

KLASIFIKASI INTONASI BAHASA JAWA KHAS PONOROGO MENGGUNAKAN ALGORITMA MULTILAYER PERCEPTRON NEURAL NETWORK

Oddy Virgantara Putra^{*1}, Aziz Musthafa², Moch. Kholil³

^{1,2,3}Universitas Darussalam Gontor; Jl. Raya Siman km.6, Kec. Siman,

Kab. Ponorogo; 0352483762

^{*}oddy@unida.gontor.ac.id

ABSTRAK

Intonasi merupakan salah satu keunikan dalam gaya berbicara. memberikan penekanan tinggi dan rendah pada kata-kata tertentu dalam sebuah kalimat. Tuturan penutur Bahasa Jawa Ponoragan berbeda dengan penutur di penjuru lain yaitu adanya nada naik di awal tuturan dan turun di bagian akhir. Beberapa penelitian sebelumnya, Automatic Speech Recognition (ASR) telah dikembangkan dalam bahasa Inggris dan Indonesia, tetapi tidak untuk bahasa daerah, terutama bahasa Jawa. Tujuan peneliti dalam makalah ini adalah pengembangan Automatic Speech Recognition dalam bahasa Jawa Ponoragan. Dataset yang digunakan adalah rekaman suara yang diperoleh dari mahasiswa dan pekerja di Universitas Darussalam Gontor yang merupakan penduduk asli Ponorogo. Selain itu, peneliti menggunakan Mel-frequency Cepstral Coefficient (MFCC) untuk ekstraksi fitur. Model yang digunakan untuk pelatihan dataset adalah Multilayer Perceptron (MLP). Setelah menguji dataset, model Multilayer Perceptron dengan MFCC feature extraction berhasil mencapai akurasi hingga 0.8125.

Kata Kunci : *Intonasi, Ponoragan, MFCC, Gontor, MLP.*

PENDAHULUAN

Dalam kehidupan sehari-hari, bahasa adalah elemen terpenting dalam berkomunikasi. Dengan menggunakan bahasa, seseorang dapat memahami apa yang menjadi maksud dan tujuan kita. Bahasa dapat dikatakan sebagai salah satu media kita untuk berinteraksi satu dengan yang lainnya. Baik itu menggunakan bahasa dalam bentuk lisan maupun tulisan. Dalam hal ini, bunyi dan ujaran dihasilkan oleh alat ucap manusia. Alat ucap dan ciri fisik organ tubuh setiap manusia berbeda satu sama lain, sehingga menghasilkan bunyi dan ujaran yang berbeda pula. Salah satu faktor yang membuat pendengar memahami sebuah ujaran, baik itu kalimat tanya atau perintah, adalah intonasi dari penutur atau pembicara.

Intonasi itu sendiri merupakan gejala prosodi, hubungannya erat dengan struktur kalimat dan interelasi kalimat dalam sebuah wacana, khususnya akan dibahas dalam fonetik akustik, yang menyelidiki ciri-ciri fisik bunyi bahasa (Herti et al., 2015). Oleh karena itu, Bunyi atau ujaran kalimat bahasa Jawa akan terdengar berbeda apabila diucapkan oleh penutur asing yang belajar bahasa Jawa. Salah satu kabupaten di

Jawa timur yang mempunyai bahasa Jawa dengan Intonasi yang Khas adalah Ponorogo.

Kabupaten Ponorogo merupakan sub etnis budaya yang tidak termasuk dalam kawasan budaya Mataraman, melainkan adalah budaya etnis Jawa Ponoragan. Bahasa Jawa dialek Ponoragan sendiri masih terbagi menjadi empat yaitu dialek Brang Kidul, Brang Lor, Brang Wetan, dan Brang Kulon(Sholihah et al., 2019) . salah satu keunikan bahasa Jawa Ponoragan adalah Tuturan penutur Bahasa Jawa Ponoragan yang berbeda dengan penutur di penjuru lain yaitu adanya nada naik di awal tuturan dan turun di bagian akhir.

Pengenalan bahasa lisan (*Spoken language Recognition*) adalah prosedur terprogram yang menentukan karakter bahasa yang diekspresikan secara verbal dalam pengucapan . Kemajuan terbaru dalam *Machine Learning* telah memperluas cakrawala untuk *human-computer interaction* hanya dengan menggunakan suara mereka. Teknologi saat ini sering kali membutuhkan bahasa yang diproses untuk dideklarasikan secara eksplisit. Kemampuan untuk memproses sampel ucapan masukan yang dinamis dari berbagai bahasa akan memperluas kegunaan teknologi *speech processing* yang ada dan membuka berbagai fungsi tambahan (Boussard et al., 2017) .

Spoken language Recognition mengatasi masalah ini dengan mempelajari cara mengekstrak informasi dari sinyal audio ucapan dan menggunakannya untuk memprediksi bahasa yang digunakan ucapan tersebut. Dalam makalah ini, peneliti mengusulkan pendekatan untuk mengidentifikasi bahasa Jawa, dengan mengklasifikasi intonasi bahasa Jawa khas Kabupaten Ponorogo yaitu Ponoragan.

Penelitian ini akan mengolah data suara yang diperoleh dari Mahasiswa dan pekerja di Universitas Darussalam Gontor yang termasuk penduduk asli Ponorogo. data suara yang didapat akan divalidasi dan kemudian diklasifikasi berdasarkan kelas intonasi. Tantangan dari penelitian ini adalah pengembangan *Spoken language Recognition* dan implementasinya pada bahasa jawa khas Ponorogo.

Pengumpulan data suara dikumpulkan menggunakan *Recorder*. Data berjumlah 78 dengan durasi dua sampai tiga detik pada setiap data suara. Format audio yang digunakan adalah *wav* serta menggunakan 44100 Hz *Sample Rate*. Data dibagi sesuai kelas keluaran yaitu low [1], mid [2] dan high [3] pada kata "Padang" dan "Peteng". Selanjutnya, data di *trim* menggunakan Audacity untuk mengambil kata yang memiliki intonasi saja dan akan dilabeli sesuai dengan kelas keluaran. Kemudian, data akan diolah menggunakan *Mel-Spectrum Cepstral Coefficient* (MFCC) untuk mendapatkan informasi data suara yang akan diklasifikasi. Selanjutnya, data di bagi 80% untuk *train* dan 20% untuk *test*.

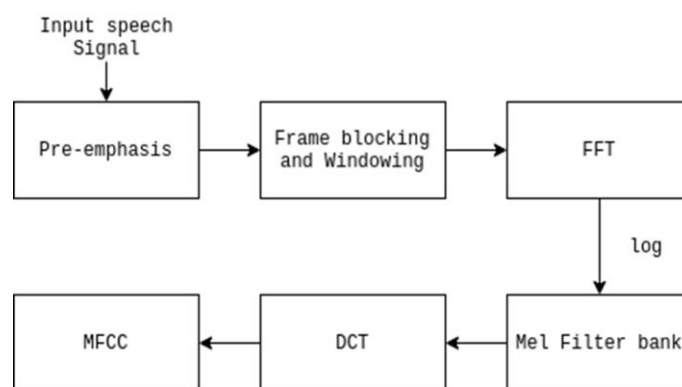
Model yang digunakan adalah *Multilayer Perceptron* (MLP) yang merupakan salah satu pemodelan dalam teknologi jaringan saraf tiruan (JST). Peneliti menggunakan empat layer yaitu satu *input layer*, dua *hidden layer* dan satu *output layer*. Dengan ini, peneliti mengusulkan pemodelan untuk mengklasifikasi Intonasi bahasa Jawa Ponoragan dan mencoba mendapatkan hasil melalui makalah ini.

METODE

Makalah ini berfokus pada metode eksplorasi untuk mengklasifikasi Intonasi bahasa Jawa Ponoragan. Model MLP digunakan untuk melatih hasil ekstraksi fitur dari *Mel-frequency Cepstral Coefficient* (MFCC).

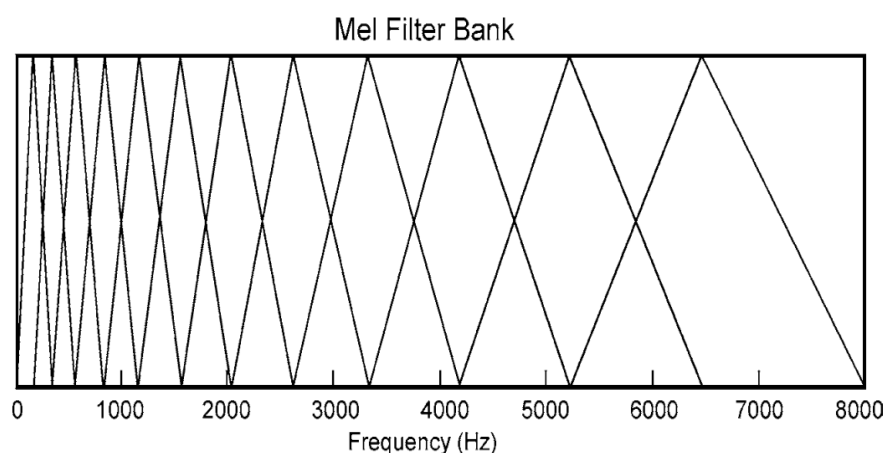
1. Mel-frequency Cepstral Coefficient

MFCC (Mel-frequency Cepstrum Coefficients), merupakan salah satu metode yang banyak digunakan dalam bidang speech technology baik speaker recognition maupun speech recognition. Metode ini digunakan untuk melakukan feature extraction, sebuah proses yang mengkonversikan signal suara menjadi beberapa parameter (Nasution et al, 2012). Alur ekstraksi fitur MFCC akan dijelaskan dengan gambar dibawah ini.



Gambar 1. Alur Ekstraksi MFCC

Pada proses Pre-emphasis sinyal suara akan diseimbangkan dengan menerapkan filter pada setiap frekuensi tinggi. Karena frekuensi tinggi biasanya memiliki besaran lebih kecil dibandingkan dengan frekuensi rendah. Setelah filter Pre-emphasis digunakan maka tahap selanjutnya adalah Frame blocking dengan membuat bingkai-bingkai pada sinyal suara setiap 10ms. Fungsi frame blocking adalah mengatasi nilai yang berubah-ubah pada satu sinyal suara. Selanjutnya, windowing diterapkan untuk mencegah nilai yang terus berkelanjutan pada ujung setiap sinyal suara. Kemudian algoritma *Discrete Fourier Transform* (DFT) digunakan pada *Fast Fourier Transform* untuk mengeluarkan informasi di dalam *frequency domain* menjadi *magnitude spectrum*. Setelah *magnitude spectrum* didapatkan, *filter bank* akan dihitung sebanyak 40, 60, 80 dan 80 filter pada setiap *power spectrum* untuk mengeluarkan *frequency bands*. Mel-scale bertujuan untuk meniru persepsi telinga manusia, agar lebih diskriminatif pada frekuensi yang lebih tinggi. Mel-scale dapat diperhatikan dalam gambar di bawah ini.



Gambar 2. Mel Filter Bank

Setelah menghitung *Filter bank*, *Discrete cosine transform* (DCT) digunakan, fungsinya adalah menjadikan hasil *Filter bank* tidak korelasi satu sama lain. Alasan membuang koefisien karena menunjukkan perubahan yang sangat cepat dalam *filter bank coefficient* dan detail ini tidak digunakan pada *Automatic Speech Recognition* (ASR).

2. Multilayer Perceptron

Multilayer Perceptron (MLP) adalah bagian dari *feedforward Artificial Neural Network* (ANN). MLP terdiri dari setidaknya tiga node layer: input layer, hidden layer, dan output layer. Kecuali node input, setiap node merupakan neuron yang menggunakan fungsi aktivasi nonlinier (Palo et al., 2017). MLP menggunakan teknik *supervised learning* yang disebut *backpropagation* untuk training. Proses *backpropagation* digunakan untuk mengoptimalkan bobot, pertama data diproses melalui bobot acak kemudian keluarannya dibandingkan dengan nilai ground values. Kemudian penurunan *stochastic gradient* digunakan untuk mengoptimalkan nilai bobot. Nilai bobot dioptimalkan sampai tidak bisa lebih rendah yang dikenal sebagai konvergensi.

Peneliti menggunakan model Sequential yang terdiri dari empat layer, input layer, dua hidden layer dan satu output layer. Semua layer memiliki dense yang mana itu adalah layer umum di dalam neural network. Tiga layer pertama mempunyai 256 nodes dan output layer mempunyai kelas intonasi yang telah ditentukan. Peneliti menggunakan fungsi aktivasi ReLU pada tiga layer pertama. Fungsi ReLU sudah terbukti bekerja dengan baik di dalam neural network. Peneliti juga menggunakan 50% dropout pada tiga layer pertama. Ini akan mengecualikan 50% unit secara acak di setiap iterasi saat melatih model agar model tidak overfitted serta meningkatkan generalisasinya. Kemudian, fungsi aktivasi *Softmax* digunakan karena fungsi *Softmax* membuat jumlah keluaran menjadi 1, sehingga dapat menjadikan output sebagai probabilitas tertinggi sebagai prediksi.

Model: "sequential_21"

Layer (type)	Output Shape	Param #
dense_84 (Dense)	(None, 256)	20736
activation_84 (Activation)	(None, 256)	0
dropout_63 (Dropout)	(None, 256)	0
dense_85 (Dense)	(None, 256)	65792
activation_85 (Activation)	(None, 256)	0
dropout_64 (Dropout)	(None, 256)	0
dense_86 (Dense)	(None, 256)	65792
activation_86 (Activation)	(None, 256)	0
dropout_65 (Dropout)	(None, 256)	0
dense_87 (Dense)	(None, 3)	771
activation_87 (Activation)	(None, 3)	0

Total params: 153,091
Trainable params: 153,091
Non-trainable params: 0

Gambar 3. Model Multilayer Perceptron

HASIL DAN PEMBAHASAN

Hasil dari pengujian model MLP dengan ekstraksi fitur MFCC pada Intonasi bahasa Jawa Ponoragan akan ditunjukkan disini. Setelah melakukan pengujian dengan model MLP, peneliti memutuskan untuk menggunakan 250 epochs pada pengujian tersebut. Dan akurasi yang didapatkan pada setiap ekstraksi fitur MFCC adalah sebagai berikut.

Tabel 1. Hasil Ekstraksi MFCC

Trained Features Extraction Results		Feature Extraction		
		MFCC Mean	MFCC Mean, Std	MFCC Mean, Var, Max, Min, Std
Number Of MFCC	40	0.6875	0.375	0.8125
	60	0.5	0.3125	0.4375
	80	0.5	0.5	0.5
	100	0.4375	0.375	0.8125

Dengan gambar di atas peneliti dapat menentukan model dengan ekstraksi fitur terbaik dari hasil berikut. Dengan ini, peneliti menyimpulkan bahwa model MLP dengan ekstraksi fitur MFCC dapat mencapai akurasi hingga 0.8125. Dan model tersebut dapat dijadikan model klasifikasi untuk Intonasi bahasa Jawa Ponoragan.

Berikut adalah proses *Confusion matrix* untuk menganalisis, mengevaluasi kinerja dan akurasi model yang telah dibuat, confusion matrix akan membandingkan hasil prediksi terhadap hasil klasifikasi yang sebenarnya. Hasil dari implentasi algoritma confusion matrix ini dapat kita lihat pada tabel berikut.

Tabel 2. Confusion Matrix

Label	Precision	Recall	F1-score	Support Data
Low	0.96	1.00	0.98	26
Mid	0.93	1.00	0.96	26
High	1.00	0.88	0.94	26
Accuracy				0.96
Macro avg				0.96
Weighted avg				0.96

Deteksi ini menghasilkan nilai akurasi yang cukup tinggi. Dari gambar confusion matrix diatas, deteksi menunjukkan keakuratan pada kelas low dan mid, sedangkan high mempunyai tingkat deteksi lebih rendah dari yang lain. Sehingga, Dari eksperimen deteksi dan klasifikasi didapatkan nilai akurasi mencapai 0.96.

KESIMPULAN

Peneliti dapat memahami model *Multi-layer Perceptron* Neural Network pada klasifikasi Intonasi bahasa Jawa Ponoragan. Di masa mendatang, peneliti dapat mengerjakan deteksi suara secara real-time. Peneliti dapat menambah data untuk mendapatkan lebih banyak kata dalam bahasa Jawa serta dapat bereksperimen dengan model lain seperti *Convolutional Neural Network* (CNN).

DAFTAR PUSTAKA

- Boussard, J., Deveau, A., & Pyron, J. (2017). *Methods for Spoken Language Identification*. <http://cs229.stanford.edu/proj2017/final-reports/5239784.pdf>
- Herti, S., Analisis, A., & Persepsi, U. (2015). ANALISIS UJI PERSEPSI: Intonasi Kalimat Perintah Bahasa Indonesia oleh Penutur Bahasa Jepang. *Tamaddun: Jurnal Kebudayaan Dan Sastra Islam*, 15(1), 149–170.
- Nasution, T. (2012). Metoda Mel Frequency Cepstrum Coefficients (MFCC) untuk Mengenali Ucapan pada Bahasa Indonesia. *SATIN - Sains Dan Teknologi Informasi*, 1, 22–31.
- Palo, H. K., Chandra, M., & Mohanty, M. N. (2017). Emotion recognition using MLP and GMM for Oriya language. *International Journal of Computational Vision and Robotics*, 7(4), 426–442. <https://doi.org/10.1504/IJCVR.2017.084987>
- Sholihah, R. A. (2019). Prosiding Seminar Nasional Forum Linguistik 2019 “ Berbagai Pendekatan Dalam Studi Bahasa .” In Sajarwa & Suhandono (Eds.), *Berbagai Pendekatan Dalam Studi Bahasa* (pp. 439–451). Fakultas Ilmu Budaya UGM Bekerjasama Dengan Departemen Bahasa dan Sastra FIB UGM.