

KLASIFIKASI MOOD MUSIK BERDASARKAN MEL FREQUENCY CEPSTRAL COEFFICIENTS DENGAN BACKPROPAGATION NEURAL NETWORK

Patriaaji Ibrahim Maulana¹, Arik Aranta², Fitri Bimantoro³, I Gede Andika⁴

¹²³Program Studi Teknik Informatika, Fakultas Teknik, Universitas Mataram
Jl. Majapahit No.62, Gomong, Kec. Selaparang, Kota Mataram, Nusa Tenggara Barat, Indonesia

⁴Program Studi Sistem Komputer, Fakultas Teknologi dan Informatika,
Institut Bisnis dan Teknologi Indonesia
Jl. Tukad Pakerisan No.97, Panjer, Denpasar Selatan, Kota Denpasar, Bali, Indonesia

Email: patriajimaulana@gmail.com, [arikaranta, bimo]@unram.ac.id, gdandika@stiki-indonesia.ac.id

Received : April, 2022

Accepted : April, 2022

Published : April, 2022

Abstract

In processing audio signals, music can be grouped into several categories, including by categorizing genres, singers, and the mood of the music itself. The introduction or grouping of music through this point of view resulted in a new term called MIR (Music Information retrieval). with this algorithm can facilitate the process of searching for music based on certain categories such as mood for example. This is strongly supported by the algorithm that is currently widely used in audio signal processing, namely MFCC (Mel Frequency Cepstral Coefficients) where with this algorithm a music feature can be issued so that it is possible to carry out the recognition process based on certain categories. After a feature is generated, the next step is to carry out the feature classification process; without the feature classification method, it cannot be very useful in digital data processing. The classification method commonly used is BPNN (Backpropagation Neural Network) with the combination of these methods in this study succeeded in classifying music based on four types of mood based on the Thayer model with better machine learning model performance. The data used in this study were obtained from the internet with a total dataset of 200. The results obtained from this study were the classification of musical moods using BPNN based on the MFCC feature was able to produce 87.67% accuracy.

Keywords: Music, MIR, Mood, MFCC, BPNN.

Abstrak

Dalam pengolahan sinyal audio musik dapat dikelompokkan menjadi beberapa kategori diantaranya adalah melalui pengkategorian genre, penyanyi, dan mood dari musik itu sendiri. Pengenalan atau pengelompokan musik melalui sudut pandang tersebut menghasilkan istilah baru yang disebut dengan MIR (Music Information retrieval). dengan adanya algoritma ini dapat memudahkan dalam proses pencarian musik berdasarkan kategori tertentu seperti suasana hati contohnya. hal ini sangat didukung dengan algoritma yang saat ini banyak digunakan dalam pemrosesan sinyal audio yakni MFCC(Mel Frequency Cepstral Coefficients) dimana dengan algoritma ini suatu musik dapat dikeluarkan fiturnya sehingga memungkinkan untuk dilakukan proses pengenalan berdasarkan kategori tertentu. setelah suatu fitur dihasilkan langkah berikutnya adalah melakukan proses klasifikasi fitur, tanpa adanya metode klasifikasi fitur tersebut belum dapat begitu bermanfaat dalam proses pengolahan data digital. Metode klasifikasi yang umum digunakan adalah BPNN (Backpropagation Neural Network) dengan

penggabungan metode tersebut pada penelitian ini berhasil melakukan klasifikasi musik berdasarkan empat jenis mood berdasarkan model thayer dengan kinerja model machine learning yang lebih baik. Data yang digunakan dalam penelitian ini diperoleh dari internet dengan total dataset 200. Hasil yang diperoleh dari penelitian ini adalah klasifikasi mood musik menggunakan BPNN berdasarkan fitur MFCC mampu menghasilkan 87,67% ketepatan.

Kata Kunci: Music, MIR, Mood, MFCC, BPNN.

1. PENDAHULUAN

Musik adalah sekumpulan nada yang disusun sehingga menghasilkan suatu kesatuan komposisi suara yang memiliki irama dan keselarasan. Musik yang dimainkan dengan komposisi terpadu dapat memberikan pengaruh terhadap emosi dan mood. Orang-orang mendengarkan musik untuk berbagai tujuan, seperti memberi semangat, untuk relaksasi, atau untuk menghibur. Dengan adanya hal ini maka terbukti bahwa musik memiliki hubungan yang erat dengan psikologi manusia, hal ini dapat dilihat dari kecenderungan mendengarkan suatu musik seseorang dipengaruhi oleh suasana hati atau mood orang tersebut. Selain itu juga, musik dapat mempengaruhi suasana hati pendengarnya yang sedang merasa sedih, dengan mendengarkan musik bernada gembira, suasana hati pendengar ikut merasakan kegembiraan[1].

Suasana hati atau mood dan emosi memiliki konsep yang terkait namun memiliki perbedaan. Perbedaan ini adalah mood merupakan keadaan emosi manusia yang bertahan lama. Mood biasanya dipicu oleh beberapa rentetan pengalaman yang emosional yang berulang kali, berbeda dengan emosi yang mudah dipicu oleh lingkungan sekitar dan bersifat kurang spesifik[2]. Dalam sebuah penelitian yang dilakukan Robert Thayer, diajukan model dua dimensi yang mengangkat teori yang menyatakan dua faktor yang mempengaruhi emosi dan suasana hati seseorang atau mood, yaitu : tingkat dari stress (senang dan cemas) dan *energi* yang menyimbolkan (santai dan energetic) dari signal audio. Dari model tersebut dalam kembali sehingga menghasilkan menjadi empat kelompok, yaitu; kepuasan atau disebut *contentment*, gembira atau disebut *exuberance*, depresi atau *depression*, dan cemas atau disebut *anxious*; kalut[3]. Dalam dunia industri musik, pengelompokan jenis musik baik dalam mood dan genre merupakan hal yang penting dalam menghasilkan informasi kepada pendengar dalam memilih musik yang ingin didengarkan. Pengelompokan musik umumnya

dilakukan dengan cara manual yaitu dengan cara mendengarkan setiap musik secara langsung. Kelebihan dari cara ini yaitu keakuratannya lebih terjamin, namun dalam hal kecepatan hal ini membutuhkan banyak waktu, termasuk untuk mengolah data musik yang sangat besar seperti di platform musik streaming. Maka dari itu sangat penting dibutuhkan suatu metode dengan bantuan komputer yang melakukan pengelompokan dengan jumlah data musik yang besar. *Music Information Retrieval (MIR)* adalah bidang penelitian yang muncul untuk memenuhi kebutuhan tersebut.

MIR yaitu sebuah bidang kajian yang mengolah informasi yang dihasilkan dari file musik berupa metadata dari suatu signal audio ataupun kandungan yang dapat dihasilkan . Hal tersebut meliputi : *artist identification (identifikasi penutur)*, *genre classification (pengenalan genre)*, *emotion recognition (emosi dari suatu audio)*, *cover song recognition (pengolahan suara dari penyanyi asli)*, *instrument recognition, and music annotations*. Ekstraksi fitur dan pelatihan klasifikasi merupakan kunci utama dari MIR. Dengan pertumbuhan internet, sistem aplikasi layanan multimedia yang berhubungan dengan basis data musik semakin penting dan permintaan untuk aplikasi MIR meningkat[4].

Dalam mendapatkan sinyal fitur suara dari musik, *Mel Frequency Cepstral Coefficients (MFCC)* adalah metode yang saat ini banyak digunakan baik dalam ilmu pemrosesan suara baik dalam bidang MIR maupun *Voice recognition*[5][6][7]. MFCC menggunakan persepsi sensitivitas pendengaran manusia sebagai bahan pertimbangan dan karena itu menjadi metode terbaik untuk *speaker/speech recognition*[4].

Neural Network adalah sebuah metode dalam *Machine Learning* yang umum digunakan untuk memproses klasifikasi data.. Desain pemrosesan data berupa tiruan dari bagaimana otak manusia bekerja dengan sel-sel syaraf tiruan (*neuron*)

yang digunakan untuk menentukan sebuah keputusan dibangun berdasarkan prinsip organisasi otak manusia. *Backpropagation Neural Network* (BPNN) dengan mempertimbangkan tingkat error BPNN dapat menyesuaikan bobot berdasarkan luaran yang dihasilkan. Dikatakan *multilayer* karena arsitekturnya memiliki 3 *layer* dalam proses pelatihannya, berbeda dengan *Neural Network* biasanya yang hanya memiliki 2 yaitu input dan *output*, BPNN memiliki 3 *layer* yaitu, adlah layer masukan atau *input*, *layer hidden* atau layer tersembunyi dan layer output atau luaran BPNN telah digunakan pada bidang MIR salah satunya yaitu pada klasifikasi *genre* lagu[6].

Berdasarkan uraian di atas, penulis mengajukan untuk melakukan penelitian untuk membuat model untuk melakukan klasifikasi mood musik berdasarkan MFCC dengan BPNN. Penelitian dilakukan menggunakan 200 data dengan mood *contentment*, *exuberance*, *depression*, dan *anxious* masing-masing memiliki 50 data yang dimana seluruh data bersumber dari internet. Penelitian ini bertujuan menguji penggabungan metode MFCC dan BPNN dalam performa untuk dapat digunakan untuk melakukan pengelompokan menjadi empat buah mood musik atau suasana hati model Thayer. Model mood Thayer dapat dilihat pada Gambar 1.



Gambar 1. Model mood Thayer [3]

2. TINJAUAN PUSTAKA

Pada penelitian sejenis, yaitu klasifikasi mood musik, pernah dilakukan sebelumnya dengan hasil akurasi sebesar 88.65%. Penelitian ini dilakukan dengan cara membagi mood menjadi 4 dengan model Thayer dimana setiap mood music memiliki 50 dataset dengan setiap data durasi sepanjang 45 detik. MFCC digunakan sebagai fitur pada penelitian ini dengan menggabungkan fitur Spektoral dan Temporal dengan total fitur yang cukup banyak yaitu sebanyak 14 fitur yang digunakan. Pada penelitian ini digunakan ANN sebagai classifier[7].

Pada penelitian lainnya, percobaan serupa pernah dilakukan dengan merubah sinyal suara menjadi bentuk dengan metode *Fast Fourier Transform* yang umum digunakan mengolah data Spectral Analysis untuk menghasilkan 9 jenis *Spectral* Fitur dimana hasilnya kemudian dilakukan perhitungan dengan *classifier K-Means*. yang dimana hasilnya dapat

menghasilkan rata-rata *silhouette coefficient* yang menghasilkan 40 – 400 data uji menunjukkan nilai 0,6268807 yang tergolong struktur menengah yang tergolong baik [8].

Pada bidang MIR lainnya, penelitian klasifikasi genre musik pernah dilakukan menggunakan MFCC dengan SVM dan BPNN sebagai classifier dan mendapatkan hasil akurasi sebesar secara berurutan 83% dan 95%[6]. Pada penelitian sejenis lainnya didapatkan hasil akurasi sebesar 97.8% dengan klasifikasi *Deep Neural Network*. Penelitian tersebut memerlukan perhitungan yang cukup banyak yaitu menggunakan 350 hidden layer[4]. K-NN juga pernah digunakan untuk penelitian sejenis, dengan menggunakan 250 data dengan 5 genre yang berbeda dan mendapatkan hasil akurasi sebesar 52,4%[9].

Penggunaan fitur lain dalam bidang MIR selain MFCC juga pernah dilakukan diantaranya dengan menggunakan beberapa fitur yaitu *Short Time energy* yang umum digunakan untuk

mengakuisisi energi berdurasi pendek, *Zero Crossing Rate* yang digunakan untuk melihat nilai selisih suatu sampel sinyal, *Spectral Centroid* yang digunakan untuk menemukan pusat spektrum, dan *Spectral Flux* yang merupakan indikator kecepatan suatu *spectral* berubah dengan *classifier SVM*. Penelitian ini menggunakan total 113 data musik dari setiap genre. Penelitian ini menghasilkan akurasi sebesar 58.33% [10].

Klasifikasi jenis musik Sufi dan Klasik pernah dilakukan dengan cara menggunakan 8 fitur diantaranya, *Beats, Periodicity, Loudness, energy, Speechiness, Acousticness, Valence, Danceability, dan Discrete Wavelet Transform (DTW)*. Masing- masing jenis musik memiliki 200 dataset dengan pembagian 60% digunakan untuk training set dan 40% untuk testing yang dilakukan dengan *Neural Network*. Penelitian ini mendapatkan hasil akurasi sebesar 85% [11].

Pada bidang pengenalan suara, MFCC dan DTW digunakan dalam sistem pembuka pintu otomatis dimana keduanya digunakan untuk mendapatkan fitur suara dari pengguna. Pada penelitian tersebut didapatkan akurasi sebesar 81,25% [5]. Pada penelitian sejenis lain yaitu pengenalan suara sebagai pengontrol lampu otomatis menggunakan MFCC sebagai metode ekstraksi fitur dan dengan metode *Adaptive Neuro-Fuzzy Inference System* atau disebut dengan (Anfis) mendapatkan hasil untuk suatu data yang tidak terlatih 98,57 % dan untuk data tidak terlatih menghasilkan nilai 95,90% [12].

Berdasarkan tinjauan pustaka tersebut diatas penulis mengajukan penelitian dengan menggunakan MFCC dan klasifikasi BPNN. Adapun pertimbangan menggunakan MFCC karena MFCC yang menggunakan persepsi sensitivitas pendengaran manusia sudah terbukti memiliki hasil yang tinggi dari penelitian-penelitian sebelumnya. Begitu juga dengan BPNN yang memiliki hasil yang tinggi.

2.1 Musik

Musik Adalah sebuah nada atau bunyi bunyian yang mengandung suatu nilai yang biasa dikaitkan dengan nilai sejarah atau ciri khas suatu daerah. yang dimana perbedaan penafsiran terkait musik juga cukup beragam (1) musik adalah suatu unsur bunyi yang memiliki beberapa unsur(2)musik yang berurutan dan

memiliki nilai seni sehingga dapat dinikmati (3)dan pada umumnya musik adalah suatu bunyi yang dihasilkan secara sengaja oleh seseorang atau kelompok . yang dimana pada umumnya suatu musik tersusun atas melodi ritme dan timbre Ketiga bagian penting tersebut, masing-masing akan memberikan pengaruh yang berbeda terhadap kehidupan manusia. Seperti bit yang akan memiliki pengaruh pada tubuh, ritme dan melodi yang akan memberikan pengaruh pada jiwa[13].

2.2 Suasana Hati (Mood)

Pada musik Suasana hati atau mood adalah keadaan emosi yang relatif tahan lama. Mood berbeda dari emosi yang sederhana dimana emosi lebih bersifat kurang spesifik, kurang intens, dan kurang mungkin dipicu oleh stimulus atau peristiwa tertentu. Mendengarkan sebuah lagu dapat memicu timbulnya perasaan suasana hati pendengar sesuai dengan musik yang didengarkan[14].

Tepatnya Tahun 1989 Robert Thayer mengenalkan sebuah model pengenalan mood yang ditampilkan dalam bentuk dua dimensi, model tersebut dapat merepresentasikan suasana hati atau mood. Model ini berbeda dengan model Hevner. Model Thayer menyatakan model ini bergantung pada dua hal utama yaitu : stress yang didalam stress ini terdiri dari (unsur happiness/ unsur kebahasaan dan unsur anxiety/ atau yang sering disebut dengan kecemasan, energi yang didalamnya mengandung (unsur calm/ atau disebut dengan unsur tenang dan dan unsur energy/ atau yang sering disebut dengan suatu mood yang berenergi) yang mana dari kedua model tersebut dijabarkan kembali menjadi: contentment/ atau disebut dengan kepuasan, yang merepresentasikan musik yang memiliki unsur mood tenang dan dapat dikatakan bahagia; yang berikutnya adalah depression/ atau sering disebut dengan unsur depresi, dimana mood ini merepresentasikan mood musik yang cemas dan dapat dikatakan depresi; yang berikutnya adalah exuberance/ atau dapat dikatakan unsur kegembiraan atau lebih cenderung bahagia dan dapat dikatakan energik; dan yang terakhir adalah anxiety/ atau sebuah unsur yang mengandung kecemasan atau menggambarkan suasana musik yang cemas, dan atau dapat dikatakan kalut[3].[3].

Model mood Thayer dapat dilihat pada Gambar 1.

2.3 Music Information Retrieval (MIR)

Music Information Retrieval (MIR) merupakan sebuah bidang ilmu untuk mengambil dan mengolah informasi dari file musik yang berupa metadata ataupun konten. Penelitian yang telah dilakukan oleh Zhouyu Fu dkk [15], (MIR) Music Information Retrieval adalah sebuah ilmu yang mempelajari metadata dari suatu musik yang dilakukan Zhouyu Fu dkk. dimana pada umumnya penelitian ini melakukan mapping penelitian pada bidang MIR dimana hal ini mengacu pada (MIREX) The Music Information Retrieval Evaluation eXchange, pengenalan Genre, pengenalan Mood, pengenalan Artis, pengenalan pada Instrumen musik, dan anotasi pada Music. menurut Chai MIR terbagi menjadi tiga bidang utama, yaitu pencarian musik dan *query* musik berdasarkan contoh, klasifikasi Music, dan segmentasi Music dan musik summarisation [16].

Dalam penelitian dengan konsentrasi pencarian musik dan contoh *query*, yang mana proses ini MIR berperan dalam bidang melakukan pencarian musik berdasarkan kandungan dari suatu musik. pencarian musik adalah salah satu topik utama MIR. klasifikasi ini digunakan untuk melakukan pencarian berdasarkan genre, atau berdasarkan penyanyi, ataupun pencarian musik berdasarkan emosi/mood. sedangkan Music Summarization untuk menemukan bagian inti dari suatu musik, hal ini disebut intro atau refrain dari suatu musik.

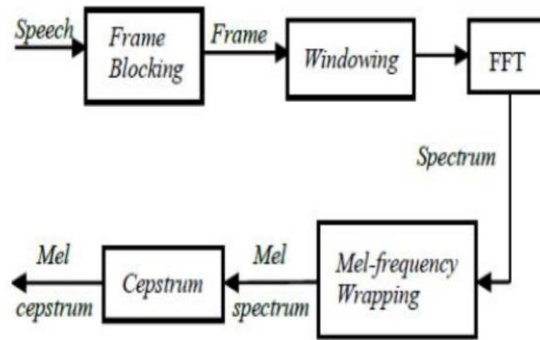
2.4 Mel Frequency Cepstral Coefficients (MFCC)

Mel Frequency Cepstral Coefficients (MFCC) awalnya dikembangkan untuk sistem pengenalan suara, namun belakangan ini MFCC banyak digunakan pada bidang MIR. MFCC menggunakan persepsi sensitivitas pendengaran sehubungan dengan frekuensi dan maka dari itu MFCC adalah metode terbaik untuk *speaker/speech recognition* [4]. Fitur MFCC berawal pada ketidaksesuaian yang diakui dari *bandwidth* kritis telinga manusia. Telinga manusia dapat lebih baik untuk membedakan frekuensi suara rendah daripada frekuensi tinggi. Contohnya, manusia dapat dengan mudah membedakan antara 500 dan 1000 Hz, namun akan lebih sulit untuk membedakan antara 10.000 dan 10.500 Hz walaupun kedua pasangan tersebut memiliki jarak yang sama. Maka dari itu MFCC

menggunakan filter frekuensi yang ditempatkan secara linier pada frekuensi rendah dan secara logaritmik pada frekuensi tinggi digunakan untuk mempertahankan properti penting secara fonetik dari sinyal ucapan. Sinyal ucapan biasanya berisi nada dengan frekuensi yang bervariasi, setiap nada dengan frekuensi aktual, f (Hz) dan nada subjektif dihitung pada skala Mel. Skala frekuensi mel memiliki jarak frekuensi linier di bawah 1000 Hz dan jarak *logaritmik* di atas 1000 Hz. Pitch nada 1 kHz dan 40 dB di atas ambang batas suara yang dipersepsi didefinisikan sebagai 1000 *mels*, dan digunakan sebagai titik referensi [17]. Blok diagram dan proses MFCC dapat dilihat pada Gambar 2.

2.4 Mel Frequency Cepstral Coefficients (MFCC)

Mel Frequency Cepstral Coefficients (MFCC) awalnya dikembangkan untuk sistem pengenalan suara, namun belakangan ini MFCC banyak digunakan pada bidang MIR. MFCC menggunakan persepsi sensitivitas pendengaran sehubungan dengan frekuensi dan maka dari itu MFCC adalah metode terbaik untuk *speaker/speech recognition* [4]. Fitur MFCC berawal pada ketidaksesuaian yang diakui dari *bandwidth* kritis telinga manusia. Telinga manusia dapat lebih baik untuk membedakan frekuensi suara rendah daripada frekuensi tinggi. Contohnya, manusia dapat dengan mudah membedakan antara 500 dan 1000 Hz, namun akan lebih sulit untuk membedakan antara 10.000 dan 10.500 Hz walaupun kedua pasangan tersebut memiliki jarak yang sama. Maka dari itu MFCC menggunakan filter frekuensi yang ditempatkan secara linier pada frekuensi rendah dan secara logaritmik pada frekuensi tinggi digunakan untuk mempertahankan properti penting secara fonetik dari sinyal ucapan. Sinyal ucapan biasanya berisi nada dengan frekuensi yang bervariasi, setiap nada dengan frekuensi aktual, f (Hz) dan nada subjektif dihitung pada skala Mel. Skala frekuensi mel memiliki jarak frekuensi linier di bawah 1000 Hz dan jarak *logaritmik* di atas 1000 Hz. Pitch nada 1 kHz dan 40 dB di atas ambang batas suara yang dipersepsi didefinisikan sebagai 1000 *mels*, dan digunakan sebagai titik referensi [17]. Blok diagram dan proses MFCC dapat dilihat pada Gambar 2.



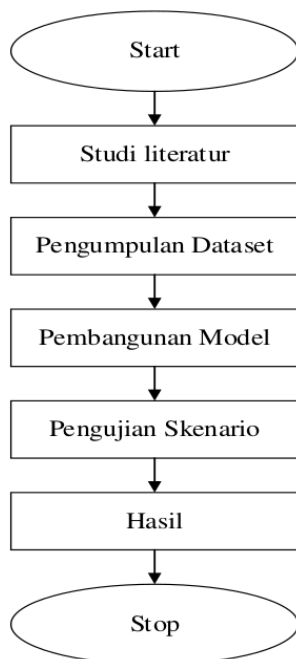
Gambar 2. Blok diagram proses MFCC [17]

3. METODE PENELITIAN

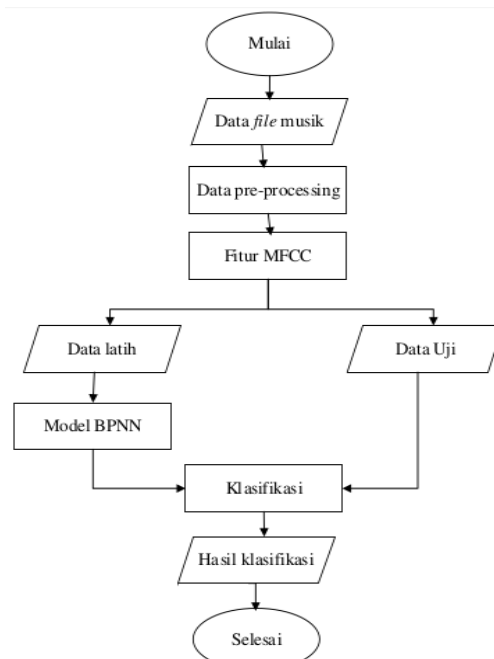
3.1 Diagram Alir penelitian

Terdapat beberapa tahapan dalam pelaksanaan penelitian ini. Pertama dilakukan studi literatur untuk memperoleh wawasan atau dasar dalam melakukan penelitian. Studi literatur dilakukan dengan membaca jurnal, buku, dan penelitian-penelitian terkait pemrosesan suara dan MIR. Selanjutnya dilakukan pengumpulan dataset, dimana dataset tersebut akan digunakan dalam pembuatan model, baik sebagai data latih

maupun data uji. Setelah dataset berhasil dikumpulkan, dilakukan pembuatan model *Machine Learning* BPNN untuk pengenalan sinyal suara musik. Setelah model selesai dibuat dilanjutkan dengan melakukan skenario uji antara lain, variasi panjang durasi musik secara keseluruhan, variasi panjang frame serta overlap pada proses FFT, dan variasi jumlah hidden layer pada proses klasifikasi. Setelah pengujian skenario selesai, maka dilakukan analisis terhadap hasil pengujian skenario yang selanjutnya dapat diambil suatu kesimpulan skenario model manakah yang memiliki akurasi terbaik pengujian. Diagram alir penelitian dapat dilihat pada Gambar 3.



Gambar 3. Diagram Alir Penelitian



Gambar 4. Diagram alir perancangan model

3.2 Pengumpulan Dataset

Dataset yang digunakan dalam penelitian ini adalah file musik instrumental dengan 4 model kategori *mood Thayer* yaitu *contentment, depression, exuberance, dan, anxiety*. Setiap kategori mood memiliki 50 data menjadikan total keseluruhan dataset berjumlah 200 dataset. Dataset didapatkan dari berbagai sumber di internet.

3.3 Pembangunan Model

Dalam pembangunan model Machine Learning menggunakan fitur MFCC dengan klasifikasi BPNN, terdapat berbagai tahapan yang dilakukan, yaitu, ekstraksi ciri, pembagian data, dan pembuatan model BPNN. Diagram alir pembangunan model dapat dilihat pada Gambar 4.

3.3.1 Data Preprocessing

Setelah seluruh file musik diimpor dilakukan *Preprocessing* pada file musik, dimana dalam Langkah ini seluruh data file musik yang awalnya dalam format .wav diubah ke dalam bentuk sinyal amplitudo dengan kemudian dilakukan *trimming*. Proses *trimming* ini memotong bagian awal dan akhir dari musik yang biasanya terdapat kekosongan sinyal amplitudo.

3.3.2 Ekstraksi Fitur MFCC

1. Frame Blocking pada file audio

Tahap ini melakukan segmentasi signal pada sebuah frame dengan durasi 20 ms sampai dengan 40ms. Frame berpengaruh besar dalam keberhasilan dalam analisa pada sinyal audio. disatu sisi panjang frame yang memiliki ukuran lebih panjang akan dapat memberikan hasil lebih baik dalam menunjukkan resolusi waktu yang baik. namun perlu diperhatikan panjangnya frame akan berpengaruh pada efisiensi waktu yang digunakan. Dalam pengolahannya sinyal suara dibagi menjadi beberapa bagian dari frame N sampel, yang selanjutnya frame yang memiliki jarak berdekatan akan dipisahkan oleh M ($M < N$) [6]. N menggambarkan panjangnya frame dari sinyal dan M merupakan titik dimulainya untuk frame berikutnya. sehingga menghasilkan kumpulan frame yang overlap. Ukuran panjang dari daerah overlap dari suatu sinyal adalah 30-50 yang berasal dari panjang dari sebuah frame. Overlapping digunakan untuk mempertahankan ciri dari

suatu sampel suara yang di proses dalam bidang pemrosesan audio dalam bidang MIR [18].

2. Windowing pada audio file

Windowing adalah sebuah proses yang berguna untuk mengatasi yang disebut kebocoran spektral dan aliasing pada suatu sinyal dan berperan untuk menahan agar tidak terjadinya pembentukan sinyal baru dari suatu audio yang diproses dengan cara mengenkalkan nilai awal dan nilai akhir dari suatu frame[18]. Metode *windowing* memiliki beberapa jenis, yang paling sering kita temui di sumber internet pada umumnya adalah *Hamming window*. semanagimana persamaan (1)

$$w(n) = 0.54 - 0.46 \cos \left[\frac{2\pi n}{N-1} \right], 0 \leq n \leq N \quad (1)$$

Metode Hamming window bekerja menghubungkan setiap blok hamming yang berada pada rantai pengolahan fitur signal, sehingga setiap blok sinyal dapat terintegrasi.

3. FFT

Fast Fourier Transform (FFT) adalah suatu metode yang umum digunakan pada proses pengolahan sinyal digital berjenis audio, peran dari metode ini adalah merubah file yang mulanya berdomain waktu diubah menjadi berdomain frekuensi dimana persamaan yang umum digunakan pada dalam metode FFT adalah SK adalah sampel signal, N adalah jumlah sampel pada frame, $k=0,1,2,\dots,(N-1)$, sedangkan j adlah imaginer ($v-1$), dan $n=1,2,3,\dots,(N-1)$ [18] sebagaimana persamaan (2).

$$x_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi jkn}{N}}, k = 0,1,2, \dots, N-1 \quad (2)$$

4. Mel – Frequency Wrapping

Mel – *Frequency Wrapping* Persepsi sistem pendengaran manusia terhadap frekuensi Tidak dapat diukurnya bentuk persepsi pendengaran manusia dalam bentuk linier pada setiap nada yang memiliki frekuensi aktual, yang di simbolkan dengan nilai f, serta dilakukan pengukuran dalam satuan Hz, yang mana sebuah penilaian subjektif pitch diukur dalam sebuah skala yang disebut dengan “Mel”. Skala mel- frequency adalah sebuah frekuensi rendah yang bersifat linier 1000 Hz atau sebuah sinyal logaritmik yang memiliki ukuran di atas 1000 Hz [18]. sebagaimana persamaan (3) dan persamaan (4)

$$mel(f) = 1125 \times \ln \left(1 + \frac{f}{700} \right) \quad (3)$$

$$mel^{-1}(mel) = 700(\exp(\frac{mel}{1125}) - 1) \quad (4)$$

Perumusan berikut untuk melakukan konversi frekuensi biasa (linear) ke frekuensi Mel Proses mel-frequency wrapping dimulai dengan menentukan jumlah filter yang akan digunakan. Melalui cara menentukan banyaknya filter yang dibutuhkan. ukuran filter yang digunakan pada umumnya berukuran 20 hingga 40[19]. sebagaimana persamaan (5) dan persamaan (6)

$$f_b(m) = \hat{f}_{mel}^{-1} \left(\hat{f}_{mel}(f_{lin-low}) + m \cdot \frac{\hat{f}_{mel}(f_{lin-high}) - \hat{f}_{mel}(f_{lin-low})}{P+1} \right) \quad (5)$$

$$H_s(k) = \begin{cases} 0 & \text{for } k < f_s(m-1) \\ \frac{k - f_s(m-1)}{f_s(m) - f_s(m-1)} & \text{for } f_s(m-1) \leq k \leq f_s(m) \\ \frac{f_s(m+1) - k}{f_s(m+1) - f_s(m)} & \text{for } f_s(m) \leq k \leq f_s(m+1) \\ 0 & \text{for } k > f_s(m+1) \end{cases} \quad (6)$$

berdasarkan filter iniliha filter bank akan dibangkitkan, seiring dengan terbentuknya filler bank maka tahapan berikutnya adalah frequency warping untuk menghasilkan output nilai log-mel frequency cepstrum. sebagaimana persamaan (7).

$$X_i = \log_{10} \left(\sum_{k=0}^{N-1} |X(k)| H_i(k) \right) \quad (7)$$

5. Cepstrum

Tahap akhir Cepstrum adalah proses menghasilkan fitur dari proses ekstraksi fitur yang telah dilakukan, fitur yang dihasilkan berasal dari perubahan frekuensi domain menjadi waktu dirubah menggunakan DCT (*Discrete Cosine Transform*) jumlah dari koefisien yang digunakan sebagai ciri. Dimana jumlah yang digunakan dari nilai i minimum yang umum dipilih dalam pengenalan suara adalah 12 dan maksimum setara dengan jumlah yang didefinisikan [19]. sebagaimana persamaan (8)

$$C_j = \sum_{i=1}^M X_i \cos \left(j(i-1) / 2 \frac{\pi}{M} \right) \quad (8)$$

3.3.3 Backpropagation Neural Network (BPNN)

BPNN atau disebut dengan (Backpropagation Neural network) adalah metode umum yang digunakan untuk proses klasifikasi, dimana metode ini mengenali suatu pola inputan tertentu dan melakukan perbaikan bobot yang digunakan sebagai penghantar antar lapisan. yang pada dasarnya algoritma ini bekerja sebagaimana syaraf biologis pada jaringan

manusia bekerja [19]. Proses training pada metode BPNN memiliki 3 buah tahapan yakni tahap maju yang kemudian dilanjutkan dengan propagasi mundur dan tahapan perubahan bobot. Adapun contoh sederhana dari penyelesaian satu lapisan hidden yang dilengkapi dengan fungsi aktivasi sigmoid biner dapat dilihat pada persamaan sebagaimana berikut:

Proses 1 : dimulai dengan proses pemberian inisial dari setiap bobot dengan angka random, tahap berikutnya adalah proses pengecekan kondisi, apabila kondisi penghentian sistem belum juga ditemui oleh sistem yang bekerja maka proses dilanjutkan ke langkah dua sampai dengan langkah delapan,

Proses 2 : Pada langkah kedua untuk setiap data pelatihan, lakukan langkah tiga sampai dengan langkah kedelapan

Tahap I : Forward propagation

Proses 3 :pada tahap satu propagasi maju pada langkah tiga : terjadi proses penerimaan sinyal yang kemudian diteruskan pada *unit hidden* layer yang terdapat di atasnya

Proses 4 :pada langkah empat terjadi prose perhitungan dari semua luaran pada uni hidden layer z_j ($j = 1, 2, \dots, p$) (9) dan (10) berikut.

$$z_{net_j} = v_{j0} \sum_{i=0}^n x_i v_{ji} \quad (9)$$

$$z_j = f(z_{net_j}) = \frac{1}{1 + e^{-z_{net_j}}} \quad (10)$$

Proses 5 : Terjadinya proses perhitungan semua output neuron y_k ($y = 1, 2, \dots, m$) penyelesaian sebagaimana berikut (11) dan (12) berikut :

$$y_{net_k} = w_{k0} + \sum_{j=1}^p z_j w_{kj} \quad (11)$$

$$y_k = f(y_{net_k}) = \frac{1}{1 + e^{-y_{net_k}}} \quad (12)$$

Tahap II : Backward propagation

Proses 6 : Berdasarkan error yang didapat lakukan penghitungan output dari semua hasil uni keluaran y_k ($y = 1, 2, \dots, m$) dengan Persamaan (13) berikut :

$$\delta k = (tk - yk) f'(y_{net_k}) = (tk - yk) yk (1 - yk) \quad (13)$$

δk Merupakan error unit yang digunakan dalam pembentukan suatu nilai perubahan bobot pada layer dibawahnya (proses 7) dilajutkan dengan prose perhitungan nilai dari w_{kj} (yang akan digunakan untuk melakukan melanjutkan

proses berikutnya dalam merubah nilai bobot w_{kj} dengan percepatan yang digunakan adalah

α . Perhitungan dilakukan dengan Persamaan (14) berikut :

$$\Delta w_{kj} = \alpha \delta_k z_j ; k = 1, 2, \dots, m ; j = 0, 1, 2, \dots, p \quad (14)$$

Proses 7 : lakukan perhitungan nilai dari faktor δ dari *hidden unit* beracuan dari error pada hidden unit z_j ($j = 1, 2, \dots, p$) dengan Persamaan dengan Persamaan (15) berikut :

$$\delta_{net_j} = \sum_{k=1}^m \delta_j w_{kj} \quad (15)$$

Faktor δ *hidden unit* dengan persamaan (16).

$$\delta_j = \delta_{net_j} f'(z_{net_j}) = \delta_{net_j} z_j (1 - z_j) \quad (16)$$

lakukan perhitungan dari perubahan nilai v_{ji} (yang berikutnya digunakan untuk menghitung v_{ji}) dengan Persamaan (17).

$$\Delta v_{ji} = \alpha \delta_j x_i ; j = 1, 2, \dots, p ; i = 0, 1, 2, \dots, m \quad (17)$$

Tahap III : Perubahan bobot

Proses 8 :Lakukan penghitungan pada semua nilai perubahan nilai garis yang berikunta mengarah pada *hidden unit* (18).

$$w_{kj} (baru) = w_{kj} (lama) + \Delta w_{kj} (k = 1, 2, \dots, m ; j = 0, 1, 2, \dots, p) \quad (18)$$

Pergantian nilai garis yang mengarah ke unit tersembunyi dengan menggunakan format berikut adapun persamaan dapat dilihat pada (19).

$$v_{ji} (baru) = v_{ji} (lama) + \Delta v_{ji} (j = 1, 2, \dots, p ; i = 0, 1, 2, \dots, n) \quad (19)$$

ketika proses training telah usai dilakukan, maka proses yang selanjutnya adalah melakukan pengenalan pola dari suatu sinyal suara atau audio. Prose 4 dan 5 adalah yang dipilih untuk menentukan output neuron. hal ini berlaku apabila fungsi yang digunakan adalah bukan sigmoid biner maka langkah ke 4 dan 5 haruslah digunakan

3.4 Teknik Pengujian

Confusion Matrix digunakan bertujuan untuk melakukan grouping pada empat bagian yang digunakan untuk menemukan nilai besarnya akurasi dari suatu pengujian [23]. Tabel Confusion Matrix dapat dilihat dalam Tabel 1.

Tabel 1: Confusion Matrix

		Nilai Sebenarnya	
		Benar	Salah
Prediksi	Benar	True Positive	False Positive
	Salah	False Negative	True Negative

True Positives (TP) menunjukkan prediksi bernilai benar sesuai dengan kondisi yang sesungguhnya bernilai benar. *False Positives* (FP) menunjukkan prediksi bernilai benar dimana pada kondisi yang sesungguhnya bernilai salah. *True Negatives* (TN) menunjukkan prediksi bernilai salah sesuai dengan kondisi yang sesungguhnya bernilai salah. *False Negatives* (FN) menunjukkan prediksi bernilai salah dimana pada kondisi yang sesungguhnya bernilai benar. Berdasarkan output *Confusion Matrix* sehingga diperoleh akurasi, presisi, dan recall [24].

Akurasi merupakan rasio prediksi benar (positif dan negatif) dengan keseluruhan data. Adapun

persamaan yang digunakan seperti pada persamaan (20)

$$akurasi = \frac{TP+TN}{TP+FN+FP+TN} \quad (20)$$

Presisi Adalah rasio prediksi dari *true positif* yang dilakukan proses perbandingan dengan keseluruhan hasil yang mendapatkan nilai positif, sebagaimana persamaan (21)

$$presisi = \frac{TP}{TP+FP} \quad (21)$$

Sedangkan recall merupakan rasio prediksi benar positif yang nantinya dilakukan proses

perbandingan dengan semua data yang dinyatakan benar positif, sebagaimana persamaan (22)

$$recall = \frac{TP}{TP+FN} \quad (22)$$

4. HASIL DAN PEMBAHASAN

Berdasarkan skenario pengujian yang dilakukan, antara lain sebagai berikut.

- Jumlah layer tersembunyi atau dikatakan hidden layer memanfaatkan nilai uji coba 1 (satu), 2 (dua), dan 3 (tiga) hidden layer.
- Sedangkan jumlah neuron yang digunakan pada setiap layer tersembunyi adalah 64, 128, 256, dan 512.
- Besar learning rate pada pengujian ini adalah 0.1 ~0.5.
- Besar dropout yang digunakan yaitu antara 0.2, 0.3, 0.4, dan 0.5.
- Durasi musik, dimana yang digunakan proses pengujian terdiri dari beberapa variasi durasi musik yaitu 30s, 45s, dan 60s.

- Panjang frame, dimana pada tahap ini diujikan beberapa variasi panjang frame yaitu sebesar 20 ms, 30ms, dan 40ms.
- Panjang overlap, dimana pada tahap ini diujikan beberapa variasi overlap frame yaitu 30%, 40%, dan 50%.

Skenario uji dijalankan secara berurutan dari pertama hingga terakhir dengan menggunakan Jupyter Notebook, dimana nilai terbaik untuk pembentukan model BPNN, yaitu pada skenario 1,2,3 dan 4 akan digunakan untuk melakukan skenario pengujian 5,6 dan 7. Adapun hasil terbaik masing-masing skenario mengacu pada nilai akurasi, presisi, dan recall-nya.

4.1 Pengujian Pengaruh

Proses ini adalah proses pengujian pada 3 komponen utama metode klasifikasi yang telah dilakukan karena ketentuan untuk nilainya sendiri tidak memiliki ketetapan sehingga harus dilakukan pengujian sehingga dapat menghasilkan hasil model yang terbaik.

Tabel 2: Hasil Pengujian Neuron 64

Hidden layer	Dropout	Accuracy (%)	Precision (%)	Recall (%)
1	0,2	75,00	76,27	75,00
1	0,3	71,66	71,66	71,66
1	0,4	76,67	76,67	76,67
1	0,5	71,67	72,88	71,67
2	0,2	73,33	74,58	73,33
2	0,5	73,33	73,33	73,33
3	0,2	70,00	70,00	70,00
3	0,5	71,67	71,67	71,67

Dilakukan pengujian menggunakan 1, 2, dan 3 hidden layer yang dimana masing-masing hidden layer diuji menggunakan neuron sebanyak 64, 128, 256, dan 512 dengan pengujian dropout 0.2, 0.3, 0.4, dan 0.5. dengan penggunaan learning rate 0,001 dan epoch 1000. Pada pengujian ini digunakan satu saja skenario pada fitur musiknya, yaitu dengan durasi 30s, panjang frame 20ms, dan overlap sebesar 30%.

Adapun pengujian dilakukan dengan perbandingan data train:test sebesar 7:3. Hasil

dari pengujian yaitu sebagai berikut. Pada pengujian di tabel 2 dapat dilihat bahwa 64 neuron adalah dengan 1 hidden layer dan drop out 76,67%, presisi sebesar 76,67%, dan recall sebesar 76,67%.

Berdasarkan Tabel 3, hasil tertinggi yang didapatkan pada pengujian 128 neuron adalah menggunakan 2 layer tersembunyi dan dropout sebesar 0.3, mendapatkan akurasi 76,67%, presisi sebesar 76,67%, dan recall sebesar 76,67%.

Tabel 3: Hasil Pengujian Neuron 128

<i>Hidden layer</i>	<i>Dropout</i>	<i>Accuracy (%)</i>	<i>Precision (%)</i>	<i>Recall (%)</i>
1	0,2	75,00	75,00	75,00
1	0,5	71,67	71,67	71,67
2	0,2	73,33	72,88	72,88
2	0,3	76,67	76,67	76,67
2	0,4	73,33	73,33	73,33
2	0,5	71,67	71,67	71,67
3	0,2	68,33	69,49	69,49
3	0,5	70,00	70,00	70,00

Berdasarkan Tabel 4, Hasil tertinggi yang didapatkan pada pengujian 256 neuron adalah dengan menggunakan 1 hidden layer dan dropout sebesar 0.4, menghasilkan akurasi sebesar 78,33%, presisi sebesar 79,66%, dan recall sebesar 78,33%. Berdasarkan Tabel 5, hasil tertinggi yang didapatkan pada pengujian 512 neuron adalah dengan menggunakan 3

hidden layer dan dropout sebesar 0.4, menghasilkan akurasi sebesar 78,33%, presisi sebesar 77,97%, dan recall sebesar 76.67%. Dari hasil seluruh pengujian diatas, hasil tertinggi didapat oleh model BPNN dengan 1 hidden layer, 256 neuron, dan 0.4 dropout. Hasil ini akan digunakan untuk pengujian selanjutnya.

Tabel 4: Hasil Pengujian Neuron 256

<i>Hidden layer</i>	<i>Dropout</i>	<i>Accuracy (%)</i>	<i>Precision (%)</i>	<i>Recall (%)</i>
1	0,2	75,00	75,00	75,00
1	0,3	76,67	76,67	76,67
1	0,4	78,33	79,66	78,33
1	0,5	76,67	76,67	76,67
2	0,2	71,67	71,67	71,67
2	0,5	75,00	75,00	75,00
3	0,2	61,67	61,67	61,67
3	0,5	75,00	75,00	75,00

Tabel 5: Hasil Pengujian Neuron 512

<i>Hidden layer</i>	<i>Dropout</i>	<i>Accuracy (%)</i>	<i>Precision (%)</i>	<i>Recall (%)</i>
1	0,2	75,00	75,00	75,00
1	0,5	76,67	76,67	76,67
2	0,2	76,67	76,67	76,67
2	0,5	76,67	76,67	76,67
3	0,2	71,67	71,19	70,00
3	0,3	78,33	77,97	76,67
3	0,4	73,33	73,33	73,33
3	0,5	76,6	76,67	76,67

4.2 Pengujian Learning Rate

Untuk mengetahui pengaruh atau peran dari learning rate Pengujian ini bertujuan untuk mengetahui pengaruh dari learning rate dalam sebagai parameter training untuk mengoreksi

weight (bobot) pada proses training. Pada Learning rate yang dipilih untuk digunakan sebesar 0.001-0.005. Pada pengujian ini digunakan model dengan hasil terbaik sebelumnya. Adapun hasil dari pengujian learning rate sebagai berikut.

Tabel 6: Hasil Pengujian Learning Rate

Learning rate	Accuracy (%)	Precision (%)	Recall (%)
0.001	78,33	77,97	76,67
0.002	76,67	76,67	76,67
0.003	76,67	76,67	76,67
0.004	76,67	76,67	76,67
0.005	76,67	76,27	75,00

Berdasarkan Tabel 6, hasil tertinggi yang didapatkan pada pengujian di atas yaitu adalah nilai learning rate menunjukkan 0.001, menghasilkan akurasi prediksi pada sistem sebesar 78,33%, presisi sebesar 77,97%, dan recall sebesar 76,67%. Hasil tersebut akan digunakan untuk membuat model pengujian selanjutnya.

4.3 Pengujian Durasi, Panjang frame, dan Besar overlap

Dari beberapa pengujian sebelumnya dapat disimpulkan bahwa model BPNN yang menghasilkan hasil paling tinggi yaitu dengan menggunakan 1 hidden layer 256 neuron, 0.4 dropout, 1000 epoch dan learning rate 0.001. pada pengujian akan dilakukan pengujian untuk semua dataset berdasarkan durasi (30s, 45s, 60s), panjang frame(20ms, 30ms, 40ms), dan besar overlap (30%, 40%, 50%). Pengujian akhir ini penggunaan *K-fold Cross Validation* dengan nilai K atau fold senilai 10 dengan perbandingan data train:test sebesar 7:3. Adapun hasil yang didapat adalah sebagai berikut.

Tabel 7: Output Pengujian Durasi, Panjang Frame, dan Overlap

Durasi (s)	Frame (ms)	Overlap (%)	Accuracy (%)	Precision (%)	Recall (%)
60	40	50	86.33	87.07	2.285,99
60	40	30	86.83	88.22	1.277,67
60	40	40	87.67	87.81	1.475,39
60	40	50	87.5	88.37	1.720,37

Berdasarkan Tabel 7, pengujian durasi, panjang frame, dan besar overlap, diperoleh hasil tertinggi menggunakan durasi 60s, panjang frame 40ms, dan overlap 40% dengan nilai akurasi 87,67%, presisi 87,81%, dan recall 86,50% dengan waktu komputasi 1.475,39 detik.

5. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan pada penelitian ini maka kesimpulan yang dapat disusun adalah:

1. Pada penelitian yang dilakukan, diperoleh model BPNN terbaik yaitu dengan menggunakan 1 hidden layer 256 neuron, 0.4 dropout, 1000 epoch dan learning rate 0.001 dengan data yang memiliki durasi 60s, panjang frame 40ms, dan overlap 40% dengan akurasi 87.67%.
2. Pada ekstraksi MFCC perbedaan durasi, panjang frame, dan besar overlap

mempengaruhi banyak fitur yang dihasilkan yang mempengaruhi hasil dari klasifikasi.

Ada beberapa saran yang penulis berikan jika dilakukannya pengembangan atas penelitian ini yaitu:

1. Menggabungkan MFCC dengan fitur lainnya yang biasa digunakan bidang MIR seperti fitur spektral dan temporal.
2. Memperbanyak jumlah data dan varian mood yang beragam.
3. Menerapkan Dimensionality Reduction agar beban komputasi dapat lebih ringan sehingga komputer dapat lebih cepat.

PERNYATAAN PENGHARGAAN

Terimakasih saya ucapkan kepada seluruh rekan-rekan saya yang telah membantu dan berkontribusi dalam penelitian saya ini.

DAFTAR PUSTAKA

- [1] R. S. Job, "The effect of mood on helping behavior," *J. Soc. Psychol.*, vol. 127, no. 4, pp. 323–328, 1987, doi: 10.1080/00224545.1987.9713711.
- [2] K. M. Heilman and K. M. Heilman, "Emotions and Mood," *Athl. Brain*, vol. 14, no. 2, pp. 77–80, 2018, doi: 10.4324/9780429428029-8.
- [3] I. G. Hersemadi, "Implementasi Fast Fourier Transform Pada Ekstraksi Fitur Mood," in *Prosiding Seminar Nasional Multidisiplin Ilmu*, 2017, pp. 121–129.
- [4] G. Jawaharlal Nehru, S. Jothilakshmi, T. Nadu, and T. Nadu, "Music Genre Classification using Deep Neural Networks," *IJSRSET*, vol. 4, no. 4, 10pp. 935–940, 2018.
- [5] Z. Effendi, T. Erlina, and R. Aishwarya, "Pengenalan Suara Menggunakan Metode MFCC (Mel Frequency Cepstrum Coefficients) dan DTW (Dynamic Time Warping) untuk Sistem Penguncian Pintu ISBN : 979-26-0280-1 ISBN : 979-26-0280-1," in *Seminar Nasional Teknologi Informasi dan Komunikasi Terapan (SEMANTIK) 2015*, 2015, pp. 239–243.
- [6] G. Kour and N. Mehan, "Music Genre Classification using MFCC, SVM and BPNN," *Int. J. Comput. Appl.*, vol. 112, no. 6, pp. 12–14, 2015.
- [7] S. Masood, J. S. Nayal, R. K. Jain, M. N. Doja, and M. Ahmad, "MFCC, Spectral and Temporal Feature based Emotion Identification in Songs," *Int. J. Hybrid Inf. Technol.*, vol. 10, no. 5, pp. 29–40, 2017, doi: 10.14257/ijhit.2017.10.5.03.
- [8] I. G. Harsemadi and I. M. Sudarma, "Penggolongan Musik Terhadap Suasana Hati Menggunakan Metode K-Means," in *Konferensi Nasional Sistem & Informatika 2017*, 2017, pp. 49–54.
- [9] P. D. Prasetyo, I. G. P. Suta Wijaya, and A. Yudo Husodo, "Klasifikasi Genre Musik Menggunakan Metode Mel-Frequency Cepstrum Coefficients dan K-Nearest Neighbors Classifier," *J. Teknol. Informasi, Komputer, dan Apl. (JTika)*, vol. 1, no. 2, pp. 189–197, 2019, doi: 10.29303/jtika.v1i2.41.
- [10] E. Jaya and Santosa, "Klasifikasi Genre Musik Menggunakan Metode Support Vector Machine," *J. Ilm. Fasilkom*, pp. 81–82, 2016.
- [11] A. Goel, M. Sheezan, S. Masood, and A. Saleem, "Genre Classification of Songs Using Neural Network," in *Proceedings - 5th IEEE International Conference on Computer and Communication Technology, ICCCT 2014, 2015*, pp. 285–289, doi: 10.1109/ICCCT.2014.7001506.
- [12] W. S. M. Sanjaya and Z. Salleh, "Implementasi Pengenalan Pola Suara Menggunakan Mel-Frequency Cepstrum Coefficients (MFCC) dan Adaptive Neuron-Fuzzy Inferense System.
- [13] L. Halimah, "Musik Dalam Pembelajaran," *EduHumaniora J. Pendidik. Dasar Kampus Cibiru*, vol. 2, no. 2, pp. 1–9, Jul. 2016, doi: 10.17509/eh.v2i2.2763.
- [14] W. Chijioke, "Predicting Listener'S Mood Based on Music Genre: an Adapted Reproduced Model of Russell and Thayer," *J. Technol. Manag. Bus.*, vol. 4, no. 1, pp. 39–58, 2017.
- [15] Z. Fu, G. Lu, K. M. Ting, and D. Zhang, "Learning naive bayes classifiers for music classification and retrieval," in *Proceedings - International Conference on Pattern Recognition, 2010*, pp. 4589–4592, doi: 10.1109/ICPR.2010.1121.
- [16] C. Wei, *Automated Analysis of Musical Structure*. Massachusetts Institute of Technology, 2005.
- [17] S. Ajibola Alim and N. Khair Alang Rashid, "Some Commonly Used Speech Feature Extraction Algorithms," in *From Natural to Artificial Intelligence - Algorithms and Applications*, IntechOpen, 2018, pp. 1–19.
- [18] H. D. Arani, "Identifikasi Ayat pada Bacaan Menggunakan Metode Dynamic Time Warping Berdasarkan Fitur Mel Frequency Cepstral Coefficient untuk Sistem Tutorial Hafalan Al-Quran," Jul. 2017.
- [19] A. H. Abriyono, "Pengenalan Ucapan Suku Kata Bahasa Lisan Menggunakan Ciri LPC, MFCC, dan JST," *IJCCS*, vol. 6, no. 2, pp. 23–34, 2013, doi: 10.22146/ijccs.2149.
- [20] A. Andana, R. Widiyati, and M. Irzal, "Pengenalan Citra Tulisan Tangan Dengan Metode Backpropagation," *JMT J. Mat. dan Terap.*, vol. 2, no. 1, pp. 36–44, May 2018, Accessed: Sep. 01, 2020. [Online]. Available: <http://journal.unj.ac.id/unj/index.php/jmt/article/view/7166>.
- [21] M. A. F. Azlah, L. S. Chua, F. R. Rahmad, F. I. 11Abdullah, and S. R. W. Alwi, "Review on techniques for plant leaf classification and recognition," *Computers*, vol. 8, no. 4, 2019, doi: 10.3390/computers8040077.

- [22] Julpan, E. B. Nababan, and M. Zarlis, "Analisis Fungsi Aktivasi Sigmoid Biner Dan Sigmoid Bipolar Dalam Algoritma Backpropagation Pada Prediksi Kemampuan Siswa," *J. Teknovasi*, vol. 02, no. 1, pp. 103–116, 2015.
- [23] Y. Marchel and J. Nasri, "Perbandingan Tingkat Akurasi Support Vector Machine dengan Naive Bayes pada Studi Kasus Okupansi Lahan Berdasarkan Kondisi Cuaca Comparison of Accuracy Level of Support Vector Machine with Naive Bayes on Land Occupancy Case Study Based on Weather Conditio," vol. 4, no. 3, pp. 4946–4949, 2017.
- [24] J. Davis and M. Goadrich, "The Relationship Between Precision-Recall and ROC Curves Jesse," in *The SAGE Handbook of Social Geographies*, 2010, pp. 546–559, doi: 10.4135/9780857021113.n29.
- [25] "Mengenal Accuracy, Precision, Recall dan Specificity serta yang diprioritaskan dalam Machine Learning | by Resika Arthana | Medium."
<https://medium.com/@rey1024/mengenal-accuracy-precision-recall-and-specificity-septa-yang-diprioritaskan-b79ff4d77de8>
(accessed Oct. 12, 2020).