Klasifikasi Mood pada Musik Pop dan Jazz dengan Menggunakan Mel Frequency Cepstral Coefficients dan K-Nearest Neighbor

I Gusti Bagus Putrawan[a1], I Ketut Gede Suhartana[a2]

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia
[1]putrawan.2208561133@student.unud.ac.id
[2]ikg.suhartana@unud.ac.id

Abstract

This research discusses mood classification in pop and jazz music using Mel Frequency Cepstral Coefficients (MFCC) and the K-Nearest Neighbor (KNN) algorithm. The dataset used consists of 900 songs with mood labels angry, happy, relaxed, and sad obtained from Kaggle. The data was processed by extracting 13 MFCC features and then continuing with classification using KNN. The research results show that the best accuracy reaches 64% with K=9. Accuracy at K=7 obtained a value of 60%, while at K=11 an accuracy of 58% was obtained. Evaluation was carried out using accuracy, precision, recall and f1-score metrics, with the best results found at K=9. This research emphasizes the importance of selecting K parameters for optimizing mood classification models.

Keywords: *Mood Clasification, MFCC, K-Nearest Neighor, Music Emotion Recognition*

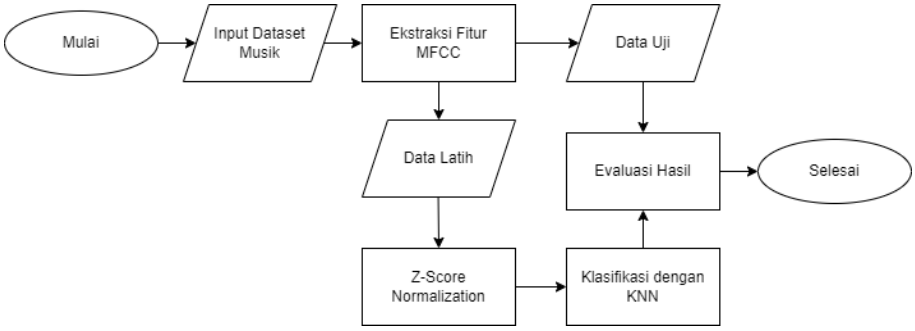
1. Pendahuluan

Musik merupakan seni yang menggabungkan irama, melodi, dan harmoni yang tercipta melalui kombinasi nada atau suara. Saat ini, musik telah menjadi simbol ekspresi emosional manusia. Misalnya, ketika seseorang merasa sedih, mereka cenderung mendengarkan musik yang melankolis, sedangkan saat bahagia, mereka memilih musik yang sesuai dengan suasana hati mereka. Musik telah menjadi bagian integral dari kehidupan manusia yang tidak dapat dipisahkan, mencerminkan dan mempengaruhi perasaan dan suasana hati dalam berbagai situasi. Penelitian terdahulu yang melakukan klasifikasi mood pada musik oleh Maulana memiliki dataset berjumlah 200 dari masing-masing 50 dataset mood, yaitu contentment, depression, exuberance, dan, anxiety. Pada penelitian ini dijelaskan bahwa menggunakan mfcc diperoleh akurasi sebesar 87,67 % dengan pemodelan Backpropagation Neural Network (BPNN) menggunakan 1 hidden layer 256 neuron, 0,4 dropout, 1000 epoch, learning rate 0.001, frame 40ms, dan overlap 40% dengan durasi 60 detik [1]. Pada penelitian yang berjudul Sistem Identifikasi Arti Tangisan Bayi Menggunakan Metode MFCC, DWT, dan KNN pada Raspberry Pi yang dilakukan oleh Prasetiyo, dengan mengklasifikasikan suara bayi ke dalam 5 kelas, yaitu pada saat kondisi bayi merasa tidak nyaman, lapar, masuk angin, sendawa, dan mengantuk. Jumlah keseluruhan yang digunakan dalam penelitian ini adalah sebanyak 74 data dengan 4 data bukan tangisan bayi dengan perbandingan data latih sebesar 50, data uji sebesar 20. Diperoleh akurasi terbaik 90% dengan menggunakan frame sebanyak 512 data per frame, 39 fitur MFCC, dan klasifikasi menggunakan KNN dengan nilai K=5 [2].

2. Metode Penelitian

Adapun tahapan-tahapan yang dilakukan dalam penelitian adalah sebagai berikut.

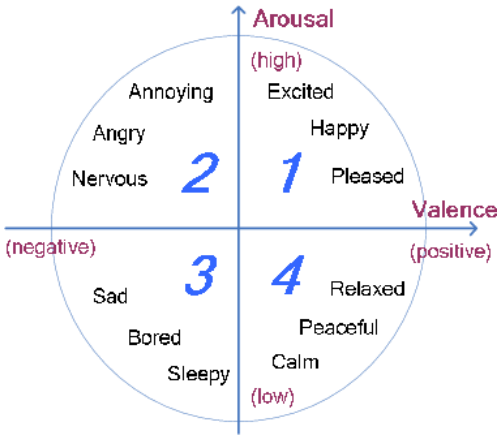
p-ISSN: 2986-3929
e-ISSN: 3032-1948



Gambar 1. Flowchart klasifikasi Musik Berdasarkan Mood

2.1 Music Emotion Recognition (MER)

Dalam musik, suasana hati adalah keadaan emosi yang bertahan lama. Suasana hati berbeda dengan emosi sederhana karena suasana hati kurang spesifik, kurang intens, dan kecil kemungkinannya dipicu oleh rangsangan atau peristiwa tertentu. Mendengarkan sebuah lagu dapat memberikan suasana hati bagi pendengarnya tergantung dari musik yang didengarkannya. MER atau Music Emotion Recognition merupakan salah satu cabang pengetahuan dari Music Information Retrieval yang mempelajari tentang pengenalan proses emosi atau mood pada musik. Terdapat banyak perumusan yang dilakukan oleh ahli psikologi di zamannya. Oleh karena itu, dalam penelitian ini akan digunakan pengklasifikasian mood pada music yang terbagi ke dalam 4 kuadran, yaitu happy (kuadran I), angry (kuadran II), sad (kuadran III), dan relax (kuadranIV) berdasarkan definisi dari Robert E. Thayer dalam artikelnya yang berjudul The Biopsychologyof Mood and Arousal tahun 1989 [3].



Gambar 2. Klasifikasi Mood Berdasarkan Definisi Robert E. Thayer [3]

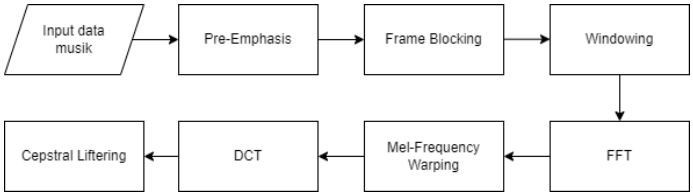
2.2 Dataset

Dalam penelitian, digunakan dataset musik berjumlah 900 data yang sudah diberi label angry, happy, relax, dan sad dengan format .mp3 yang diperoleh dari kaggle yang berjudul 4Q audio emotion dataset (Russell's model) [4]. Dalam penelitian, dataset musik yang digunakan akan berfokus pada dataset lagu pop dan jazz barat dengan jumlah data sebesar 900. Dataset yang digunakan sebagai data latihan dan data uji masing-masing adalah 85:15.

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

2.3 Mel Frequency Cepstral Coefficient (MFCC)

Data audio yang sudah dilakukan preprocessing dilanjutkan dengan melakukan ekstraksi fitur MFCC. Secara umum, proses ekstraksi fitur MFCC dibagi ke dalam beberapa tahapan, antara lain:



Gambar 3. Tahapan-Tahapan Ekstraksi Fitur MFCC

2.3.1 Pre-Emphasis

Pre-emphasis adalah proses untuk menerima dan mempertahankan amplitudo berfrekuensi tinggi agar meningkatkan kualitas dari sinyal audio. Pre-emphasis memiliki beberapa fungsi, diantaranya adalah untuk mengurangi noise pada sinyal, meningkatkan resolusi frekuensi sinyal, dan menyeimbangkan spektrum dari sinyal audio [5]. Pre-emphasis dirumuskan dalam persamaan berikut.

$$y(n) = s(n) - \alpha s(n-1) \tag{1}$$

Keterangan:

y(n) = sinyal setelah dilakukan pre-emphasiss(n) = sinyal awal
α = konstanta pre-emphasis dimana biasanya digunakan pada rentang 0,95 < α < 0,99

2.3.2 Frame Blocking

Pada tahapan ini disebut tahapan segmentasi, yaitu sinyal audio akan dibagi menjadi beberapa frame yang lebih kecil dimana masing-masing frame mewakili sejumlah sampel audio. Sifat dari frame ini saling tumpang tindih atau overlap untuk menghindari kehilangan informasi yang terdapat pada ujung frame. Panjang ukuran frame yang digunakan adalah 25ms dengan panjang tiap frame bernilai sama.

2.3.3 Windowing

Proses windowing digunakan untuk mengurangi terjadinya efek aliasing setelah sinyal diproses pada proses frame blocking. Aliasing adalah sinyal baru dengan frekuensi yang berbeda dari frekuensi sinyal asli yang dapat menyebabkan sinyal terputus [6]. Pada tahap ini, sinyal dari proses frame blocking dikalikan dengan nilai Hamming windowing. Nilai dari Hamming windowing dituliskan dalam persamaan berikut dengan N merupakan jumlah sampel setiap frame.

$$w_n = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right), \quad 0 \le n \le N-1 \tag{2}$$

Pada persamaan windowing dituliskan sebagai berikut dengan x_n adalah frame ke-n.

$$y_n = x_n \cdot w_n \tag{3}$$

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

2.3.4 Fast Fourier Transform

FFT atau Fast Fourier Transform digunakan untuk mengubah domain waktu menjadi domain frekuensi pada frekuensi setiap frame [7]. Dalam tahapan ini, dihasilkan spektrum atau spektral yang merepresentasikan amplitudo dari berbagai frekuensi di dalam sinyal audio. FFT dirumuskan dalam persamaan berikut

$$y_k = \sum_{n=0}^{N-1} y_n e^{-\frac{2\pi i k n}{N}}, \quad n = 0, 1, 2, \ldots, N-1 \tag{4}$$

2.3.5 Mel Frequency Warping

Sinyal yang sudah melalui proses FFT selanjutnya akan melewati proses Mel Frequency Warping dengan menggunakan filterbank. Filterbank adalah bentuk filter yang dirancang untuk menentukan ukuran energi dari suatu frekuensi tertentu dalam domain frekuensi. Dalam proses ini, sinyal hasil FFT dikelompokkan menggunakan filter triangular filter file. Setiap nilai FFT kemudian dikalikan dengan nilai filter yang sesuai dan hasil perkalian dijumlahkan. Tujuan dari proses tersebut adalah untuk mendapatkan representasi energi dari sinyal suara dalam skala mel, yang lebih sesuai dengan persepsi pendengaran manusia [7]. Frekuensi suara diukur menggunakan skala mel, yang bersifat linier untuk frekuensi di bawah 1000 Hz dan logaritmik untuk frekuensi di atas 1000 Hz.

2.3.6 Discrete Cosine Transform

DCT (Discrete Cosine Transform) adalah langkah terakhir dari proses utama ekstraksi fitur MFCC. DCT digunakan untuk mengonversi spektrum magnitudo menjadi domain waktu agar dapat direpresentasikan lebih baik dan akan menghasilkan Mel-Frequency Coefficient Cepstrum dalam bentuk vektor [6]. Jumlah koefisien dari cepstral yang digunakan dalam penelitian umumnya berjumlah 13. DCT dirumuskan dalam bentuk persamaan berikut.

$$C_m = \sum_{k=1}^{N} \cos\left[n\left(k - \frac{1}{2}\right)\frac{\pi}{N}\right] \log E_k, \quad n = 1, 2, \ldots, L \tag{5}$$

Keterangan:

C_m = koefisien MFCC ke-k
E_k = output dari Mel Frequency Warping
N = jumlah Mel-Frequency Coefficient Cepstrum
L = jumlah Cepstral Coefficient

2.3.7 Cepstral Liftering

Cepstral Liftering diimplentasikan ketika diperoleh hasil dari DCT dalam bentuk fitur cepstral dengan menggunakan persamaan window sehingga memperoleh hasil model yang lebih baik. Persamaan windows dirumuskan sebagai berikut.

$$w_n = 1 + 2\sin\left(\frac{n\pi}{L}\right) \tag{6}$$

Keterangan:

W_n = Koefisien lifting untuk Cepstral Coefficient ke-nn = index ke-n Cepstral Coefficient
L = jumlah Cepstral Coefficient

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

2.4 Z-Score Normalization

Standard Scaling atau disebut juga dengan Z-Score Normalization adalah metode yang digunakan dalam mengatur data berdasarkan nilai rata-rata atau mean dan standar deviasi dari data. Dataset dinormalisasi dengan mengubah skala pada distribusi nilai rata-rata fitur menjadi 0dan untuk standar deviasi fitur menjadi 1 [8]. Persamaan dari Z-Score Normalization dirumuskansebagai berikut.

$$x_{\text{Scaled}} = \frac{x - \mu}{\sigma} \tag{7}$$

2.5 K-Nearest Neighor (KNN)

Algoritma KNN biasa digunakan dalam supervised learning dimana model mempelajari dari labelyang sudah diberikan sebelumnya. KNN sendiri bekerja dengan cara mengklasifikasikan data yang baru dikenali berdasarkan mayoritas dari nilai k-tetangga terdekat. KNN menyimpan semua data latih dan label diikuti dengan menentukan nilai K atau jumlah tetangga terdekat yang akan digunakan untuk mengklasifikasikan data baru. Pada setiap data baru yang muncul, KNN menghitung jarak antara data tersebut dengan setiap data dalam data latih menggunakan metode seperti Euclidean, Manhattan, dan Minkowski. Dilanjutkan dengan pemilihan K data latih dengan jarak terdekat sebagai tetangga terdekat [9]. KNN dengan euclidean distance function dirumuskan dalam persamaan berikut.

$$d(x_i, y_i) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{8}$$

2.6 Confusion Matrix

Pada pengevaluasian model dilakukan dengan menggunakan confusion matrix yang bertujuan melakukan pengelompokan berdasarkan empat kelas yang terdiri atas True Positive (TP), FalsePositif (FP), False Negative (FN), dan True Negative (TN). TP terjadi ketika model dapat memprediksi suatu peristiwa yang sesuai dengan kondisi yang sesungguhnya. FP menunjukkanprediksi yang benar, akan tetapi pada kondisi yang sebenarnya bernilai salah. TN menunjukkan prediksi bernilai salah tepat dengan keadaan sesungguhnya yang bernilai salah. FN memprediksi kejadian bernilai salah dimana peristiwa sesungguhnya bernilai benar. Berdasarkan output Confusion Matrix sehingga diperoleh akurasi [10]. Berdasarkan definisi tersebut, confusion matrixdapat dirumuskan sebagai berikut.

Tabel 1. Confusion Matrix

Predicted Label		
	Positive	Negative
Positive	True Positive (TP)	False Positive (FP)
Actual Label		
Negative	False Negative (TN)	True Negative (TN)

Akurasi adalah perbandingan antara akumulasi data TP dan TN dengan keseluruhan data sehingga dapat dirumuskan dalam persamaan berikut.

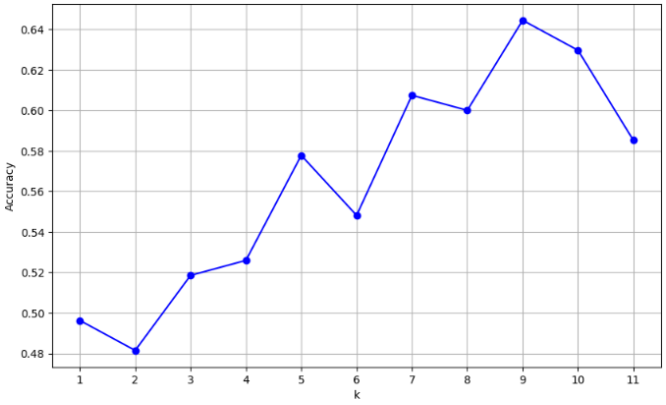
$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \tag{9}$$

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

3. Hasil dan Pembahasan

3.1 Klasifikasi dengan K-Nearest Neighbor dan MFCC

Dataset yang telah terkumpul terdiri dari 900 musik dengan genre pop yang berbeda. Dilanjutkan dengan melakukan ekstraksi fitur pada dataset dengan menggunakan MFCC sehingga tiap data musik berisi 13 fitur audio MFCC. Data musik tersebut akan dipisahkan menjadi data latih dan data uji dengan perbandingan sebesar 85:15. Proses klasifikasi data latih menggunakan K- Nearest Neighbor dan pengujian dengan metrik akurasi, presisi, recall, dan f1-score.



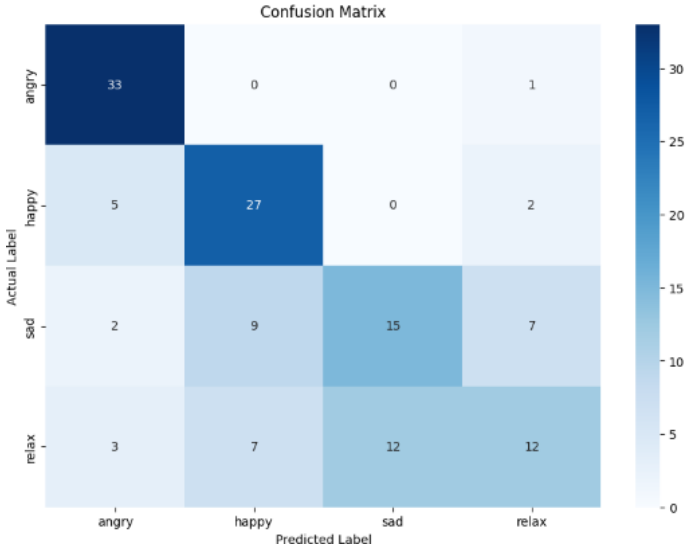
Gambar 4. Hasil Perbandingan Nilai K pada K-Nearest Neighbor

	precision	recall	f1-score	support
angry	0.77	0.97	0.86	34
happy	0.63	0.79	0.70	34
relax	0.55	0.35	0.43	34
sad	0.56	0.45	0.50	33
accuracy			0.64	135
macro avg	0.62	0.64	0.62	135
weighted avg	0.62	0.64	0.62	135

Gambar 5. Hasil Akurasi, Presisi, Recall, dan F1-Score dengan K=9

Berdasarkan Gambar 5, diperoleh hasil terbaik dari model K-Nearest Neighbor dengan parameter K = 9, dimana pemilihan K berpengaruh pada tingkat akurasi, akan tetapi ketika parameter K > 9, menghasilkan model yang kurang optimal dikarenakan akurasinya mulai menurun. Nilai akurasi dari data uji yang diprediksi benar oleh model sebesar 64%, yaitu sebanyak 87 data dan prediksi salah sebesar 36% dengan jumlah data 48 data. Mood marah diklasifikasikan benar dengan jumlah 33 data, mood happy dengan jumlah 27 data, mood sedih dengan jumlah 15 data,dan mood santai dengan jumlah 12 data.

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



Gambar 6. Evaluasi dengan Confusion Matrix

4. Kesimpulan

Klasifikasi mood pada lagu pop dan jazz dilakukan menggunakan fitur MFCC dan K-Nearest Neighbor. Evaluasi akurasi menggunakan metrik akurasi, presisi, recall, dan f1-score. Akurasi terbaik diperoleh dengan perbandingan data latih dengan data uji adalah 85:15. Klasifikasi menggunakan K-Nearest Neighbor dan 13 fitur MFCC memberikan akurasi sebesar 64% dengan parameter K = 9, akurasi 60% pada K = 7, dan 58% pada K = 11. Jumlah mood yang diprediksi benar sebanyak 87 data sedangkan jumlah yang diprediksi salah sebanyak 48 data.

Referensi

[1] P. I. Maulana, A. Aranta, F. Bimantoro, and I. G. Andika, “Klasifikasi Mood Musik Berdasarkan Mel Frequency Cepstral Coefficients Dengan Backpropagation Neural Network,” Jurnal Resistor (Rekayasa Sistem Komputer), vol. 5, no. 1, pp. 72–85, Apr. 2022, doi: 10.31598/jurnalresistor. v5i1.1089.

[2] Y. Yohannes and R. Wijaya, “Klasifikasi Makna Tangisan Bayi Menggunakan CNN Berdasarkan Kombinasi Fitur MFCC dan DWT,” JATISI (Jurnal Teknik Informatika dan Sistem Informasi), vol. 8, no. 2, pp. 599–610, Jun. 2021, doi: 10.35957/jatisi. v8i2.470.

[3] R. E. Thayer, The Biopsychology of Mood and Arousal. Oxford University PressNew York, NY, 1990. [Online]. Available: http://dx.doi.org/10.1093/oso/9780195068276.001.0001

[4] R. Panda, R. Malheiro, and R. P. Paiva, “Novel Audio Features for Music Emotion Recognition,” IEEE Transactions on Affective Computing, vol. 11, no. 4, pp. 614–626, Oct. 2020, doi: 10.1109/taffc.2018.2820691.

[5] S. Y. Yusdiantoro and T. B. Sasongko, “Implementasi Algoritma MFCC dan CNN dalam Klasifikasi Makna Tangisan Bayi,” Indonesian Journal of Computer Science, vol. 12, no. 4, Aug. 2023, doi: 10.33022/ijcs. v12i4.3243.

[6] S. P. Dewi, A. L. Prasasti, and B. Irawan, “Analysis of LFCC Feature Extraction in Baby Crying Classification using KNN,” in 2019 IEEE International Conference on Internet of Things and Intelligence System (IoTaIS), Nov. 2019. Accessed: May 24, 2024. [Online]. Available: http://dx.doi.org/10.1109/iotais47347.2019.8980389

[7] I. D. G. y A. Wibawa and I. D. M. B. A. Darmawan, “Implementation of audio recognition using mel frequency cepstrum coefficient and dynamic time warping in wirama praharsini,” Journal of Physics: Conference Series, vol. 1722, no. 1, p. 012014, Jan. 2021, doi: 10.1088/1742-6596/1722/1/012014.

[8] P. I. Maulana, A. Aranta, F. Bimantoro, and I. G. Andika, “Klasifikasi Mood Musik Berdasarkan Mel Frequency Cepstral Coefficients Dengan Backpropagation Neural

Network," Jurnal Resistor (Rekayasa Sistem Komputer), vol. 5, no. 1, pp. 72–85, Apr. 2022, doi: 10.31598/jurnalresistor. v5i1.1089.

[9] I. N. Y. T. Giri, L. A. A. Rahning Putri, G. A. V. Mastrika Giri, I. G. N. Anom Cahyadi Putra, I. M. Widiartha, and I. W. Supriana, "Music Genre Classification Using Modified K-Nearest Neighbor (MK-NN)," JELIKU (Jurnal Elektronik Ilmu Komputer Udayana), vol.10, no. 3, p. 261, Feb. 2022, doi: 10.24843/jlk. 2022.v10.i03.p02.

[10] J. Davis and M. Goadrich, "The relationship between Precision-Recall and ROC curves," in Proceedings of the 23rd international conference on Machine learning - ICML '06, 2006. [Online]. Available: http://dx.doi.org/10.1145/1143844.1143874