MIND (Multimedia Artificial Intelligent Networking Database) Journal

MIND Journal | ISSN (p): 2528-0015 | ISSN (e): 2528-0902 | Vol. 6 | No. 2 | Halaman 194 - 204
DOI : https://doi.org/10.26760/mindjournal.v6i2.194-204 Desember 2021

Peningkatan Kemampuan Pengenalan Emosi melalui Suara dalam Bahasa Indonesia

FATAN KASYIDI[1], RIDWAN ILYAS[2], NIDA MUTHI ANNISA[3]

[1,2]Program Studi Teknik Informatika, Universitas Jenderal Achmad Yani
[3]Program Studi Psikologi, Universitas Informatika dan Bisnis Indonesia
Email : fatan.kasyidi@lecture.unjani.ac.id

Received 22 Oktober 2021 **|** *Revised* 23 November 2021 **|** *Accepted* 29 November 2021

ABSTRAK

Interaksi manusia dengan komputer merupakan fenomena yang terus berkembang diikuti oleh meningkatnya penggunaan komputer yang sering digunakan dalam ranah sosial manusia. Manusia saling berinteraksi dengan melibatkan emosi untuk memahami seseorang. Emosi manusia seringkali terwakili melalui cara berbicara. Penelitian tentang pengenalan emosi melalui suara telah banyak dilakukan, namun terdapat upaya peningkatan pengenalan emosi melalui suara, terutama masalah korpus yang menjadi salah satu faktor yang menjadikan pengenalan emosi ini belum menghasilkan akurasi pengenalan yang optimal, khususnya berkaitan dengan imbalance data. Penelitian ini dilakukan untuk meningkatkan performa pengenalan emosi untuk mengenali lima kelas emosi yaitu senang, marah, sedih dan kepuasan serta netral menggunakan algoritma boosting. Selain itu, digunakan pula metode seperti CNN dan RNN untuk dapat dilakukan perbandingan serta penerapan SMOTE untuk korpusnya. Setelah eksperimen, dapat dihasilkan akurasi pengenalan mencapai 65% untuk akurasi untuk data tes berdasarkan konfigurasi 22050 Hz sebagai sampling rate, MFCCs dan oversampling SMOTE.

Kata kunci: *Imbalance data, Algoritma Boosting, CNN, RNN, SMOTE*

ABSTRACT

Human interaction with computers are a growing phenomenon followed by the increasing use of computers which are often utilized in human social activities. Humans interact with one another by involving emotions. Plenty of research on speech emotion recognition has been established. Nevertheless, there are still efforts to enhance speech emotion recognition, especially the corpus problem which is one of the factors that the model does not in an optimal performance, especially about imbalance data. This study was conducted to enhance the performance of emotion recognition to recognize five class emotions: happiness, angry, sadness, contentment, and neutral. Furthermore, we employed CNN, RNN, and Boosting Algorithms. Lastly, we applied SMOTE to the corpus. After the experiment, the accuracy reached 65% with 22050 Hz configuration as rate, MFCCs, and SMOTE oversampling.

***Keywords**: Data Imbalance, Boosting Algorithms, CNN, RNN, SMOTE*

1. PENDAHULUAN

Emosi adalah suatu aspek penting dalam kehidupan manusia berupa perasaan dan memiliki banyak pengaruh pada perilaku manusia. Emosi dapat dikatakan sebagai level afektif yang paling mudah dikenali oleh sesama manusia terutama agar dapat saling memahami ketika berkomunikasi. Pada dasarnya ketika ingin memahami lebih jauh terkait dengan emosi tentunya tidak terlepas dari unsur *affect* dan *mood*. Diantara *affect, mood* dan emosi merupakan hal yang saling berkaitan dan tidak dapat terpisahkan **(Russell, 2003)**. Dalam hal ini keadaan emosional yang dialami oleh individu dalam jangka waktu yang singkat dan bersifat tidak menetap dinamakan emosi. Sedangkan *affect* mencakup konsep yang lebih luas dan didalamnya terdapat emosi. Pembahasan terkait emosi biasanya akan mengacu kepada dua teori utama. Teori yang pertama adalah teori *basic emotion* dimana emosi identifikasi emosi berdasarkan karakteristik masing-masing emosi yang yang unik **(Ekman, 1992)**. Teori yang kedua adalah teori *circumplex model of affect* yang menyatakan bahwa emosi inti individu dapat dibagi menjadi dua kuadran yaitu valensi dan *arousal.* Kedua emosi inti (*core affect*) akan merujuk pada kondisi psikologis dan fisiologis individu. Keadaan psikologis yang dirujuk valensi ini terdiri dari emosi positif dan negatif. Keadaan fisiologis yang dirujuk oleh *arousal* dijelaskan dengan keadaan tenang dan bersemangat. Keadaan psikologis yang merupakan interaksi antara valensi dan arousal dijelasakan sebagai emosi **(Russell, 2003)**. Di zaman sekarang ini tentunya dalam kehidupan sehari-hari manusia tidak terlepas dari berbagai hal yang menyangkut teknologi termasuk dalam ranah komunikasi. Emosi sebagai salah satu hal yang penting dalam berkomunikasi pun dapat dikaitkan dengan komputer salah satunya dalam pengembangan pengenalan emosi manusia.

Emosi dapat diekspresikan dengan berbagai cara, salah satunya adalah melalui suara. Suara yang dimaksud merupakan suara yang keluar pada saat percakapan dua orang atau lebih dalam kondisi spontan. Percakapan secara spontan dalam merepresentasikan lebih baik emosi yang diekspresikan sehingga komputer dapat mendeteksi secara natural **(Lubis et al., 2014)**. Penelitian untuk pengembangan model pengenalan emosi melalui suara telah dilakukan, khususnya dalam bahasa Indonesia. Beberapa penelitian yang telah dilakukan tersebut menggunakan korpus yang telah dibangun sebelumnya dan metode yang digunakan untuk pengenalannya diantaranya menggunakan *Naive Bayes*, *Support Vector Machine* dan *Random Forest* **(Kasyidi & Lestari, 2018**). Konstruksi korpus tersebut mengikuti skenario yang telah dilakukan pada penelitian sebelumnya dengan mengambil data dari acara TV *talk show* dalam bahasa indonesia melibatkan empat kelas emosi utama yaitu marah, senang, sedih dan puas serta ditambah dengan kelas emosi netral **(Lubis, dkk., 2014)**. Kemudian korpus tersebut digunakan kembali untuk dilakukan peningkatan kemampuan pengenalan emosi menggunakan *Long Short Term Memory* (LSTM) ditambahkan dengan skenario *sampling* data menggunakan *Random Over Sampler* (ROS). Pada penelitian tersebut, didapatkan model pengenalan yang *overfit*. Hal tersebut dapat terjadi dikarenakan mekanisme *sampling* yang tidak menghasilkan sebaran data yang merata untuk setiap kelas **(Lasiman & Puji Lestari, 2018)**. Pada penelitian lainnya juga terjadi *imbalance data* pada setiap kelasnya sehingga tidak mencapai akurasi pengenalan yang baik **(Kasyidi & Lestari, 2018)**, walaupun disisi lain, dapat diidentifikasi kata-kata yang mewakili emosi tersebut dalam bahasa Indonesia diantaranya kata "selamat", "semoga" dan "cinta" yang menjadi representasi kata yang sering diucapkan dalam kondisi emosi senang.

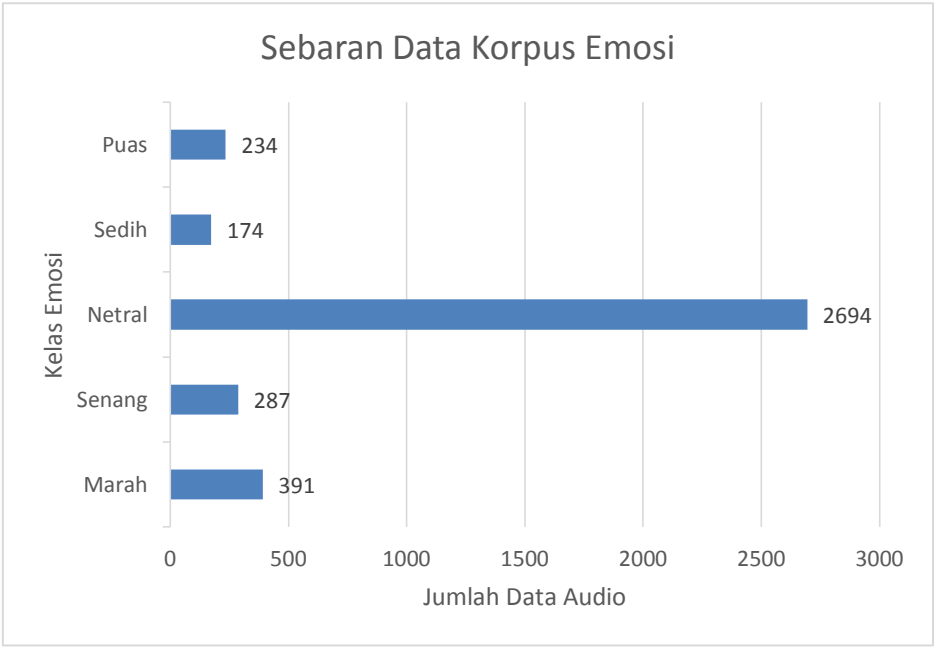
Berdasarkan permasalahan pada penelitian terdahulu, maka dilakukan serangkaian skenario agar dapat meningkatkan performa pengenalan emosi melalui suara dalam bahasa Indonesia. Skenario tersebut lebih fokus kepada penggunaan metode-metode yang dapat

meningkatkan akurasi tanpa mengubah atau menambahkan korpus yang digunakan. Terdapat beberapa cara diantaranya menggunakan teknik algoritma *boosting* untuk meningkatkan *weak learner* menjadi *strong learner* **(Wu, dkk., 2018)**, dilakukan juga perubahan konfigurasi *sampling rate* mulai dari 8000 Hz, 16000 Hz dan 22050 Hz **(Atmaja & Akagi, 2019) (Tarunika, dkk., 2018) (Tzirakis, dkk., 2018) (Umamaheswari & Akila, 2019)** . Ketiga *sampling rate* tersebut digunakan atas beberapa pertimbangan yaitu 8000 Hz merupakan *sampling rate* yang kualitas suaranya setara dengan suara dari telepon. Kemudian 16000 Hz setara dengan kualitas suara dari *microphone* yang sering digunakan untuk merekam suara, sedangkan 22050 Hz merupakan *sampling rate* untuk suara yang keluar dari hasil rekaman studio untuk memproduksi lagu **(Jurafsky & Martin, 2013).** Variasi tersebut digunakan kepada audio yang tersedia dalam korpus. Penerapan koefisien *Mel Frequency Cepstral Coefficients* (MFCCs) dengan tiga variasi yaitu 13, 26 dan 39 koefisien **(Aouani & Ayed, 2018) (Kasyidi & Lestari, 2018) (Winursito, dkk., 2018)**. Untuk mengatasi *imbalance data*, metode *Synthetic Minority Oversampling Technique* (SMOTE) dapat diterapkan sebagai upaya peningkatan akurasi pengenalan emosi **(Sarakit, dkk., 2015)**.

2. METODE PENELITIAN

2.1 Dataset

Dataset yang digunakan adalah korpus yang telah dibangun pada penelitian sebelumnya **(Kasyidi & Lestari, 2018)**. Sebaran data korpus tersebut dapat dilihat pada Gambar 1.



Gambar 1. Sebaran Data Korpus Emosi (Kasyidi & Lestari, 2018)

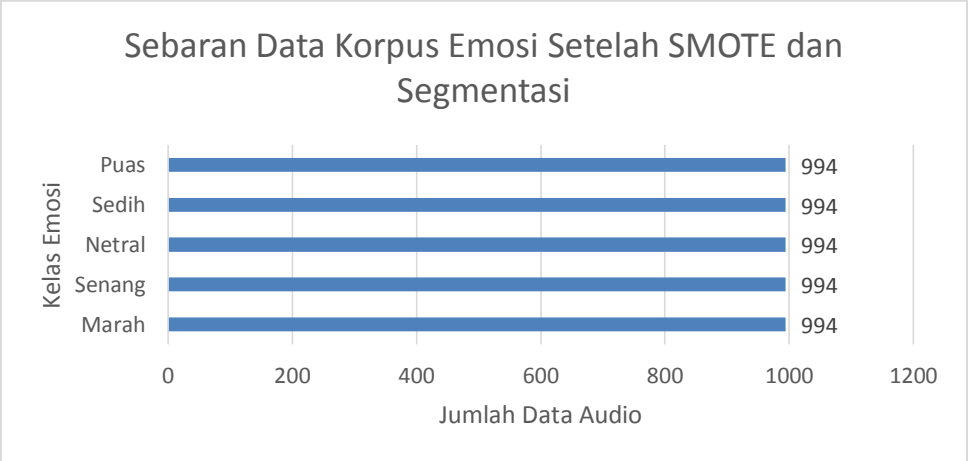
Pada korpus tersebut akan diterapkan skenario *over sampling* menggunakan SMOTE untuk mengatasi *imbalance data*. SMOTE akan membuat sebaran data pada setiap kelas emosi menjadi sama. Sebelum SMOTE diterapkan, setiap data audio disegmentasi ke dalam 5 segmen per satu detik, apabila terdapat data yang lebih dari satu detik, maka tetap disegmentasi hanya satu detik dari total durasi dan tidak menghitung lebihnya. Sebaran dataset setelah SMOTE dapat dilihat pada Gambar 2. Dataset yang telah dilakukan *over*

sampling akan dibagi ke dalam beberapa konfigurasi fitur yaitu jumlah *sampling rate* dan jumlah koefisien MFCCs.

Teknik SMOTE ini menerapkan *over sampling* terhadap kelas yang jumlahnya kecil dengan cara membuat *instance* baru yang disebut *synthetic sample*. *Sample* tersebut dibangkitkan berdasarkan similaritas dari ruang fitur data diantara seluruh kelas minoritas yang ada. *Sample* dipilih menggunakan *K-Nearest Neighbors* (KNN) secara acak dan selanjutnya diterapkan interpolasi lanjar untuk membangkitkan *synthetic sample* yang baru. Misalkan *sample* x_i, kemudian *synthetic sample* baru akan dibangkitkan berdasarkan perhitungan KNN. Dari hasil KNN, *sample* acak x_{zi} dipilih. Kemudian *sample* baru dibangkitkan berdasarkan Persamaan (1) berikut ini **(Rathpisey & Adji, 2019)**.

$$x_{baru} = x_i + \lambda \times (x_{zi} - x_i) \quad (1)$$

Dimana λ merupakan angka acak pada *range* [0,1].



Gambar 2. Sebaran Data Korpus Emosi Setelah SMOTE

Tabel 1. Contoh Dataset Korpus Emosi (Kasyidi & Lestari, 2018)

Nama File	Transkrip	Label Emosi	Keterangan
P004001	dia sekarang lagi suka banget sama patung HI	Senang	ada peningkatan intonasi dan terdengar sangat senang dan antusias ketika bercerita
P004002	tiap kali di rumah dia maunya yang kayak patung HI jadi nyebutnya HI gitu	Senang	intonasi tinggi dan terdengar senang dan antusias bercerita
P004003	setiap kali keluar rumah tuh ngasih lihat kayak ini Que ini namanya patung HI	Senang	intonasi tinggi dan terdengar senang dan antusias bercerita
P004004	jadi setiap kali di rumah begitu baru masuk mobil dia langsung gini HI HI	Senang	intonasi tinggi dan terdengar senang dan antusias bercerita
P004005	jadi kayak oh kok dia beda ya kayak setelah minum ASI tuh memang daya tahan tubuhnya kan jadi kuat banget	Netral	-
P004006	kuat banget gitu jarang sakit <tepuk> terus juga cepat	Puas	kalimat menunjukan emosi tersebut dan intonasi cukup

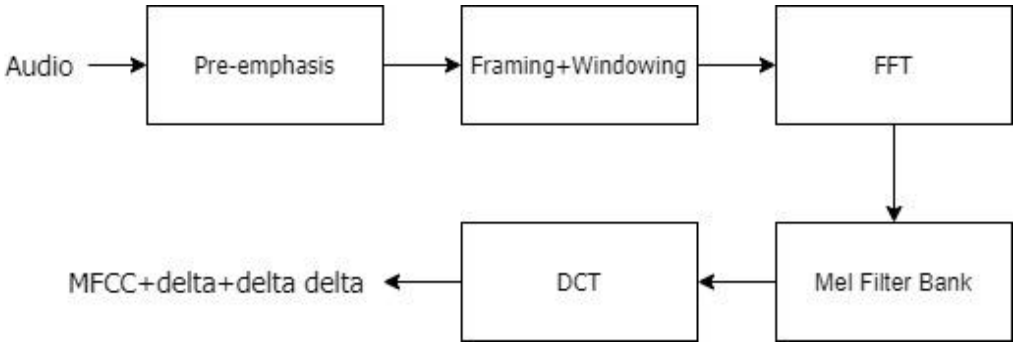
Nama File	Transkrip	Label Emosi	Keterangan
	nangkep apa apa		tinggi
P004007	restoran <eum> <tepuk> jadi kebetulan memang restoran tidak terlalu jauh letaknya dari rumah	Netral	-

Setelah SMOTE diimplementasikan pada korpus. Dilakukan pengelompokkan data menjadi dua kelompok, yaitu data latih dan data uji dengan rasio 75:25. Kemudian, untuk setiap data dilakukan proses segmentasi dengan membaginya menjadi 5 segmen per data audio bertipe wav.

2.2 MFCC

Metode yang digunakan untuk mengekstraksi fitur suara yaitu *Mel Frequency Ceptral Coeffisient* (MFCC). MFCC sangat baik dalam mengekstraksi fitur suara yang merepresentasikan suara manusia maupun musik dan terbukti sangat bermanfaat untuk *speech recognition* **(Winursito, dkk., 2018)**. Pada penelitian ini menggunakan beberapa variasi koefisien MFCC yaitu 13, 26 dan 39. Pada prosesnya, MFCC menghasilkan 13 koefisien MFCC + 13 delta koefisien + 13 delta delta koefisien. Delta berarti melakukan penurunan dari 13 koefisien utama. Penambahan delta koefisien dapat meningkatkan akurasi pengenalan suaranya **(Aouani & Ayed, 2018)**.

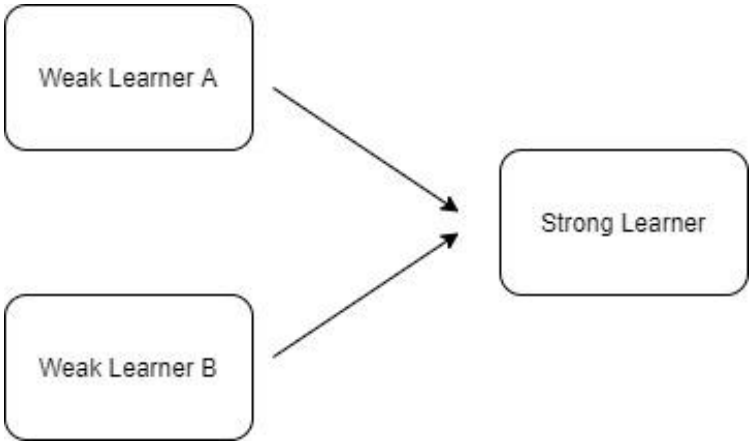
Untuk menghitung koefisien MFCC, metode yang digunakan *Inverse Fourier Transform* (IFT) untuk dimasukkan pada logaritma dari *Fast Fourier Transform* (FFT). Langkah ini selanjutnya diikuti dengan filtering berdasarkan *Mel Scale* **(Aouani & Ayed, 2018)**. Proses mendapatkan seluruh koefisien MFCC dapat dilihat pada Gambar 3.



Gambar 3. Proses Ekstrasi Koefisien MFCC (Aouani & Ayed, 2018)

2.2 Algoritma Boosting

Boosting merupakan salah satu dari *ensemble decision tree* yang bertujuan untuk meningkatkan akurasi dari beberapa *classifier* yang dikategorikan sebagai *weak learner* berubah menjadi *strong learner*. Konsep dari *weak learner* akan menentukan performa dari *strong learner*. Definisi dari *weak learner* merupakan *classifier* yang akurasinya berada disekitar 50% **(Wu, dkk., 2018)**. Alasan digunakannya *weak learner* untuk menyeimbangkan *error rate* dari *weak learner* tersebut. Akibatnya, *learning accuracy* dapat meningkat yang direpresentasikan oleh *strong learner*. Pada penelitian ini, algoritma *boosting* digunakan sebagai upaya meningkatkan akurasi yang menjadi persoalan pada penelitian sebelumnya **(Kasyidi & Lestari, 2018)**.



Gambar 4. Ensemble Learning (Wu dkk., 2018)

2.3 Convolutional Neural Network.

Pada penelitian ini, *Convolutional Neural Network* (CNN) digunakan untuk mengenali emosi melalui suara dengan membangun model CNN. Hal ini melihat performa dari CNN untuk mengenali emosi melalui suara menghasilkan akurasi yang cukup baik pada penelitian lain **(Abdul Qayyum, dkk., 2019)**. Konfigurasi yang digunakan pada CNN ini yaitu menggunakan *Adam optimizer* dengan *learning rate* 0.0001, *dense* 64, aktivasi menggunakan *Rectified Linear Unit* (ReLU) pada *dense layer*, 2D *convolution layer*. Untuk bagian *output layer* menggunakan fungsi *softmax*. Model dilatih sebanyak 30 *epochs* dengan ukuran *batch*=32.

2.4 Long-Short Term Memory

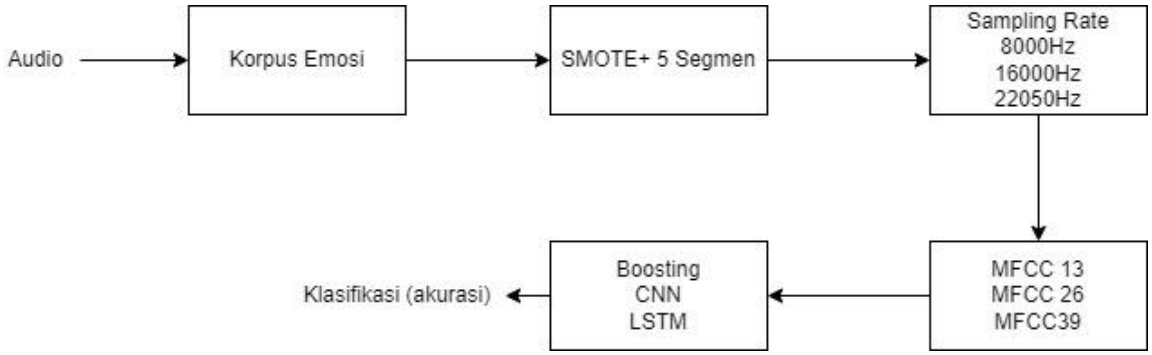
Model lain dibangun menggunakan *Long-Short Term Memory* (LSTM). LSTM dibangun melalui *Recurrent Neural Network* (RNN) dengan tujuan menerima data sekuensial khususnya yang memiliki domain waktu yang erat di mana keterkaitan informasi saat ini dengan sebelumnya tidak hilang dan tetap terhubung. Konfigurasi yang digunakan adalah *learning rate* 0.0001, 256 *layer* pada *dense hidden layer*, dan fungsi aktivasi *softmax* pada *output layer*. Untuk ukuran *batch* sebesar 64 dengan 100 *epochs* pada proses pembangunan model. Beberapa konfigurasi ini merujuk pada penelitian sebelumnya dengan mengganti beberapa nilai sebagai upaya peningkatan akurasi untuk model pengenalnya **(Atmaja & Akagi, 2019)**.

2.5 Skenario Eksperimen

Upaya untuk meningkatkan performansi dari *Speech Emotion Recognition* (SER) dapat dilakukan dengan beberapa cara. Hal tersebut dapat dilihat dari sudut pandang data di mana korpus yang dibangun masih terdapat kekurangan, terutama dalam hal jumlah data pada setiap kelas emosi. SER yang menggunakan bahasa Indonesia sepenuhnya masih terdapat beberapa kendala seperti *classifier*, sebaran data yang kurang baik dan konfigurasi-konfigurasi seperti *sampling rate* dan jumlah koefisien MFCC masih harus diatur dan merujuk ke eksperimen yang telah dilakukan sebelumnya.

Eksperimen yang dilakukan secara umum akan terbagi berdasarkan beberapa aspek yaitu *sampling* data pada korpus, *sampling rate* dan metode pemodelan pengenalan emosi. Pertama, dataset akan melalui proses *over sampling* menggunakan SMOTE. Kedua, dilakukan eksperimen terhadap beberapa konfigurasi *sampling rate* yaitu 8000 Hz, 16000 Hz dan 22050 Hz. Ketiga, konfigurasi MFCCs yang memiliki tiga kelompok koefisien yaitu 13

MFCC, 26 MFCC dan 39 MFCC. Keempat, pemodelan yang diterapkan menggunakan metode seperti *Convolutional Neural Network* (CNN), *Recurrent Neural Network* (RNN) dengan LSTM dan boosting menggunakan *Decision Tree Learning* (DTL). Pada Gambar 5 dapat dilihat skenario lengkap eksperimen yang akan dilakukan.



Gambar 5. Bagan Skenario Eksperimen

2.6 Hasil Eksperimen

Berikut merupakan hasil eksperimen yang telah dilakukan berdasarkan skenario eksperimen yang telah dijelaskan sebelumnya.

Tabel 2. Hasil Eksperimen Dari Korpus Setelah SMOTE

Skenario	Koefisien MFCC	Sampling Rate	Metode	Akurasi		
				Train	Validation	Test
1	13	8000 Hz	Boosting	87 %	90 %	55 %
2	13	16000 Hz	Boosting	89 %	89 %	52 %
3	13	22050 kHz	Boosting	86 %	85 %	63 %
4	26	8000 Hz	Boosting	89 %	92 %	54 %
5	26	16000 Hz	Boosting	92 %	94 %	52 %
6	26	22050 kHz	Boosting	90 %	90 %	**65 %**
7	39	8000 Hz	Boosting	90 %	94 %	50 %
8	39	16000 Hz	Boosting	93 %	92 %	61 %
9	39	22050 kHz	Boosting	91 %	92 %	63 %
10	13	8000 Hz	CNN	41,95 %	41,45 %	37,82 %
11	13	16000 Hz	CNN	52,12 %	39,39 %	-
12	13	22050 kHz	CNN	54,39 %	45,56 %	42,6 %
13	26	8000 Hz	CNN	52,48 %	41,23 %	40,74 %

Skenario	Koefisien MFCC	Sampling Rate	Metode	Akurasi		
				Train	Validation	Test
14	26	16000 Hz	CNN	64,70 %	45,89 %	44,96 %
15	26	22050 kHz	CNN	71,77 %	48,48 %	47,72 %
16	39	8000 Hz	CNN	57,43 %	41,67 %	42,43 %
17	39	16000 Hz	CNN	71,28 %	48,59 %	46,33 %
18	39	22050 kHz	CNN	78,58 %	47,73 %	49,22 %
19	13	8000 Hz	RNN-LSTM	70,68 %	21,65 %	15,01 %
20	13	16000 Hz	RNN-LSTM	75,19 %	24,13 %	17,02 %
21	13	22050 kHz	RNN-LSTM	73,62 %	25,22 %	21,30 %
22	26	8000 Hz	RNN-LSTM	70,14 %	28,14 %	16,43 %
23	26	16000 Hz	RNN-LSTM	86,90 %	30,09 %	18,32 %
24	26	22050 kHz	RNN-LSTM	87,02 %	27,71 %	21,23 %
25	39	8000 Hz	RNN-LSTM	91,67 %	24,57 %	21,31 %
26	39	16000 Hz	RNN-LSTM	91,75 %	24,46 %	27,36 %
27	39	22050 kHz	RNN-LSTM	90,99 %	26,52 %	23,96 %

Berdasarkan hasil eksperimen yang telah dilakukan, terlihat bahwa akurasi model pengenalan emosi melalui suara dapat meningkat dengan mengatasi data *imbalance*, menentukan *sampling rate* dan koefisien MFCC yang tepat serta metode *machine learning* yang digunakan. Terlihat bahwa LSTM dan CNN tidak terlalu signifikan meningkatkan akurasi dikarenakan faktor *overfitting* **(Lasiman & Puji Lestari, 2018)**. Oleh karena itu, metode *boosting* dipilih agar dapat meningkatkan akurasi sehingga hasilnya terlihat menunjukan peningkatan sesuai pada Tabel 2. Peningkatan akurasi tersebut disebabkan oleh kemampuan *boosting* yang dapat mengatasi permasalahan *overfitting* **(Wu, dkk., 2018)** yang terjadi pada metode lain yang digunakan pada tahap eksperimen.

Beberapa pertimbangan lain untuk meningkatkan akurasi adalah menggunakan koefisien MFCC yang lebih banyak sesuai dengan yang telah dijelaskan sebelumnya. Hal tersebut dilakukan karena dapat meningkatkan akurasi pengenalan emosi **(Hadjadji, dkk., 2019) (Winursito, dkk., 2018)**. Seperti yang terlihat pada tabel 1, peningkatan akurasi berbanding lurus dengan penambahan koefisien MFCC. Pertimbangan terakhir adalah dengan menggunakan tiga konfigurasi *sampling rate*. *Sampling rate* dapat mempengaruhi seberapa baik fitur dapat diekstraksi **(Hokking, dkk., 2016)**.

Jika melihat hasil eksperimen pada beberapa skenario seperti skenario pertama, keempat, kelima, ketujuh dan kesembilan menunjukkan akurasi pada validasi yang lebih tinggi daripada akurasi pelatihan. Hal tersebut dapat terjadi dikarenakan *weighted majority voting* yang menghasilkan pilihan model yang akurasinya dapat lebih tinggi dari model pelatihannya

(Wu, dkk., 2018). Kemudian hal tersebut pula dapat dipengaruhi bias yang tinggi yang mengakibatkan berkurangnya fleksibilitas dan kemampuan belajar dari *metode boosting* itu sendiri terhadap data pelatihan.

3. KESIMPULAN

Berdasarkan eksperimen yang telah dilakukan, upaya peningkatan kemampuan pengenalan emosi melalui suara dalam bahasa Indonesia menghasilkan kenaikan menjadi 65 % daripada penelitian sebelumnya yang menerapkan SMOTE sebagai teknik *over sampling* **(Kasyidi & Lestari, 2018) (Lasiman & Puji Lestari, 2018)**. Akurasi tersebut mencakup semua kelas emosi, dikarenakan akurasi tersebut didapatkan berdasarkan *confusion matrix* yang merepresentasikan *predicted* label. Peningkatan tersebut didapatkan dari penentuan konfigurasi lain pada *sampling rate*, penambahan koefisien MFCC dan penerapan *boosting*. Meskipun begitu, permasalahan *overfitting* masih terjadi, sehingga diperlukan strategi lain. Hal ini akan ditangani kedepannya.

DAFTAR RUJUKAN

Abdul Qayyum, A. B., Arefeen, A., & Shahnaz, C. (2019). Convolutional Neural Network (CNN) Based Speech-Emotion Recognition. *2019 IEEE International Conference on Signal Processing, Information, Communication Systems (SPICSCON)*, (pp. 122–125). https://doi.org/10.1109/SPICSCON48833.2019.9065172

Aouani, H., & Ayed, Y. B. (2018). Emotion recognition in speech using MFCC with SVM, DSVM and auto-encoder. *2018 4th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, (pp. 1–5). https://doi.org/10.1109/ATSIP.2018.8364518

Atmaja, B. T., & Akagi, M. (2019). Speech Emotion Recognition Based on Speech Segment Using LSTM with Attention Model. *2019 IEEE International Conference on Signals and Systems (ICSigSys)*, (pp. 40–44). https://doi.org/10.1109/ICSIGSYS.2019.8811080

Ekman, P. (1992). An Argument for Basic Emotions. *Cognition and Emotion*, *6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Hadjadji, I., Falek, L., Demri, L., & Teffahi, H. (2019). Emotion recognition in Arabic speech. *2019 International Conference on Advanced Electrical Engineering (ICAEE)*, (pp. 1–5). https://doi.org/10.1109/ICAEE47123.2019.9014809

Hokking, R., Woraratpanya, K., & Kuroki, Y. (2016). Speech recognition of different sampling rates using fractal code descriptor. *2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, (pp. 1–5). https://doi.org/10.1109/JCSSE.2016.7748895

Jurafsky, D., & Martin, J. H. (2013). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition, 2/E.* (Second). Pearson Education.

Kasyidi, F., & Lestari, D. P. (2018). Identification of four class emotion from Indonesian spoken language using acoustic and lexical features. *Journal of Physics: Conference Series*, *971*, 012048. https://doi.org/10.1088/1742-6596/971/1/012048

Lasiman, J. J., & Puji Lestari, D. (2018). Speech Emotion Recognition for Indonesian Language Using Long Short-Term Memory. *2018 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*, (pp. 40–43). https://doi.org/10.1109/IC3INA.2018.8629525

Lubis, N., Lestari, D., Purwarianti, A., Sakti, S., & Nakamura, S. (2014). Emotion recognition on Indonesian television talk shows. *2014 IEEE Spoken Language Technology Workshop (SLT)*, (pp. 466–471). https://doi.org/10.1109/SLT.2014.7078619

Rathpisey, H., & Adji, T. B. (2019). Handling Imbalance Issue in Hate Speech Classification using Sampling-based Methods. *2019 5th International Conference on Science in Information Technology (ICSITech)*, (pp. 193–198). https://doi.org/10.1109/ICSITech46713.2019.8987500

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172. https://doi.org/10.1037/0033-295x.110.1.145

Sarakit, P., Theeramunkong, T., & Haruechaiyasak, C. (2015). Improving emotion classification in imbalanced YouTube dataset using SMOTE algorithm. *2015 2nd International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA)*, (pp. 1–5). https://doi.org/10.1109/ICAICTA.2015.7335373

Tarunika, K., Pradeeba, R. B., & Aruna, P. (2018). Applying Machine Learning Techniques for Speech Emotion Recognition. *2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, (pp. 1–5). https://doi.org/10.1109/ICCCNT.2018.8494104

Tzirakis, P., Zhang, J., & Schuller, B. W. (2018). End-to-End Speech Emotion Recognition Using Deep Neural Networks. *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (pp. 5089–5093). https://doi.org/10.1109/ICASSP.2018.8462677

Umamaheswari, J., & Akila, A. (2019). An Enhanced Human Speech Emotion Recognition Using Hybrid of PRNN and KNN. *2019 International Conference on Machine Learning,*

Big Data, Cloud and Parallel Computing (COMITCon), (pp. 177–183). https://doi.org/10.1109/COMITCon.2019.8862221

Winursito, A., Hidayat, R., & Bejo, A. (2018). Improvement of MFCC feature extraction accuracy using PCA in Indonesian speech recognition. *2018 International Conference on Information and Communications Technology (ICOIACT)*, (pp. 379–383). https://doi.org/10.1109/ICOIACT.2018.8350748

Wu, Y., Mao, J., & Li, W. (2018). Predication of Futures Market by Using Boosting Algorithm. *2018 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, (pp. 1–4). https://doi.org/10.1109/WiSPNET.2018.8538586