

## Klasifikasi Senyawa Kimia dengan Notasi *Simplified Molecular Input Line Entry System* (SMILES) menggunakan Metode *Extreme Learning Machine* (ELM)

Isti Marlisa Fitriani<sup>1</sup>, Dian Eka Ratnawati<sup>2</sup>, Syaiful Anam<sup>3</sup>

<sup>1,2</sup>Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya

<sup>3</sup>Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Brawijaya

Email: <sup>1</sup>istimarlisa@gmail.com, <sup>2</sup>dian\_ilkom@ub.ac.id, <sup>3</sup>syaiful @ub.ac.id

### Abstrak

Indonesia mempunyai potensi alam yang besar dengan ditemukannya berbagai spesies tanaman dan hewan. Hal ini menjadi keuntungan bagi masyarakat Indonesia dalam memanfaatkan kekayaan alamnya terutama dalam bidang farmakologi. Dalam bidang farmakologi, senyawa aktif mempunyai manfaat untuk pencegahan dan penyembuhan penyakit. Oleh karena itu, di dalam bidang informatika dilakukan penelitian dengan menciptakan sistem pengelompokan senyawa aktif untuk mengetahui manfaat farmakologinya. SMILES adalah notasi senyawa kimia yang digunakan dalam penelitian ini. Fitur SMILES yang digunakan ada 15, diantaranya atom B, C, N, O, P, S, F, Cl, Br, I, OH, @, =, #, dan (). ELM adalah metode JST yang dapat menjalankan generalisasi dengan bagus dan waktu yang singkat. Pengujian jumlah *hidden neuron* dengan metode *k-fold cross validation* terhadap 2 kelas menghasilkan akurasi paling baik 85,03% pada skenario kelas Metabolisme dan Radang dengan jumlah 5, 10, dan 15 *hidden neuron*. Pengujian jumlah *hidden neuron* dengan metode *k-fold cross validation* terhadap 3 kelas menghasilkan akurasi paling baik 55,06% pada skenario kelas Metabolisme, Radang, dan Kanker dengan jumlah 300 *hidden neuron*. Dihasilkan akurasi terbaik pada pengujian 15 fitur sebesar 55,06% dengan jumlah 300 *hidden neuron* sedangkan pada pengujian 11 fitur diperoleh akurasi terbaik sebesar 49,18% dengan jumlah 400 *hidden neuron*.

**Kata Kunci:** senyawa aktif, SMILES, klasifikasi, ELM

### Abstract

Indonesia has a huge natural's potential by the existence of various plants and animals discovery. This issue brings a good for Indonesian people through taking advantage of nature, especially in pharmacology. In pharmacology, active compounds can be used to prevent and cure disease. Therefore, a research is conducted in informatics's field by making an active compounds' classification system to determine its pharmacological benefits. SMILES is a chemical compound notation used in this research. SMILES's features which are used as many as 15, namely B, C, N, O, P, S, F, Cl, Br, I, OH, @, =, #, and (). ELM is an ANN method that can do a generalization better than conventional methods in a limited time. A number of hidden neurons test which were conducted using *k-fold cross validation* method in 2 classes produced the best accuracy, 85,03%, in Metabolism and Inflammation class scenario with a total of 5, 10, and 15 hidden neurons. A number of hidden neurons' test use *k-fold cross validation* method which were conducted in 3 classes produced the best accuracy, 55,06%, in Metabolism, Inflammation, and Cancer class scenario with a total of 300 hidden neurons. The best accuracy was obtained as many as 55,06% by testing 15 features with 300 hidden neurons, while in 11 features's test with 400 hidden neurons was found a number of 49,18% as the best accuracy.

**Keywords:** active compounds, SMILES, classification, ELM

## 1. PENDAHULUAN

Senyawa diartikan sebagai zat kimia yang mempunyai minimal dua unsur pembentuk yang

nantinya dapat dipecah lagi menjadi unsur terkecil yaitu atom (Brown, dkk, 2009). Di alam semesta ini terdapat jutaan senyawa yang dikelompokkan menjadi senyawa aktif dan

senyawa tidak aktif. Senyawa aktif mempunyai banyak manfaat seperti pengobatan, produk kesehatan, suplemen makanan, dan kosmetik (Melissa, & Muchtaridi, 2017). Senyawa aktif dapat diperoleh dari ekstraksi hewan dan tanaman. Salah satu negara yang mempunyai potensi alam terbesar adalah Indonesia, yaitu dengan ditemukannya berbagai spesies tanaman dan hewan. Hal ini menjadi keuntungan tersendiri bagi masyarakat Indonesia dalam memanfaatkan kekayaan alamnya terutama dalam bidang farmakologi.

Dalam bidang farmakologi, senyawa aktif mempunyai manfaat untuk pencegahan dan penyembuhan suatu penyakit. Namun, senyawa aktif yang sudah diidentifikasi manfaatnya masih sedikit sekali jika dibandingkan dengan keberadaannya di alam semesta ini. Senyawa aktif dapat diidentifikasi manfaatnya dengan serangkaian penelitian di dalam laboratorium oleh kimiawan. Fitokimia merupakan salah satu uji lab yang penting untuk mendapatkan informasi golongan senyawa kimia yang ada pada tanaman atau hewan yang menjadi objek penelitian. Terdapat dua faktor penting dalam uji fitokimia, yaitu pemilihan metode ekstraksi dan pelarut (Kristanti, dkk, 2008).

*Simplified Molecular Input Line Entry System* adalah notasi kimia satu dimensi yang dibuat sebagai pemrosesan informasi kimia modern untuk menggambarkan struktur kimia dengan karakter-karakter *American Standard Code for Information Interchange* dan dibaca dengan variabel *string* (Weininger, 1988). Notasi SMILES dipilih karena sederhana namun tetap memiliki informasi yang sama dengan notasi senyawa kimia lain sehingga dapat dibaca dengan baik oleh manusia maupun komputer. Oleh karena itu, notasi SMILES dipilih dalam penelitian ini untuk mempermudah komputasi dalam sistem klasifikasi. Dalam penelitian ini digunakan fitur sebanyak 15, yaitu atom B, C, N, O, P, S, F, Cl, Br, I, OH, @, =, #, dan ().

ELM merupakan metode JST *feed-forward* menggunakan satu *hidden layer* (Sun, dkk, 2008). Kelebihan ELM ada pada kecepatan belajar dan memiliki akurasi lebih bagus daripada metode konvensional (Agustina, 2010). Bahkan ELM dapat menjalankan generalisasi dengan bagus dan dengan waktu yang singkat daripada *Backpropagation* dan *Support Vector Machine* (SVM) (Huang, 2006).

## 2. LANDASAN KEPUSTAKAAN

### 2.1 Notasi *Simplified Molecular Input Line Entry System*

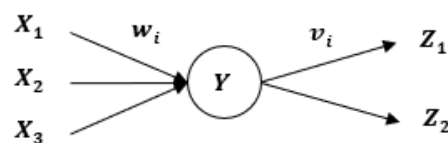
SMILES yaitu notasi kimia yang digunakan untuk pemrosesan informasi dari kimia modern agar dapat menggambarkan struktur kimia dengan karakter-karakter ASCII (Weininger, 1988). Notasi SMILES adalah tatanama senyawa kimia yang sederhana sehingga dapat dibaca dengan baik oleh manusia maupun komputer. Oleh karena itu, notasi SMILES dapat digunakan sebagai pertukaran data dalam bidang informatika. Atom yang digunakan dalam notasi SMILES adalah B, C, N, O, P, S, F, Cl, Br, I, OH, @, =, #, dan ().

### 2.2 Klasifikasi

Klasifikasi tidak sama dengan prediksi, dalam *machine learning* klasifikasi adalah bagian dari prediksi. Klasifikasi bertujuan untuk memprediksi suatu kelas (Thabtah, 2007). Klasifikasi menggunakan data *training* dan data *testing*. Data *testing* dikelompokkan berdasarkan data *training* yang sudah ada.

### 2.3 Jaringan Saraf Tiruan

JST meniru cara kerja dari sel-sel saraf pada makhluk hidup. Otak manusia memiliki jutaan sel saraf (neuron) yang bertugas untuk memroses informasi. Neuron saling terhubung membentuk jaringan saraf untuk mendukung kemampuan kerja otak manusia. Seperti halnya super komputer yang memiliki kumpulan prosesor saling berhubungan. JST adalah prosesor yang tersebar secara sejajar yang dibuat dari proses sederhana dan menyimpan pengetahuan yang disebut sebagai pengalaman serta bisa digunakan untuk proses yang lain (Haykin, 2009). Jadi, JST meniru cara kerja otak manusia terutama kemampuan untuk belajar. Arsitektur JST ada pada Gambar 1 (Fausett, 1994).



**Gambar 1 Rancangan JST**

Keterangan:

$X_1, X_2, X_3$  : *input neuron*

$Y$  : *hidden neuron*

$Z_1, Z_2$  : *output neuron*

$w_i, v_i$  : bobot

### 2.4 Fungsi Aktivasi JST

Fungsi ini digunakan untuk menetapkan keluaran neuron, input yang digunakan adalah gabungan linier input serta bobot yang digunakan (Siang, 2009). Macam-macam fungsi aktivasi pada JST adalah:

#### 1. Linier

Fungsi linier memungkinkan nilai keluaran sama dengan nilai input. Fungsi ini menggunakan Persamaan 1.

$$y = f(x) \tag{1}$$

#### 2. Threshold

Fungsi *threshold* atau fungsi tangga biner atau fungsi step digunakan untuk mengubah input variabel kontinu menjadi *output* biner (1 atau 0). Fungsi ini menggunakan Persamaan 2.

$$y = f(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases} \tag{2}$$

#### 3. Sigmoid Biner

Fungsi ini mempunyai nilai pada interval 0 sampai 1. Persamaan yang digunakan adalah Persamaan 3.

$$y = f(x) = \frac{1}{1+e^{-x}} \tag{3}$$

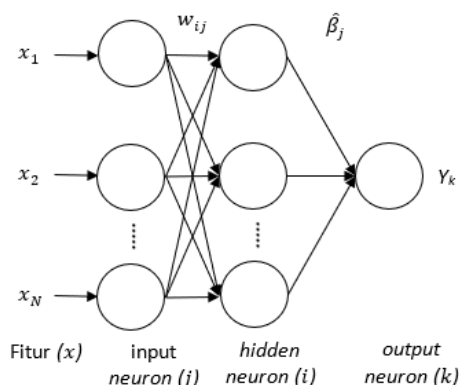
#### 4. Sigmoid Bipolar

Fungsi *sigmoid* bipolar mempunyai nilai interval 1 sampai -1. Fungsi ini menggunakan Persamaan 4.

$$y = f(x) = \frac{1-e^{-x}}{1+e^{-x}} \tag{4}$$

### 2.5 Extreme Learning Machine (ELM)

ELM biasa disebut *Single Hidden Layer Feedforward neural Networks* (SLFNs) (Sun, dkk, 2008). ELM digunakan untuk menyelesaikan permasalahan pada JST, yang paling utama adalah dalam hal kecepatan belajar. ELM tidak melakukan pelatihan pada bobot input tapi melakukan pelatihan untuk mendapatkan bobot *output* dengan *moore-penrose invers* pada sistem linier dan mendapatkan *node* yang memiliki nilai *output* maksimal. Nilai *w* ditentukan secara *random*. Sehingga algoritme ELM mempunyai kecepatan belajar dan kinerja generalisasi yang baik. Rancangan ELM terdapat pada Gambar 2.



Gambar 2 Rancangan ELM

### 2.6 Algoritme ELM

Algoritme ELM mempunyai tahapan untuk menghitung data latih dan data uji (Huang, dkk, 2016). Terdapat 5 langkah dalam *training* ELM, yaitu:

**Langkah 1:** Menentukan nilai bobot ( $w_{ij}$ ) secara *random* dengan  $i$  adalah *hidden neuron* dan  $j$  adalah *input neuron*. Nilai  $W$  dibangkitkan antara -0,5 sampai 0,5.

**Langkah 2:** Menentukan matriks awal *output hidden neuron* ( $H_{init}$ ), dengan  $X$  adalah matriks data latih dengan panjang  $j$  *input neuron* (fitur) dikali  $N$  data *training* dan  $W^T$  yaitu matriks bobot yang ditransposisi.

$$H_{init} = X \cdot W^T \tag{5}$$

**Langkah 3:** Menentukan matriks *output hidden neuron* ( $H$ ) menggunakan fungsi aktivasi *sigmoid* biner sesuai Persamaan 3 sehingga nilai  $H$  dihitung menggunakan Persamaan 6.

$$H = \frac{1}{(1+\exp(-H_{init}))} \tag{6}$$

**Langkah 4:** Menghitung matriks  $H^\dagger$ , dengan  $H^T$  adalah matriks nilai  $H$  yang sudah ditransposisi dan  $(H^T \cdot H)^{-1}$  adalah nilai *invers* dari  $(H^T \cdot H)$ .

$$H^\dagger = (H^T \cdot H)^{-1} \cdot H^T \tag{7}$$

**Langkah 5:** Menentukan bobot keluaran ( $\hat{\beta}$ ),  $\hat{\beta}$  adalah bobot keluaran dan  $Y$  adalah target atau kelas.

$$\hat{\beta} = H^\dagger \cdot Y \tag{8}$$

Dengan menggunakan nilai bobot ( $w_{ij}$ ) dan  $\hat{\beta}$  yang diperoleh dari *training* ELM, maka lakukan 4 langkah dalam *testing* ELM berikut.

**Langkah 1:** Menghitung matriks inialisasi

output hidden neuron ( $H_{init}$ ),  $X$  adalah matriks data uji dengan panjang  $j$  input neuron (fitur) dikali  $N$  data uji serta  $W^T$  adalah matriks nilai bobot yang di-*transpose*. Persamaan untuk menghitung  $H_{init}$  sama dengan Persamaan 5.

**Langkah 2:** Menentukan matriks output hidden neuron ( $H$ ) dengan menggunakan Persamaan 6.

**Langkah 3:** Menentukan  $\hat{Y}$  prediksi menggunakan Persamaan 9.

$$\hat{Y} = H \cdot \beta \tag{9}$$

**Langkah 4:** Menghitung kelas prediksi dengan mencari selisih terkecil antara  $\hat{Y}$  prediksi dengan kelas sebenarnya ( $Y$ ) yang terdapat pada data latih.

$$\text{Kelas prediksi} = |\hat{Y} - Y| \tag{10}$$

### 2.6 Evaluasi Kinerja Klasifikasi

Evaluasi kinerja klasifikasi digunakan untuk mengukur seberapa akurat sistem untuk mengelompokkan kelas. *Confusion matrix* digunakan untuk membantu perhitungan evaluasi kinerja (Han, dkk, 2012). *Confusion matrix* dapat dilihat pada Tabel 1.

**Tabel 1 Confusion Matrix**

| Actual Class \ Predicted Class | Yes | No |
|--------------------------------|-----|----|
|                                | Yes | TP |
| No                             | FP  | TN |

#### 1. Akurasi

Akurasi digunakan untuk mengetahui seberapa akurat sistem melakukan klasifikasi dengan baik. Untuk menghitung nilai akurasi digunakan Persamaan 11.

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{11}$$

#### 2. Precision

*Precision* digunakan untuk mengetahui kemampuan sistem agar tidak memanggil kembali data tidak relevan dengan apa yang diinginkan *user* dari basis data. Untuk menghitung nilai *precision* digunakan Persamaan 12.

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\% \tag{12}$$

#### 3. Recall

*Recall* digunakan untuk mengetahui

kemampuan sistem untuk memanggil kembali data yang relevan dengan apa yang diminta pengguna dari basis data. Untuk menghitung nilai *precision* digunakan Persamaan 13.

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \tag{13}$$

### 3. METODOLOGI

Berdasarkan sistem yang dibuat, maka data yang dibutuhkan adalah data latih berupa senyawa yang terdiri dari nama senyawa, formula molekul, notasi SMILES, dan manfaatnya. Senyawa tersebut didapatkan dari database yang dapat diakses di link <https://pubchem.ncbi.nlm.nih.gov/search/search.cgi#>.

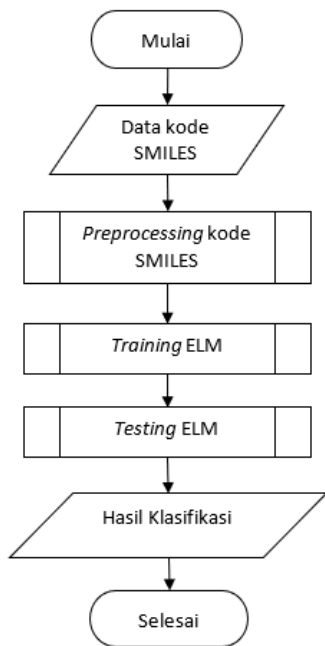
Implementasi sistem menggunakan bahasa pemrograman Python versi 2.7 dan menggunakan *Integrated Development Environment* (IDE) Spyder. Sistem menggunakan fitur atau *input neuron* sebanyak 15.

Data latih yang digunakan sebanyak 840 yaitu 501 data pada kelas Metabolisme, 247 data pada kelas Kanker, dan 92 data pada kelas Radang. Data uji yang digunakan sebanyak 209 yaitu 125 data pada kelas Metabolisme, 62 data pada kelas Kanker, dan 22 data pada kelas Radang.

Sistem melakukan klasifikasi dengan menerima data berupa notasi SMILES. Sistem kemudian melakukan *preprocessing* notasi SMILES untuk mendapatkan nilai 15 fitur, lalu dilakukan proses perhitungan menggunakan metode ELM yang terdiri dari *training* ELM dan *testing* ELM. Setelah perhitungan selesai maka akan didapatkan hasil klasifikasi berupa kelas dari data notasi SMILES. Diagram alir sistem terdapat pada Gambar 3.

### 4. HASIL PENGUJIAN

Dilakukan pengujian dan analisis terhadap hasil perhitungan yang telah dilakukan oleh sistem serta evaluasi kinerja sistem. Pengujian dilakukan berdasarkan perancangan pengujian dengan dilakukan 2 pengujian, yaitu pengujian jumlah *hidden neuron* menggunakan metode *k fold cross validation* dan pengujian jumlah fitur dengan metode *k fold cross validation*.



Gambar 3 Diagram Alir Sistem

### 4.1 Pengujian Jumlah Hidden Neuron dengan Metode K-fold Cross Validation

Metode *k fold cross validation* digunakan untuk mengevaluasi kinerja ELM pada data SMILES. Pengujian jumlah *hidden neuron* dengan *k fold cross validation* adalah pengujian jumlah *hidden neuron* memakai 20% data uji dari keseluruhan data yang dipilih secara bergantian dengan nilai  $k=5$  sehingga diperlukan 5 kali pengujian menggunakan data *training* dan data *testing* yang berbeda lalu hasil akurasi dari kelima pengujian tersebut dirata-rata. Setiap pengujian dilakukan 5 kali *running* pada program lalu hasilnya dirata-rata.

#### 4.1.1 Pengujian 2 Kelas

Pengujian 2 kelas terdiri dari 3 skenario yang berbeda, yaitu kelas Metabolisme dengan Kanker, Radang dengan Kanker, dan Metabolisme dengan Radang. Pengujian jumlah *hidden neuron* dikerjakan sebanyak 8 kali, yaitu dengan jumlah *hidden neuron* 5, 10, 15, 20, 25, 30, 35, dan 40. Pengujian *k fold cross validation* memakai 20% data uji dari keseluruhan data yang dipilih secara bergantian dengan nilai  $k=5$  sehingga diperlukan 5 kali pengujian menggunakan data *training* dan data *testing* yang berbeda lalu hasil akurasi kelima pengujian tersebut dirata-rata.

##### 4.1.1.1 Kelas Metabolisme dan Kanker

Hasil pengujian pada kelas Metabolisme dan Kanker terdapat pada Tabel 2.

Tabel 2 Perancangan Pengujian Jumlah Hidden Neuron Kelas Metabolisme dan Kanker

| Cross Validation Jumlah Hidden Neuron | 1     | 2     | 3     | 4     | 5     | Rata-Rata Akurasi |
|---------------------------------------|-------|-------|-------|-------|-------|-------------------|
| 5                                     | 67.70 | 65.56 | 67.27 | 66.63 | 67.85 | 67.00             |
| 10                                    | 66.85 | 64.92 | 67.49 | 64.49 | 68.71 | 66.49             |
| 15                                    | 68.13 | 63.21 | 67.81 | 65.99 | 67.96 | 66.62             |
| 20                                    | 70.80 | 64.81 | 69.52 | 66.85 | 67.74 | 67.94             |
| 25                                    | 69.31 | 65.95 | 69.20 | 68.02 | 65.59 | 67.61             |
| 30                                    | 69.95 | 65.35 | 67.70 | 66.20 | 65.48 | 66.94             |
| 35                                    | 69.20 | 66.42 | 68.45 | 65.99 | 65.70 | 67.15             |
| 40                                    | 66.85 | 65.13 | 67.59 | 67.91 | 65.91 | 66.68             |

##### 4.1.1.2 Kelas Radang dan Kanker

Hasil pengujian pada kelas Radang dan Kanker dapat dilihat pada Tabel 3.

Tabel 3 Perancangan Pengujian Jumlah Hidden Neuron Kelas Radang dan Kanker

| Cross Validation Jumlah Hidden Neuron | 1     | 2     | 3     | 4     | 5     | Rata-Rata Akurasi |
|---------------------------------------|-------|-------|-------|-------|-------|-------------------|
| 5                                     | 73.81 | 73.81 | 73.33 | 73.81 | 73.01 | 73.55             |
| 10                                    | 72.62 | 73.81 | 73.81 | 73.81 | 73.01 | 73.41             |
| 15                                    | 72.62 | 72.86 | 73.33 | 73.57 | 70.12 | 72.50             |
| 20                                    | 71.67 | 72.38 | 73.10 | 74.29 | 69.40 | 72.17             |
| 25                                    | 72.38 | 73.10 | 73.57 | 73.57 | 69.40 | 72.40             |
| 30                                    | 74.29 | 75.24 | 73.81 | 74.05 | 66.75 | 72.83             |
| 35                                    | 73.81 | 74.76 | 73.81 | 70.24 | 65.79 | 71.68             |
| 40                                    | 76.90 | 75.95 | 74.76 | 73.09 | 65.30 | 73.20             |

##### 4.1.1.3 Kelas Metabolisme dan Radang

Hasil pengujian pada kelas Metabolisme dan Radang terdapat pada Tabel 4.

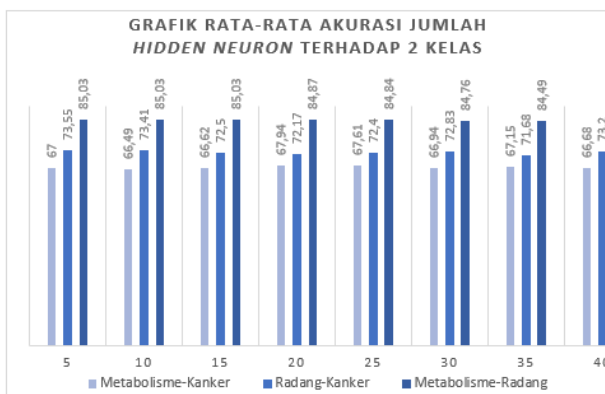
Tabel 4 Perancangan Pengujian Jumlah Hidden Neuron Kelas Metabolisme dan Radang

| Cross Validation Jumlah Hidden Neuron | 1     | 2     | 3     | 4     | 5     | Rata-Rata Akurasi |
|---------------------------------------|-------|-------|-------|-------|-------|-------------------|
| 5                                     | 85.03 | 85.03 | 85.03 | 85.03 | 85.03 | 85.03             |
| 10                                    | 85.03 | 85.03 | 85.03 | 85.03 | 85.03 | 85.03             |
| 15                                    | 85.03 | 85.03 | 85.03 | 85.03 | 85.03 | 85.03             |
| 20                                    | 84.89 | 84.49 | 85.03 | 85.03 | 84.89 | 84.87             |
| 25                                    | 84.49 | 84.62 | 85.03 | 85.17 | 84.89 | 84.84             |
| 30                                    | 84.49 | 84.49 | 84.62 | 85.17 | 85.03 | 84.76             |



|    |       |       |       |       |       |       |
|----|-------|-------|-------|-------|-------|-------|
| 35 | 84.35 | 83.81 | 84.35 | 85.17 | 84.76 | 84.49 |
| 40 | 84.76 | 83.94 | 83.40 | 85.03 | 84.62 | 84.35 |

Berdasarkan ketiga tabel di atas, yaitu Tabel 2, Tabel 3, dan Tabel 4 diperoleh rata-rata akurasi jumlah *hidden neuron* terhadap skenario kelas Metabolisme dan Radang, Radang dan Kanker, serta Metabolisme dan Radang. Untuk perbandingan nilai rata-rata akurasi dari ketiga skenario tersebut terdapat pada Gambar 4.



Gambar 4 Grafik Rata-Rata Akurasi Jumlah *Hidden Neuron* Terhadap 2 Kelas

4.1.2 Pengujian 3 Kelas

Pengujian 3 kelas terdiri dari 1 skenario yaitu kelas Metabolisme, Radang, dan Kanker. Pengujian jumlah *hidden neuron* dikerjakan sebanyak 4 kali, yaitu dengan jumlah *hidden neuron* 100, 200, 300, dan 400. Pengujian *k fold cross validation* menggunakan 20% data uji dari keseluruhan data yang dipilih secara bergantian dengan nilai  $k=5$  sehingga diperlukan 5 kali pengujian lalu hasil akurasi dari kelima pengujian tersebut dirata-rata.

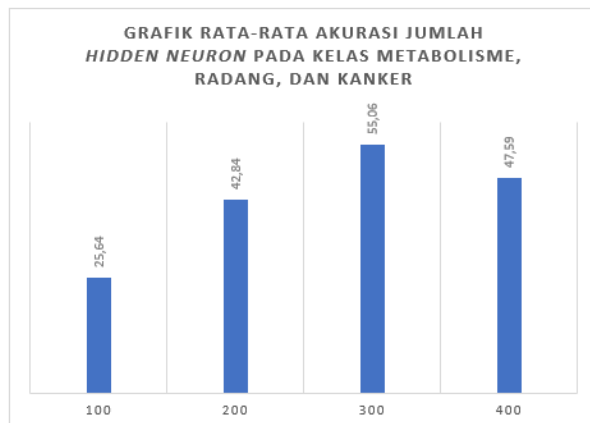
4.1.2.1 Kelas Metabolisme, Radang, dan Kanker

Hasil pengujian pada kelas Metabolisme, Radang, dan Kanker terdapat pada Tabel 5.

Tabel 5 Perancangan Pengujian Jumlah *Hidden Neuron* Kelas Metabolisme, Radang, dan Kanker

| Jumlah <i>Hidden Neuron</i> | Cross Validation |       |       |       |       | Rata-Rata Akurasi |
|-----------------------------|------------------|-------|-------|-------|-------|-------------------|
|                             | 1                | 2     | 3     | 4     | 5     |                   |
| 100                         | 25.74            | 24.50 | 22.87 | 22.87 | 32.21 | 25.64             |
| 200                         | 37.99            | 50.43 | 44.98 | 34.93 | 45.87 | 42.84             |
| 300                         | 53.50            | 55.98 | 59.52 | 51.10 | 55.19 | 55.06             |
| 400                         | 53.88            | 44.50 | 50.81 | 41.05 | 47.70 | 47.59             |

Perbandingan rata-rata akurasi dari dari Tabel 4 terdapat pada Gambar 5.



Gambar 5 Grafik Rata-Rata Akurasi Jumlah *Hidden Neuron* pada Kelas Metabolisme, Radang, dan Kanker

4.2 Pengujian Jumlah Fitur dengan Metode *K-fold Cross Validation*

Pada *k fold cross validation* menggunakan 20% data uji dari keseluruhan data yang dipilih secara bergantian dengan nilai  $k=5$  sehingga diperlukan 5 kali pengujian lalu hasil akurasi dari kelima pengujian tersebut dirata-rata. Pada setiap pengujian dilakukan 5 kali *running* pada program lalu hasilnya dirata-rata.

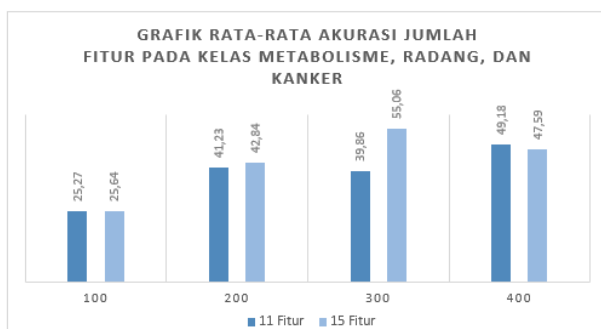
4.2.1 Pengujian 11 Fitur

Pada pengujian 11 fitur digunakan fitur B, C, N, O, P, S, F, Cl, Br, I, dan OH. Hasil pengujian 11 fitur terdapat pada Tabel 6.

Tabel 6 Perancangan Pengujian 11 Fitur pada Kelas Metabolisme, Radang, dan Kanker

| Jumlah <i>Hidden Neuron</i> | Cross Validation |       |       |       |       | Rata-Rata Akurasi |
|-----------------------------|------------------|-------|-------|-------|-------|-------------------|
|                             | 1                | 2     | 3     | 4     | 5     |                   |
| 100                         | 22.87            | 25.36 | 23.73 | 27.56 | 26.83 | 25.27             |
| 200                         | 41.44            | 47.36 | 29.76 | 40.86 | 46.73 | 41.23             |
| 300                         | 41.34            | 37.23 | 37.70 | 28.42 | 54.62 | 39.86             |
| 400                         | 41.05            | 58.09 | 46.12 | 47.37 | 53.27 | 49.18             |

Grafik perbandingan rata-rata akurasi antara Tabel 4 dan Tabel 5 terdapat pada Gambar



Gambar 6 Grafik Rata-Rata Akurasi Jumlah Fitur pada Kelas Metabolisme, Radang, dan Kanker

### 5. EVALUASI KINERJA SISTEM

Evaluasi kinerja terdiri dari perhitungan akurasi, *precision*, dan *recall*. Hasil pengujian yang dievaluasi adalah pada pengujian 3 kelas yaitu Metabolisme, Radang, dan Kanker dengan 15 fitur dan 300 *hidden neuron*. Perhitungan evaluasi kinerja akan diterapkan pada masing-masing kelas, dan karena menggunakan multipel kelas maka hasil evaluasi dari ketiga kelas akan dirata-rata.

#### 1. Kelas Metabolisme

Nilai *confusion matrix* yang terdiri dari TP, TN, FP, dan FN pada kelas Metabolisme dapat dilihat pada Tabel 7.

Tabel 7 *Confusion Matrix* Kelas Metabolisme

| <i>Actual Class</i> \ <i>Predicted Class</i> | Yes | No  |
|--|-----|-----|
|  | Yes | 123 |
| No   | 82  | 2   |

##### 1. Akurasi

Untuk menghitung nilai akurasi digunakan Persamaan 11.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$$

$$Akurasi = \frac{123 + 2}{123 + 2 + 82 + 2} \times 100\%$$

$$Akurasi = 59,8\%$$

##### 2. Precision

Untuk menghitung nilai *precision* digunakan Persamaan 12.

$$Precision = \frac{TP}{TP+FP} \times 100\%$$

$$Precision = \frac{123}{123 + 82} \times 100\%$$

$$Precision = 60\%$$

##### 3. Recall

Untuk menghitung nilai *recall* digunakan Persamaan 13.

$$Recall = \frac{TP}{TP+FN} \times 100\%$$

$$Recall = \frac{123}{123 + 2} \times 100\%$$

$$Recall = 98,4\%$$

#### 2. Kelas Radang

Nilai *confusion matrix* pada kelas Radang dapat dilihat pada Tabel 8.

Tabel 8 *Confusion Matrix* Kelas Radang

| <i>Actual Class</i> \ <i>Predicted Class</i> | Yes | No  |
|--|-----|-----|
|  | Yes | 1   |
| No   | 63  | 124 |

##### 1. Akurasi

Untuk menghitung nilai akurasi digunakan Persamaan 11.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$$

$$Akurasi = \frac{1 + 124}{1 + 124 + 63 + 21} \times 100\%$$

$$Akurasi = 59,8\%$$

##### 2. Precision

Untuk menghitung nilai *precision* digunakan Persamaan 12.

$$Precision = \frac{TP}{TP+FP} \times 100\%$$

$$Precision = \frac{1}{1 + 63} \times 100\%$$

$$Precision = 1,56\%$$

##### 3. Recall

Untuk menghitung nilai *recall* digunakan Persamaan 13.

$$Recall = \frac{TP}{TP+FN} \times 100\%$$

$$Recall = \frac{1}{1 + 21} \times 100\%$$

$$Recall = 4,55\%$$

#### 3. Kelas Kanker

Nilai *confusion matrix* pada kelas Kanker dapat dilihat pada Tabel 9.

Tabel 9 *Confusion Matrix* Kelas Kanker

| <i>Actual Class</i> \ <i>Predicted Class</i> | Yes | No |
|--|-----|----|
|  | Yes | 1  |

|    |    |     |
|----|----|-----|
| No | 23 | 124 |
|----|----|-----|

1. Akurasi

Untuk menghitung nilai akurasi digunakan Persamaan 11.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$$

$$Akurasi = \frac{1 + 124}{1 + 124 + 23 + 61} \times 100\%$$

$$Akurasi = 59,8\%$$

2. Precision

Untuk menghitung nilai precision digunakan Persamaan 12.

$$Precision = \frac{TP}{TP+FP} \times 100\%$$

$$Precision = \frac{1}{1 + 23} \times 100\%$$

$$Precision = 4,17\%$$

3. Recall

Untuk menghitung nilai recall digunakan Persamaan 13.

$$Recall = \frac{TP}{TP+FN} \times 100\%$$

$$Recall = \frac{1}{1 + 61} \times 100\%$$

$$Recall = 1,61\%$$

4. Hasil Evaluasi Kinerja Klasifikasi Multipel Kelas

Untuk mendapatkan nilai hasil evaluasi kinerja klasifikasi multipel kelas maka hasil perhitungan akurasi, precision, dan recall masing-masing kelas Metabolisme, Radang, dan Kanker dirata-rata.

$$Akurasi = \frac{59,8\% + 59,8\% + 59,8\%}{3}$$

$$= 59,8\%$$

$$Precision = \frac{60\% + 1,56\% + 4,17\%}{3}$$

$$= 21,91\%$$

$$Recall = \frac{98,4\% + 4,55\% + 1,61\%}{3}$$

$$= 34,85\%$$

Berdasarkan hasil perhitungan evaluasi kinerja sistem pada 3 kelas, maka diperoleh nilai akurasi sebesar 59,8%, hal ini menunjukkan bahwa tingkat akurasi sistem cukup baik. Nilai precision sebesar 21,91%, hal ini menunjukkan bahwa kemampuan sistem untuk tidak memanggil kembali data tidak relevan dengan apa yang diminta user masih kurang baik. Nilai recall sebesar 34,85% hal ini menunjukkan

kemampuan sistem untuk memanggil kembali data yang relevan dengan apa yang diminta user masih kurang baik

Hasil evaluasi kinerja pada perhitungan precision dan recall masih kurang baik karena hasil klasifikasi sistem cenderung menghasilkan kelas 1 atau kelas Metabolisme yang memiliki jumlah data latih lebih banyak daripada kelas Radang dan kelas Kanker. Hal ini dibuktikan dari hasil perhitungan precision dan recall pada kelas Metabolisme lebih besar yaitu 60% dan 98,4 % dibandingkan hasil perhitungan precision dan recall pada kelas Radang dan Kanker.

6. KESIMPULAN

Berdasarkan hasil pengujian dan analisis, didapatkan kesimpulan sebagai berikut.

1. Dari hasil pengujian jumlah hidden neuron dengan metode k-fold cross validation pada 2 kelas, ELM dapat melakukan klasifikasi dengan akurasi terbaik sebesar 85,03% pada skenario kelas Metabolisme dan Radang dengan jumlah 5, 10, dan 15 hidden neuron. Dan pada 3 kelas, ELM dapat melakukan klasifikasi dengan akurasi terbaik sebesar 55,06% pada skenario kelas Metabolisme, Radang, dan Kanker dengan jumlah 300 hidden neuron.
2. ELM dapat bekerja lebih maksimal pada 2 kelas daripada 3 kelas, hal ini dikarenakan semakin banyak kelas maka ELM harus lebih banyak belajar untuk membedakan data antar kelas, sehingga diperlukan lebih banyak fitur sebagai pembeda data antar kelas agar data dapat dikenali dengan maksimal. Hal ini ditunjukkan bahwa ELM dapat menghasilkan akurasi terbaik pada pengujian 15 fitur sebesar 55,06% dengan jumlah 300 hidden neuron sedangkan pada pengujian 11 fitur diperoleh akurasi terbaik sebesar 49,18% dengan jumlah 400 hidden neuron.

7. DAFTAR PUSTAKA

Agustina, I.D., Anggraeni, W., dan Mukhlason A., 2010 Penerapan Metode Extreme Learning Machine untuk Peramalan Permintaan, [online] Tersedia di: <<http://digilib.its.ac.id/public/ITS-Undergraduate-9832-Paper.pdf>> [Diakses 20 Agustus 2018]

Brown, Theodore L., LeMay, H E., Bursten, Bruce E., Murphy, Catherine J.,



- Woodward, dan Patrick M., 2009. *Chemistry: The Central Science*. 11th Edition. Upper Saddle River, NJ: Prentice Hall.
- Fadilla I., Adikara P. P., dan Perdana R. S., 2018. *Klasifikasi Penyakit Chronic Kidney Disease (CKD) dengan menggunakan Metode Extreme Learning machine (ELM)*, [e-journal] 2(10). Tersedia melalui: <<http://j-ptiik.ub.ac.id/index.php/j-ptiik>> [Diakses 20 Oktober 2018]
- Fausett, L., 1994. *Fundamentals of Neural Networks (Architectures, Algorithms, and Applications)*. New Jersey : Prentice Hall.
- Han, J., Kamber M., dan Pei, J., 2012. *Data Mining Concepts and Techniques. Third Edition*. United States of America: Elsevier.
- Haykin, S., 2009. *Neural Networks and Learning Machines. Third Edition*. Upper Saddle River, New Jersey: Pearson Education, Inc.
- Huang, G., Zhou, H., Ding X., dan Zhang R., 2012. *Extreme Learning Machine for Regression and Multiclass Classification*, [online] <<https://ieeexplore.ieee.org/document/6035797>> [Diakses 25 Februari 2019]
- Huang, G., Zhu, Q. dan Siew, C., 2006. *Extreme learning machine: Theory and applications*, [online] <<http://axon.cs.byu.edu/~martinez/classes/678/Presentations/Yao.pdf>> [Diakses 20 Oktober 2018]
- Kristanti, A. N., Aminah, N. S., Tanjung, M., dan Kurniadi, B., 2008. *Buku Ajar Fitokimia*. Surabaya: Airlangga University Press.
- Melissa, dan Muchtaridi M., 2017. *Senyawa Aktif dan Manfaat Farmakologis Ageratum Conyzoides*, [online] Tersedia di: <<http://jurnal.unpad.ac.id/farmaka/article/view/13302/pdf>> [Diakses 4 Desember 2018]
- Rochman, E. M. S., dan Rachmad, A., 2017. *Kecerdasan Komputasional Konsep Dan Aplikasi*. Malang: Media Nusa Creative.
- Siang, J. J., 2009. *Jaringan Syaraf Tiruan Dan Pemrogramannya Menggunakan Matlab*. Yogyakarta: Andi.
- Sun, Z. L., Choi, T. M., Au, K. F. dan Yu, Y., 2008. *Sales Forecasting using Extreme Learning Machine with Application in Fashion Retailing*, [e-journal]. Tersedia melalui <<https://www.sciencedirect.com/>> [Diakses 20 Oktober 2018]
- Thabtah F., 2007. *A Review Of Associative Classification Mining*, [online] Tersedia di: <<http://eprints.hud.ac.uk/id/eprint/269/1/ThabtahReview.pdf>> [Diakses 20 Agustus 2018]
- Tigusti R. R. W., Ratnawati D. E., dan Anam, S., 2018. *Fungsi Senyawa Berdasarkan Simplified Molecular Input Line Entry System (SMILES)*, [e-journal] 3(1). Tersedia melalui: <<http://j-ptiik.ub.ac.id/index.php/j-ptiik>> [Diakses 20 Oktober 2018]
- Weininger, D., 1988. *SMILES, a Chemical Language and Information System. Introduction to Methodology and Encoding Rules*. [online] <<http://organical.org/seminario/smile1988.pdf>> [Diakses 20 Oktober 2018]
- Witanto S., Ratnawati D. E., dan Anam, S., 2018. *Pengelompokan Fungsi Aktif Senyawa Data SMILES (Simplified Molecular Input Line Entry System) Menggunakan Metode K-Means Dengan Inisialisasi Pusat Klaster Menggunakan Metode Heuristic  $O(N \log N)$* , [e-journal] 3(1). Tersedia melalui: <<http://j-ptiik.ub.ac.id/index.php/j-ptiik>> [Diakses 20 Oktober 2018]