

MENDETEKSI JENIS BURUNG BERDASARKAN POLA SUARANYA
Budi Darma Setiawan¹, Imam Cholissodin², Rekyan Regasari MP.³

^{1,2,3}Fakultas Ilmu Komputer Universitas Brawijaya Malang
Email: ¹s.budidarma@ub.ac.id, ²imamcs@ub.ac.id, ³rekyan.rmp@ub.ac.id

(Naskah masuk: 20 Mei 2016, diterima untuk diterbitkan: 20 Juni 2016)

Abstrak

Ilmuwan biologi terutama di bidang biodiversitas, terus melakukan penelitian tentang spesies hewan yang ada di dunia. salah satu hewan yang spesiesnya memiliki banyak variasi adalah burung. Tiap jenis burung memiliki perbedaan-perbedaan, mulai dari bentuk anggota tubuhnya, prilakunya, makanannya hingga suaranya. Ilmuwan sering juga mengalami kesulitan untuk melakukan pengamatan di alam. Misalnya, untuk mengetahui spesies burung apa saja yang ada di suatu daerah, mereka harus hadir di suatu wilayah, dan menelusuri setiap pelosok. kadang kala kehadiran mereka di tempat tersebut dalam jangka waktu lama, malah mengusik burung yang ada, dan burung-burung malah pergi meninggalkan tempat, sebelum berhasil diamati. Salah satu cara untuk mendeteksi burung apa saja yang ada di suatu wilayah, tanpa harus mengusik keberadaan burung adalah dengan menggunakan alat bantu. Bisa dengan menggunakan kamera video untuk mengambil gambar lingkungan sekitar, atau dengan perekam suara, untuk merekam suara burung yang ada di sana. Untuk itu penelitian ini ditujukan untuk membuat sebuah pengklasifikasi suara burung secara otomatis. Fitur yang digunakan adalah *rhythm*, *pitch*, *mean*, *varian*, *min*, *max*, dan *delta* dari suara burungnya. Metode klasifikasi yang digunakan adalah *Ekstreme learning Machine* (ELM). Dari hasil klasifikasi 4 jenis burung, didapatkan hasil rata-rata akurasi terbaik sebesar 88.82%.

Kata Kunci: *suara burung, klasifikasi, rhythm, pitch*

Abstract

Many of Biology scientist, especially in the field of biodiversity, conduct research on the animal species that exist in the world. One of the animal which is largely diverse in species is bird. Each species of birds have differences, from the shape of his body, his behavior, his food to it's voice. Scientists often find it difficult to make observations in nature. For example, to determine which species of birds present in an area, they should be present in an area, and explore every corner. Sometimes their presence in that place for a long time, even disturb the bird, and they leaving the place, before been observed. One way to detect any bird that is in an area, without having to disturb the presence of birds is to use the automatic tools. For example to use a video camera to take pictures of the surrounding environment, or with voice recorders to record the sound of the birds that were there. This study is aimed to create a classifier bird sound automatically. Features used in this research are rhythm, pitch, mean, variance, min, max, and delta of the bird sound samples. For classifying, we used Extreme Learning Machine (ELM). The average results of the classification of four types of birds, showed the best average of accuracy is 88.82%.

Keywords: *bird song, classification, rhythm, pitch.*

1. Pendahuluan

Ilmuwan biologi terutama di bidang biodiversitas, terus melakukan penelitian tentang spesies hewan yang ada di dunia. salah satu hewan yang spesiesnya memiliki banyak variasi adalah burung. Tiap jenis burung memiliki perbedaan-perbedaan, mulai dari bentuk anggota tubuhnya, prilakunya, makanannya hingga suaranya. Ilmuwan sering juga mengalami kesulitan untuk melakukan pengamatan di alam. Misalnya, untuk mengetahui spesies burung apa saja yang ada di suatu daerah, mereka harus hadir di suatu wilayah, dan menelusuri setiap pelosok. kadang kala kehadiran mereka di tempat tersebut dalam jangka waktu lama, malah mengusik burung yang ada, dan burung-burung malah pergi meninggalkan tempat,

sebelum berhasil diamati. Salah satu cara untuk mendeteksi burung apa saja yang ada di suatu wilayah, tanpa harus mengusik keberadaan burung adalah dengan menggunakan alat bantu. Bisa dengan menggunakan kamera video untuk mengambil gambar lingkungan sekitar, atau dengan perekam suara, untuk merekam suara burung yang ada di sana.

Permasalahan selanjutnya adalah, jika kita menggunakan perekam suara, maka sistem yang kita pasang juga harus bisa mengenali bahwa suatu suara kicauan burung adalah milik burung spesies tertentu. Untuk itu perlu dibuat sebuah sistem dengan menggunakan algoritma tertentu, untuk mengenali suara kicauan burung dengan baik. Penelitian sebelumnya sudah pernah dilakukan, misalnya: Somervuo (2004) menggunakan analisis

histogram pada *syllable* (suku kata) dari kicauan burung. Pada tahun 2011, Wei Chu menganalisa *syllable* kicauan burung dengan menggunakan *Hidden Markov Model*. Dan di tahun 2013, Sha-sha Chen menganalisa tekstur dari frekuensi berbasis waktu dari *syllable* kicauan burung. Dari beberapa penelitian yang sudah dilakukan, sebagian besar menggunakan *syllable* untuk dianalisis. Penelitian ini akan menggunakan analisis frekuensi dari setiap *syllable* kicauan burung, menganalisa jarak antara gelombang pada *syllable*, serta keberadaan perubahan *pitch* dari jenis burung tertentu. Pengamatan dan analisis ini ditekankan pada penggunaan fitur-fitur, yaitu *rhythm*, *pitch*, *mean*, *varian*, *max*, *min*, dan *delta*.

Penelitian ini diharapkan menjadi langkah awal dalam membuat tool otomatis untuk dapat mendeteksi jenis burung. Dari latar belakang tersebut, dapat dirumuskan rumusan masalah sebagai berikut:

1. Fitur apa saja yang digunakan untuk dapat membedakan jenis burung melalui pengenalan suaranya.
2. Seberapa baik akurasi yang dihasilkan dengan menggunakan kombinasi fitur dan metode tersebut.

2. Kajian Pustaka

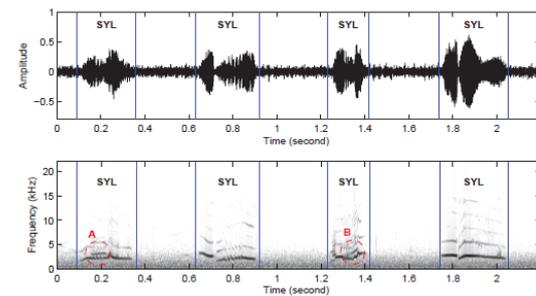
Studi dan penelitian mengenai pengenalan suara burung sudah banyak dilakukan. Metode yang digunakan juga bervariasi. Sebagian besar meneliti suara burung berdasarkan *Syllable*. Untuk memecah suara burung menjadi *syllable* dilakukan segmentasi. Cara yang pertama untuk memisahkan *syllable* ini dengan menggunakan *change point detection* yang melihat perubahan bentuk gelombang dari suara burung (Wang et al., 2013). Cara lainnya dengan menggunakan *time-frequency segmentation* pada lingkungan suara dengan banyak noise. metode ini menrubah gelombang suara burung ke dalam bentuk spektrogram dengan *Short-time FFT* dan kemudian melakukan segmentasi untuk memisahkan noise dan suara burung (Neal et al., 2011).

Dari *Syllable* yang didapatkan, kemudian dilakukan analisis dengan berbagai metode. Salah satunya adalah dengan menggunakan metode pengenalan kata dari suara manusia, yaitu dengan menggunakan *Hidden Markov Model* (Chu & Blumstein, 2011). dari penelitian tersebut diperoleh nilai *Precision* terbesar yaitu 75,2% dan *Recal* terbesar yaitu 76%. Metode lain yang digunakan yaitu dengan menganalisis tekstur dari spektrogram tiap *syllable*. Akurasi yang diperoleh sebesar 96,5% dari 10 spesies burung yang dipelajari (Chen & Li, 2013). Di tahun 2004 *syllable* yang didapatkan dianalisis dengan menggunakan histogramnya, dan menghasilkan akurasi terbesar yaitu 80% (Somervuo & Harma, 2004).

Penelitian kali ini akan metode untuk mengambil *syllable* akan menggunakan metode *change point detection* (Wang et al., 2013). Dan analisis *syllable*-nya menggunakan analisis tekstur. Untuk meningkatkan akurasi, akan diamati pula fitur lainnya, yaitu ritme suara burung. Hal ini dilakukan dengan mengamati tempo. Selain itu juga akan mengamati tinggi rendahnya nada (*pitch*), pengulangan *syllable* dan nadanya.

3. Suara Burung

Jika didengarkan, variasi suara burung lebih sederhana jika dibandingkan dengan suara manusia. Namun demikian, suara burung juga dibagi menjadi beberapa bagian, yaitu frasa, suku kata (*syllable*) dan elemen (Catchpole & Slater, 1995). Frasa adalah kumpulan *syllable* yang biasanya muncul membentuk pola tertentu. Setiap burung memiliki bentuk kumpulan *syllable* dan frasa yang berbeda. *Syllable* adalah suara yang dikeluarkan oleh burung dalam satu kali tiupan udara dari paru-parunya. (Somervuo et al., 2006). *Syllable* ini yang biasa didengarkan sebagai satu siulan. Jika digambarkan dalam gelombang, *syllable* bisa dilihat seperti pada gambar 1.



Gambar 1 SYL = Syllable dalam bentuk gelombang (atas), dan spektrogram (bawah). (Chu & Blumstein, 2011)

4. Ritme

Ritme menggambarkan tempo ketukan dari suara burung. ada yang bertempo cepat, dan ada yang bertempo lambat. Ritme dapat diartikan waktu yang dibutuhkan untuk membunyikan 1 syllable. Ritme diukur dengan menghitung rasio waktu dengan jumlah syllable yang muncul pada saat itu.

5. Pitch

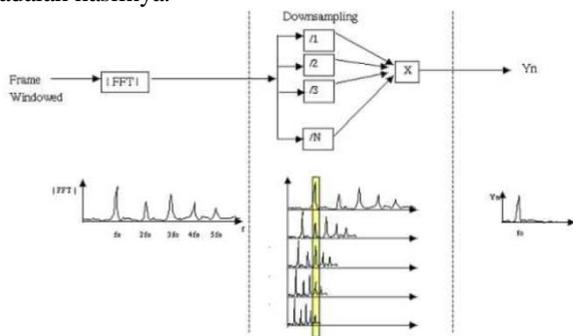
Beberapa jenis burung memiliki *pitch* yang khas. ada yang memiliki *pitch* sangat tinggi sampai yang sangat rendah. Ada jenis burung yang *pitch*-nya rata/datar, dan bahkan ada jenis burung, yang *pitch*-nya berubah-ubah dalam satu kali siulan.

Untuk mengambil nilai *pitch*, digunakan metode *Harmonic Product Spectrum* (HPS). Prinsip kerja HPS adalah:

1. Masukan yang berupa sample suara, dibagi menjadi beberapa *window*. Pada penelitian ini, isi dari tiap *window* adalah *syllable*.

2. Tiap *window*, ditransformasi dengan menggunakan *Fast Fourier Transform* (FFT) untuk mencari frekuensi penyusun gelombang suara tiap *window*.
3. Tiap *window*, hasil FFT tadi dilakukan *downsampling* 2 kali, dan 3 kali
4. Mengalikan *magnitude* dari setiap frekuensi yang bersesuaian antara hasil FFT awal, hasil *downsapling* 2 kali, dan hasil *downsampling* 3 kali.

Ilustrasinya bisa dilihat pada gambar 2. Bagian kiri adalah masukan yang ditransformasi dengan menggunakan FFT, bagian tengah adalah proses *downsampling* dan perkalian *magnitude* dari frekuensi yang bersesuaian, dan bagian kanan adalah hasilnya.



Gambar 2. Proses Harmonic Product Spectrum

Contoh downsampling 3 kali:

	sample									
FFT	1	2	30	4	52	3	70	4	1	30
FFT downsampling 2	1	30	52	70	1	0	0	0	0	0
FFT downsampling 3	1	4	70	30	0	0	0	0	0	0

hasil	1	240	109200	8400	0	0	0	0	0	0
Frequensi (Hz)	0	20	40	60	80	100	120	140	160	180

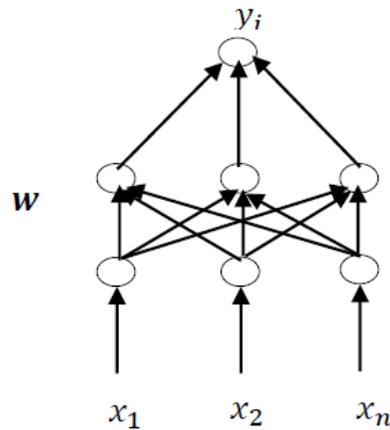
dari contoh didapatkan *magnitude* tertinggi sebesar 109200 yaitu terletak pada frekuensi 40 Hz. dengan demikian pitchnya adalah 40 Hz.

6. Extreme Learning Machine (ELM)

Metode ELM ini pertama kali diperkenalkan oleh Huang (2004). ELM merupakan jaringan syaraf tiruan (JST) *feedforward* dengan *single hidden layer* atau disebut dengan *Single Hidden Layer Feedforward Neural Networks* (SLFNs) (Sun et al., 2008). Metode pembelajaran ELM dibuat untuk mengatasi beberapa kelemahan dari jaringan syaraf tiruan *feedforward*, terutama dalam hal *learning speed*.

Menurut Huang (2004), JST *feedforward* masih memiliki kelemahan dalam *learning speed* karena:

1. Menggunakan *slow gradient based learning* untuk melakukan training.
2. Semua parameter pada jaringan ditentukan secara iteratif dengan menggunakan metode pembelajaran satu persatu untuk setiap data.



Gambar 3 Struktur ELM

Pada proses pembelajaran JST, semua parameternya harus ditentukan secara manual (Huang et al., 2005), yaitu *input weight* dan *hidden bias* yang menghubungkan antara *layer* yang satu dengan yang lain untuk setiap data training, sehingga membutuhkan *learning speed* yang lama. Sedangkan pada ELM parameter-parameter tersebut dipilih secara acak, sehingga ELM memiliki *learning speed* yang cepat dan mampu menghasilkan *good generalization performance*. Hal itu menjadikan ELM cepat dari algoritma *feedforward* BP. Struktur ELM dapat dilihat pada gambar 3.

Berikut model matematis dari ELM pada data (x_i, y_i) .

$$\sum_{i=1}^{\bar{N}} \beta_i g(w_i \cdot x_i + b_i) = y_j \tag{1}$$

dimana

$$x_i = [x_{i1} \ \dots \ x_{iN}]^T \in R^N \text{ dan}$$

$$y_i = [y_{i1} \ \dots \ y_{im}]^T \in R^m$$

kemudian

$$w_i = [w_{i1} \ \dots \ w_{iN}]^T \text{ dan}$$

$$\beta_i = [\beta_{i1} \ \dots \ \beta_{im}]^T$$

Persamaan (1) dapat disederhanakan menjadi persamaan (2), yang kemudian menggunakan *Moore-Penrose* (MPinv(HT)) = (inv((HT*H))* HT) sebagai *generalized inverse*.

$$H\beta = Y \tag{2}$$

Keterangan :

H = Matrik fungsi aktivasi hidden neuron

\bar{N} = jumlah hidden neuron yang diaktivasi dengan fungsi $g(x)$

β = Matrik bobot output

Y = Matrik output dimana,

$$H(w_1, \dots, w_N, b_1, x_1, \dots, x_N) = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \dots & g(w_{\bar{N}} \cdot x_{\bar{N}} + b_{\bar{N}}) \\ \vdots & \ddots & \vdots \\ g(w_1 \cdot x_N + b_1) & \dots & g(w_{\bar{N}} \cdot x_N + b_{\bar{N}}) \end{bmatrix}_{\bar{N} \times N}$$

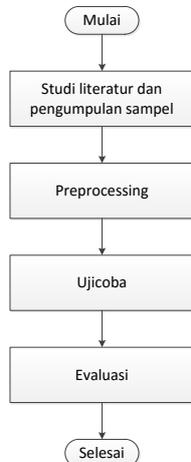
$$\beta = [\beta_1^T \ \dots \ \beta_{\bar{N}}^T]_{\bar{N} \times m} \text{ dan}$$

$$Y = [y_1^T \ \dots \ y_{\bar{N}}^T]_{\bar{N} \times m} \tag{3}$$

$$\hat{\beta} = MP\text{inv}(H^T)Y \tag{4}$$

7. Metode Penelitian

Penelitian ini dilakukan dengan beberapa tahapan. Dimulai dari pengumpulan data, hingga proses ujicoba. Diagramnya dapat dilihat pada Gambar 4.



Gambar Error! No text of specified style in document. Alur Pelaksanaan Penelitian

8. Hasil dan Pembahasan

8.1. Hasil Preprocessing dan Smoothing

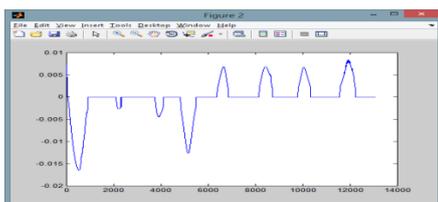
Hasil dari preprocessing adalah dapat menemukan bagian-bagian dari gelombang yang merupakan *syllable*. Pada gambar 5, bentuk gelombang asli di bagian atas diproses dengan proses smoothing dengan menggunakan metode *robust lowest* yang menggunakan regresi linier.



Gambar Error! No text of specified style in document.. Hasil Preprocessing

8.2. Hasil deteksi syllable

Setelah itu, sinyal hasil *smoothing* gelombang dengan simpangan tertentu dijadikan *syllable*. Hasilnya dapat dilihat pada gambar 6.

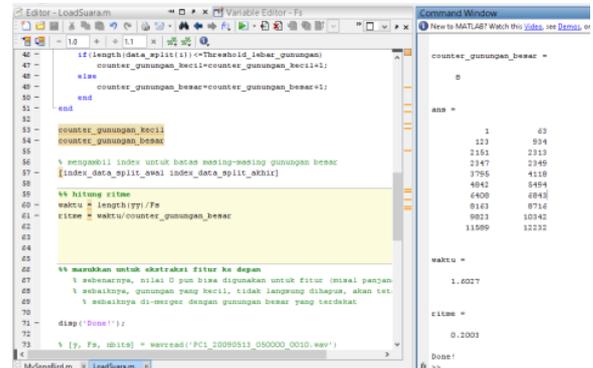


Gambar 6 Hasil *Smoothing* Signal untuk diambil setiap *syllable*

8.3. Hasil Ekstraksi Fitur

8.3.1. Fitur Ritme

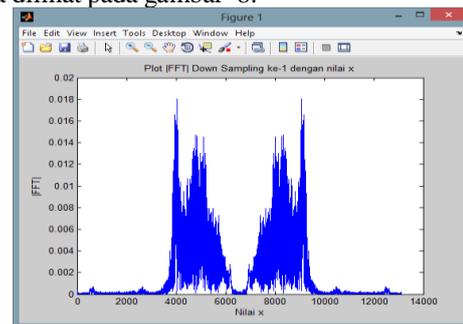
Setelah *syllable* terdeteksi, memungkinkan untuk menghitung jumlah *syllable* yang muncul dan ritmenya. Hasilnya bisa dilihat pada gambar 7.



Gambar 7 Hasil Ekstraksi Fitur Ritme

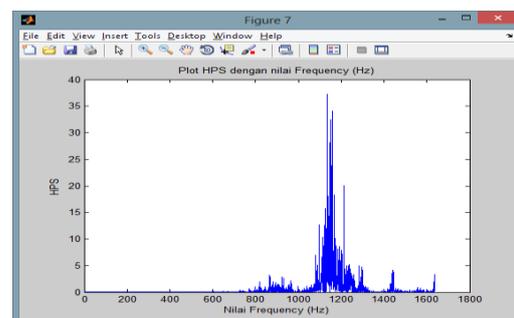
8.3.2. Fitur Pitch

Untuk mendapatkan fitur *pitch*, gelombang suara ditransformasi dahulu ke dalam domain frekuensi menggunakan transformasi FFT. hasilnya bisa dilihat pada gambar 8.



Gambar 8 Hasil FFT(y)

Setelah ditransformasi, dilakukan *downsampling* sebanyak 4 kali. Kemudian hasil *downsampling* dan hasil FFT dikalikan semua, untuk mendapatkan *Harmonic Product Spectrum* (HPS). Nilai Frekuensi terbesar dalam HPS ini yang kemudian dijadikan nilai *pitch*. grafik HPSnya bisa dilihat pada gambar 9. Dari situ, didapatkan nilai *pitch* berupa frekuensi dengan nilai HPS tertinggi.



Gambar 9. Plot HPS dalam Frequency (Hz)

8.4. Lingkungan Uji Coba

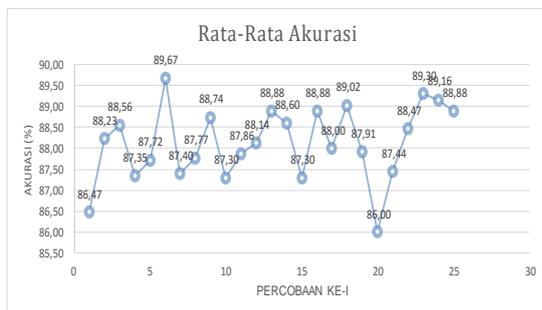
Pada penelitian ini, data suara yang digunakan berupa file (.wav). Data diunduh dari internet

(youtube) dengan panjang 13129 *sample*. Data yang digunakan sebanyak 175 file *.wav dengan 4 kelas (*Burung Blackthroat, Celilin, Cendet, dan Kutilang*). Spesifikasi dari perangkat keras yang digunakan dalam implementasi perangkat lunak terdiri dari prosesor berjenis *Intel(R) Core(TM) i7*, memori berkapasitas 4 GB dan kapasitas *harddisk* 750 GB. Sedangkan untuk spesifikasi perangkat lunak yang digunakan dalam implementasi perangkat lunak yakni *Matlab 2009*.

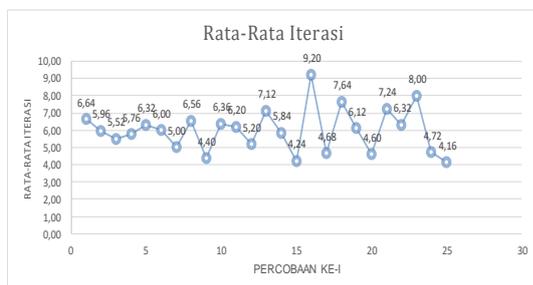
8.5. Hasil Uji Coba

8.5.1. Pengujian Akurasi terhadap K-Fold (K=2, Fold Ke-1)

Pengujian akurasi digunakan untuk mengetahui performa dari sistem klasifikasi suara burung menggunakan algoritma *Extreme Learning Machine* (ELM). Pengujian ini dilakukan dengan cara membandingkan kelas aktual dengan kelas prediksi. Data hasil dari nilai rata-rata akurasi dari keseluruhan pengujian dan jumlah iterasinya dapat dilihat pada grafik Gambar 10 dan 11. Berdasarkan hasil pengujian akurasi pada Gambar 5.14 menggunakan K=2 Fold Ke-1 memiliki nilai akurasi rata – rata yakni 88.12%. Dari hasil ini disimpulkan bahwa algoritma ELM mampu melakukan klasifikasi dengan baik.



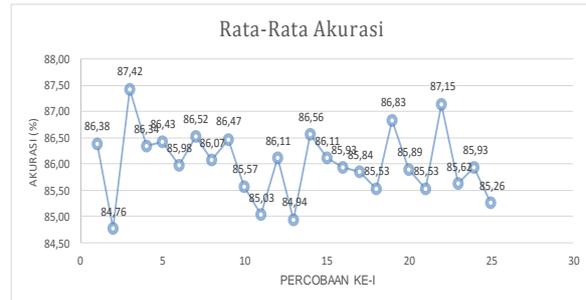
Gambar 10 Grafik Pengujian Akurasi terhadap 2-Fold Ke-1



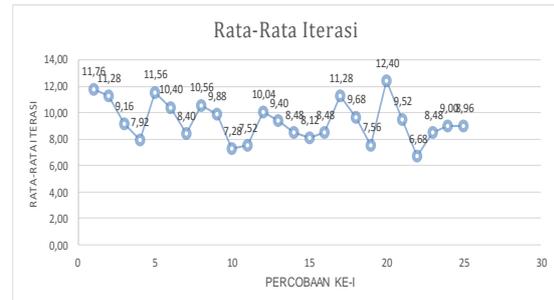
Gambar 11 Grafik Iterasi terhadap 2-Fold Ke-1

8.5.2. Pengujian Akurasi terhadap K-Fold (K=2, Fold Ke-2)

Data hasil dari nilai rata-rata akurasi dari keseluruhan pengujian dan jumlah iterasinya, dapat dilihat pada Gambar 12 dan 13. Didapatkan nilai akurasi rata – rata sebesar 86.01%.



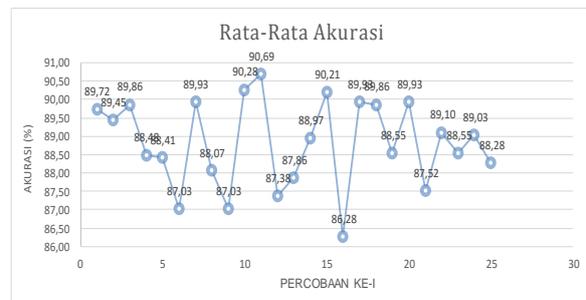
Gambar 12 Grafik Pengujian Akurasi terhadap 2-Fold Ke-2



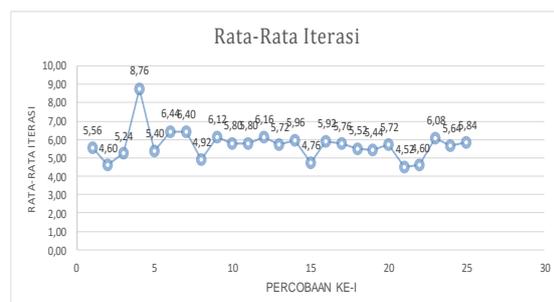
Gambar 13 Grafik Iterasi terhadap 2-Fold Ke-2

8.5.3. Pengujian Akurasi terhadap K-Fold (K=3, Fold Ke-1)

Data hasil dari nilai rata-rata akurasi dari keseluruhan pengujian ini dan rata-rata iterasinya dapat dilihat pada Gambar 14 dan Gambar 15. Berdasarkan hasil pengujian akurasi pada Gambar 5.18 memiliki nilai akurasi rata – rata yakni 88.82%.



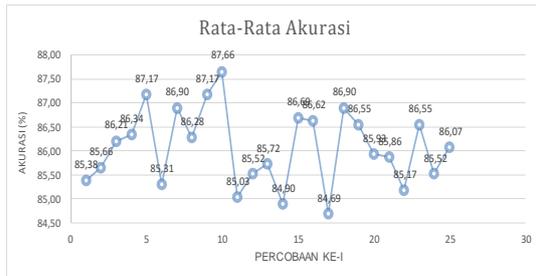
Gambar 14 Grafik Pengujian Akurasi terhadap 3-Fold Ke-1



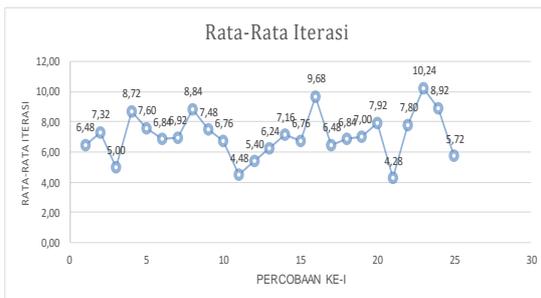
Gambar 15 Grafik Iterasi terhadap 3-Fold Ke-1

8.5.4. Pengujian Akurasi terhadap K-Fold (K=3, Fold Ke-2)

Gambar 16 menunjukkan grafik hasil dari nilai rata-rata akurasi dari keseluruhan pengujian. Dan gambar 17 menunjukkan rata-rata iterasi yang dilakukan. Berdasarkan hasil pengujian akurasi pada Gambar 5.20 memiliki nilai akurasi rata – rata yakni 86.07%. Dari hasil ini disimpulkan bahwa algoritma ELM mampu melakukan klasifikasi dengan baik.



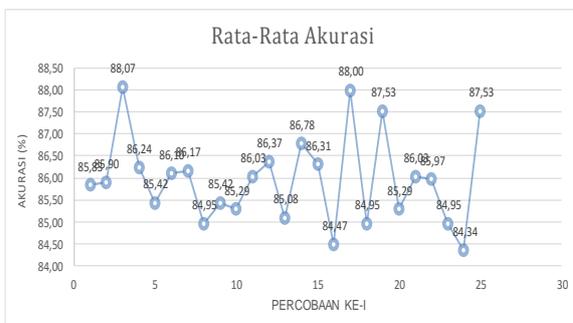
Gambar 16 Grafik Pengujian Akurasi terhadap 3-Fold Ke-2



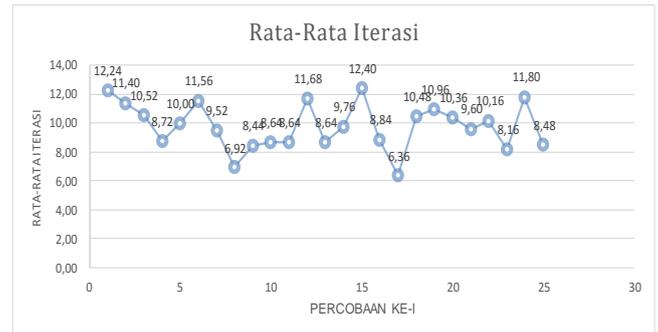
Gambar 17 Grafik Iterasi terhadap 3-Fold Ke-2

8.5.5. Pengujian Akurasi terhadap K-Fold (K=3, Fold Ke-3)

Gambar 18 menunjukkan grafik hasil dari nilai rata-rata akurasi dari keseluruhan pengujian, dan gambar 19 menunjukkan rata-rata iterasinya. Berdasarkan hasil pengujian ini diketahui nilai akurasi rata – rata yakni 85.96%. Dari hasil ini disimpulkan bahwa algoritma ELM mampu melakukan klasifikasi dengan baik.



Gambar 18 Grafik Pengujian Akurasi terhadap 3-Fold Ke-3



Gambar 19 Grafik Iterasi terhadap 3-Fold Ke-3

9. Kesimpulan

1. Implementasi *Extreme Learning Machine* (ELM) dapat digunakan untuk mengklasifikasikan suara burung dengan melakukan proses ekstraksi fitur dari ritme, pitch, mean, varian, min, max dan delta. Hasil ekstraksi fitur tersebut kemudian dilakukan proses pelatihan dan pengujian terhadap dataset yang sudah disediakan.
2. Hasil rata-rata akurasi terbaik dari klasifikasi suara burung dengan menggunakan ELM sebesar 88,82% dari 3-Fold ke-1 dengan banyaknya iterasi yang dibutuhkan sebesar 6 iterasi. Kemudian sistem sempat mampu menghasilkan nilai akurasi tertinggi 100% saat beberapa percobaan pada 3-Fold ke-1 dan pada 3-Fold ke-3.

10. Daftar Pustaka

Birds, A. A. (2015). *The Cornell lab of Ornithology - All About Birds*. Retrieved March 1, 2015, from <http://www.allaboutbirds.org/page.aspx?pid=1059>

Catchpole, C. K., & Slater, P. J. (1995). *Bird Song: Biological Themes and Variations*. Cambridge: Cambridge Univ. Press.

Chen, S.-s., & Li, y. (2013). Automatic Recognition of Bird Songs Using Time-Frequency Texture. Mathura: 5th International Conference on Computational Intelligence and Communication Networks (CICN).

Chu, W., & Blumstein, D. T. (2011). Noise robust bird song detection using syllable pattern-based hidden Markov models. Prague: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

Huang, G.B., Zhu, Q.Y., dan Siew, C.K. (2004). Extreme Learning Machine : A New Learning Scheme of Feddforward neural Networks. Proceeding of International Joint Conference on Neural Networks. Hungary, 25-29 Juli.

Huang, G.B., Zhu, Q.Y., dan Siew, C.K. (2005). Extreme Learning Machine : Theory and applications. Elsevier science : Neurocomputing 70 (2006) 489-501.

- Neal, L., Brigs, F., Raich, R., & Fern, X. Z. (2011). Time-frequency segmentation of bird song in noisy acoustic environments. Prague: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).
- Somervuo, P., & Harma, A. (2004). Bird song recognition based on syllable pair histograms. IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. (ICASSP '04).
- Somervuo, P., Harma, A., & Fagerlund, S. (2006). Parametric Representations of Bird Sounds for Automatic Species Recognition. *IEEE Transaction on Audio, Speech, and Language Processing*, 14(6).
- Sun, Z.L., Choi, T.M., Au, K.F., dan Yu, Y. (2008). Sales Forecasting using Extreme Learning Machine with Application in Fashion Retailing. Elsevier Decision Support Systems 46 (2008) 411-419.
- Wang, N.-C., Hudson, R. E., Tan, L. N., Taylor, C. E., Alwan, A., & Yao, K. (2013). Change point detection methodology used for segmenting bird songs. Beijing: IEEE China Summit & International Conference on Signal and Information Processing (ChinaSIP).