PERANCANGAN SISTEM PENCARIAN LAGU INDONESIA MENGGUNAKAN QUERY BY HUMMING BERBASIS LONG SHORT-TERM MEMORY

Henry Hartono [1] Viny Christanti Mawardi, M.Kom [2] Janson Hendryli S. Kom. M.Kom.[3]

[1] [2] [3] Teknik Informatika Universitas Tarumanagara
Jl. Letjen S. Parman No. 1, Jakarta 11440 Indonesia
Goolman1231@gmail.com[1], vinym@fti.untar.ac.id[2] , jansonh@fti.untar.ac.id[3]

ABSTRACT

Song identification dan query by humming is an application that is developed using Mel-frequency cepstral coefficients *(MFCC) and Long Short-Term Memory (LSTM) algorithm.The application purpose is to detect and recognize humming from the input data. In this application the humming input will be divided into two parts, namely the training audio and test audio. For the training audio, the training audio will be divided into two process stages, namely recognizing humming and searching for the unique features of a humming audio.*

To recognize the humming feature, the humming will be processed using the MFCC method. After obtaining a part of the MFCC Features, the MFCC features will be saved as a vector model. The feature that has been extracted will be learned by the LSTM method. For the test audio of the stages carried out as in the training audio, after the MFCC Feature is detected, an introduction will be made based on learning that has been done with the LSTM method to obtain output in the form of a song name that is successfully recognized and detected will be labeled by the application.

Key words

Mel Frequency Cepstral Coefficient (MFCC), Long Short-Term Memory (LSTM),Query By Humming (QBH).

1. Pendahuluan

Musik Merupakan Sebuah Alat hiburan Yang Dapat Dinikmati oleh Semua Kalangan Mulai dari yang umur Muda hingga yang Tua,Sehingga musik sudah menjadi sebuah kebutuhan penting dalam kehidupan manusia, baik dari segi kegiatan upacara, keagamaan maupun sebagai sarana hiburan [1].

Indonesia sendiri merupakan salah satu negara yang multikultur terbesar di dunia. Sebagai bangsa yang besar, dengan jumlah penduduk yang sangat besar, kekayaan alam yang melimpah dan wilayah yang sangat besar. Indonesia sendiri merupakan negara kekayaan dengan budaya dan bahasa. Namun Musik indonesia dengan kekayaan budaya masih tidak mampu bersaing popularitas dengan lagu luar negeri seperti korea dan amerika.

Padahal banyak masyarakat Indonesia masih menikmati lagu Indonesia seperti dangdut atau keroncong tetapi tidak dapat mengingat nama tersebut oleh karena itu dibuatlah sistem Query by Humming (QbH). Dengan sistem QbH ini, lagu dapat dicari dengan menyenandung melodi dari lagu tersebut. Pada dasarnya sistem ini dibuat dengan mencari kemiripan melodi dari senandungan (query) dengan melodi dari dataset humming yang ada di database.

MFCC Digunakan untuk merepresentasikan frekuensi data audio menjadi data frekuensi yang dapat didengarkan oleh manusia, dengan ekstraksi frekuensi yang penting dalam audio tersebut maka dapat mendapat data humming tersebut.

Deep learning adalah salah satu bidang yang terdapat pada machine learning. Kata *deep* tersebut disebut dikarenakan jumlah dan struktur jaringan saraf algoritma tersebut sangat banyak dan bisa mencapai hingga lebih dari ratusan lapisan (layer). [2] Dengan begitu maka mesin juga dapat melakukan proses pembelajaran sehingga dapat disebut mempunyai "kecerdasan" layaknya manusia. Mesin juga dinilai lebih teliti dalam melakukan komputasi.

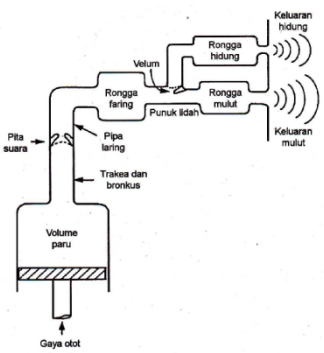
Merancang sebuah sistem pengocokan vektor menggunakan metode Long Short-Term Memory (LSTM) yang merupakan jenis dari Recurrent Neural Network (RNN). RNN akan memproses data melalui beberapa lapisan dengan masukan berupa data sekuensial. Hasil Data MFCC Merupakan sebuah vektor, maka data ini merupakan data sekuensial yaitu data yang perlu diproses secara berurutan dan memiliki hubungan satu dengan yang lainnya.

Metode LSTM dipilih karena pada metode ini terdapat tambahan sel memori (*memory cell*) jangka

panjang yang tidak dimiliki oleh RNN. LSTM dapat mempelajari data mana saja yang akan disimpan dan data mana saja yang akan dibuang, karena pada setiap neuron LSTM memiliki beberapa gerbang yang mengatur memori pada setiap neuron itu sendiri. [3]

2. Dasar Teori

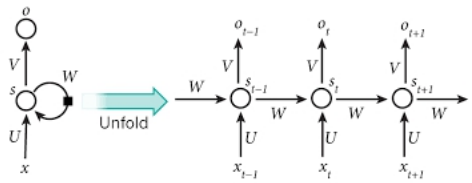
Produksi suara gelombang suara bervariasi mulai dari variasi dari tekanan media perantara seperti udara. Suara diciptakan oleh getaran dibuat dari suatu objek, yang jadi sebabnya udara yang ada disekitar bergetar.Yang pertama adalah Produksi suara karena suara adalah mekanisme terjadinya suara.suara manusia dapat digolongkan kedalam kelompok alat musik tiup.sebelum menyanyi, harus memompa atau menghirup udara melalui hidung(inhalasi) masuk ke dalam paru–paru dibantu oleh otot perut,otot dada,otot sisi tubuh dan otot diafragma.kemudian paru–paru mengalirkan kembali udara keluar atau dihembuskan(ekshalasi) sedemikian rupa sehingga membentuk pita suara yang terdapat didalam larynx(tenggorokan) bentuk pita suara ini seperti selaput yang terbelah dibagian tengahnya. Pita suara terbuka pada saat menghirup udara dan akan menutup dan bergetar pada saat bersuara (bernyanyi/berbicara) menjadi suara yang jelas dan indah di dalam rongga mulut.sebenarnya pita suara ini tidak menutup secara total tetapi masih ada celah kecil sehingga akibat tekanan udara dari bawah membuat pita suara ini bergetar.getaran ini diperkuat dan diperbesar oleh rongga resonansi yang ada pada tubuh.[4].



Gambar 1. Alur Kerja Produksi Suara Manusia

2.1. Recurrent Neural Network (RNN)

Recurrent neural network (RNN) merupakan bagian dari neural network berfungsi untuk memproses sejumlah data yang data bersambung atau sequential data. RNN tersebut merupakan pengembangan dari Jaringan Syaraf Tiruan (JST) , Arsitektur RNN mirip dengan Multi Layer Perceptron (MLP) [5].



Gambar 2. Alur kerja RNN

Persamaan yang digunakan untuk menghitung *hidden state* dari **t** sampai **T** ditunjukan pada persamaan (1), (2), dan (3). [6].

$$h_t = \tanh(W \cdot h_{t-1} + U \cdot x_t) \quad (1)$$

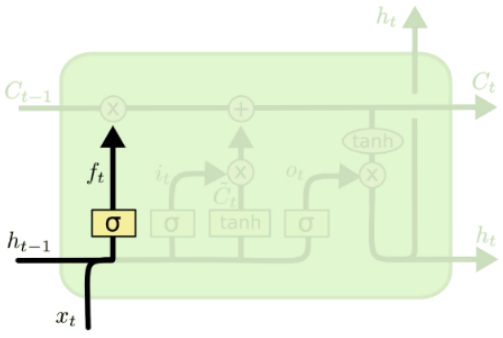
$$O_t = V \cdot h_t \quad (2)$$

$$\hat{y}_t = \text{softmax}(O_t) \quad (3)$$

2.2. Long Short-Term Memory (LSTM)

Long Short Term Memory networks (LSTM) merupakan perkembangan dari arsitektur RNN,Arsitektur LSTM diperkenalkan pertama kali oleh Hochreiter & Schmidhuber pada tahun 1997. Penelitian ini dilakukan banyak oleh para peneliti yang terus mengembangkan arsitektur LSTM ini dalam berbagai seperti dalam bidang forecasting atau speech recognition [7].

Forget Gate: Fungsi dari gerbang ini adalah untuk menentukan berapa informasi lama dari cell sebelumnya yang masih ingin disimpan

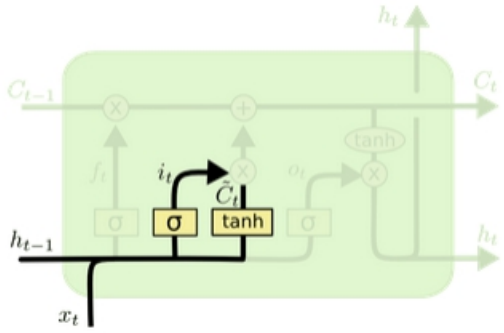


Gambar 3. Forget gate

$$f_t = \sigma(wf_h \cdot h_{t-1} + wf_x \cdot x_t + b_f) \quad (4)$$

Input Gate dan Candidate Gate: Bagian state memerlukan dua gerbang yakni input gate dan

candidate gate. Input gate yang memutuskan nilai mana yang ingin diperbaharui. Gerbang kedua adalah candidate gate yang berisi kandidat nilai yang ditambahkan ke dalam cell state.

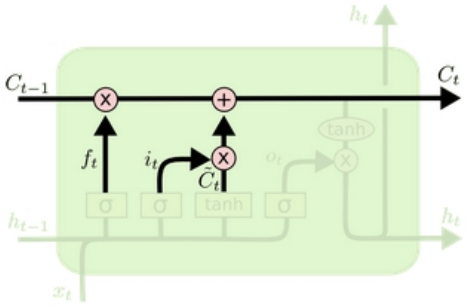


Gambar 4. Input Gate dan Candidate Gate pada LSTM cell

$$i_t = \sigma(wi_h \cdot h_{t-1} + wi_x \cdot x_t + b_i) \quad (5)$$

$$\widetilde{C}_t = \tanh(w\widetilde{C}_h \cdot h_{t-1} + w\widetilde{C}_x \cdot x_t + b_{\widetilde{C}}) \quad (6)$$

Input Gate dan Candidate Gate : Cell state menunjukkan informasi yang melewati setiap LSTM cell dan dapat dilihat pada Gambar 5 Setiap informasi tersebut dikontrol oleh gerbang yang terstruktur.



Gambar 5. Cell State pada LSTM Cell

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \quad (7)$$

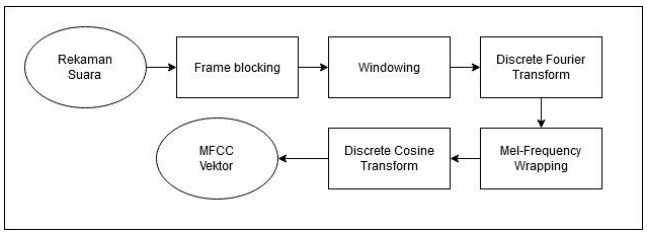
2.2. Mel-frequency cepstral coefficients

Mel Frequency Cepstrum Coefficient (MFCC) adalah sebuah metode ekstraksi yang menggunakan transformasi fourier untuk memperoleh nilai cepstrum dari sinyal suara yang masuk dalam bentuk gelombang transversal [8].
Metode ini digunakan sebagai feature extraction, yaitu sebuah proses mengkonversikan data sinyal suara menjadi beberapa parameter. Keunggulan dari metode ini adalah sebagai berikut [9]:

1. Mampu mengambil karakteristik suara yang sangat penting dalam pengenalan suara atau mengambil informasi-informasi penting yang terkandung dalam sinyal suara tersebut.
2. Menghasilkan data seminimal mungkin, tanpa membuang informasi-informasi penting dari sinyal suara.

3. Mengadaptasi telinga manusia untuk melakukan persepsi terhadap sinyal suara

Input dari metode tersebut berupa sinyal suara yang hasilnya keluaran berupa fitur MFCC. Satu *frame* menghasilkan satu vektor fitur sehingga satu sinyal suara akan menghasilkan beberapa baris vektor fitur.



Gambar 6. Alur kerja MFCC

Frame blocking: Framing merupakan tahapan berfungsi untuk memisahkan sinyal suara menjadi dalam beberapa frame. Kemudian sinyal suara tersebut menjadi berbentuk frame-frame yang menyimpan informasi-informasi untuk nantinya dikonversi dalam bentuk vektor akustik. [10].

Windowing: digunakan dalam meminimalisir diskontinuitas sinyal pada permulaan dan pengakhiran tiap frame. Dengan penggunaan *window*, sinyal suara pada awal dan akhir *frame* menjadi bentuk runcing. Prosesnya adalah sinyal suara dikalikan dengan *window*. [11] Tahap perhitungan *windowing* ditunjukkan dengan persamaan

$$Y_t(n) = X_t(n)w(n) \qquad 0 \le n \le N-1 \quad (8)$$

Banyak fungsi window yang bisa digunakan seperti blackman, gaussian, dan hamming window. Namun pada metode MFCC, window yang sering dipakai adalah hamming window

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right) \quad (9)$$

Discrete Fourier Transform: Bentuk sinyal data suara biasanya dalam domain waktu. Data sinyal dalam domain waktu masih berupa data mentah sehingga

harus diubah menjadi domain frekuensi.[12] Discrete Fourier Transform (DFT) merupakan metode untuk mengubah sinyal domain waktu menjadi domain frekuensi.

$$X(k) = \sum_{n=0}^{N-1} x(n) e\left(\frac{-j2\pi nk}{N}\right), k = 0,1,\ldots, N-1 \qquad (10)$$

Mel-Frequency Wrapping:.Dalam mendapat nilai mel-frequency wrapping dibuat filter bank terlebih dahulu, Skala mel frekuensi merupakan skala frekuensi linier di bawah 1000 Hz dan skala logaritmik di atas 100 Hz, Filter bank tersebut dibuat dengan menggunakan persamaan berikut [13]:

$$f_{mel}(f) = 2495 * \log_{10}(1 + \frac{f}{700}) \qquad (11)$$

Discrete Cosine Transform: koefisien spektrum mel yakni X_i X_i pada tahap sebelumnya masih dalam domain frekuensi. Oleh karena itu nilai koefisien spektrum mel diubah ke domain waktu.. Hasil akhir dari MFCC berupa bilangan rill, oleh karena itu dilakukan Discrete Cosine Transform (DCT) untuk mengubah dari domain frekuensi ke domain waktu. [14].

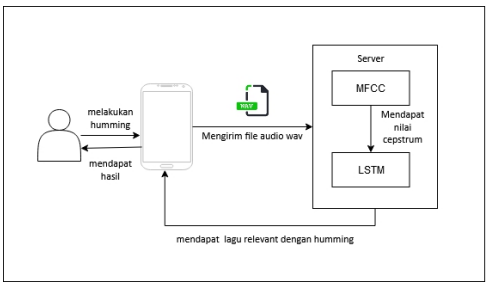
$$C_j = \sum_{i=1}^{P} X_i \cdot \cos\left[i\left(i - \frac{1}{2}\right) \cdot \frac{\pi}{P}\right], untuk\ j = 1,2,\ldots, j \qquad (12)$$

3.Rancangan

Aplikasi tersebut dirancang merupakan aplikasi yang dapat mendeteksi humming(senandung) untuk mendapat sebuah lagu dengan nada yang sama. Aplikasi ini diharapkan dapat membantu orang yang ingin mendapatkan lagu indonesia hanya dengan humming.

Aplikasi ini sendiri dirancang dengan menggunakan sistem berbasis aplikasi dimana untuk *user interface*-nya menggunakan Android. dan menggunakan bahasa pemrograman Python sebagai mesin pencocok lagu dan ekstraksi suara.

Sistem pada aplikasi ini menggunakan langkah langkah dari metode System Development Life Cycle (SDLC). Ada empat tahapan pada metode ini, antara lain Perencanaan, Analisis, Perancangan, serta Implementasi dan Pengujian. Tiga dari empat tahapan tersebut akan dijelaskan pada subbab ini, antara lain tahap Perencanaan, tahap Analisis dan tahap Perancangan.



Gambar 7. Alur kerja Sistem

4. Pengujian

Dataset query by humming terdapat 7 lagu yang bertotal 319 file WAV file tersebut berdurasi 20-30 detik. Berjumlah 8 orang, 6 perempuan dan 2 laki-laki. 6 perempuan terdapat sekolah musik purwacaraka umur antara 15 - 23 berprofesi sebagai pelajar dan mahasiswi. 6 orang tersebut menyanyikan dalam studio dan ruangan sunyi, lagu tersebut dinyanyikan dalam 5 bagian seperti chorus, intro, verse, bridge dan ending.Dalam Pelatihan model digunakan hyperparameter dan model sebagai berikut.

Tabel 1 Hyperparameter yang digunakan.

No.	OPTIMIZER	BATCH SIZE	LEARNING RATE	EPOCH
1.	ADAMAX	8	0.0001	100
2.	ADAMAX	16	0.0001	100
3.	ADAMAX	32	0.0001	100

Tabel 2 Dua Model Arsitektur yang di Uji Coba

Nama Model	Tipe Layer	Parameter	Nilai Parameter	Output Shape
Model 1	LSTM	Dimensi output	128	(none,128)
	LSTM	Dimensi output	64	(none,64)
	Dense	Jumlah Unit	13	(none,13)
	Dense	Jumlah Unit	5	(none,5)
Model 2	LSTM	Dimensi output	64	(none,64)
	LSTM	Dimensi output	64	(none,64)
	Dense	Jumlah Unit	13	(none,13)
	Dense	Jumlah Unit	5	(none,5)

Dalam Pengujian Model Query by Humming tersebut dilakukan secara 2 eksperimen, eksperimen pertama yaitu melakukan pembagian dataset 60% Data

Testing dan 40% Data Validasi, Eksperiment kedua yaitu membagi dataset 80% Data Testing dan 20% Data training

Skenario			Hasil Presentase(%)	
No.	Hyperparameter	Model	Latih	Uji
1.	1	1	0.8936%	0.6790%
2.	2	1	0.8887%	0.7042%
3.	3	1	0.9147%	0.6989%
4.	1	2	0.9066%	0.6618%
5.	2	2	0.8432%	0.6260%
6.	3	2	0.8018%	0.6645%

Gambar 8. Eksperiment Pertama

Skenario			Hasil Presentase(%)	
No.	Hyperparameter	Model	Latih	Uji
1.	1	1	0.7156%	0.6842%
2.	2	1	0.7039%	0.6687%
3.	3	1	0.7091%	0.6710%
4.	1	2	0.7256%	0.6141%
5.	2	2	0.5156%	0.5156%
6.	3	2	0.4656%	0.4156%

Gambar 9. Eksperiment Kedua

Untuk pengujian selanjutnya digunakan 2 orang untuk menyanyikan bagian awal sebuah lagu.berikut merupakan hasil yang didapat:

Tabel 3 Hasil Pengujian

Jenis Orang	Jumlah Data	Jumlah lagu benar	Jumlah lagu Salah	Akurasi Kebenaran
Pria 1	14 data	7 audio	7 audio	50%
Pria 2	14 data	5 audio	9 audio	35%

Pria 1 memiliki suara nada yang jelas dan menyanyi dalam ruangan sunyi sedangkan Pria 2 menyanyi dengan nada yang datar dan menyanyi dalam ruangan dengan banyak noise, noise tersebut berdampak dalam hasil pencarian lagu.

5. Kesimpulan

Berdasarkan pengujian program "Perancangan Sistem Pencarian Lagu Indonesia Menggunakan Query By Humming Berbasis Long Short-term Memory", maka didapatkan kesimpulan sebagai berikut:

1. Bagus atau tidaknya model tidak dapat dilihat hanya dari persentase pelatihan dan validasi.
2. Banyaknya jumlah noise mempengaruhi hasil pencarian humming
3. Kejelasan suara mempengaruhi sistem pencarian.

REFERENSI

[1] Wisnu Mintargo. Budaya Musik Indonesia. (Yogyakarta: Kanisius, 2018) .

[2] Wayan Dadang, Memahami Kecerdasan Buatan berupa Deep Learning dan Machine Learning, https://warstek.com/2018/02/06/deepmachine learning/,Diakses 21 oktober 2020

[3] Muhammad Wildan Putra Aldi.et,al. "Analisis dan Implementasi Long Short Term Memory Neural Network untuk Prediksi Harga Bitcoin", e-Proceeding of Engineering, Vol.5 , Nomor 2, (Agustus, 2018)

[4] Arintyo Archamadi,"Analisis Dan Simulasi Identifikasi Judul Lagu Dari Senandung Manusia Menggunakan Ekstraksi Ciri DCT (Discrete Cosine Transform)", E-proceeding Of Engineering : Vol.III, Nomor.3 (December 2016)

[5] Kuncoro Yoko, Viny Christanti M, Janson Hendryli, "Sistem Peringkas Otomatis Abstraktif Dengan Menggunakan Recurrent Neural Network", Journal of Computer Science and Information Systems, Vol.2, Nomor 1, (April, 2018)

[6] Mulyanti , Sistem Klasifikasi Irama Detak Jantung Menggunakan Deep Learning Long Short-term Memory, Jakarta: Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Tarumanagara (Skripsi tidak Dipublikasikan), 2019

[7] Belajar pembelajaran mesin Indonesia,Pengenalan Recurrent Neural Network (RNN) – Bagian 1, https://indoml.com/2018/04/04/pengenalan-rnn-bag-1/,diakses pada 27 September 2020

[8] Budiman,Fajar et al.,"Pengenalan Suara Burung Menggunakan Melfrequency Cepstrum Coefficient Dan Jaringan Syaraf Tiruan Pada Sistem Pengusir Hama Burung ",Jurnal Nasional Teknik Elektro,Vol: 5, No. 1,(Maret 2016)

[9] Nasution,Torkis."Metoda Mel Frequency Cepstrum Coefficients (MFCC) untuk Mengenali Ucapan pada Bahasa Indonesia",Jurnal Sains dan Teknologi Informasi, Vol. 1, No. 1, (Juni 2012)

[10] Risah Prayogi,Yanuar.Modifikasi Metode Ekstraksi Fitur Mel Frequency Cepstral Coefficient Untuk Identifikasi Pembicara Pada Lingkungan Berderau Menggunakan Residu Endpoint Detection,Surabaya:Program Magister Jurusan Teknik Informatika Fakultas Teknologi Informasi Institut Teknologi Sepuluh Nopember(Skrispi),2015

[11] Tri Ramadhani ,Implementasi Algoritma Mel-frequency Cepstral Coefficients–vector Quantization (Mfcc-vq)untuk Deteksi Suara Burung Pemakan Padi Di Sawah,Medan: Fakultas Ilmu Komputer Dan Teknologi Informasi universitas Sumatera Utara(Skripsi),2018

[12] Muhammad Gufindo Alenra,Penerapan Mel Frequency Cepstrum Coefficients (Mfcc) Dan Backpropagation Neural Network (Bpnn) Untuk Pengenalan Huruf Hijaiyah,riau:universitas Islam Negeri Sultan Syarif Kasim,2018

[13] Marwa A.Nasr, et.al., "Speaker identification based on ormalized pitch frequency and Mel Frequency Cepstral Coefficient", International Journal of Speech Technology , Vol.21, Issue 4,(Desember, 2018)
[14] Agus Buono et.al.,"Perluasan Metode MFCC 1D ke 2D sebagai Ekstraksi Ciri pada Sistem Identifikasi Pembicara Menggn Hidden Markov Model (HMM)", Makara Sains, Vol. 13, Nomor 1,(April 2009)

Henry Hartono, seorang mahasiswa pada proram studi Fakultas Teknologi Informasi di Universitas Tarumanagara

Viny Christanti Mawardi, S.Kom., M.Kom., memperoleh gelar S.Kom Dari Universitas Tarumanagara dan gelar M.Kom dari Universitas Indonesia. Saat ini sebagai Dosen program studi Teknik Informatika, Universitas Tarumanagara.

Janson Hendryli S. Kom. M.Kom., memperoleh gelar S.Kom Dari Universitas Tarumanagara dan gelar M.Kom dari Universitas Indonesia. Saat ini sebagai Dosen program studi Teknik Informatika, Universitas Tarumanagara.