier Transform* (FFT) dan ekstraksi berbagai fitur akustik penting seperti MFCC, Chroma, dan ZCR. Fitur yang telah diekstraksi selanjutnya diproses melalui tahap standarisasi dan reduksi dimensi menggunakan PCA, sebelum dimanfaatkan oleh algoritma *Support Vector Machine* (SVM) untuk melakukan klasifikasi gender. Hasil klasifikasi tersebut kemudian dievaluasi menggunakan metrik kinerja seperti akurasi,

presisi, recall, dan F1-score. Alur kerja sistem tersebut dapat dilihat pada Gambar 2.



Gambar 2. Alur Kerja Sistem FFT-SVM

Gambar 2 menunjukkan alur kerja sistem klasifikasi suara berbasis FFT-SVM yang digunakan dalam penelitian ini. Proses dimulai dari tahap akuisisi data suara, di mana file audio berformat .wav dimuat sebagai masukan sistem. Tahap selanjutnya adalah pra-pemrosesan sinyal, yang mencakup normalisasi amplitudo dan reduksi noise untuk meningkatkan kualitas sinyal sebelum analisis frekuensi dilakukan.

Selanjutnya, sinyal yang telah melalui proses pembersihan ditransformasikan dari domain waktu ke domain frekuensi menggunakan *Fast Fourier Transform* (FFT). Transformasi ini memungkinkan sistem menangkap informasi spektral yang menjadi dasar analisis karakteristik suara. Pada tahap ekstraksi fitur, sistem kemudian menghitung sejumlah parameter utama seperti *mean amplitude*, *standard deviation*, *spectral centroid*, *bandwidth*, dan *roll-off*. Selain itu, fitur akustik tambahan seperti MFCC (*Mel-Frequency Cepstral Coefficients*), Chroma, serta *Zero Crossing Rate* (ZCR) juga diambil untuk memperkaya representasi karakteristik sinyal suara.

Data hasil ekstraksi kemudian dipisahkan menjadi data pelatihan dan data pengujian. Pada tahap pelatihan, algoritma *Support Vector Machine* (SVM) digunakan dengan bantuan *GridSearchCV* untuk menentukan kombinasi parameter terbaik, sementara *Principal Component Analysis* (PCA) diterapkan guna mereduksi dimensi agar proses klasifikasi lebih efisien. Model yang telah dilatih selanjutnya digunakan untuk memprediksi data uji, dan hasil klasifikasi dievaluasi menggunakan metrik performa seperti *Confusion Matrix*, *Accuracy*, *Precision*, *Recall*, serta *F1-Score*. Melalui tahapan ini, sistem diharapkan mampu mengklasifikasikan gender suara secara otomatis dengan tingkat akurasi tinggi dan performa yang stabil.

## 2.2.2 Ekstraksi Fitur Spektral dan Akustik

*Fast Fourier Transform* (FFT) diterapkan untuk mentransformasikan sinyal dari domain waktu ke domain frekuensi sehingga karakteristik spektralnya dapat dianalisis dengan lebih mudah. Secara matematis, FFT merupakan versi efisien dari *Discrete Fourier Transform* (DFT) yang dapat dinyatakan dengan persamaan berikut:

$$F(k) = \frac{1}{N} \sum_{n=0}^{N-1} X(n) e^{-j2\pi \frac{nk}{N}} \quad (1)$$

Keterangan:

$F(k)$  = representasi sinyal pada domain frekuensi ke- $k$

$X(n)$  = sinyal pada domain waktu pada indeks  $n$

$N$  = jumlah total sampel sinyal

$n$  = indeks sampel waktu ( $n = 0, 1, 2, \dots, N-1$ )

$k$  = indeks frekuensi diskrit ( $k = 0, 1, 2, \dots, N-1$ )

$e^{-j2\pi \frac{nk}{N}}$  = faktor kompleks yang berfungsi sebagai basis transformasi

Transformasi tersebut menghasilkan spektrum frekuensi yang menunjukkan distribusi energi sinyal pada setiap komponen frekuensinya. Nilai magnitudo  $|X[k]|$  digunakan untuk mengekstraksi fitur utama, sedangkan nilai fasenya diabaikan karena tidak relevan untuk klasifikasi gender suara. Beberapa fitur utama yang diambil meliputi:

1. **Mean Amplitude ( $\mu$ )**: nilai rata-rata amplitudo dari spektrum frekuensi.
2. **Standard Deviation ( $\sigma$ )**: ukuran penyebaran energi sinyal terhadap frekuensi *Standard Deviation* ( $\sigma$ ).
3. **Spectral Centroid ( $f_c$ )**: menunjukkan pusat massa spektrum, menggambarkan "kecerahan" suara.
4. **Spectral Bandwidth (BW)**: lebar pita frekuensi tempat sebagian besar energi terkonsentrasi.
5. **Spectral Roll-off ( $f_r$ )**: frekuensi di mana 85% energi spektrum terkandung.

Selain fitur yang diperoleh dari FFT, sistem juga mengintegrasikan fitur akustik tambahan seperti *Mel-Frequency Cepstral Coefficients* (MFCC), *Chroma Feature*, dan *Zero Crossing Rate* (ZCR). MFCC berfungsi untuk meniru cara pendengaran manusia dalam mengenali skala mel, Chroma merepresentasikan distribusi energi pada tiap nada musik (C-B), sementara ZCR digunakan untuk membedakan pola sinyal *voiced* dan *unvoiced*. Seluruh fitur tersebut kemudian dihitung nilai rata-ratanya (*mean*) serta simpangan bakunya (*standard deviation*) guna membentuk satu vektor numerik berdimensi tinggi yang akan digunakan pada proses klasifikasi.

### 2.2.3 Reduksi Dimensi dan Pelatihan Model

Fitur yang telah diekstraksi kemudian dinormalisasi menggunakan *StandardScaler()* agar seluruh fitur berada pada skala yang sebanding. Setelah itu, diterapkan *Principal Component Analysis* (PCA) dengan jumlah komponen yang mempertahankan 95% variansi data untuk mengurangi redundansi serta mempercepat proses pelatihan. Proses klasifikasi dilakukan menggunakan *Support Vector Machine* (SVM) dengan kernel *Radial Basis Function* (RBF) Optimasi parameter dilakukan dengan menggunakan *GridSearchCV* dengan kombinasi:

$$C = [1, 10, 100, 1000]$$

$$\gamma = ['scale', 'auto', 0.01, 0.001]$$

Proses validasi silang dilakukan menggunakan *5-fold cross-validation* untuk memilih parameter terbaik berdasarkan skor akurasi rata-rata tertinggi.

### 2.3. Evaluasi Kinerja Sistem

Evaluasi kinerja dilakukan untuk menilai seberapa akurat sistem berbasis *Fast Fourier Transform* (FFT) dan *Machine Learning Support Vector Machine* (SVM) dalam mengklasifikasikan suara, dengan membandingkan hasil prediksi model terhadap label sebenarnya pada data uji menggunakan *confusion matrix*. *Confusion matrix* terdiri dari empat parameter utama, yaitu:

- TP (True Positive):** jumlah data Male yang berhasil diklasifikasikan dengan benar sebagai Male.
- TN (True Negative):** jumlah data Female yang diklasifikasikan dengan benar sebagai Female.
- FP (False Positive):** jumlah data Female yang keliru diklasifikasikan sebagai Male.
- FN (False Negative):** jumlah data Male yang salah terklasifikasi sebagai Female.

Dari nilai-nilai tersebut, dihitung beberapa metrik evaluasi utama sebagai berikut:

#### 1. Akurasi (Accuracy)

Akurasi merepresentasikan persentase data yang berhasil diklasifikasikan dengan benar dari total keseluruhan data uji. Semakin tinggi nilai akurasi, semakin rendah tingkat kesalahan yang terjadi dalam proses klasifikasi. Perhitungan akurasi dapat ditunjukkan pada rumus (2) berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

#### 2. Presisi (Precision)

Presisi menunjukkan tingkat ketepatan sistem dalam mengklasifikasikan data positif (Male). Presisi tinggi menunjukkan bahwa sebagian besar data yang diprediksi positif benar-benar termasuk ke dalam kelas tersebut. Perhitungan presisi ditunjukkan pada rumus (3) sebagai berikut:

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

#### 3. Sensitivitas (Recall)

Recall digunakan untuk menilai sejauh mana sistem mampu mengenali seluruh data positif yang sebenarnya ada. Nilai recall yang tinggi menunjukkan bahwa sistem dapat mendeteksi sebagian besar data yang benar-benar termasuk kategori positif. Perhitungan recall atau sensitivitas ditunjukkan pada rumus (4) berikut:

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

#### 4. F1-Score

F1-Score adalah rata-rata harmonik antara nilai presisi dan recall digunakan untuk mengukur keseimbangan antara tingkat ketepatan dan kelengkapan dalam proses klasifikasi. Metrik ini sangat berguna terutama ketika distribusi data antar kelas tidak seimbang. Perhitungan F1-Score ditunjukkan pada rumus (5) berikut:

$$F1-Score = \frac{TP}{TP+FP} \quad (5)$$

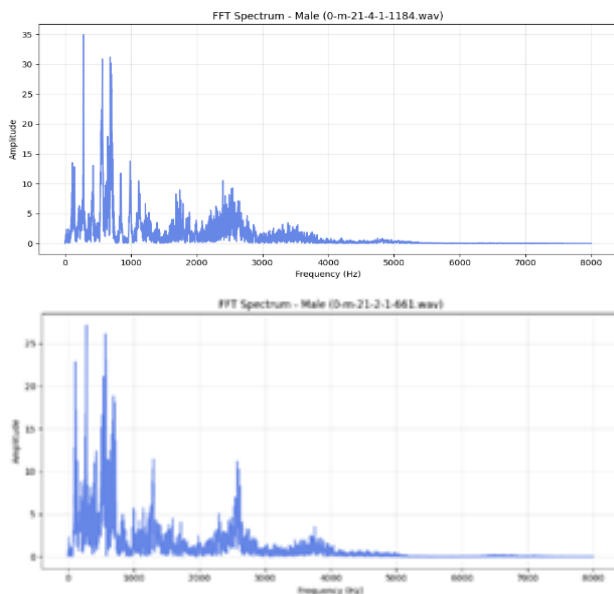
### 3. Hasil dan Pembahasan

Hasil yang diperoleh meliputi analisis spektrum suara untuk membedakan karakteristik antara suara Male dan Female, hasil ekstraksi fitur spektral dan akustik yang menjadi representasi numerik dari sinyal suara, serta performa sistem klasifikasi yang diukur menggunakan algoritma *Support Vector Machine* (SVM). Pembahasan dilakukan dengan meninjau keterkaitan antara hasil pengujian, karakteristik spektrum frekuensi, dan kinerja model dalam mengenali pola suara berdasarkan gender.

#### 3.1 Hasil Spektrum Suara Male dan Female

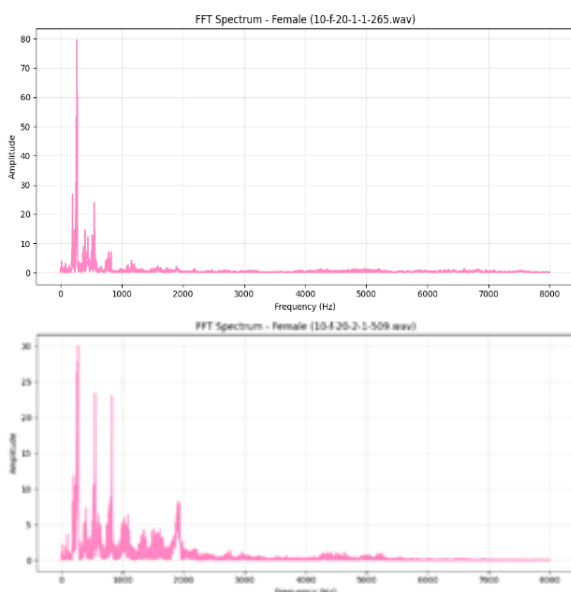
Hasil Spektrum dapat digunakan untuk memahami perbedaan karakteristik suara antara laki-laki dan perempuan. Spektrum suara memperlihatkan distribusi energi pada berbagai frekuensi, sehingga dapat memberikan gambaran mengenai frekuensi dasar, harmonik, serta pola distribusi amplitudo yang khas pada masing-masing gender. Dibawah ini hasil Spektrum male dan female secara sampling, serta rata-rata dari laki-laki dan perempuan.

Spektrum suara laki-laki umumnya menunjukkan energi yang terpusat pada frekuensi rendah. Frekuensi dasar (F0) pria dewasa biasanya berada di kisaran sekitar 85–180 Hz (dalam beberapa sumber terluas 65–260 Hz), sehingga puncak amplitudo utama pada FFT sering muncul di sekitar 95–120 Hz. Selain fundamental, terdapat deretan harmonik dan formant—resonansi saluran vokal—yang juga berkontribusi pada bentuk Spektrum; namun secara umum energi menurun ketika frekuensi naik, sehingga Spektrum pria tampak “lebih berat” atau lebih dominan di frekuensi rendah. Dimana hasil Spektrum Male dapat dilihat pada Gambar 3.



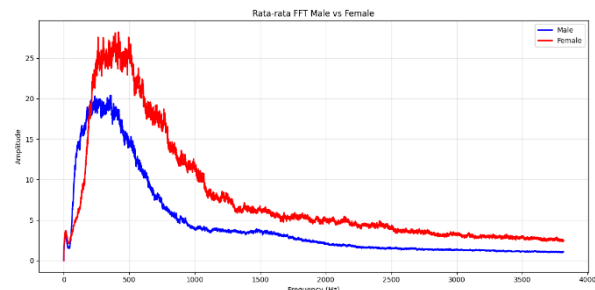
Gambar 3. Hasil spektrum Male

Spektrum suara perempuan cenderung bergeser ke frekuensi yang lebih tinggi. F0 perempuan dewasa biasanya di kisaran 165–255 Hz (rentang umum bisa 100–525 Hz), sehingga puncak dominan pada contoh sering berada di sekitar 199–210 Hz. Karena pita suara perempuan lebih pendek dan tipis, getarannya lebih cepat sehingga lebih banyak energi di frekuensi menengah-tinggi dan harmonik atas yang relatif lebih kuat menghasilkan karakter suara yang lebih “nyaring” atau “terang” pada spektrum. Hasil Spektrum dapat dilihat pada Gambar 4.



Gambar 4. Hasil spektrum Female

Spektrum suara Male dan Female telah dihasilkan dan menampilkan hasil spektrum satu file .wav menghasilkan satu spektrum. Selanjutnya, dilakukan penghitungan rata-rata antara suara Male dan Female yang ditunjukkan pada Gambar 5.



Gambar 5. Hasil rata-rata spektrum Male dan Female

Berdasarkan Gambar 5, terlihat perbedaan jelas antara spektrum suara Male dan Female. Suara Female memiliki amplitudo puncak lebih tinggi dengan dominasi energi pada frekuensi menengah (200–500 Hz), sedangkan suara Male menunjukkan puncak pada frekuensi rendah (100–250 Hz). Hal ini menunjukkan bahwa suara perempuan memiliki frekuensi dasar lebih tinggi dan sebaran harmonisa yang lebih luas dibandingkan laki-laki. Perbedaan pola spektrum inilah yang menjadi dasar utama dalam proses ekstraksi fitur dan klasifikasi oleh model SVM.

### 3.2 Hasil Ekstraksi Fitur

Tahap ekstraksi fitur dilakukan untuk mengubah sinyal frekuensi hasil FFT menjadi parameter numerik yang dapat digunakan oleh model klasifikasi. Fitur utama yang diekstraksi meliputi *Mean Amplitude*, *Standard Deviation*, *Spectral Centroid*, *Spectral Bandwidth*, dan *Spectral Roll-off*. Selain itu, sistem juga menambahkan fitur akustik tambahan seperti MFCC, Chroma, dan *Zero Crossing Rate* (ZCR). Hasil ekstraksi fitur ditunjukkan pada Tabel 1.

Tabel 1. Rata-rata Fitur Spektral Suara Male dan Fe Male

Fitur	Female	Male
Mean Amplitude	5.263	2.993
Standard Deviation	12.247	7.236
Spectral Centroid (Hz)	1782.98	1579.85
Spectral Bandwidth (Hz)	1702.99	1643.43
Roll-off (Hz)	3704.67	3256.80

Tabel 1 menjelaskan bahwa nilai Mean Amplitude dan *Spectral Centroid* pada suara Female lebih tinggi dibandingkan Male. Hal ini menunjukkan bahwa suara perempuan memiliki energi dominan di frekuensi yang lebih tinggi. Sebaliknya, nilai *Mean Amplitude* pada suara Male lebih rendah dengan persebaran energi yang cenderung berada di frekuensi rendah. Nilai *Spectral*

*Bandwidth* dan *Roll-off* yang sedikit lebih besar pada Fe Male juga menandakan bahwa suara perempuan memiliki sebaran frekuensi yang lebih luas dibandingkan suara laki-laki. Perbedaan nilai-nilai fitur ini menunjukkan bahwa hasil ekstraksi berbasis FFT berhasil menangkap karakteristik khas dari masing-masing gender dan dapat menjadi representasi numerik yang efektif bagi algoritma SVM.

### 3.3 Hasil Klasifikasi

Tahap klasifikasi dilakukan menggunakan algoritma *Support Vector Machine* (SVM) dengan kernel *Radial Basis Function* (RBF). Optimasi parameter dilakukan melalui *GridSearchCV*, dan diperoleh parameter terbaik pada kombinasi  $C = 1$ ,  $\gamma = \text{'auto'}$ , serta kernel = 'rbf'. Model diuji menggunakan data uji sebanyak  $\pm 75\text{--}80\%$  dari total dataset dengan metode stratified sampling untuk menjaga keseimbangan antara kelas Male dan Female. Hasil evaluasi performa sistem ditunjukkan pada Tabel 2, yang memuat nilai *precision*, *recall*, dan F1-score untuk masing-masing kelas suara.

Tabel 2. Confusion Matrix Hasil Klasifikasi

Asli \ Prediksi	Female	Male	Total
Female	6	748	754
Male	555	14	569
Total	561	762	1323

Berdasarkan hasil pada Tabel 2, dari 754 sampel suara Male, sebanyak 748 (99,2%) diklasifikasikan dengan benar dan 6 (0,8%) salah dikenali sebagai Female. Sementara itu, dari 569 sampel suara Female, sebanyak 555 (97,5%) terklasifikasi dengan benar dan 14 (2,5%) salah terprediksi sebagai Male. Distribusi hasil prediksi yang hampir sempurna menunjukkan bahwa sistem klasifikasi berbasis SVM mampu mengenali pola karakteristik suara dari kedua kelas dengan tingkat kesalahan yang sangat kecil. Total akurasi 98,54%, sistem terbukti memiliki performa klasifikasi yang sangat tinggi, menandakan keberhasilan proses ekstraksi fitur dengan spektral dan akustik dalam merepresentasikan perbedaan nyata antara suara Male dan Female.

### 3.4 Hasil Evaluasi Kinerja

Evaluasi performa model dilakukan menggunakan metrik *Precision*, *Recall*, F1-score, dan *Accuracy*. Tabel 3 menampilkan hasil evaluasi yang dihasilkan dari pengujian model SVM terbaik menggunakan data uji sebesar  $\pm 75\text{--}80\%$  dari total dataset.

Tabel 3. Hasil Evaluasi Model SVM

Kelas	Precision	Recall	F1-Score
Female	0.99	0.98	0.98
Male	0.98	0.99	0.99
Accuracy			98.54%
Macro Avg	0.99	0.98	0.99
Weighted Avg	0.99	0.99	0.99

Dari Tabel 3, didapatkan bahwa nilai akurasi 98,54% menunjukkan bahwa model bekerja dengan sangat baik dalam mengenali pola suara antar gender. Nilai *Precision* dan *Recall* yang hampir sempurna menandakan bahwa sistem minim kesalahan dalam prediksi baik terhadap kelas Male maupun Female. Tingginya F1-score ini juga mengindikasikan keseimbangan baik antara kemampuan sistem mendeteksi dan menghindari kesalahan klasifikasi. Dengan demikian, model SVM berbasis fitur FFT dan fitur akustik tambahan (MFCC, Chroma, dan ZCR) terbukti mampu memberikan hasil klasifikasi yang stabil, akurat, dan konsisten. Kinerja tinggi ini didukung oleh penerapan pra-pemrosesan sinyal yang optimal, teknik augmentasi data, dan tuning parameter SVM yang tepat.

## 4. Kesimpulan

Berdasarkan hasil analisis spektrum, ekstraksi fitur, dan pengujian klasifikasi yang telah dilakukan, dapat disimpulkan bahwa sistem klasifikasi gender suara yang dikembangkan mampu membedakan karakteristik suara Male dan Female secara efektif. Transformasi *Fast Fourier Transform* (FFT) berhasil menampilkan perbedaan pola spektrum frekuensi, di mana suara Male didominasi oleh energi pada frekuensi rendah, sedangkan suara Female menunjukkan dominasi energi pada frekuensi menengah hingga tinggi. Perbedaan ini selaras dengan hasil ekstraksi fitur spektral seperti *spectral centroid*, *bandwidth*, dan *roll-off* yang secara konsisten bernilai lebih tinggi pada kelas Female.

Penggabungan fitur spektral berbasis FFT dengan fitur akustik tambahan berupa MFCC, Chroma, dan *Zero Crossing Rate* (ZCR) terbukti mampu merepresentasikan karakteristik suara secara lebih komprehensif. Penerapan *StandardScaler* dan reduksi dimensi menggunakan *Principal Component Analysis* (PCA) dengan mempertahankan 95% variansi data berhasil menekan redundansi fitur tanpa mengurangi informasi penting, sehingga meningkatkan efisiensi dan stabilitas proses klasifikasi.

Hasil pengujian menunjukkan bahwa model *Support Vector Machine* (SVM) dengan kernel *Radial Basis Function* (RBF) dan parameter optimal ( $C = 1$ ,  $\gamma = \text{'auto'}$ ) mencapai akurasi sebesar 98,54%, dengan nilai *precision*, *recall*, dan F1-score di atas 98% untuk kedua kelas. *Confusion Matrix* juga menunjukkan tingkat kesalahan



klasifikasi yang sangat rendah, menandakan bahwa model mampu mengenali pola suara *Male* dan *Female* secara konsisten.

Dengan demikian, pendekatan FFT-SVM yang dikombinasikan dengan ekstraksi fitur spektral dan akustik, reduksi dimensi PCA, serta optimasi hiperparameter melalui GridSearchCV dapat dinyatakan efektif dan andal untuk tugas klasifikasi gender berbasis suara. Sistem ini berpotensi untuk dikembangkan lebih lanjut dan diterapkan pada aplikasi pengenalan suara, autentikasi biometrik berbasis audio, maupun sistem cerdas yang memanfaatkan karakteristik suara manusia.

## Referensi

- [1]. P. Paidil and S. Sari, "Peran Teknologi Terbaru Membentuk Kehidupan Di Era Digital," *JUPEIS : Jurnal Pendidikan dan Ilmu Sosial*, vol. 4, no. 1, pp. 8–15, Jan. 2025, doi: 10.57218/jupeis.Vol4.Iss1.1326.
- [2]. V. Karenina, M. F. Erinsyah, and D. S. Wibowo, "Klasifikasi Rentang Usia Dan Gender Dengan Deteksi Suara Menggunakan Metode Deep Learning Algoritma Cnn (Convolutional Neural Network)," *Komputika : Jurnal Sistem Komputer*, vol. 12, no. 2, pp. 75–82, Sep. 2023, doi: 10.34010/komputika.v12i2.10516.
- [3]. D. T. Kusuma, "Fast Fourier Transform (FFT) Dalam Transformasi Sinyal Frekuensi Suara Sebagai Upaya Perolehan Average Energy (AE) Musik," *PETIR*, vol. 14, no. 1, pp. 28–35, Oct. 2020, doi: 10.33322/petir.v14i1.1022.
- [4]. R. A. N. Diaz, N. L. G. P. Suwirmayanti, and K. Budiarta, "Perbandingan Kualitas Pengenalan Suara Untuk Ekstraksi Fitur Menggunakan Mfcc Dan Spectral," *Naratif: Jurnal Nasional Riset, Aplikasi dan Teknik Informatika*, vol. 6, no. 1, pp. 58–63, Jul. 2024, doi: 10.53580/naratif.v6i1.281.
- [5]. F. Zalkow and M. Muller, "CTC-Based Learning of Chroma Features for Score-Audio Music Retrieval," *IEEE/ACM Trans Audio Speech Lang Process*, vol. 29, pp. 2957–2971, 2021, doi: 10.1109/TASLP.2021.3110137.
- [6]. P. Shetty, S. Singraddi, M. K. M. A. V. A. S., and B. K. P., "Speech Signal Segmentation using Zero Crossing Rate and Short-Time Energy for Speech Synthesis," in *2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS)*, IEEE, Aug. 2024, pp. 1–5. doi: 10.1109/IACIS61494.2024.10721744.
- [7]. D. Hosseinzadeh and S. Krishnan, "On the Use of Complementary Spectral Features for Speaker Recognition," Nov. 02, 2022. doi: 10.32920/21428718.v1.
- [8]. V. K. Kumar, B. Maram, and S. Chandre, "Optimal Feature Selection for Dimensionality Reduction with Preserved Predictive Performance in High-Dimensional Data," in *2025 International Conference on Computing Technologies & Data Communication (ICCTDC)*, IEEE, Jul. 2025, pp. 1–6. doi: 10.1109/ICCTDC64446.2025.11158774.
- [9]. B. A. Putra *et al.*, "Optimisasi Algoritma K-Means Dengan Metode Reduksi Dimensi Untuk Pengelompokan Big Data Dalam Arsitektur Cloud Computing," *Journal of Software Engineering and Information Systems*, vol. 5, no. 1, pp. 1–8, Dec. 2021, doi: 10.37859/seis.v5i1.7616.
- [10]. G. N. C. Nugroho, "Klasifikasi Gender Berdasarkan Sidik Jari Menggunakan Principal Component Analysis dan Support Vector Machine," *Journal of Information Engineering and Educational Technology*, vol. 8, no. 1, pp. 45–53, Jun. 2024, doi: 10.26740/jieet.v8n1.p45-53.
- [11]. [11] W. Aprilliandhika and F. F. Abdulloh, "Comparison Of K-Nearest Neighbor And Support Vector Machine Algorithm Optimization With Grid Search Cv On Stroke Prediction," *Jurnal Teknik Informatika (Jutif)*, vol. 5, no. 4, pp. 991–1000, Jul. 2024, doi: 10.52436/1.jutif.2024.5.4.1951.
- [12]. L. N. Valli and N. Sujatha, "Predictive Modeling and Decision-Making in Data Science: A Comparative Study," in *2024 5th International Conference on Recent Trends in Computer Science and Technology (ICRTCST)*, IEEE, Apr. 2024, pp. 603–608. doi: 10.1109/ICRTCST61793.2024.10578364.
- [13]. D. G. K. Vardhni, D. R., and A. Mahadevan, "A Comparative and Analytical Study of Text Classification Models using Various Metrics and Visualizations," in *2023 OITS International Conference on Information Technology (OCIT)*, IEEE, Dec. 2023, pp. 1–6. doi: 10.1109/OCIT59427.2023.10430972.
- [14]. F. Abakarim and A. Abenaou, "Voice Gender Recognition Using Acoustic Features, MFCCs and SVM," 2022, pp. 634–648. doi: 10.1007/978-3-031-10522-7\_43.
- [15]. H. Patel, R. Raman, M. Jawarneh, A. S. Ansari, H. Pallathadka, and D. T. Sanchez, "Machine Learning for Automatic Speech Recognition," in *Conversational Artificial Intelligence*, Wiley, 2024, pp. 147–168. doi: 10.1002/9781394200801.ch10.
- [16]. K. Akgün and Ş. A. Sadık, "Unified voice analysis: speaker recognition, age group and gender estimation using spectral features and machine learning classifiers," *Journal of Scientific Reports-A*, no. 057, pp. 12–26, Jun. 2024, doi: 10.59313/jsr-a.1422792.