

Momentum *Backpropagation* Untuk Klasifikasi Fungsi Senyawa Aktif Berdasarkan Notasi SMILES (*Simplified Molecular Input Line Entry System*)

Nyimas Ayu Widi Indriana¹, Dian Eka Ratnawati², Syaiful Anam³

^{1,2}Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya

³Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Brawijaya
Email: ¹nyimasayu27@gmail.com, ²dian_ilkom@ub.ac.id, ³syaiful@ub.ac.id

Abstrak

Senyawa aktif merupakan senyawa yang dapat dimanfaatkan untuk pembuatan obat-obatan tertentu dan sangat penting dalam bidang kesehatan. Pengklasifikasian senyawa aktif merupakan hal yang paling utama dalam pembuatan obat-obatan. Setelah tahap pengklasifikasian senyawa aktif, maka dapat dilanjutkan dengan proses pembuatan dan pengujian obat-obatan yang membutuhkan berbagai macam alat. Biaya dan waktu yang diperlukan untuk pembuatan dan pengujian obat-obatan ini cukup tinggi. Hal ini lah yang menjadi kendala utama bagi para ahli medis untuk pembuatan obat-obatan tertentu. Dengan memanfaatkan teknologi saat ini dapat dibuat suatu sistem untuk melakukan proses klasifikasi senyawa aktif, sehingga kinerja para ahli medis dalam pembuatan obat-obatan tertentu dapat lebih cepat. Proses klasifikasi dapat dilakukan dengan menggunakan computer dan memanfaatkan notasi SMILES. Penggunaan notasi SMILES memungkinkan suatu senyawa dapat di proses oleh computer. Metode momentum *Backpropagation* dapat digunakan untuk melakukan proses klasifikasi dengan baik. Berdasarkan program yang sudah dibuat, dilakukan 4 macam pengujian dengan menggunakan jumlah data latih sebanyak 522 dan data uji sebanyak 131 menghasilkan akurasi terbaik sebesar 70,99% dengan *learning rate* sebesar 0,00001, *max epoch* sebesar 100, momentum sebesar 0,25 dan *neuron hidden layer* sebanyak 4.

Kata kunci: SMILES, *backpropagation*, momentum *backpropagation*

Abstract

Active compounds can be used to make certain drugs and very important in the medical sector. Classification of active compounds is the most important thing in making medicines. After classifying the active compound, it is continued with the process of making and testing drugs that require a variety of tools. The cost of making and testing these drugs requires a high cost and time. This is a major obstacle for medical experts to make certain medicines. By utilizing current technology, a system can be made to classification process of active compounds, so the performance of medical experts for making certain drugs can be faster. The classification process can be done by using a computer and utilizing the SMILES notation. SMILES notation allows a compound to be processed by a computer. The momentum Backpropagation method can be used to perform the classification process properly. Based on the program that has been made, there are 4 types of testing using 522 training data and 131 test data producing, the best accuracy of 70,99% with a learning rate of 0,00001, max epoch of 100, momentum of 0,25 and hidden layer neurons of 4.

Keywords: SMILES, *backpropagation*, momentum *backpropagation*

1. PENDAHULUAN

Senyawa merupakan suatu zat yang tersusun dari dua atau lebih unsur yang berbeda dan membentuk suatu ikatan kimia. Senyawa dibagi dua macam yaitu senyawa aktif dan

senyawa tidak aktif. Senyawa aktif merupakan senyawa yang dapat memicu terjadinya aktivitas biologi dalam organisme hidup atau sering kita sebut dengan istilah biokimia (Sumardjo, 2006), sedangkan senyawa tidak aktif memiliki peran sebagai pendukung suatu reaksi kimia. Senyawa

aktif juga dapat berfungsi sebagai farmakologis dan banyak digunakan untuk pembuatan obat-obatan yang sangat dibutuhkan dalam kehidupan manusia untuk menyembuhkan berbagai penyakit.

Proses pembuatan obat-obat ini harus sesuai dengan standart yang sudah di tetapkan oleh Badan Pengawas Obat dan Makanan (BPOM) dan sudah melalui berbagai macam pengujian. Kendala utama para ahli medis dalam pembuatan obat-obatan ini adalah, tingginya waktu dan biaya yang dibutuhkan untuk melakukan suatu penelitian dalam pembuatan obat tertentu sehingga dapat menghambat kinerja para ahli medis, selain itu banyaknya jenis senyawa aktif dapat menimbulkan terjadinya kesalahan saat proses klasifikasi. Oleh karena itu, proses klasifikasi fungsi senyawa aktif perlu dilakukan untuk mempermudah para ahli medis, sehingga waktu yang dibutuhkan menjadi lebih singkat dan hasil yang didapatkan dalam klasifikasi fungsi senyawa aktif lebih akurat.

Klasifikasi dilakukan untuk mengelompokan suatu objek berdasarkan atribut-atributnya (Susilowati, Sabariah, & Gozali, 2015). Dengan memanfaatkan perkembangan teknologi, proses klasifikasi suatu senyawa aktif dapat menggunakan notasi *Simplified Molecular Input Line Entry System* (SMILES). Penggunaan notasi SMILES dapat memungkinkan suatu senyawa untuk diproses oleh komputer (Junaedi, 2011).

Proses klasifikasi dapat dilakukan dengan menggunakan metode *Backpropagation* yang terdapat pada Jaringan Syaraf Tiruan. Terdapat beberapa penelitian yang sudah dilakukan sebelumnya dengan menggunakan algoritma *Backpropagation*. Penelitian pertama dilakukan pada tahun 2016 dengan menggunakan objek berupa citra daun untuk melakukan klasifikasi varietas cabai. Penelitian ini menghasilkan nilai rata-rata akurasi sebesar 97,92 % pada pengujian *K-Fold cross validation* dengan menggunakan parameter $K=3$, *learning rate* 0,2, *neuron hidden layer* sebanyak 5 dan *max epoch* sebanyak 100.000 (Syaban & Harjoko, 2016). Penelitian kedua dilakukan pada tahun 2016 dengan menggunakan objek berupa citra sidik jari untuk melakukan klasifikasi pola sidik jari yang digunakan untuk Analisa karakteristik seseorang. Penelitian ini menghasilkan nilai akurasi sebesar 83 % dengan menggunakan

parameter *max epoch* sebesar 1000, *learning rate* sebesar 0,5 dan *neuron hidden layer* sebanyak 160 (Setiawan & Agung, 2016). Berdasarkan kedua penelitian sebelumnya, nilai *epoch* yang dibutuhkan untuk mencapai konvergensi sistem terlalu banyak, sehingga waktu komputasi yang dibutuhkan menjadi semakin lama. Oleh karena itu, pada penelitian ini algoritma momentum *Backpropagation* akan ditambahkan dengan momentum. Penambahan momentum berguna untuk meningkatkan kecepatan konvergensi sistem, sehingga waktu komputasi yang dibutuhkan menjadi lebih sedikit (Maharani, 2009).

Penelitian sebelumnya dengan menggunakan momentum *Backpropagation* untuk klasifikasi data menghasilkan nilai akurasi 96 %. Pada penelitian ini menunjukkan bahwa konstanta momentum dapat mempercepat kecepatan konvergensi sistem (Maharani, 2009). Berdasarkan uraian diatas, maka penelitian ini akan membahas tentang momentum *Backpropagation* untuk klasifikasi fungsi senyawa aktif berdasarkan notasi SMILES (*Simplified Molecular Line Entry System*).

2. TINJAUAN PUSTAKA

2.1 Preprocessing

Tahap awal dalam *preprocessing* data bertujuan untuk menghilangkan *noise* sehingga data layak untuk digunakan (Andono, Nurtantio, Sutojo, & Muljono, 2017). *Preprocessing* dalam penulisan ini menggunakan *regular expression* (regex) untuk mengambil notasi SMILES. Penggunaan regex sangat membantu dalam pengambilan huruf, angka, kata atau simbol yang diinginkan (Muliantara, 2009).

2.2 SMILES (*Simplified Molecular Input Line Entry System*)

SMILES merupakan notasi kimia yang dirancang untuk pemrosesan kimia modern. Penulisan spesifikasi kimia pada notasi SMILES menggunakan karakter *American Standart code for Information Interchange* (ASCII) untuk merepresentasikan simbol atom dan ikatan yang nantinya akan di simpan ke dalam variabel string (Junaedi, 2011). Penggunaan karakter ASCII pada notasi SMILES memudahkan senyawa aktif dapat di proses oleh komputer dengan menggunakan memori yang lebih sedikit. Contoh penulisan notasi SMILES dapat dilihat

pada Tabel 1.

Tabel 1. Contoh Notasi SMILES

Rumus molekul	Notasi SMILES
C10H18N4O5	C(CC(C(=O)O)NC(=O)CCC(=O)O)CN=C(N)N
C9H9NO3	C1=CC=C(C=C1)C(=O)NCC(=O)O
C9H21N3O	CC(=O)NCCCNCCCCN

2.3 Backpropagation

Backpropagation merupakan algoritma *supervised learning* dan merupakan salah satu algoritma dari jaringan syaraf tiruan. Algoritma ini digunakan untuk melatih jaringan sehingga dapat menghasilkan nilai *error* minimum. *Backpropagation* memiliki 3 layer utama yaitu, *input layer*, *hidden layer*, dan *output layer* (Kusumadewi, 2003). Pada algoritma *Backpropagation* terdapat beberapa fungsi aktivasi, diantaranya (Kusumadewi, 2003).

- Linier/ Purelin
Fungsi linier ini memiliki fungsi untuk membawa nilai *input* ke *output* yang sebanding.
- Tansig
Fungsi tansig merupakan fungsi sigmoid tangen. Fungsi ini akan membawa nilai *input* ke *output* dengan rumus *hyperbolic* tangen sigmoid dengan nilai *output* maksimal 1 dan nilai *output* minimal -1.
- Logsig
Fungsi Logsig merupakan fungsi yang membawa nilai *input* ke *output* dengan perhitungan *log - sigmoid*, dengan nilai *output* antara -1 hingga 1.

2.4 Momentum Backpropagation

Perhitungan momentum *backpropagation* dapat dilihat pada tahap berikut ini (Andrian & Putra, 2014):

1. Inisialisasi bobot awal *V* dan *W* dengan range nilai 0 sampai 1.
2. Menentukan nilai maksimum *epoch*, *learning rate* (α), momentum, dan *target error*.
3. Lakukan langkah ke- 4 sampai ke- 8 selama kondisi $epoch < maksimum\ epoch$ dan $MAPE > Target\ error$.

4. Lakukan langkah ke-6 sampai ke-10 selama proses *training*.
5. Lakukan langkah ke-6 sampai ke-7 untuk proses pengujian.

Perambatan maju (Feedforward)

6. Penjumlahan bobot pada *hidden layer*
 $Z_{in(j)} = V_{0j} + \sum_{i=1}^n X_i \cdot V_{ij}$ (2.1)

Melakukan operasi aktivasi penjumlahan terbobot dengan fungsi aktivasi sigmoid.

$$Z_j = \frac{1}{1+e^{-Z_{in(j)}}}$$
..... (2.2)

7. Menghitung nilai output layer
 $Y_{in(k)} = W_{0k} + \sum_{j=1}^n Z_j \cdot W_{jk}$ (2.3)

Menghitung sinyal *output* dengan fungsi aktivasi:

$$Y_k = Y_{in(k)}$$
..... (2.4)

Perambatan mundur (Backward)

8. Menghitung nilai *error*, berdasarkan nilai *error* dan nilai target.

$$\delta_k = (t_k - Y_k)$$
..... (2.5)

Menghitung nilai koreksi bobot W_{jk} :

$$\Delta W_{jk}(t) = \alpha \delta_k Z_j + \Delta W_{jk}(t - 1)\mu$$
 (2.6)

Menghitung nilai koreksi bias W_{0k} :

$$\Delta W_{0k}(t) = \alpha \delta_k + \Delta W_{0k}(t - 1)\mu$$
 ... (2.7)

9. Penjumlahan delta unit untuk setiap *neuron hidden layer*

$$\delta_{in\ j} = \sum_{k=1}^m \delta_k W_{jk}$$
 (2.8)

Menghitung nilai error:

$$\delta_j = \delta_{in\ j}(Z_j)(1 - Z_j)$$
 (2.9)

Menghitung nilai koreksi bobot V_{ij} :

$$\Delta V_{ij}(t) = \alpha \delta_j X_i + \Delta V_{ij}(t - 1)\mu$$
(2.10)

Menghitung nilai koreksi bobot bias V_{0j} :

$$\Delta V_{0j}(t) = \alpha \delta_j + \Delta V_{0j}(t - 1)\mu$$
(2.11)

10. Perbaiki nilai bobot dan nilai bias
Perbaiki nilai bobot dan nilai bias pada *output layer*.

$$W_{jk}(baru) = W_{jk}(lama) + \Delta W_{jk}$$
 ..(2.12)

$$W_{0k}(baru) = W_{0k}(lama) + \Delta W_{0k}$$
 (2.13)

Perbaiki nilai bobot dan nilai bias pada

hidden layer.

$$V_{ij}(\text{baru}) = V_{ij}(\text{lama}) + \Delta V_{ij} \dots\dots(2.14)$$

$$V_{0j}(\text{baru}) = V_{0j}(\text{lama}) + \Delta V_{0j} \dots\dots(2.15)$$

Keterangan:

V = Bobot input

V_{0j} = Bias pada neuron hidden

W = Bobot neuron pada hidden layer

W_{0k} = Bias pada output neuron

X = Data

Y_k = Output neuron k

Z_j = Neuron hidden j

i = Data ke-

j = Jumlah neuron pada hidden layer

k = Jumlah target

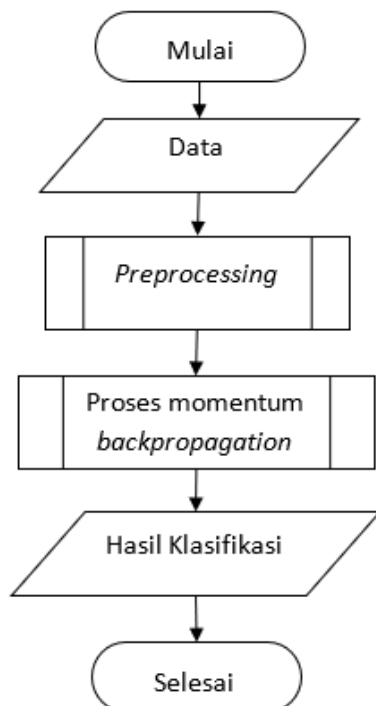
μ = Momentum

2.5 MAPE

Persentase kesalahan rata-rata antara data yang diinginkan dengan data hasil prediksi dapat dihitung dengan menggunakan rumus sebagai berikut:

$$MAPE = \frac{\sum \frac{|X_i - F_i|}{X_i} \times 100\%}{n}$$

3. METODE



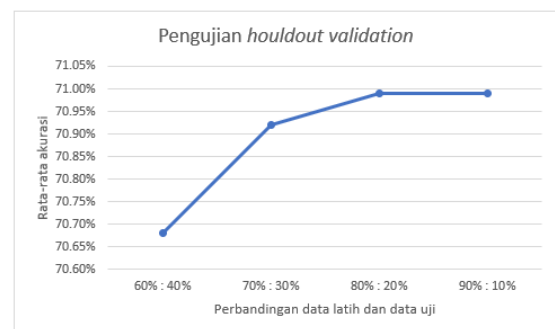
Gambar 1. Diagram Alir Sistem

Penelitian ini menggunakan modifikasi *Backpropagation* dengan menambahkan momentum pada perhitungan delta W , delta bias W , delta V , delta bias V . Alur program dalam penelitian ini dimulai dengan melakukan proses *preprocessing* untuk mengambil 11 fitur yang ada pada semua data. Semua data *training* dan data *testing* diambil dari <https://pubchem.ncbi.nlm.nih.gov>. Fitur yang sudah didapatkan dari proses *preprocessing* akan digunakan sebagai *input* pada proses momentum *Backpropagation* sehingga didapatkan hasil klasifikasi fungsi senyawa aktif. Diagram alir sistem secara umum dapat dilihat pada Gambar 1.

4. PENGUJIAN DAN PEMBAHASAN

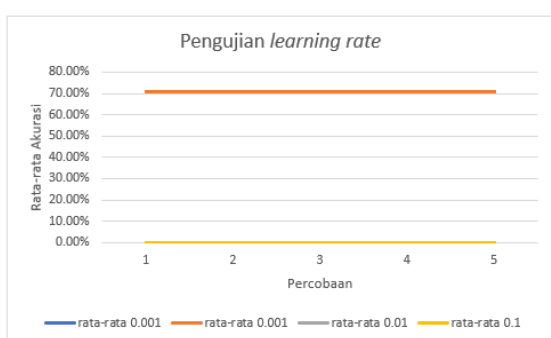
Pengujian dalam penelitian ini ada 4 macam yaitu, pengujian jumlah data latih dan data uji, pengujian *learning rate*, pengujian momentum dan pengujian *K-Fold cross validation*. Pengujian ini menggunakan data sebanyak 653 yang terdiri dari 464 data kelas metabolisme dan 189 kelas kanker.

Pengujian pertama bertujuan untuk mengetahui tingkat akurasi sistem dengan jumlah data latih dan data uji yang berbeda. Proses pengujian ini dilakukan menggunakan perbandingan *holdout validation* dengan menggunakan parameter *learning rate* sebesar 0,0001, *neuron hidden layer* sebanyak 4, momentum sebesar 0,25, dan *max epoch* sebanyak 50. Hasil pengujian menunjukkan nilai akurasi terbaik sebesar 70,99 % pada perbandingan 80 % : 20 % dan 90 % : 10 %. Hal ini menunjukkan bahwa nilai minimum untuk perbandingan data latih dan data uji sebesar 80 % : 20 %. Hasil pengujian perbandingan jumlah data latih dan data uji dapat dilihat pada Gambar 2.



Gambar 2. Hasil Pengujian Holdout Validation

Pengujian kedua bertujuan untuk mengetahui pengaruh learning rate terhadap nilai akurasi sistem. Pengujian dilakukan dengan menggunakan data latih dan data uji pada pengujian sebelumnya dengan *neuron hidden layer* sebanyak 4, momentum sebesar 0,25, dan *max epoch* sebanyak 100. Hasil pengujian menunjukkan bahwa *learning rate* berpengaruh terhadap nilai akurasi. Hal ini dikarenakan semakin besar nilai *learning rate* yang digunakan, maka akan semakin besar pula nilai *error* yang dihasilkan sehingga *neural network* tidak bisa mencapai hasil yang konvergen. Hasil pengujian *learning rate* dapat dilihat pada Gambar 3.

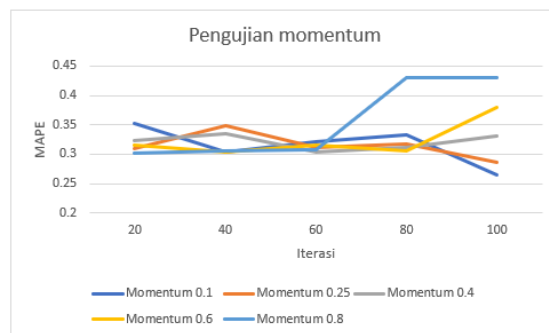


Gambar 3. Hasil Pengujian Learning Rate

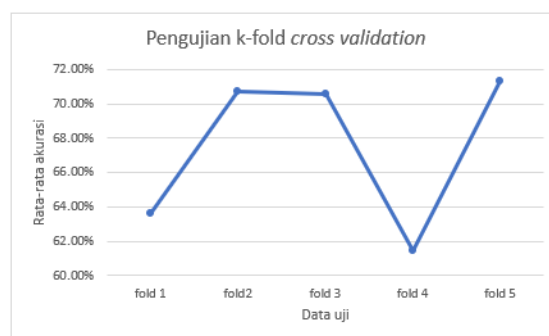
Pengujian ke tiga bertujuan untuk mengetahui pengaruh momentum terhadap nilai MAPE. Pengujian dilakukan dengan menggunakan data latih dan data uji pada pengujian sebelumnya dengan *neuron hidden layer* sebanyak 4 dan *learning rate* sebesar 0,00001. Hasil pengujian nilai momentum menghasilkan nilai MAPE yang bervariasi untuk setiap iterasi. Hasil penelitian menunjukkan jika nilai momentum kecil maka, jumlah iterasi yang dibutuhkan semakin banyak tetapi nilai kesalahan yang dihasilkan semakin kecil, sedangkan jika menggunakan nilai momentum besar walaupun menggunakan iterasi besar nilai kesalahannya tidak mengecil. Hasil pengujian momentum dapat dilihat pada Gambar 4.

Pengujian ke empat bertujuan untuk mengetahui tingkat akurasi sistem dengan jumlah data yang sama. Pada pengujian ini, nilai *K* yang digunakan sebanyak 5 dan data akan dibagi menjadi 5 kelompok data. Setiap kelompok data memiliki jumlah yang sama. Hasil pengujian *K-Fold cross validation* bervariasi. Rata-rata akurasi sebesar 67,32%, sedangkan hasil akurasi tertinggi sebesar 71,32%. Hal ini dikarenakan jumlah data pada

kelas metabolisme lebih banyak dari pada jumlah data pada kelas kanker, sehingga metode momentum *Backpropagation* tidak dapat melakukan pembelajaran dengan baik terhadap pola kelas kanker. Hasil pengujian *K-Fold cross validation* dapat dilihat pada Gambar 5.



Gambar 4. Hasil Pengujian Momentum



Gambar 5. Hasil Pengujian K-Fold Cross Validation

5. KESIMPULAN

Berdasarkan hasil pengujian yang telah dilakukan. Penerapan metode momentum *Backpropagation* untuk klasifikasi fungsi senyawa aktif berdasarkan notasi SMILES (*Simplified Molecular Input Line Entry System*) dapat dilakukan dengan melakukan tahap *preprocessing* terlebih dahulu untuk mendapatkan 11 fitur yang terdiri dari atom B ,C ,N ,O ,P ,S ,F ,Cl ,Br ,I ,OH. Nilai akurasi tertinggi yang di dapatkan menggunakan metode momentum *Backpropagation* sebesar 70.99% dengan menggunakan parameter *learning rate* sebesar 0,00001, *max epoch* sebesar 100, *neuron hidden layer* sebesar 4, dan momentum sebesar 0,25. Hasil klasifikasi lebih banyak mengenali kelas metabolisme dari pada kelas kanker. Hal ini terjadi karena jumlah data yang di gunakan untuk kelas metabolisme lebih banyak dari pada

kelas kanker.

6. DAFTAR PUSTAKA

- Andono, Nurtantio, P., Sutojo, & Muljono. (2017). *Pengolahan citra digital*. Semarang: Andi.
- Andrian, Y., & Putra, P. H. (2014). Analisis penambahan momentum pada proses prediksi curah hujan kota medan menggunakan metode backpropagation neural network. *Seminar nasional informatika*, 165-172.
- Junaedi, H. (2011). Penggambaran rantai karbon dengan menggunakan simplified molecular input line system(SMILES). *IdeaTech*, 219-226.
- Kusumadewi, S. (2003). *Artificial intelligence*. Yogyakarta: Graha ilmu.
- Maharani, W. (2009). Klasifikasi menggunakan jst backpropagation momentum dengan adaptive learning rate. *Seminar nasional informatika*, 25-31.
- Muliantara. (2009). *Penerapan regular expression dalam melindungi alamat email dari spam robot pada konten wordpress*. Denpasar: Universitas Udayana.
- Setiawan, A. F., & Agung, A. K. (2016). Klasifikasi pola sidik jari menggunakan jaringan syaraf tiruan backpropagation untuk analisa karakteristik seseorang. *Jurnal Ilmiah dan Teknik Informatika*, 50-55.
- Sumardjo, D. (2006). *Pengantar kimia : buku panduan kuliah mahasiswa kedokteran dan program strata I fakultas bioeksakta*. Jakarta: EGC.
- Susilowati, E., Sabariah, M. K., & Gozali, A. A. (2015). Implementasi Metode Support Vector Machine untk Melakukan Klasifikasi Kemacetan Lalu Lintas Pada Twitter. *e-Proceeding of Engineering*, 1478.
- Syaban, K., & Harjoko, A. (2016). Klasifikasi varietas cabai berdasarkan morfologi daun menggunakan backpropagation neural network. *IJCCS*, 161-172.