



Klasifikasi Dialek Pengujar Bahasa Inggris Menggunakan Random Forest

Muhamad Azhar^{1,*}, Hilman F Pardede^{1,2}

¹ Ilmu Komputer, Ilmu Komputer, STMIK Nusa Mandiri, Jakarta, Indonesia

² Pusat Penelitian Informatika, Lembaga Ilmu Pengetahuan Indonesia, Bandung, Indonesia

Email: ^{1,*}14002310@nusamandiri.ac.id, ²hilman@nusamandiri.ac.id

Email Penulis Korespondensi: 14002310@nusamandiri.ac.id

Abstrak—Pengenalan suara merupakan salah satu bidang riset yang cukup penting yang dimana saat ini banyak digunakan secara luas untuk digunakan berbagai aplikasi. Akan tetapi, kinerja pengenalan suara dipengaruhi oleh dialek pengujar. Oleh karena itu, pengenalan dialek sering digunakan sebagai fitur tambahan pada pengenalan suara. Proses pengenalan dialek tidaklah mudah. Saat ini, teknologi *Machine Learning* banyak diterapkan dalam pengenalan dialek. Salah satu tantangan dalam pengenalan dialek berbasis *machine learning* adalah ketidakseimbangan kelas dan tumpang tindih dalam banyak nya ragam pada tehnik klasifikasi. Penelitian ini menerapkan teknologi oversampling berbasis Random Forest untuk pengenalan dialek. Untuk optimasi *hyper-parameter* dari algoritma random forest, kami menerapkan *Grid Search*. Percobaan pada data *Speech Accent Archive* menggunakan fitur MFCC menghasilkan akurasi 0.91 dan AUC 0.95.

Kata Kunci: Pengenalan Suara; *Imbalance*; Klasifikasi; MFCC; *Random Forest*

Abstract—Speech recognition is one of the important research fields which is currently widely used for various applications. However, speech recognition performance is affected by the dialect of the speaker. Therefore, dialect recognition is often used as an additional feature in speech recognition. The process of recognizing dialects is not easy. Currently, Machine Learning technology is widely applied in dialect recognition. One of the challenges in the introduction of machine learning-based dialects is the imbalance of classes and overlaps in a wide variety of classification techniques. This study applies Random Forest-based oversampling technology for dialect recognition. For hyper-parameter optimization of the random forest algorithm, we apply the Grid Search method. Experiments on Speech Accent Archive data using the MFCC feature resulted in an accuracy of 0.91 and AUC of 0.95.

Keywords: Dialect recognition; Speech Recognition; Imbalance Class; MFCC; Random Forest

1. PENDAHULUAN

Pengenalan suara merupakan salah satu bidang riset atau penelitian yang cukup penting dimana saat ini sudah banyak digunakan secara luas untuk keperluan berbagai aplikasi [1]. Pengenalan suara itu sendiri merupakan suatu pengembangan teknik untuk mengidentifikasi dan menterjemahkan sumber suara yang diucapkan melalui perangkat teknologi [2] kedalam bentuk teks. Hasil konversi suara ke dalam bentuk teks dapat dimanfaatkan untuk berbagai aplikasi [3], misalnya sistem diktasi, sistem antar-muka elektronik diantaranya untuk menyalakan lampu menggunakan perintah suara, absensi dan *request* lagu. Selain itu, identifikasi dari kata yang diucapkan dapat digunakan untuk menyimpan pembicaraan atau kalimat dalam bentuk teks ataupun dikembangkan untuk robot sahabat [4] dan penelitian yang memanfaatkan pengenalan suara juga antara lain seperti identifikasi nyamuk *culex* dan *aedes aegypti*, dimana suara yang dihasilkan dari kepanakan nyamuk memiliki ciri tersendiri yang menggambarkan dari jenis nyamuk tersebut dan jenis kelaminnya [5], dan masih banyak lagi pemanfaatan pengenalan suara pada bidang dan penelitian lainnya.

Dalam proses pengenalan suara dibutuhkan suatu metode untuk ekstraksi fitur dari suara tersebut, hasil dari ekstraksi tersebut yang nantinya akan diproses dan dilakukan pencocokan dengan pemodelan tertentu. Ekstraksi fitur adalah tahapan pengambilan karakter dari sebuah objek sehingga dapat mendeskripsikan identitas dari objek tersebut [6]. Dan pendekatan ekstraksi fitur suara dapat didefinisikan juga sebagai tahapan mengubah sinyal ucapan menjadi urutan kecil vektor fitur, vektor ini membawa karakteristik atau ciri penting dari sinyal ucapan tertentu untuk mengidentifikasikan antara pembicara dengan suara mereka [7].

Mel-frequency cepstral coefficients (MFCC) adalah salah satu fitur yang paling sering dipakai pada pengenalan suara. Fitur ini diperkenalkan pertama kali tahun 1980 [8]. MFCC merupakan metode ekstraksi fitur suara yang didasarkan dengan prinsip karakteristik pendengaran manusia [9] yang ternyata tidak linear. pada MFCC sinyal suara diekspresikan dalam skala Mel, dimana filter linear digunakan pada frekuensi dibawah 1000 Hz dan filter logaritmik digunakan pada frekuensi di atas 1000 Hz.

Namun, kinerja sistem pengenalan suara sangat dipengaruhi dialek pengujarnya. Sistem yang dilatih dengan pengujar asli (*native speaker*) mungkin tidak akan terlalu baik jika digunakan oleh pengujar yang *non-native*. Hal ini dikarenakan adanya perbedaan dialek antar pengujar. Oleh karena itu, sistem pengenalan dialek pengujar sering digunakan sebagai fitur tambahan pengenalan suara. Dengan mengetahui dialek pengujar, sistem dapat diadaptasi terhadap dialek pengujar.

Saat ini *Machine Learning* menjadi solusi untuk pengenalan dialek. Namun, proses identifikasi pengenalan suara khususnya pengenalan dialek tidaklah mudah, hal ini disebabkan oleh bagaimana pembicara dalam melakukan penuturan dialek. Perbedaan ini umum terjadi pada bukan penutur bahasa asli terlepas dari penutur bahasa asli (misal, kecepatan berbicara), dan ada beberapa bukti bahwa ucapan bukan penutur bahasa asli mungkin



lebih bervariasi baik di dalam maupun diluar penutur bahasa asli [10]. Dan juga proses klasifikasi menjadi sulit karena adanya keragaman dan ketidakseimbangan kelas.

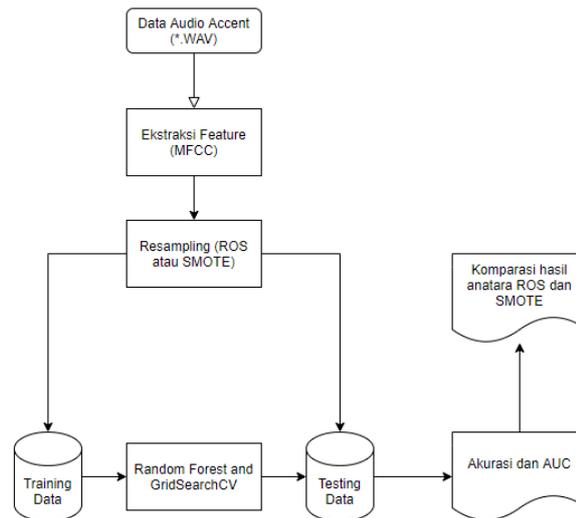
Fokus penelitian ini adalah mengatasi ketidakseimbangan kelas pada sistem pengenalan dialek. Secara teknis, kumpulan suatu data dikatakan tidak seimbang jika distribusi antar kelas di dalam dataset tidak merata atau seragam. Dalam kondisi seperti ini satu kelas dataset yang digambarkan hanya oleh sejumlah kecil contoh atau kelas minoritas dan kelas lain membentuk sebagian besar data atau kelas mayoritas [13]. Akibatnya, sering hasil prediksi cenderung kepada dataset yang mayoritas. Oleh karena itu, perlunya suatu metode komputasi yang dapat menjawab beberapa permasalahan tersebut. Hal ini tentu dapat mempengaruhi kinerja pemodelan analisis klasifikasi [12]. Sehingga perlu dilakukan penyeimbangan jumlah data terutama untuk kelas minoritas sehingga kelas minoritas tidak menjadi terabaikan dalam klasifikasi [14].

Solusi untuk mengatasi ketidakseimbangan dapat dikategorikan menjadi beberapa kelompok. Pertama adalah dengan cara pendekatan level data (*Sampling*) yaitu pengambilan ulang sampel data untuk mengubah data yang tidak seimbang menjadi seimbang. Pendekatan level data (*Sampling*) dapat digunakan untuk modifikasi distribusi kelas dari data latih untuk menyeimbangkan data [13], pendekatan level data itu sendiri adalah tahapan *preprocessing* yang dilakukan sebelum membuat pemodelan machine learning [12]. Pendekatan kedua adalah dengan cara penyesuaian *Cost Sensitive* pada data aslinya [13], *Cost Sensitive* merupakan pembelajaran machine learning dalam mempertimbangkan kesalahan klasifikasi [13].

Pada penelitian ini, kami mengusulkan Random Forest (RF) untuk sistem pengenalan dialek dengan menggunakan *Random Over Sampling* (ROS) dan SMOTE untuk mengatasi ketidakseimbangan data. Untuk mencari *hyper-parameter* terbaik pada algoritma RF, kami menggunakan *Grid Search*. Penelitian ini menggunakan dataset *speech accent* yang dapat diunduh pada repository <http://accent.gmu.edu/> (Diakses pada 13 Oktober 2020).

2. METODOLOGI PENELITIAN

Gambar 1 adalah diagram metode penelitian ini. Metode penelitian meliputi penerapan fitur ekstraksi MFCC, algoritma pendekatan level data atau resampling dengan ROS atau SMOTE dengan algoritma Random Forest. Dimana hasil akhir akan melakukan perbandingan algoritma resampling untuk hasil pengujian dan analisis terhadap model prediksi yang memiliki tingkat akurasi dan AUC yang paling tinggi atau terbaik dalam mengenal dialek pembicara.



Gambar 1. Kerangka Pemikiran

Sesuai dari pemodelan yang diusulkan yakni pendekatan level data (*Resampling*) dengan melakukan perbandingan dua buah algoritma yaitu ROS dibandingkan dengan SMOTE. Pendekatan level data tersebut dimaksudkan untuk menyeimbangkan kelas pada dataset dialek berbicara kalimat bahasa Inggris, yang pada umumnya bersifat tidak ada keseimbangan data.

Untuk hasil pemodelan yang digunakan adalah menggunakan *Random Forest* dan dengan melakukan *Hyperparameter* menggunakan *GridSearchCV* dan evaluasi metrik yang digunakan adalah nilai akurasi dan *AUC*.

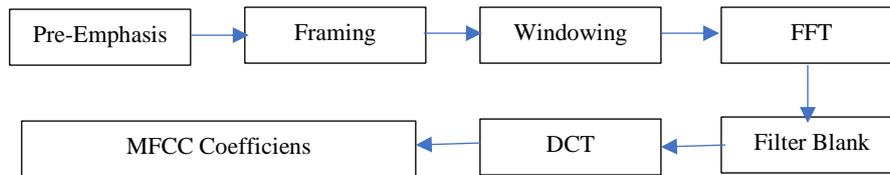
2.1 MFCC

Gambar 2 adalah proses ekstraksi MFCC. Untuk ekstraksi MFCC meliputi pre-emphasis, framing, windowing, fast Fourier Transform (FFT), filter-bank, dan discrete cosine transform (DCT). Pre-Emphasis adalah proses meningkatkan energi spektrum frekuensi tinggi dengan melakukan operasi sebagai berikut: menyaring sinyal suara agar membuat sinyal tersebut menjadi lebih jelas saat melanjutkan ke tahap berikut nya:



$$s'_n = s_n - as_{n-1} \tag{1}$$

Dimana, s'_n adalah hasil sinyal dari proses Pre-Emphasis, s_n adalah inputan sinyal yang belum Pre-Emphasis pada waktu n , a adalah konstanta dengan nilai 0.97 dan s_{n-1} adalah sinyal input sebelum sinyal s_n .



Gambar 2. Proses Ekstraksi MFCC

Setelah pre-emphasis, sinyal suara dipotong menjadi beberapa frame dimana setiap frame N sampel sinyal yang saling berdekatan [17]. Panjang setiap frame pada umumnya sangat singkat berkisaran antara 20 – 40 ms. Keuntungan dari penggunaan framing adalah memudahkan analisis dan mengurangi alokasi atau penggunaan memori [18]. Setelah itu, operasi windowing dilakukan untuk setiap frame. Teknik yang umum digunakan pada windowing adalah *hamming window* $H(n)$ pada waktu n , dimana perumusannya seperti pada persamaan (2) berikut [18]:

$$H(n) = 0,54 - 0,46 \cos\left(2\pi \frac{n}{N-1}\right) \tag{2}$$

Dimana N adalah jumlah sampel tiap frame. Setelah itu, sinyal suara akan dikonversi ke domain frekuensi menggunakan fast Fourier transform (FFT). Lalu, skala mel dengan persamaan berikut [3] diterapkan untuk memperoleh skala frekuensi mel dari sinyal suara:

$$Mel(f) = 2595 \left(\log\left(1 + \frac{f}{700}\right) \right) \tag{3}$$

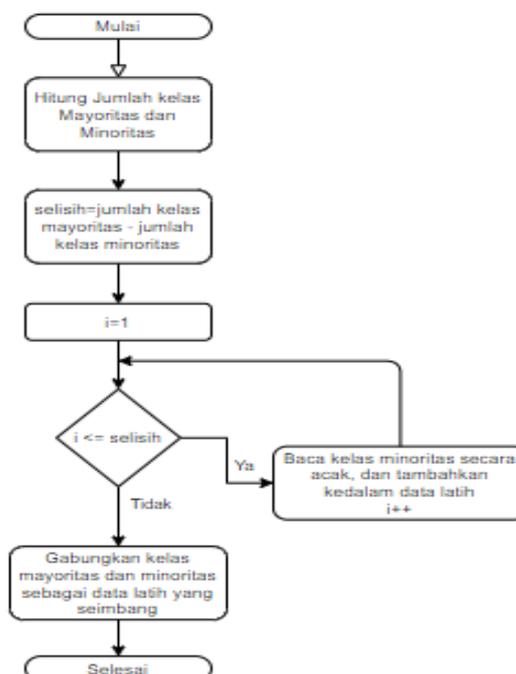
Dimana f adalah frekuensi linear hasil FFT. Lalu setiap spektrum mel yang diperoleh akan dikonversi ke domain cepstrum menggunakan DCT dengan formulasi [18]:

$$C_n = \sum_{k=1}^k (f'_k) \cos\left[n(k-0,5)\frac{\pi}{k}\right] \tag{4}$$

Dimana k adalah jumlah *mel frequency filter blank*, f'_k adalah hasil dari transformasi non linear.

2.2 Resampling

Metode *Oversampling* ini dilakukan dengan cara menambahkan sample kelas minoritas sampai kepada sekumpulan data yang tidak seimbang. Berikut Gambar 2 flowchart untuk *Oversampling* [17] :



Gambar 3. Flowchart Oversampling

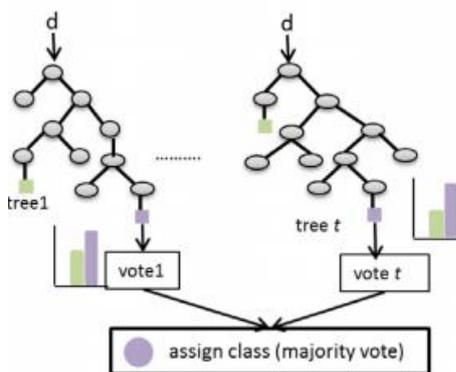


Gambar 3 diatas menjelaskan tahapan-tahapan bagaimana algoritma *Oversampling* bekerja, sehingga kelas yang tadinya tidak seimbang menjadi seimbang. Dan pada penelitian ini akan menggunakan dua algoritma *Oversampling* yang akan saling dibandingkan yaitu *ROS* dan *SMOTE*.

Pada algoritma *resample SMOTE (Synthetic Minority Oversampling Technique)* konfigurasi yang digunakan adalah *random_state* bernilai 0, *k_neighbors* bernilai 2, *n_jobs* bernilai 1 dan *sampling_strategy* menggunakan *not majority*. Sedangkan pada algoritma *ROS (Random Over-Sampling)* konfigurasi yang digunakan adalah *random_state* bernilai 0 dan *sampling_strategy* menggunakan *not majority* sehingga *ratio* data yang dibawah data *major* (kelas terbanyak) akan mengalami *resample* hingga mencapai nilai maksimum atau menstarakan jumlah kelas terbanyak dari dataset tersebut.

2.3 Random Forest dan Grid Search

Random Forest (RF) adalah sebuah algoritma *Ensemble Classifier* yang terdiri dari banyak pohon keputusan (*Deccision Trees*), dimana kelas prediksi akhir untuk contoh pengujian didapatkan dari penggabungan prediksi semua pohon individu [18], seperti ilustrasi pada Gambar 4 berikut:



Gambar 4. Random Forest[19]

RF forest bekerja dengan membangun banyak pohon keputusan dimana setiap pohon keputusan dibentuk dari sampling data dan fitur keseluruhan. Dari semua pohon yang terbentuk akan dilakukan voting untuk mencari kelas dari data yang diberikan [18]. Untuk mendapatkan setting parameter terbaik, dapat dilakukan dengan melakukan variasi *hyperparameter* seperti jumlah estimator, kedalaman tree, kriteria untuk memilih fitur dan lain-lain. Oleh karena itu, untuk membantu mencari *hyper-parameter* terbaik, kami menggunakan Grid search.

Grid Search adalah sebuah algoritma yang akan mencari kombinasi hyper-parameter yang menghasilkan performa terbaik ketika diberi kandidat pilihan pencarian parameter-parameter tersebut. Di dalam *Grid Search* biasanya didefinisikan kamus atau *dictionary* untuk menyimpan parameter yang perlu dicari terlebih dahulu dan kemudian algoritma ini akan melakukan perhitungan model berdasarkan semua parameter yang tersimpan tersebut dan mendapatkan parameter yang terbaik dan mendapatkan performa validasi silang terbaik pada semua dataset latihan [20].

3. HASIL DAN PEBAHASAN

3.1 Dataset

Jenis data yang digunakan adalah data sekunder dengan mengunduh data dari situs <https://accent.gmu.edu/> pada tanggal 13 Oktober 2020 dataset ini juga digunakan pada penelitian [15], dan dataset ini berupa file audio berekstensi MP3 yang kemudian di konversi menjadi WAV dan jumlah dataset yang digunakan sebanyak 1231 yang terdiri dari 5 negara, dapat dilihat pada Tabel 1 berikut:

Tabel 1. Dataset yang digunakan

| Bahasa-Dialek | Jumlah Data Sampel |
|---------------|--------------------|
| Inggris | 627 |
| Spanyol | 220 |
| Arab | 172 |
| Mandarin | 132 |
| Perancis | 80 |

Pada Tabel 1 diatas menggambarkan jumlah penyebaran data set yang di gunakan pada penelitian ini, dimana Inggris sebagai kelas mayoritas dan Spanyol, Arab, Mandarin dan Perancis sebagai minoritas.



3.2 Setting Grid Search

Masing – masing eksperimen dilakukan konfigurasi parameter pada algoritma *Grid Search* seperti pada tabel 2 dan 3 berikut:

Tabel 2. Konfigurasi GridSearchCV

| Parameter | Nilai |
|------------------|---|
| Estimator | <i>Random Forest</i> dengan nilai <i>Random State</i> 0 |
| Param_grid | Tabel 4 |
| n_jobs | -1 |
| Cross Validation | 10 |

Pada Tabel 2 menggambarkan detail konfigurasi pada algoritma *Grid Search*, konfigurasi pada *GridSearch* menggunakan estimator *Random Forest*, parameter yang digunakan sesuai pada Tabel 3, n_jobs bernilai -1. Untuk *Grid Search*, k-fold cross validation dengan k=10 diterapkan untuk mencari performa terbaik.

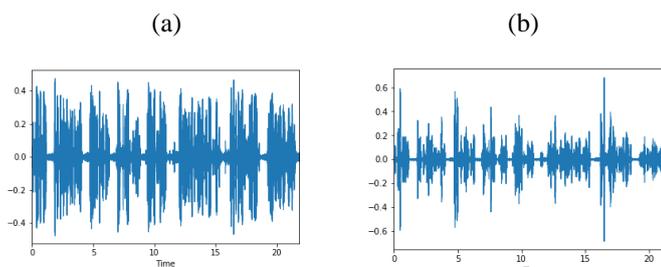
Tabel 3. Parameter yang diujikan

| Parameter | Nilai |
|--------------|--------------------|
| Estimators | [100, 200] |
| Max Features | [auto, sqrt, log2] |
| Max Depth | [4, 5, 6, 7, 8] |
| Criterion | [gini, entropy] |

Evaluasi digunakan untuk menampilkan hasil kinerja dari eksperimen pemodelan yang diusulkan. Evaluasi di sajikan dalam bentuk tabel dan grafik berdasarkan variabel dari pengujian seperti akurasi, AUC, F1-Score, Precision dan recall.

3.3 Data Audio Hasil Ekstraksi Fitur MFCC

Adapun ilustrasi dari perbedaan sebelum dan sesudah data sinyal dilakukan preemphasis dapat dilihat pada Gambar 5 berikut ini :



Gambar 5. Waveplot sebelum dan sesudah preemphasis

Pada Gambar 5 menunjukkan bagian (a) sinyal suara yang masuk dan belum dilakukan *preemphasis*, sedangkan pada Gambar 5 bagian (b) menunjukkan hasil sinyal suara yang sudah dilakukan *preemphasis*. Dan proses Fitur Ekstraksi pada penelitian ini menggunakan MFCC, contoh hasil fitur ekstraksi pada dataset English 1.wav tersebut dapat dilihat pada Tabel 4:

Tabel 4. Hasil MFCC

| Koefisien | Nilai Koefisien |
|-----------|---------------------|
| MFCC1 | -240.09072875976562 |
| MFCC2 | 82.25692749023438 |
| MFCC3 | -6.231974124908447 |
| MFCC4 | 26.523008346557617 |
| MFCC5 | 12.450281143188477 |
| MFCC6 | 11.530415534973145 |
| MFCC7 | -9.661667823791504 |
| MFCC8 | 6.894360065460205 |
| MFCC9 | -16.78693389892578 |
| MFCC10 | -2.663843870162964 |
| MFCC11 | -4.380803108215332 |
| MFCC12 | 9.8905611038208 |
| MFCC13 | -8.933212280273438 |



Pada Tabel 4 merupakan hasil ekstraksi fitur MFCC yang di ekstrak menjadi 13 koefisien dalam 1 file audio yakni file *English1.wav*.

3.4 Hasil resampling dengan ROS ataupun SMOTE

Adapun hasil yang didapatkan untuk rasio dataset setelah di lakukan resampling baik dengan ROS atau SMOTE, adalah sebagai berikut:

Tabel 5. Hasil Resampling data dengan SMOTE ataupun ROS

| Bahasa-Dialek | Data Sample Asli | Data Resampling |
|---------------|------------------|-----------------|
| Inggris | 627 | 627 |
| Spanyol | 220 | 627 |
| Arab | 172 | 627 |
| Mandarin | 132 | 627 |
| Perancis | 80 | 627 |

Dari tabel 5 diatas menggambarkan hasil *resampling data* dari pendekatan level data dengan SMOTE ataupun ROS dimana kelas mayoritas ada pada bahasa atau dialek Inggris, sedangkan yang lainnya seperti Spanyol, Arab, Mandarin dan Perancis mengalami *resampling data* sehingga jumlah data setelah *resampling* sama dengan jumlah data mayoritas yakni 627 yang dimiliki oleh Inggris .

3.5 Hasil Data latih sebelum resampling dan sesudah resampling.

Parameter-parameter algoritma *Random Forest* yang diujikan menggunakan *Grid Search* dan menghasilkan parameter terbaik untuk ketiga eksperimen tersebut yang dapat dilihat pada tabel 6,7 dan 8 berikut ini:

Tabel 6. Parameter terbaik yang digunakan pemodelan random forest tanpa resampling

| Parameter | Nilai |
|--------------|-------|
| Estimators | 200 |
| Max Features | Auto |
| Max Depth | 8 |
| Criterion | gini |

Pada Tabel 6 menjelaskan parameter terbaik yang dapat digunakan untuk algoritma *Random Forest* untuk melakukan klasifikasi pada data uji terhadap data latih, dengan konfigurasi estimator bernilai 200, max features auto, max depth 8 dan *criterion* gini pada algoritma *Random Forest* mendapatkan nilai akurasi sebesar 0.51 untuk data latih.

Tabel 7. Parameter terbaik yang digunakan pemodelan random forest dengan SMOTE

| Parameter | Nilai |
|--------------|---------|
| Estimators | 200 |
| Max Features | Auto |
| Max Depth | 8 |
| Criterion | entropy |

Pada Tabel 7 menjelaskan parameter terbaik yang dapat digunakan untuk algoritma *Random Forest* untuk melakukan klasifikasi pada data uji terhadap data latih, dengan konfigurasi estimator bernilai 200, max features auto, max depth 8 dan *criterion* entropy pada algoritma *Random Forest* mendapatkan nilai akurasi sebesar 0.77 untuk data latih.

Tabel 8. Parameter terbaik yang digunakan untuk pemodelan random forest dengan ROS

| Parameter | Nilai |
|--------------|---------|
| Estimators | 200 |
| Max Features | Auto |
| Max Depth | 8 |
| Criterion | entropy |

Pada tabel 8 menjelaskan parameter terbaik yang dapat digunakan untuk algoritma *Random Forest* untuk melakukan klasifikasi pada data uji terhadap data latih, dengan konfigurasi estimator bernilai 200, max features auto, max depth 8 dan *criterion* entropy pada algoritma *Random Forest* mendapatkan nilai akurasi sebesar 0.87 untuk data latih.

4.4 Pengujian Model

Eksperimen dilakukan dengan melakukan pengujian model usulan terhadap dataset yang digunakan. Model yang diuji meliputi penerapan model klasifikasi *Random Forest Hyperparameter Grid Search* tidak menggunakan



proses *resampling* data, penerapan model klasifikasi *Random Forest* dengan pendekatan level data *SMOTE* dan *Hyperparameter Grid Search* dan yang terakhir penerapan model klasifikasi *Random Forest* dengan pendekatan level data *ROS* dan *Hyperparameter Grid Search* Pengujian ketiga model tersebut dilakukan terhadap kelima klasifikasi negara pada dataset tersebut.

Diketahui bahwa dataset tersebut memiliki distribusi kelas yang tidak seimbang yang digambarkan pada Gambar 6, hal ini perlu diperhatikan untuk perhitungan agar kelas yang minoritas dapat diklasifikasikan dengan baik, maka metrik yang digunakan pada penelitian ini adalah akurasi, F1-Score dan *AUC*.

Pada bagian ini akan membahas hasil pengukuran kinerja model yang diusulkan yang telah dilakukan percobaan, dapatkah algoritma *Random Forest* dapat meningkatkan pemodelan klasifikasi dialek pada dataset menggunakan ekstraksi fitur MFCC. Apakah terjadi perbedaan yang cukup signifikan antara model usulan.

Tabel 9. Pengkodean Eksperimen

| Pemodelan | Pengkodean |
|--------------------------------|------------|
| MFCC + RF + GridSearch | Eks 1 |
| MFCC + RF + SMOTE + GridSearch | Eks 2 |
| MFCC + RF + ROS + GridSearch | Eks 3 |

Pada tabel 9 dilakukan pengkodean untuk percobaan yang dilakukan pada penelitian ini, dimana Eks 1, Eks 2 dan Eks 3.

Tabel 10. Perbandingan Hasil Kinerja Pemodelan

| Eksperimen | Recall | Precision | F1-Score | Akurasi | AUC |
|------------|--------|-----------|----------|---------|------|
| Eks 1 | 0.26 | 0.52 | 0.25 | 0.57 | 0.54 |
| Eks 2 | 0.92 | 0.88 | 0.90 | 0.91 | 0.94 |
| Eks 3 | 0.93 | 0.88 | 0.90 | 0.91 | 0.95 |

Pada tabel 10 hasil evaluasi dari masing pemodelan terhadap dataset yang sama. Dimana data yang sudah dilakukan *resampling* menghasilkan tingkat akurasi yang lebih baik. Dan untuk eks 2 dan eks 3 memiliki nilai akurasi yang sama yaitu bernilai 0.91. ROS sedikit lebih baik dibanding kan SMOTE untuk nilai recall dan AUC. Terjadi perbedaan antara eks 2 dan eks 3 pada nilai Recall dan AUC dimana kedua nya memiliki selisih 0.01 untuk recall dan AUC.

4. KESIMPULAN

Pada penelitian ini, kami telah menerapkan metode random forest dengan random over sampling untuk mengatasi masalah imbalance. Teknik grid search digunakan untuk mencari hyperparameter terbaik untuk algoritma random forest. Hasil percobaan menunjukkan, penerapan teknik *resampling* meningkatkan performa sistem pengenalan dialek secara signifikan. Kedepannya, kami akan melakukan eksplorasi penerapan pengenalan dialek menggunakan metode yang mempertimbangkan relasi temporal data. Sinyal suara adalah contoh data temporal dan RF memiliki keterbatasan untuk itu. Teknik seperti conditional random field, hidden markov model, maupun recurrent neural networks adalah beberapa teknik yang dapat dipertimbangkan dimasa depan. Selain itu, teknik *resampling* berbasis deep learning juga dapat menjadi salah satu arah penelitian. Metode ROS dan SMOTE pada prinsipnya hanya melakukan perbanyakan data dengan melakukan penggandaan sample yang ada, sehingga tidak mempertimbangkan variasi data. Metode seperti adversarial networks mungkin dapat menjadi solusi untuk penggandaan data berbasis model generative sehingga lebih merepresentasikan variasi data.

UCAPAN TERIMAKASIH

Terima kasih disampaikan kepada pihak-pihak yang telah mendukung terlaksananya penelitian ini.

REFERENCES

- [1] R. B. Handoko and S. Suyanto, "Klasifikasi Gender Berdasarkan Suara Menggunakan Support Vector Machine," *Indones. J. Comput.*, 2019, doi: 10.21108/indojc.2019.4.1.244.
- [2] I. S. Permana, Y. Indrawaty, and A. Zulkarnain, "IMPLEMENTASI METODE MFCC DAN DTW UNTUK PENGENALAN JENIS SUARA PRIA DAN WANITA," *MIND J.*, vol. 3, no. 1, pp. 61–76, Jan. 2019, doi: 10.26760/mindjournal.v3i1.61-76.
- [3] A. Setiawan, A. Hidayatno, and R. Isnanto, Rizal, "Aplikasi Pengenalan Ucapan dengan Ekstraksi Mel-Frequency Cepstrum Coefficients (MFCC) Melalui Jaringan Syaraf Tiruan (JST) Learning Vector Quantization (LVQ) untuk Mengoperasikan Kursor Komputer," *Apl. Pengenalan Ucapan dengan Ekstraksi Mel-Frequency Cepstrum Coefficients Melalui Jar. Syaraf Tiruan Learn. Vector Quantization untuk Mengoperasikan Kursor Komput.*, 2011, doi: 10.12777/transmisi.13.3.82-86.
- [4] N. Nurhamidah, E. C. Djamal, and R. Ilyas, "Perintah Menggunakan Sinyal Suara dengan Mel- Frequency Cepstrum



- Coefficients dan Learning Vector Quantization,” *Semin. Nas. Apl. Teknol. Inf.* 2017, 2017.
- [5] A. Lukman and W. T. Saputro, “IDENTIFIKASI NYAMUK CULEX DAN AEDES AEGYPTI BETINA MENGGUNAKAN LINIER PREDICTIVE CODING DAN JARINGAN SYARAF TIRUAN LEARNING VECTOR QUANTIZATION,” *JIKO (Jurnal Inform. dan Komputer)*, vol. 1, no. 2, Sep. 2016, doi: 10.26798/jiko.2016.v1i2.33.
- [6] D. Satria and M. Mushthofa, “Perbandingan Metode Ekstraksi Ciri Histogram dan PCA untuk Mendeteksi Stoma pada Citra Penampang Daun Freycinetia,” *J. Ilmu Komput. dan Agri-Informatika*, 2013, doi: 10.29244/jika.2.1.20-28.
- [7] A. K. H. Al-Ali, D. Dean, B. Senadji, V. Chandran, and G. R. Naik, “Enhanced Forensic Speaker Verification Using a Combination of DWT and MFCC Feature Warping in the Presence of Noise and Reverberation Conditions,” *IEEE Access*, vol. 5, pp. 15400–15413, 2017, doi: 10.1109/ACCESS.2017.2728801.
- [8] Kunxia Wang, Ning An, Bing Nan Li, Yanyong Zhang, and Lian Li, “Speech Emotion Recognition Using Fourier Parameters,” *IEEE Trans. Affect. Comput.*, vol. 6, no. 1, pp. 69–75, Jan. 2015, doi: 10.1109/TAFFC.2015.2392101.
- [9] L. Juvela *et al.*, “Speech Waveform Synthesis from MFCC Sequences with Generative Adversarial Networks,” in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2018, doi: 10.1109/ICASSP.2018.8461852.
- [10] T. Bent, E. Atagi, A. Akbik, and E. Bonifield, “Classification of regional dialects, international dialects, and nonnative accents,” *J. Phon.*, 2016, doi: 10.1016/j.wocn.2016.08.004.
- [11] B. S. Raghuvanshi and S. Shukla, “SMOTE based class-specific extreme learning machine for imbalanced learning,” *Knowledge-Based Syst.*, vol. 187, p. 104814, Jan. 2020, doi: 10.1016/j.knosys.2019.06.022.
- [12] S. Vluymans, “Learning from Imbalanced Data,” in *Studies in Computational Intelligence*, 2019, pp. 81–110.
- [13] A. Sonak, R. Patankar, and N. Pise, “A new approach for handling imbalanced dataset using ANN and genetic algorithm,” in *International Conference on Communication and Signal Processing, ICCSP 2016*, 2016, doi: 10.1109/ICCSP.2016.7754521.
- [14] Shuo Wang and Xin Yao, “Using Class Imbalance Learning for Software Defect Prediction,” *IEEE Trans. Reliab.*, vol. 62, no. 2, pp. 434–443, Jun. 2013, doi: 10.1109/TR.2013.2259203.
- [15] Y. Singh, A. Pillay, and E. Jembere, “Features of Speech Audio for Accent Recognition,” in *2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*, Aug. 2020, pp. 1–6, doi: 10.1109/icABCD49160.2020.9183893.
- [16] S. Helmiyah, I. Riadi, R. Umar, and A. Hanif, “Ekstraksi Fitur Pengenalan Emosi Berdasarkan Ucapan Menggunakan Linear Predictor Ceptral Coeffecient Dan Mel Frequency Cepstrum Coefficients,” *Mob. Forensics*, vol. 1, no. 2, p. 48, Dec. 2019, doi: 10.12928/mf.v1i2.1259.
- [17] A. Syukron and A. Subekti, “Penerapan Metode Random Over-Under Sampling dan Random Forest Untuk Klasifikasi Penilaian Kredit,” *J. Inform.*, vol. 5, no. 2, pp. 175–185, Sep. 2018, doi: 10.31294/ji.v5i2.4158.
- [18] K. R. Gray, P. Aljabar, R. A. Heckemann, A. Hammers, and D. Rueckert, “Random forest-based similarity measures for multi-modal classification of Alzheimer’s disease,” *Neuroimage*, vol. 65, pp. 167–175, Jan. 2013, doi: 10.1016/j.neuroimage.2012.09.065.
- [19] M. Belgiu and L. Drăguț, “Random forest in remote sensing: A review of applications and future directions,” *ISPRS J. Photogramm. Remote Sens.*, vol. 114, pp. 24–31, Apr. 2016, doi: 10.1016/j.isprsjprs.2016.01.011.
- [20] Y. Shuai, Y. Zheng, and H. Huang, “Hybrid Software Obsolescence Evaluation Model Based on PCA-SVM-GridSearchCV,” in *2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)*, Nov. 2018, pp. 449–453, doi: 10.1109/ICSESS.2018.8663753.