

Seleksi Fitur Dengan *Artificial Bee Colony* Untuk Optimasi Klasifikasi Data Teh Menggunakan *Support Vector Machine*

Suhaila*¹, Danang Lelono², Yunita Sari³

¹Program Studi Elektronika dan Instrumentasi; FMIPA UGM, Yogyakarta, Indonesia

^{2,3}Departemen Ilmu Komputer dan Elektronika, FMIPA UGM, Yogyakarta, Indonesia

e-mail: *suhaila@mail.ugm.ac.id, danang@ugm.ac.id, yunita.sari@ugm.ac.id

Abstrak

Teh dapat dikenal kualitasnya melalui aroma yang dihasilkan. Penelitian klasifikasi teh menggunakan *e-nose* umumnya hanya mendeteksi kualitas aroma menggunakan general sensor gas. Namun, adanya redundansi fitur sensor dapat menyebabkan penurunan performa sistem *e-nose*. Oleh karena itu diperlukan sebuah sistem yang dapat menyeleksi fitur sehingga performa klasifikasi menjadi lebih optimal. Pada penelitian ini dibentuk sistem perangkat lunak yang mampu menyeleksi fitur untuk mengoptimalkan performa klasifikasi. Data input untuk sistem adalah respon sensor *e-nose* terhadap 3 kualitas teh hitam dengan jumlah sampel 300. Fitur yang diseleksi berupa sensor-sensor pada instrumen *e-nose*. Proses seleksi fitur dilakukan dengan pendekatan wrapper, algoritma ABC digunakan untuk seleksi fitur, kemudian hasil fitur yang terpilih dievaluasi dengan klasifikasi menggunakan SVM. Hasil sistem ABC-SVM kemudian dibandingkan dengan sistem SVM tanpa seleksi fitur. Hasil penelitian menunjukkan bahwa dari 12 sensor *e-nose*, sensor yang paling mencirikan teh hitam kualitas 1-3 yaitu sensor TGS 2600, TGS 813, TGS 825, TGS 2602, TGS 2611, TGS 832, TGS 2612, TGS 2620 dan TGS 822. Sedangkan untuk sensor MQ-7, TGS 826 dan TGS 2610 merupakan sensor yang redundan pada sistem dikarenakan gas yang dideteksi oleh 3 sensor tersebut dapat diwakili oleh sensor lainnya. Dengan berkurangnya fitur menjadi 9, performa akurasi klasifikasi meningkat 16,7%.

Kata kunci— Seleksi Fitur, *Artificial Bee Colony*, *Support Vector Machine*, Klasifikasi.

Abstract

Tea quality can be recognized through the aroma it produces. Tea classification using *e-nose* generally only detects aroma using a general gas sensor. However, redundancy of sensor features can cause a decreasing in the system performance. Therefore we need a system that can select features so the classification performance becomes optimal. A software system of feature selection was built to optimize classification performance. Input data for the system is *e-nose* sensor response to 3 black tea qualities. The features are sensors on the *e-nose* instrument. Feature selection is implemented using wrapper approach, ABC algorithm is used for feature selection, then the selected features are evaluated by SVM classification. The results of the ABC-SVM system are then compared with the SVM only system. The results showed that from 12 *e-nose* sensors, sensors that most characterized black tea quality were TGS 2600, TGS 813, TGS 825, TGS 2602, TGS 2611, TGS 832, TGS 2612, TGS 2620 and TGS 822. Meanwhile, MQ-7, TGS 826 and TGS 2610 sensors are redundant in the system because the gas detected by the 3 sensors can be represented by other sensors. With the reduction in features to 9, the classification accuracy performance increased by 16.7%.

Keywords— Feature Selection, ABC, SVM, Classification

1. PENDAHULUAN

Faktor utama penentu kualitas teh adalah aroma, warna dan rasa, dengan faktor aroma merupakan faktor yang paling penting [1]. Umumnya di industri teh, untuk meng-klasifikasi teh dilakukan oleh tester teh berdasarkan pertimbangan aroma, warna, rasa, dll. Namun penentuan kualitas dan juga jenis teh oleh tester teh ini cenderung bersifat subjektif dan relatif akurasi. Oleh karena itu seiring berkembangnya zaman, diperlukan sebuah instrumen alternatif yang dapat menentukan kualitas yaitu electronic nose (e-nose)[2][3].

E-nose telah banyak digunakan dalam berbagai bidang di antaranya pada bidang kesehatan yaitu identifikasi kanker paru-paru berdasarkan deteksi nafas, pada industri makanan yaitu identifikasi teh hijau, kopi, daging sapi dan teh hitam. Pada sistem e-nose menggunakan metode pengenalan pola memungkinkan sistem untuk menganalisa aroma kompleks dari sinyal sensor [4]. Dalam proses pengenalan pola, respon sensor di preproses untuk menghilangkan noise, kemudian diekstraksi fitur yang mencirikan sampel untuk dilakukan identifikasi atau klasifikasi sampel.

Beberapa penelitian yang telah melakukan klasifikasi pada data e-nose yaitu klasifikasi kualitas roti dengan KNN, klasifikasi kopi dengan LDA. Penelitian selanjutnya membandingkan performa klasifikasi SVM dengan LDA [5] dan juga performa klasifikasi SVM dengan K-NN [6][7]. Berdasarkan penelitian tersebut didapatkan bahwa SVM mengungguli LDA dan K-NN dalam segi performa, akurasi, dan kemampuan mengklasifikasikan. Kelebihan metode SVM diantaranya memiliki performa klasifikasi yang baik berdasarkan pada akurasi tinggi dan error pembacaan yang rendah serta memiliki waktu komputasi yang lebih cepat dibandingkan dengan metode lainnya dengan hasil akurasi yang masih unggul [8][7].

Kemampuan e-nose tidak hanya terletak pada mesin pengenalan pola untuk klasifikasi sampel. Bahkan mesin pengenalan pola akan menghasilkan performa yang optimal jika sinyal respon dari setiap sensor dapat diketahui ciri spesifiknya. Namun, pada proses ekstraksi ciri adanya kesamaan respon sensor menyebabkan penentuan ciri menjadi tidak spesifik serta redundansi fitur sensor, sehingga dapat menurunkan performa mesin pengenalan pola dalam melakukan proses klasifikasi. Performa klasifikasi dapat ditingkatkan dengan menghapus fitur yang redundant dan tidak relevan dari fitur awal [9]. Feature selection mengekstraksi informasi yang relevan dan paling berguna serta mempercepat proses klasifikasi. Oleh karena itu dibutuhkan suatu metode untuk melakukan pemilihan subset fitur yang relevan dan mengandung paling banyak informasi signifikan saja [10].

Beberapa algoritma optimasi yang biasa digunakan yaitu PSO, ACO dan GA. Salah satu algoritma optimasi lainnya yaitu ABC (Artificial Bee Colony) yang merupakan algoritma optimasi berdasarkan kecerdasan perilaku lebah madu [11]. Penelitian sebelumnya telah banyak melakukan perbandingan performa ABC dengan algoritma optimasi lainnya seperti DE, PSO, ACO serta GA, dan didapatkan ABC unggul dalam akurasi, convergence time, computational time [12][13].

Pada penelitian ini dilakukan optimasi sistem klasifikasi SVM dengan mengimplementasikan algoritma ABC untuk seleksi fitur sensor yang redundant, sehingga diharapkan performa klasifikasi yang dihasilkan lebih optimal.

2. METODE PENELITIAN

2.1 Analisis Sistem

Dalam proses klasifikasi teh oleh tester teh, terdapat faktor-faktor yang mempengaruhi keakuratan hasil klasifikasi yaitu kesubjektifan setiap orang berbeda-beda, pengalaman yang berbeda, serta kondisi fisik yang berbeda. Untuk itu diperlukan instrument yang dapat melakukan klasifikasi teh secara lebih akurat.

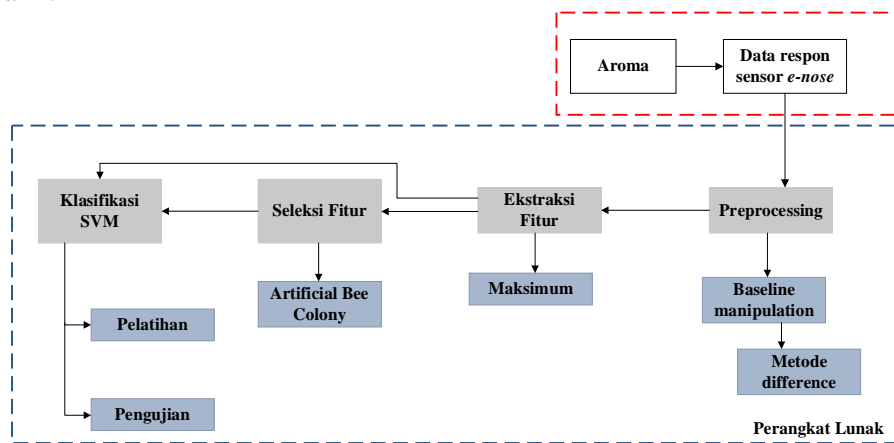
E-nose merupakan instrument yang bekerja layaknya hidung manusia. E-nose terdiri dari larik sensor yang telah terintegrasi untuk mendeteksi gas. Dalam e-nose dibutuhkan sistem pengenalan pola untuk mengklasifikasi gas yang telah dideteksi.

Aroma yang ditangkap oleh array sensor gas diubah ke dalam bentuk tegangan menjadi respon sensor, respon sensor tersebut dipreproses dan diekstraksi fiturnya untuk pemrosesan sinyal lebih lanjut. Namun, adanya kesamaan respon sensor menyebabkan redundansi fitur sehingga dapat menurunkan kinerja klasifikasi. Selanjutnya diperlukan pemilihan respon sensor yang redundan dan pada data yang kompleks sulit dilakukan pemilihan sensor secara manual.

Oleh karena itu diperlukan sebuah mesin yang mampu menyeleksi sensor yang redundan hingga diperoleh sensor-sensor yang mencirikan kualitas teh hitam dengan spesifik, sehingga dapat meningkatkan performa dari mesin pengenalan pola dalam meng-klasifikasi sampel.

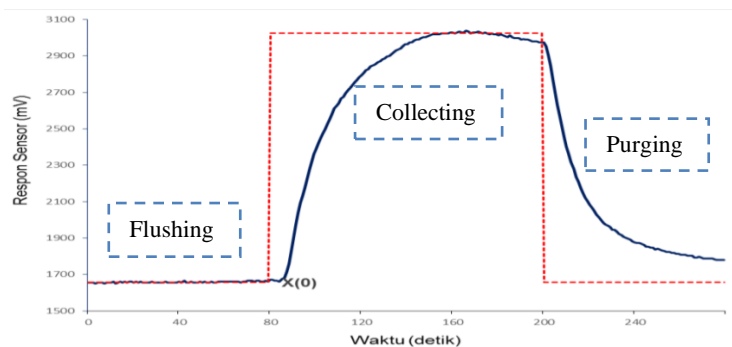
2. 2 Perancangan Sistem

Instrumen e-nose terdiri dari perangkat keras dan perangkat lunak. Perangkat keras e-nose tersusun atas perangkat elektronik untuk mengubah aroma menjadi sinyal keluaran dalam bentuk respon sensor. Sedangkan perangkat lunak merupakan sistem yang dirancang untuk pemrosesan respon sensor lebih lanjut hingga sistem dapat mengidentifikasi aroma sampel yang diberikan. Gambaran instrumen e-nose pada sistem ditunjukkan dalam bentuk diagram blok pada Gambar 1.



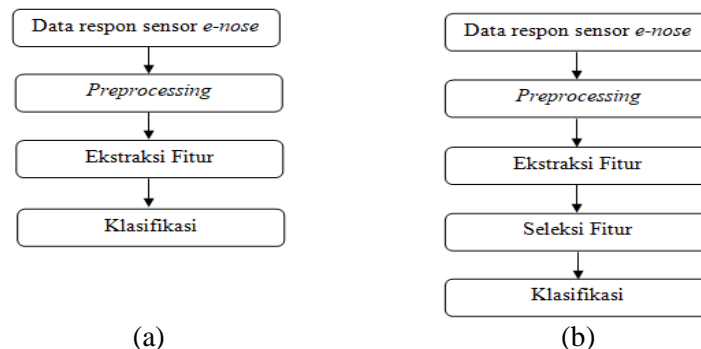
Gambar 1 Diagram Blok E-nose

Sensor-sensor yang digunakan pada instrumen e-nose berjumlah 12 yang terdiri dari MQ-7, TGS 2600, TGS 813, TGS 825, TGS 2602, TGS 826, TGS 2610, TGS 2611, TGS 832, TGS 2612, TGS 2620 dan TGS 822. Respon sensor terdiri dari sinyal respon yang terbentuk dari aroma yang berhasil ditangkap oleh masing-masing sensor (S_1, S_2, \dots, S_{12}) dan diubah menjadi besaran tegangan dengan waktu setiap proses sniffing. Proses sniffing merupakan proses yang membentuk respon sensor agar terlihat responnya terhadap aroma. Respon sensor e-nose dapat dilihat pada Gambar 2.



Gambar 2 Respon sensor e-nose [14]

Respon sensor pada Gambar 2 nantinya akan berjumlah 12 dikarenakan jumlah sensor yang digunakan yaitu 12. Respon sensor ini kemudian diolah berdasarkan rancangan sistem pada Gambar 3.



Gambar 3 Rancangan sistem (a) tanpa dan (b) dengan seleksi fitur

Terdapat dua rancangan sistem yaitu tanpa seleksi fitur dan dengan seleksi fitur. Sistem yang tanpa seleksi fitur diawali dengan pembacaan data respon sensor e-nose. Kemudian dilakukan preprocessing dengan mengurangi respon sensor dengan nilai baseline agar data respon sensor lebih berskala. Setelah preprocessing, dilakukan ekstraksi fitur menggunakan metode maksimum. Lalu setelah ekstraksi fitur, dilakukan klasifikasi menggunakan metode SVM. Pada sistem dengan seleksi fitur, setelah ekstraksi fitur dilakukan seleksi fitur menggunakan metode ABC. Seleksi fitur dilakukan untuk menyeleksi fitur sensor yang redundant sehingga diperoleh fitur yang mencirikan kualitas teh hitam paling spesifik. Seleksi fitur digunakan dengan pendekatan wrapper, dimana setiap kombinasi fitur yang terpilih dievaluasi dengan dimasukkan ke proses klasifikasi yang menggunakan metode SVM. Hasil seleksi fitur berupa kombinasi sensor yang dapat menghasilkan akurasi klasifikasi maksimal. Klasifikasi dilakukan menggunakan metode SVM. Terdapat dua tahap dalam proses klasifikasi yaitu, tahap pelatihan dan pengujian. Proses pelatihan bertujuan untuk mendapatkan nilai bobot yang mencirikan masing-masing sampel teh dan akan menjadi acuan sistem untuk melakukan pengelompokan. Setelah didapatkan nilai bobot yang tepat bagi semua sampel pelatihan, selanjutnya dilakukan proses pengujian dengan keluaran berupa hasil klasifikasi.

2. 2.1 Seleksi Fitur

Proses seleksi fitur yang dilakukan pada penelitian ini yaitu menggunakan algoritma ABC. Seleksi fitur dilakukan dengan pendekatan wrapper, dimana dalam pendekatan ini dibutuhkan suatu algoritma model pembelajaran dan seleksi fitur dilakukan dengan tujuan meningkatkan kemampuan generalisasi model pembelajaran tersebut [15]. Seleksi fitur dilakukan untuk menyeleksi fitur yang relevan serta signifikan berpengaruh terhadap klasifikasi sehingga dapat meningkatkan performa klasifikasi.

Berikut merupakan contoh alur proses seleksi fitur menggunakan algoritma ABC sesuai dengan masalah yang penulis angkat dengan nilai parameter berbeda.

Pertama, dilakukan inisialisasi parameter kontrol berupa colony size (C), maksimum siklus (T), dan limit. Contoh : $C = 6$; $T = 2$; $limit = 2$. Dimensi dari permasalahan yang akan diselesaikan adalah 12.

Kemudian dilakukan inisialisasi populasi awal (P) yang berisi nilai random yang merepresentasikan solusi dengan ukuran matriks $Np \times D$, dimana $Np = C/2$.

$$P = \begin{bmatrix} -0,5 & 0,1 & 1,2 & -2,3 & 0,4 & 1,7 & 0,8 & -0,3 & -1,7 & 1 & -0,4 & 1,5 \\ 2,1 & 1,5 & -1,7 & -0,2 & 0,8 & 3,5 & -1,2 & 0,8 & 3,7 & 0,5 & 2 & 0,9 \\ -3,5 & -0,7 & 2,4 & 1,3 & -0,2 & -2,2 & -0,6 & 1,4 & 0,7 & -0,1 & -1,6 & 3 \end{bmatrix}$$

Populasi awal diatas kemudian nilainya dijadikan representasi pemilihan sensor. Apabila bilangan pada populasi bernilai < 0 maka sensor tersebut tidak dipilih atau representasi nilai

tersebut menjadi biner 0. Apabila bernilai ≥ 0 , maka sensor tersebut dipilih atau representasi nilai tersebut menjadi biner 1.

$$\begin{aligned} y(1) &= [0 \ 1 \ 1 \ 0 \ 1 \ 1 \ 1 \ 0 \ 0 \ 1 \ 0 \ 1] \\ y(2) &= [1 \ 1 \ 0 \ 0 \ 1 \ 1 \ 0 \ 1 \ 1 \ 1 \ 1 \ 1] \\ y(3) &= [0 \ 0 \ 1 \ 1 \ 0 \ 0 \ 0 \ 1 \ 1 \ 0 \ 0 \ 1] \end{aligned}$$

Setelah nilai pada populasi awal telah direpresentasikan menjadi biner yang mewakili pemilihan sensor, dilakukan perhitungan objective function (f) dan fitness value (fit). Objective function pada permasalahan merupakan fungsi yang nilainya akan dioptimalkan yaitu akurasi klasifikasi SVM.

$$y(1) = [0 \ 1 \ 1 \ 0 \ 1 \ 1 \ 1 \ 0 \ 0 \ 1 \ 0 \ 1]$$

Sensor 2, 3, 5, 6, 7, 10 dan 12 dipilih

Klasifikasi dilakukan dengan data respon sensor yang digunakan hanya data respon sensor yang dipilih. Misal didapatkan akurasi sebagai berikut :

$$acc.train = 96$$

$$acc.test = 90$$

$$f = \frac{1}{acc.train + acc.test}$$

$$fit = \frac{1}{1 + obj.func}$$

$$f(1) = 0,0053$$

$$fit(1) = 0,9947$$

$$f(2) = 0,0062$$

$$fit(2) = 0,9938$$

$$f(3) = 0,0051$$

$$fit(3) = 0,9949$$

Berdasarkan nilai di atas solusi yang menghasilkan akurasi terbaik berada pada P(3). Selanjutnya masuk ke 3 fase pada algoritma ABC yaitu *employed bee*, *onlooker bee* dan *scout bee*. Pada fase *employed bee* dilakukan update solusi untuk semua solusi pada populasi awal. Proses pembentukan solusi baru dapat dilihat pada berikut.

$$P(1) = [-0,5 \ 0,1 \ 1,2 \ -2,3 \ 0,4 \ 1,7 \ 0,8 \ -0,3 \ -1,7 \ 1 \ -0,4 \ 1,5]$$

$$n = 1 \quad j = 1 \quad p = 3$$

Pertama dilakukan pemilihan secara random indeks variable yang akan diubah (j) serta solusi partner (p) untuk pembentukan solusi baru, n merupakan indeks solusi yang akan diperbarui. Setiap pemilihan p , dilakukan pengecekan nilai $p \neq n$ karena solusi partner tidak boleh sama dengan solusi yang akan diperbarui. Variabel \emptyset merupakan nilai random antara (-1,1).

$$x_{new} = x_{ij} + \emptyset(x_{ij} - x_{pj})$$

$$x_{new} = (-0,5) + 0,9((-0,5) - (-3,5))$$

$$x_{new} = 2,2$$

$$P(1) = [2,2 \ 0,1 \ 1,2 \ -2,3 \ 0,4 \ 1,7 \ 0,8 \ -0,3 \ -1,7 \ 1 \ -0,4 \ 1,5]$$

$$y(1) = [1 \ 1 \ 1 \ 0 \ 1 \ 1 \ 1 \ 0 \ 0 \ 1 \ 0 \ 1]$$

Kemudian dihitung objective function (f) dan fitness value (fit) dari kandidat solusi baru. Misal didapatkan sebagai berikut:

$$f(1) = 0,0051$$

$$fit(1) = 0,9949$$

Selanjutnya dilakukan *greedy selection* antara solusi lama dan kandidat solusi baru. Karena kandidat solusi baru nilai *fitness*-nya lebih besar dari solusi lama, maka kandidat solusi baru dimasukkan dalam populasi, sehingga populasi menjadi seperti berikut:

$$P = \begin{bmatrix} 2,2 & 0,1 & 1,2 & -2,3 & 0,4 & 1,7 & 0,8 & -0,3 & -1,7 & 1 & -0,4 & 1,5 \\ 2,1 & 1,5 & -1,7 & -0,2 & 0,8 & 3,5 & -1,2 & 0,8 & 3,7 & 0,5 & 2 & 0,9 \\ -3,5 & -0,7 & 2,4 & 1,3 & -0,2 & -2,2 & -0,6 & 1,4 & 0,7 & -0,1 & -1,6 & 3 \end{bmatrix}$$

Apabila kandidat solusi baru *fitness*-nya lebih kecil dari solusi lama, maka kandidat solusi tersebut tidak dimasukkan dalam populasi dan nilai *trial* solusi tersebut bertambah 1.

Pada fase *onlooker bee*, solusi yang akan diupdate dipilih berdasarkan nilai probability nya. Apabila probabilitynya lebih besar dari angka random yang sudah ditentukan, maka solusi tersebut diperbarui. Pada fase ini juga dilakukan *greedy selection*. Prosedur pada fase *onlooker bee* dapat dilihat sebagai berikut:

$$probability = a \times \frac{fit}{\max(fit)} + b, \quad a + b = 1$$

$$probability(2) = 0,9 \times \frac{0,9938}{0,9949} + 0,1$$

$$probability(2) = 0,999$$

$$rand = 0,851$$

Karena kondisi $rand < probability$ terpenuhi, maka solusi $P(2)$ akan diperbarui. Prosedur pembentukan solusi baru pada fase *onlooker bee* sama dengan fase *employed bee*.

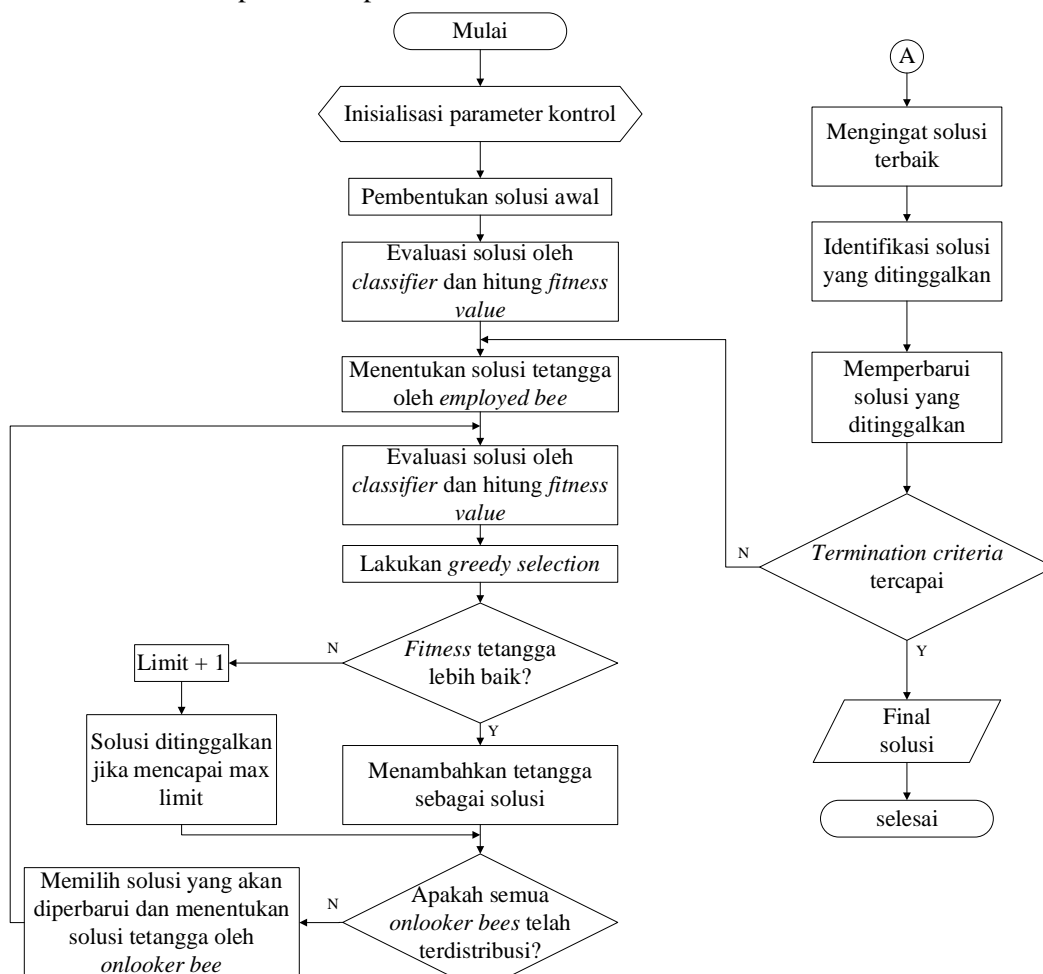
Pada fase *scout bee*, solusi yang diperbarui merupakan solusi yang telah melebihi nilai maximum limit dalam trial update solusi. Pada fase ini, dibentuk solusi baru secara random seperti pada inisialisasi populasi awal.

Inputan untuk sistem seleksi fitur merupakan matriks hasil ekstraksi ciri dengan ukuran 300×12 . Pada setiap fase ABC akan disimpan solusi yang menghasilkan akurasi paling bagus sehingga dicari nilai objective function minimum, karena objective function dihasilkan dari persamaan (1). Output dari seleksi fitur menggunakan ABC yaitu berupa fitur sensor terpilih yang menghasilkan objective function paling minimum. Persamaan (2) merupakan persamaan untuk menghitung fitness value.

$$f = \frac{1}{(accuracy.model.train + accuracy.model.test)} \quad (1)$$

$$fit_i = \begin{cases} \frac{1}{1+f_i} & \text{if } f_i \geq 0 \\ 1 + abs(f_i) & \text{if } f_i < 0 \end{cases} \quad (2)$$

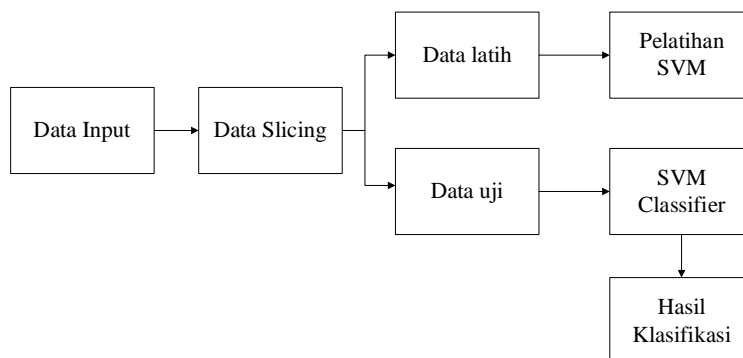
Jumlah koloni lebah pada ABC diinisiasikan sebesar 60, maksimum limit diinisiasikan sebesar 100 dan maksimum siklus sebesar 50 sebagai termination criteria. Flowchart rancangan sistem seleksi fitur dapat dilihat pada Gambar 4.



Gambar 4 Flowchart rancangan sistem seleksi fitur

2. 2.2 Klasifikasi

Pada penelitian ini proses klasifikasi dilakukan menggunakan algoritma support vector machine (SVM). Dalam proses klasifikasi terdapat 2 tahap yaitu pelatihan dan pengujian. Untuk itu pada proses ini digunakan holdout cross validation dengan partisi data latih 90% dan data uji 10%. Inputan pada proses klasifikasi terdapat 2 macam yaitu data hasil seleksi fitur dan data hasil ekstraksi fitur. Blok diagram rancangan sistem klasifikasi dapat dilihat pada Gambar 5.



Gambar 5 Diagram blok rancangan sistem klasifikasi

2. 2.3 Prosedur Pengumpulan Data

Data yang digunakan pada penelitian ini menggunakan data teh hitam dari penelitian sebelumnya [14]. Data yang digunakan merupakan pengukuran intensitas aroma menggunakan 12 array sensor yaitu MQ-7, TGS 2600, TGS 813, TGS 825, TGS 2602, TGS 826, TGS 2610, TGS 2611, TGS 832, TGS 2612, TGS 2620 dan TGS 822. Data yang digunakan diperoleh dari respon 12 sensor e-nose untuk teh hitam dengan tiga jenis kualitas, dimana setiap sampel teh (BPO, BOP-II dan Bohea) diolah menggunakan e-nose dengan waktu tertentu.

Interval pengambilan nilai respon sensor yang digunakan pada e-nose adalah setiap satu detik. Durasi fase collecting dan purging adalah 140 detik, sehingga nilai respon yang diambil dalam satu kali siklus penciuman adalah 140 nilai untuk setiap respon sensor. Setiap siklus sniffing menghasilkan data matriks yang berdimensi 140×12 (nilai respon sensor x jumlah sensor). Setiap kualitas terdiri dari 10 file dan dalam 1 file pembacaan sensor terdapat 10 siklus. Sehingga untuk setiap kualitas, data sampel berukuran 14000×12 .

Respon sensor kemudian di preproses dengan mengurangi data pada setiap siklus dengan baseline masing-masing siklus, sehingga hasil dari preprocessing berupa data yang berskala dengan nilai awal 0 untuk semua siklus.

Setelah preprocessing, data diekstraksi fiturnya untuk setiap siklus sniffing dengan diambil nilai maksimum pembacaan sensor. Sehingga dimensi data untuk setiap siklus menjadi 1×12 . Setiap kualitas menghasilkan data dengan dimensi 100×12 , sehingga total data hasil ekstraksi fitur berukuran 300×12 .

Data hasil ekstraksi fitur kemudian dijadikan input seleksi fitur dan juga input sistem klasifikasi tanpa seleksi fitur. Data dipartisi menggunakan holdout cross validation dengan partisi data latih 90% dan data uji 10%. Sehingga dari 300 data, 270 data digunakan untuk pelatihan dan 30 data untuk pengujian.

Output dari sistem klasifikasi dengan seleksi fitur berupa fitur sensor yang terpilih serta performa prediksi klasifikasi dengan fitur sensor yang digunakan hanya yang terpilih saja. Performa klasifikasi berupa nilai akurasi didapatkan dari confusion matrix.

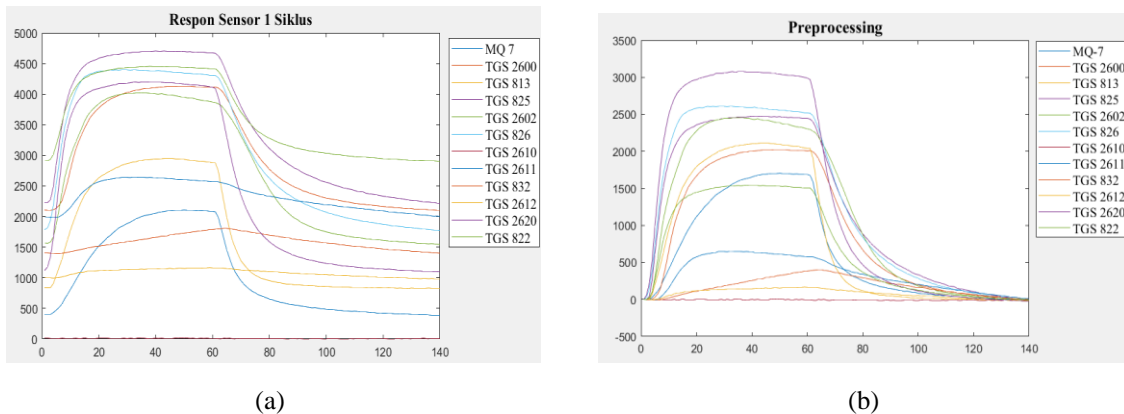
Output dari sistem klasifikasi tanpa seleksi fitur berupa performa prediksi klasifikasi dengan menggunakan semua fitur sensor yang berjumlah 12 sensor. Performa klasifikasi berupa

nilai akurasi didapatkan dari confusion matrix. Pada pengujian dilakukan variasi parameter SVM berupa fungsi kernel dan parameter ABC berupa jumlah koloni lebah serta variasi partisi pada holdout cross validation.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Preprocessing dan Ekstraksi Fitur

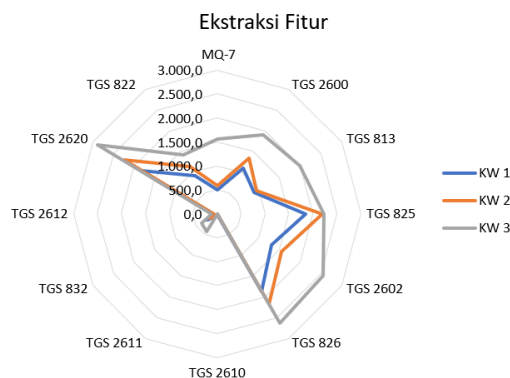
Preprocessing respon sensor dilakukan dengan manipulasi baseline. Metode yang digunakan yaitu metode difference, dimana nilai setiap sensor akan dikurangi dengan baseline masing-masing sensor. Hal ini dilakukan agar respon sensor memiliki skala awal yang sama, sehingga dapat dilihat perubahan nilai sensor. Hasil preprocessing dapat dilihat pada Gambar 6.



Gambar 6 Respon sensor (a) sebelum dan (b) setelah preprocessing

Pada Gambar 6 (b) dapat dilihat bahwa hasil dari preprocessing yaitu skala awal untuk setiap sensor menjadi sama untuk semua sensor yaitu pada nilai 0, sehingga respon sensor menjadi berskala. Preprocessing ini dilakukan untuk semua data yang digunakan pada penelitian.

Data hasil preprocessing kemudian diekstraksi fiturnya menggunakan metode maksimum. Metode maksimum dilakukan dengan mengambil nilai puncak dari masing-masing sensor. Dengan dilakukannya ekstraksi fitur, dimensi data yang sebelumnya 140×12 menjadi 1×12 untuk setiap siklus collecting dan purging. Total dimensi data hasil ekstraksi fitur menjadi 300×12 untuk ketiga kualitas teh. Hasil ekstraksi fitur dapat dilihat pada Gambar 7.



Gambar 7 Hasil ekstraksi fitur

Dari gambar di atas terlihat bahwa teh kualitas 1 dan 2 polanya cenderung sama, sedangkan untuk teh kualitas 3 polanya berbeda karena cenderung lebih besar dibandingkan dengan dua kualitas lainnya. Maka dalam proses klasifikasi, prediksi teh kualitas 3 akan lebih mudah dikarenakan polanya cukup berbeda dibandingkan teh kualitas 1 dan 2 yang polanya mirip.

3.2 Hasil Pengujian Sistem

Seleksi fitur dilakukan dengan control parameter sesuai dengan rancangan sistem dimana colony lebah berjumlah 60, limit 100 dan maksimum iterasi 50. Seleksi fitur dilakukan pada data hasil ekstraksi fitur. Proses pengujian sistem dilakukan dengan 4 data partisi yang berbeda untuk variasi kernel SVM dan variasi colony size. Pertama dilakukan pengujian dengan $C = 60$ untuk 4 data variasi, kemudian dilakukan pengujian dengan variasi $C = 20$, $C = 100$ dan $C = 200$. Hal ini dilakukan untuk mengetahui apakah sistem yang dibuat menghasilkan performa yang konsisten pada kombinasi data uji dan colony size yang berbeda, serta mengetahui bagaimana pengaruh variasi tersebut terhadap performa sistem.

Setelah dilakukan pengujian dengan 4 variasi colony size dan partisi data, diperoleh hasil bahwa masing-masing data partisi memberikan hasil optimal pada colony size yang berbeda. Data partisi 1 dan 2 memberikan hasil optimal pada $C = 200$, data partisi 3 pada $C = 60$ dan data partisi 4 mampu memberikan hasil optimal pada $C = 20$. Oleh karena itu akan dipakai hasil pada $C = 200$ dikarenakan pada variasi tersebut keempat data partisi sudah menghasilkan performa paling optimal. Hasil variasi kernel SVM pada 4 data partisi dengan $C = 200$ dapat dilihat pada Tabel 1-4.

Tabel 1 Variasi Kernel pada Data Partisi 1

Variasi Kernel	ABC-SVM		SVM		Best Value
	TRAIN	TEST	TRAIN	TEST	
linear	75,6%	80%	78,1%	63,3%	2-3-6-8-9-12
rbf	97,4%	80%	100%	56,7%	1-3-4-5-8-9-11-12
polynomial	99,6%	86,7%	100%	76,7%	1-3-4-6-7-9-10-11-12

Tabel 2 Variasi Kernel pada Data Partisi 2

Variasi Kernel	ABC-SVM		SVM		Best Value
	TRAIN	TEST	TRAIN	TEST	
linear	80%	83,3%	81,1%	66,7%	1-2-3-4-5-8-9-10-11-12
rbf	96,7%	90%	100%	76,7%	1-2-3-4-5-6-8-9
polynomial	99,6%	96,7%	99,6%	70%	2-3-4-5-6-8-9-10-11-12

Tabel 3 Variasi Kernel pada Data Partisi 3

Variasi Kernel	ABC-SVM		SVM		Best Value
	TRAIN	TEST	TRAIN	TEST	
linear	75,6%	90,0%	77,8%	80,0%	1-2-4-7-9-10-11-12
rbf	95,2%	87%	100%	66,7%	1-2-4-7-8-11-12
polynomial	98,5%	96,7%	100%	76,7%	2-4-5-6-8-9-10-11-12

Tabel 4 Variasi Kernel pada Data Partisi 4

Variasi Kernel	ABC-SVM		SVM		Best Value
	TRAIN	TEST	TRAIN	TEST	
linear	74,8%	90%	77,4%	80%	1-2-3-4-6-9-11-12
rbf	98,9%	90%	100%	73,3%	1-2-4-5-6-8-9-11-12
polynomial	98,5%	96,7%	100%	80%	2-3-4-5-8-9-10-11-12

Variasi kernel dilakukan untuk 3 kernel SVM yaitu linear, rbf dan polynomial. Berdasarkan variasi kernel dapat dilihat bahwa kernel yang menghasilkan performa baik pada sistem adalah kernel rbf dan polynomial dengan jumlah fitur yang dikurangi berkisar antara 3-5 fitur. Rata-rata performa sistem hasil variasi untuk 4 data partisi dapat dilihat pada Tabel 5.

Tabel 5 Rata-Rata Performa Sistem Hasil Variasi pada 4 Data Partisi

Variasi	ABC-SVM	SVM
---------	---------	-----

Kernel	TRAIN	TEST	TRAIN	TEST
linear	76,5%	85,8%	78,6%	72,5%
rbf	97,1%	87%	100%	68,4%
polynomial	99%	94,2%	99,9%	75,9%

Berdasarkan Tabel 5 dapat dilihat bahwa saat kernel polynomial digunakan kemampuan prediksi sistem ABC-SVM menghasilkan performa paling bagus dengan rata-rata akurasi pengujian 94,2%. Kemudian dilakukan variasi partisi data pada holdout cv yaitu untuk data latih sebanyak 70%, 80% dan 90%. Hasil variasi dapat dilihat pada Tabel 6.

Tabel 4 Variasi Holdout CV

Variasi Holdout CV	ABC-SVM		SVM		Best Value
	TRAIN	TEST	TRAIN	TEST	
70% latih	98,6%	80%	100%	68,9%	1-2-3-4-5-6-9-11-12
80% latih	96,7%	85%	100%	73,3%	1-3-4-5-6-9-11-12
90% latih	98,5%	96,7%	100%	80%	2-3-4-5-8-9-10-11-12

Variasi partisi data pada holdout CV dilakukan dengan parameter lainnya menggunakan parameter terbaik yang telah didapatkan pada variasi sebelumnya. Dari Tabel 6.6 dapat dilihat bahwa variasi partisi dengan data latih 90% menghasilkan akurasi prediksi yang lebih baik dibandingkan dengan partisi data latih 70% dan 80%. Pada partisi data latih 90% sistem ABC-SVM menghasilkan akurasi prediksi yang paling baik dengan jumlah fitur yang dikurangi berjumlah 3 fitur.

Berdasarkan variasi yang telah dilakukan maka kombinasi parameter dengan hasil performa sistem paling optimal yaitu pada $C = 200$, kernel polynomial dan holdout CV dengan data latih 90%. Hasil performa sistem paling optimal dapat dilihat pada Gambar 8 dan 9

<i>fitur terbaik : 2 3 4 5 8 9 10 11 12</i>					
$cm =$	90	3	0	$cm =$	8 0 0
	1	87	0		1 10 0
	0	0	89		0 0 11
<i>akurasi data train ABC – SVM = 98,5%</i>			<i>akurasi data test ABC – SVM = 96,7%</i>		

Gambar 8 Hasil Sistem Klasifikasi dengan Seleksi Fitur

$cm =$	91	0	0	$cm =$	7 4 0
	0	90	0		2 6 0
	0	0	89		0 0 11
<i>akurasi data train SVM = 100%</i>			<i>akurasi data test SVM = 80%</i>		

Gambar 9 Hasil Sistem Klasifikasi tanpa Seleksi Fitur

Berdasarkan Gambar 8 dan 9 dapat dilihat bahwa sistem klasifikasi dengan seleksi fitur berhasil meningkatkan performa klasifikasi yang ditunjukkan pada akurasi data test yaitu dari nilai 80% meningkat menjadi 96,7%. Peningkatan akurasi sebesar 16,7% dengan hasil fitur terbaik berjumlah 9 yaitu sensor 2, 3, 4, 5, 8, 9, 10, 11 dan 12.

Maka sensor-sensor yang paling mencirikan teh hitam kualitas 1-3 yaitu sensor TGS 2600, TGS 813, TGS 825, TGS 2602, TGS 2611, TGS 832, TGS 2612, TGS 2620 dan TGS 822. Sedangkan untuk sensor MQ-7, TGS 826 dan TGS 2610 merupakan sensor yang redundant pada sistem dikarenakan gas yang dideteksi oleh 3 sensor tersebut dapat diwakili oleh sensor lainnya.

4. KESIMPULAN

Berdasarkan hasil pengujian sistem dapat diambil kesimpulan bahwa telah berhasil dirancang sebuah sistem perangkat lunak yang dapat melakukan seleksi fitur untuk optimasi

klasifikasi data teh hitam. Seleksi fitur oleh algoritma Artificial Bee Colony berhasil menyeleksi fitur yang redundant dengan mengurangi jumlah fitur dari 12 menjadi 8 fitur. Peningkatan performa yang diperoleh pada sistem klasifikasi dengan seleksi fitur yaitu sebanyak 16,7%, dengan nilai akurasi sistem 96,7% dibandingkan dengan sistem klasifikasi tanpa seleksi fitur dengan nilai akurasi 80%.

5. SARAN

Dilakukan validasi menggunakan k-fold cross validation agar tidak perlu dilakukan running beberapa kali untuk menguji sistem dengan data partisi yang berbeda, namun diperhatikan waktu komputasinya.

Mencoba algoritma optimasi yang lain dengan performa lebih baik dari ABC atau sama, dikarenakan algoritma ABC memerlukan waktu pemrosesan yang lama serta spesifikasi processor yang tinggi. Pemilihan algoritma yang lain dilakukan agar dapat mengefisiensi waktu pemrosesan dengan hasil performa yang baik.

Penelitian selanjutnya juga dapat mengimplementasikan sistem ABC-SVM untuk kasus yang lain untuk mengetahui bagaimana performa algoritma dalam kasus lainnya.

DAFTAR PUSTAKA

- [1] N. Bhattacharyya, B. Tudu, R. Bandyopadhyay, M. Bhuya, and R. Mudi, "Aroma characterization of orthodox black tea with electronic nose," *IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON*, vol. B, pp. 427–430, 2004, doi: 10.1109/tencon.2004.1414623. <https://ieeexplore.ieee.org/abstract/document/1414623> [Accessed: 23-Nov-2020]
- [2] A. Dutta, B. Tudu, R. Bandyopadhyay, and N. Bhattacharyya, "Black tea quality evaluation using electronic nose: An artificial bee colony approach," *2011 IEEE Recent Adv. Intell. Comput. Syst. RAICS 2011*, no. 2, pp. 143–146, 2011, doi: 10.1109/RAICS.2011.6069290. <https://ieeexplore.ieee.org/document/6069290> [Accessed: 23-Nov-2020]
- [3] R. Banerjee, P. Chattopadhyay, B. Tudu, N. Bhattacharyya, and R. Bandyopadhyay, "Artificial flavor perception of black tea using fusion of electronic nose and tongue response: A Bayesian statistical approach," *J. Food Eng.*, vol. 142, pp. 87–93, 2014, doi: 10.1016/j.jfoodeng.2014.06.004. https://www.sciencedirect.com/science/article/abs/pii/S0260877418304102?casa_token=WUzxSYUrZmEAAAAA:oZMha9xUU95Oj0w_EJo2AEBAtURTPYyPrETeZ8Np3i6yfZddyNjGsn0rAoHsboAXHmaGpzZuaQ [Accessed: 14-Nov-2019]
- [4] R. Dutta, E. L. Hines, J. W. Gardner, K. R. Kashwan, and M. Bhuyan, "Tea quality prediction using a tin oxide-based electronic nose: An artificial intelligence approach," *Sensors Actuators, B Chem.*, vol. 94, no. 2, pp. 228–237, 2003, doi: 10.1016/S0925-4005(03)00367-8. https://www.researchgate.net/publication/223495802_Tea_Quality_Prediction_Using_a_Tin_Oxide-based_Electronic_Nose_an_Artificial_Intelligence_Approach [Accessed: 26-Des-2020]
- [5] B. Santosa and D. R. Hanum, "Studi komparasi metode klasifikasi dua kelas," *Pros. Semin. Nas. Manaj. Teknol. V*, 2007. <https://adoc.pub/studi-komparasi-metode-klasifikasi-dua-kelas.html> [Accessed: 27-Sep-2019]
- [6] Q. Chen, J. Zhao, Z. Chen, H. Lin, and D. A. Zhao, "Discrimination of green tea quality using the electronic nose technique and the human panel test, comparison of linear and nonlinear classification tools," *Sensors Actuators, B Chem.*, vol. 159, no. 1, pp. 294–300, 2011, doi: 10.1016/j.snb.2011.07.009. <https://www.sciencedirect.com/science/article/abs/pii/S0925400511006393> [Accessed: 4-Okt-2019]

- [7] S. Jenicka and A. Suruliandi, "Comparative study of classification algorithms with modified multivariate local binary pattern texture model on remotely sensed images," *Int. Conf. Recent Trends Inf. Technol. ICRTIT 2011*, pp. 848–852, 2011, doi: 10.1109/ICRTIT.2011.5972312. <https://ieeexplore.ieee.org/abstract/document/5972312> [Accessed: 27-Sep-2019]
- [8] R. Prakash, V. P. Tharun, and S. Renuga Devi, "A Comparative Study of Various Classification Techniques to Determine Water Quality," *Proc. Int. Conf. Inven. Commun. Comput. Technol. ICICCT 2018*, no. Iccct, pp. 1501–1506, 2018, doi: 10.1109/ICICCT.2018.8473168. <https://ieeexplore.ieee.org/document/8473168> [Accessed: 17-Sep-2019]
- [9] Y. T. Liu and K. T. Tang, "A Minimum Distance Inlier Probability (MDIP) Feature Selection Method to Improve Gas Classification for Electronic Nose Systems," *IEEE Access*, vol. 8, pp. 133928–133935, 2020, doi: 10.1109/ACCESS.2020.3010788. <https://ieeexplore.ieee.org/abstract/document/9145749> [Accessed: 27-Dec-2020]
- [10] Y. Lu, I. Cohen, X. S. Zhou, and Q. Tian, "Feature selection using principal feature analysis," in *Proceedings of the 15th international conference on Multimedia - MULTIMEDIA '07*, 2007, p. 301, doi: 10.1145/1291233.1291297. <https://ieeexplore.ieee.org/document/5640135> [Accessed: 4-Nov-2019]
- [11] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm," *J. Glob. Optim.*, vol. 39, no. 3, pp. 459–471, 2007, doi: 10.1007/s10898-007-9149-x. https://www.researchgate.net/publication/225392029_A_powerful_and_efficient_algorithm_for_numerical_function_optimization_Artificial_bee_colony_ABC_algorithm [Accessed: 5-Apr-2019]
- [12] B. Basu and G. K. Mahanti, "A comparative study of modified particle swarm optimization, differential evolution and artificial bees colony optimization in synthesis of circular array," *ICPCES 2010 - Int. Conf. Power, Control Embed. Syst.*, pp. 1–5, 2010, doi: 10.1109/ICPCES.2010.5698614. https://www.researchgate.net/publication/251987264_A_comparative_study_of_Modified_Particle_Swarm_Optimization_Differential_Evolution_and_Artificial_Bees_Colony_optimization_in_synthesis_of_circular_array [Accessed: 10-Sep-2019]
- [13] M. Meguellati, F. Srairi, F. Djeflal, and L. Saidi, "Performance analysis of swimming microrobot using GA, ABC and PSO based-optimization techniques," *2015 4th Int. Conf. Syst. Control. ICSC 2015*, pp. 310–314, 2015, doi: 10.1109/ICoSC.2015.7153277. <https://ieeexplore.ieee.org/document/7153277> [Accessed: 13-Sep-2019]
- [14] D. Lelono, "Pengembangan Instrumentasi Sistem Electronic Nose untuk Uji Teh Hitam Lokal. Universitas Gadjah Mada," Universitas Gadjah Mada, 2017.
- [15] M. Monirul Kabir, M. Monirul Islam, and K. Murase, "A new wrapper feature selection approach using neural network," *Neurocomputing*, vol. 73, no. 16–18, pp. 3273–3283, 2010, doi: 10.1016/j.neucom.2010.04.003. <https://www.sciencedirect.com/science/article/abs/pii/S0925231210001979> [Accessed: 11-Dec-2020]