

Klasifikasi Fungsi Senyawa Aktif berdasarkan Data *Simplified Molecular Input Line Entry System* (SMILES) menggunakan Metode *Support Vector Machine* (SVM)

Dwi Febry Indarwati¹, 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: ¹dwifeebb@gmail.com, ²dian_ilkom@ub.ac.id, ³syaiful@ub.ac.id

Abstrak

Senyawa dapat dibedakan sebagai senyawa aktif atau yang biasa disebut senyawa bioaktif dan senyawa tidak aktif atau yang biasa disebut senyawa pasif. Saat ini sebagian senyawa aktif sudah diketahui peran farmakologinya namun masih banyak pula yang belum diketahui peranannya atau masih dalam tahap penelitian, sehingga dibuatnya sistem untuk mengelompokkan fungsi senyawa aktif diharapkan dapat menunjang penelitian yang dilakukan oleh kimiawan di laboratorium. Metode *Support Vector Machine* (SVM) digunakan untuk mengelompokkan fungsi senyawa aktif karena SVM memiliki kemampuan generalisasi yang tinggi tanpa membutuhkan *dataset* tambahan. Untuk memudahkan proses komputasi pada sistem akan digunakan data *Simplified Molecular Input Line Entry System* (SMILES), dimana data SMILES menggambarkan senyawa kimia dalam notasi baris, dari data SMILES tersebut nantinya akan digunakan 15 fitur. Penelitian dilakukan dengan menggunakan objek sebanyak 3 kelas fungsi senyawa aktif, diantaranya adalah metabolisme, infeksi, dan anti radang. Hasil dari pengujian terbaik ketika menggunakan kernel Gaussian RBF, menggunakan nilai λ sebesar 5, nilai *complexity* sebesar 0,1, nilai σ sebesar 0,5, dan dengan jumlah iterasi sebanyak 5 mendapatkan akurasi sebesar 83,33%.

Kata kunci: Senyawa Aktif, SMILES, *Support Vector Machine* (SVM).

Abstract

Chemical compounds can be distinguished into active compounds or commonly called bioactive compounds and inactive compounds or commonly called passive compounds. At this time there are still many active compounds that the pharmacological role does not known yet, so the system being made for classify the functions of active compounds that expected to support chemists research in the laboratory. To simplify the process of making the system, the representation of molecular structure must be easily processed by a computer so that the SMILES notation will be used, the SMILES notation describes chemical formula in a row notation. This system is using the SVM (Support Vector Machine) method because the SVM method has high generalization capabilities without requiring additional datasets. In this research uses as many as 15 features and objects as many as 3 classes of active compound functions, including metabolism, infection, and anti-inflammation. The best test result is 83.33% when using the Gaussian kernel RBF, using a lambda value (λ) of 5, the complexity value is 0.1, the sigma value (σ) is 0.5, and with the number of iterations is 5.

Keywords: Active compound, SMILES, *Support Vector Machine* (SVM).

1. PENDAHULUAN

Senyawa merupakan zat tunggal yang terdiri dari minimal dua atau beberapa unsur berbeda yang saling mengikat. Senyawa dapat dibedakan menjadi dua, yaitu senyawa aktif atau biasa disebut bioaktif dan senyawa tidak aktif atau

biasa disebut senyawa pasif. Senyawa aktif dapat mencakup terapeutik dan farmakologi. Farmakologi merupakan ilmu mengenai cara mencampur dan membuat obat yang digunakan untuk pencegahan perawatan dan pengobatan penyakit (Noviani, 2017).

Saat ini sebagian senyawa aktif sudah

diketahui peran farmakologinya namun masih banyak pula yang belum diketahui peranannya dan masih dalam tahap penelitian. Indonesia sendiri termasuk negara yang memiliki ragam flora yang melimpah dimana sebagian besarnya memiliki potensi untuk dimanfaatkan sebagai tanaman obat-obatan. Untuk mengetahui peran farmakologi pada senyawa aktif tersebut dilakukan suatu penelitian di laboratorium oleh kimiawan dimana membutuhkan banyak biaya dan waktu yang cukup lama seperti ekstraksi yang dapat dilakukan dengan metode maserasi, perkolasi, reflux, destilasi uap, dll (Mukhriani, 2014). Sehingga dibutuhkan sebuah sistem yang dapat digunakan untuk mengelompokkan fungsi senyawa aktif sehingga dapat menunjang penelitian yang ada di laboratorium.

Untuk mempermudah proses pembuatan sistem agar dapat mengklasifikasi fungsi senyawa aktif penggambaran struktur molekul harus dapat mudah diolah oleh komputer, sehingga akan digunakan data SMILES. SMILES menggambarkan struktur molekul pada senyawa kimia dalam notasi baris sehingga dapat dikatakan bahwa data SMILES dirancang khusus untuk diolah oleh komputer. Keunggulan lain dari data SMILES diantaranya adalah tidak sulit diproses dan tidak membutuhkan ruang penyimpanan memori yang besar (Weininger, 1988).

Penelitian mengenai SMILES juga sudah dilakukan. Diantaranya dengan menggunakan algoritme C4.5 menghasilkan nilai akurasi sebesar 79.34% dengan menggunakan objek kelas sebanyak 2 kelas fungsi senyawa aktif (Rochman, 2018). Pada penelitian-penelitian yang telah dilakukan sebelumnya data SMILES dibagi menjadi 11 fitur yaitu 'B', 'C', 'N', 'O', 'P', 'S', 'F', 'Cl', 'Br', 'I', 'OH' dan panjang dari data SMILES itu sendiri. Sedangkan pada penelitian yang akan dilakukan adalah dengan menambahkan 4 fitur, yaitu '@', '=', '#', dan '()' sehingga fitur yang digunakan menjadi 15 fitur. Diharapkan dengan ditambahnya jumlah fitur dapat memperoleh nilai akurasi yang lebih optimal.

Pada penelitian ini, metode yang akan digunakan untuk melakukan klasifikasi terhadap fungsi senyawa aktif adalah metode SVM karena SVM memiliki kemampuan generalisasi yang tinggi tanpa memerlukan dataset tambahan, bahkan dengan dimensi yang tinggi dari input space (Burges, 1998). Sehingga perkembangan pada metode SVM memberikan dorongan pada penelitian yang berada di bidang pengenalan

pola untuk mengembangkan potensi SVM baik secara teoritis maupun aplikasi. Salah satu contoh penelitian yang telah dilakukan yaitu untuk mengelompokkan penyakit gigi dan mulut dengan menggunakan kernel *Gaussian* RBF mendapatkan rata akurasi tertinggi adalah 93,32% dengan *dataset* yang digunakan sejumlah 122 data (Puspitasari, 2018).

2. LANDASAN KEPUSTAKAAN

2.1 SMILES

SMILES adalah penggambaran notasi kimia modern dalam bentuk notasi baris yang dikenalkan di tahun 1980, dimana SMILES didesain dengan menggunakan sebuah konsep dari graph (Weininger, 1988). Dalam menuliskan data SMILES harus benar-benar memperhatikan huruf kapital dan non-kapitalnya karena penulisan SMILES bersifat *case sensitive*.

2.2 Klasifikasi

Klasifikasi merupakan teknik yang digunakan untuk melakukan prediksi kelas dari setiap data itu sendiri, dalam setiap data tentunya akan memiliki kelas dan atribut, sehingga dilakukannya klasifikasi untuk dapat menentukan atau mengetahui *class* dan atribut dari suatu data yang sebelumnya belum diketahui. Metode yang digunakan dalam klasifikasi sangat beragam, mulai dari C45, k-NN, naïve bayes, ataupun *Support Vector Machine* (Subiyakto, 2008).

2.3 Preprocessing

Teknik *Preprocessing* digunakan untuk merubah struktur data yang terdapat pada SMILES agar sesuai dengan kebutuhan fitur yang digunakan. Sedangkan *regular expression* (regex) merupakan konstruksi bahasa yang dapat dipakai untuk mengenali sebuah pola atau *pattern* dalam pencarian teks. Sehingga regex digunakan untuk mengolah data SMILES agar dapat dikomputasi oleh sistem seperti panjang senyawa data SMILES dan jumlah dari karakter yang ada pada masing-masing fitur yang telah ditentukan sebelumnya. Hasil akhir dari proses *preprocessing* berupa jumlah tiap fitur yang dibagi dengan panjang senyawa notasi itu sendiri.

2.4 SVM (Support Vector Machine)

Pada dasarnya metode SVM hanya digunakan untuk model linier, namun saat ini SVM juga dapat diimplementasikan pada model yang bersifat non-linier karena saat ini permasalahan yang ada lebih banyak yang bersifat non-linier. Permasalahan yang bersifat non-linier dapat dipecahkan dengan modifikasi pada metode SVM dengan menggunakan fungsi kernel di dalamnya sehingga akan sangat mempermudah dalam menentukan nilai support vector. Nilai akurasi pada model yang dihasilkan menggunakan metode SVM ini sangat dipengaruhi oleh fungsi kernel yang digunakan dan nilai-nilai parameter yang dipakai (Siagian, 2011).

2.5 Sequential Training Pada SVM

Sequential Training merupakan algoritme yang sederhana tanpa menghabiskan waktu yang banyak sehingga dapat meminimalkan waktu ketika mencari *hyperplane* yang terbaik dalam SVM yang digunakan pada proses pelatihan. *Sequential Training* merupakan algoritme yang sederhana Metode *Sequential Training* yang dikembangkan oleh Vijayakumar akan ditunjukkan pada proses sebagai berikut (Vijayakumar, 1999):

- 1) Memberikan nilai awal pada parameter alpha $\alpha_1 = 0$ yang akan digunakan untuk mendapatkan nilai *support vector*. Kemudian menghitung nilai Matriks *Hessian* dengan Persamaan 1.

$$D_{ij} = y_i y_j (K(x_i, x_j) + \lambda^2) \tag{1}$$

Dimana i dan $j = 1, 2, 3, \dots, n$

Keterangan:

D_{ij} = matriks *Hessian*

$y_i y_j$ = kelas level data ke- i dan data ke- j

λ = *lambda*

Langkah selanjutnya yang dilakukan adalah melakukan perhitungan variabel γ (*gamma*) atau *learning rate* dengan persamaan 2.

$$\gamma = \frac{0,01}{\max D_{ij}} \tag{2}$$

Keterangan:

γ = *gamma*

$\max D_{ij}$ = nilai tertinggi yang ada pada matriks hessian

- 2) Mengulangi Persamaan 3 dan Persamaan 4 hingga mencapai nilai iterasi yang telah ditentukan.

- a) Nilai *error rate* dihitung dengan menggunakan Persamaan 3:

$$E_i = \sum_{j=1}^i \alpha_i D_{ij} \tag{3}$$

Keterangan:

E_i = *error rate*

α_i = *alpha*

- b) Mencari selisih *alpha* atau $\delta \alpha_i$ dengan memakai Persamaan 4:

$$\delta \alpha_i = \min\{\max[\gamma(1 - E_i), -\alpha_i], C - \alpha_i\} \tag{4}$$

Keterangan:

C = *Complexity*

$\delta \alpha_i$ = *Selisih nilai alpha*

- c) Meng-*update* nilai *alpha* (α_i) dengan Persamaan 5:

$$\alpha_i = \alpha_i + \delta \alpha_i \tag{5}$$

- 3) Setelah mendapatkan nilai α_i , langkah yang dilakukan selanjutnya adalah menghitung kernel positif dan kernel negatif dengan Persamaan 6 dan Persamaan 7:

$$wx^+ = \sum_{i=1}^n \alpha_i y_i K(x_i, x^+) \tag{6}$$

$$wx^- = \sum_{i=1}^n \alpha_i y_i K(x_i, y_i^-) \tag{7}$$

Keterangan:

wx^+ = kernel positif y_i = kelas level

wx^- = kernel negatif

Tahap terakhir yang dilakukan adalah dengan menghitung nilai bias dengan Persamaan 8:

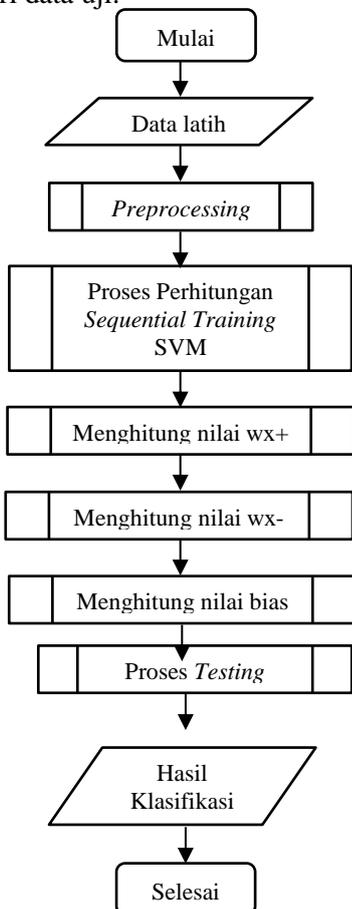
$$bias = -\frac{1}{2} \sum (wx^+ + wx^-) \tag{8}$$

3. METODOLOGI PENELITIAN

Pada bagian ini akan dibahas mengenai tahapan kerja yang akan dilaksanakan untuk membangun metode SVM pada pengelompokan fungsil senyawa aktif dengan data SMILES. Pada gambar 1 menunjukkan alur terhadap proses klasifikasi secara umum.

Penjabaran mengenai alur pada Gambar 1 diantaranya adalah hal pertama yang dilakukan yaitu memasukkan data latih yang berupa kode SMILES, agar dapat diproses oleh sistem data SMILES akan di-*preprocessing* untuk mendapatkan nilai fitur yang sesuai. Kemudian digunakannya *sequential training* untuk melakukan Perhitungan pada data latih, Kemudian dilanjutkan dengan Proses menghitung nilai wx^+ (kernel positif) dan wx^- (kernel negatif) lalu proses menghitung nilai bias. Tahap terakhir yang dilakukan adalah

proses perhitungan pengujian pada SVM untuk memutuskan *class* prediksi dari data uji dengan memakai proses *One-Against-All*. Hasil dari klasifikasi yang dikeluarkan berupa *class* prediksi dari data uji.



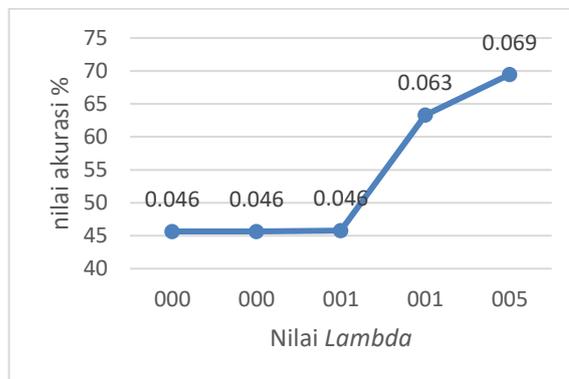
Gambar 1 Diagram alir proses algoritme SVM

4. HASIL PENGUJIAN

Pengujian yang akan digunakan pada penelitian yang dilakukan guna menguji sistem diantaranya adalah pengujian terhadap parameter *lambda*, *complexity*, *sigma*, jumlah *itermax*, serta pengujian terhadap jenis kernel dan terhadap jumlah kelas. Keseluruhan data yang dipakai yaitu sejumlah 818, dari data tersebut diantaranya 656 merupakan data latih dan 162 merupakan data uji. Parameter yang digunakan untuk pengujian di sistem ini diantaranya adalah $\alpha = 0$, $\lambda=5$, $\sigma = 1$, iterasi = 2, $C = 0,05$, k-fold = 5, dengan pemakaian perbandingan rasio data latih yaitu 80% sedangkan untuk data uji sebesar 20%, sehingga pengujian dilakukan dengan mengganti nilai parameter yang sedang diuji.

4.1 Pengujian Terhadap Nilai *Lambda* (λ)

Hasil pengujian terhadap nilai *lambda* ditunjukkan pada Gambar 2.

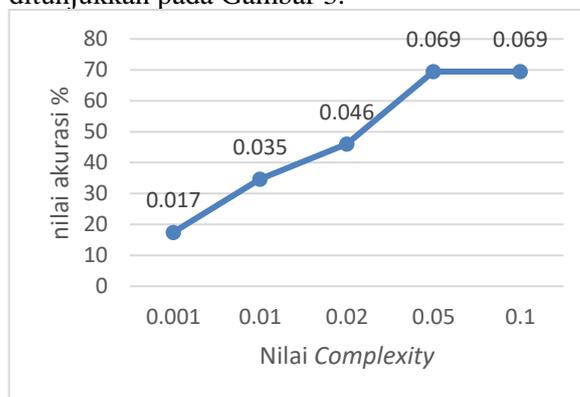


Gambar 2 Hasil Pengujian terhadap nilai *lambda* (λ)

Bersumber pada grafik dari hasil pengujian pada parameter *lambda* diperoleh nilai akurasi tertinggi sebesar 83,3% dengan rata-rata akurasi sebesar 69,461279% saat menggunakan nilai *lambda* sebesar 5, Kesimpulan yang bisa ditarik berdasarkan grafik analisis pengujian terhadap nilai *lambda* adalah semakin besarnya nilai *lambda* maka dapat membuat nilai akurasi menjadi lebih baik. Namun penggunaan *lambda* yang berlebihan besarnya juga bisa memperpanjang waktu ketika kalkulasi matriks yang diakibatkan oleh *augmented factor* dimana dapat menyebabkan sistem melambat untuk menggapai konvergensi (Vijayakumar, et al., 1999). Rata-rata nilai akurasi mulai konvergen ketika menggunakan nilai *lambda* sebesar 5, oleh karena itu untuk pengujian yang akan dilakukan selanjutnya dipakai nilai *lambda* sebesar 5.

4.2 Pengujian Terhadap Nilai *Complexity* (C)

Hasil pengujian terhadap nilai *complexity* ditunjukkan pada Gambar 3.



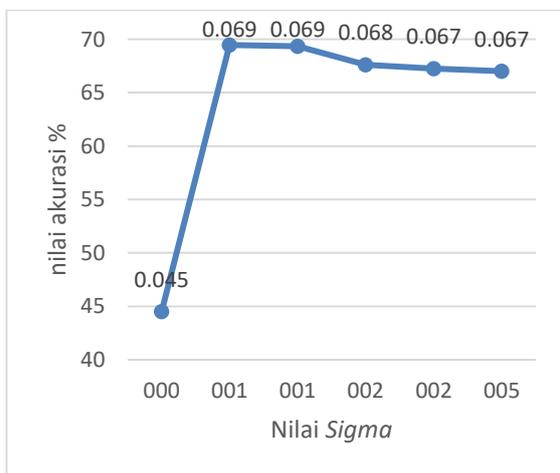
Gambar 3 Hasil pengujian terhadap nilai *Complexity* (C)

Berdasarkan grafik pada Gambar 3 didapatkan nilai akurasi tertinggi adalah 83,3% dengan rata-rata nilai akurasi sebesar 69,461% saat menggunakan nilai *complexity* sebesar 0,05. Nilai *complexity* digunakan untuk mengurangi variabel *error* pada tahap pelatihan ketika menghitung nilai *w*, oleh karena itu sangat diperlukannya untuk mengoptimalkan *margin* dan meminimalkan jumlah *slack* (Puspitasari, 2018). Rata-rata akurasi mulai konvergen ketika menggunakan nilai *complexity* 0,05, kemudian pada pengujian setelahnya akan dipakai parameter *complexity* sebesar 0,05.

Penggunaan nilai *C* yang semakin besar juga akan berdampak pada nilai *error*. Namun jika nilai *C* berdekatan dengan angka 0, maka luas dari *margin* yang ada di *hyperplane* akan menjadi maksimum kemudian disaat bersamaan data yang sedang dilatih baik ada di dalam *margin* atau berada di posisi yang salah tidak akan diperdulikan, perihal tersebut tentunya akan menurunkan nilai akurasi pada tahap pelatihan yang akan berdampak pada tahap pengujian.

4.3 Pengujian Terhadap Nilai Sigma (σ)

Hasil pengujian terhadap nilai sigma ditunjukkan pada Gambar 4.



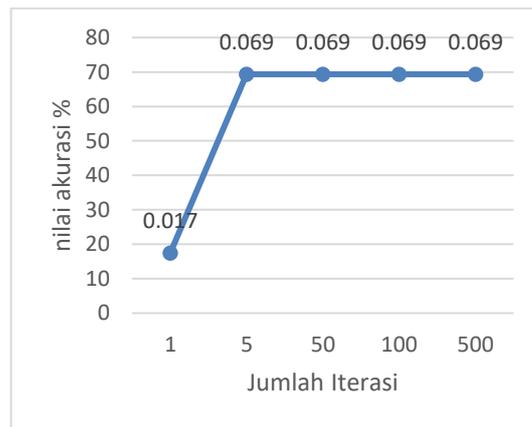
Gambar 4 Hasil pengujian terhadap nilai sigma (σ)

Berdasarkan grafik yang dapat dilihat pada Gambar 4 diperoleh nilai akurasi tertinggi adalah 83,3% dengan rata-rata nilai akurasi sebesar 69,461% saat menggunakan nilai sigma sebesar 0,5. Parameter *sigma* memiliki fungsi mencari nilai optimal pada setiap *dataset*, nilai *sigma* juga berpengaruh dalam pengelompokan dari data yang dibentuk sehingga penggunaan

parameter sigma yang sangat berlebihan jumlahnya akan menimbulkan risiko kernel yang *overfitting*. Berdasarkan dari hasil pengujian terhadap nilai *sigma*, maka nilai sigma 0,5 yang akan dipakai selanjutnya.

4.4 Pengujian Terhadap Jumlah Iterasi

Hasil pengujian terhadap nilai sigma ditunjukkan pada Gambar 5.

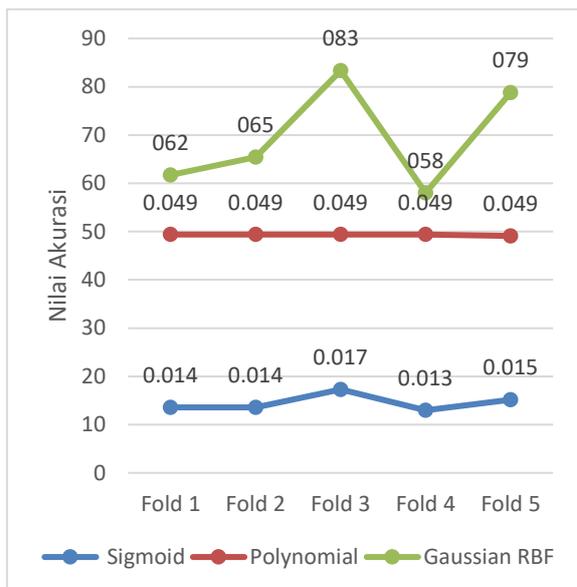


Gambar 5 Hasil pengujian terhadap jumlah iterasi

Berdasarkan Grafik hasil uji pada Gambar 5 di peroleh nilai akurasi tertinggi adalah 83,3% dengan rata-rata akurasi sebesar 69,337% saat menggunakan jumlah iterasi sebanyak 5. Hasil akurasi naik pada saat jumlah iterasi=5. Banyaknya jumlah iterasi sangat mempengaruhi kinerja nilai α , iterasi sendiri akan berhenti ketika telah mencapai *itermax*. Rata-rata akurasi mulai konvergen ketika menggunakan jumlah iterasi sebanyak 5. Penambahan jumlah iterasi juga dapat membentuk *support vector* berlangsung secara proporsional dan setiap fitur data mempunyai lokasi yang dekat dengan *hyperplane* (Puspitasari, et al., 2018).

4.5 Pengujian Terhadap Jenis Kernel

Pengujian terhadap jenis kernel dilakukan dengan menggunakan kernel *sigmoid* (*tangen hiperbolik*), *polynomial*, dan *Gaussian RBF*. Hasil pengujian terhadap jenis kernel ditunjukkan pada Gambar 6.

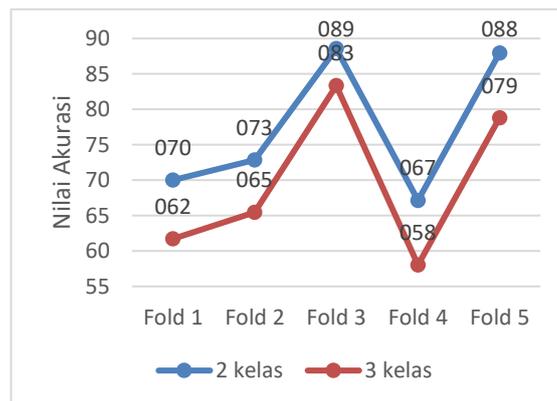


Gambar 6 Hasil pengujian terhadap jenis kernel

Dari grafik yang dapat dilihat pada Gambar 6 ditunjukkan bahwa kernel *Gaussian RBF* memiliki nilai rata-rata tertinggi sedangkan kernel *sigmoid* memiliki rata-rata terendah, sehingga kernel *Gaussian RBF* nantinya akan digunakan dalam kode program klasifikasi fungsi senyawa aktif. Salah satu penyebab kernel *Gaussian RBF* dapat memperoleh tingkat akurasi lebih optimal karena kernel *Gaussian RBF* memiliki lebih sedikit jumlah *hiperparameter* dari pada kernel *polynomial* dan *sigmoid*, karena kompleksitas pemilihan model sangat dipengaruhi oleh jumlah *hiperparameter* (Hsu, et al., 2016).

4.6 Pengujian Terhadap Jumlah Kelas

Jumlah kelas yang di gunakan adalah 2 kelas dan 3 kelas. Untuk jumlah kelas sebanyak 2, kelas yang dipakai adalah metabolisme dan infeksi, sedangkan untuk jumlah kelas sebanyak 3 kelas yang dipakai ditambahkan dengan kelas anti radang. Hasil pengujian terhadap jumlah kelas ditunjukkan pada Gambar 7.



Gambar 7 Hasil pengujian terhadap jumlah kelas

Grafik yang terdapat pada Gambar 7 memperlihatkan kalau penggunaan kelas berjumlah 2 akan mendapatkan nilai rata-rata akurasi lebih tinggi, hal ini disebabkan karena penggunaan kelas berjumlah 2 lebih sederhana pemodelannya sehingga akan mempermudah pada tahap baik pelatihan maupun pengujian. Untuk penggunaan kelas berjumlah 3 akan lebih rumit daripada yang berjumlah di bawahnya karena jumlah model yang digunakan lebih banyak sehingga pada tahap pelatihan sistem harus dapat dengan tepat memodelkan masing-masing kelas yang digunakan agar menampilkan hasil akurasi yang baik.

5. PENUTUP

Kesimpulan yang dapat di peroleh dari penelitian menggunakan metode SVM untuk pengelompokkan fungsi senyawa aktif adalah:

1. Siasat untuk mengolah data SMILES agar dapat di pakai di proses klasifikasi fungsi senyawa aktif yaitu dengan dilakukannya *preprocessing*. *Preprocessing* yang dilakukan adalah dengan menggunakan *regular expression (regex)*, *regex* digunakan untuk mencari jumlah dari setiap fitur yang terdapat dalam data SMILES dan menghitung jumlah panjang dari data SMILES. Fitur yang digunakan diantaranya adalah atom B, C, N, O, P, S, F, Cl, Br, I, OH, @, =, #, () dan panjang dari data SMILES itu sendiri. Ketika sudah didapatkan jumlah dari setiap fitur, jumlah dari tiap fitur tersebut dibagi dengan panjang data SMILES dan hasil dari pembagian tersebut yang digunakan untuk komputasi sistem klasifikasi fungsi senyawa aktif.

2. Cara kerja metode SVM yaitu melakukan tahap pelatihan terlebih dahulu. Tahap pelatihan diawali dengan proses perhitungan kernel *Gaussian RBF*. Kemudian kernel *Gaussian RBF* digunakan untuk dilakukan komputasi matriks *Hessian*. Lalu akan dilakukan tahapan *sequential training* yang akan mendapatkan nilai *alpha* terbaru yang akan digunakan di proses selanjutnya yaitu menghitung kernel wx^+ dan wx^- . Setelah mendapatkan kernel wx^+ dan wx^- , proses perhitungan *bias* akan dilakukan. Lalu tahap yang dilakukan setelah pelatihan adalah pengujian pada SVM yang disertai dengan hasil keputusan untuk prediksi class dari data uji.
3. Nilai akurasi tertinggi yang diperoleh dari penelitian ini didapatkan sebesar 83,33333333% dengan rata-rata akurasi pada semua percobaan sebesar 69,46127946%. Setelah dilakukan beragam percobaan untuk pengujian di sistem, parameter-parameter yang mendapatkan hasil yang paling optimal diantaranya adalah dengan menggunakan kernel *Gaussian RBF*, jumlah objek kelas yang dipakai adalah 3, penggunaan nilai *lambda* sebesar 5, nilai *complexity* sebesar 0,05, nilai *sigma* sebesar 0,5, dan menggunakan total iterasi sebanyak 5.

Untuk pembangunan pada penelitian yang akan di lakukan selanjutnya, saran yang bisa dikutip melalui eksperimen ini yaitu untuk meningkatkan hasil akurasi yang ada di sistem adalah dengan meningkatkan jumlah data latih. Untuk saran selanjutnya agar lebih sempurna lagi yaitu menambahkan fitur yang lebih banyak.

6. DAFTAR PUSTAKA

- Hsu, C.-W., Chang, C.-C. & Lin, C.-J., 2016. A Practical Guide to Support Vector Classification. *Department of Computer Science National Taiwan University*.
- Noviani, N. dan Nurilawari, V. 2017. *Farmakologi*. Jakarta Selatan: Pusat Pendidikan Sumber Daya Manusia Kesehatan.
- Puspitasari, A. M., Ratnawati, D. E. dan Widodo, A. W. 2018. *Klasifikasi Penyakit*

Gigi Dan Mulut Menggunakan Metode Support Vector Machine. S1. Universitas Brawijaya.

- Ratnawati, D. E., Marjono, M., Anam, S. 2018. *Prediction of Active Compounds from SMILES Codes Using Backpropagation Algorithm*. AIP Conference Proceedings 2021, 060009 (2018); doi: 10.1063/1.5062773.
- Siagian, R. Y. 2011. *Klasifikasi Parket Kayu Jati Menggunakan Metode Support Vector Machines (SVM)*. Jawa Barat: Universitas Gunadarma.
- Subiyakto, A. 2008. *Penggunaan Algoritma Klasifikasi dalam Data Mining*. Jakarta: Universitas Islam Negeri Jakarta.
- Vijayakumar, S., & Wu, S. 1999. *Sequential Support Vector Classifiers and Regression*. Japan: RIKEN Brain Science Institute.
- Weininger, David. 1988. *SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules*. California: Pomona College.