

DOI: doi.org/10.21009/03.SNF2020.01.FA.22

# ANALISIS CLUSTER BAHAN HERBAL BERDASARKAN FITUR RESPON E-NOSE

Anggiyansah Sitompul<sup>a)</sup>, Bambang Heru Iswanto<sup>b)</sup>,

Widyaningrum Indrasari<sup>c)</sup>

*Program Studi Fisika, FMIPA, Universitas Negeri Jakarta, Jalan Pemuda No. 10 Jakarta Timur, 13220, Indonesia*

Email: <sup>a)</sup>anggiyansahsitompul\_3225164873@mhs.unj.ac.id, <sup>b)</sup>bhi@unj.ac.id,  
<sup>c)</sup>widyaningrum-indrasari@unj.ac.id

## Abstrak

Electronic nose (e-nose) merupakan alat yang tersusun atas berbagai sensor kimia elektronik dengan sensitivitas parsial dan sistem pengenalan pola yang tepat serta mampu mengenali bau yang sederhana maupun kompleks. Dalam perkembangannya, e-nose berfungsi menggantikan keterbatasan hidung manusia dalam mengenali aroma tertentu secara cepat dan tepat. Namun, e-nose yang terdiri dari sejumlah larik sensor menghasilkan data yang sangat besar sehingga membutuhkan metode ekstraksi fitur yang tepat dalam mengenali pola dari respons e-nose. Data respon e-nose terhadap lima bahan herbal yang terdiri dari jahe (ZO), kencur (KG), kunyit (CL), lengkuas (LG), dan temulawak (CX) telah dianalisis dalam penelitian ini. Dua metode ekstraksi fitur, yaitu relative amplitude (RA) dan surface (S) digunakan untuk mendapatkan fitur terbaik untuk clustering data respon e-nose kelima bahan herbal tersebut. Pada proses analisis data, metode cluster analysis yaitu k-means clustering digunakan untuk clustering dataset respons yang diekstraksi menggunakan metode RA, dan S. Dua kriteria eksternal validasi cluster yaitu entropy dan purity digunakan dalam mengukur kualitas cluster yang dihasilkan. Nilai entropy minimum pada penelitian ini adalah 0,53 diperoleh pada fitur RA dan purity maksimum adalah 0,83 yang diperoleh pada fitur RA. Dari hasil tersebut, fitur yang lebih efektif dalam menghasilkan solusi cluster terbaik untuk membedakan kelima bahan herbal adalah fitur RA.

**Kata-kata kunci:** *electronic nose*, bahan herbal, ekstraksi fitur, *cluster analysis*.

## Abstract

Electronic nose (e-nose) is a device composed of various electronic chemical sensors with partial sensitivity and a precise pattern recognition system capable of recognizing simple and complex odors. In its development, the e-nose serves to replace the limitations of the human nose in recognizing certain aromas quickly and precisely. However, the e-nose which consists of a number of sensor arrays produces very large data, so it requires the right feature extraction method in recognizing the pattern of the e-nose response. E-nose response data to five herbal ingredients consisting of *Zingiber officinale* (ZO), *Kaempferia galanga* (KG), *Curcuma longa* (CL), *Languas galanga* (LG), and *Curcuma xanthorrhiza roxb* (CX) were analyzed in this study. Two feature extraction methods, namely relative amplitude (RA) and surface (S), were used to obtain the best features for clustering the e-nose response data of the five herbal ingredients. In the data analysis process, the cluster analysis method, namely k-means clustering, was used for clustering the response dataset which was extracted using the RA and S

methods. Two external criteria for cluster validation, namely entropy and purity, were used to measure the quality of the resulting clusters. The minimum entropy value in this study was 0.53 obtained for the RA feature and the maximum purity was 0.83 obtained for the RA feature. From these results, the feature that is more effective in producing the best cluster solution to differentiate the five herbal ingredients is the RA feature.

**Keywords:** electronic nose, herbal ingredients, feature extraction, cluster analysis.

## PENDAHULUAN

Tanaman obat dan produknya telah digunakan sebagai sumber obat hampir di semua budaya sejak dahulu kala. Di Indonesia, beberapa spesies tanaman digunakan sebagai obat dan dalam beberapa tahun terakhir, perdagangan tanaman obat meningkat beberapa kali lipat. Perkembangan perdagangan tanaman obat sangat dipengaruhi oleh perkembangan perdagangan obat tradisional dan perkembangan industri obat tradisional. Ada beberapa aspek penting yang perlu diperhatikan dalam pengembangan bahan obat herbal, salah satunya kandungan berbagai senyawa aktif di dalamnya. Senyawa aktif herbal terkenal karena berguna untuk kesehatan manusia seperti antibiotik, antikanker, anti-penuaan, penyakit anti-hepatitis, perawatan pasca melahirkan [1]. Jenis bahan obat herbal yang beragam dengan aroma yang mirip menjadi tantangan tersendiri bagi produsen jamu tradisional maupun kemasan dalam mengidentifikasi bahan obat herbal secara akurat. Indera penciuman manusia tidak mampu untuk mengenali masing-masing bahan obat herbal sebab bau yang ditimbulkan sangat identik. Saat ini, analisis kuantitatif herbal biasanya dilakukan dengan menggunakan instrumen analitis seperti GC / MS dan spektrofotometer FTIR [2]. Metode-metode tersebut mahal dan juga rumit. Oleh sebab itu, dibutuhkan alat deteksi cepat dan mudah yang bernama electronic nose untuk membedakan setiap bahan obat herbal.

Electronic nose merupakan alat yang tersusun atas berbagai sensor kimia elektronik dengan sensitivitas parsial dan sistem pengenalan pola yang tepat serta mampu mengenali bau yang sederhana maupun kompleks [3]. Konsep electronic nose sebagai alat untuk mengklasifikasikan bau diperkenalkan pertama kali oleh Persaud dan Dodd pada tahun 1982. Pengembangan electronic nose untuk membedakan dan mengklasifikasikan sampel berdasarkan deteksi bau telah banyak dilakukan. Seiring berjalannya waktu pengaplikasian electronic nose semakin luas digunakan pada berbagai objek. Pada penelitian sebelumnya electronic nose telah digunakan untuk menguji makanan [4], produk pertanian [5], urin manusia [6], dan bahan herbal [7]. Dari penelitian sebelumnya, permasalahan paling umum yang dihadapi praktisi e-nose adalah menangani ukuran dimensi data keluaran e-nose yang sangat besar. Hal ini berhubungan dengan pemilihan metode ekstraksi fitur yang tepat. Tujuan dari ekstraksi fitur adalah untuk mengekstrak informasi yang kuat dari karakteristik respon sensor dengan redundansi yang lebih sedikit, yang dapat mewakili pola yang berbeda dengan baik. Selain untuk mereduksi dimensi data, metode ekstraksi fitur juga bertanggung jawab terhadap tingkat keefektifan algoritma pengenalan pola yang digunakan selanjutnya [8]. Oleh karena itu, pemilihan metode ekstraksi yang tepat akan mampu membedakan objek yang diteliti menggunakan electronic nose.

## METODOLOGI

Metode yang digunakan pada penelitian ini adalah metode penelitian sekunder atau dikenal juga dengan metode penelitian analisis data sekunder. Data yang digunakan pada penelitian ini diperoleh dari penelitian [9] berupa respon *e-nose* terhadap lima macam bahan herbal. Adapun sampel yang digunakan berjumlah 100 sampel dari lima buah bahan obat herbal, yang terdiri dari jahe (ZO), kencur (KG), kunyit (CL), lengkuas (LG), dan temulawak (CX).

Secara garis besar penelitian ini melakukan *clustering* data yaitu respon *e-nose* sampel lima macam bahan herbal dan membandingkan metode ekstraksi fitur respon *e-nose* sensor seperti *relative amplitude* (RA) dan *surface* (S). Hasil *cluster* dari masing-masing fitur kemudian dianalisa menggunakan *cluster validity* dengan kriteria *entropy* dan *purity* sehingga dapat diketahui fitur

terbaik untuk mengidentifikasi kelima macam bahan herbal tersebut. Adapun metode yang ditempuh dalam penelitian ini adalah sebagai berikut :

### Ekstraksi fitur

Ekstraksi fitur merupakan proses mengekstrak informasi yang penting dari karakteristik respon sensor dengan redundansi yang lebih sedikit, yang dapat mewakili pola yang berbeda dengan baik, untuk memastikan keefektifan algoritma pengenalan pola berikutnya.

#### a. Relative Amplitude (RA)

Metode Relative Amplitude (RA) mengekstraksi data dari hasil respon sensor-sensor pada e-nose dengan cara menghitung selisih dari nilai maksimum dan nilai minimum kemudian hasilnya dibagi dengan nilai minimum. Untuk memperoleh hasil ekstraksi dari metode RA dapat menggunakan persamaan berikut [10]:

$$RA = \frac{R_{\max} - R_o}{R_o} \quad (1)$$

#### b. Surface (S)

Pada metode ini, luasan yang berada dibawah kurva dari respon sensor dihitung untuk mendapatkan ciri dari respon sensor e-nose. Persamaan yang digunakan metode surface adalah sebagai berikut [11]:

$$S = \sum_{t_1}^{t_2} R(t)\Delta t = \sum_{t_1}^{t_2} V(t)\Delta t \quad (2)$$

R(t) merupakan respon sensor yang digambarkan sebagai fungsi waktu, sedangkan  $\Delta t$  merupakan selisih waktu dalam 1 detik.

### Cluster Analysis (CA)

Cluster Analysis termasuk dalam teknik analisis multivariate metodh interdependen yang mewakili seluruh hubungan interdependensi, tidak ada perbedaan variabel bebas dan variabel tak bebas (independent and dependent variables) dalam analisis ini. Teknik ini bertujuan untuk mengatur informasi / meringkas data dengan cara mengelompokkan objek-objek berdasarkan kesamaan karakteristik tertentu. Dengan demikian, analisis cluster atau clustering memiliki tujuan untuk mengelompokkan data dari serangkaian pengamatan ke subset disebut cluster [12]. Banyak metode clustering yang telah dikembangkan dan masing-masing metode menggunakan prinsip yang berbeda-beda [13]. Dalam penelitian ini menggunakan metode yang paling umum yaitu metode K-means clustering. Algoritma K-means clustering ini akan mengelompokkan data atau objek ke dalam k buah cluster, setiap cluster terdapat titik pusat (centroid) yang merepresentasikan cluster tersebut. Pusat dari setiap cluster dapat dihitung berdasarkan nilai rata-rata dari semua objek data milik setiap cluster. Dalam menentukan nilai pusat dapat menggunakan persamaan berikut :

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q \quad (3)$$

$\mu_k$  merupakan titik *centroid* dari *cluster* ke-k,  $N_k$  adalah banyaknya data pada *cluster* ke-k, dan  $x_q$  data ke-q pada *cluster* ke-k. Berikut tahapan algoritma K-Means *clustering* :

- i Membuat partisi sejumlah  $k$  dari segmentasi yang akan dibentuk, dalam penelitian ini  $k=5$ .
- ii Inisialisasi  $K$  pusat *cluster* (*centroid*) sebagai *seed points* (prototipe *cluster* awal). *Centroid* ini dapat diperoleh secara acak atau dipilih dari  $K$  objek data pertama.
- iii Menghitung jarak setiap data yang lain dengan pusat *cluster*.
- iv Mengisi setiap objek dalam himpunan data ke dalam pusat *cluster* terdekat.
- v Kalkulasi dan ubah kembali *centroid* tiap *cluster* sebagai rata-rata dari seluruh anggota kelompok (group) *cluster* tersebut.

vi Ulangi langkah c sampai e hingga data di dalam *cluster* (anggota kelompok) tidak berubah (konvergen).

### Validasi Cluster

Pada penelitian ini, dua kriteria validasi cluster digunakan, yaitu validasi cluster entropy dan purity [14]. Validasi cluster entropy merupakan kriteria cluster yang melihat bagaimana setiap kelas-kelas yang telah ditentukan sebelumnya terdistribusi ke dalam setiap cluster. Untuk mengukur entropy dari suatu cluster  $S_r$  dengan ukuran cluster  $n_r$  digunakan persamaan berikut:

$$E(S_r) = -\frac{1}{\log 2} \sum_{i=1}^q \frac{n_r^i}{n_r} \log \frac{n_r^i}{n_r} \quad (4)$$

$q$  merupakan jumlah kelas dalam data,  $n_r^i$  adalah jumlah anggota kelas ke- $i$  dalam cluster ke- $r$ . Entropy keseluruhan cluster diberikan pada persamaan berikut:

$$entropy = \sum_{r=1}^k \frac{n_r}{n} E(S_r) \quad (5)$$

$k$  merupakan jumlah cluster. Solusi cluster dikatakan terbaik apabila setiap cluster yang ada hanya terisi objek-objek dari satu kelas. Semakin kecil nilai entropy (minimum = 0) yang diperoleh maka semakin baik solusi cluster tersebut. Sedangkan pada validasi cluster purity, masing-masing cluster akan ditentukan kelas mana yang paling dominan dalam masing-masing cluster berdasarkan kelas dengan jumlah anggota terbanyak yang masuk ke dalam cluster. Kemudian keakuratan dari penentuan kelas dominan akan diukur dengan cara membagi jumlah anggota kelas dominan dengan jumlah seluruh objek atau data. Untuk lebih jelas, persamaan yang digunakan dalam mengukur purity ditunjukkan pada persamaan berikut:

$$purity(\Omega, C) = \frac{1}{N_k} \sum_k \max_j |\omega_k \cap c_j| \quad (6)$$

Dengan  $\Omega = \{\omega_1, \omega_2, \dots, \omega_k\}$  adalah set clusters dan  $C = \{c_1, c_2, \dots, c_k\}$  adalah set kelas. Secara umum, semakin besar nilai purity (maksimum = 1), semakin baik cluster tersebut.

### HASIL DAN PEMBAHASAN

Hasil clustering dari kedua fitur dengan metode K-means clustering kemudian divalidasi berdasarkan nilai entropy dan purity. Entropy dan purity memvalidasi anggota dari cluster-cluster yang terbentuk apakah hanya terdiri dari anggota kelas yang sama berdasarkan anggota kelas yang sudah ditentukan sebelumnya (Jahe, Kencur, Kunyit, Lengkuas, dan Temulawak).

**TABEL 1.** Hasil evaluasi *cluster* K-means clustering dengan fitur RA dan S dari respon *e-nose* terhadap 5 macam bahan herbal

Metode Ekstraksi Fitur	E	P	K
RA	0.53	0.83	5
S	0.57	0.80	5

Keterangan: E = entropy, P = purity, K = jumlah cluster pada K-Means

TABEL 1. menampilkan hasil evaluasi cluster berupa nilai entropy dan purity dari solusi cluster yang dihasilkan oleh K-means clustering dengan fitur RA dan S terhadap respon e-nose lima bahan herbal. Nilai k merupakan jumlah cluster yang ditentukan  $k=5$  sesuai dengan banyak kelas atau jenis bahan herbal yang digunakan yaitu jahe (ZO), kencur (KG), kunyit (CL), lengkuas (LG), dan temulawak (CX). Cluster yang dihasilkan oleh metode K-means clustering terhadap lima bahan herbal dengan fitur RA memiliki nilai entropy 0,53 dan purity 0,83. Sedangkan cluster yang dihasilkan oleh metode K-means clustering terhadap lima bahan herbal dengan fitur S memiliki nilai

entropy 0,57 dan purity 0,80. Nilai entropy yang tidak mencapai minimum 0 dan purity tidak mencapai maksimum 1 masih menunjukkan bahwa sampel belum semuanya terbedakan berdasarkan kelas awal di mana masing-masing bahan sampel herbal berasal. Seperti yang diketahui bahwa kriteria entropy mengukur homogenitas dari setiap cluster yang terbentuk dengan kelas yang sudah ditentukan sebelumnya. Ketika cluster yang terbentuk terisi oleh hanya satu kelas yang sama, maka nilai entropy dari cluster tersebut adalah 0. Sedangkan purity dari cluster merupakan nilai anggota kelas paling dominan dari setiap cluster kemudian dibagi dengan jumlah seluruh objek atau data. Dalam purity, semakin besar nilai purity (maksimum = 1) semakin baik cluster tersebut.

Berdasarkan kriteria entropy, semakin kecil nilai entropy (minimum = 0) yang diperoleh maka semakin baik solusi cluster tersebut. Sedangkan kriteria purity, semakin besar nilai purity (maksimum = 1), semakin baik cluster tersebut. Dapat dilihat pada tabel 1, nilai entropy terkecil dan purity terbesar diberikan highlight berwarna hijau. Nilai entropy fitur RA lebih mendekati 0 dibandingkan nilai entropy fitur S. Nilai purity fitur RA lebih mendekati 1 dibandingkan purity fitur S. Hal tersebut menunjukkan bahwa distribusi sampel di ruang fitur cukup lebih baik ketika dilakukan clustering menggunakan fitur RA dibandingkan fitur S.

### SIMPULAN

Pada penelitian ini telah dianalisis cluster lima jenis bahan herbal berdasarkan respon e-nose menggunakan fitur RA dan S. Telah dibuktikan bahwa metode tersebut dapat mengekstrak fitur respon e-nose dan mengurangi dimensi data ke dalam bentuk yang lebih spesifik secara efektif. Dari hasil evaluasi cluster menunjukkan bahwa metode ekstraksi fitur relative amplitude (RA) memberikan solusi cluster terbaik terhadap pengelompokan lima jenis bahan herbal dengan K-means clustering berdasarkan kriteria entropy dan purity.

### REFERENSI

- [1] S. Evans, "Changing the knowledge base in Western herbal medicine," *Soc. Sci. Med.*, vol. 67, no. 12, pp. 2098-2106, 2008.
- [2] C. M. Buré dan N. M. Sellier, "analysis of the essential oil of indonesian patchouli (*pogostemon cabin benth.*) using GC/MS (EI/CI)," *J. Essent. Oil Res.*, vol. 16, no. 1, pp. 17-19, 2004.
- [3] L. Capelli, S. Sironi dan R. Del Rosso, "Electronic noses for environmental monitoring applications," *Sensors (Switzerland)*, vol. 14, no. 11, pp. 19979-20007, 2014.
- [4] H. Tian, F. Li, L. Qin, H. Yu dan X. Ma, "Quality Evaluation of Beef Seasonings Using Gas Chromatography-Mass Spectrometry and Electronic Nose: Correlation with Sensory Attributes and Classification According to Grade Level," *Food Anal. Methods*, vol. 8, no. 6, pp. 1522-1534, 2015.
- [5] S. Srivastava, G. Mishra dan H. N. Mishra, "Fuzzy controller based E-nose classification of *Sitophilus oryzae* infestation in stored rice grain," *Food Chem.*, vol. 283, pp. 604-610, 2019.
- [6] N. Firmawati dan K. Triyana, "Kelayakan Teknologi Electronic Nose untuk Mendeteksi Urin yang Mengandung Metadon Dengan Menggunakan Principal Component Analysis (PCA)," vol. 8, no. 1, pp. 45-51, 2016.
- [7] F. Hardoyono, B. H. Iswanto dan K. Triyana, "Comparative analysis of feature extraction methods in the clustering of electronic nose response correlated with GC/MS analysis," *AIP Conf. Proc.*, vol. 1755, July 2016.
- [8] J. Yan, X. Guo, S. Duan, P. Jia, L. Wang, C. Peng dan Songlin Zhang, "Electronic Nose Feature Extraction Methods: A Review," pp. 27804-27831, 2015.

- [9] F. Hardoyono, K. Triyana dan B. H. Iswanto, "Rapid Discrimination of Indonesian Herbal Medicines by Using Electronic Nose Based on Array of Commercial Gas Sensors," *Appl. Mech. Mater*, vol. 771, pp. 209-212, 2015.
- [10] K. Brudzewski dan J. Ulaczyk, "An effective method for analysis of dynamic electronic nose responses," *Sensors Actuators, B Chem*, vol. 140, no. 1, pp. 43-50, 2009.
- [11] C. A. Lintang, T. W. Widodo dan D. Lelono, "Rancang Bangun Electronic Nose untuk Mendeteksi Tingkat Kebusukan Ikan Air Tawar," *IJEIS (Indonesian J. Electron. Instrum. Syst*, vol. 6, no. 2, p. 129, 2017.
- [12] A. Wijayanto, "Penggunaan X-Means Clustering Method untuk Mengelompokkan Potensi Sekolah Menengah Unggul di Kabupaten Banyumas," *J. Informatics, Inf. Syst. Softw. Eng. Appl*, vol. 2, no. 1, pp. 80-88, 2019.
- [13] K. Jiawei, Han. Jian, Pei. Cambeline, "DATA MINING 3rd Edition," 2011.
- [14] O. Maimon dan L. Rokach, "Data Mining and Knowledge Discovery Handbook 2ed," 2010.